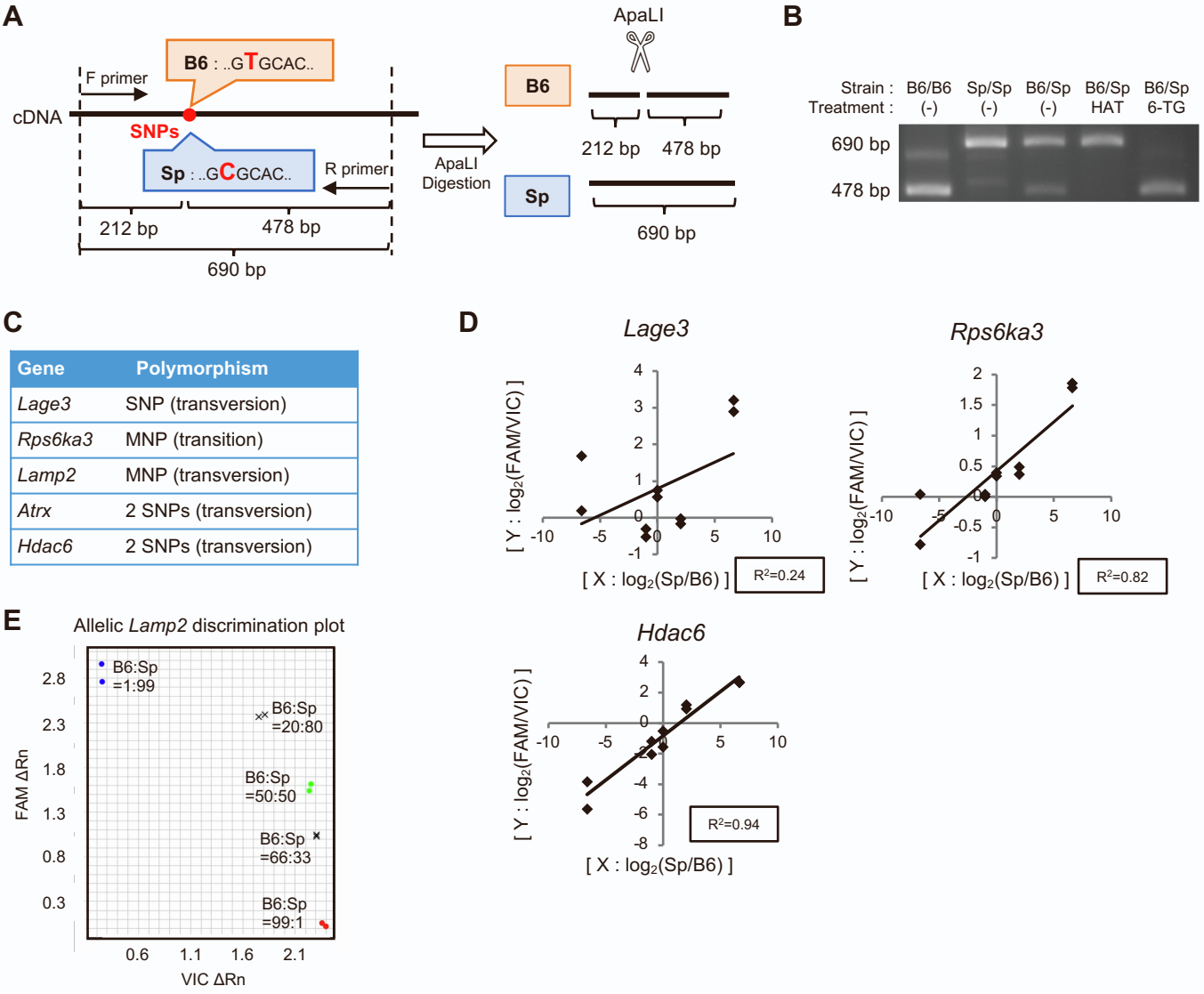
Stem Cell Reports, Volume *17*

Supplemental Information

Early reactivation of clustered genes on the inactive X chromosome during somatic cell reprogramming

Shiho Aizawa, Ken Nishimura, Gonzalo Seminario Mondejar, Arun Kumar, Phuong Linh Bui, Yen Thi Hai Tran, Akihiro Kuno, Masafumi Muratani, Shin Kobayashi, Tsukasa Nabekura, Akira Shibuya, Eiji Sugihara, Taka-Aki Sato, Aya Fukuda, Yohei Hayashi, and Koji Hisatake



Gene	Polymorphism
Lage3	SNP (transversion)
Rps6ka3	MNP (transition)
Lamp2	MNP (transversion)
Atrx	2 SNPs (transversion)
Hdac6	2 SNPs (transversion)

Figure S1. Detection and determination of allele-specific transcription. Related to Figures 1 and 3

(A) Experimental scheme of *Apa*LI digestion of amplified DNA to distinguish *Lamp2* alleles. cDNA was synthesized from RNA extracted from bsMEFs, and a 690-bp region encompassing the SNP was amplified by PCR from the cDNA. *Apa*LI digests only PCR products from the B6 allele, which produces two shorter fragments. (B) Evaluation of HAT and 6-TG selections by *Apa*LI digestion. cDNA from RNAs extracted from bsMEFs treated with either HAT or 6-TG for 14-20 days, were used for *Apa*LI digestion as described in (A). Fragments were separated by agarose gel electrophoresis. The 690-bp fragment derives from the Sp allele while the 478-bp fragment derives from the B6 allele. (C) Polymorphisms in *Lage3*, *Rps6ka3*, *Lamp2*, *Atrx,* and *Hdac6*, for which TaqMan probes were tested. (D) Graphic plots of abundance of allele-specific transcripts and the proportions of two fluorescence intensities for *Lage3*, *Rps6ka3*, *Atrx,* and *Hdac6*. cDNAs derived from homozygous B6 and Sp mice were mixed at five different ratios (B6 : Sp = 1 : 99, 20 : 80, 50 : 50, 66 : 33, 99 : 1), and qRT-PCR was performed with TaqMan probes for indicated genes. R^2 indicates the square of the correlation coefficient between $\log_2$ of (FAM intensity / VIC intensity) and $\log_2$ of (Sp allelic abundance / B6 allelic abundance). (E) Discrimination of allele-specific *Lamp2* transcripts. cDNAs derived from homozygous B6 and Sp mice were mixed as (D), and qRT-PCR was performed with TaqMan probes for the *Lamp2* gene. VIC ΔRn and FAM ΔRn were plotted on the X and Y axes, respectively. ΔRn means normalized intensity of gained fluorescence for each reporter dye.

