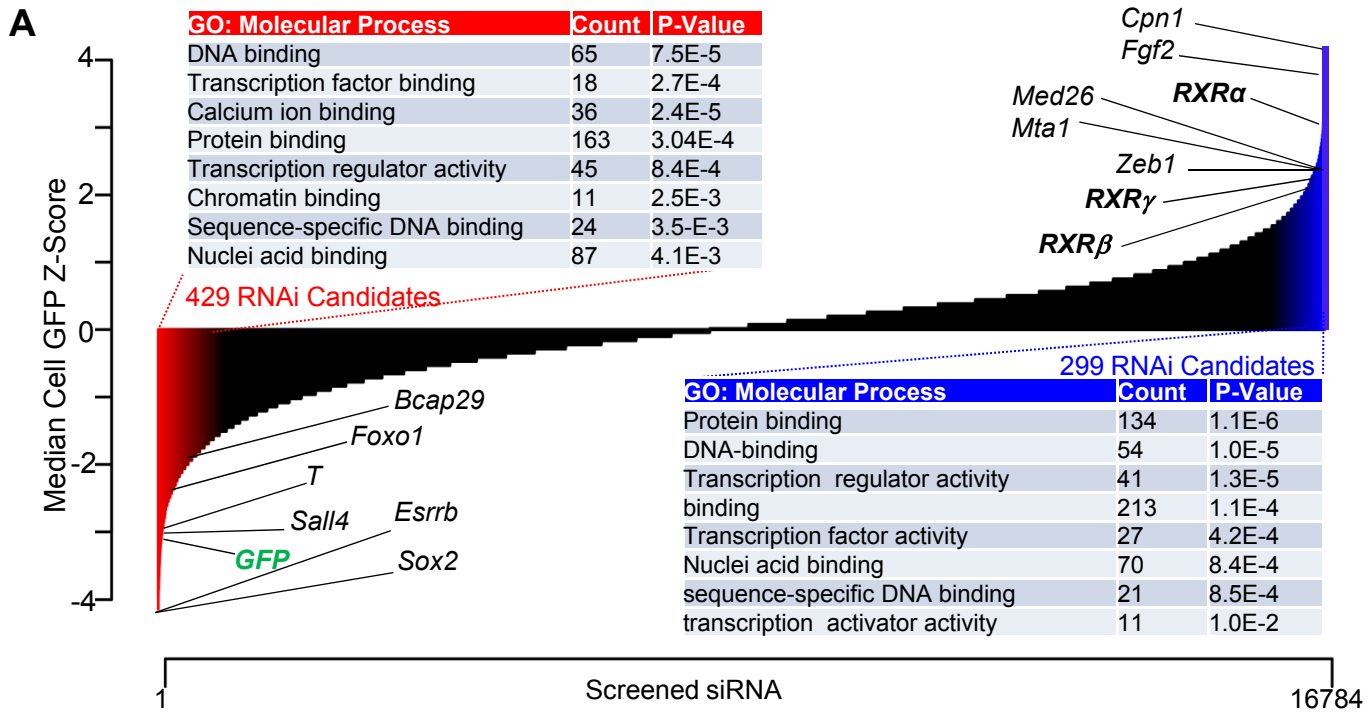
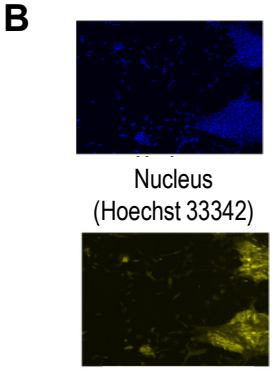
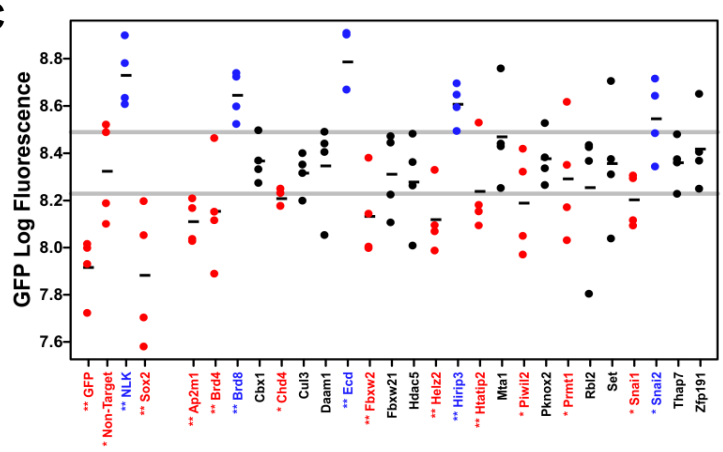

A
GO: Molecular Process | Count | P-Value
DNA binding | 65 | 7.5E-5
Transcription factor binding | 18 | 2.7E-4
Calcium ion binding | 36 | 2.4E-5
Protein binding | 163 | 3.04E-4
Transcription regulator activity | 45 | 8.4E-4
Chromatin binding | 11 | 2.5E-3
Sequence-specific DNA binding | 24 | 3.5-E-3
Nuclei acid binding | 87 | 4.1E-3
Cpn1
Fgf2
Med26
RXRα
Mta1
Zeb1
RXRγ
RXRβ
429 RNAi Candidates
299 RNAi Candidates
GO: Molecular Process | Count | P-Value
Protein binding | 134 | 1.1E-6
DNA-binding | 54 | 1.0E-5
Transcription regulator activity | 41 | 1.3E-5
binding | 213 | 1.1E-4
Transcription factor activity | 27 | 4.2E-4
Nuclei acid binding | 70 | 8.4E-4
sequence-specific DNA binding | 21 | 8.5E-4
transcription activator activity | 11 | 1.0E-2
Median Cell GFP Z-Score
4
2
0
-2
-4
Bcap29
Foxo1
T
Sall4
Esrrb
GFP
Sox2
1
Screened siRNA
16784


B
Nucleus
(Hoechst 33342)
Nanog-GFP


C
GFP Log Fluorescence
8.8
8.6
8.4
8.2
8.0
7.8
7.6
** GFP
* Non-Target
** NLK
** Sox2
** Ap2m1
** Brd4
** Brd8
Cbx1
* Chd4
Cul3
Daam1
** Ecd
** Fbxw2
Fbxw21
Hdac5
** Helz2
** Hirip3
** Htatip2
Mta1
* Piwil2
Pknox2
* Prmt1
Rbl2
Set
* Snai1
* Snai2
Thap7
Zfp191

