筑　波　大　学

博 士 （医 学） 学 位 論 文

Analysis of X Chromosome Reactivation during Reprogramming

(リプログラミング時のＸ染色体再活性化の解析)

2018

筑波大学大学院博士課程人間総合科学研究科

TRAN THI HAI YEN

Contents

Chapter I. General overview

1.1 Background and Purpose

Induced pluripotent stem cells (iPSCs) are applicable tools for modeling of diseases, drug development and transplantation medicine. However, several issues regarding to its generation will limit the potential in medicine, such as a low number of high quality iPSCs. Previous studies have shown that the epigenetic state of X chromosomes in female iPSCs is closely linked to pluripotency, an indicator for high quality of iPSCs. Female mouse iPSCs with two active X chromosomes (XaXa) exhibit higher pluripotency than iPSCs with one active and one inactive X chromosome (XaXi). In order to visualize the X chromosome status in live cells and analyze the mechanism of X chromosome reactivation (XCR), I established a novel live cell imaging system of XCR.

1.2 Materials and Methods

I used the CRISPR/Cas9 system to generate mouse embryonic stem cell (mESC) lines that carry two different fluorescent protein genes (EGFP and humanized Kusabira Orange (hKO)) on both X chromosomes. Female mESCs were transfected with two different donor template plasmids, each of which contains one of the fluorescent protein genes and a drug-resistant gene between the homologous sequences in the targeted site, together with the expression vector of Cas9 and a guide RNA. After selection of drug-resistant mESC clones that expressed two fluorescence proteins ($EGFP^+/hKO^+$), PCR analysis of their genome DNA was performed to confirm that these mESC clones have the expected inserts at the targeted sites of both X chromosomes. In order to visualize X chromosome status in live cells, candidate mESC clones were differentiated through embryoid bodies (EBs) formation and their derived differentiated single-color cells were reprogrammed to generate iPSCs to monitor X chromosome inactivation (XCI) and X chromosome reactivation (XCR), respectively.

1.3 Results

In this study, I observed that the female $EGFP^+/hKO^+$ ESC clones express only one of the two fluorescent proteins ($EGFP^+$ or hKO^+) upon differentiation, indicating that the inserted fluorescent protein genes are subject to XCI. Furthermore, when the single-color somatic cells were reprogrammed into iPSCs, the iPS colonies displayed double colors ($EGFP^+/hKO^+$) (“double” signal). These results indicate that my system can also detect XCR during reprogramming in a predicted manner. Interestingly, I found out a correlation between

the extent of XCR and the level of pluripotency of iPSCs. Colonies with complete XCR expressed higher levels of pluripotency marker genes than those with partial XCR. It indicates that my system provides a simple method for distinguishing high and low quality iPSCs.

1.4 Conclusions

- I established a novel detection system of XCR, which can be utilized for visualizing the X chromosome status in live cells:
  - Tracking X chromosome inactivation (XCI) upon differentiation.
  - Monitor X chromosome reactivation (XCR) during reprogramming.
- My detection system of XCR during reprogramming provides a simple method for isolating high quality iPSCs, which are promising materials for regenerative therapy research.

1.5 Abbreviations

2-ME	2-Mercaptoethanol
c-Myc	Myelocytomatosis oncogene
Cdh1	Cadherin-1
CRISPR/Cas9	Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR associated protein 9
DD	Destabilizing Domain
DMEM	Dulbecco's modified Eagle medium
DNA	Deoxynucleic Acid
EpiSC	Epiblast Stem Cell
ESC	Embryonic Stem Cell
EGFP	Enhanced Green Fluorescent Protein
Esrrb	Estrogen Related Receptor Beta
FBS	Fetal Bovine Serum
GSK3βi	Glycogen Synthase Kinase 3β Inhibitor
hKO	Humanized Kusabira Orange
H3K27me3	H3K27 trimethylation
iPSC	Induced Pluripotent Stem Cell
IRES	Internal Ribosome Entry Site

Klf4	Krupple-like factor 4
KSR	Knockout Serum Replacement
LIF	Leukemia Inhibitory Factor
MEF	Mouse Embryonic Fibroblast
MEKi	MAPK/ERK Kinase inhibitor
NEAA	Non-Essential Amino Acid
Oct4	Octamer-binding Transcription factor 4
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
$Puro^R$	Puromycin Resistance
Rex1	Reduced Expression Protein 1
RNA	Ribonucleic acid
SeVdp	Sendai Virus defective and persistent
Sox2	SRY (Sex determining region Y)-box 2
Xa	Active X chromosome
Xi	Inactive X chromosome
XCI	X Chromosome Inactivation
XCR	X Chromosome Reactivation
Xist	X-inactive specific Transcript
Zeo^R	Zeocin Resistance

Chapter II. X Chromosome reactivation during reprogramming

2.1 Introduction

2.1.1 Induced pluripotent stem cells (iPS cells)

Embryonic stem cells (ESCs) are stem cells that are derived from the inner cell mass of a blastocyst. ESCs are able to divide and renew themselves indefinitely (self-renewal) and to differentiate into all types of cell in the embryo proper (pluripotency) [1, 2]. With their abilities of self-renewal and pluripotency, ESCs might be used to treat various diseases, such as diabetes, spinal cord injury and Parkinson's disease [3]. ESCs could be also used to understand the mechanisms of and screen drugs for various diseases. However, due to ethical controversies, it is almost impossible to use human embryos for these purposes. Moreover, it is difficult to derive patients' own ESCs, which are immune-compatible for transplantation. These barriers can be overcome by iPSCs, which are induced by from somatic cells by somatic cell reprogramming.

Somatic cells reprogramming occurs when somatic cells are fused with ESCs or when their nuclear contents are transferred into somatic cells [4-6], indicating that ESCs contain factors that can confer pluripotency to somatic cells. Therefore, these factors are hypothesized to play important functions in maintenance and induction of pluripotency in somatic cells. These factors include Oct3/4 [7, 8], Sox2 [9], Nanog [10, 11], Klf4 [12] and c-Myc [13], and they have been show to function in both early embryos and ESCs and contribute to the long-term maintenance and rapid proliferation of ESCs in culture. By combining these factors, it is possible to generate pluripotent stem cells directly from somatic cells such as mouse embryonic or adult fibroblasts.

iPSCs are a type of pluripotent stem cells that can be generated from somatic cells by induction of four specific genes encoding transcription factors – Klf4, Oct4, Sox2 and c-Myc [14]. iPSCs are similar to ESCs with regards to cell morphology, high expression of pluripotency marker genes, and ability to form embryoid bodies, teratoma and chimeric mice [15, 16]. Moreover, tissue-specific cells generated from iPSCs escape from immunological rejection upon transplantation since they are derived from a patients' own cells. In addition, since iPSC generation does not require destruction of human embryos, iPSCs pose little ethical concerns. Thus, iPSCs can be ideal sources for cell-based therapies to cure diseases for which there is currently no effective treatment.

However, to generate high-quality and safe iPSCs reproducibly, several obstacles need to be overcome. The initial iPSC generation system utilized retroviral vectors. Retroviral vectors integrate transgenes into the host genome, leading to alterations in the

genome sequence. In particular, the c-Myc gene, which is a proto-oncogene, increases the risk of tumor formation after retrovirus-mediated integration into the genomes [17]. Moreover, it is still unclear whether iPSCs are genetically and epigenetically equivalent to ESCs. A recent study showed that reprogramming is accompanied by copy number variations at a high frequency, giving rise to genetic mosaicism of iPSCs [18]. Whole genome profiles of DNA methylation in five human iPSC lines as well as ESCs, somatic cells and differentiated iPSCs indicate that iPSCs may have the variability in somatic memory and DNA methylation, raising the possibility of an unfaithful epigenetic reprogramming [19]. To avoid this problem, various methods for integration-free human iPSC generation have been reported. One of them is Sendai virus vectors encoding reprogramming factors used to induce human fibroblasts and obtained iPSCs [20]. Sendai viruses express reprogramming factors stably and generate integration-free iPSCs since they are RNA viruses, which replicate their genome in cytoplasm of infected cells.

Two other major problems in iPSC generation for research or therapy are the inefficiency of reprogramming fibroblasts into iPSCs (less than 10% of initial somatic cells can be successfully transformed into iPSCs) and the low number of high quality iPSCs in a population of reprogrammed cells (less than 0.02%) [21, 22]. In order to overcome these problems, the molecular mechanisms that underlie behind somatic cell reprogramming need to be clarified.

2.1.2 X chromosome inactivation and X chromosome reactivation

2.1.2.1 Dosage compensation

In mammals, there is a major difference in the composition of chromosomes between the two different sexes. Besides pairs of autosomes present in both males and females, there are sex chromosomes, X and Y, which differ between male and female. Male has an X chromosome and a Y chromosome whereas female has two X chromosomes. The chromosomal difference is critical for determining the sex of individual mammals.

The X chromosome is large and gene rich, possessing more than 1,000 annotated genes while the Y chromosome is small and gene poor. The two sexes also differ in the copy number of X-linked genes, which leads to imbalance in the amount of gene products between male and female. Imbalanced expression of several X-linked genes presumed to be lethal. Therefore, it needs to be dealt with by a mechanism to compensate for gene dosage (dosage compensation), by which expression of X-lined genes between different sexes is equalized. The dosage compensation is achieved by three main mechanisms. Firstly, X-linked gene

expression from a single male X chromosome is up-regulated by two fold [23]. Secondly, X-linked gene expression from female X chromosomes is down-regulated by two fold [24]. Finally, one of two X chromosomes is completely inactivated in female.

Two-fold up-regulation of X-linked genes in male was first reported by Muller in 1932 [25]. He tracked differences in *Drosophila melanogaster*'s eye color, which is under control of an X-linked gene, among male and female mutants. Muller mutated a gene that led to the loss or reduction of pigment in the eyes of a fly. Compared with females with two copies of the mutant gene, males with one copy of the mutant gene showed a similar degree of pigmentation, which he termed as "dosage compensation". In 1965, more advanced autoradiography experiments were performed to further confirm Muller's observed phenomenon of dosage compensation. Mukherjee and Beermann designed an experiment to visualize [3H]-Uridine incorporation into RNAs expressed from X chromosomes. Levels of [3H]-Uridine incorporation in the single male X chromosome was equal to the two female X chromosomes [26]. These results confirmed Muller's hypothesized dosage compensation. Recent technologies, microarrays and high throughput RNA sequencing have also provided strong evidence for this hypothesis in mammals [27-29].

However, in *Caenorhabditis elegans*, the system of sex determination is different [30]. *C. elegans* with two X chromosomes (XX) are hermaphrodites, and those with one X chromosome (XO) are males [31]. Similar to the XX/XY sex determination system, the difference in the number of X chromosome between sexes leads to differences in the expression levels of X-linked genes. In hermaphrodites, the normal expression of genes from both X chromosomes is lethal. To compensate dosage, hermaphrodites broadly reduce the expression levels of genes on both X chromosome. In the XX hermaphrodites, this repression occurs by two-fold down-regulation of transcription from both X chromosomes. This down-regulation is achieved by a specific complex on the X chromosomes in XX hermaphrodites, termed dosage compensation complex (DCC). Components of this complex are homologous to the condensin protein complex, which plays a central role in chromosome condensation and segregation during mitosis and meiosis. This homologue has led to the hypothesis that the DCC achieves X-linked gene repression by partially condensing the X chromosomes.

In 1949, Barr and Bertram observed a structure in the nuclei under a light microscope. They observed various mammalian species to find that this structure was present in nuclei of only female cells and therefore named it sex chromatin body. Thereafter, in 1959, Ohno showed that this structure was from one of the two female X chromosomes. He called these

structures Barr bodies. Ohno's studies of Barr bodies in female mammals revealed that these females used Barr bodies to inactivate one of their X chromosomes. Later, in 1961, Mary Lyon performed experiments on the expression of coat color genes on the X chromosome in female mice. Lyon proved that every cell of the female body inactivated either maternal or paternal X chromosome in a random mode. This result confirmed the heterogeneous patterns she observed in her mosaic mice. This process is known as random X chromosome inactivation [32]. Shortly thereafter, skin fibroblasts from a female who is heterozygous at the glucose-6-phosphate dehydrogenase G6PD) locus were grown in culture to isolate fibroblast clones. Only one allele was expressed in each clone originating from a single fibrablast, demonstrating the inactive state is inherited from one cell generation to the next, and random X chromosome inactivation occurred in human females [33, 34]. Thus, in mammals, dosage compensation is achieved by silencing one of two female X chromosomes via X chromosome inactivation.

2.1.2.2 X chromosome inactivation (XCI)

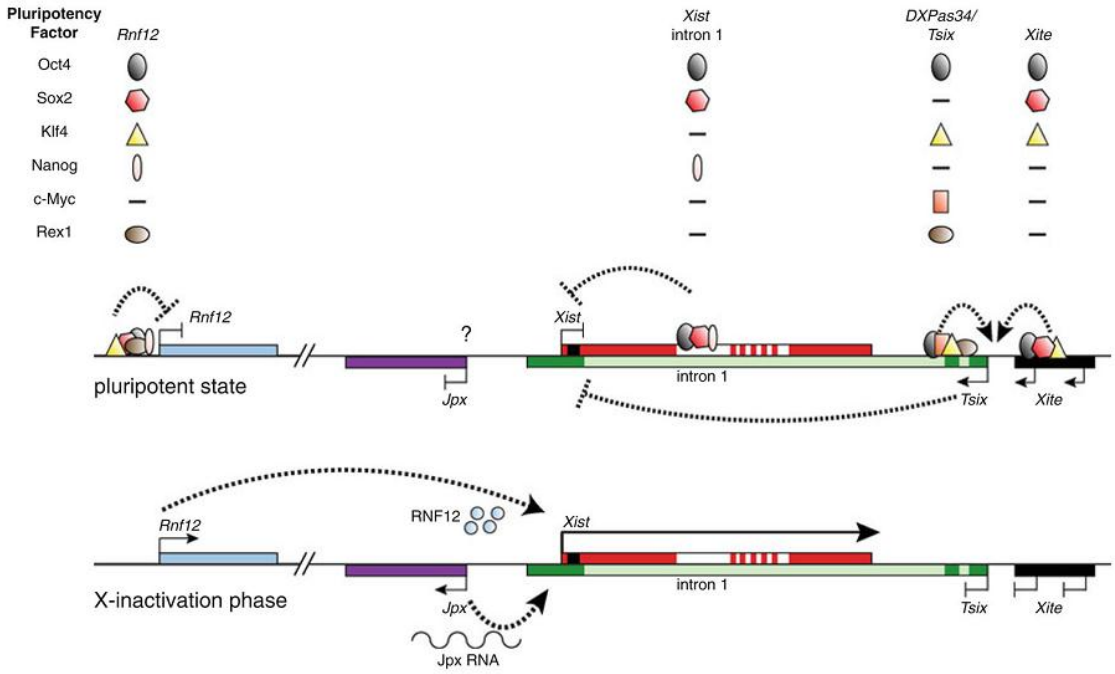
As compared with males, female mammals have a special epigenetic gene regulatory mechanism termed X chromosome inactivation (XCI) to deal with the gene dosage imbalance between females (XX) and males (XY). It is a compensation mechanism that occurs early in female embryonic development. Transcriptional silencing of one X chromosome is accomplished in diploid cells in epiblasts, and once established, the inactivated X chromosome is stably inherited through subsequent cell divisions. In the embryo, XCI occurs at random and leads to a mosaic distribution of cells which express either the maternal or paternal X chromosome [32].

Although random XCI occurs in all mammals, researchers have largely focused on the mouse system to investigate the underlying mechanisms of XCI. The inactivated X chromosomes (Xi) is condensed into a structure termed Barr body, and firmly maintained in a silent state upon subsequent cell divisions [35]. X inactivation center (Xic), a region on the X chromosome, controls initiation of XCI. This process depends on the upregulation of X inactive specific transcript (Xist) on one of the X chromosomes. Xist produces a long noncoding RNA (ncRNA) that accumulates and coats over the Xi chromosome in *cis* and triggers its transcriptional silencing [36, 37]. Thereafter, it mediates chromatin modifications, such as loss of RNA polymerase II and nascent transcripts, gain of chromatin marks associated with Polycomb group (PcG) complex, loss of histone H3 lysin 4 trimethylation (H3K4me3) and acetylation of histone H3 (H3ac) [38]. A mono-ubiquitination of histone

H2A lysine 119 (ubH2A) is mediated by polycomb repressive complex 1 (PRC1) and trimethylation of histone H3 lysine 27 (H3K27me3) is catalyzed by PRC2 on the Xi regulation [39-44].

Tsix, a non-coding RNA that is antisense to the Xist RNA, is known to partly mediate regulation of Xist transcription [45]. In undifferentiated ESCs, Xist expression is strongly repressed to ensure the active status of X chromosomes [46]. Tsix is involved in maintaining low levels of the Xist RNA in these cells. Targeted deletion of the Tsix gene has revealed that the Tsix RNA represses transcription of the Xist gene in *cis*. Any alteration in the Tsix gene that results in a lowered level of the Tsix RNA leads to skewed XCI with silencing of the X chromosome [47].

A previous study has shown that the Tsix gene may not be directly involved in regulation of Xist expression [46]. Bothe Xist and Tsix genes are targets of pluripotency factors. In undifferentiated mESC, Nanog, Oct4 and Sox2 have been reported to bind the first intron of the Xist gene [48]. Male ES cells being knocked-out of Oct4 and Nanog showed upregulation of the Xist gene, and this upregulation was independent of repression by the Tsix RNA. These results indicate that Oct4 and Nanog repress Xist transcription in mESCs independently of the Tsix.



Reference Figure 1. Model for molecular link between Xist repression and pluripotency factors [49]

While targeted deletions of pluripotency factors revealed the importance of their Xist intron 1 binding site [50], it also results in down-regulation of Xist activators such as Ring finger protein 12 (Rnf12) [51]. Consistently, Oct4, Sox2 and Nanog bind to the 5' region of the Rnf12 gene [52, 53]. In ESCs, Oct4 and Sox2 have also involved in Tsix upregulation by binding the Tsix enhancer regions [54]. In addition, Tsix expression in ESCs depends on Rex1 in addition to c-Myc and Klf4, all of which bind 5' regulatory regions of the Tsix gene as well. In fact, Rex1 is found to be important for transcriptional elongation of the Tsix gene [55]. Besides, when ES cells are differentiated, Rex1 expression is lost rapidly, suggesting that activation of the Tsix gene is reduced through loss of Rex1, and this reduction contributes to the initiation of XCI.

2.1.2.2 X chromosome reactivation (XCR)

X chromosome status changes dynamically during early mouse development [38] [56]. Paternal X chromosome (X_p) is inactivated at the 4-cell stage, which is commonly known as imprinted XCI. Imprinted X_p chromosome is activated in the inner cell mass (ICM) of blastocysts, resulting in two active X chromosomes (XaXa). This phenomenon is termed X chromosome reactivation (XCR). In epiblasts, either the paternal (X_p) or maternal (X_m) X chromosome is silenced by random XCI.

Germ cells are the only cells that escape stable somatic inheritance of X chromosome inactivation. Therefore, the mechanisms of Xi reactivation has been focused on primordial germ cells (PGCs) and investigated by several studies. It is reported that the process of XCR can be divided into three main steps, (1) Xist expression is repressed, (2) histone mark H3K27me3 dissipates from the Xi, and (3) X-linked genes are reactivated from the Xi. Reactivation of genes on the Xi occurs gradually and this process also overlaps with changes in chromatin modifications and DNA demethylation [57]. Interestingly, the timing of Nanog expression and the initiation of Xist repression appears in E7.5 PGCs [58], consistent with Nanog expression in ICM which is essential for the establishment of pluripotency [59]. Expression of the Xist RNA from imprinted paternal Xi is also repressed in $Nanog^+$ cells of the ICM [60]. Furthermore, in 2008, Silva and colleagues have reported that Xist repression and Xi reactivation at the transition from pre-iPS cells to iPS cells correlates with Nanog expression [61].

It has been proven that in ESCs, recruitment of Polycomb group (PcG) depends on the Xist RNA, therefore, loss of Xist expression could be an explanation for loss of H3K27me3 from the Xi [42] [62]. Nonetheless, it remains unclear whether repression of the Xist gene is

necessary for Xi reactivation in PGCs. During PGC development, several prominent events occurs such as changes in epigenetic modifications, including histone modifications and DNA demethylation. These events are thought to regulate reprogramming of the germline cells [57]. This makes PGCs an interesting system to investigate the mechanism and epigenetic processes of Xi reactivation.

By reprogramming female somatic cells into iPSCs, XCR can also be achieved [63]. XCR is a late event during reprogramming process to generate induced pluripotent stem cells (iPSCs) [64] [65]. Of note, XCI has attracted much attention in basic research in developmental biology, stem cell research and regenerative biology. However, its reverse process, XCR, remains one of questions that require further investigations.

2.1.2.3 X chromosome activities and human diseases

Human diseases could be strongly influenced by the establishment of X inactivation during development, indicating that any abnormal process of XCI can influence phenotype severity of recessive X-linked mutations [66, 67]. Tumorigenic process is also linked to XCI since an unisomy of X chromosomes which can profit recessive X-linked mutation during tumorigenesis [68]. Moreover, loss of Xi or amplification of Xa have been observed in ovarian and breast cancers [69].

To date, mouse model has been widely used to study XCI, therefore, XCI mechanism in other mammalian species is less well known. Ethical and technical issues related to human embryos and its derived ESCs have been hindrances to investigation of XCI mechanism in human development. The difficulty to obtain biological materials from early development has limited the study of XCI in humans.

Scientists have been currently focusing in Xi reactivation in order to define the reprogrammed pluripotent cell state and understand chromatin changes during reprogramming. Xi reactivation has also been considered for investigating human genetic diseases in female patients, which are caused by mutations on one of the two X chromosomes. Therefore, studies on Xi reactivation become a prominent topic for basic research as well as potential clinical applications in the future.