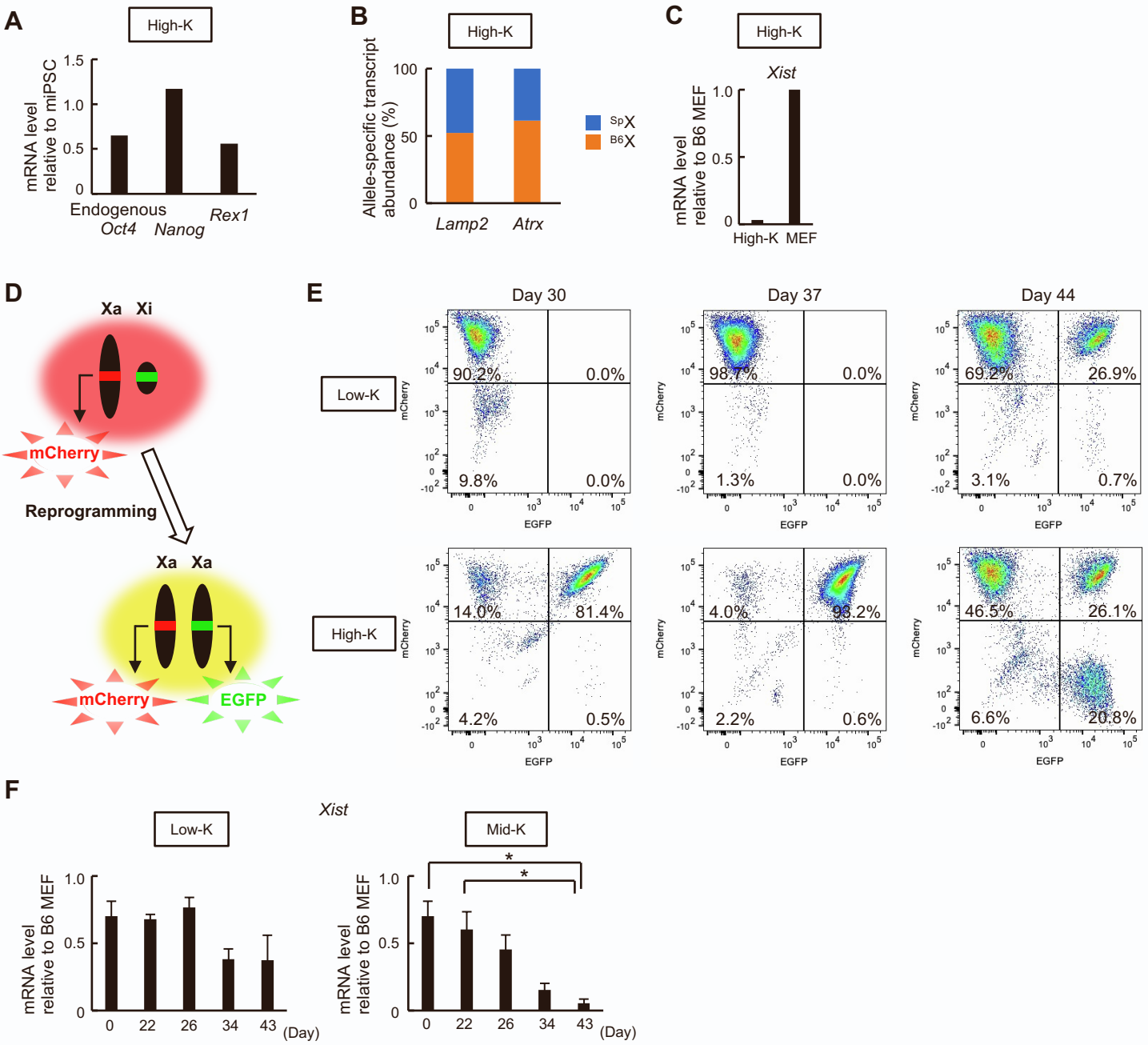


Figure S2. Induction of XCR under the High-K condition. Related to Figure 2

(A) Expression of pluripotency markers in HAT-bsMEFs reprogrammed under the High-K condition. The levels of endogenous *Oct4*, *Nanog*, and *Rex1* mRNA were determined by qRT-PCR at day 30 of reprogramming. Data consist of a single experiment. (B) Relative abundance of allele-specific transcripts in HAT-bsMEFs reprogrammed under the High-K condition. Relative abundance of *Lamp2* and *Atrx* transcripts were determined by TaqMan probes for each gene in the same cells as (A). Data consist of a single experiment. (C) *Xist* RNA levels in HAT-bsMEFs reprogrammed under the High-K condition. The cells described in (A) as well as MEFs were used for qRT-PCR. Data consist of a single experiment. (D) Monitoring of XCR during reprogramming of Momiji mouse-derived MEFs, in which the mCherry or GFP gene was integrated into each allele of *Hprt*. In addition to mCherry expressed from the Xa, reprogrammed mCherry(+) MEFs express GFP from the reactivated Xi. (E) Instability of XCR when iPSCs are fully reprogrammed. mCherry(+) Momiji MEFs were reprogrammed under the Low-K or High-K condition, and fluorescence was observed by FACS at indicated days of reprogramming. (F) Kinetics of the expression level of *Xist* RNA during reprogramming. The expression level of *Xist* RNA was determined by qRT-PCR from the same cells described in Figure 2B. Data represent means ± SEM of at least three independent experiments. *p < 0.05.