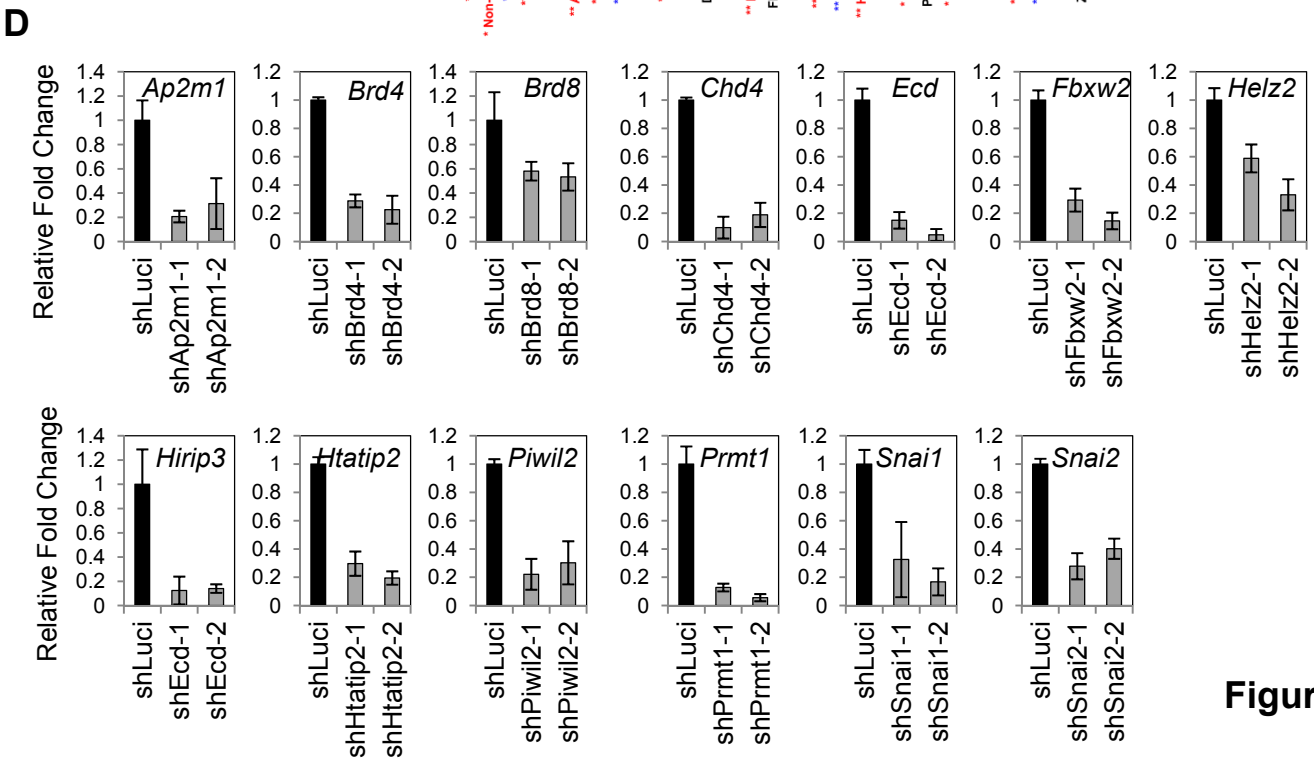
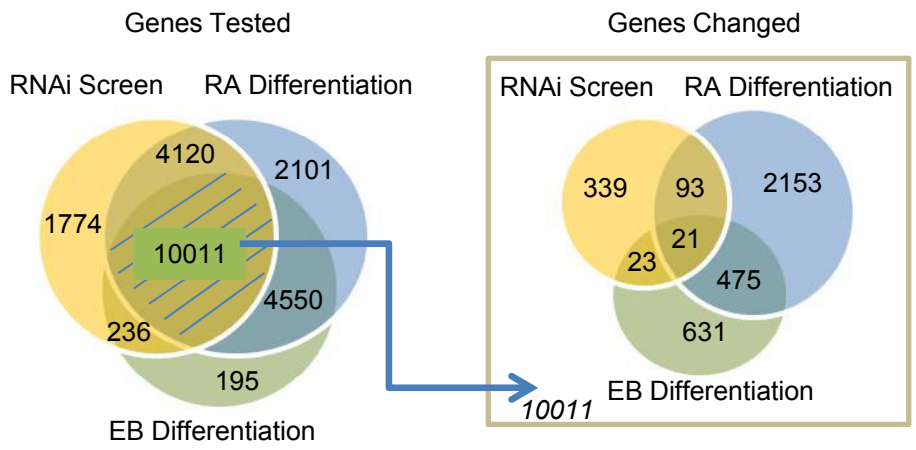

D
Relative Fold Change
Ap2m1
shLuci
shAp2m1-1
shAp2m1-2
Brd4
shBrd4-1
shBrd4-2
Brd8
shBrd8-1
shBrd8-2
Chd4
shChd4-1
shChd4-2
Ecd
shEcd-1
shEcd-2
Fbxw2
shFbxw2-1
shFbxw2-2
Helz2
shHelz2-1
shHelz2-2
Hirip3
Htatip2
shHtatip2-1
shHtatip2-2
Piwil2
shPiwil2-1
shPiwil2-2
Prmt1
shPrmt1-1
shPrmt1-2
Snai1
shSnai1-1
shSnai1-2
Snai2
shSnai2-1
shSnai2-2

Figure S1

E



F

		Changed in RA Differentiation +	-	Total
RNAi Candidate	**+**	114	362	476
	-	2628	6907	9535
	Total	2742	7269	10011

		Changed in EB Differentiation +	-	Total
RNAi Candidate	**+**	44	432	476
	-	1106	8429	9535
	Total	1150	8861	10011

Figure S1

Figure S1. Additional Unpooled siRNA and shRNA Validation of Candidates with Immunofluorescence and Flow Cytometry. Related to Figure 1.

(A) Results from the RNAi primary screen on median cell fluorescence. A selected subset of the outlier and control genes is labeled.

(B) Representative fluorescent images from a Non-Targeting siRNA-treated well. For each well treated with a given siRNA, nuclei-stained cells are identified and levels of GFP are measured.

(C) Levels of GFP for each siRNA-treated well and controls. Dots denote effect averaged across technical triplicate. Short black horizontal lines are average of 4 unpooled siRNA effects. Territory outside of long horizontal lines in grey denotes FDR <0.05 using the empty controls. Single star before gene name indicates 2 of 4 distinct siRNAs are significant with FDR<0.05, with two stars if 3-4 of distinct siRNAs are significant. Further details are in the Supplementary Information.

(D) Relative knockdown efficiencies of shRNAs in pLKO.1 vector. Values are normalized to shLuciferase (shLuci). Error bars indicate average ± SD.

(E) Venn diagram for genes assayed (left) and changed >2-fold following 2 days 2 μM RA or EB differentiation or considered RNAi screen candidates (right). Genes in shaded central region on left are assayed in all datasets.

(F) Contingency table for genes assayed in all datasets and changed following RA or EB differentiation or considered RNAi screen candidates. "+" and "-" indicate inclusion (overlapping) and exclusion (non-overlapping), respectively, in the pairwise comparisons.