2.1.3 XCR – The key to unlock roadblock for high quality iPSCs

During early development stages from pre-implantation to post-implantation, pluripotent stem cells can be established from ICM and post-implantation epiblast cells. The embryonic stem cells (ESCs) are established from pre-implantation ICM and is termed naïve

pluripotent stem cells (PSCs). Epiblast stem cells (EpiSCs) are established from post-implantation stage epiblasts and termed primed PSCs. Primed EpiSCs represent a more differentiated state than naïve ESCs [65]. These two types of pluripotent stem cells differ in their optimal culture conditions, which reflect different signaling pathways involved in maintaining self-renewal and pluripotency. The maintenance of mESCs depends on LIF and BMP4, or inhibition of the MAP and GSK3 kinase pathways [70, 71], whereas the maintenance of EpiSCs requires both Activin and FGF2 [72, 73].

Naïve ESCs exhibit unlimited self-renewal capacity. When injected to the preimplantation embryos, they are readily incorporated into the epiblasts and enter embryonic development to produce mouse chimeras [74]. The EpiSCs also express some core pluripotency factors such as Oct4 and Sox2. However, they still differ from ESCs in expression patterns of several other transcription factors. EpiSCs show a "primed" pluripotent state and have capability in differentiating into various cell types *in vitro*, such as teratoma formation. Nevertheless, they fail to contribute to blastocyst chimeras, which indicates the ability of differentiation in EpiSCs is lower than naïve ESCs [75, 76]. Interestingly, X chromosome status is a most important difference between ESCs and EpiSCs. In female cells, naïve ESCs have two active X chromosomes (XaXa), while primed EpiSCs have one active and one inactive X chromosome (XaXi). These epigenetic differences could be considered as markers to distinguish two different states of PSCs. In addition, introduction of Klf4 allows EpiSCs to reprogram to naïve pluripotency state together with reactivation of the X chromosome [75]. Therefore, XCR could be a marker for good quality of female PSCs.

Furthermore, it is suggested that by experimentally inducing a chromatin environment related to mouse ESCs, a faster and more efficient reprogramming could be done [77]. It has been reported that the efficiency of mouse cloning is increased by deletion of the Xist gene [78], suggesting that XCR may pose a roadblock to efficient reprogramming. Previous studies have also shown that the epigenetic state of X chromosome in female iPSCs is tightly linked to pluripotency. Female mouse iPSCs with two active X chromosome (XaXa) exhibit higher pluripotency than iPSCs that have only one active X chromosome (XaXi). Thus, XCI and XCR are tightly linked to the loss and gain of pluripotency [79].

The generation of iPSCs has opened huge promising applications in regenerative medicine. It also provides a unique tool to study genetic diseases *in vitro*. Moreover, in order to establish animal models for human diseases which link to X chromosome, high-quality female iPSCs are preferable. As mentioned above, the presence of two active X chromosomes is considered as an indicator for the quality of iPSCs. Thus, deep

understanding on the XCR mechanism will enable us to overcome the roadblock of iPSC generation.

2.1.4 Model to detect X chromosome status in live cells

XCI has been focused for more than 50 years and to support XCI researches, various methods or techniques have been developed for its analysis. Common XCI detection methods are summarized in ***Reference Table 1***. X-linked green fluorescent protein (GFP)-expressing transgenic mice are widely used to detect XCI in live cells [80-83]. Easy distinction between inactive and active X chromosome by observing green fluorescent makes these transgenic mice useful for XCI research. The fact that no pre-treatment is required for the samples is an advantage of this model. Nonetheless, since they can only be used to monitor the activity of one X chromosome, these mice show limitations for some research purposes.

XCR has been attracting more attentions in recent years. Although XCR is genetically considered as a marker to distinguish naïve from primed state PSCs, the common XCR detection method is also fixation of cells or tissues to detect the H3K27me3 immunostaining pattern. Consequently, it is impossible to further characterize these cells. Common XCR detection methods are also summarized in ***Reference Table 2***. Due to the lack of a live imaging system to monitor XCI and XCR, there is a major obstacle in stem cell research using live PSCs. In order to visualize the X chromosome status in live cells and further extensively examine the mechanism of XCR, I established a novel live cell imaging system of XCR. The live imaging approach will enable us to monitor the changes in X chromosome status in these cells. More precise observations *in vitro* will enable us to understand the mechanisms of XCI and XCR *in vivo*.

Reference Table 1. Summary of XCI detection methods (Summarized from Reference [84])

Method	Pretreatment of samples	Note	Reference
Observation of Barr body	Fixation	Xi dark staining	[85]
Replication timing	Fixation	Xi shows late replication within the S phase of the cell cycle	[86]
Enzymatic activity	Cell extraction	Activities of X-linked enzymes (Hprt, Pgk1)	[87, 88]
RNA fluorescent in situ hybridization (FISH)	Fixation	Detection of X-linked gene expression	[60] [89, 90]
Allele-specific expression analysis	RNA extraction	Allele-specific expression analysis using DNA polymorphism	[91, 92]
Antibodies	Fixation	Immunostaining of H3K27me3, DNA polymerase II	[90]
Transgenic ESCs	Noninvasive	Ezh2-Venus transgenic ESCs to detect Xi	[93, 94]
Transgenic mice	Fixation	HMG-lacZ transgene inserted into X chromosome to detect its activity	[95]
	Noninvasive	CAG-eGFP transgene inserted into X chromosome to detect its activity	[80, 81]
	Noninvasive	microH2A-eGFP transgene inserted into autosomes to detect Xi	[82]
	Noninvasive	CAG-GFP and CAG-tomato transgenes inserted into X chromosomes to detect their activities	[83]
	Noninvasive	CAG-GFP and CAG-mCherry transgenes inserted into X chromosomes to detect their activities	[96]

Ezh2: enhancer of zeste homologue 2; HMG-lacZ: 3-hydroxy-3-methylglutaryCoA (HMG) promoter driving the *Escherichia coli* beta-galactosidase (lacZ) gene; Hprt: hypoxanthine phosphoribosyltransferase; Pgk1: phosphoglycerate kinase 1

Reference Table 2. Summary of XCR detection methods (Summarized from Reference [97])

Method	Pretreatment of samples	Note	Reference
Replication timing	Fixation	Reactivation of X-linked gene (Pgk1)	[98]
RNA fluorescent in situ hybridization (FISH)	Fixation	Tsix expression, Xist repression, loss of H3K27me3, X-linked gene reactivation (Pgk1)	[61] [63]
DNA methylation	Cell extraction	Methylation status on Xist promoter	[99]
Microarray analysis	RNA extraction	Xist repression and increasing X-linked genes	[100]
Transgenic cells	RNA extraction	CMV-GFP transgene inserted into X chromosome to detect its activity. Tsix expression, Xist repression, X-linked gene reactivation (X-linked-GFP transgene)	[101, 102]
Transgenic mice	Noninvasive	CAG-GFP and CAG-mCherry transgenes inserted into X chromosomes to detect their activities	[96]

2.2 Materials and Methods

2.2.1 Plasmids and guide RNAs

The pX330-U6-Chimeric_BB-Cbh-hSpCas9 (#42230) and pCAG-EGxxFP (#50716) were purchased from Addgene. pPyCAG-EGFP-IP and pPyCAG-EGFP-IZ were generous gifts from Dr. Hitoshi Niwa (RIKEN CDB). Guide RNAs (gRNAs) were designed using CRISPRdirect (https://crispr.dbcls.jp), and the gRNAs that had the minimum potential off-target effects were chosen for the *S* and *T* site as shown in Table 1. The B6N mouse Bac clones B6Ng01-177J10 (for the *S* site) and B6Ng01-316J16 (for the *T* site) were provided by the RIKEN BRC through the National Bio-Resource Project of the MEXT, Japan.

2.2.2 Construction of plasmids

To construct pCAG-EGxxFP-based validation plasmids, 600 bp~1700 bp genomic-DNA fragment containing the gRNA target sequences for *S* or *T* site was amplified from the mouse Bac clone and inserted into the multiple cloning site of the vector (Table 2). Complementary pairs of oligonucleotides encoding the gRNAs were annealed and inserted into the *BbsI* site of pX330-U6-Chimeric_BB-Cbh-hSpCas9 to prepare the Cas9/gRNA-expression vectors. The targeting vectors for knock-in of the fluorescent protein-coding genes into the *S* or *T* site were constructed using pPyCAG-EGFP-IP and pPyCAG-EGFP-IZ. The CAG promoter was replaced by the human elongation factor alpha-1 (EF-1) promoter, and the EGFP gene in pPyCAG-EGFP-IZ was replaced by hKO gene. The DNA fragment around the target site of gRNA1 (*S* site) or gRNA5 (*T* site) were isolated from the B6N mouse Bac clones and inserted into the upstream of the fluorescent protein-coding gene and downstream of the drug-resistant gene.

2.2.3 Transfection of HEK293 cells

Human embryonic kidney (HEK293) cells (RIKEN BRC, RBRC-RCB1637) were seeded at 1.5 x 10^5 cells/well in a 24-well plate and cultured at 37˚C under 5% CO_2 for 12 hours in Dulbecco's modified Eagle's medium (DMEM) (Nacalai tesque, Inc.) supplemented with 10% fetal bovine serum (FBS) (Thermo Fisher Scientific Inc.). Five hundred nanogram each of pX330-Cas9/gRNA-expression vector and pCAG-EGxxFP containing the gRNA target sequence were mixed with 2 µl of Lipofectamine® 2000 (Thermo Fisher Scientific Inc.), and the mixture was added to the HEK293 cells according to the manufacturer's protocol. EGFP expression was observed under a fluorescent microscope on the second day after transfection.

2.2.4 Transfection of mouse ES cells

Female mESCs (RIKEN BRC, AES0010) were seeded at $5x10^5$ cells/well on SNL feeder cells harboring the puromycin-resistant gene in a 6-well plate and cultured at 37°C under 5% CO_2 for 5 hours in DMEM supplemented with 1 mM sodium pyruvate (Nacalai tesque, Inc.), 15% KnockOut Serum Replacement (KSR) (Thermo Fisher Scientific Inc.), 0.1 mM nonessential amino acids (NEAA) (Wako pure chemical industries, Ltd.), 0.1 mM 2-mercaptoethanol (2-ME) (Thermo Fisher Scientific Inc.) and 1,000 U/ml LIF (Oriental Yeast Co., LTD.). Two microgram each of pX330-Cas9/gRNA expression vector and two different targeting vectors (phEF1-EGFP-IP-Syap1 and phEF1-hKO-IZ-Syap1, or phEF1-EGFP-IP-Taf1 and phEF1-hKO-IZ-Taf1) were mixed with 10 µl of Lipofectamine® 2000, and the mixture was added to the mESCs. The medium was changed to fresh same medium 5 hours after transfection to minimize the cell toxicity. Then the cells were treated with 1 µg/ml puromycin for 5 days followed by treatment with 50 µg/ml zeocin for 3 days to isolate EGFP/hKO-double positive mESC clones.

2.2.5 Genotyping analysis of isolated ESC clones

Genomic DNAs were extracted from the isolated EGFP/hKO-positive mESC clones and used as templates for PCR. To avoid contamination with feeder cells, the EGFP/hKO-positive mESCs were cultured without feeder cells for 5 days prior to DNA extraction. The location of primer sets used for PCR were shown in Fig. 4 and 5, and their sequences are listed in Table 3.

2.2.6 Differentiation of the $EGFP^+/hKO^+$ mESC clones

The isolated $EGFP^+/hKO^+$ mESC clones were grown on SNL feeder cells in a 100mm dish until the density becomes 80% confluent and trypsinized to suspend in the DMEM supplemented with 20% FBS, 0.1 mM NEAA and 0.1 µM 2-ME. The cell suspension was transferred to a 100mm cell culture dish and incubated for 20 minutes to remove feeder cells which attach to the dish quickly. Then, the supernatant containing the EGFP/hKO-positive mESCs was collected and plated into a 100 mm non-coated bacterial dish (AGC TECHNO GLASS CO., LTD.) for formation of embryoid bodies (EBs). After 5 days, EBs were trypsinized and filtrated through a 100 µm cell strainer (BD Falcon). The filtrated cells were cultured on a collagen Type I-coated dish (AGC TECHNO GLASS CO., LTD. and Sumitomo Bakelite Co., Ltd) in the presence of 50 µg/ml zeocin to select hKO-positive single-colored cells.

2.2.7 Reprogramming of the EGFP$^+$/hKO$^+$ mESC-derived differentiated cells

The isolated hKO$^+$ differentiated cells were seeded in a 24-well plate at 2.5x10^4 cells/well in DMEM plus 10% FBS and cultured at 37°C under 5% CO_2 for 12 hours. The cells were infected with the Sendai virus which expresses Klf4, Oct4, Sox2 and c-Myc (SeVdp(KOSM)) for 16 hours at 32°C to induce reprograming. The virus-infected cells were trypsinized and cultured on SNL-feeder cells in Knockout DMEM (Thermo Fisher Scientific Inc.) supplemented with 15% KSR, 2 mM GlutaMAX (Thermo Fisher Scientific Inc.), 0.1 mM NEAA, 55 μM 2-ME, 100 units/ml penicillin, 100 μg /ml streptomycin (Nacalai tesque, Inc.) and 1000 U/mL LIF for 7 days. The culture medium was replaced by 2i medium (1:1 mixture of DMEM/F12 (Nacalai tesque, Inc.) and Neurobasal medium (Thermo Fisher Scientific Inc.) supplemented with N2 supplement (Thermo Fisher Scientific Inc.), B27 supplement (Thermo Fisher Scientific Inc.), 2 mM GlutaMax (Thermo Fisher Scientific Inc.), NEAA, 0.1 mM 2-ME, 0.05% BSA (Thermo Fisher Scientific Inc.), 1,000 U/ml LIF, 1 μM MEK inhibitor PD0325901, 3 μM GSK3β inhibitor CHIR99021, 100 units/ml penicillin and 100 μg/ml streptomycin) or DMEM supplemented with 1 mM sodium pyruvate, 15% KSR, 0.1 mM NEAA, 0.1 mM 2-ME and 1,000 U/ml LIF for continuous culture of iPSCs.

2.2.8 Reverse transcription and quantitative real-time PCR

Total RNA was extracted from the EGFP$^+$/hKO$^+$ mESCs, embryoid bodies, differentiated cell and EGFP$^+$/hKO$^+$ iPSCs using Sepasol-RNA I Super G (Nacalai tesque, Inc.) according to the manufacture's instruction. To avoid contamination with feeder cells, the EGFP$^+$/hKO$^+$ mESCs and hKO$^+$ cells-derived iPSCs were cultured without feeder cells for 5 days prior to RNA extraction. Reverse transcription was performed using Superscript III First-Strand Synthesis System (Thermo Fisher Scientific Inc.), and the synthesized first-strand cDNA was used to measure the mRNA level of various marker genes by quantitative real-time PCR using GoTaq qPCR Master Mix (Promega Corp.). The mRNA level of γ-tubulin was used to normalize the obtained data.

2.2.9 Statistical analysis

The data were analyzed by t-test analysis to perform the statistical analyses. P values of <0.05 were considered significant. The data are presented as the mean ± standard error (SE).

2.3 Results

2.3.1 Establishment of detection system of X chromosome status

2.3.1.1 Novel detection system of X chromosome status in live cells

In order to visualize the X chromosome status in live cells and analyze the mechanism of X chromosome reactivation (XCR), a novel detection system of XCR is established. Two reporter genes encoding two different fluorescent proteins, enhanced green fluorescent protein (EGFP) and humanized Kusabira orange (hKO), are inserted into each X chromosome of female mouse embryonic stem cells (mESCs). These cells are expected to initially display both green and orange fluorescence owing to two active X chromosomes. Upon differentiation, they show either green or orange fluorescence, indicating that the mESC clones undergo X chromosome inactivation (XCI). Single-color cells are reprogrammed into iPSCs. High quality iPSCs display two fluorescence, indicating XCR upon reprogramming (Fig. 1).

2.3.1.2 Knock-in fluorescent protein-coding genes into X chromosomes

To visualize XCR in live cells during somatic cell reprogramming, I first generated female ESCs that express EGFP from one X chromosome and hKO from the other. To insert the EGFP and hKO genes into the genome, I avoided protein-coding genes as an insertion site because of their potential effect as a facilitator or inhibitor on the reprogramming process when iPSCs are generated [103]. Instead, I chose two intergenic sites near the *Syap1* or *Taf1* gene on the X chromosomes (Fig. 2). These sites were chosen because the insertion sites, which I term *S* and *T* sites, are near the genes, *Syap1* and *Taf1*, respectively, that are subject to XCI [104]. In addition, database search of National Center for Biotechnology Information (NCBI) showed that the genes surrounding the *S* site (*Syap1*, *Txlng*, *Rbbp7*, and *Ctps2*) and the *T* site (*Taf1*, *Nono*, *Zmym3*, and *Med12*) do not exhibit strong tissue- or developmental stage-specific expression pattern. Moreover, the GeneProf database (http://www.geneprof.org/) [105] showed that these sites are sandwiched between CTCF binding sites together with at least one of these surrounding genes. Thus, the EGFP and hKO genes that are inserted into the *S* and *T* sites were expected to obey XCI and XCR in a similar manner to the surrounding genes.

By using the CRISPR/Cas9 system, I insert two reporter gene cassettes encoding two different fluorescent proteins into specific sites. I expected that different insertion sites might provide different monitoring results as a recent research has suggested that XCI is not

uniform throughout chromosomes [106], therefore, XCR might be different. At each site, different guide RNAs (gRNA) specific for each site were designed (Fig. 2).

To validate the working of gRNA sequences for targeting *S* site, I cotransfected the pCAG-EGxxFP target sequence and pX330-hSpCas9-gRNA plasmids into HEK293 cells and then the reconstituted EGFP fluorescent was observed 48 hours after transfection. In the presence of gRNA sequences, the transfected cells became fluorescent. EGFP expression levels were also different between gRNAs. Among three tested gRNAs, $gRNA_1$ showed the strongest fluorescent signal while compared with other gRNAs (Fig. 3A). This result indicated that $gRNA_1$ has better efficiency on targeting genomic sequence and it is able to be used to target desired genomic DNA region. Similarly, on *T* site, I also tested gRNAs to target this site efficiently. As shown in Fig. 3B, $gRNA_5$ was selected for targeting female mESCs.

Following the validation of gRNA in HEK293 cells, I prepared different targeting vectors, each of which contained one of the fluorescent protein genes and a drug-resistant gene between the homologous sequences in the targeted site (*S* site: Fig. 4A, 4B) (*T* site: Fig. 5A, 5B). To generate mouse embryonic stem cell (mESC) lines that carry two different fluorescent protein genes on each allele on both X chromosomes, I transfected targeting vectors together with the expression vector of Cas9 and a guide RNA into female mESCs. In this study, I tested two different strategies to knock in fluorescent protein-coding genes into X chromosomes. The first strategy is step-by-step method by which mESC clones harboring the EGFP gene in one allele of X chromosome are isolated and then used to insert the hKO gene into the other allele (Fig. 4 and Fig. 5). The second strategy is simultaneous delivery of two different fluorescent protein coding genes into cells (Fig. 7 and Fig. 8).

a. Step-by-step method for knocking in fluorescent coding genes into X chromosomes

The first strategy is step-by-step method by which mESC clones harboring the EGFP gene in one allele of X chromosome are isolated and then used to insert the hKO gene into the other allele (Fig. 4C). Though CRISPR/Cas9 system has been widely used in genome editing, it has not been well-established procedure to generate gene knocked-in ESC lines. Therefore, I first determined conditions suitable for establishing ESC lines with minimized non-specific integration. For 24 well plate scale, I transfected female mESCs Cas9/gRNA expressing plasmid together with different amount of EGFP targeting vectors. Following transfection process, the transfected cells were selected by addition of puromycin. After

seven days of transfection, the $EGFP^+$ colonies were counted and compared between samples. As shown in Fig. 4D and 4E, increasing amount of targeting vector increased $EGFP^+$ colonies. Nonetheless, non-specific integrated $EGFP^+$ colonies also increased (samples without gRNA) although colony number in 0.5 µg, 1.0 µg, 2.0 µg targeting vector samples were almost similar. Therefore, to minimize random integration of EGFP gene, I decided to use 0.5 µg targeting vector together with 0.5 µg Cas9/gRNA expressing vector.

To generate mESC clones harboring the EGFP gene in one allele of X chromosome, I transfected optimized combination of Cas9/gRNA expressing plasmid together with EGFP targeting vector into female ESCs. Puromycin were added to cultured cells to select transfected cells. Green colonies after puromycin selection were collected to validate EGFP insertion in one allele of X chromosome (Fig. 4F and 4G). On *S* site, as a suggestive of homologous recombinant (HR) mediated genome editing, colonies were detected to carry the expected 9.5 kb fragment. PCR analysis with indicated primer set (shown in Fig 4A) confirmed candidate ESC clone with expected insertion (Clone #19). This clone was used to insert hKO coding gene into second allele of X chromosome. Zeocin was added after two days of transfection. After second transfection, I could observe double-colored (EGFP and hKO expression) colonies as indication for hKO insertion (Fig. 4H). PCR analysis with primers detecting full length of insertion showed initial heterozygous mESCs became homozygous ones (clones marked with red rectangles). However, hKO gene could not be inserted into another allele of X chromosome (Fig. 4K). The similar number of double-colored colonies between samples transfected with and without gRNA (Fig. 4L) indicated that hKO gene was randomly integrated into genome. Using this method, I failed to knock-in two fluorescent protein coding genes into both alleles of X chromosomes.

