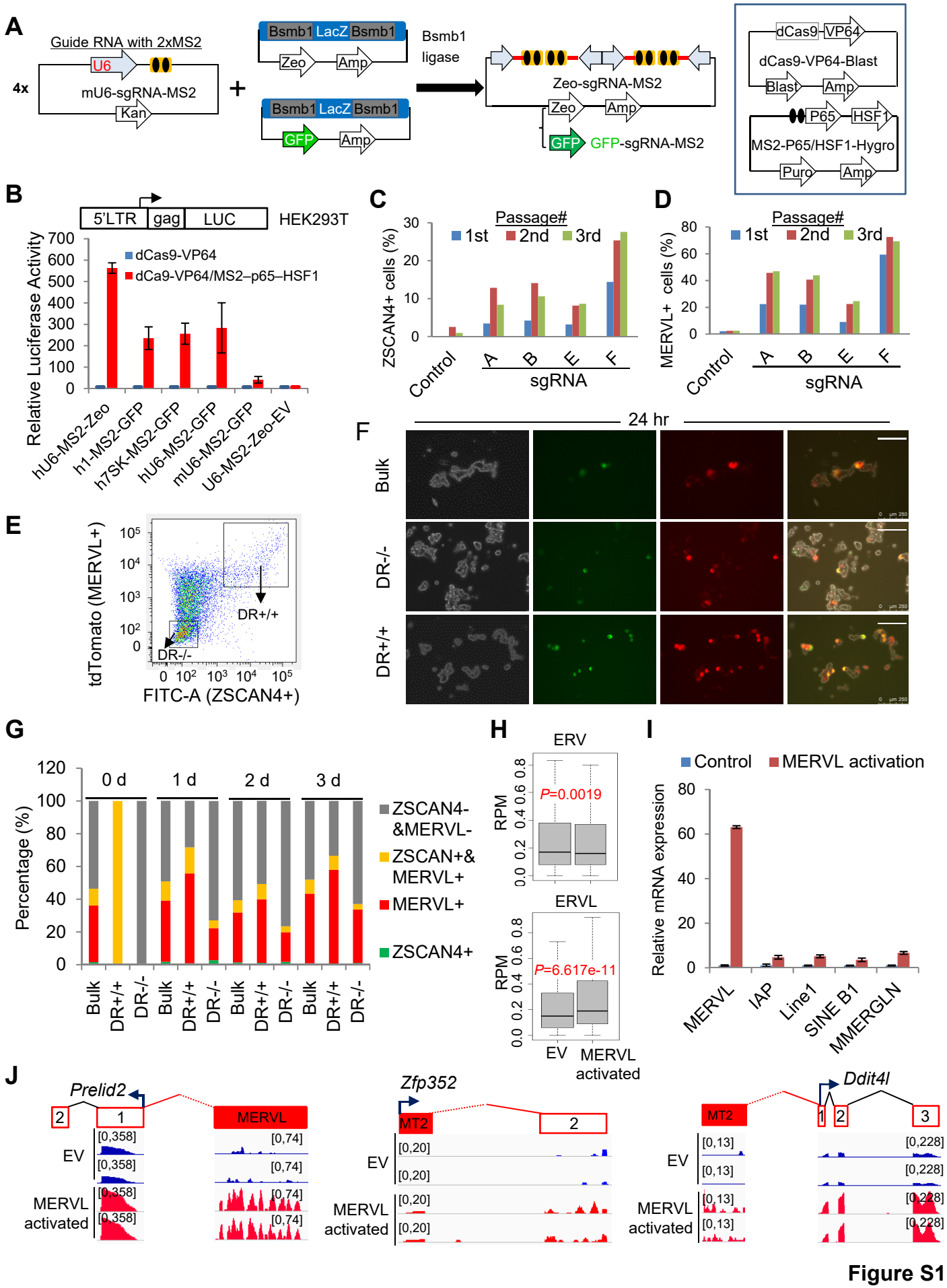

A
Guide RNA with 2xMS2
U6
4x
mU6-sgRNA-MS2
Kan
+
Bsmb1 LacZ Bsmb1
Zeo
Amp
Bsmb1 LacZ Bsmb1
GFP
Amp
Bsmb1 ligase
Zeo-sgRNA-MS2
Zeo
Amp
GFP
GFP-sgRNA-MS2
dCas9 VP64
dCas9-VP64-Blast
Blast
Amp
P65
HSF1
MS2-P65/HSF1-Hygro
Puro
Amp
B
5'LTR
gag
LUC
HEK293T
dCas9-VP64
dCa9-VP64/MS2–p65–HSF1
Relative Luciferase Activity
hU6-MS2-Zeo
h1-MS2-GFP
h7SK-MS2-GFP
hU6-MS2-GFP
mU6-MS2-GFP
U6-MS2-Zeo-EV
C
Passage#
1st
2nd
3rd
ZSCAN4+ cells (%)
Control
A
B
E
F
sgRNA
D
Passage#
1st
2nd
3rd
MERVL+ cells (%)
Control
A
B
E
F
sgRNA
E
tdTomato (MERVL+)
DR+/+
DR-/-
FITC-A (ZSCAN4+)
F
24 hr
Bulk
DR-/-
DR+/+
G
Percentage (%)
0 d
1 d
2 d
3 d
Bulk
DR+/+
DR-/-
ZSCAN4-&MERVL-
ZSCAN+&MERVL+
MERVL+
ZSCAN4+
H
ERV
RPM
P=0.0019
ERVL
RPM
P=6.617e-11
EV
MERVL activated
I
Control
MERVL activation
Relative mRNA expression
MERVL
IAP
Line1
SINE B1
MMERGLN
J
Prelid2
2
1
MERVL
EV
MERVL activated
[0,358]
[0,74]
Zfp352
MT2
2
EV
MERVL activated
[0,20]
Ddit4l
MT2
1
2
3
EV
MERVL activated
[0,13]
[0,228]

Figure S1

Figure S1. MERVL activation is sufficient to induce 2C-like cells. Related to Figure 1.

(A) Strategy of golden gate assembly to create lentiviral vectors encoding CRISPR/dCas9 and multiple sgRNA expression cassettes.