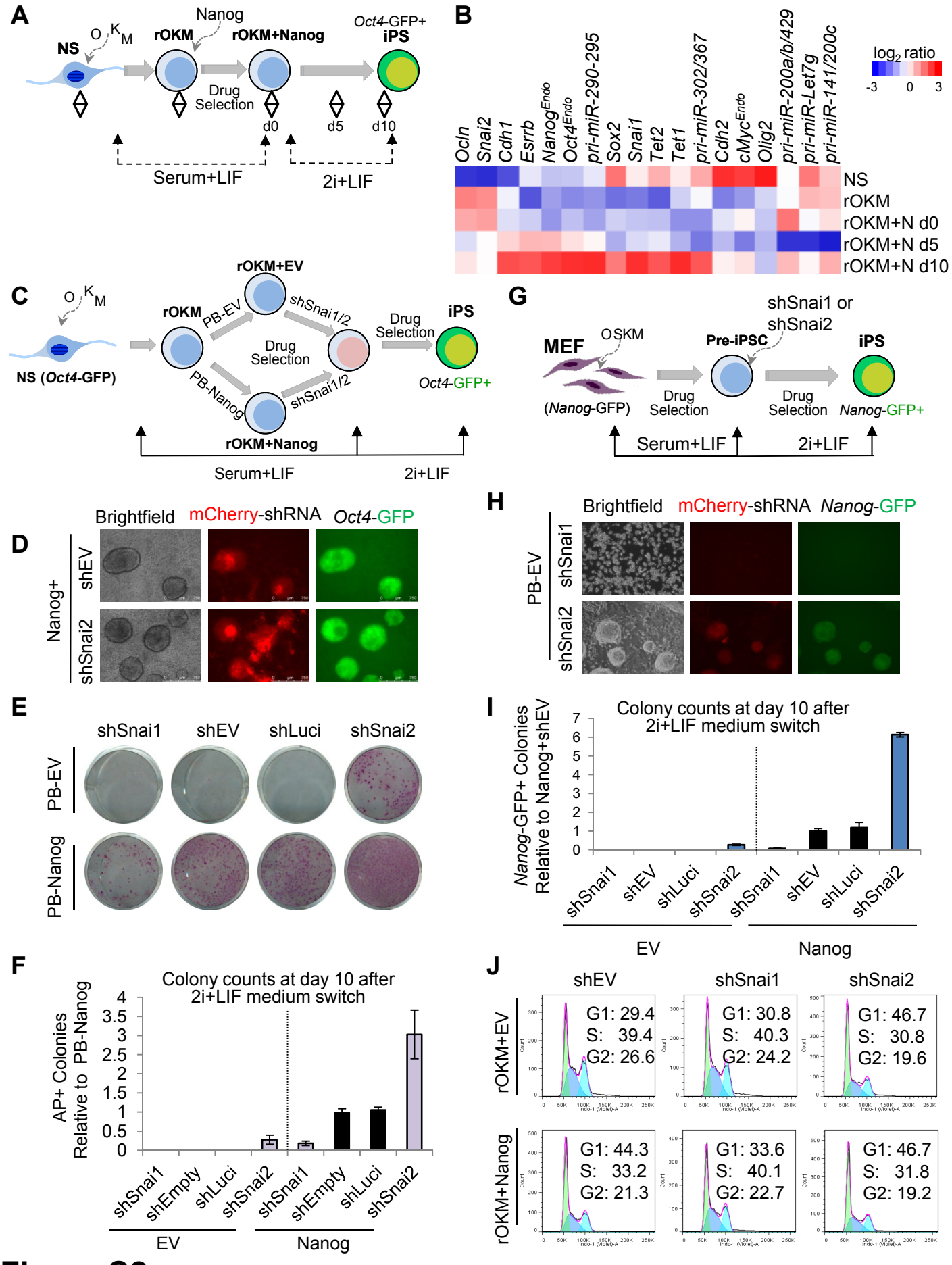

A
NS
O K M
rOKM
Nanog
rOKM+Nanog
Oct4-GFP+
iPS
Drug Selection
d0
d5
d10
Serum+LIF
2i+LIF
B
log2 ratio
-3 0 3
Ocln
Snai2
Cdh1
Esrrb
NanogEndo
Oct4Endo
pri-miR-290-295
Sox2
Snai1
Tet2
Tet1
pri-miR-302/367
Cdh2
cMycEndo
Olig2
pri-miR-200a/b/429
pri-miR-Let7g
pri-miR-141/200c
NS
rOKM
rOKM+N d0
rOKM+N d5
rOKM+N d10
C
NS (Oct4-GFP)
rOKM
rOKM+EV
PB-EV
shSnai1/2
Drug Selection
PB-Nanog
rOKM+Nanog
iPS
Oct4-GFP+
Serum+LIF
2i+LIF
G
MEF
(Nanog-GFP)
OSKM
Pre-iPSC
shSnai1 or shSnai2
iPS
Drug Selection
Nanog-GFP+
Serum+LIF
2i+LIF
D
Brightfield
mCherry-shRNA
Oct4-GFP
Nanog+
shEV
shSnai2
H
Brightfield
mCherry-shRNA
Nanog-GFP
PB-EV
shSnai1
shSnai2
E
shSnai1
shEV
shLuci
shSnai2
PB-EV
PB-Nanog
I
Colony counts at day 10 after 2i+LIF medium switch
Nanog-GFP+ Colonies Relative to Nanog+shEV
shSnai1
shEV
shLuci
shSnai2
EV
Nanog
F
Colony counts at day 10 after 2i+LIF medium switch
AP+ Colonies Relative to PB-Nanog
shSnai1
shEmpty
shLuci
shSnai2
EV
Nanog
J
shEV
shSnai1
shSnai2
rOKM+EV
G1: 29.4
S: 39.4
G2: 26.6
G1: 30.8
S: 40.3
G2: 24.2
G1: 46.7
S: 30.8
G2: 19.6
rOKM+Nanog
G1: 44.3
S: 33.2
G2: 21.3
G1: 33.6
S: 40.1
G2: 22.7
G1: 46.7
S: 31.8
G2: 19.2

Figure S2

Figure S2. Loss-of-Function Studies of Snai1 and Snai2 in Nanog-Driven Pre-iPSC Reprogramming. Related to Figure 2.

(A) Schematic overview of Nanog-dependent NS pre-iPSC reprogramming.

(B) Relative gene expression during the pre-iPSC reprogramming process. The time points of gene expression analysis are indicated. Relative gene expression in NS, rOKM and rOKM+Nanog (N) cells was measured by qRT-PCR. Data are normalized to *Gapdh*.

(C) Schematic overview of pre-iPSC reprogramming.

(D) Representative images of Nanog±shSnai2 reprogramming of *Oct4*-GFP reporter pre-iPSCs. shRNA is marked by mCherry. Data from one shRNA are shown.

(E) Images of AP stained colonies generated as indicated in D.

(F) Quantitation of AP+ colonies at day 10 after reprogramming with indicated shRNAs. Data are represented as average ± SD of 2 independent shRNAs against each Snai1 and Snai2.

(G) Schematic overview of pre-iPSC reprogramming in MEFs harboring the *Nanog*-GFP reporter transgene.

(H) Representative images of shSnai2 reprogramming of MEF-derived pre-iPSCs. shRNA vector is marked by mCherry and final iPSCs are positive for *Nanog*-GFP reporter activity.

(I) Quantitation of *Nanog*-GFP+ colonies at day 10 after reprogramming with indicated shRNAs. Data are represented as average ± SD of 2 independent shRNAs against each Snai1 and Snai2.

(J) Cell cycle analyses by flow cytometry upon Snai1/2 depletion in pre-iPSCs (rOKM) with or without ectopic Nanog expression.