At *T* site, I also proceeded same procedure to generate $EGFP^+/hKO^+$ ESC lines. After first transfection, I could generate three clones (#34, #40, #43) with EGFP insertion in one allele (Fig. 5F). These three clones were both used to insert hKO gene into second allele of X chromosome. However, I could not detect any hKO insertion in double-colored colonies. Similar to *S* site, step-by-step method was not suitable for knocking-in fluorescent protein coding genes into specific site.

b. Effect of medium components on $EGFP^+/hKO^+$ mESC generation

During generation of $EGFP^+/hKO^+$ mESCs, taken advantage of double-color ESC system, I found out that my current ESC culture medium (DMEM supplemented with KSR and LIF) is not suitable to maintain pluripotency of mESCs. Candidate double-colored mESC

clones could not survive under drug selection, indicating one of X chromosomes was inactivated during culturing. 2i medium (Serum-free medium supplemented with two inhibitors – MEK and GSK inhibitors) is proven to be the standard for ES cell culturing. However, a drawback with 2i medium is that it generates cultures that are very difficult for transfection. Therefore, I tested transfection efficiency with different medium: 2i medium; E medium (current medium), E+2i(s) medium, 50/50 medium. Components of each medium were listed in Fig. 6A. Experimental procedure was followed as shown in Fig. 6B. Female ESCs were cultured in four different media. The cells were transfected Cas9/gRNA plasmid together with two targeting vectors. After two days of transfection, puromycin was added to cultured medium. Morphology and fluorescent signal of transfected cells were monitored daily during transfection. Double-colored colonies surviving after drug selection were then collected, validated both EGFP and hKO insertion at both alleles of X chromsomes. Recombination was confirmed by polymerase chain reaction (PCR) analysis of their genome DNA (Fig. 6E-G) with different primer sets as shown in red arrows in Fig. 4A and 4B. On *S* site, as a suggestive of homologous recombinant (HR) mediated editing genome, colonies were detected to carry the expected 9.5 kb fragment (Primers: a+b). mESC clones having the expected EGFP and hKO inserts at both alleles of X chromosomes (Primers: a+c, a+d) were marked in red. Targeting efficiency in each medium were also calculated (Fig. 6H). On tested *S* site, cells responsed differently to medium. The number of double-colored colonies after transfection were different between medium. 2i medium seems to prevent random integration of targeting vector (without gRNA) outside of targeted region. However, this medium is too severe for transfected cells, few cells could survive. In E+2i(s), higher number of double-colored colonies were obtained, it still resulted in high percentage of random integrated colonies. 50/50 medium showed highest targeting efficiency among mediums. For the observation during transfection, this medium also supported transfected colonies growing with round shapes and maintaining fluorescent signals than other medium. Perhaps, it is combination of 2i medium and E medium. One is known to maintain stem cells at naïve state, and one is a standard serum-containing medium with LIF, which could support cell expansion faster than other medium. In my study, it is recommended to culture the cells in this combined medium for transfection.

c. Simultaneous delivery of two different fluorescent protein coding genes into cells

With optimal combined medium for transfection, I tested second strategy to knock-in EGFP and hKO gene into specific location. I simultaneously delivered two different

fluorescent protein coding genes into cells. Two days after transfection, the transfected cells become fluorescent. The single-color ESCs were removed by sequential selections with puromycin and zeocin to obtain $EGFP^+/hKO^+$ ESC colonies (*S* site: Fig. 7A and *T* site: Fig. 8A).

These mESC clones that expressed two fluorescent signals in second method were also collected (Fig. 7B and 8B). Recombination was confirmed by polymerase chain reaction (PCR) analysis of their genome DNA (Fig. 7C and Fig. 8C) with different primer sets as shown in red arrows in Fig. 4A, 4B and Fig. 5A, 5B. On *S* site, among the isolated 50 clones, 33 clones that grew well were genotyped, and most of the isolated clones had the inserted gene(s) at the *S* site (Primers: a+b). However, only five ESC clones (No. 20, 21, 29, 36, and 40) had the EGFP and hKO gene at each *S* site on the X chromosome while other clones had only the EGFP gene or hKO gene on both X chromosomes (Primers: a+c, a+d) (marked in red). These results indicated heterozygous female mESCs ($EGFP^+/hKO^+$) has been established (Fig. 7C). In addition, to further confirm non-specific integration of neither EGFP nor hKO, DNA genome from these five clones were also used as template for PCR analysis with primers on targeting vectors and EGFP or hKO sequences (Fig. 7D). Two out of five clones, (clone No. 20 and No. 29, hereafter called "S20" and "S29") showed single copy of fluorescent reporter genes at targeted site (Fig. 7D). Similarly, on *T* site, I also established $EGFP^+/hKO^+$ mESCs which carry single copy of fluorescent reporter genes at targeted site (clone No. 36, hereafter called "T36") (Fig. 8). Conclusively, at each site, I successfully generated heterozygous $EGFP^+/hKO^+$ mESCs (*S* site: S20 and S29, *T* site: T36). These clones were used for my further experiments.

d. Generation of negative selection marker containing $EGFP^+/hKO^+$ mESCs

In this study, I aim to establish system to monitor X chromosome reactivation during reprogramming and further investigate XCR mechanism. For more convenient monitoring, I collect single-color differentiated cells. Moreover, it is difficult to achieve XCI completely without any remained undifferentiated cells. Therefore, a strategy for selecting single-color cells without contaminated undifferentiated cells is developed. In this study, I tested negative selection by Thymidine kinase (TK) gene. For that purpose, I aim to establish mESC clones whose one allele carry EGFP coding gene together with TK gene, another allele carries hKO gene. During differentiation, negative selection with Ganciclovir will be performed to remove TK^+ cells. Therefore, only hKO^+ differentiated cells (TK^- cells) were selected for reprogramming (Fig. 9A). TK gene is constructed together with puromycin resistance gene in

EGFP targeting vector (pEF1-EGFP-IP.TK targeting vector) (Fig. 9B and 9C). Unexpectedly, female mESCs being transfected EGFP-IP.TK targeting vector showed weak fluorescent signal and could not survive under puromycin and zeocin selection during transfection (Fig. 9D and 9E). Knock-in of TK gene into mESCs might be harmful for cells.

2.3.2 Detection of X chromosome inactivation during differentiation

The integration site of a reporter is important if I are about to monitor X chromosome inactivation (XCI) because some genes on X chromosomes are known to escape from XCI process. In order to evaluate XCI subjection of the inserted reporter gene, three candidate mESC clones are differentiated through embryoid bodies (EBs) and observed changing of fluorescent signal in monolayer cells as shown in Fig. 10A. At mESC stage, initial cells are expected to express double signals, implying 2 active X chromosome. Upon differentiation, at EB stage, they generate chimeric cluster of cells displaying either green or orange fluorescence. EB-derived cells express only one of the two fluorescent proteins ($EGFP^+$ or hKO^+). In my established candidate clones (S20, S29 and T36), they initially showed double fluorescent pattern at mESC stage, chimeric pattern at EB stage and expressing either green or orange pattern at single cell state (Fig. 10B and Fig. 10C). This indicated that the inserted fluorescent protein genes are subjected to X chromosome inactivation in live cells. It also confirmed there is no random integration of EGFP or hKO along genome, which are consistent with random integration confirmation in Fig. 7D and Fig. 8D. The silencing is initiated by the long noncoding RNA, X inactive specific transcript (*Xist*), which coats Xi. The *Xist* gene is exclusively expressed in inactive X chromosome (Xi) and accumulates within the territory of the Xi. Therefore, I further confirmed XCI in monolayer cells by checking expression of *Xist*. During differentiation from ESCs to monolayer cells with transient EBs, *Xist* were significantly increased (Fig. 10D). *Xist* expression confirmed inactive X chromosome (Xi) status. Taken together, these results suggested that these cells enable monitoring of XCI by the fluorescent signals in live cells.

Moreover, XCI process in *T* site $EGFP^+/hKO^+$ cells occurred earlier than *S* site cells. This observation might be resulted from insertion location. As I showed insertion sites for each site in Fig. 2, *T* site is closer to *Xist* than *S* site, hence, XCI might proceed faster in *T* site. Notably, during XCI process, some somatic cells differentiated from the T36 clone were found to lose expression of both EGFP and hKO (marked in white dot line), suggesting that the gene inserted into the *T* site may be repressed independent of XCI and monitoring of XCI is affected by insertion site. Taken together, these results indicate that the fluorescent protein

genes driven by the human EF-1α promoter are subject to XCI even when inserted into intergenic sites of the X chromosome, and the fluorescent genes at the *S* site may be more suitable than those at the *T* site for observing XCI. Therefore, I will focus on *S* site clone in further experiments. Summary of these clones is described in Table 5.

2.3.3 Detection of X chromosome reactivation during reprogramming

For further observation or tracking X chromosome reactivation (XCR), I collect hKO^+ cells derived from S29 ESC clone for reprogramming. As shown in Fig. 11A, I induce $EGFP^+/hKO^+$ mESCs to differentiate through embryoid bodies (EBs). Then, EBs will be suspended into single cells. hKO^+ differentiated cells will be isolated by zeocin. hKO^+ cells are reprogrammed into iPSCs. iPSCs display two fluorescent signals, indicating XCR upon reprogramming.

After isolating hKO^+ differentiated cells, quantitative RT-PCR was performed to check differentiation status. The result showed significantly reduced level of pluripotency marker gene (*Oct4*) and increased level of differentiation marker genes (*Cdh2*, *Tgfb1*) (Fig. 11B). These data implied that hKO^+ cells were undergone XCI, maintained one active X chromosome (XaXi).

2.3.3.1 Effect of medium component on XCR observation

XCR is a reversal process of inactivation which is occurred during reprogramming. To examine whether establishing system could be used for detecting XCR *in vitro* or not, I induced iPSCs from hKO^+ cells ($Xa^{hKO}Xi^{EGFP}$ cells) by infecting SeVdp(KOSM), which express 4 Yamanaka factors (Klf4, Oct4, Sox2 and c-Myc). Different stem cell laboratories rely on different culture conditions to support the expansion and maintenance of pluripotent stem cells. Culture of stem cells under undefined conditions are not able to enhance pluripotency of stem cells. Hence, it might affect to my observation during reprogramming. Therefore, I first tested effect of medium components on XCR observation. I used two different media (Serum-free medium: 2i medium and Standard serum containing medim with LIF: ES medium) for culturing reprogramming cells. Following reprogramming, fluorescent signal is monitored to track XCR (EGFP reactivation - $Xa^{hKO}Xa^{EGFP}$). In 2i medium, around day 15 of reprogramming, colonies reprogrammed from hKO^+ cells started to show heterogenous pattern of double signal, indicating partial XCR. Around day 17, I observed homogenously double-colored colonies, indicating completed XCR (Fig. 11C). In standard serum-containing medium with LIF, initiation of XCR timepoint was slightly delayed. It also could not support completed XCR, only heterogeneously double-colored colonies were

formed. Moreover, compact round shape colony is a well-known marker for distinguishing naïve and primed pluripotent stem cells. Serum-free medium (2i medium) maintained reprogrammed colonies in rounder shape than serum medium. Therefore, in the purpose of using fluorescence as an indicator for high quality iPSCs as well as for more convenient observation, I decided to use serum-free medium during reprogramming.

I further determined whether SeV infected cells were successfully reprogrammed by checking expression of pluripotency genes (*Cdh1*, *Oct4*, *Rex1*). These genes were upregulated during iPSC generation and showed similar expression level of mESCs. Especially, *Rex1*, a marker for high quality PSCs, expressed as high as mESCs (Fig. 11E). I also evaluated *Xist* expression to confirm X chromosome status in generated iPSCs since XCR occurs after silencing of *Xist* RNA [79]. *Xist* expression was significantly decreased in iPSCs although its expression was still detected (Fig. 11F). Since population of generated iPSCs is heterogenous, it is hard to detect *Xist* silencing completely. The downregulation of *Xist* is sign for XCR. Conclusively, my system could be used for detecting XCR during reprogramming.

2.3.3.2 Correlation between pluripotency and XCR

Using this system, it is obvious to observe heterogeneity of reprogrammed cells. In population of reprogrammed colonies, some colonies showed homogenous or heterogenous pattern of double signal, some of them maintained orange signal. I obtained iPSC colonies reprogrammed from hKO^+ somatic cells that were derived from *S* site ESCs, that were morphologically indistinguishable but nonetheless showed different expression patterns of EGFP. Fig. 12A shows fluorescent pattern of picked-up colonies in reprogrammed population. It indicated that XCR has not occurred simultaneously in all formed colonies.

Additionally, in 2014, Pasque *et al* showed XCR occurred at late phase of reprogramming. In this study, $Nanog^+$ cells exhibited biallelic expression of X linked genes, a sign of XCR. Therefore, reactivation of X chromosome is closely linked to pluripotency or quality of iPSCs. In order to examine the relationship between XCR and pluripotency, expression level of pluripotency marker genes was determined in different miPSC clones (Fig. 12B). Interestingly, I found out correlation between pluripotency and XCR pattern. Clones maintaining orange signal expressed lower level of pluripotency marker genes than heterogenous or homogenous pattern of double fluorescent signals (Fig. 12B). It indicated XCR is correlated to pluripotency.

In particular, I found out that homogenously double-colored iPSCs (iPSC clone #6), which indicated completed XCR, showed higher *Rex1* expression (a marker for high pluripotency stem cells) than heterogeneously double-colored (#3) or orange iPSCs (#20) (Fig. 12C). *Xist* expression were also determined to gain more evidences on difference on XCR process between these iPSC clones. Expectedly, I found correlation between *Xist* expression level and pattern of fluorescence in iPSC clones. *Xist* is significantly repressed in homogenously double-colored iPSCs (iPSC #6) compared to heterogenous double-colored iPSC clone or orange clone (Fig. 12D). *Xist* expression was consistent with *Rex1* expression level between iPSC clones. This data further confirmed correlation between XCR and pluripotency. Although further experiments should be done to confirm pluripotency of generated double-colored iPSCs *in vivo*, this system provides simple methods for detecting heterogeneity of reprogramming cells' population and isolating high and low quality iPSC by monitoring XCR during reprogramming.

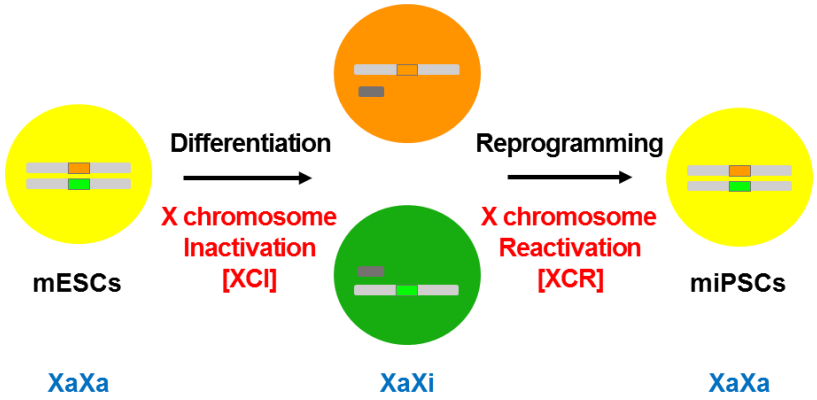


Figure 1. Detection system of X chromosome inactivation (XCI) and reactivation (XCR)

Two different fluorescent reporter genes (EGFP and hKO) are inserted into X chromosomes of female mESCs. Before differentiation, mESCs express double (EGFP and hKO) fluorescence from their two active X chromsomes (XaXa). Upon differentiation, they express either orange or green signal due to XCI, which occurs randomly in one of the two X chromosomes. During reprogramming, which is reverse process of differentiation, the inactive X chromosome is reactivated by the process termed X chromosome reactivation (XCR), which can be monitored by the presence of two fluorescence signals.

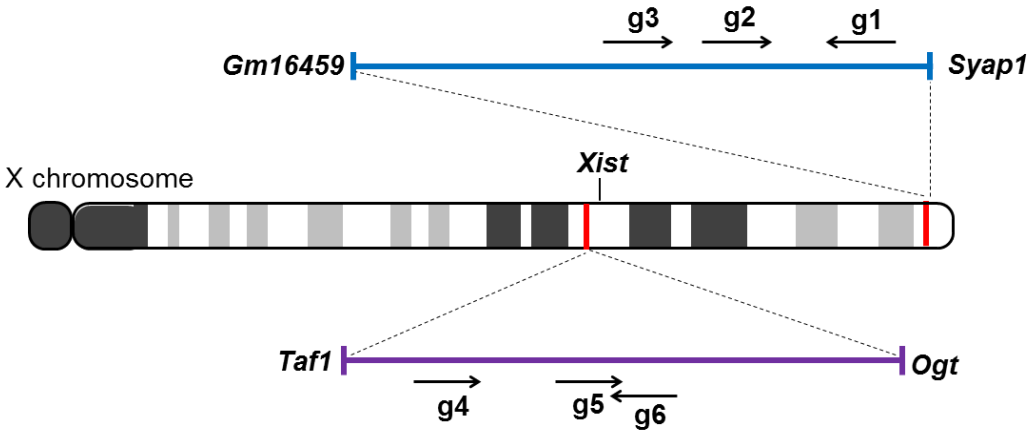


Figure 2. X chromosomal locations where fluorescent protein coding genes were inserted

Two sites on the X chromosome were used for insertion of reporter genes encoding fluorescent protein markers. Intergenic regions close to *Syap1* (*S* site) and *Taf1* (*T* site) were chosen. Guide RNAs (gRNAs) for targeting each site were also shown.

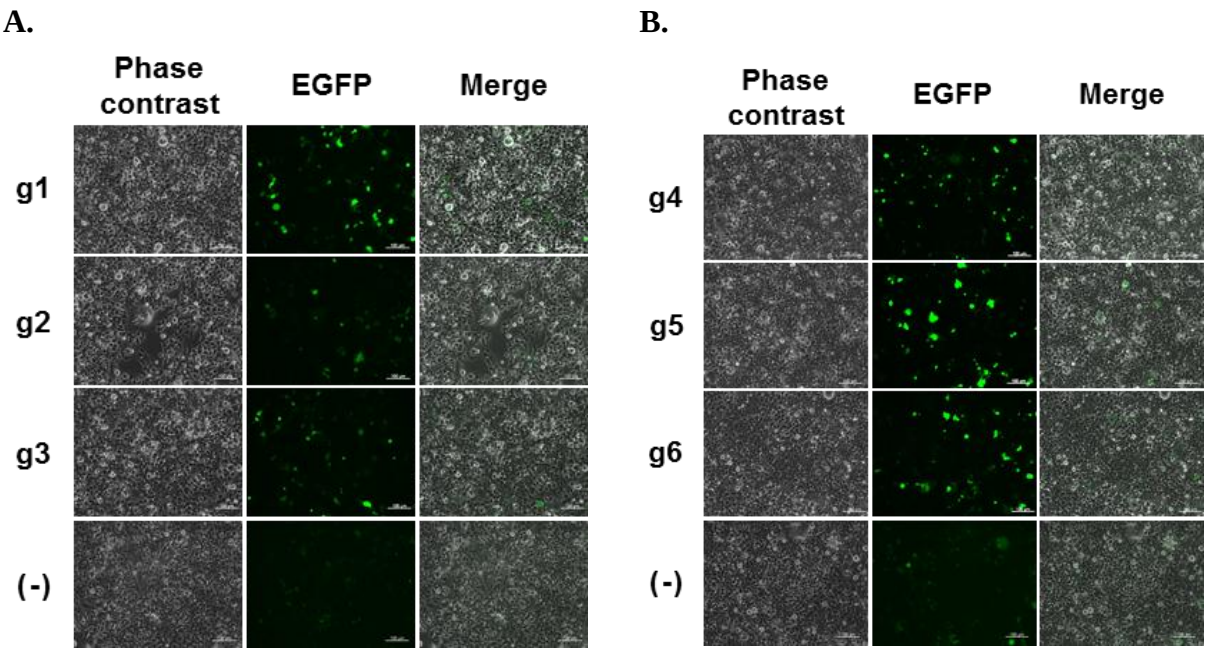
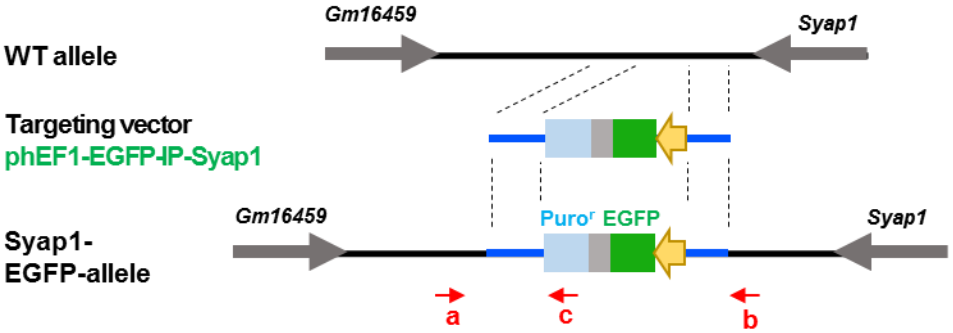
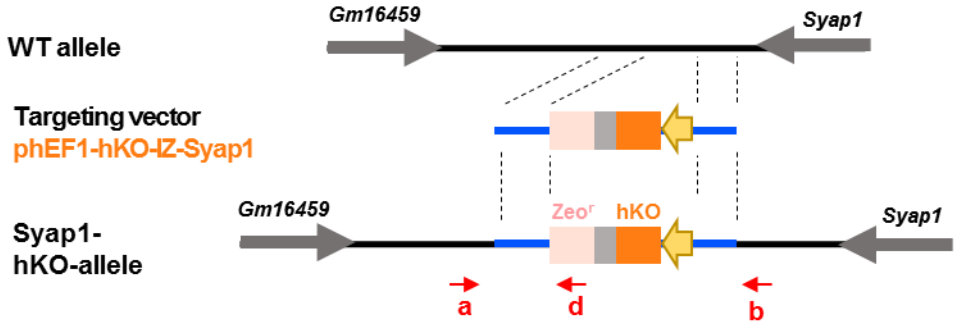


Figure 3. Validation of gRNA efficiency. Cas9-gRNA expression vectors and validation vectors (pCAG-EGxxFP) containing the corresponding gRNA target sequence were transfected into HEK293 cells. The efficiency of targeting the two sites by each gRNA is validated by EGFP expression. Scale bars: 100μm.