(B) Relative luciferase activity by U6/h1-driven 1x "F" sgRNA with (red) or without (blue) MS2 in HEK293T cells. Fold activation denotes relative luciferase activity normalized to an empty sgRNA expression vector. Data are presented as average ± SD.

(C-D) FACS profiles of Zscan4+ (C) and MERVL+ (D) populations after MERVL activation by different sgRNAs after 3 passages. ESCs were passed every three days.

(E) FACS plot of the pZscan4c-GFP (x-axis) and MERVL::tdTomato (y-axis) dual-reporter ESC line (DR+/+).

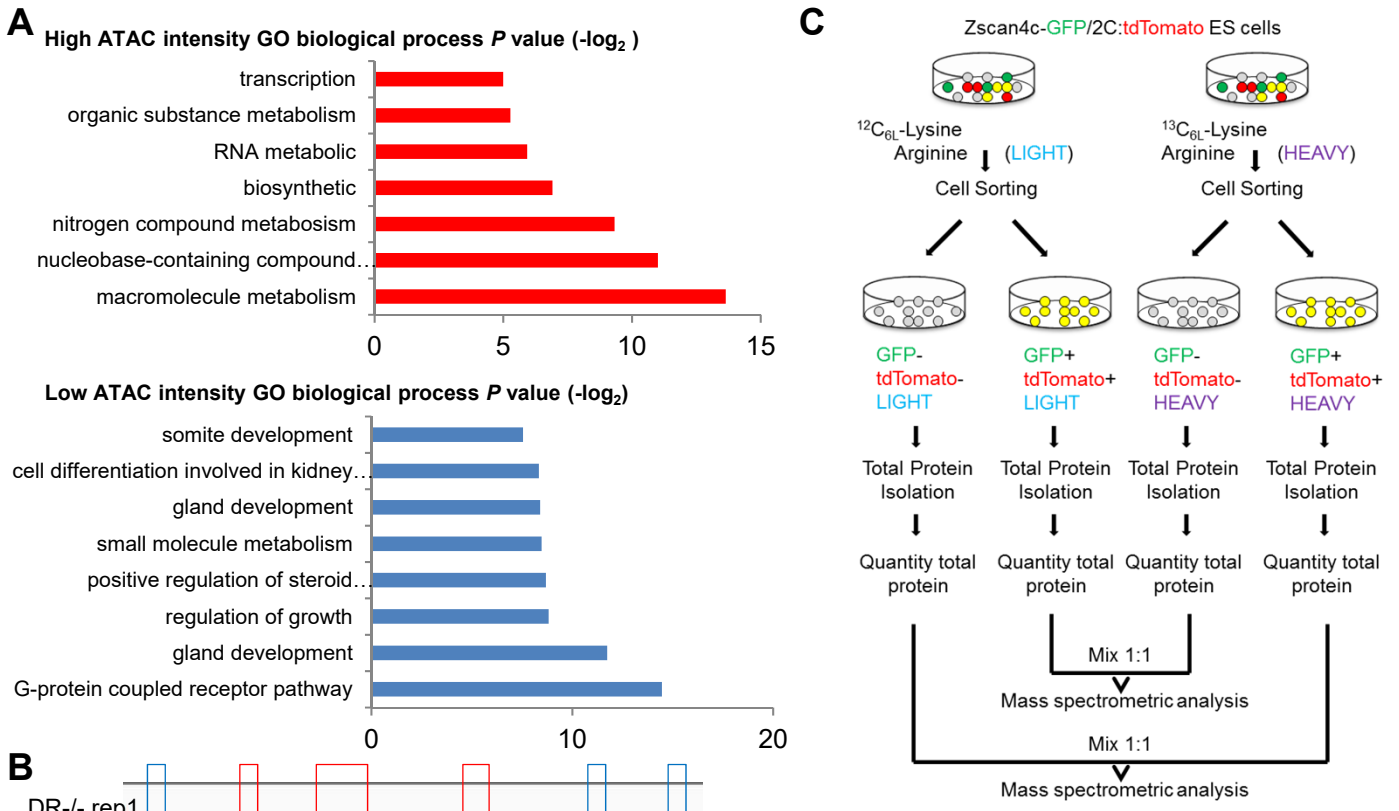
(F) Images of bulk (top), FACS-sorted DR-/- (middle), and DR+/+ (bottom) cells replated separately, then cultured for 24h. Scale bars, 250 μm.

(G) The proportion of MERVL+, Zscan4+, MERVL+Zscan4+ (DR+/+), and MERVL-Zscan4- (DR-/-) cells of bulk and FACS-sorted DR-/- and DR+/+ cells separately cultured for 3 days.

(H) Expression of total ERV and ERVL elements in the empty vector (EV) and MERVL-activated cells. P-value is from Mann-Whitney test.

(I) Relative MERVL induction levels in control and MERVL-activated cells analyzed by RT-qPCR. Multiple ERV families were tested. Data are presented as average ± SD.

(J) Examples of induced expression of MERVL-proximal genes *Zfp352, Prelid2, and Ddit4l* in MERVL-activated cells compared with empty vector (EV) treated cells. Locations of the coding exons and the nearby MERVL solo LTR (MT2, for *Ddit4l*, *Zfp352*) or entire MERVL (for *Prelid2*) regions are connected by dotted red lines. RNA-seq peaks were shown.



Gene	Oocyte Mean	2C Mean	Fold Change	P-value
Gm5662	40.68	1223.68	4.91	7.70E-16
Zscan4b	108.00	2382.12	4.46	2.05E-16
Zscan4c	206.00	2640.00	3.68	1.09E-12
Zscan4d	184.67	2438.12	3.72	1.13E-12
Zscan4e	109.24	2377.89	4.44	2.33E-16
Zscan4f	186.55	2677.18	3.84	1.46E-13

Figure S2

Figure S2. Discovery of miR-344 as a totipotency-associated miRNA. Related to Figure 2.

(A) Gene Ontology (GO) analysis of biological processes of genes with significantly enriched ATAC-seq signals between DR-/- and DR+/+ cells.

(B) The genome browser view of ATAC-seq signals enriched at the *Zscan4* gene cluster.

(C) The strategy of SILAC labeling and LC-MS/MS analysis. Two biological replicates with reciprocal labeling were performed to compare the proteome of DR-/- and DR+/+ cells.