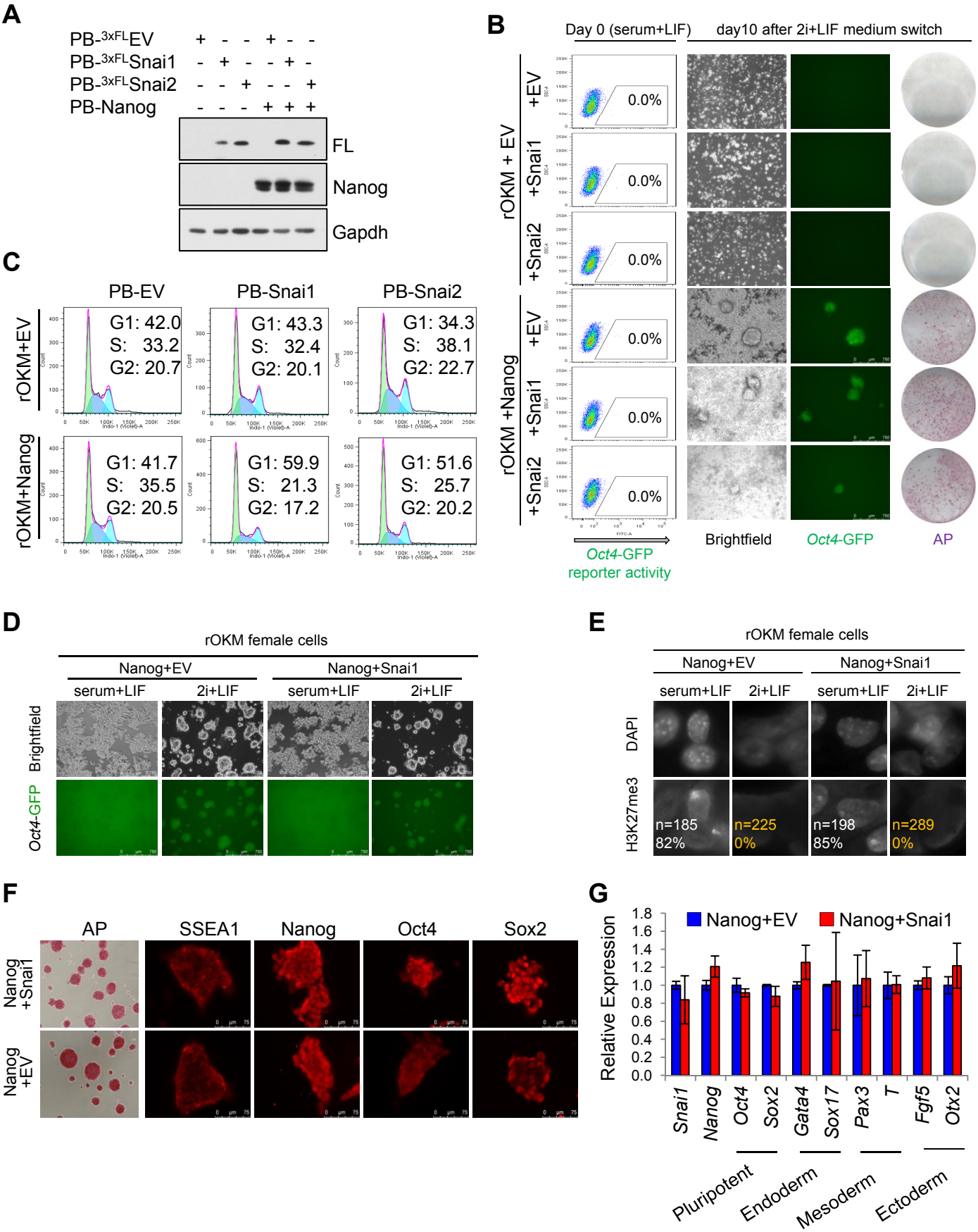
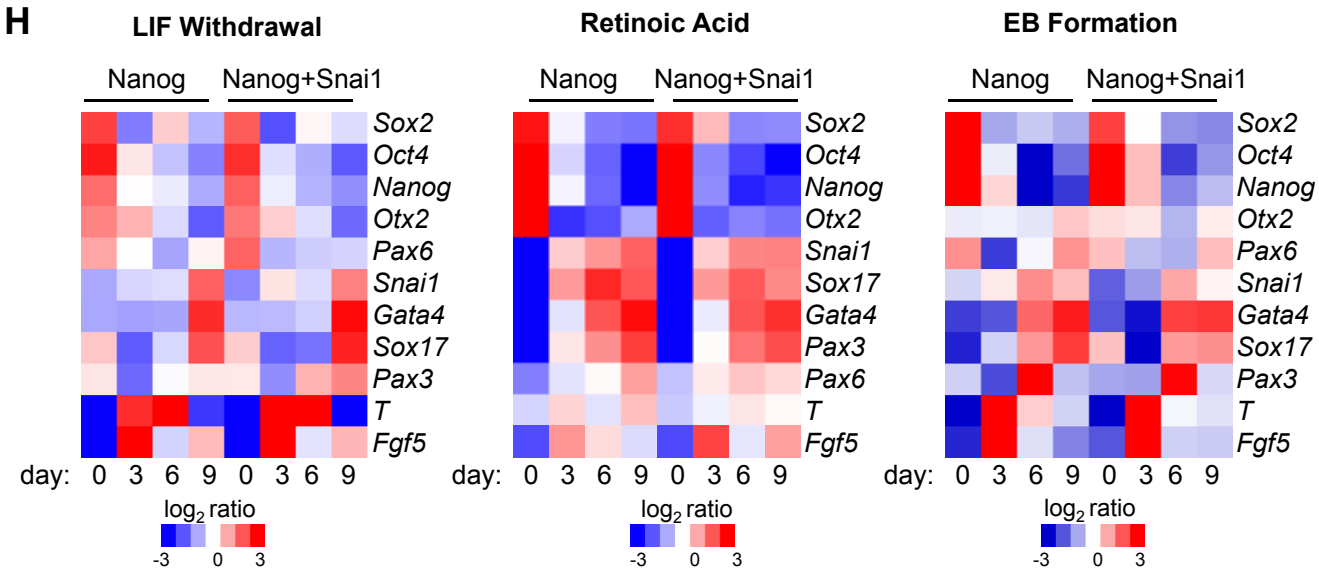
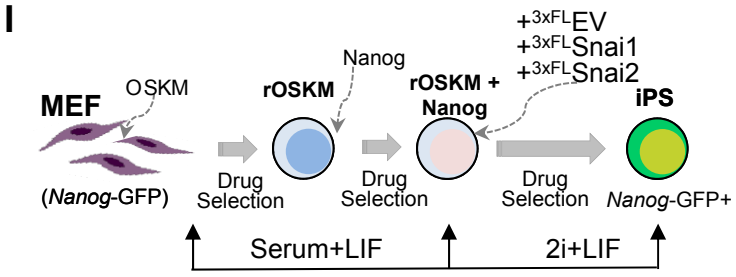
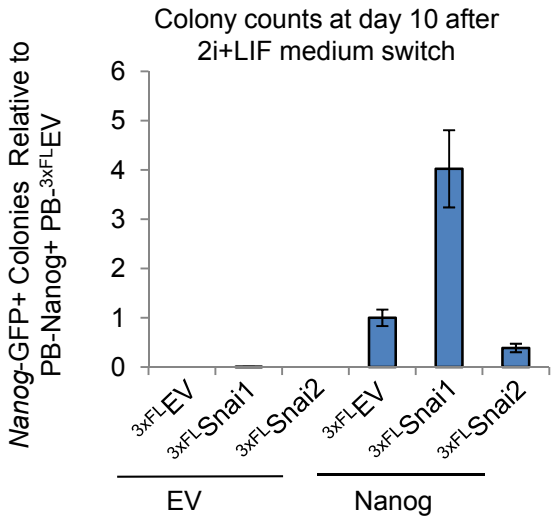