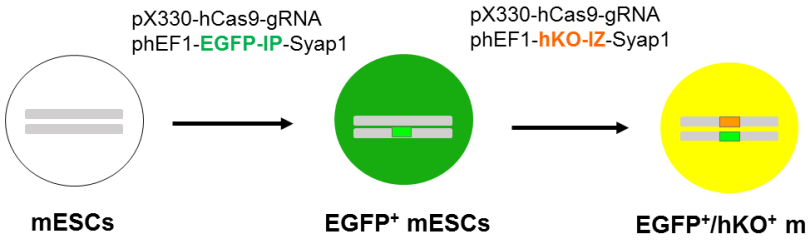
A.



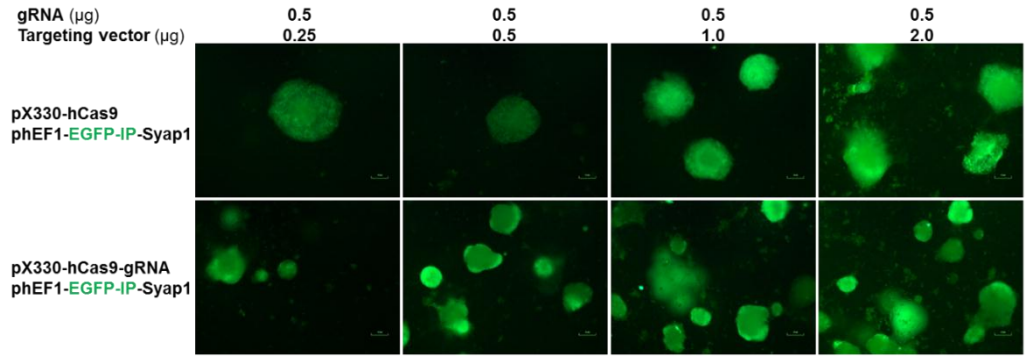
B.



C.



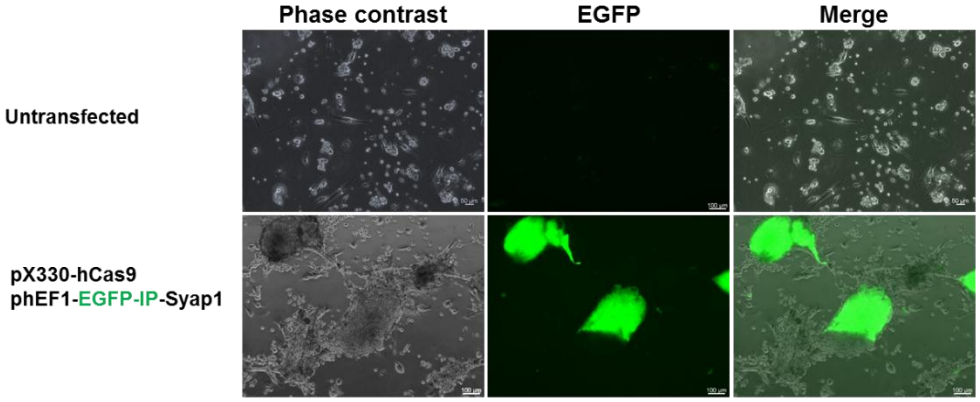
D.



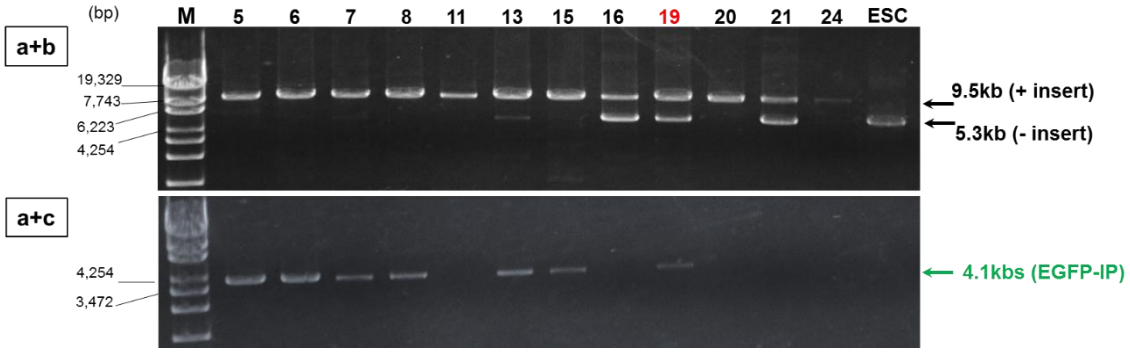
E.

Cas9-gRNA vector (μg) **Targeting vector (μg)**	**0.5** **0.25**	**0.5** **0.5**	**0.5** **1.0**	**0.5** **2.0**
pX330-hCas9 phEF1-EGFP-IP-Syap1	7	18	41	35
pX330-hCas9-gRNA phEF1-EGFP-IP-Syap1	69	86	100	97

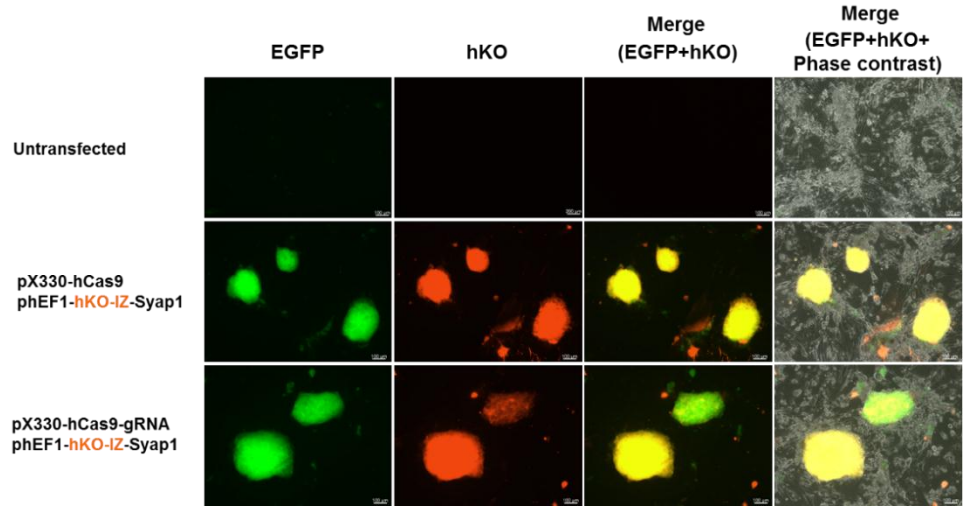
F.


Phase contrast
EGFP
Merge
Untransfected
pX330-hCas9
phEF1-EGFP-IP-Syap1

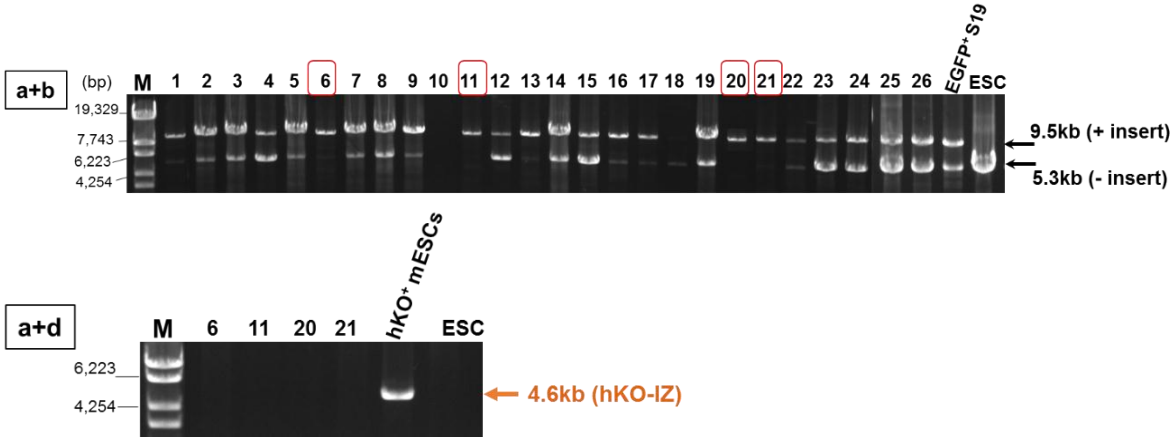
G.


(bp)
M 5 6 7 8 11 13 15 16 19 20 21 24 ESC
a+b
19,329
7,743
6,223
4,254
9.5kb (+ insert)
5.3kb (- insert)
a+c
4,254
3,472
4.1kbs (EGFP-IP)

H.


EGFP
hKO
Merge (EGFP+hKO)
Merge (EGFP+hKO+ Phase contrast)
Untransfected
pX330-hCas9
phEF1-hKO-IZ-Syap1
pX330-hCas9-gRNA
phEF1-hKO-IZ-Syap1

K.

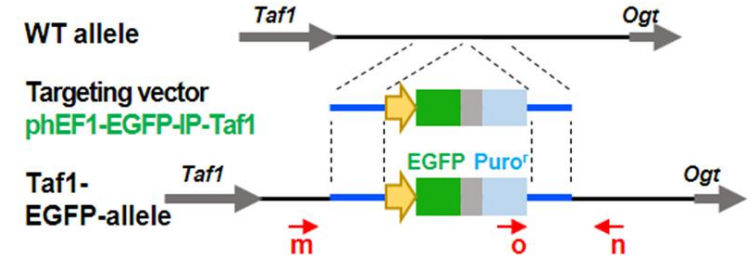

a+b
(bp)
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 EGFP+ S19 ESC
19,329
7,743
6,223
4,254
9.5kb (+ insert)
5.3kb (- insert)
a+d
M 6 11 20 21 hKO+ mESCs ESC
6,223
4,254
4.6kb (hKO-IZ)

L.

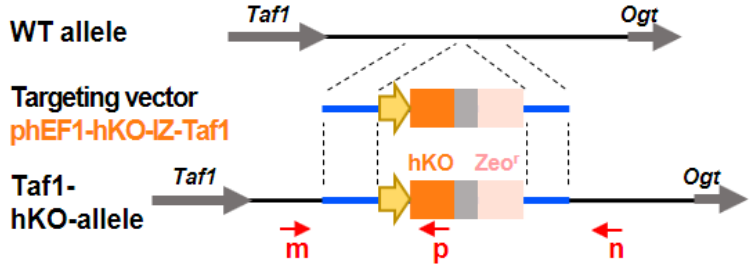
	Number of EGFP$^+$/hKO$^+$ colonies
pX330-hCas9 phEF1-hKO-IZ-Syap1	446
pX330-hCas9-gRNA phEF1-hKO-IZ-Syap1	441

Figure 4. Generation of mESCs that possess EGFP$^+$ and hKO$^+$ at the *S* site. (A) EGFP targeting vector for the *S* site. (B) hKO targeting vector for the *S* site. Each targeting vector carried a fluorescent protein gene and a drug-resistant gene between the homologous sequences in the targeted site. Expression of fluorescent protein coding genes and drug-resistant genes was driven by the EF1α promoter. (C) Scheme for establishing EGFP$^+$/hKO$^+$ mESCs by a step-by-step method. Female mESCs were transfected with the Cas9/gRNA expression vector and EGFP targeting vector. EGFP$^+$ clone with validated EGFP insertion in one allele were used for transfection with the Cas9/gRNA expression vector and hKO targeting vector. Double-positive (EGFP$^+$/hKO$^+$) colonies were collected, and hKO insertion into the other allele was confirmed. (D) (E) Effect of the amount of the targeting vector on the efficiency of generating mESCs with the knocked-in fluorescent protein coding gene. Female mESCs were transfected with various amount of the Cas9/gRNA expression vector. The numbers of EGFP$^+$ colonies were compared among different groups to evaluate the effect of plasmid amounts on the efficiency of knocking-in the EGFP coding gene into mESCs. (F) EGFP$^+$ mESC generation. Two days after transfection, puromycin was added to select for puromycin-resistant mESCs. (G) Colonies with green fluorescence were picked up and PCR analysis was performed to confirm the correct genome editing in these colonies. The positions of the primer sets are shown by red arrows in (A) and (B). The clone with the targeted insertion is marked in red. (H) EGFP$^+$/hKO$^+$ mESC generation. Candidate clones (marked in red color – No. 19) were used for a second transfection with the Cas9/gRNA expression vector and hKO targeting vector. Two days after transfection, zeocin was added to select for zeocin-resistant mESCs. (K) Double-colored mESC clones were collected and analyzed to confirm the insertion of the hKO-coding gene by the indicated primer sets. (L) The number of double-colored colonies that survived zeocin selection after the second transfection. The colony numbers were counted for the mESCs transfected with or without gRNA. Scale bars: 100μm.

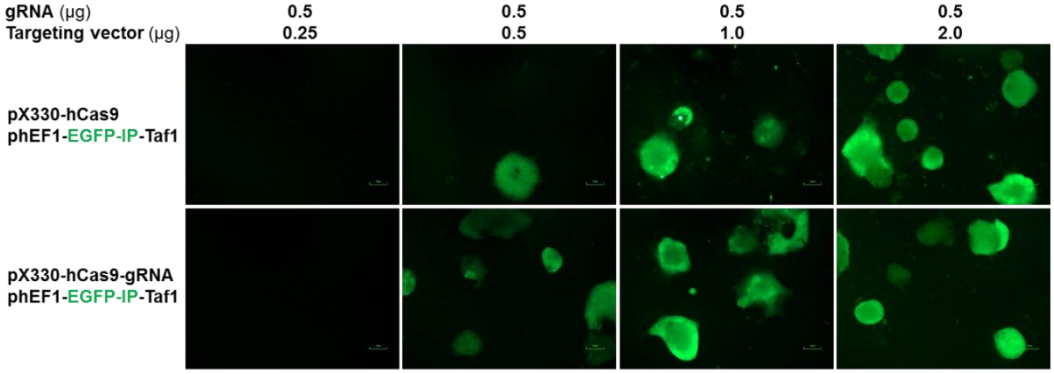
A.



B.



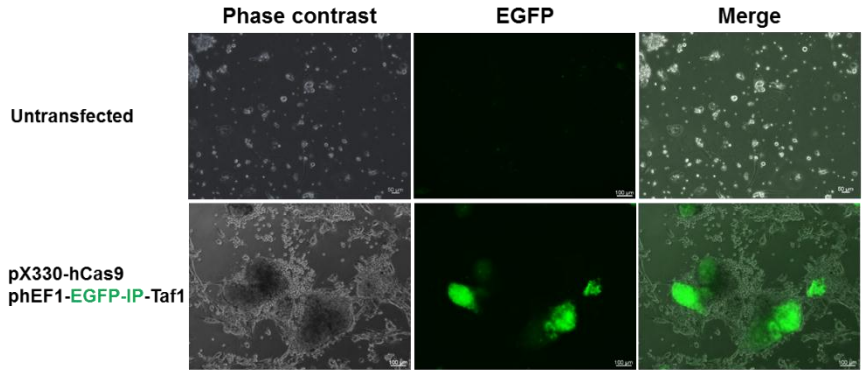
C.



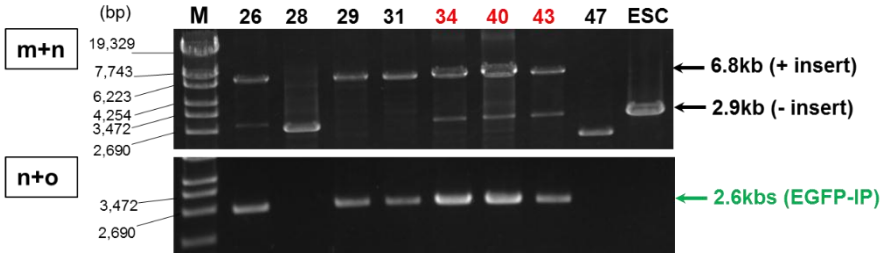
D.

Cas9-gRNA vector (μg) **Targeting vector (μg)**	**0.5** **0.25**	**0.5** **0.5**	**0.5** **1.0**	**0.5** **2.0**
pX330-hCas9- phEF1-EGFP-IP-Taf1	1	20	33	20
pX330-hCas9-gRNA phEF1-EGFP-IP-Taf1	0	52	78	65

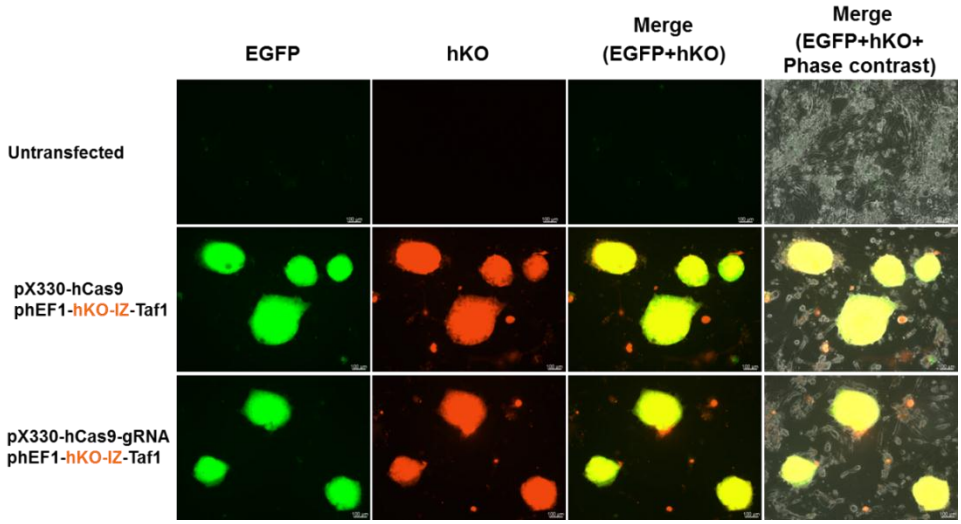
E.



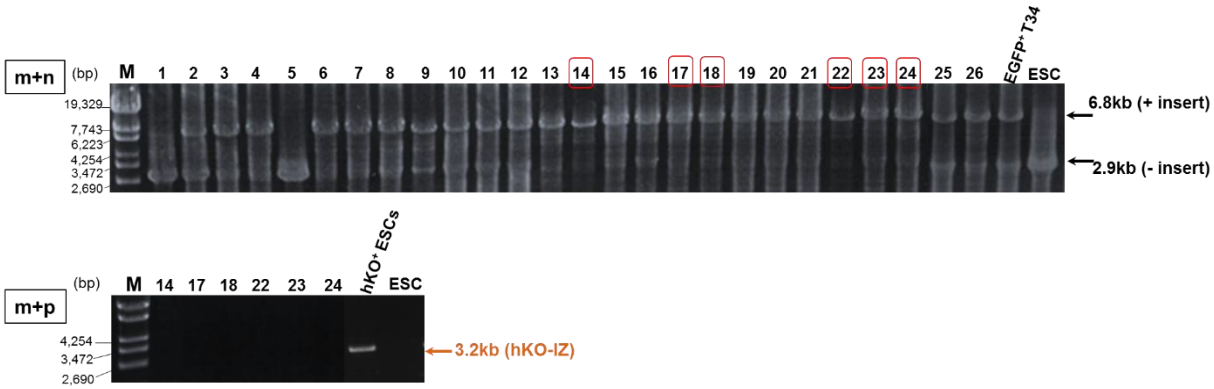
F.



G.



H.



K.

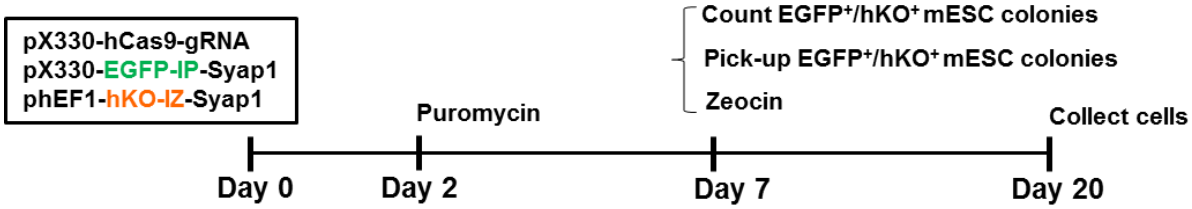
	Number of EGFP$^+$/hKO$^+$ colonies		
	#34	**#40**	**#43**
pX330-hCas9 phEF1-hKO-IZ-Taf1	210	315	205
pX330-hCas9-gRNA phEF1-hKO-IZ-Taf1	187	320	213

Figure 5. Generation of mESCs that possess EGFP$^+$ and hKO$^+$ mESCs at the *S* site. (A) EGFP targeting vector for the *T* site. (B) hKO targeting vector for the *T* site. Each targeting vector carried a fluorescent protein gene and a drug-resistant gene between the homologous sequences in the targeted site. Expression of fluorescent protein-coding genes and drug-resistant genes was driven by the EF1α promoter. (C) (D) Effect of the amount of the targeting vector on the efficiency of generating knocked-in fluorescent protein-coding gene. Female mESCs were transfected with various amount of the Cas9/gRNA expression vector and EGFP targeting vector. The numbers of EGFP$^+$ colonies were compared between different groups to evaluate the effect of plasmid amount on the efficiency of knocking-in the EGFP coding gene into mESCs. (E) EGFP$^+$ mESC generation. Two days after transfection, puromycin was added to select for puromycin-resistant mESCs. (F) Colonies with green fluorescence were picked up and PCR analysis was performed to confirm the correct genome editing in these colonies. The positions of the primer sets are shown by red arrows in (A) and (B). The clones with the targeted insertion were marked in red. (G) EGFP$^+$/hKO$^+$ mESC generation. Candidate clones (marked in red color – No. 34, 40, 43) were used for a second transfection with the Cas9/gRNA expression vector and hKO targeting vector. Two days after transfection, zeocin was added to select for zeocin-resistant mESCs. (H) Double-colored mESC clones were collected and analyzed to confirm the insertion of the hKO-coding gene by the indicated primer sets. (K) The number of double-colored colonies that survived zeocin selection after the second transfection. The colony numbers were counted for the mESCs transfected with or without gRNA. Scale bars: 100μm.