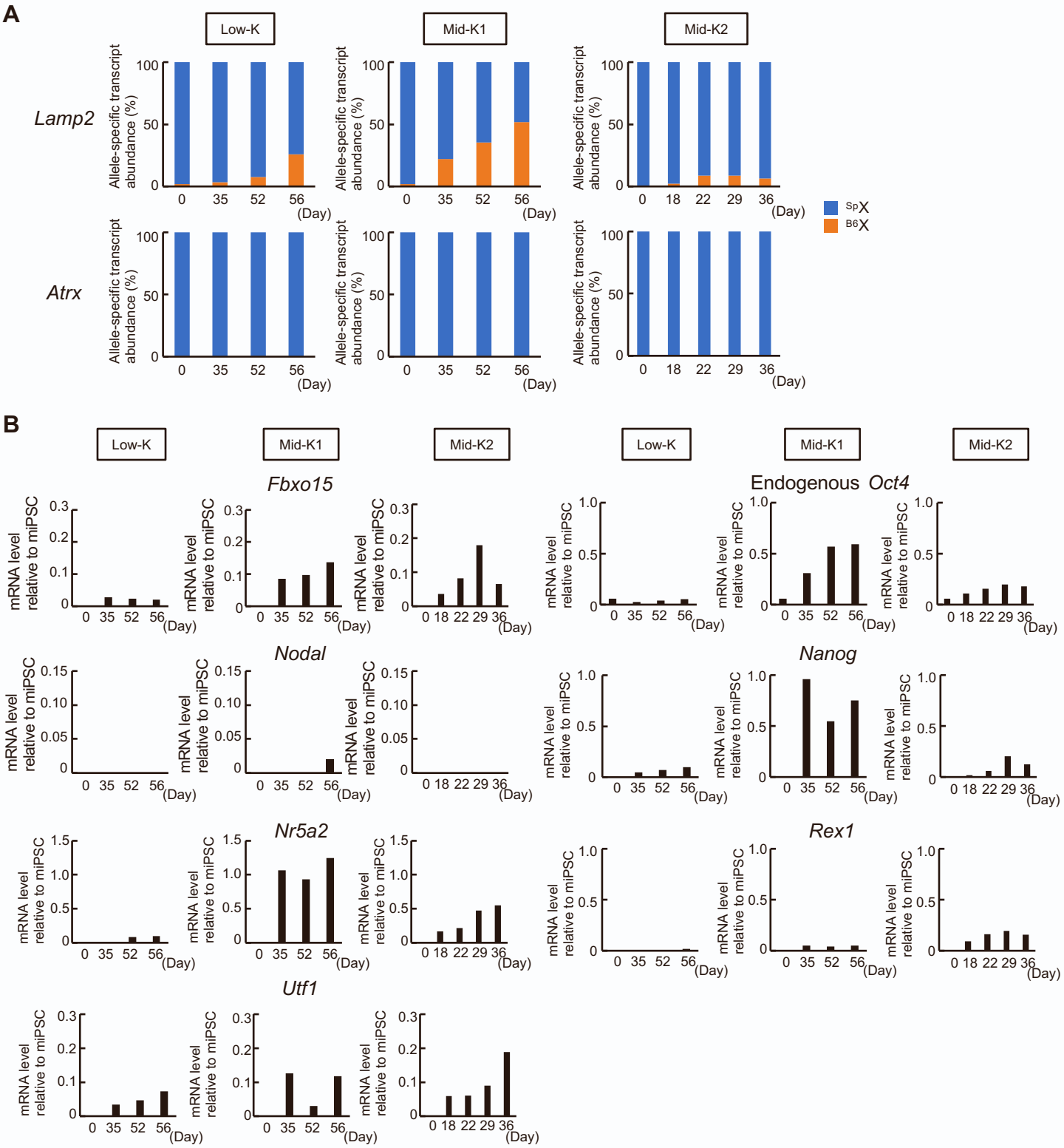


Figure S3. Reactivation of X-linked genes on the Xi in partially reprogrammed iPSCs analyzed in RNA-seq. Related to Figure 3

(A) Relative abundance of allele-specific transcripts of *Lamp2* and *Atrx*. HAT-bsMEFs were reprogrammed under the Low-K and Mid-K conditions, and allele-specific transcripts were analyzed by TaqMan probes for *Lamp2* and *Atrx* at indicated days of reprogramming.
(B) Expression kinetics of pluripotency markers in partially reprogrammed iPSCs. The levels of *Fbxo15, Nodal, Nr5a2, Utf1*, endogenous *Oct4*, *Nanog*, and *Rex1* transcripts in the same cells prepared in (A) were determined by qRT-PCR at indicated days of reprogramming.
All data consist of a single experiment.