A
PB-3xFLEV + - - + - -
PB-3xFLSnai1 - + - - + -
PB-3xFLSnai2 - - + - - +
PB-Nanog - - - + + +
FL
Nanog
Gapdh
B
Day 0 (serum+LIF)
day10 after 2i+LIF medium switch
rOKM + EV
+EV
+Snai1
+Snai2
rOKM +Nanog
+EV
+Snai1
+Snai2
0.0%
Oct4-GFP reporter activity
Brightfield
Oct4-GFP
AP
C
PB-EV
PB-Snai1
PB-Snai2
rOKM+EV
G1: 42.0
S: 33.2
G2: 20.7
G1: 43.3
S: 32.4
G2: 20.1
G1: 34.3
S: 38.1
G2: 22.7
rOKM+Nanog
G1: 41.7
S: 35.5
G2: 20.5
G1: 59.9
S: 21.3
G2: 17.2
G1: 51.6
S: 25.7
G2: 20.2
D
rOKM female cells
Nanog+EV
Nanog+Snai1
serum+LIF
2i+LIF
serum+LIF
2i+LIF
Brightfield
Oct4-GFP
E
rOKM female cells
Nanog+EV
Nanog+Snai1
serum+LIF
2i+LIF
serum+LIF
2i+LIF
DAPI
H3K27me3
n=185
82%
n=225
0%
n=198
85%
n=289
0%
F
AP
SSEA1
Nanog
Oct4
Sox2
Nanog +Snai1
Nanog +EV
G
Nanog+EV
Nanog+Snai1
Relative Expression
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Snai1
Nanog
Oct4
Sox2
Gata4
Sox17
Pax3
T
Fgf5
Otx2
Pluripotent
Endoderm
Mesoderm
Ectoderm

Figure S3


H
LIF Withdrawal
Retinoic Acid
EB Formation
Nanog
Nanog+Snai1
Sox2
Oct4
Nanog
Otx2
Pax6
Snai1
Gata4
Sox17
Pax3
T
Fgf5
day: 0 3 6 9 0 3 6 9
log2 ratio
-3 0 3


I
MEF
OSKM
(Nanog-GFP)
rOSKM
Nanog
rOSKM + Nanog
+3xFLEV
+3xFLSnai1
+3xFLSnai2
iPS
Nanog-GFP+
Drug Selection
Drug Selection
Drug Selection
Serum+LIF
2i+LIF


J
Colony counts at day 10 after 2i+LIF medium switch
Nanog-GFP+ Colonies Relative to PB-Nanog+ PB-3xFLEV
0 1 2 3 4 5 6
3xFLEV
3xFLSnai1
3xFLSnai2
3xFLEV
3xFLSnai1
3xFLSnai2
EV
Nanog

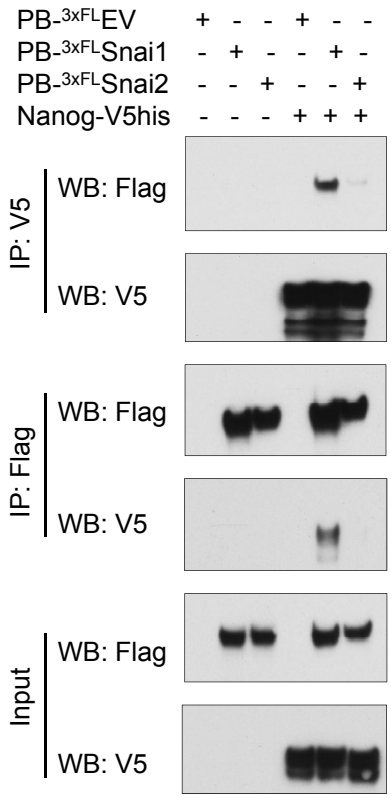

K
PB-3xFLEV + - - + - -
PB-3xFLSnai1 - + - - + -
PB-3xFLSnai2 - - + - - +
Nanog-V5his - - - + + +
IP: V5
WB: Flag
WB: V5
IP: Flag
WB: Flag
WB: V5
Input
WB: Flag
WB: V5

Figure S3

Figure S3. Gain-of-Function Studies of Snai1 and Snai2 in Nanog-Driven Pre-iPSC Reprogramming. Related to Figure 3.

(A) Combinations of empty vector (EV), Nanog, Snai1, and Snai2 were used to generate stable pre-iPSC clones. Western blots confirm transgenic Nanog, 3xFlag-Snai1 or 3xFlag-Snai2 protein expression in pre-iPSCs cultured in serum+LIF.

(B) Representative images of iPSCs generated from overexpression of Snai1 and Snai2 in the presence or absence or ectopic Nanog for both *Oct4*-GFP (left) and AP staining (right) at day 10 of 2i/LIF treatment.

(C) Cell cycle analyses by flow cytometry upon Snai1/2 ectopic expression in pre-iPSCs (rOKM) with or without ectopic Nanog expression.

(D) Fluorescence and bright-field images of transgene-free Nanog and Nanog+Snai1 iPSCs cultured under serum+LIF or 2i/LIF for 30 days.

(E) Immunofluorescence analysis against H3K27me3 in female transgene-free iPSCs under serum/LIF or 2i/LIF culture. Nuclei were stained with DAPI.

(F) Alkaline phosphatase (AP) and immunofluorescence staining for SSEA-1, Nanog, Oct4, and Sox2 protein levels in transgene-free Nanog and Nanog+Snai1 iPSCs.

(G) Similar expression of pluripotency and lineage specific marker genes in transgene-free Nanog and Nanog+Snai1 iPSCs. Error bars indicate average ± SD. Note that the similar Snai1 levels between Nanog+EV and Nanog+Snai1 samples indicate the complete removal of the ectopic Snai1 in Nanog+Snai1 iPSCs.

(H) Similar gene expression pattern of transgene-free Nanog and Nanog+Snai1 iPSCs during differentiation induced by LIF-withdrawal, retinoic acid (RA) treatment, or embryoid body (EB) formation. Gene expression analysis was measured by qRT-PCR, and data were normalized to *Gapdh*.

(I) Schematic overview of the reprogramming strategy in MEF-derived pre-iPSCs with Snai1/2 gain-of-function and Nanog ectopic expression.

(J) Quantification of *Nanog*-GFP+ colonies for noted treatments.

(K) CoIP validation of the Nanog-Snai1 interaction in 293T cells.