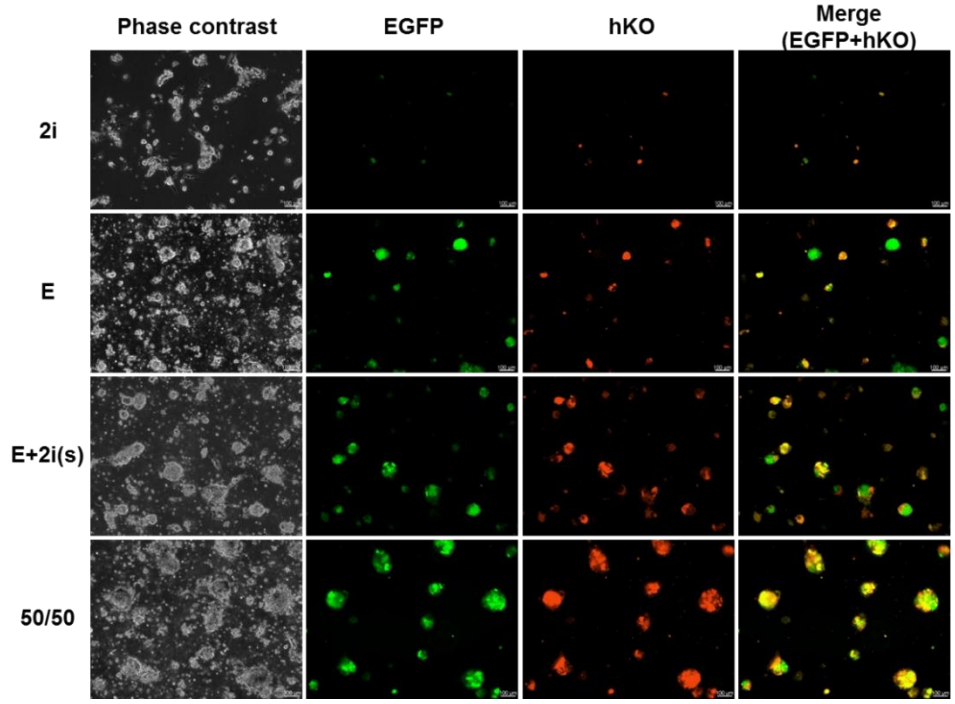
A.

Medium	**2i**	**E**	**E+2i(s)**	**50/50**
Components	- DMEM/F12 - Neurobasal medium - N2 supplement - B27 supplement - LIF - 2 inhibitors (MEK and GSK inhibitors)	- DMEM - KSR - LIF	- DMEM - KSR - LIF - 2 inhibitors	50% 2i medium + 50% E medium

B.



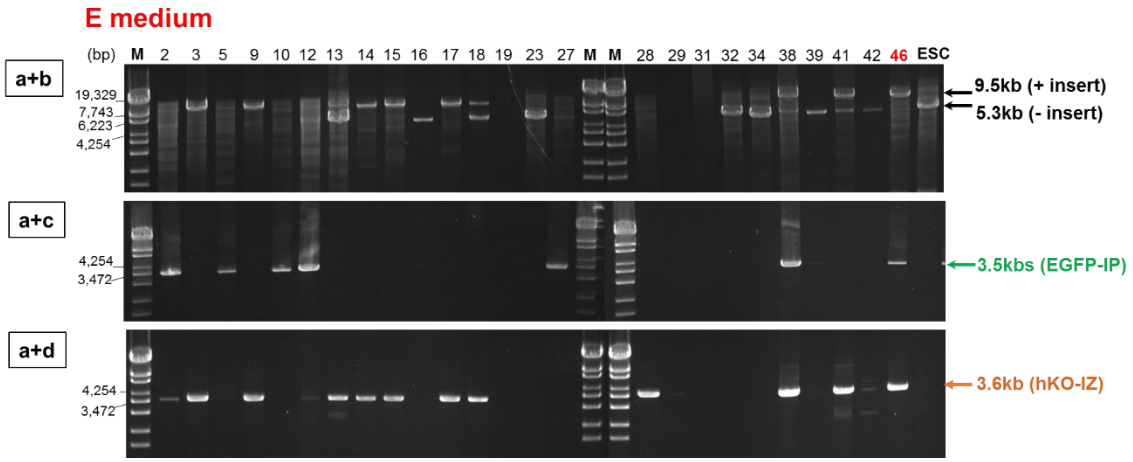
C.



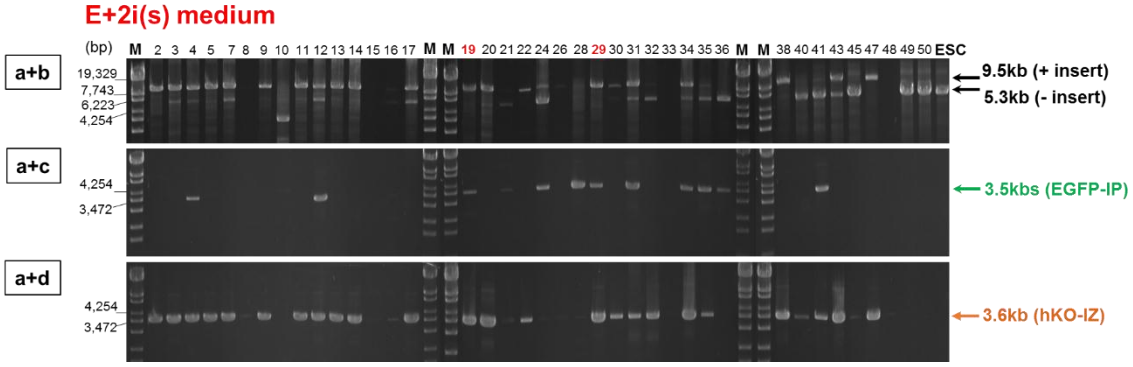
D.

	Number of EGFP$^+$/hKO$^+$ colonies	
	+ gRNA	- gRNA
2i	**20**	0
E	**107**	16
E+2i(s)	**423**	51
50/50	**148**	10

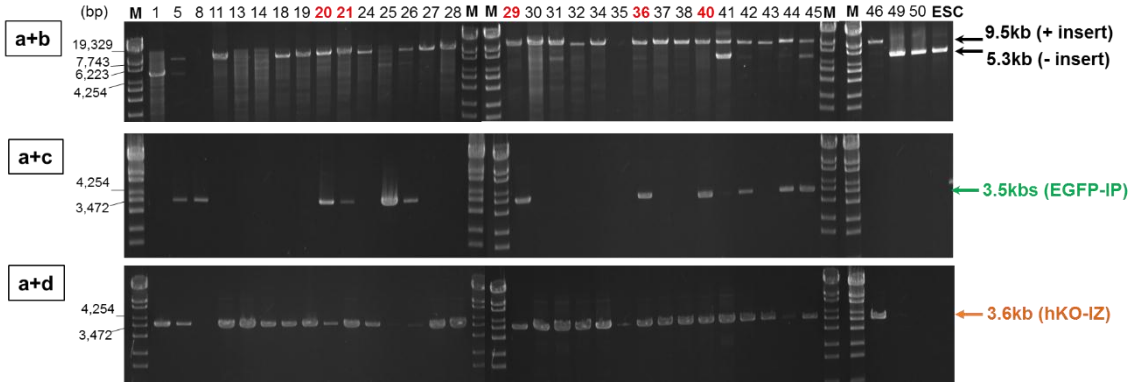
E.



F.



G.



H.

	Picked up colonies	Survival clones	Clones with targeted insertion on both alleles	Targeting efficiency (%)
2i	20	0	0	0%
E	50	25	1	2%
E+2i(s)	50	40	2	4%
50/50	50	34	5	10%

Figure 6. Effect of medium components on generation of EGFP^{+}/hKO^{+} mESCs. (A) Four different media were tested. Components of each medium are listed. (B) Female mESCs were transfected with the Cas9/gRNA expression vector and two different *S* site-targeting vectors. The transfected cells were cultured in different medium and selected sequentially by puromycin and then by zeocin. (C) Morphology and fluorescent pattern of transfected colonies after 4 days of transfection. (D) Number of double-colored colonies after puromycin selection. The colony number were counted on samples transfected with and without gRNA. (E-G) Double-colored colonies were picked up and continuously cultured in these media together with zeocin. Surviving clones were collected and analyzed to validate the correct recombination on the targeted chromosome. The clones were cultured in different medium – E, E+2i(s) and 50-50 medium as shown in (H). Clones with the expected insertion were marked in red. (H) Summary of the targeting efficiency of cells cultured in different medium. Scale bars: 100μm.

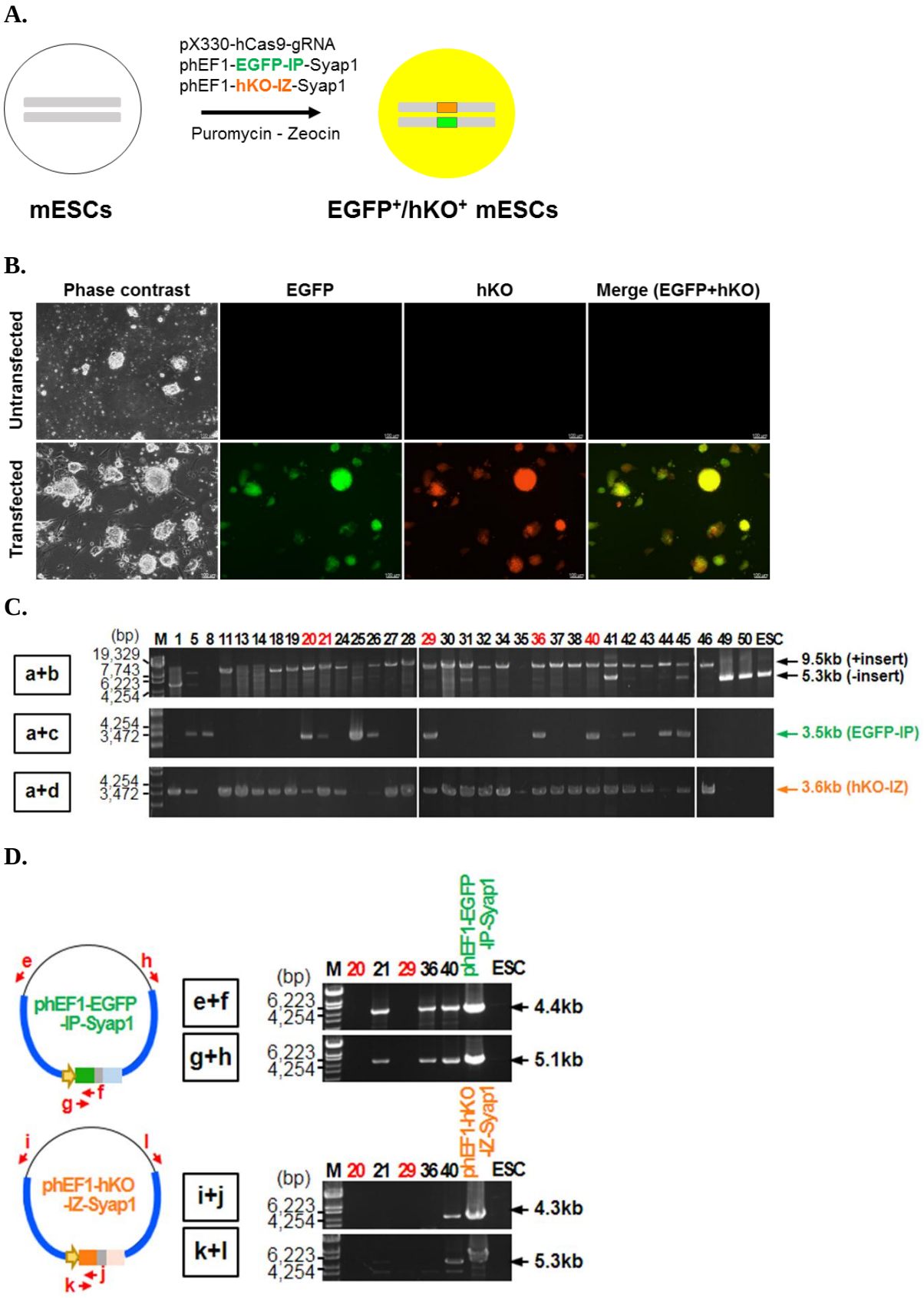


Figure 7. Generations of *S* site EGFP$^+$/hKO$^+$ mESCs by the simultaneous delivery method. (A) Scheme for establishing EGFP$^+$/hKO$^+$ mESCs. Female mESCs were transfected

with the Cas9/gRNA expression vector and two different targeting vectors. The transfected cells were then selected by puromycin and zeocin. (B) Transfected mESC colonies expressed double fluorescence ($EGFP^+/hKO^+$) after drug selection. (C) Genomic PCR analysis of the fluorescent protein gene inserted at the specific site of isolated clones. Primer sets used in analyses are shown in Figs. 4A and 6B. Clones with the expected insertion are marked in red. (D) Confirmation of non-specific insertion. Different primer sets were used to detect random integration of the EGFP or hKO gene into the genome of isolated clones. Scale bars: 100 μm.

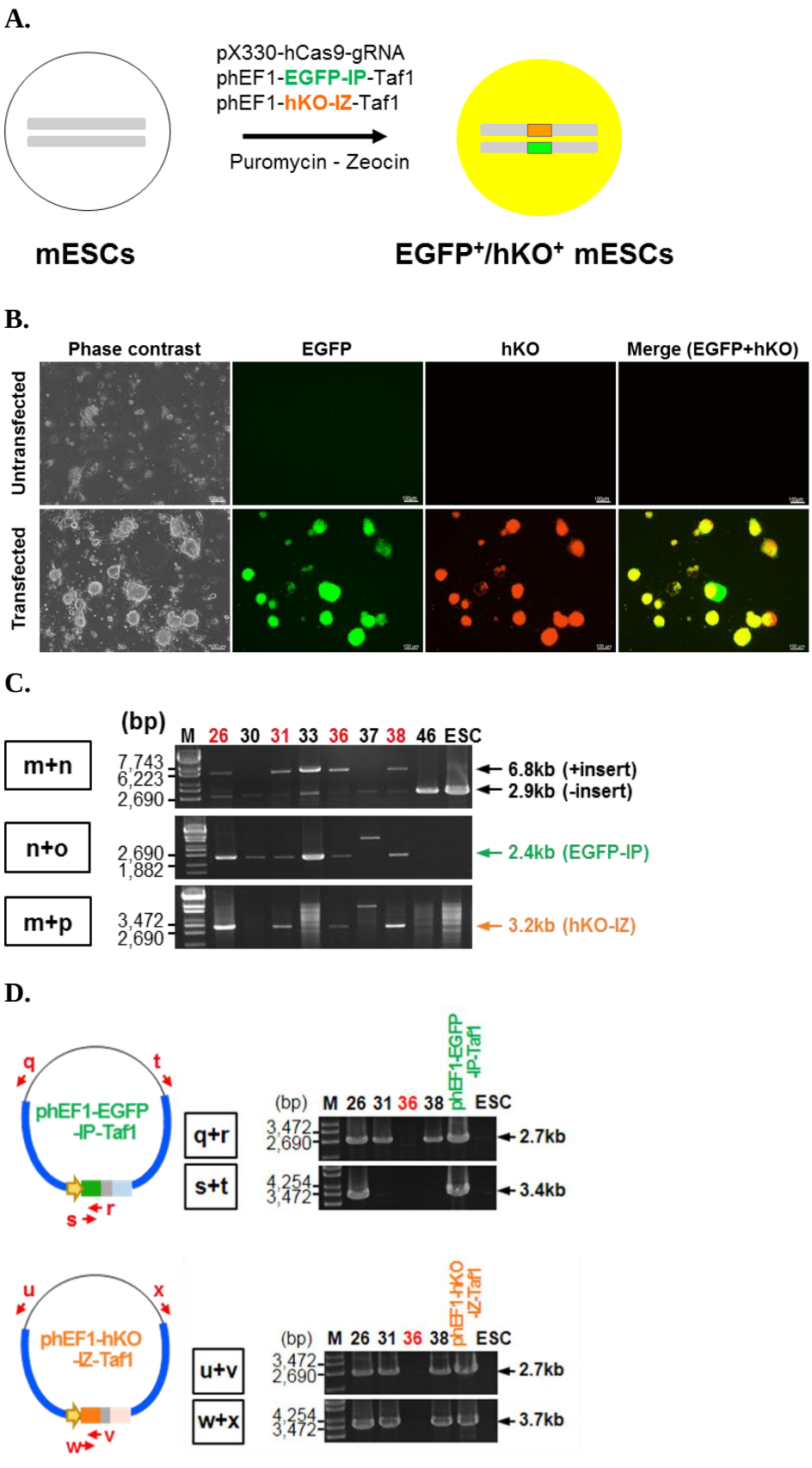
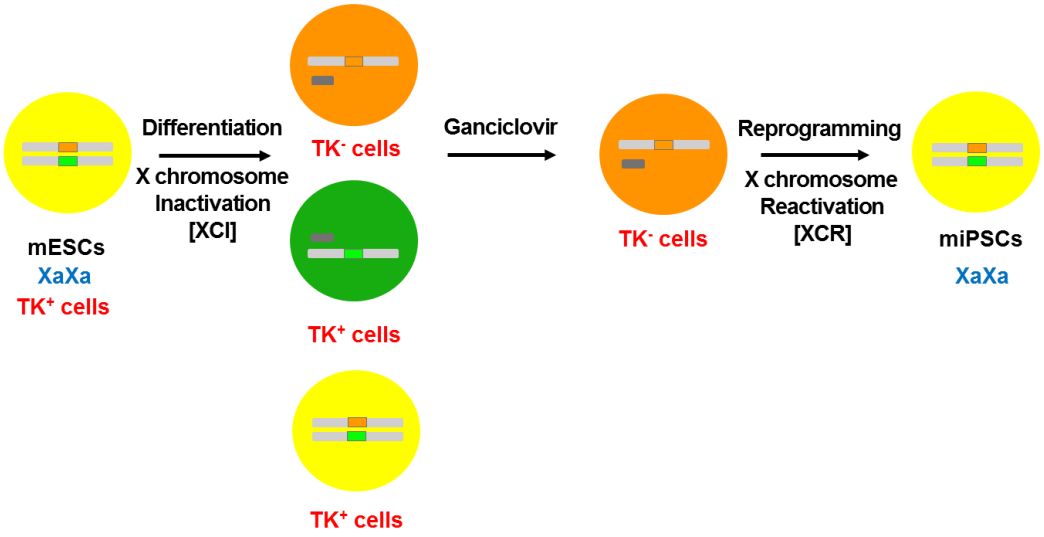


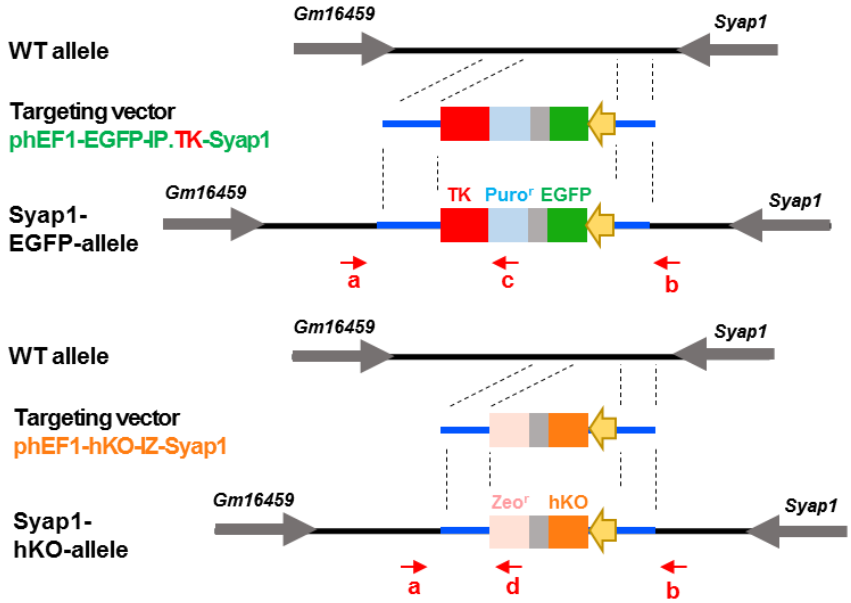
Figure 8. Generations of *T* site EGFP$^+$/hKO$^+$ mESCs by the simultaneous delivery method. (A) Scheme for establishing EGFP$^+$/hKO$^+$ mESCs. Female mESCs were transfected with the Cas9/gRNA expression vector together and two different targeting vectors. The

transfected cells were then selected by puromycin and zeocin. (B) Transfected mESC colonies expressed double fluorescence ($EGFP^+/hKO^+$) after drug selection. (C) Genomic PCR analysis of the fluorescent protein gene inserted at the specific site of isolated clones. Primer sets used in each analyses are shown in Figs. 4A and 6B. Clones with the expected insertion are marked in red. (D) Confirmation of non-specific insertion. Different primer sets were used to detect random integration of the EGFP or hKO gene into the genome of isolated clones. Scale bars: 100 μm.