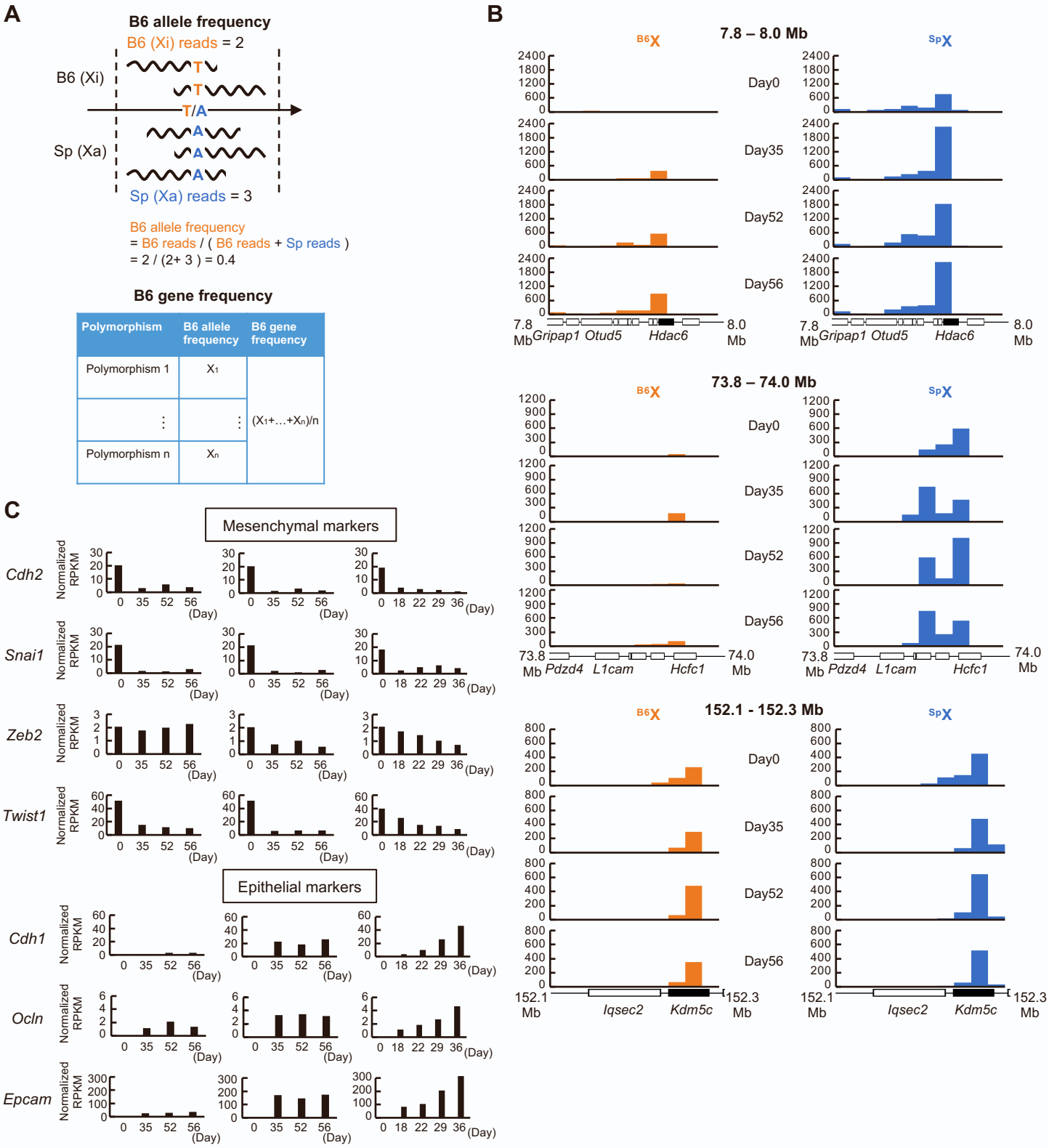


Figure S4. Gene expression data from RNA-seq analyses. Related to Figure 3

(A) Calculation of B6 allele frequency and B6 gene frequency from RNA-seq data. DNA sequence of B6 mouse genome was used as the reference genome. The reads which cannot be aligned to the B6 reference genome were assigned as an Sp-derived transcript. B6 allele frequency is calculated as the number of B6 reads / the number of total reads at each polymorphic position. B6 gene frequency is the average of B6 allele frequencies at all polymorphic positions in the gene. (B) Transcripts from B6 and Sp alleles in representative regions where transcription from the Xi showed increase (7.8-8.0 Mb region: Xtreme) or little change (73.8-74.0 Mb region). A region of an escapee gene is also shown (152.1-152.3 Mb region). Location of *Hdac6* (one of the 7 genes) and *Kdm5c* (escapee) are highlighted. Reads from Mid-K1 cells were assigned to B6 or Sp allele, and the assigned reads were summed up in a 20-kbp window. (C) Kinetics of gene expression of mesenchymal and epithelial markers. Normalized expression values were determined by RNA-seq.

HAT-bsMEF (n=10)

Cell	Abundance of allele-specific transcripts						*Xist* expression level relative to control MEF
	Hdac6		*Lamp2*		*Atrx*		
	B6	Sp	B6	Sp	B6	Sp	
#1	6.3	93.7	0.0	100.0	0.0	100.0	5.471
#2	0.1	99.9	0.0	100.0	0.0	100.0	5.183
#3	0.0	100.0	0.0	100.0	0.0	100.0	4.007
#4	0.1	99.9	0.0	100.0	0.0	100.0	2.653
#5	15.2	84.8	19.5	80.5	0.0	100.0	0.609
#6	0.8	99.2	0.0	100.0	0.0	100.0	2.651
#7	1.1	98.9	0.0	100.0	0.0	100.0	2.771
#8	0.0	100.0	0.0	100.0	0.0	100.0	3.684
#9	22.1	77.9	0.1	99.9	7.0	93.0	2.437
#10	18.9	81.1	0.0	100.0	0.0	100.0	1.053

Mid-K Day 26 (n=16)

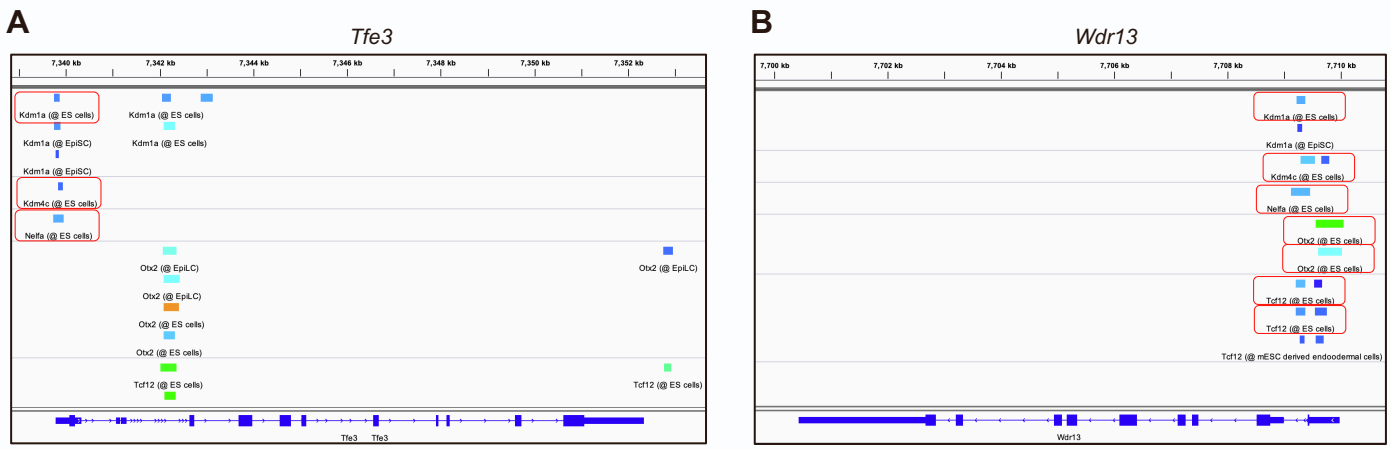
Cell	Abundance of allele-specific transcripts						*Xist* expression level relative to control MEF
	Hdac6		*Lamp2*		*Atrx*		
	B6	Sp	B6	Sp	B6	Sp	
#1	5.5	94.5	0.0	100.0	29.2	70.8	0.000
#2	2.8	97.2	2.0	98.0	1.1	98.9	0.337
#3	23.0	77.0	13.1	86.9	0.0	100.0	0.000
#4	35.1	64.9	0.0	100.0	0.0	100.0	0.113
#5	30.6	69.4	0.0	100.0	4.0	96.0	0.000
#6	56.0	44.0	0.0	100.0	4.6	95.4	0.357
#7	24.9	75.1	28.2	71.8	0.0	100.0	0.549
#8	32.4	67.6	9.2	90.8	11.1	88.9	0.000
#9	33.1	66.9	1.1	98.9	0.8	99.2	0.000
#10	43.0	57.0	73.0	27.0	0.0	100.0	0.029
#11	46.9	53.1	1.0	99.0	10.2	89.8	0.000
#12	0.3	99.7	2.0	98.0	0.0	100.0	0.254
#13	11.8	88.2	21.8	78.2	13.7	86.3	0.000
#14	7.9	92.1	0.0	100.0	0.0	100.0	0.000
#15	41.1	58.9	40.1	59.9	0.0	100.0	0.260
#16	26.7	73.3	0.0	100.0	6.9	93.1	0.000