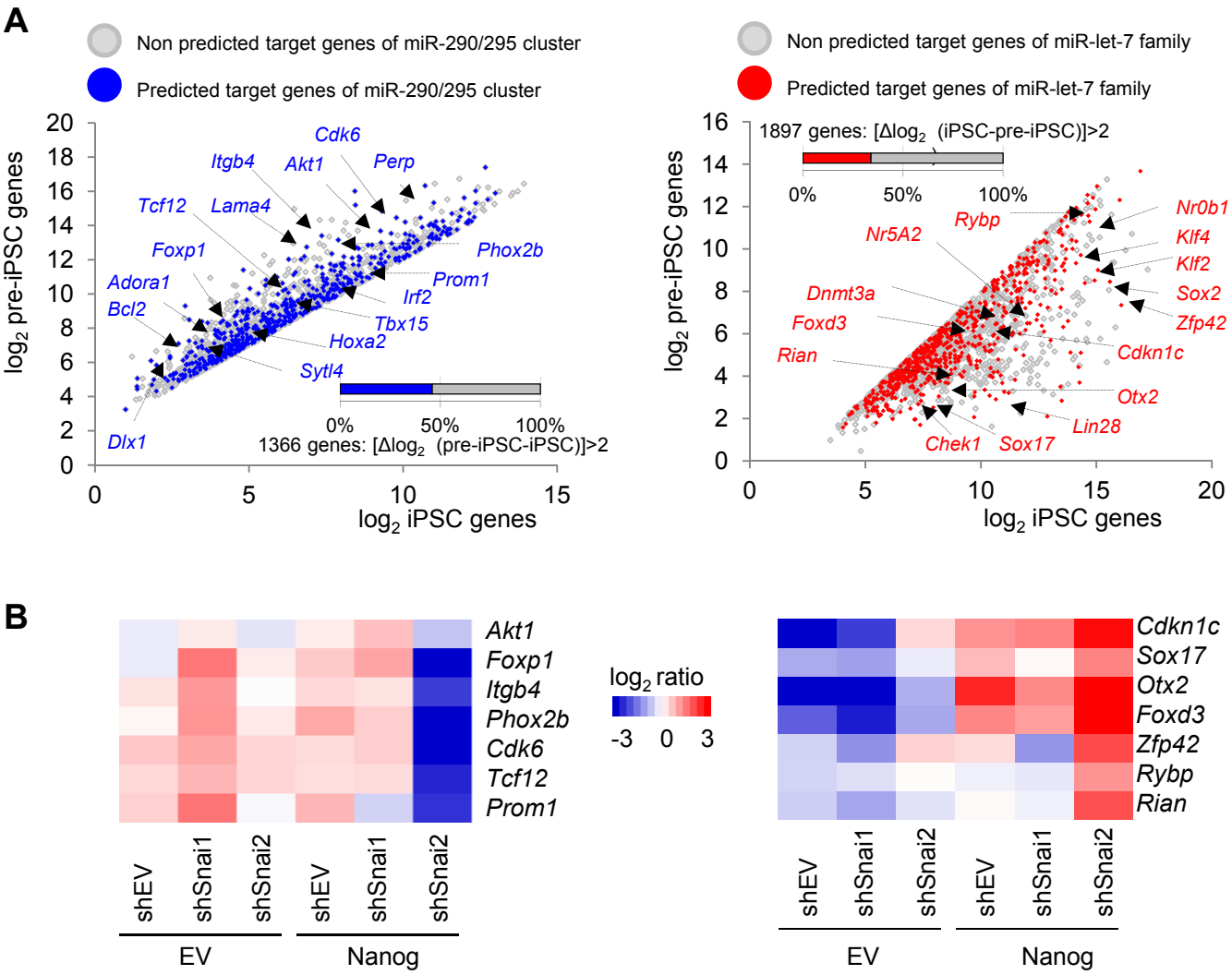

A
Non predicted target genes of miR-290/295 cluster
Predicted target genes of miR-290/295 cluster
log2 pre-iPSC genes
log2 iPSC genes
Cdk6
Itgb4
Akt1
Perp
Tcf12
Lama4
Foxp1
Phox2b
Adora1
Prom1
Bcl2
Irf2
Tbx15
Hoxa2
Sytl4
Dlx1
0%
50%
100%
1366 genes: [Δlog2 (pre-iPSC-iPSC)]>2
Non predicted target genes of miR-let-7 family
Predicted target genes of miR-let-7 family
1897 genes: [Δlog2 (iPSC-pre-iPSC)]>2
Nr0b1
Rybp
Nr5A2
Klf4
Klf2
Dnmt3a
Sox2
Foxd3
Zfp42
Rian
Cdkn1c
Otx2
Lin28
Chek1
Sox17
B
Akt1
Foxp1
Itgb4
Phox2b
Cdk6
Tcf12
Prom1
log2 ratio
-3 0 3
shEV
shSnai1
shSnai2
EV
Nanog
Cdkn1c
Sox17
Otx2
Foxd3
Zfp42
Rybp
Rian

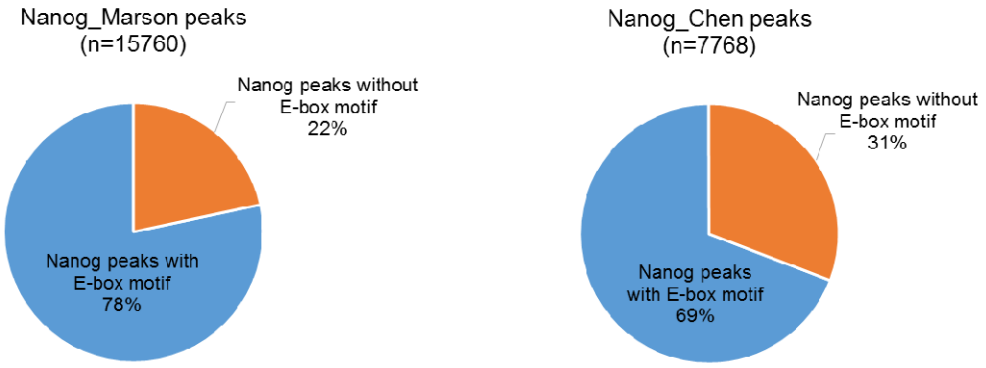
Figure S4

Figure S4. The Nanog and Snai1 Partnership in Transcriptional Regulation of Pluripotency-Associated Genes and miRNAs. Related to Figure 4.

(A) Relative expression levels of RNAi hit genes that are potential targets of miR-290-295 (left) and miR-Let-7 (right) in pre-iPSCs and iPSCs. Scatter plots of global gene expression arrays from published microarrays (Sridharan et al., 2009) showing highest ("pre-iPSC genes", left panel) or lowest ("iPSC genes", right panel) relative expression (log_2>2) in pre-iPSCs compared to iPSCs. Predicted target genes for the mouse miR-290-295 cluster are shown as blue dots, and those predicted for the mouse miR-Let7 family are shown as red dots. Blue and red bars indicate percentages of total pre-iPSC genes and iPS genes that are predicted targets of miR-290-295 and miR-Let7, respectively. Specific genes are indicated with arrows.

(B) Relative expression of the "pre-iPS genes" (left panel) or "iPSC genes" (right panel) measured by quantitative RT-PCR in the NS-derived reprogramming intermediates transduced with the indicated shRNA and cultured in serum/LIF.

A



Motif Sequence	Number E-box	Motif Frequency
CANN**TG**	**26362**	**100.0**
CAGC**TG**	2800	10.6
CAAA**TG**	2625	10.0
CATT**TG**	2595	9.8
CAGG**TG**	2195	8.3
CACC**TG**	2157	8.2
CAGA**TG**	1965	7.5
CATC**TG**	1925	7.3
CAAG**TG**	1553	5.9
CACT**TG**	1534	5.8
CAGT**TG**	1337	5.1
CAAT**TG**	1334	5.1
CAAC**TG**	1252	4.7
CACA**TG**	927	3.5
CATG**TG**	893	3.4
CATA**TG**	890	3.4
CACG**TG**	380	1.4

Motif Sequence	Number E-box	Motif Frequency
CANN**TG**	**10047**	**100.0**
CAGC**TG**	1068	10.6
CACC**TG**	999	9.9
CAGG**TG**	974	9.7
CATT**TG**	870	8.7
CAAA**TG**	859	8.5
CAGA**TG**	724	7.2
CATC**TG**	687	6.8
CAAT**TG**	588	5.9
CAAG**TG**	572	5.7
CACT**TG**	519	5.2
CAAC**TG**	512	5.1
CAGT**TG**	505	5.0
CACA**TG**	374	3.7
CATG**TG**	359	3.6
CATA**TG**	302	3.0
CACG**TG**	135	1.3