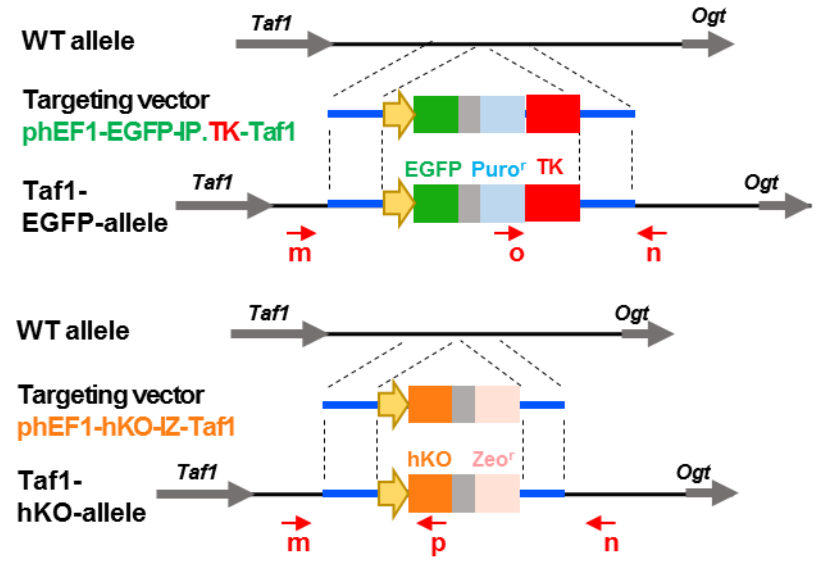
A.


Differentiation
X chromosome Inactivation [XCI]
Ganciclovir
Reprogramming
X chromosome Reactivation [XCR]
mESCs
XaXa
TK+ cells
TK- cells
TK+ cells
TK+ cells
TK- cells
miPSCs
XaXa

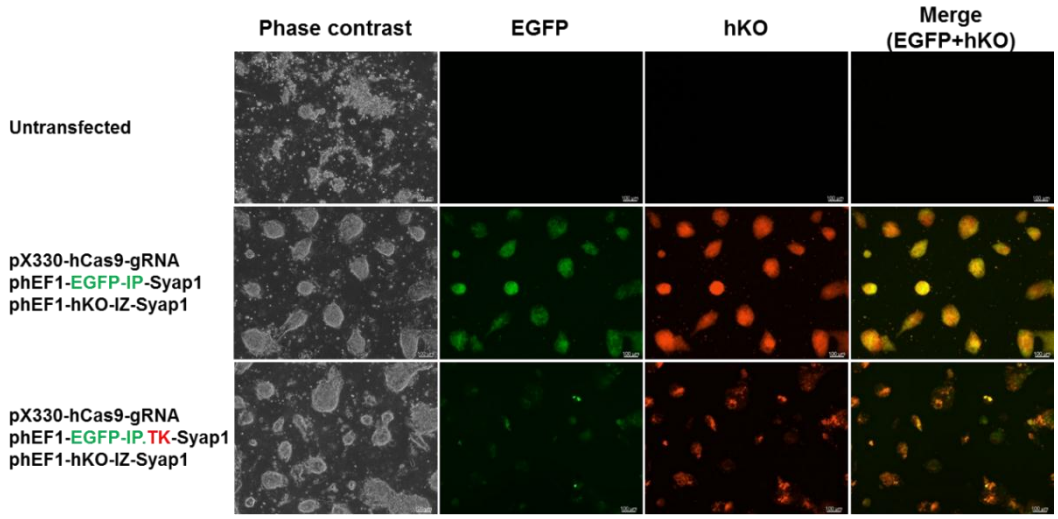
B.


WT allele
Gm16459
Syap1
Targeting vector
phEF1-EGFP-IP.TK-Syap1
Syap1-
EGFP-allele
TK
Puror
EGFP
a
c
b
WT allele
Targeting vector
phEF1-hKO-IZ-Syap1
Syap1-
hKO-allele
Zeor
hKO
a
d
b

C.


WT allele
Taf1
Ogt
Targeting vector
phEF1-EGFP-IP.TK-Taf1
Taf1-
EGFP-allele
EGFP
Puror
TK
m
o
n
WT allele
Targeting vector
phEF1-hKO-IZ-Taf1
Taf1-
hKO-allele
hKO
Zeor
m
p
n

D.



E.

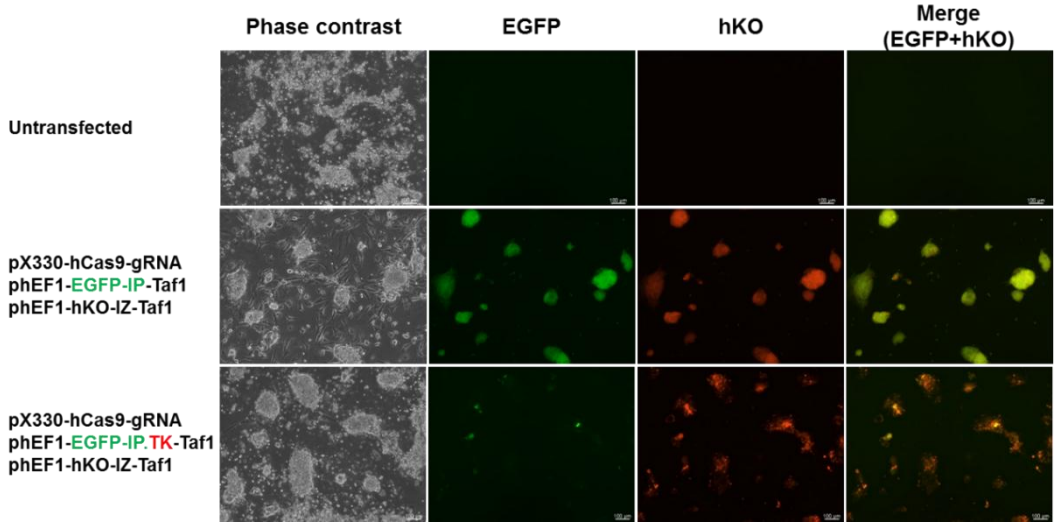


Figure 9. Generation of negative selection marker containing EGFP$^+$/hKO$^+$ mESCs. (A) (A) In order to observe XCI, double-colored mESCs were allowed to differentiate. In a population of differentiated cells, some cells remained undifferentiated. Negative selection (thymidine kinase (TK) selection) was done during differentiation to remove double-colored cells and isolate single-color cells (hKO$^+$ cells), which were then used for further reprogramming. Ganciclovir was added to the differentiated cells to remove TK$^+$ cells and select only hKO$^+$ (TK$^-$) cells for reprogramming. (B) (C) Structure of the thymidine kinase targeting vector for each site. The puromycin resistance gene in the EGFP targeting vector was replaced by the puromycin resistance gene (pEF1-EGFP-IP.TK targeting vectors). (D) Morphology and fluorescent pattern of transfected colonies after puromycin and zeocin selection, at the *S* and *T* site, respectively. Scale bars: 100μm.

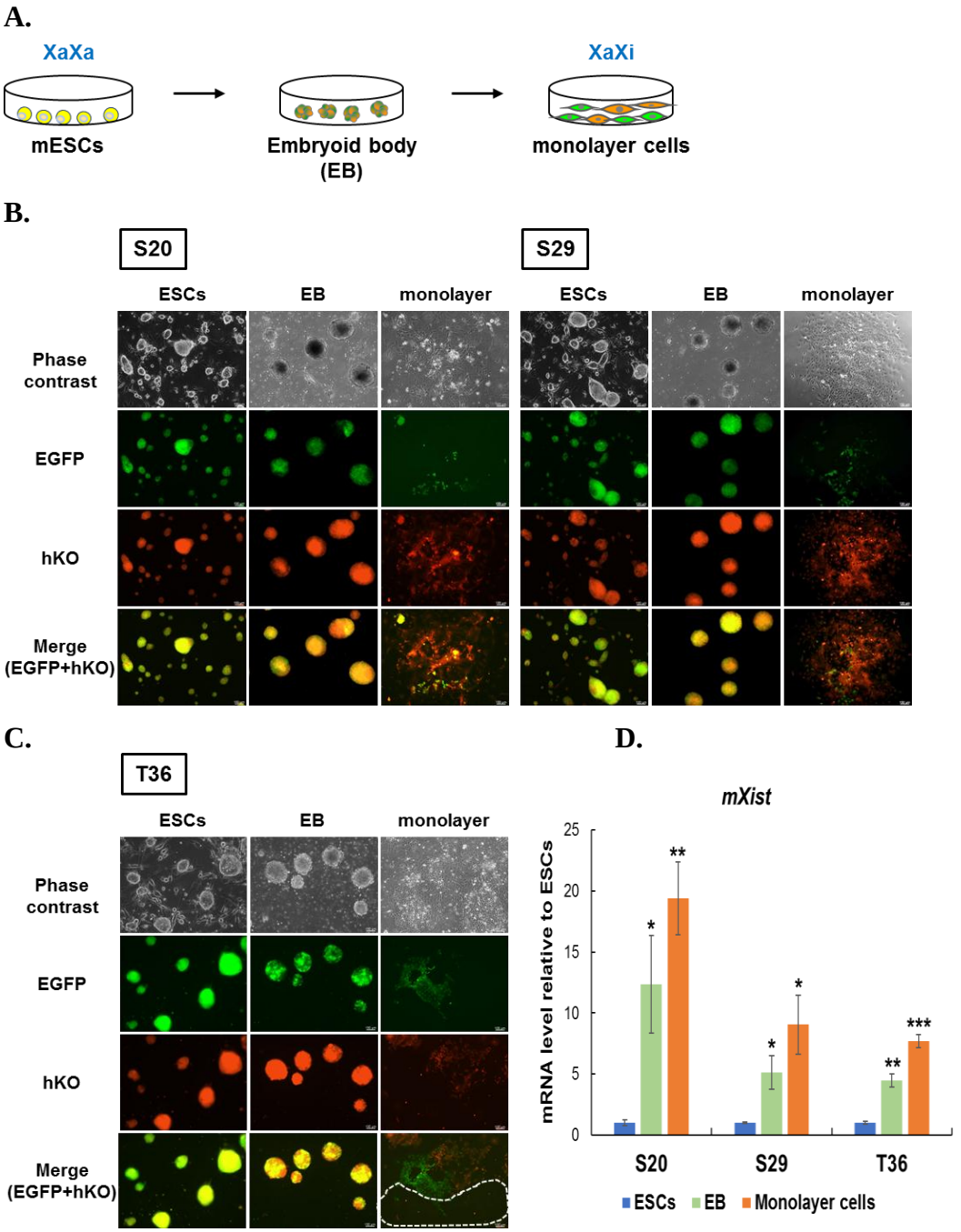
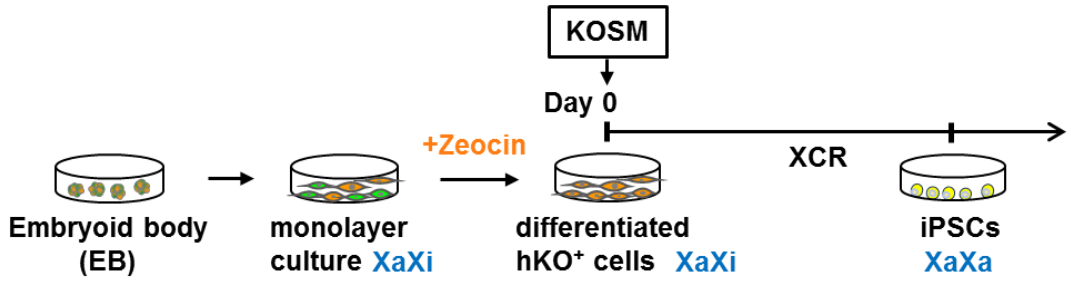


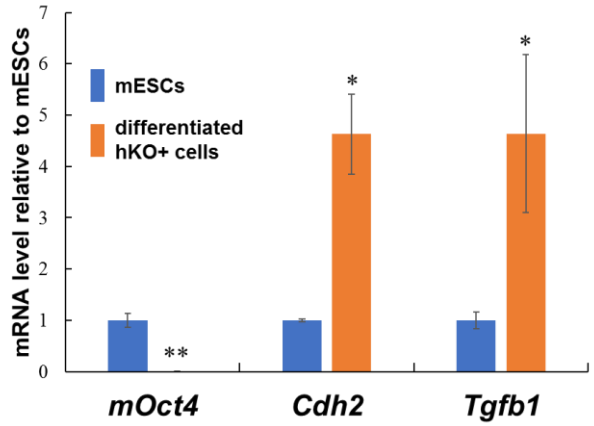
Figure 10. Detection of X chromosome inactivation (XCI) during differentiation. (A) Scheme of ESC differentiation and induction of XCI. mESC clones, which express double fluorescence ($EGFP^+/hKO^+$), were differentiated via embryoid body (EB) formation. At the EB stage, the cells showed a heterogeneous pattern of fluorescent signals. Subsequently, EB-derived cells formed monolayer and ceased to one of the two fluorescent proteins ($EGFP^+$ or hKO^+), indicating X chromosome inactivation (XCI) during differentiation. (B) (C) Fluorescent patterns of selected clones. White dotted line indicates that monolayer cells derived from T36 ESC clone lost expression of both EGFP and hKO (D) *Xist* expression at

each stage of selected clones. Data represent means ±SEM of three biologically independent experiments. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Scale bars: 100µm.

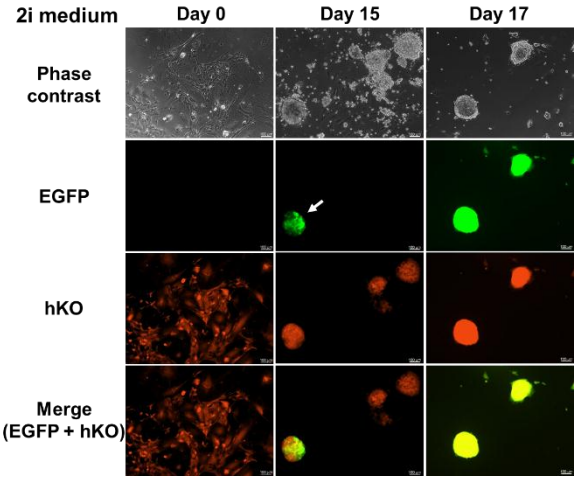
A.


KOSM
Day 0
+Zeocin
XCR
Embryoid body (EB)
monolayer culture XaXi
differentiated hKO+ cells XaXi
iPSCs XaXa

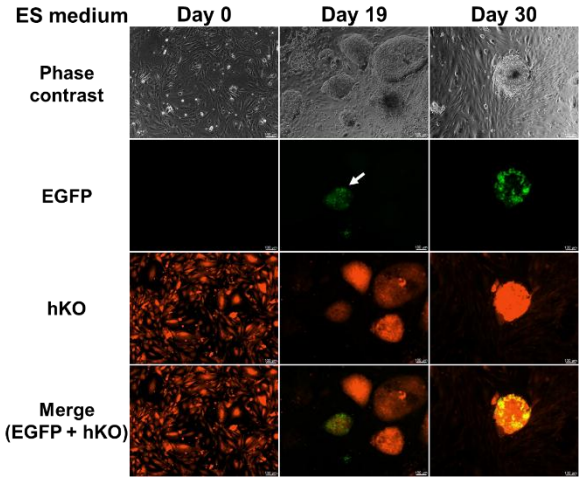
B.


mRNA level relative to mESCs
mESCs
differentiated hKO+ cells
mOct4
Cdh2
Tgfb1

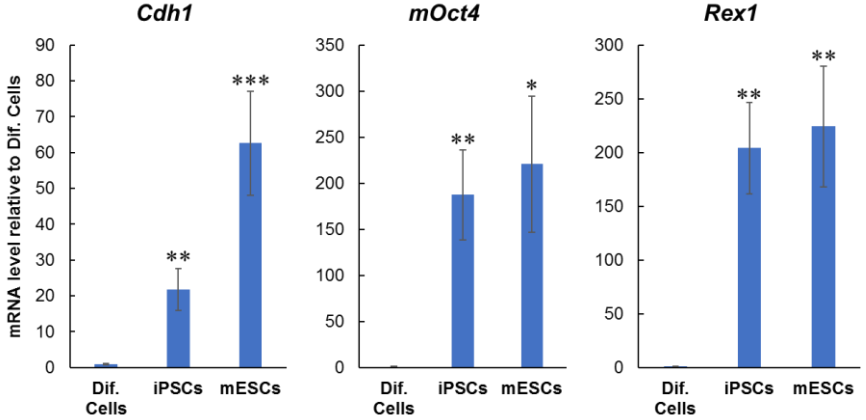
C.


2i medium
Day 0
Day 15
Day 17
Phase contrast
EGFP
hKO
Merge (EGFP + hKO)

D.


ES medium
Day 0
Day 19
Day 30
Phase contrast
EGFP
hKO
Merge (EGFP + hKO)

E.


Cdh1
mOct4
Rex1
mRNA level relative to Dif. Cells
Dif. Cells
iPSCs
mESCs

F.

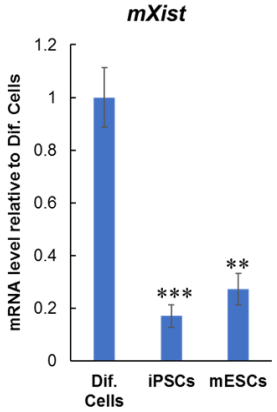

mXist
mRNA level relative to Dif. Cells
Dif. Cells
iPSCs
mESCs

Figure 11. Detection of X chromosome reactivation during reprogramming. (A) Scheme of iPSC generation to detect XCR. The hKO^{+} somatic cells were collected from a population of monolayer cells by zeocin selection. iPSCs were generated from hKO^{+} EB-derived cells by using SeVdp(KOSM). After the start of reprogramming, live cells were monitored continuously by fluorescence to detect XCR. Some of the iPSC colonies displayed double ($EGFP^{+}/hKO^{+}$) fluorescence, indicating that XCR occurred during iPSC generation. (B) Expression of pluripotency (*mOct4*) and differentiation markers (*Cdh2*, *Tgfb1*) in hKO^{+} somatic cells. (C) (D) Effect of different media on detection of X chromosome reactivation. During reprogramming of isolated hKO^{+} differentiated cells, Sendai virus-infected cells were cultured in serum-free medium (2i medium) or standard serum-containing medium with LIF (ES medium). (E) Expression of pluripotency markers (*Cdh1*, *mOct4*, *Nanog*) of generated iPSCs cultured in 2i medium. (F) Expression of *mXist* of generated iPSCs cultured in 2i medium. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Scale bars: 100μm.

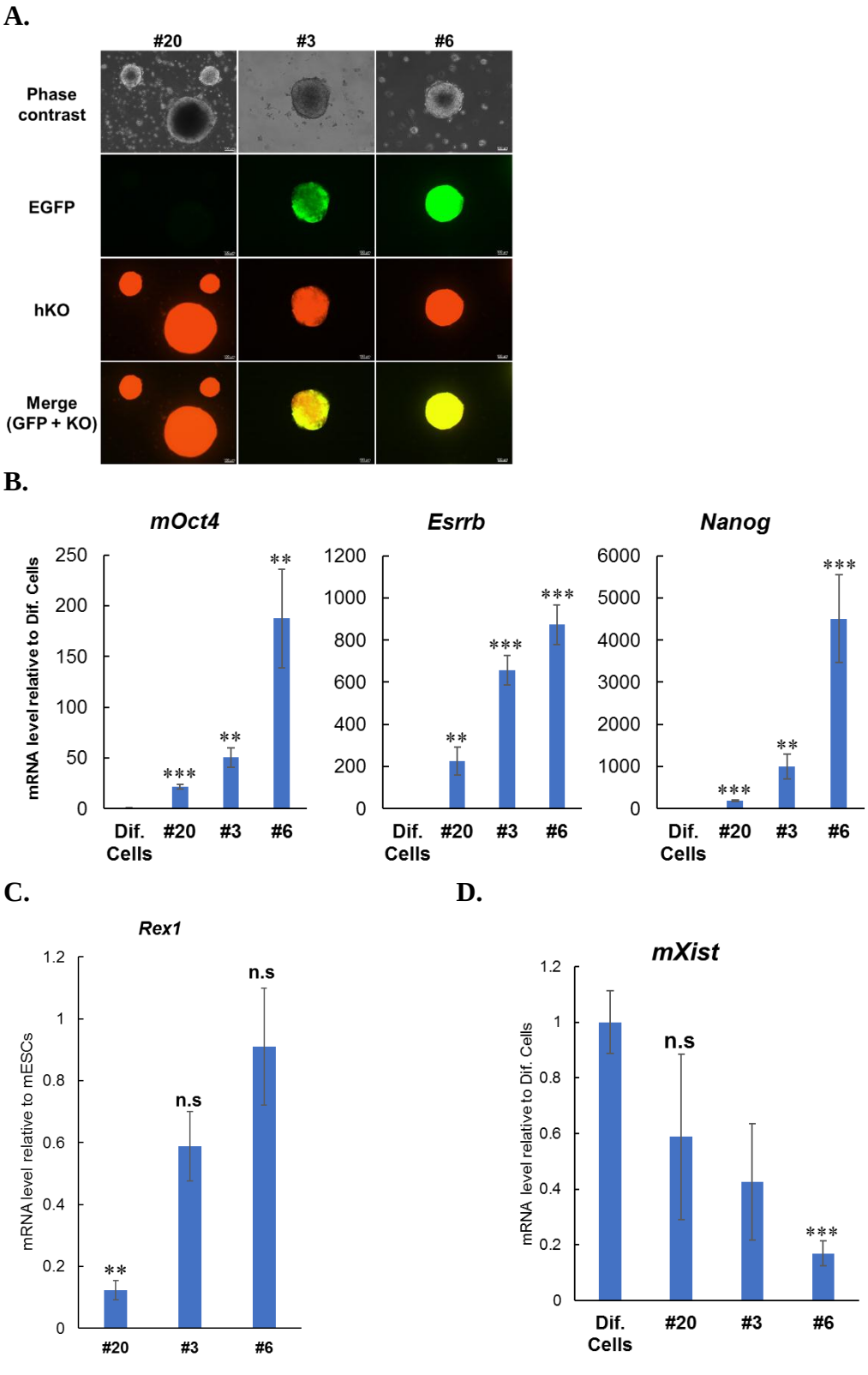


Figure 12. Detection of a heterogeneous population of iPSCs. (A) Different fluorescent patterns of iPSC clones were observed in a population of reprogrammed cells. iPSC colonies with different fluorescent patterns were picked up. Three representative clones were shown (#20, #3, #6). (B) Correlation between the expression levels of pluripotency marker genes

and the patterns of fluorescent protein expression. Representative markers (*mOct4*, *Esrrb, Nanog*) were shown. (D) *Rex1* expression in iPSCs with different fluorescent patterns. (H) mXist expression in iPSCs with different fluorescent patterns. Data represent means ±SEM of three biologically independent experiments. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Scale bars: 100μm.