Mid-K Day 34 (n=13)

Cell	Abundance of allele-specific transcripts						*Xist* expression level relative to control MEF
	Hdac6		*Lamp2*		*Atrx*		
	B6	Sp	B6	Sp	B6	Sp	
#1	22.7	77.3	0.0	100.0	0.0	100.0	0.457
#2	60.8	39.2	22.8	77.2	0.0	100.0	0.337
#3	55.7	44.3	8.3	91.7	1.5	98.5	0.000
#4	24.7	75.3	35.6	64.4	7.2	92.8	0.113
#5	54.2	45.8	0.0	100.0	28.9	71.1	0.277
#6	88.3	11.7	0.0	100.0	8.4	91.6	0.000
#7	43.5	56.5	20.3	79.7	39.9	60.1	0.357
#8	0.2	99.8	42.4	57.6	3.6	96.4	0.000
#9	7.0	93.0	0.0	100.0	0.0	100.0	0.000
#10	45.0	55.0	0.0	100.0	27.0	73.0	0.029
#11	68.8	31.2	0.0	100.0	11.0	89.0	0.000
#12	12.7	87.3	0.0	100.0	0.0	100.0	0.000
#13	55.2	44.8	0.1	99.9	0.0	100.0	0.000

Figure S5. Raw data of transcriptional reactivation on the Xi and *Xist* expression in single cells. Related to Figure 4

HAT-bsMEFs were reprogrammed under the Mid-K condition, and single cells were isolated. Relative abundance of allele-specific transcripts were analyzed as in Figure 4A, using RT-RamDA and the TaqMan probes for *Lamp2*, *Atrx*, and *Hdac6*. More than 25% of the B6-specific transcript is regarded as transcriptional reactivation (highlighted in yellow). *Xist* expression levels were determined by qPCR of the same RT-RamDA products. Cells that showed *Xist* expression are highlighted in yellow.



C

Transcriptional regulator	7 genes					
	Gpkow	*Gm45208 /Wdr45*	*Tfe3*	*Hdac6*	*Wdr13*	*Ftsj1*
KDM1A	+		+	+	+	+
KDM4C	+	+	+	+	+	+
NELFA	+	+	+	+	+	+
OTX2		+		+	+	+
TCF12	+	+		+	+	+

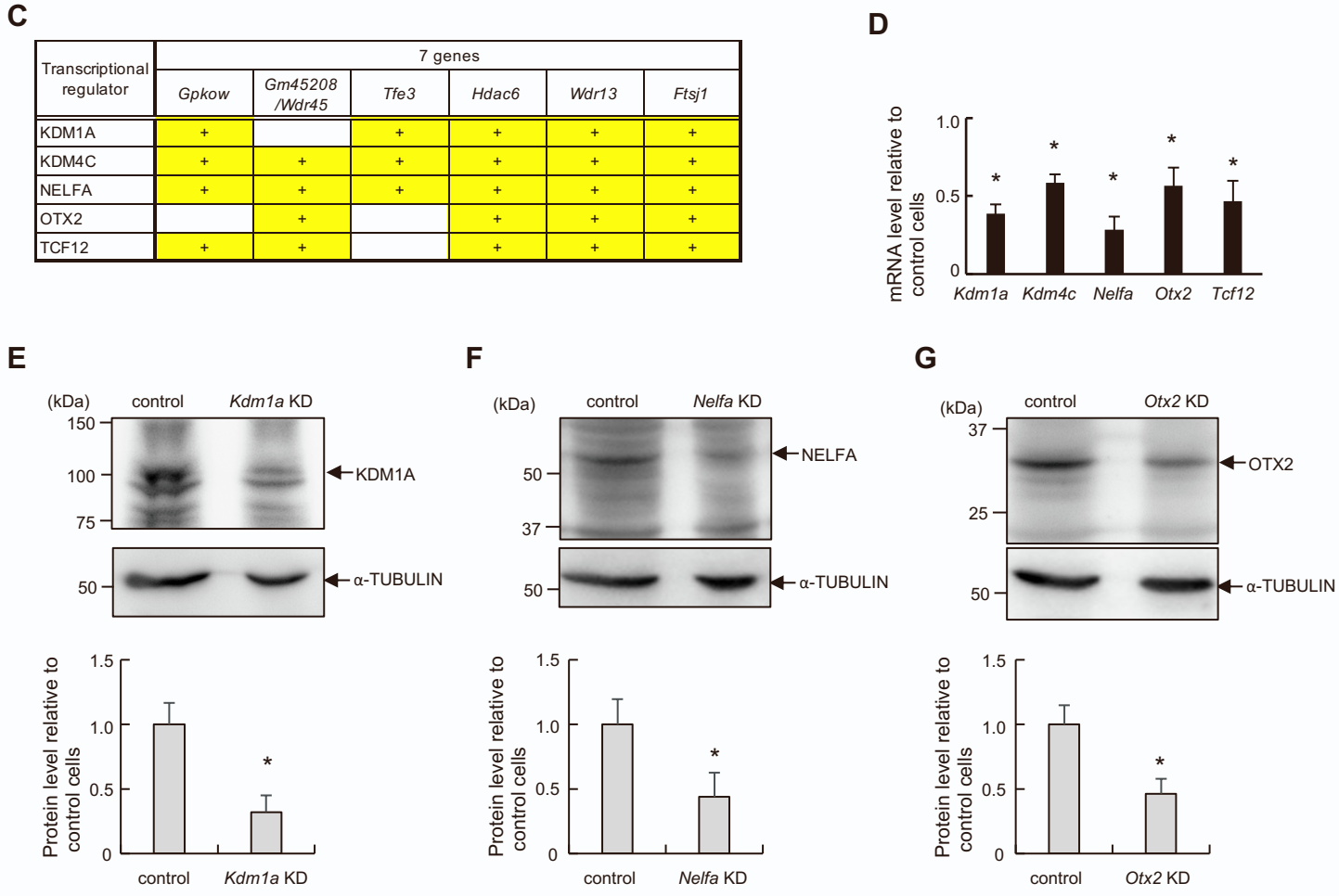


Figure S6. Screening and functional confirmation of candidate transcriptional regulators. Related to Figure 5