(D) Expression of 2C-specific genes *Gm5662* and *Zscan4b/c/d/e/f* in the oocyte and 2C embryo. Data were from a public RNA-seq study (Macfarlan et al., 2012).

(E) RT-qPCR analysis of totipotency related genes in DR+/+ and DR-/- cells. Data are presented as average ± SD.

(F) Expression of miR-344 during mouse preimplantation embryo. Data were from a public miRNA array study (Liu et al. 2012).

(G) Expression of MERVL elements MERVL-int and MT2_Mm during mouse preimplantation development. Data were from a public RNA-seq study (Xue et al., 2013).

Figure S3

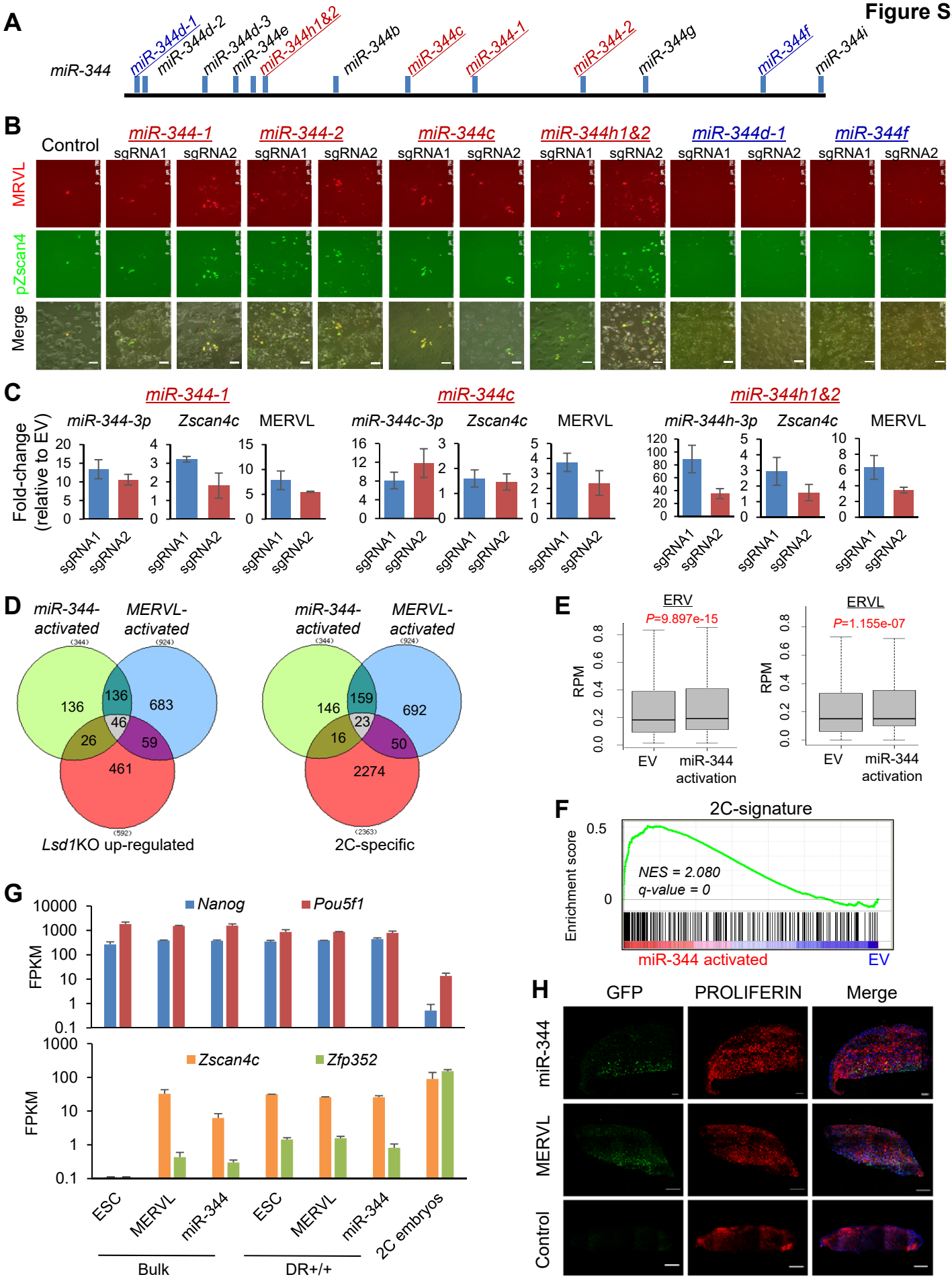

A
miR-344
miR-344d-1
miR-344d-2
miR-344d-3
miR-344e
miR-344h1&2
miR-344b
miR-344c
miR-344-1
miR-344-2
miR-344g
miR-344f
miR-344i
B
Control
miR-344-1
miR-344-2
miR-344c
miR-344h1&2
miR-344d-1
miR-344f
sgRNA1
sgRNA2
MRVL
pZscan4
Merge
C
miR-344-1
miR-344c
miR-344h1&2
miR-344-3p
miR-344c-3p
miR-344h-3p
Zscan4c
MERVL
Fold-change (relative to EV)
sgRNA1
sgRNA2
D
miR-344-activated
MERVL-activated
(344)
(924)
136
136
683
46
26
59
461
(592)
Lsd1KO up-regulated
miR-344-activated
MERVL-activated
(344)
(924)
146
159
692
23
16
50
2274
(2363)
2C-specific
E
ERV
P=9.897e-15
ERVL
P=1.155e-07
RPM
EV
miR-344 activation
F
2C-signature
Enrichment score
NES = 2.080
q-value = 0
miR-344 activated
EV
G
Nanog
Pou5f1
FPKM
Zscan4c
Zfp352
ESC
MERVL
miR-344
2C embryos
Bulk
DR+/+
H
GFP
PROLIFERIN
Merge
miR-344
MERVL
Control

Figure S3. Endogens *miR-344* activation by CRISPR^{SAM} induces 2C-like cells with expanded potency in vivo. Related to Figure 3.

(A) Schematic diagram of the mouse *miR-344* cluster loci, which contains 13 individual miR genes. 12 sgRNAs were designed to activate 6 *miR-344* genes (2 sgRNAs for each gene), including miR-344-1/2/c/h (red) that are expressed in DR+/+ cells, and miR-344-d/f (blue) as negative controls.