B

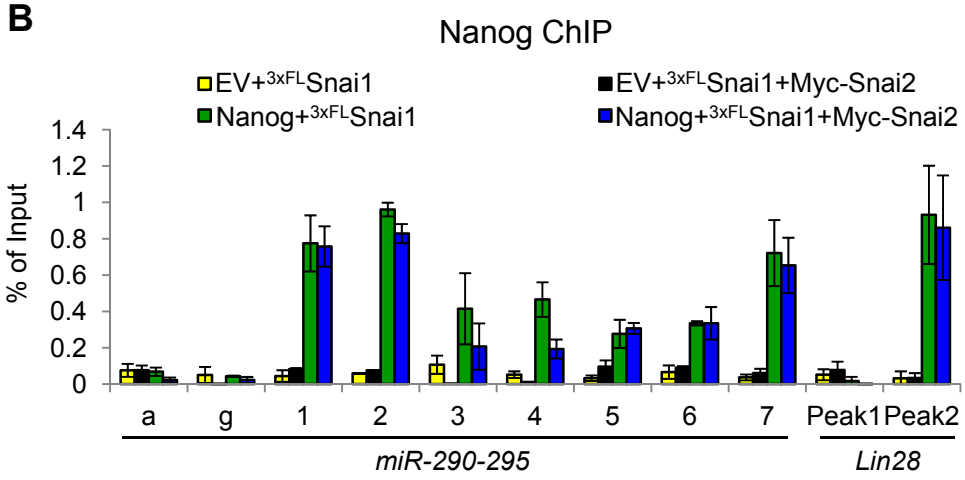


Figure S5

Figure S5. Snai2 Does Not Compete with Snai1 in Binding to the Nanog Sites at the *Lin28* and *miR-290-295* Loci. Related to Figure 5.

(A) Pie plots showing the percentage of Nanog peaks containing the consensus E-Box motif (CANNTG) in two published Nanog ChIP-Seq datasets. Number and frequency of individual E-Box motifs within the Nanog peaks are listed.

(B) Nanog ChIP and qPCR analyses on *the miR-290-295* and *Lin28* loci. Snai2 does not affect Nanog binding to the regulatory regions in both *miR-290-295* and *Lin28a* genes in pre-iPSCs ectopically expressing the indicated factors. Peaks "1-7" and “Peak2” are Nanog positive binding peaks on *miR290-295* and *Lin28a* regulatory regions, respectively. Peaks "a", "g" and “Peak1” are negative control regions. Error bars indicate average ± SD.

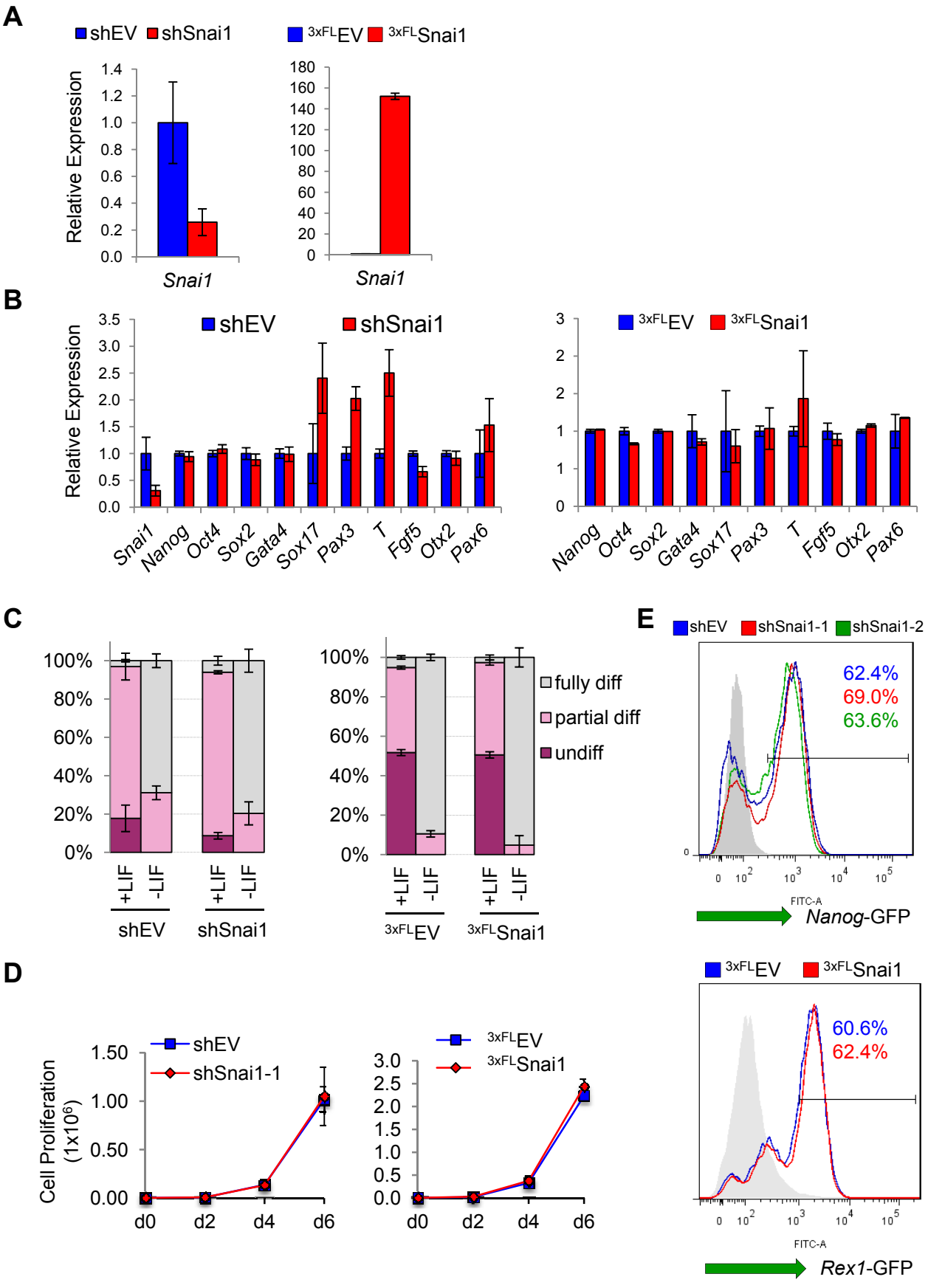

A
shEV
shSnai1
3xFLEV
3xFLSnai1
Relative Expression
Snai1
Snai1
B
shEV
shSnai1
3xFLEV
3xFLSnai1
Relative Expression
Snai1
Nanog
Oct4
Sox2
Gata4
Sox17
Pax3
T
Fgf5
Otx2
Pax6
C
fully diff
partial diff
undiff
+LIF
-LIF
shEV
shSnai1
3xFLEV
3xFLSnai1
D
shEV
shSnai1-1
3xFLEV
3xFLSnai1
Cell Proliferation (1x10^6)
d0
d2
d4
d6
E
shEV
shSnai1-1
shSnai1-2
62.4%
69.0%
63.6%
FITC-A
Nanog-GFP
3xFLEV
3xFLSnai1
60.6%
62.4%
FITC-A
Rex1-GFP

Figure S6

Figure S6. Snai1 is Dispensable for ESC Maintenance. Related to Figure 6.

(A) Snai1 knockdown (left panel) and transgenic (right) efficiency by shRNA and ectopic expression in CCE ESCs. Knockdown efficiency represents the average of two independent shRNAs against Snai1. Error bars indicate average ± SD.