(A, B) Representative ChIP-Atlas browser tracks that show occupancies of transcriptional regulators within 1 kb of the TSS of *Tfe3* or *Wdr13* in ESCs. Binding peaks near TSS of *Tfe3* (A) or *Wdr45* (B) in ESCs are enclosed by a red box in the images from ChIP-Atlas peak browser (https://chip-atlas.org/peak_browser). (C) Summary of the occupancies (highlighted in yellow) of transcriptional regulators in the 7 genes. (D) Effect of shRNA on mRNA expression of each transcriptional regulator. The mRNA level was determined in HAT-bsMEFs (*Kdm1a*, *Kdm4c*, *Nelfa*, and *Tcf12*) or bsMEF-derived iPSCs (*Otx2*) after transduction of a retrovirus expressing shRNA against each transcriptional regulator. Data represent means ± SEM of at least three independent experiments. *p < 0.05 versus cells infected with control hKO-expressing retrovirus. (E-G) Effect of shRNA on the protein level of its target gene. MEFs treated with shRNA for *Kdm1a* (E) or *Nelfa* (F) or ESC for *Otx2* (G) were selected by puromycin for 3 days. The protein level was determined using the whole cell extract from the selected cells. The lower graphs show the intensity of each band normalized against α-TUBULIN. Data are represented as means SEM of three independent experiments. *p < 0.05 versus mock infected control cells.

Table S1. Primers and probes used for allele distinction. Related to Figures 1, 2, 3, 4, 5, 6, and 7, and Figures S1, S2, S3, and S5.

Target	Types of oligonucleotide		Sequence
Lamp2	Primers for allele-specific digestion	Forward primer	GGCTCAGCTTTCAACATTTCC
		Reverse primer	GGCATCTTAGGTTAGGATCCCA
	Primers for single cell analysis	Forward primer	TACCTGACAAGGCGACACAC
		Reverse primer	GCACTGCAGTCTTGAGCTGT
	Primer	Forward primer	CACCTGCAAGCTTTTGTCCA
		Reverse primer	GTGTGAATGATGGGTGCCAC
	Probe	B6	AAGACCAAACTccCACCAC
		Sp	AAGACCAAACTgtCACCAC
Atrx	Primers for single cell analysis	Forward primer	CGACGACGACAATGATCCTGA
		Reverse primer	GTGAAGAAAACTAACCACCTGGA
	Primer	Forward primer	GTTCAAGAGGATTCCTCCAGTGA
		Reverse primer	TTCTGCCTTTTCCAGGTGACTT
	Probe	B6	TTCTGAGGAGGAcAAgAA
		Sp	TGAGGAGGAaAAaAA
Hdac6	Primers for single cell analysis	Forward primer	CAGCAGGATTTGCCCACCTA
		Reverse primer	CACGATTAGGATCCGCAGGG
	Primer	Forward primer	CCTTCTGGGAGGTCCTGGA
		Reverse primer	GCTGAGGTGTCTCAGGGTGAT
	Probe	B6	AAGAAGCCgTgCTAGAAG
		Sp	TGGAAGAAGCCaTtCTAGAAG
Rps6ka3	Primer	Forward primer	CCATCCAACATTTTGTATGTTTTTG
		Reverse primer	GCCATCTCTAAGGTTTGCCG
	Probe	B6	TAAATAGTTCACctGGATTAA
		Sp	TAAATAGTTCACtgGGATTAA
Lage3	Primer	Forward primer	TTTAACCCAAAGGCCTGGAA
		Reverse primer	GCGATTTCTGCCTCCAAAGA
	Probe	B6	ACATCACAGATTCtCCCTCT
		Sp	ATCACAGATTCgCCCTCT

Table S2. Primers used for genotyping. Related to Figure 1

Target	Types of primer	Sequence
Zfy1 (chr. Y)	Forward primer	TGGTGATTCAATAATGTCTGTTCAGCTG
	Reverse primer	GGACAAACTTTACGTGTCTATCCTTGC
Hprt	Forward primer	GAGAGAACTGCTACCACTGAGAG
	Reverse primer for *Hprt*(-)	GTTATTGGTGGAGATGATCTCTCAAC
	Reverse primer for *Hprt*(+)	ATCAGAGCAGCCGATTGTCTGTTG

Table S3. Primers used for qRT-PCR. Related to Figures 2, 4, 5, and 7, and Figures S2, S3, and S6

Target	Types of primer	Sequence
Fbxo15	Forward primer	CTCATCTGTCACGAAGCAGC
	Reverse primer	AGGTCACCGCATCCAAGTAA
Kdm1a	Forward primer	CTCCTGGCCCCTCAATTC
	Reverse primer	TGTGTGTTCTCCAGCAAAGAA
Kdm4c	Forward primer	GCGGGTTCATGCAAGTTGTT
	Reverse primer	GTTTCAGAGCACCTCCCCTC
Nanog	Forward primer	ACCTGAGCTATAAGCAGGTTAAGAC
	Reverse primer	GTGCTGAGCCCTTCTGAATCAGAC
Nelfa	Forward primer	CTGGCCTGGTATCCACACAG
	Reverse primer	AGTCTCAGAGCTGGGGATGT
Nodal	Forward primer	ACATGTTGAGCCTCTACCGAG
	Reverse primer	GTGAAAGTCCAGTTCTGTCCG
Nr5a2	Forward primer	CAAAGTGGAGACGGAAGCC
	Reverse primer	ATCGCCACACACAGGACATA
endogenous *Oct4*	Forward primer	CTGTTCCCGTCACTGCTCTG
	Reverse primer	AACCCCAAAGCTCCAGGTTC
Otx2	Forward primer	CTCGACGTTCTGGAAGCTC
	Reverse primer	GGCCTCACTTTGTTCTGACC
Rex1	Forward primer	TTGATGGCTGCGAGAAGAG
	Reverse primer	ACCCAGCCTGAGGACAATC
Tbp	Forward primer	TATCTGCTGGCGGTTTGGC
	Reverse primer	TGAAATAGTGATGCTGGGCAC
Tcf12	Forward primer	AGACACAAACCTGGCAGGAG
	Reverse primer	TGCAGAAGCGACACACTGAT
Utf1	Forward primer	GTCCCTCTCCGCGTTAGC
	Reverse primer	GGGGCAGGTTCGTCATTT
Xist (B6)	Forward primer	GGTTCTCTCTCCAGAAGCTAGGAAAG
	Reverse primer	TGGTAGATGGCATTGTGTATTATATGG