(B) Images of ESCs transfected with double (pZscan4c-GFP/2C::tdTomato) reporters and sgRNAs for activating different *miR-344* cluster genes, with 2 sgRNAs for each gene. Refer to Figure 3A. Scale bars, 250 μm.

(C) Relative expression of mature miR-344, 2C-specific genes *Zscan4c* and *MERVL* in *miR-344-1-*, *miR-344c-/miR-344h1&2*-activated cells. Data are presented as average ± SD

(D) Venn diagrams showing overlapped genes up-regulated in MERVL- and miR-344-activated ESCs, and genes upregulated in *Lsd1*KO ESCs or 2C-specific genes.

(E) Box plots showing the expression of total ERV and ERVL elements in the empty vector (EV) and *miR-344-2*-activated cells. P-value is from Mann-Whitney test.

(F) Gene set enrichment analysis (GSEA) indicating that upregulated genes by miR-344 activation were highly enriched in the 2-cell embryo gene set (Wu et al., 2016). Red, upregulated genes; blue, downregulated genes.

(G) Expression of pluripotency genes (*Nanog*, *Pou5f1*) and 2C genes (*Zscan4c*, *Zfp352*) in bulk or DR+/+ sorted ESCs or MERVL- and miR344-activated ESCs. Data are presented as average ± SD.

(H) Representative whole-placenta confocal images showing both GFP-labeled miR-344-activated and MERVL-activated mESCs can contribute to trophoblastic lineages in chimeric E12.5 placentae. Scale bars, 200 μm.

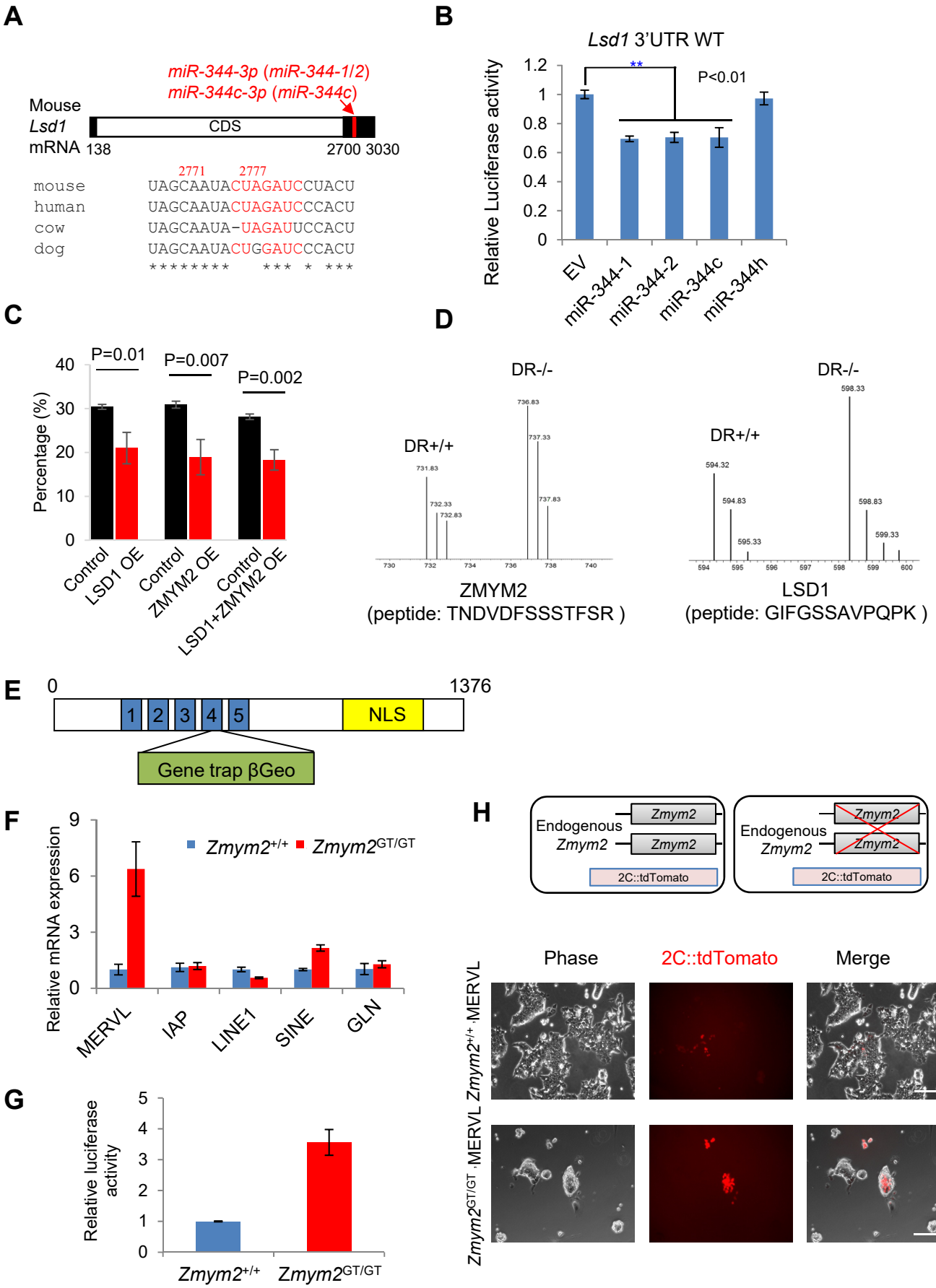

A
miR-344-3p (miR-344-1/2)
miR-344c-3p (miR-344c)
Mouse
Lsd1
mRNA
CDS
138
2700 3030
2771
2777
mouse UAGCAAUACUAGAUCCUACU
human UAGCAAUACUAGAUCCCACU
cow UAGCAAUA-UAGAUUCCACU
dog UAGCAAUACUGGAUCCCACU
B
Lsd1 3'UTR WT
Relative Luciferase activity
**
P<0.01
EV
miR-344-1
miR-344-2
miR-344c
miR-344h
C
Percentage (%)
P=0.01
P=0.007
P=0.002
Control
LSD1 OE
Control
ZMYM2 OE
Control
LSD1+ZMYM2 OE
D
DR-/-
DR+/+
ZMYM2
(peptide: TNDVDFSSSTFSR)
LSD1
(peptide: GIFGSSAVPQPK)
E
0
1376
1
2
3
4
5
NLS
Gene trap βGeo
F
Relative mRNA expression
Zmym2+/+
Zmym2GT/GT
MERVL
IAP
LINE1
SINE
GLN
G
Relative luciferase activity
Zmym2+/+
Zmym2GT/GT
H
Endogenous Zmym2
Zmym2
2C::tdTomato
Phase
2C::tdTomato
Merge
Zmym2+/+ ;MERVL
Zmym2GT/GT ;MERVL

Figure S4

Figure S4. Zmym2 and Lsd1 are the targets of miR-344. Related to Figure 4.