2.4 Discussion

2.4.1 Generation of EGFP$^+$/hKO$^+$ mESCs

❖ **Effect of gRNA sequence on homologous recombination**

In the current study, my goal was to establish a novel cell line that permits live cell imaging system of XCR during reprogramming and further analyze the mechanism of XCR. For this purpose, I aimed to establish female mESC lines that carry reporter genes on both X chromosomes and chose EGFP and hKO as fluorescent markers. These two fluorescent proteins have distinct emission wavelengths with minimal overlap, enabling identification of cells that express one or both fluorescent proteins under microscopy. These fluorescent proteins should be inserted into the same positions of both X chromosomes, so that the effect of chromosome position on their expression levels could be minimized. This would enable easy identification of XCI during differentiation and XCR during reprogramming at a single-cell level without using any invasive methods.

To do this, several genome-engineering techniques were taken into consideration. Zinc-finger nucleases (ZFN) and transcription activator-like effector nucleases (TALEN) have opened the window for mammalian genome engineering. The system utilizes the DNA sequence-specific binding of each zinc-finger domain that contains specific amino acid residues in it. By combining four zinc-finger domains in tandem, one could make a factor that binds to the specific position within the genome. By attaching a nuclease to this tandem repeats of the zinc finger domains, it is possible to make an enzyme that binds to a specific genome position and induce a targeted DNA double strand breaks (DSB), which stimulates error-prone nonhomologous end joining (NHEJ) or HR. However, difficulties in designing these enzymes and complexity in preparing them prevent its widespread use.

Recently, the CRISPR/Cas9 system has been widely used for genome editing to generate genetically modified organisms or cells to study the function of genes or their regulatory mechanisms. The CRISPR/Cas9 system allows a simple, easy, and inexpensive way to edit the genome with an unprecedented flexibility and specificity. The specificity of targeting the genome is conferred by the sequence of a guide RNA (gRNA), which, as compared to four tandem repeats of zinc finger domains, enables far more flexible and specific targeting. In a genome targeting, it is possible to test different gRNAs and choose the one with the highest efficiency. In my initial experiment, I tested several different gRNAs to find their targeting efficiency. Expression levels of EGFP, which is reconstituted by homologous recombination after gRNA-mediated DNA cleavage, were different among the testes gRNAs. Having a look back at these gRNA sequences, I found that 20-bp genome

target of $gRNA_1$ and $gRNA_5$, which showed the strongest EGFP expression, starts with the base guanine (G). This G base is part of the promoter sequence required for the initiation of transcription of human U6 promoter. Bases other than G at the 5' end of a gRNA may affect transcription efficiency of the U6 promoter, which may lead to low gRNA expression and lower efficiency of gRNA-mediated DNA cleavage.

❖ **Position of knock-in reporter genes**

XCI are known to occur not uniformly throughout the X chromosome, and some X-linked genes are known to escape XCI. Furthermore, spreading of silencing along the X chromosome depends on the three-dimensional conformation of X chromosome during XCI *in vitro* [107]. This suggests that the position of a reporter gene within the X chromosome can influence monitoring of XCI. Therefore, the integration site of the fluorescent genes is important if XCI is to be monitored accurately. In this project, I chose two different sites as integration sites that fulfilled the criteria: 1) the site is within an intergenic region, 2) the surrounding genes are expressed ubiquitously, and 3) the surrounding genes are subject to XCI. I chose two sites close to the *Syap1* or *Taf1* gene and inserted two different fluorescent genes into each allele of the X chromosome. Each site is located in the intergenic region, surrounded by genes that show ubiquitous expression throughout all developmental stages (intergenic regions for insertion are shown in Fig. 2). At the ESC stage, generated *T* site $EGFP^+/hKO^+$ mESC clones showed unstable expression of fluorescence, which indicates that the pluripotency of ES cells may be also unstable despite the indistinguishable morphological features. The *T* mESC clones could not maintain their pluripotency for long-term culture and could not survive well under drug selection, as compared with the *S* clones. It might be that a gene inserted at the *T* site may be subject to silencing independent of XCI. It could also be possible that the *T* site is subject to XCI earlier than the *S* site since the *T* site is closer to the *Xist* gene or X chromosome inactivation center (XIC). Moreover, I observed half of cells express neither EGFP nor hKO gene in the *T* clones during differentiation. This may indicate the intrinsic instability of expression of the gene at the *T* site.

❖ **Delivery methods of reporter genes**

In this study, I tested two kinds of strategies to introduce desired inserts into X chromosomes. The first strategy is a step-by-step method, by which mESC clones harboring the EGFP gene in one allele of X chromosome were isolated and then used to insert the hKO gene into the other allele. The second strategy is a simultaneous delivery method, by which two different fluorescent protein-coding genes were inserted into mESCs simultaneously in a single transfection experiment. In the first strategy, hKO insertion can be identified by

observing double-colored colonies after the second transfection. PCR analysis with primers detecting the full length of insertion showed that the initial EGFP heterozygous mESCs became EGFP homozygous ones instead of having both EGFP and hKO, indicating that the hKO gene could not be inserted into the second allele at either *T* or *S* site. This phenomenon may be because the homologous recombination between two X chromosomes occurs more readily than that between the X chromosome and the donor template plasmid: The second allele, which is the uninserted allele after the first transfection, tends to use the edited allele, where the EGFP has been inserted, as a template for homology-based repair rather than the free hKO donor template plasmid. The much longer homologous regions and proper alignment of two X chromosomes may prevent the donor template plasmid to function as a repair template despite the much higher copy number of the donor template. It may also explain the much higher number of single-colored mESCs possessing only EGFP or hKO on both X chromosome. In any case, I succeeded in generating $EGFP^+/hKO^+$ mESCs using the second strategy. Thus, it is recommended to use the simultaneous delivery method when two different fluorescent genes are inserted into the same locus of two chromosomes.

❖ **Effect of cell culture medium on genome editing efficiency**

Different stem cell laboratories rely on different methods to support expansion and maintenance of mESCs. Culture of stem cells under inappropriate conditions can induce their differentiation and reduce reproducibility of experiments. In recent years, new ES cell culture protocols, using better-defined conditions, have been published. 2i medium containing two small molecule inhibitors (CHIR99021 and PD0325901) is proven to be the standard condition for culturing ESCs. However, 2i medium renders ESCs more resistant to transfection, perhaps due to the densely packed ESCs within colonies formed under this condition. Another drawback of 2i medium is its high cost. Therefore, I tested the transfection efficiency, using different media (2i medium; E medium: DMEM+KSR+LIF; E+2i(s) medium: E medium + 2 inhibitors; 50/50 medium: combination of 2i and E medium). In the experiments targeting the *S* site, ESCs responded differently to different media. The numbers of double-colored colonies after transfection varied between different media. 2i medium seems to prevent random integration of the targeting vector, which occurs independent of gRNA, outside the targeted *S* site. However, this medium was too severe for the transfected cells to survive, and most of the cells died. When E+2i(s) was used, by contrast, a higher number of double-colored colonies were obtained. However, it resulted in a higher percentage of colonies with random integration. When the 50/50 medium was used, the targeting efficiency was higher than those observed when other media were used.

Observation of the cells during transfection and selection, the 50/50 medium supported reasonably good growth of transfected colonies while maintaining round shapes and fluorescent signals. The combination of the 2i medium and E medium may have retained the properties of both media; to maintain ESCs in the naïve state by the 2i medium and to support fast cell expansion by the E medium.

❖ **Effect of reporter genes' structures**

Previous studies showed that the CAG promoter inserted in the X chromosome is subjected to XCI and completely inactivated at post-implantation stages [81], showing that the CAG promoter does not interfere with XCI and thus can be used for monitoring XCI using a CAG promoter-driven gene. Different promoters have different abilities to express the reporter gene depending on the stage of development and type of cells [108]. The EF1α promoter is known to show high transcriptional activities in either ESCs or EBs, and in my project, I tested the EF1α and CAG promoters to drive the expression of reporter genes. In female mESCs, the EF1α promoter appeared to show a stable activity than the CAG promoter (data not shown). Therefore, I used the EF1 promoter to express the inserted reporter genes on the X chromosome.

In order to collect single-color differentiated cells for further monitoring XCR during reprogramming, I differentiated $EGFP^+/hKO^+$ mESCs into monolayer cells. Upon differentiation, the cells expressed either green or orange fluorescent signal. However, a population of differentiated cells contained undifferentiated double-colored cells. Since these double-color cells grew faster than single-color ones, they may obscure XCR observation during reprogramming. To overcome this, the thymidine kinase (TK) gene was placed in the vector together with the puromycin resistance gene in the EGFP targeting vector (EGFP-IP.TK). During differentiation, negative selection with Ganciclovir was attempted to remove TK^+ cells so that only hKO^+ differentiated cells (TK^- cells) could be selected for reprogramming. Unexpectedly, female mESCs transfected with the EGFP-IP.TK targeting vector showed a weak fluorescent signal and could not survive under puromycin and zeocin selection after transfection. Knock-in of the TK gene into mESCs might be harmful for some unknown reasons.

2.4.2 3S reprogramming system for analyzing mechanism of XCR

Several reports indicated that the efficiency of reprogramming, occurrence of XCR as well as the characteristics of iPSCs are influenced by the expression levels and stoichiometry of reprogramming factors. Our group has developed a stage specific reprogramming system

(3S system), which generated a series of iPSC populations with distinct degrees of reprogramming. A Sendai virus vector termed SeVdp (fK-OSM), in which KLF4 is tagged N terminally with a destabilization domain (DD), facilitates degradation of KLF4 and decreases its expression level. Because a small chemical termed Shield1 regulated the degradation of KLF4 by the DD, adding different concentrations of Shield1 to cell culture medium allows generation of paused iPSCs at specific time points of reprogramming in a reproducible and predictable manner. The paused iPSCs are stable and retain the ability to resume reprogramming after a long period time in the cell culture. According to our group's published paper, the pause iPSCs are relatively homogeneous and phenotypically stable, and free from silencing of transgenes [109]. Thus, this system has a significant advantage for further analyzing the process of reactivation of X chromosome in iPSC generation, using the differentiated cells obtained from double-colored ESCs.

2.4.3 XCR and acquisition of pluripotency

Around day 15 after SeV infection, $EGFP^{+}/hKO^{+}$ colonies started to appear from $Xa^{hKO}Xi^{EGFP}$ differentiated cells that were infected with SeVdp(KOSM)., showing that XCR occurs at this time point. As colonies showed different patterns of fluorescence expression, I used RT-qPCR to investigate expression of pluripotency marker genes and Xist expression and found a correlation between XCR and pluripotency. The cells that have undergone complete XCR (homogenously double signal) showed higher expression of pluripotency marker genes. This result provides evidence that tracking XCR could help to distinguish between low and high qualities of iPSCs.

In 2014, Pasque *et al* showed XCR is a very late event of reprogramming that occurs after *Xist* RNA coating has disappeared and occurs only after the expression of *Nanog.* Later in reprogramming, $Nanog^{+}$ cells showed biallelic expression of X-linked genes, which indicate XCR. In Pasque's experiment, they used FISH analysis of single cells to detect expression of the Xist RNA and chose some specific genes on the X chromosome. There are thousand genes along the X chromosome, and therefore, whether only detecting the reactivation of some specific genes is sufficient to indicate reactivation of the whole X chromosome. If XCR does not occur simultaneously at these genes, it is possible that some genes can be reactivated before or after the acquisition of pluripotency. Therefore, to understand the mechanism of XCR more clearly and its relationship to the acquisition of pluripotency, it is essential to investigate where XCR initiates and how XCR spreads along the entire X chromosome.

As mentioned above, different stem cell laboratories depend on different methods to expand and maintain mESCs or pluripotent stem cells. Culture of stem cells under different conditions can elicit different states of XCI or XCR. In my own experiments, two different media (2i medium and standard serum-containing medium with LIF) were used to culture iPSCs during reprogramming. Appearance of XCR was obviously affected by the choice of medium. In the 2i medium, initiation of XCR occurred around day 15 after SeVdp(KOSM) infection and soon proceeded to completed XCR with appearance of many colonies with homogenous double-color fluorescence. However, in the standard serum-containing medium with LIF, initiation of XCR was delayed as compared to cells culture in the 2i medium, and these colonies showed only heterogeneously double-colored fluorescence. Moreover, a compact and round shape of colonies is a well-known marker for distinguishing naïve and primed pluripotent stem cells. Serum-free medium (2i medium) maintained reprogrammed iPSC colonies in rounder shape than serum medium. The use of fluorescence as an indicator for XCR and high-quality iPSCs is not only convenient but also reveals a difference between different media and iPSC colonies, which otherwise would be very difficult to observe.

2.4.4 Future application of XCR research

My live imaging approach of ESCs/iPSCs will help us to follow cell derivation and observation of the X chromosome status, which changes during culture of these cells *in vitro*. With the precise observation of the X chromosome status *in vitro*, it is possible to investigate the mechanisms of XCI and XCR *in vivo*.

During normal development, only early embryonic cells and germ cells undergo XCR, and this process is a rare event *in vivo* [39]. However, abnormality of XCI increases with aging and in cancer cells, presumably due to the improper maintenance of XCI [68] [110] [111], and reactivation of X chromosome might occur in other events as well. It is worth evaluating whether cells undergo XCR during development, aging, tumorgenesis, and dedifferentiation in a whole organism. It would be interesting to explore the prevalent characteristics between newly discovered cells undergoing XCR, as they could be a new type of pluripotent stem cells. Analysis of these cells enable us to clarify the mechanism and biologically important role of the reactivation of inactive X chromosome *in vivo*.

In the ICM of a human embryo, XCR is also observed, suggesting that it could be used as a marker to distinguish naïve PSCs from primed PSCs [112]. Therefore, my live cell imaging system of XCR enables us to improve culture conditions for human naïve PSCs. My established system also contributes in analyzing genomic reprogramming which would help

us to investigate the biology of naïve cells in order to maintain its pluripotent state stably. Furthermore, it also enables us to clarify or establish an effective system to convert primed human ESCs to of naïve iPSCs.

In summary, my work establishes a novel system for visualizing the X chromosome status in live cells *in vitro*. Generated female $EGFP^+/hKO^+$ mESC lines can be utilized for tracking XCI upon differentiation and/or monitoring XCR during reprogramming. The detection system of XCR during reprogramming provides a simple method for isolating high-quality iPSCs, which will be a promising material for regenerative therapy research. My data suggest that this system also provides an important tool for tracing the reprogramming process and clarifying how XCR participates in this process. It also will help to clarify how to convert/rescue primed cells to naïve cells. This clarification will help me to effectively establish a novel method to convert primed human ESCs to naïve iPSCs. Although my initial results have just pointed out the time point of reactivation, this finding enables scientists to move closer to understand the molecular mechanism that underlies reprogramming. Moreover, establishment of cell lines which were generated by using a new technique CRISPR/Cas9 has built a new storage of cell lines which are useful in XCI or XCR related experiments.

Chapter III. Conclusions and Perspectives

3.1 Conclusions

My work reveals a novel system for visualizing X chromosome status in live cells *in vitro*. Generated female $EGFP^+/hKO^+$ mESC lines can be utilized for tracking XCI upon differentiation and/or monitoring XCR during reprogramming.

My data also pointed out correlation between XCR and pluripotency. The completed XCR iPSCs also expressed higher level of *Rex1* than partial XCR iPSCs. *Xist* expression gave further evidence on X chromosome status in different pattern iPSCs with different state of XCR.

3.2 Perspective

The detection system of XCR during reprogramming provides a simple method for isolating high quality iPSCs, which will be promising materials for regenerative therapy research. My data suggest that this system also provides a powerful tool for tracing the reprogramming and will help to clarify how XCR participates in this process. It also will help to clarify how to convert/rescue primed cells to naïve cells. Clarification of reprogramming process, including conversion from primed to naïve state, will help in the effective establishment of naïve human iPSCs from primed ESCs. Although my initial results have just pointed out the time point of reactivation, this finding enables scientists move closer steps to understand molecular mechanism that underlies reprogramming. Moreover, establishment of cell lines which are generated by using new technique CRISPR/Cas9 for the first time has built a new storage of cell lines which are used in XCI or XCR related experiments.

List of tables

Table 1. Guide RNA sequences used in this study

	Sequence (5' → 3')	**Site in X chromosome**
gRNA1	CCGGGCCAGCGGGTATGCAG	*Syap1*
gRNA2	GGGGGTTAGAGAGAATAGTG	
gRNA3	CCTGACGTCCACACATGGGG	
gRNA4	TTTGGTGTCTGCAGATCGAA	*Taf1*
gRNA5	GTCATGGGGTCCCATTACCG	
gRNA6	CCGACTATTGGGAGCCATTA	

Table 2. Location of the intergenic regions in X chromosome used in this study

pCAG-EGxxFP

Site	**Location (length)**
Syap1 (gRNA1, gRNA2)	162850793 - 162852499 (1707bp)
Syap1 (gRNA3)	162853330 - 162853981 (652bp)
Taf1 (gRNA4 ~ gRNA6)	101625420 - 101626319 (900bp)

Targeting vectors

Vector name	**Location (left arm)**	**Location (right arm)**
phEF1-EGFP-IP-Syap1 phEF1-hKO-IZ-Syap1	162851428 - 162853765 (2338bp)	162853781 - 162856225 (2445bp)
phEF1-EGFP-IP-Taf1 phEF1-hKO-IZ-Taf1	101625331 - 101626185 (855bp)	101626196 - 101627027 (832bp)

Table 3. Primer sets used in genomic PCR

Primer name	Sequence (5'→3')
a	AGGTCTCATCACGTAGCTCTGTCTTGCAACTC
b	CGCCATCACTGCCCAGCTATCTCCCAC
c	ACCTCCGCGCCCCGCAACCTCCCCTTCTAC
d	GAGTTCTGGACCGACCGGCTCGGGTTCTC
e	AGGCCCTCCGCCATCTTCTGAAGCTGAATC
f	TCTCGTTGGGGTCTTTGCTCAGGGC
g	GCCCTGAGCAAAGACCCCAACGAGA
h	GGTTTCGCCACCTCTGACTTGAGCGTC
i	AGGCCCTCCGCCATCTTCTGAAGCTGAATC
j	ATCTTCTTGGCGGCCTTGTAGGTGGTCTTGAAC
k	TGAGCGTGATCAAGCC
l	GGTTTCGCCACCTCTGACTTGAGCGTC
m	CTAGAGCAAAGAAGACTGTGGGTCAGGTCCCCTC
n	CTCCCTCCTCTGTTTCTTAATGTCAGCTCATGCAG
o	ACCTCCGCGCCCCGCAACCTCCCCTTCTAC
p	ATCTTCTTGGCGGCCTTGTAGGTGGTCTTGAAC
q	CCAAACTCATCAATGTATCTTATCATGTCTGGATCTG
r	TCTCGTTGGGGTCTTTGCTCAGGGC
s	GCCCTGAGCAAAGACCCCAACGAGA
t	GGTTTCGCCACCTCTGACTTGAGCGTC
u	TGTCCAAACTCATCAATGTATCTTATCATGTCTGG
v	ATCTTCTTGGCGGCCTTGTAGGTGGTCTTGAAC
w	TGAGCGTGATCAAGCC
x	GGTTTCGCCACCTCTGACTTGAGCGTC

Table 4. Primer sets used in real-time PCR

gene name		Sequence (5'→3')
γ-tubulin	Forward	CGGACCTGTCGCCAGTTT
	Reverse	TGCGGAACTGCTCCATGA
Cdh1	Forward	ACGTCCCCCTTTACTGCTG
	Reverse	TATCCGCGAGCTTGAGATG
Cdh2	Forward	ATCAACCCCATCTCAGGACA
	Reverse	CAATGTCAATGGGGTTCTCC
Esrrb	Forward	TGGCAGGCAAGGATGACAGA
	Reverse	TTTACATGAGGGCCGTGGGA
Nanog	Forward	ACCTGAGCTATAAGCAGGTTAAGAC
	Reverse	GTGCTGAGCCCTTCTGAATCAGAC
Oct4	Forward	CTGTTCCCGTCACTGCTCTG
	Reverse	AACCCCAAAGCTCCAGGTTC
Rex1	Forward	TTGATGGCTGCGAGAAGAG
	Reverse	ACCCAGCCTGAGGACAATC
Tgfb1	Forward	TGAGTGGCTGTCTTTTGACG
	Reverse	GGCTGATCCCGTTGATTTC
m*Xist*	Forward	GGTTCTCTCTCCAGAAGCTAGGAAAG
	Reverse	TGGTAGATGGCATTGTGTATTATATGG

Table 5. Summary of established EGFP$^+$/hKO$^+$ clones

Clone	*S* site		*T* site
	S20	**S29**	**T36**
ESCs	Express EGFP and hKO fluorescence.	Express EGFP and hKO fluorescence.	Express EGFP and hKO fluorescence.
Monolayer cells	- Express either EGFP or hKO fluorescence. - Suitable for XCI observation.	- Express either EGFP or hKO fluorescence. - Suitable for XCI observation. - Collect hKO$^+$ cells for reprogramming.	- Express either EGFP or hKO fluorescence. - Some cells lost expression of both EGFP and hKO. - Maybe unsuitable for XCI observation.
iPSCs	(not performed reprogramming yet)	Express EGFP and hKO fluorescence.	(not perform reprogramming)

Acknowledgement

During my four years studying in Japan, there are many individuals and organizations who have supported me and to whom I would like to express my gratitude.