Table S4. Annealed oligos for shRNA. Related to Figure 5 and Figure S6

Target	Sequence
Kdm1a	gatccGAGTTGAAAGAGCTTCTTAATCTCGAGATTAAGAAGCTCTTTCAACTCTTTTTGg
	aattcCAAAAAGAGTTGAAAGAGCTTCTTAATCTCGAGATTAAGAAGCTCTTTCAACTCg
Kdm4c	gatccGCAGAGTGATAGATGTGACATCTCGAGATGTCACATCTATCACTCTGCTTTTTGg
	aattcCAAAAAGCAGAGTGATAGATGTGACATCTCGAGATGTCACATCTATCACTCTGCg
Nelfa	gatccGCCAGTACCTGAACAAGAATGCTCGAGCATTCTTGTTCAGGTACTGGCTTTTTGg
	aattcCAAAAAGCCAGTACCTGAACAAGAATGCTCGAGCATTCTTGTTCAGGTACTGGCg
Otx2	gatccGCATGGACTGTGGATCTTATTCTCGAGAATAAGATCCACAGTCCATGCTTTTTGg
	aattcCAAAAAGCATGGACTGTGGATCTTATTCTCGAGAATAAGATCCACAGTCCATGCg
Tcf12	gatccGAAGGCCTTGGCATCTATTTACTCGAGTAAATAGATGCCAAGGCCTTCTTTTTGg
	aattcCAAAAAGAAGGCCTTGGCATCTATTTACTCGAGTAAATAGATGCCAAGGCCTTCg

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Isolation and selection of mouse embryonic fibroblasts

Mouse embryonic fibroblasts (MEFs) were derived from E13.5 embryos which were derived from the cross between female *Hprt*-KO *Mus musculus* (RBRC02467, RIKEN Bioresource Center) and male *Mus spretus* (RBRC00208, RIKEN Bioresource Center) and cultured in MEF medium (DMEM (Sigma) supplemented with 10% FBS (Biosera), 100 U/mL penicillin-streptomycin (Nacalai Tesque)). To select bsMEFs ($^{B6}X^{Hprt-}$, ^{Sp}X) from the isolated MEFs, their genomic DNAs were genotyped by PCR using KOD-Plus-Neo (TOYOBO). Genomic DNAs were extracted using Geno Plus Mini (VIOGENE). DNA sequences of the PCR primers are shown in Table S2.

In order to isolate subpopulations of 6TG-bsMEF (^{B6}Xa, ^{Sp}Xi) from bsMEF, the cells were cultured in MEF medium containing 40 mM of 6-TG (Sigma) for 14 days. For isolating HAT-bsMEFs (^{B6}Xi, ^{Sp}Xa), bsMEFs were cultured in HAT medium (MEF medium supplemented with 4 mg/mL Hypothanthine (Sigma), 2 mg/mL Thymidine (Sigma) and 1 μg/mL MTX (Pfizer)) for 14-20 days. Because Momiji MEFs (Kobayashi et al. 2016) possess GFP and mCherry at each allele of the *Hprt* locus, the MEFs were sorted by MoFlo XDP (Beckman coulter) to isolate a homogeneous population of mCherry(+) MEFs.

Detection of allele-specific transcripts by PCR

For allele-specific digestion, PCR was done to amplify cDNA of the *Lamp2* transcript using KOD-plus-neo. The PCR products were then digested by ApaLI (New England Biolabs) at 37˚C for 1 hour and analyzed on an agarose gel. The DNA sequences of the used primers are listed in Table S1.

Design of TaqMan probes for determination of allele-specific transcripts

TaqMan probes were designed based upon polymorphisms between B6 and Sp, which were available from Sanger Mouse Genomes Project. Target polymorphisms were chosen according to the following criteria: 1) one MNP or two SNPs that can be included in a single probe and 2) MNP or SNPs that are located within genes expressed both in MEFs and ESCs. Custom TaqMan SNP Genotyping Assays (Thermo Fisher) was used to create the probes labeled with a fluorescence reporter (VIC for B6 allele or FAM for Sp allele) and NFQ-MGB quencher.

SeVdp vector production

Blasticidin S-resistant gene together with the T2A peptide sequence was amplified and inserted after the c-MYC-coding sequence of SeVdp(fK-OSM). SeVdp vectors were produced as described previously (Nishimura et al. 2011). Titers of the SeVdp vectors were determined by

counting the number of NIH3T3 cells infected with serial dilutions of vector suspension, using immunostaining with an anti-SeV NP antibody (Nishimura et al. 2007).

cDNA synthesis and qPCR

Total RNA was extracted using ISOGEN (Nippon Gene). RNA concentration was measured by Nanodrop 2000 (Thermo Fisher). Reverse transcription was performed using Superscript III First-Strand Synthesis System (Thermo Fisher). Quantitative PCR (qPCR) was performed using QuantStudio 5 Real-Time PCR System (Applied Biosystems) with GoTaq qPCR Master Mix (Promega). As controls, we used cDNA from germline-competent mouse iPSCs generated by the SeVdp vector (Nishimura et al. 2011). The expression levels were normalized against that of TATA-box binding protein (*Tbp*). The DNA sequences of the primers are listed in Table S3.

Fluorescence activated cell sorting (FACS) for monitoring of XCR

Expression of mCherry and EGFP from reprogrammed Momiji MEFs were detected by FACS Aria SORP (BD Biosciences) with 530/30 nm and 610/20 nm fluorescent filters. Data were analyzed using FlowJo software (BD Biosciences).