(A) Schematic layout (left) of the *Lsd1* mRNA with predicted miR-344 binding sites indicated. The *miR-344-3p/miR-344c-3p* seed sequence (shown in red) are evolutionarily conserved among different species.

(B) Luciferase reporter assays in mESCs co-transfected with constructs containing the *Lsd1* 3'UTRs and *miR-344-1/2/c/h* expression vectors or empty vector (EV). *miR-344-h* is a non-targeting control. Data are presented as average ± SD. P-value is from T-test.

(C) The percentage of MERVL+ cells in *miR-344-2*-activated cells when overexpressing LSD1, ZMYM2, or LSD1+ZMYM2 (L+Z). Data are presented as average ± SD.

(D) Relative mass spectrums of ZMYM2 (top) and LSD1 (bottom) peptides in DR+/+ and DR-/- cells.

(E) Schematic diagram of the *Zmym2* allele for the gene trap (GT) knockout.

(F) RT-qPCR analysis of repetitive elements MERVL, IAP, LINE L1 (LINE1), SINE B1 (SINE), and MMERGLN (GLN) expression in Zmym2 $^{+/+}$ and Zmym2$^{GT/GT}$ ESCs. Data are presented as average ± SD.

(G) Luciferase activity assay of MERVL LTR reporter in Zmym2 $^{+/+}$ and Zmym2$^{GT/GT}$ ESCs. Data are presented as average ± SD.

(H) Diagram (top) and images (bottom) of Zmym2$^{+/+}$ and Zmym2$^{GT/GT}$ ESCs with 2C::tdTdTomato reporter. Scale bars, 100 μm.