First, I would like to express my gratefulness to my supervisor, Professor Koji Hisatake, for the kind support during my Doctoral course in University of Tsukuba. You chose me and gave me opportunity to come here. Thank you for your tremendous efforts on training my logical thinking and valuable advice on conducting research. You did reprogram me.

I also would like to give my sincere gratitude to my mentor, Associate Professor Aya Fukuda. Big Mama, thank you for always providing me your time and patience. From the beginning time of my study till now, you have been given me much valuable advice in doing research. Thank you for teaching me skillful techniques and carefulness in each experiment design.

I would like to thank Associate Professor of University of Sciences, Dang Thi Phuong Thao, who always follows and supports me kindly since I was an undergraduate student. You are the one who guided me through my first steps in science. Thank you for your passion and transferring it to our later generations.

To fulfill this dissertation, I would like to thank to the thesis committee members, Professor Fumihiro Sugiyama, Professor Satoru Takahashi, Associate Professor Takashi Matsuzaka, and Associate Professor Masafumi Muratani for the kind comments and important advice.

I would like to thank to Associate Professor Ken Nishimura and Assistant Professor Yohei Hayashi for your kind supports during my experiments and publication. Your advices have gained a lot of new ideas in my research.

I greatly appreciate all members of Gene Regulation Laboratory for your friendships. Kato, Emi, Michie, Ryota, Chen, Kaisar, Phuong Linh, Ai-chan, Shihu, Yuya, Kei, Anh, Jenny, Arun, Miho, Rie, Hsang Hsang, Sakuragi, Norie, Tomoko… Thank you for going together with me through many ups and downs. I deeply treasure invaluable memories we have spent together. I will remember your encouraging Japanese words - もっともっとがんばります and もうちょと。

I would like to appreciate to the Japanese Government for providing me the MEXT

Scholarship for supporting my study and life in Tsukuba.

I also would like to thank my Vietnamese and international friends who are besides and encourage me during the time staying here. Your hospitality and kindness let me know that I'm loved and warm me through blue days.

My deepest gratitude goes to my families who are always besides, unconditionally loving and giving me tireless supports. Thank you for always believing me, understanding and supporting my study and work. Thank you for keeping me calm whenever I'm frustrated and doubtful about my future. You are my endless energy source.

Last but not least, I would like to send my apologies to all. I know sometimes I made mistakes that bothered you or made you uncomfortable. You all forgave me and taught me the rights from wrongs. Thank you for always forgiving me and let me become more mature in life.

University of Tsukuba, July 13th, 2018

Tran Thi Hai Yen

References

1. Evans, M.J. and M.H. Kaufman, *Establishment in culture of pluripotential cells from mouse embryos.* Nature, 1981. **292**(5819): p. 154-6.
2. Martin, G.R., *Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells.* Proc Natl Acad Sci U S A, 1981. **78**(12): p. 7634-8.
3. Thomson, J.A., et al., *Embryonic stem cell lines derived from human blastocysts.* Science, 1998. **282**(5391): p. 1145-7.
4. Cowan, C.A., et al., *Nuclear reprogramming of somatic cells after fusion with human embryonic stem cells.* Science, 2005. **309**(5739): p. 1369-73.
5. Tada, M., et al., *Nuclear reprogramming of somatic cells by in vitro hybridization with ES cells.* Curr Biol, 2001. **11**(19): p. 1553-8.
6. Wilmut, I., et al., *Viable offspring derived from fetal and adult mammalian cells.* Nature, 1997. **385**(6619): p. 810-3.
7. Nichols, J., et al., *Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4.* Cell, 1998. **95**(3): p. 379-91.
8. Niwa, H., J. Miyazaki, and A.G. Smith, *Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells.* Nat Genet, 2000. **24**(4): p. 372-6.
9. Avilion, A.A., et al., *Multipotent cell lineages in early mouse development depend on SOX2 function.* Genes Dev, 2003. **17**(1): p. 126-40.
10. Chambers, I., et al., *Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells.* Cell, 2003. **113**(5): p. 643-55.
11. Mitsui, K., et al., *The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells.* Cell, 2003. **113**(5): p. 631-42.
12. Li, Y., et al., *Murine embryonic stem cell differentiation is promoted by SOCS-3 and inhibited by the zinc finger transcription factor Klf4.* Blood, 2005. **105**(2): p. 635-7.
13. Cartwright, P., et al., *LIF/STAT3 controls ES cell self-renewal and pluripotency by a Myc-dependent mechanism.* Development, 2005. **132**(5): p. 885-96.
14. Yamanaka, S. and K. Takahashi, *[Induction of pluripotent stem cells from mouse fibroblast cultures]*. Tanpakushitsu Kakusan Koso, 2006. **51**(15): p. 2346-51.
15. Amabile, G. and A. Meissner, *Induced pluripotent stem cells: current progress and potential for regenerative medicine.* Trends Mol Med, 2009. **15**(2): p. 59-68.

16. Marion, R.M., et al., *Telomeres acquire embryonic stem cell characteristics in induced pluripotent stem cells.* Cell Stem Cell, 2009. **4**(2): p. 141-54.
17. Okita, K., T. Ichisaka, and S. Yamanaka, *Generation of germline-competent induced pluripotent stem cells.* Nature, 2007. **448**(7151): p. 313-7.
18. Hussein, S.M., et al., *Copy number variation and selection during reprogramming to pluripotency.* Nature, 2011. **471**(7336): p. 58-62.
19. Lister, R., et al., *Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells.* Nature, 2011. **471**(7336): p. 68-73.
20. Fusaki, N., et al., *Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome.* Proc Jpn Acad Ser B Phys Biol Sci, 2009. **85**(8): p. 348-62.
21. Polo, J.M., et al., *A molecular roadmap of reprogramming somatic cells into iPS cells.* Cell, 2012. **151**(7): p. 1617-32.
22. Takahashi, K., et al., *Induction of pluripotent stem cells from fibroblast cultures.* Nat Protoc, 2007. **2**(12): p. 3081-9.
23. Lucchesi, J.C. and M.I. Kuroda, *Dosage compensation in Drosophila.* Cold Spring Harb Perspect Biol, 2015. **7**(5).
24. Strome, S., et al., *Regulation of the X chromosomes in Caenorhabditis elegans.* Cold Spring Harb Perspect Biol, 2014. **6**(3).
25. Muller, *Further studies on the nature and causes of gene mutations.* Proc. Sixth Int. Cong. Genet. USA 1932. **1**: p. 213-255.
26. Mukherjee, A.S. and W. Beermann, *Synthesis of ribonucleic acid by the X-chromosomes of Drosophila melanogaster and the problem of dosage compensation.* Nature, 1965. **207**(998): p. 785-6.
27. Nguyen, D.K. and C.M. Disteche, *Dosage compensation of the active X chromosome in mammals.* Nat Genet, 2006. **38**(1): p. 47-53.
28. Lin, H., et al., *Dosage compensation in the mouse balances up-regulation and silencing of X-linked genes.* PLoS Biol, 2007. **5**(12): p. e326.
29. Deng, X., et al., *Evidence for compensatory upregulation of expressed X-linked genes in mammals, Caenorhabditis elegans and Drosophila melanogaster.* Nat Genet, 2011. **43**(12): p. 1179-85.
30. Meyer, B.J., *Sex Determination and X Chromosome Dosage Compensation*, in *C. elegans II*, nd, et al., Editors. 1997: Cold Spring Harbor (NY).

31. Meyer, B.J., *Sex in the wormcounting and compensating X-chromosome dose.* Trends Genet, 2000. **16**(6): p. 247-53.
32. Lyon, M.F., *Gene action in the X-chromosome of the mouse (Mus musculus L.).* Nature, 1961. **190**: p. 372-3.
33. Davidson, R.G., H.M. Nitowsky, and B. Childs, *Demonstration of Two Populations of Cells in the Human Female Heterozygous for Glucose-6-Phosphate Dehydrogenase Variants.* Proc Natl Acad Sci U S A, 1963. **50**: p. 481-5.
34. Beutler, E., M. Yeh, and V.F. Fairbanks, *The normal human female as a mosaic of X-chromosome activity: studies using the gene for C-6-PD-deficiency as a marker.* Proc Natl Acad Sci U S A, 1962. **48**: p. 9-16.
35. Boumil, R.M. and J.T. Lee, *Forty years of decoding the silence in X-chromosome inactivation.* Hum Mol Genet, 2001. **10**(20): p. 2225-32.
36. Clemson, C.M., et al., *XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure.* J Cell Biol, 1996. **132**(3): p. 259-75.
37. Plath, K., et al., *Xist RNA and the mechanism of X chromosome inactivation.* Annu Rev Genet, 2002. **36**: p. 233-78.
38. Augui, S., E.P. Nora, and E. Heard, *Regulation of X-chromosome inactivation by the X-inactivation centre.* Nat Rev Genet, 2011. **12**(6): p. 429-42.
39. Heard, E. and C.M. Disteche, *Dosage compensation in mammals: fine-tuning the expression of the X chromosome.* Genes Dev, 2006. **20**(14): p. 1848-67.
40. Penny, G.D., et al., *Requirement for Xist in X chromosome inactivation.* Nature, 1996. **379**(6561): p. 131-7.
41. Marahrens, Y., et al., *Xist-deficient mice are defective in dosage compensation but not spermatogenesis.* Genes Dev, 1997. **11**(2): p. 156-66.
42. Plath, K., et al., *Role of histone H3 lysine 27 methylation in X inactivation.* Science, 2003. **300**(5616): p. 131-5.
43. Silva, J., et al., *Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes.* Dev Cell, 2003. **4**(4): p. 481-95.
44. Zhao, J., et al., *Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome.* Science, 2008. **322**(5902): p. 750-6.
45. Lee, J.T., L.S. Davidow, and D. Warshawsky, *Tsix, a gene antisense to Xist at the X-inactivation centre.* Nat Genet, 1999. **21**(4): p. 400-4.

46. Navarro, P., et al., *Tsix transcription across the Xist gene alters chromatin conformation without affecting Xist transcription: implications for X-chromosome inactivation.* Genes Dev, 2005. **19**(12): p. 1474-84.
47. Lee, J.T. and N. Lu, *Targeted mutagenesis of Tsix leads to nonrandom X inactivation.* Cell, 1999. **99**(1): p. 47-57.
48. Navarro, P., et al., *Molecular coupling of Xist regulation and pluripotency.* Science, 2008. **321**(5896): p. 1693-5.
49. Payer, B., J.T. Lee, and S.H. Namekawa, *X-inactivation and X-reactivation: epigenetic hallmarks of mammalian reproduction and pluripotent stem cells.* Hum Genet, 2011. **130**(2): p. 265-80.
50. Navarro, P. and P. Avner, *When X-inactivation meets pluripotency: an intimate rendezvous.* FEBS Lett, 2009. **583**(11): p. 1721-7.
51. Jonkers, I., et al., *RNF12 is an X-Encoded dose-dependent activator of X chromosome inactivation.* Cell, 2009. **139**(5): p. 999-1011.
52. Marson, A., et al., *Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells.* Cell, 2008. **134**(3): p. 521-33.
53. Chen, X., et al., *Integration of external signaling pathways with the core transcriptional network in embryonic stem cells.* Cell, 2008. **133**(6): p. 1106-17.
54. Donohoe, M.E., et al., *The pluripotency factor Oct4 interacts with Ctcf and also controls X-chromosome pairing and counting.* Nature, 2009. **460**(7251): p. 128-32.
55. Navarro, P., et al., *Molecular coupling of Tsix regulation and pluripotency.* Nature, 2010. **468**(7322): p. 457-60.
56. Jeon, Y., K. Sarma, and J.T. Lee, *New and Xisting regulatory mechanisms of X chromosome inactivation.* Curr Opin Genet Dev, 2012. **22**(2): p. 62-71.
57. Hackett, J.A., J.J. Zylicz, and M.A. Surani, *Parallel mechanisms of epigenetic reprogramming in the germline.* Trends Genet, 2012. **28**(4): p. 164-74.
58. Yamaguchi, S., et al., *Nanog expression in mouse germ cell development.* Gene Expr Patterns, 2005. **5**(5): p. 639-46.
59. Silva, J., et al., *Nanog is the gateway to the pluripotent ground state.* Cell, 2009. **138**(4): p. 722-37.
60. Mak, W., et al., *Reactivation of the paternal X chromosome in early mouse embryos.* Science, 2004. **303**(5658): p. 666-9.
61. Silva, J., et al., *Promotion of reprogramming to ground state pluripotency by signal inhibition.* PLoS Biol, 2008. **6**(10): p. e253.

62. Kohlmaier, A., et al., *A chromosomal memory triggered by Xist regulates histone methylation in X inactivation.* PLoS Biol, 2004. **2**(7): p. E171.
63. Maherali, N., et al., *Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution.* Cell Stem Cell, 2007. **1**(1): p. 55-70.
64. Payer, B., et al., *Tsix RNA and the germline factor, PRDM14, link X reactivation and stem cell reprogramming.* Mol Cell, 2013. **52**(6): p. 805-18.
65. Stadtfeld, M. and K. Hochedlinger, *Induced pluripotency: history, mechanisms, and applications.* Genes Dev, 2010. **24**(20): p. 2239-63.
66. Agrelo, R. and A. Wutz, *ConteXt of change--X inactivation and disease.* EMBO Mol Med, 2010. **2**(1): p. 6-15.
67. Amos-Landgraf, J.M., et al., *X chromosome-inactivation patterns of 1,005 phenotypically unaffected females.* Am J Hum Genet, 2006. **79**(3): p. 493-9.
68. Spatz, A., C. Borg, and J. Feunteun, *X-chromosome genetics and human cancer.* Nat Rev Cancer, 2004. **4**(8): p. 617-29.
69. Kawakami, T., et al., *Characterization of loss-of-inactive X in Klinefelter syndrome and female-derived cancer cells.* Oncogene, 2004. **23**(36): p. 6163-9.
70. Ying, Q.L., et al., *BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3.* Cell, 2003. **115**(3): p. 281-92.
71. Ying, Q.L., et al., *The ground state of embryonic stem cell self-renewal.* Nature, 2008. **453**(7194): p. 519-23.
72. Brons, I.G., et al., *Derivation of pluripotent epiblast stem cells from mammalian embryos.* Nature, 2007. **448**(7150): p. 191-5.
73. Greber, B., et al., *Conserved and divergent roles of FGF signaling in mouse epiblast stem cells and human embryonic stem cells.* Cell Stem Cell, 2010. **6**(3): p. 215-26.
74. Bradley, A., et al., *Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines.* Nature, 1984. **309**(5965): p. 255-6.
75. Guo, G., et al., *Klf4 reverts developmentally programmed restriction of ground state pluripotency.* Development, 2009. **136**(7): p. 1063-9.
76. Tesar, P.J., et al., *New cell lines from mouse epiblast share defining features with human embryonic stem cells.* Nature, 2007. **448**(7150): p. 196-9.
77. Wutz, A., *Gene silencing in X-chromosome inactivation: advances in understanding facultative heterochromatin formation.* Nat Rev Genet, 2011. **12**(8): p. 542-53.

78. Inoue, K., et al., *Impeding Xist expression from the active X chromosome improves mouse somatic cell nuclear transfer.* Science, 2010. **330**(6003): p. 496-9.
79. Pasque, V., et al., *X chromosome reactivation dynamics reveal stages of reprogramming to pluripotency.* Cell, 2014. **159**(7): p. 1681-97.
80. Hadjantonakis, A.K., et al., *An X-linked GFP transgene reveals unexpected paternal X-chromosome activity in trophoblastic giant cells of the mouse placenta.* Genesis, 2001. **29**(3): p. 133-40.
81. Takagi, N., et al., *Nonrandom X chromosome inactivation in mouse embryos carrying Searle's T(X;16)16H translocation visualized using X-linked LACZ and GFP transgenes.* Cytogenet Genome Res, 2002. **99**(1-4): p. 52-8.
82. Soma, A., K. Sato, and T. Nakanishi, *Visualization of inactive X chromosome in preimplantation embryos utilizing MacroH2A-EGFP transgenic mouse.* Genesis, 2013. **51**(4): p. 259-67.
83. Wu, H., et al., *Cellular resolution maps of X chromosome inactivation: implications for neural development, function, and disease.* Neuron, 2014. **81**(1): p. 103-19.
84. Kobayashi, S., *Live imaging of X chromosome inactivation and reactivation dynamics.* Dev Growth Differ, 2017. **59**(6): p. 493-500.
85. Rastan, S., et al., *X-chromosome inactivation in extra-embryonic membranes of diploid parthenogenetic mouse embryos demonstrated by differential staining.* Nature, 1980. **288**(5787): p. 172-3.
86. Takagi, N., O. Sugawara, and M. Sasaki, *Regional and temporal changes in the pattern of X-chromosome replication during the early post-implantation development of the female mouse.* Chromosoma, 1982. **85**(2): p. 275-86.
87. McMahon, A. and M. Monk, *X-chromosome activity in female mouse embryos heterozygous for Pgk-1 and Searle's translocation, T(X; 16) 16H.* Genet Res, 1983. **41**(1): p. 69-83.
88. Monk, M. and M.I. Harper, *Sequential X chromosome inactivation coupled with cellular differentiation in early mouse embryos.* Nature, 1979. **281**(5729): p. 311-3.
89. Huynh, K.D. and J.T. Lee, *Inheritance of a pre-inactivated paternal X chromosome in early mouse embryos.* Nature, 2003. **426**(6968): p. 857-62.
90. Okamoto, I., et al., *Epigenetic dynamics of imprinted X inactivation during early mouse development.* Science, 2004. **303**(5658): p. 644-9.
91. Marks, H., et al., *Dynamics of gene silencing during X inactivation using allele-specific RNA-seq.* Genome Biol, 2015. **16**: p. 149.

92. Sugimoto, M. and K. Abe, *X chromosome reactivation initiates in nascent primordial germ cells in mice.* PLoS Genet, 2007. **3**(7): p. e116.
93. Ng, K., et al., *A system for imaging the regulatory noncoding Xist RNA in living mouse embryonic stem cells.* Mol Biol Cell, 2011. **22**(14): p. 2634-45.
94. Guyochin, A., et al., *Live cell imaging of the nascent inactive X chromosome during the early differentiation process of naive ES cells towards epiblast stem cells.* PLoS One, 2014. **9**(12): p. e116109.
95. Tan, S.S., E.A. Williams, and P.P. Tam, *X-chromosome inactivation occurs at different times in different tissues of the post-implantation mouse embryo.* Nat Genet, 1993. **3**(2): p. 170-4.
96. Kobayashi, S., et al., *Live imaging of X chromosome reactivation dynamics in early mouse development can discriminate naive from primed pluripotent stem cells.* Development, 2016. **143**(16): p. 2958-64.
97. Ohhata, T. and A. Wutz, *Reactivation of the inactive X chromosome in development and reprogramming.* Cell Mol Life Sci, 2013. **70**(14): p. 2443-61.
98. Takagi, N., et al., *Reversal of X-inactivation in female mouse somatic cells hybridized with murine teratocarcinoma stem cells in vitro.* Cell, 1983. **34**(3): p. 1053-62.
99. Hanna, J., et al., *Human embryonic stem cells with biological and epigenetic characteristics similar to those of mouse ESCs.* Proc Natl Acad Sci U S A, 2010. **107**(20): p. 9222-7.
100. Wang, W., et al., *Rapid and efficient reprogramming of somatic cells to induced pluripotent stem cells by retinoic acid receptor gamma and liver receptor homolog 1.* Proc Natl Acad Sci U S A, 2011. **108**(45): p. 18283-8.
101. Pasque, V., et al., *Histone variant macroH2A confers resistance to nuclear reprogramming.* EMBO J, 2011. **30**(12): p. 2373-87.
102. Gillich, A., et al., *Epiblast stem cell-based system reveals reprogramming synergy of germline factors.* Cell Stem Cell, 2012. **10**(4): p. 425-39.
103. Ebrahimi, B., *Reprogramming barriers and enhancers: strategies to enhance the efficiency and kinetics of induced pluripotency.* Cell Regen (Lond), 2015. **4**: p. 10.
104. Yang, F., et al., *Global survey of escape from X inactivation by RNA-sequencing in mouse.* Genome Res, 2010. **20**(5): p. 614-22.
105. Halbritter, F., H.J. Vaidya, and S.R. Tomlinson, *GeneProf: analysis of high-throughput sequencing experiments.* Nat Methods, 2011. **9**(1): p. 7-8.

106. Cerase, A., et al., *Xist localization and function: new insights from multiple levels.* Genome Biol, 2015. **16**: p. 166.
107. Robert Finestra, T. and J. Gribnau, *X chromosome inactivation: silencing, topology and reactivation.* Curr Opin Cell Biol, 2017. **46**: p. 54-61.
108. Chung, S., et al., *Analysis of different promoter systems for efficient transgene expression in mouse embryonic stem cell lines.* Stem Cells, 2002. **20**(2): p. 139-45.
109. Nishimura, K., et al., *Manipulation of KLF4 expression generates iPSCs paused at successive stages of reprogramming.* Stem Cell Reports, 2014. **3**(5): p. 915-29.
110. Wareham, K.A., et al., *Age related reactivation of an X-linked gene.* Nature, 1987. **327**(6124): p. 725-7.
111. Chaligne, R., et al., *The inactive X chromosome is epigenetically unstable and transcriptionally labile in breast cancer.* Genome Res, 2015. **25**(4): p. 488-503.
112. Sahakyan, A., et al., *Human Naive Pluripotent Stem Cells Model X Chromosome Dampening and X Inactivation.* Cell Stem Cell, 2017. **20**(1): p. 87-101.