RNA-seq

RNA concentration was measured by Nanodrop 2000 (Thermo Fisher), and RNA quality was validated with RNA 6000 Pico kit (Agilent). RNA-seq libraries were constructed with 500 ng of total RNA using NEBNext Poly(A) mRNA Magnetic Isolation Module and NEBNext Ultra Directional RNA Library Prep Kit (New England Biolabs). Sequencing was performed with NextSeq 500 (Illumina) with paired-end mode for 2x76-base reads. FASTQ files were imported to CLC Genomics Workbench (CLC-GW, Version 10.1.1, Qiagen). Mapping to the mouse reference genome (mm10) was performed with RNA-seq analysis tool in CLC-GW. Transcription expression values were obtained in RPKM (reads per kilobase per million reads) and normalized by quantile normalization method.

Analyses of published data

To analyze ATAC-seq data of GSM1828645 and GSM1828646 (Giorgetti et al. 2016) annotated to the mouse reference genome (mm9), all peaks were shifted to mm10 using Lift Genome Annotations (https://genome.ucsc.edu/cgi-bin/hgLiftOver). Relative accessibility of Xi was calculated as 129 (Xi) reads / Cast (Xa) reads in each polymorphism. To compare chromatin accessibility in regions in the Xi, all of 129 (Xi) reads were summed up in 2-Mbs windows.

To screen for candidate factors that bind the Xtreme region, peak information was visualized and retrieved from ChIP-Atlas database using Peak Browser (Oki et al. 2018). All proteins in “TFs

and others" (332 factors out of 13,558 data sets) that show binding within ±1 kbs of TSSs of *Gpkow*, *Gm45208*/*Wdr45*, *Tfe3*, *Hdac6*, *Wdr13* and *Ftsj1* in "Embryonic Stem Cells" of "Pluripotent stem cell" were screened. The proteins that show binding to more than 5 genes were selected as candidate transcriptional regulators potentially relevant to early-onset reactivation on the Xi.

Production of retroviral vector

To prevent retrovirus silencing during reprogramming, the shRNA expression cassette from pMXs-U6-Puro plasmid (Cambridge bioscience) was inserted into pMCsΔYY1-IRES-Puro plasmid (Nishimura et al. 2017) to produce pMCΔYY-U6-Puro plasmid. Annealed DNA oligonucleotides were inserted into pMCΔYY-U6-Puro to construct the retroviral vector expressing shRNA against each target gene. The DNA sequences of the annealed DNA oligonucleotides are listed in Table S4.

PLAT-E cells were transfected with each vector using Lipofectamine 2000 (Thermo Fisher). Viral supernatant was collected 2 days after transfection and filtered through a 0.45 μm cellulose acetate filter and stored at -80 ˚C until use.

Determination of Protein Expression by Immunoblotting

MEFs or EB5 mESCs transduced with shRNA-expressing retroviral vector were collected three days after puromycin treatment, and proteins from these cells were isolated as described previously (Bui et al. 2019). Protein expression levels were quantified using FUSION FX7.EDGE and FUSION Capt Software (Vilber-Lourmat). We employed the following primary antibodies; anti-KDM1A/LSD1 (1:2,000, ab17721; Abcam), anti-NELFA (1:2,000, 10456-1-AP; Proteintech), anti-OTX2 (1:2,000, GTX133210; GeneTex), and anti-α-TUBULIN (1:10,000, ab7291; Abcam).

ChIP-seq and calculation of KDM1A occupancy

Chromatin immunoprecipitation was done as described previously (Nishimura et al., 2017) using anti-KDM1A/LSD1 (ab17721; Abcam). Libraries for ChIP-seq were prepared using KAPA HyperPlus Kit (Roche) according to the manufacturer's instructions. Briefly, 10 ng of ChIP DNA was used for library preparation with unique dual index adaptor from IDT for Illumina TruSeq DNA UD Indexes (Illumina). The libraries were quantified using Qubit 4 Fluorometer (Thermo Fisher) and analyzed for size distribution using MultiNA (Shimdzu) or TapeStation 2200 (Agilent Technologies). Paired-end sequencing of the libraries was performed with Novaseq 6000 Sequencing System using SP 100 cycle kit (Illumina).

ChIP-seq reads were mapped to the mouse reference genome (mm10) using BWA-MEM version 0.7.17 (Li and Durbin, 2010). Variant calls were conducted using piping samtools mpileup

output into bcftools call (Li et al., 2009). To remove sequence errors, only variants found in more than two samples were regarded polymorphisms between B6 and Sp. To calculate KDM1A occupancy, all the reads in enhancer, promoter, and gene body was summed up in 0.2 Mb windows. Then, the summed read numbers of the Xi were normalized against that of the Xa. We used ESC enhancers based on EnhancerAtlas 2.0 (Gao and Qian, 2020) (ESC_J1) (http://www.enhanceratlas.org/indexv2.php) as enhancers for the nearest neighbor genes. Reads within ±1 kb from TSS of each gene were assigned as reads for promoter. Data under each condition were derived from a single sample.

SUPPLEMENTAL REFERENCES

Bui, P.L., Nishimura, K., Seminario Mondejar, G., Kumar, A., Aizawa, S., Murano, K., Nagata, K., Hayashi, Y., Fukuda, A., Onuma, Y., et al. (2019). Template Activating Factor-I alpha Regulates Retroviral Silencing during Reprogramming. Cell Rep 29, 1909-1922 e1905.

Gao, T., and Qian, J. (2020). EnhancerAtlas 2.0: An updated resource with enhancer annotation in 586 tissue/cell types across nine species. Nucleic Acids Res. *48*, D58–D64.

Giorgetti L, Lajoie BR, Carter AC, Attia M, Zhan Y, Xu J, Chen CJ, Kaplan N, Chang HY, Heard E, et al. (2016). Structural organization of the inactive X chromosome in the mouse. Nature *535*: 575–579.

Kobayashi S, Hosoi Y, Shiura H, Yamagata K, Takahashi S, Fujihara Y, Kohda T, Okabe M, Ishino F. (2016). Live imaging of X chromosome reactivation dynamics in early mouse development can discriminate naïve from primed pluripotent stem cells. Development *143*: 2958–2964.

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics *26*, 589–595.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

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

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

Nishimura K, Segawa H, Goto T, Morishita M, Masago A, Takahashi H, Ohmiya Y, Sakaguchi T, Asada M, Imamura T, et al. (2007). Persistent and stable gene expression by a cytoplasmic RNA replicon based on a noncytopathic variant sendai virus. J Biol Chem *282*: 27383–27391.