Figure S5

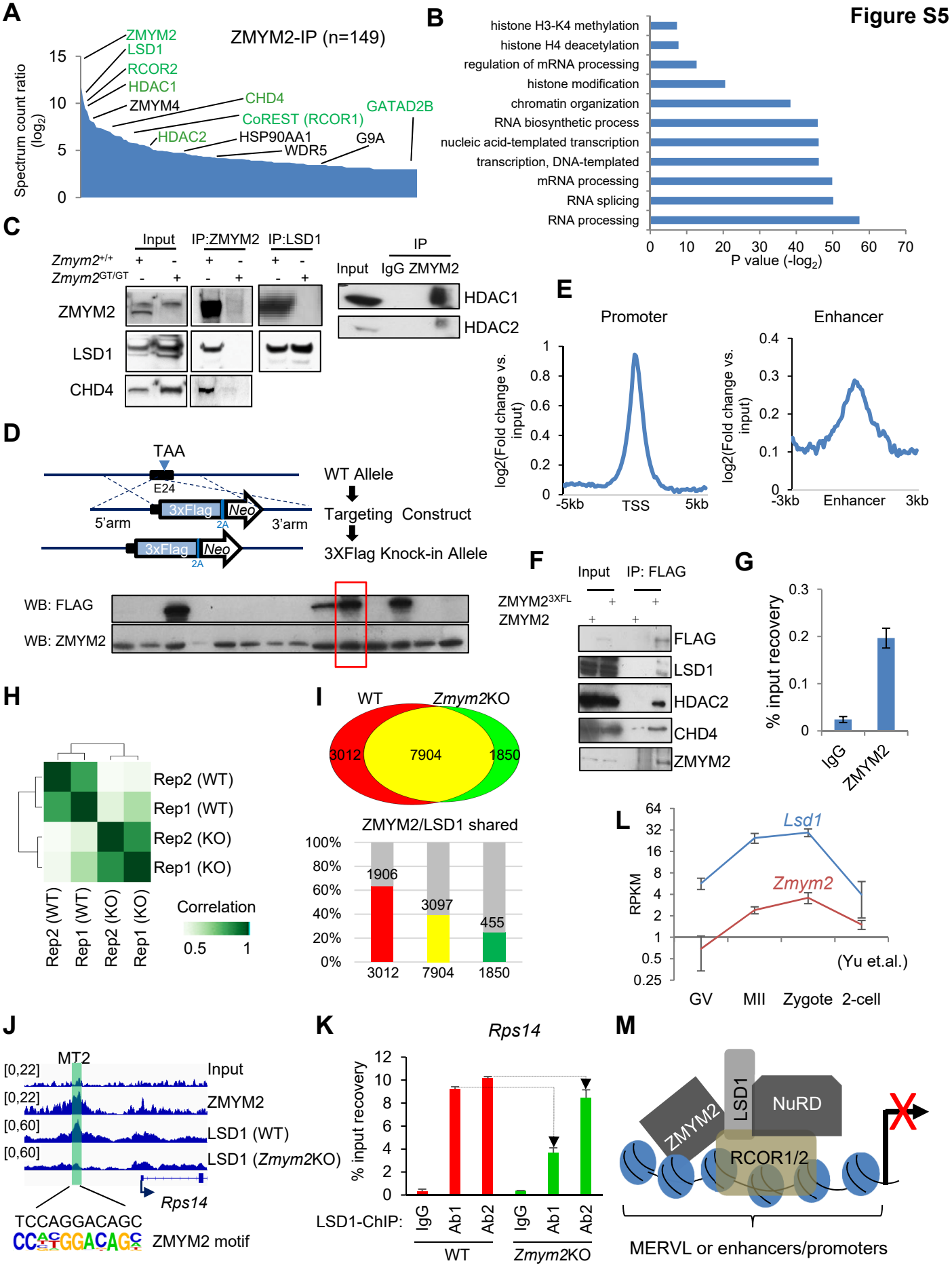

A
ZMYM2-IP (n=149)
Spectrum count ratio (log2)
ZMYM2
LSD1
RCOR2
HDAC1
ZMYM4
CHD4
CoREST (RCOR1)
HDAC2
HSP90AA1
WDR5
G9A
GATAD2B
B
histone H3-K4 methylation
histone H4 deacetylation
regulation of mRNA processing
histone modification
chromatin organization
RNA biosynthetic process
nucleic acid-templated transcription
transcription, DNA-templated
mRNA processing
RNA splicing
RNA processing
P value (-log2)
C
Input
IP:ZMYM2
IP:LSD1
Zmym2+/+
Zmym2GT/GT
ZMYM2
LSD1
CHD4
IP
Input IgG ZMYM2
HDAC1
HDAC2
D
TAA
E24
3xFlag
Neo
2A
5'arm
3'arm
WT Allele
Targeting Construct
3XFlag Knock-in Allele
WB: FLAG
WB: ZMYM2
E
Promoter
Enhancer
log2(Fold change vs. input)
-5kb
TSS
5kb
-3kb
Enhancer
3kb
F
Input
IP: FLAG
ZMYM2 3XFL
ZMYM2
FLAG
LSD1
HDAC2
CHD4
ZMYM2
G
% input recovery
IgG
ZMYM2
H
Rep2 (WT)
Rep1 (WT)
Rep2 (KO)
Rep1 (KO)
Correlation
0.5
1
I
WT
Zmym2KO
3012
7904
1850
ZMYM2/LSD1 shared
100%
80%
60%
40%
20%
0%
1906
3097
455
L
Lsd1
Zmym2
RPKM
(Yu et.al.)
GV
MII
Zygote
2-cell
J
MT2
[0,22]
Input
ZMYM2
[0,60]
LSD1 (WT)
LSD1 (Zmym2KO)
Rps14
TCCAGGACAGC
ZMYM2 motif
K
Rps14
% input recovery
LSD1-ChIP:
IgG
Ab1
Ab2
WT
Zmym2KO
M
ZMYM2
LSD1
NuRD
RCOR1/2
MERVL or enhancers/promoters

Figure S5. ZMYM2 physically associates and functionally interacts with the LSD1-NuRD corepressor complex in controlling MERVL expression. Related to Figure 5.

(A) AP-MS identification of ZMYM2-interacting partners (n=149). Data are presented by fold change ($Zmym2^{GT/GT}/Zmym2^{+/+}$) of spectrum counts for proteins identified in ESCs.

(B) Gene Ontology (GO) analysis of biological processes of ZMYM2 partners.

(C) CoIP validation of physical interactions of ZMYM2 with LSD1/CHD4 and HDAC1/2.

(D) Targeted knock-in of 3xFLAG into the *Zmym2* locus in mESCs (top) and western blotting analysis of clones. The clone indicated with a red rectangle was used for FLAG ChIP-seq.

(E) Enrichment of ZMYM2 binding around TSS regions and ESC enhancer regions.

(F) FLAG CoIP validation in ZMYM2-3xFLAG ESCs to test the interactions of ZMYM2 with LSD1, CHD4, and HDAC2.

(G) ChIP-qPCR analysis of ZMYM2 enrichment at MERVL LTR region (MT2_Mm). Data are presented as average ± SD.

(H) Hierarchy clustering analysis of LSD1 ChIP-seq in *Zmym2* WT versus KO ESCs. Two biological replicates were performed in each condition.

(I) Overlap of the LSD1 ChIP-seq peaks in *Zmym2* WT versus KO ESCs. The number and percentage of ZMYM2 peaks colocalized in each group of LSD1 peaks are indicated at the bottom.

(J) ZMYM2 and LSD1 cobind at *Rps14* promoter with an MT2 region. ChIP-seq tracks of input and ZMYM2 in ESCs, and of LSD1 in WT and *Zmym2*KO ESCs are indicated.

(K) ChIP-qPCR validation of LSD1, with 2 different antibodies, at promoters of *Rps14* in *Zmym2* WT and KO ESCs. Data are presented as average ± SD.

(L) Relative expression of *Zmym2* and *Lsd1* at different stages (germinal vesicle (GV), MII oocyte, zygote, 2-cell embryo) during ZGA. Data are from public RNA-seq study (Yu et al., 2016). Data are presented as average ± SD.

(M) A model of ZMYM2-associated LSD1/RCOR1/2-NuRD complex repressing the 2C-specific genes.

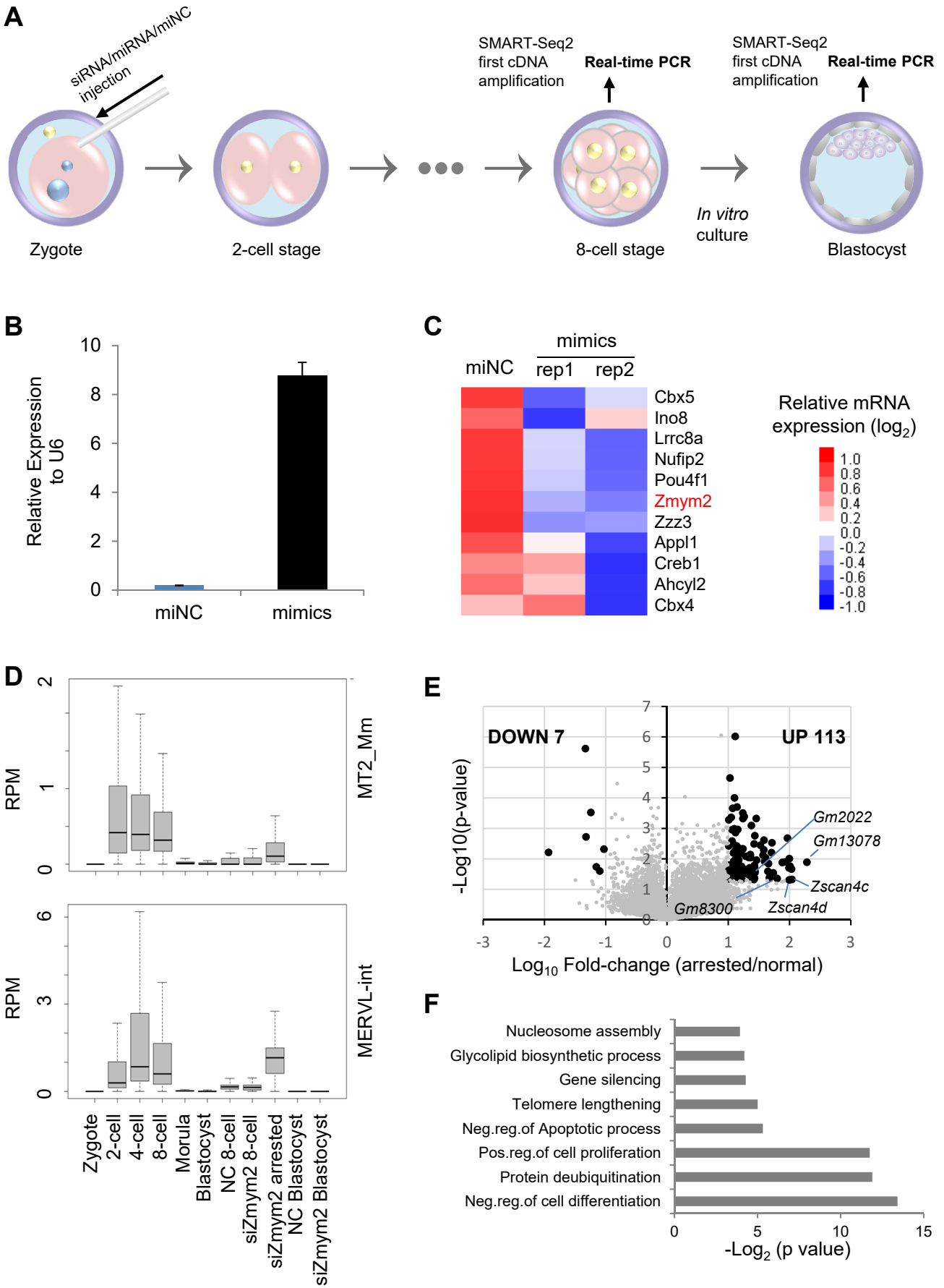

A
siRNA/miRNA/miNC injection
Zygote
2-cell stage
SMART-Seq2 first cDNA amplification
Real-time PCR
8-cell stage
In vitro culture
SMART-Seq2 first cDNA amplification
Real-time PCR
Blastocyst
B
Relative Expression to U6
10
8
6
4
2
0
miNC
mimics
C
mimics
miNC
rep1
rep2
Cbx5
Ino8
Lrrc8a
Nufip2
Pou4f1
Zmym2
Zzz3
Appl1
Creb1
Ahcyl2
Cbx4
Relative mRNA expression (log2)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
D
RPM
MT2_Mm
MERVL-int
Zygote
2-cell
4-cell
8-cell
Morula
Blastocyst
NC 8-cell
siZmym2 8-cell
siZmym2 arrested
NC Blastocyst
siZmym2 Blastocyst
E
DOWN 7
UP 113
-Log10(p-value)
Gm2022
Gm13078
Zscan4c
Gm8300
Zscan4d
Log10 Fold-change (arrested/normal)
F
Nucleosome assembly
Glycolipid biosynthetic process
Gene silencing
Telomere lengthening
Neg.reg.of Apoptotic process
Pos.reg.of cell proliferation
Protein deubiquitination
Neg.reg.of cell differentiation
-Log2 (p value)

Figure S6

Figure S6. Zygotic depletion of Zmym2 arrests development at 2C-stage embryos. Related to Figure 6.

(A) Depiction of zygotic injection with miRNA mimics, nontargeting control (miNC) or siRNA followed by in vitro culture till the blastocyst stage.

(B) RT-qPCR analysis of mature miR-344-3p expression in embryos upon injection with miR-344-3p mimics (mimics) or non-target control (miNC).

(C) Heatmap displaying relative expression of predicted miR-344 targets in embryos upon injection with miR-344-3p mimics (mimics) or non-target control (miNC).

(D) Box plots showing expression of MT2_Mm (top) and MERVL-int (bottom) from RNA-seq data of embryos with siZmym2 or non-target control (NC) injections from this study, and normal embryos at different embryonic stages (Zygote, 2-cell, 4-cell, 8-cell, Morula, Blastocyst) from a previous study (Liu et al., 2016).

(E) Volcano plot showing gene expression between morphologically normal and arrested embryos with siRNA treatment. Black dots and numbers indicate the differentially expressed genes ($Log_{10}FoldChange>10$, p-value<0.05).

(F) Gene ontology analysis for significantly upregulated genes in arrested embryos (n=113) in panel E.

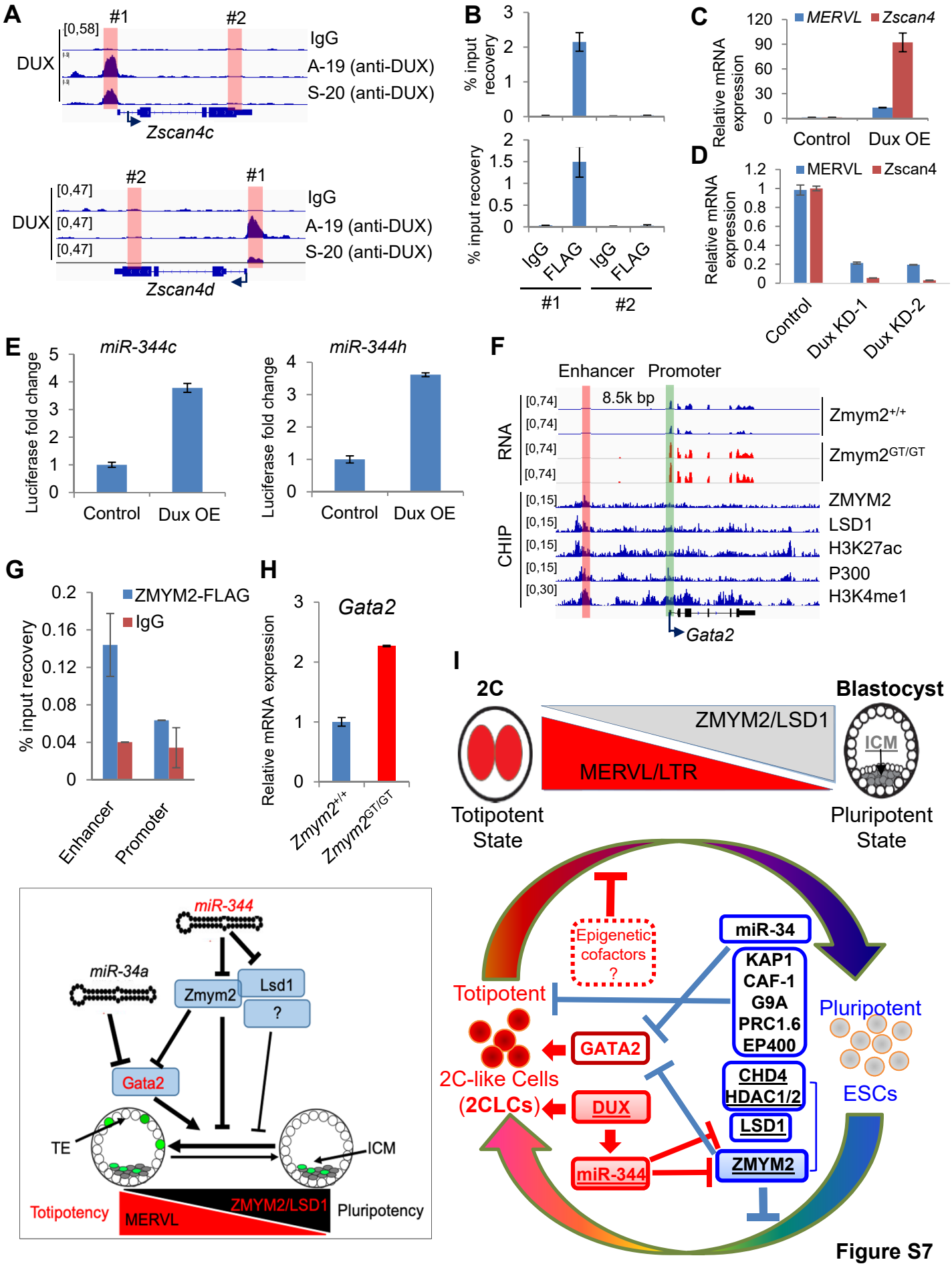

A
#1
#2
[0,58]
DUX
IgG
A-19 (anti-DUX)
S-20 (anti-DUX)
Zscan4c
#2
#1
[0,47]
[0,47]
[0,47]
DUX
IgG
A-19 (anti-DUX)
S-20 (anti-DUX)
Zscan4d
B
% input recovery
IgG
FLAG
IgG
FLAG
#1
#2
% input recovery
C
MERVL
Zscan4
Relative mRNA expression
Control
Dux OE
D
MERVL
Zscan4
Relative mRNA expression
Control
Dux KD-1
Dux KD-2
E
miR-344c
Luciferase fold change
Control
Dux OE
miR-344h
Luciferase fold change
Control
Dux OE
F
Enhancer
Promoter
8.5k bp
RNA
CHIP
[0,74]
[0,74]
[0,74]
[0,74]
[0,15]
[0,15]
[0,15]
[0,15]
[0,30]
Zmym2+/+
Zmym2GT/GT
ZMYM2
LSD1
H3K27ac
P300
H3K4me1
Gata2
G
ZMYM2-FLAG
IgG
% input recovery
Enhancer
Promoter
H
Gata2
Relative mRNA expression
Zmym2+/+
Zmym2GT/GT
I
2C
Blastocyst
ZMYM2/LSD1
MERVL/LTR
ICM
Totipotent State
Pluripotent State
miR-344
miR-34a
Zmym2
Lsd1
?
Gata2
TE
ICM
Totipotency
MERVL
ZMYM2/LSD1
Pluripotency
Epigenetic cofactors ?
miR-34
KAP1
CAF-1
G9A
PRC1.6
EP400
Totipotent
GATA2
2C-like Cells (2CLCs)
DUX
miR-344
CHD4
HDAC1/2
LSD1
ZMYM2
Pluripotent
ESCs

Figure S7

Figure S7. DUX binds to the *miR-344* promoter and activates *miR-344* expression in MERVL+ cells. Related to Figure 7.

(A) DUX ChIP-seq tracks at *Zscan4c* (top) and *Zscan4d* (bottom) loci. Locations of primers for ChIP-qPCR analysis in panel B are indicated.

(B) ChIP-qPCR of DUX enrichment at *Zscan4c* (top) and *Zscan4d* (bottom) loci. Region #1 is with DUX enrichment, whereas #2 is negative from DUX ChIP-seq data. Data are presented as average ± SD.

(C) RT-qPCR of *MERVL* and *Zscan4* expression upon Dux overexpression (OE) in mESCs. Data are presented as average ± SD.

(D) RT-qPCR of *MERVL* and *Zscan4* expression upon Dux knockdown in sorted MERVL+ mESCs. KD-1 and KD-2 are experiments with 2 independent shRNAs. Data are presented as average ± SD.

(E) Luciferase reporter assay in mESCs co-transfected with the *miR-344c* (left) or *miR-344h* (right) reporters and a DUX expression vector or a control vector. Data are presented as average ± SD.

(F) ZMYM2 represses *Gata2* expression by occupying the enhancer of MERVL regulator *Gata2*. RNA-seq expression profile (top) and ChIP-seq binding profiles (reads per million) (bottom) were shown.

(G) ChIP-qPCR of ZMYM2 enrichment at the *Gata2* enhancer and promoter regions. Data are presented as average ± SD.

(H) RT-qPCR analysis of *Gata2* expression in *Zmym2*$^{+/+}$ and *Zmym2*$^{GT/GT}$ ESCs. Data are presented as average ± SD.

(I) A summary model of molecular control of totipotency and pluripotency in vivo (top) and in culture (bottom). The DUX-*miR-344*-Zmym2/Lsd1 axis in regulating embryonic potency revealed by this study is underlined. The molecular pathways involving miR-34a, miR-344, and their downstream targets for totipotency and pluripotency control are further highlighted (bottom left panel). The detailed information is provided in Discussion.