(B) Pluripotency and lineage-specific marker analysis in ESCs transduced with two independent shRNAs against Snai1 or shEV (left panel) and in transgenic stable lines with ectopic expression of Snai1 or EV (right panel). Error bars indicate average ± SD.

(C) Colony formation assay (CFA) demonstrating lack of effect of Snai1 depletion (left panel) or enforced expression (right panel) on self-renewal. Colonies stained for AP (Alkaline Phosphatase) were scored into three categories (uniformly undifferentiated, partially differentiated, and fully differentiated). Error bars indicate average ± SD (n=3).

(D) ESC proliferation is independent of Snai1. Growth curve of CCE ESCs transduced with a shRNA against Snai1 or shEV (left panel) or with enforced Snai1 or EV expression (right panel). Cell numbers were counted every another day. Error bars indicate average ± SD (n=3).

(E) Nanog and Rex1 expression are not affected by Snai1 depletion (top panel) or enforced expression (bottom panel). Two independent ESC lines were used to measure the GFP activity under the *Nanog* (NG4 ES line) or *Rex1* (Rex1-dGFP ES line) promoter.

Table S1. RNAi Screen Studies in ESCs, Related to Figure 1.			
Screen Type	**Self-Renewal**	**Differentiation**	**Self-Renewal and Differentiation**
Published Studies	1. Ivanova et al., Nature 2006 2. Fazzio et al., Cell 2008 3. Ding et al., Cell Stem Cell 2009 4. Hu et al., Genes Dev. 2009 5. Kagey et al., Nature 2010 6. Chia et al., Nature 2010 7. Buckley et al., Cell Stem Cell 2012	1. Schaniel et al., Stem Cells 2009 2. Westerman, B.A. et al., JEM 2011 3. Yang et al., PLOS Genetics 2012 4. Buckley et al., Cell Stem Cell 2012 5. Betschinger et al., Cell 2013 6. Leeb et al., Cell Stem Cell 2014	This study
Aim	Identify factors regulating ESC self renewal and maintaining ESC identity	Identify factors required for exit of pluripotency and initiation of differentiation	Bidirectional screening for novel activators and repressors of ESC self-renewal gene expression program that can regulate early differentiation decisions and the establishment of pluripotency
Reporter Line	1. None 2. Oct4-GFP 3. Oct4-GFP 4. Oct4-GFP 5. Oct4-GFP 6. Oct4-GFP 7. Nanog-GFP	1. Nanog-GFP 2. None (AP staining) 3. Rex1-GFP and Oct4-GFP 4. Nanog-GFP 5. Oct4-GFP 6. Rex1-GFP	Nanog-GFP
Screen Conditions	1. Standard ES media (mouse) 2. Standard ES media (mouse) 3. Standard ES media (mouse) 4. Standard ES media (mouse) 5. Standard ES media (mouse) 6. Standard ES media (human) 7. Standard ES media (mouse)	1. Strong differentiation (2000 nM RA plus LIF withdrawal) 2. LIF withdrawal for 22 days 3. 2i withdrawal 4. Strong differentiation media (5000 nM RA plus LIF withdrawal) 5. 2i withdrawal and restoration 6. 2i withdrawal and restoration, haploid ESCs	Mild differentiation media (10 nM RA plus LIF withdrawal) where ESCs are poised at differentiation toward neural lineage

Table S1. RNAi Screen Comparison (Continued)			
Genome Scale?	1. No 2. No 3. Yes 4. Yes 5. No 6. Yes 7. No	1. No 2. Yes 3. Yes 4. No 5. No 6. No	Yes
Data Analysis Method	1. Self-renewal competition 2-7. Median GFP	1. GFP 2. AP staining 3-6. Median GFP	Median GFP
Main Findings	1. Esrrb/Tbx3/Tcl1 2. Tip60-p400 3. Paf1 complex 4. Cnot4/Trim28 5. Mediator complex 6. PRDM14 7. UPS-Psmd14	1. Smarcc1/Baf155 2. Mek binding protein 1 (MP1) 3. MAP kinase phosphatases 4. UPS-Fbxw7 5. Tfe3 6. Zfp706 and Pum1	•Opposing functions of mesenchymal transcription factors Snai1 and Snai2 in the Nanog-driven reprogramming •Snai1 and Snai2 differentially regulate Lin28 and miRNA-290-295 expression in the last stage of reprogramming •Snai1, but not Snai2, is activated by and interacts with Nanog in promoting transition of partially reprogrammed cells to full pluripotency
Candidate Hits Tested in Reprogramming ?	1. No 2. No 3. No 4. No 5. No 6. No 7. Yes	1. No 2. No 3. No 4. Yes 5. No 6. No	Yes with a pre-iPSC reprogramming system to specifically interrogate candidate function in the last stage of reprogramming

Supplementary Table S2, Related to Figure 1.
Summary of genome-wide RNAi candidate hits.
See Excel File.

Supplementary Table S3, Related to Figures 1-5.
List of qRT-PCR, ChIP-qPCR primer sequences as well as siRNA/shRNA sequences used in this study.
See Excel File.

Supplementary Table S4, Related to Figure S4.
List of genes differentially expressed in pre-iPSCs and iPSCs that are predicted to be targets of miR-290-295 cluster and Let-7 family miRNAs.
See Excel File.

Extended Experimental Procedures

RNAi Library and Plate Preparation

The mouse siGENOME library from Thermo Scientific, covering siRNA targets for 16,872 genes with pools of 4 sequences per gene target was used. Screen plate preparation and replication were done with Biomek FX Automated Workstation (Beckman Coulter). Library plates were hydrated to 2 μM, replicated and stored in Abgene 384-well polypropylene plates (AB-0781) from Thermo Scientific. As screened, an average of 750 ESCs per well were mixed with transfection reagent and siRNA pools in 384-well plates, cultured in LIF-containing media for 1 day and grown for 2 additional days without LIF and with 10 nM RA (Figure 1D). Confocal fluorescent microscopy on the fixed, Hoechst-33342 nucleus-stained cells provided images with cell-level resolution for each condition in the Thermo Scientific siGENOME library targeting 16,872 mouse genes along with assay-specific controls in technical triplicate. Under these conditions siRNA pools that decrease reporter fluorescence as well as those that increase it can both be identified, allowing for identification of both positive and negative regulators of pluripotency in one screen.

On the day of experiment, controls were added to the outer columns (1,2,23,24) of the library plates with 2 μM siRNA using the Janus Varispan Liquid Handling System (Perkin Elmer). The plates were diluted to 500 nM and added with 1:100 DharmaFECT 2 transfection reagent (TR) in a 3:1 ratio using the MultiDrop Combi liquid dispenser (Thermo Scientific) to bring siRNA to 125 nM and TR to 1:133 dilution. 10 μL of the siRNA/transfection mix were then replicated to the triplicate coated, gelatin-removed Aurora plates (Brooks Automation) and cells in mouse ES media added, bringing the siRNA to a final 25 nM concentration with 1:667 TR dilution.

Cell Culture

Cell culture was performed as described previously (Ivanova et al., 2006; Lee et al., 2012a; Lee et al., 2012b). NG4, CCE and *Rex1*-GFP mouse ESC lines were used as previously described (Schaniel et al., 2009; Toyooka et al., 2008). Cells were maintained feeder-free on gelatin-coated tissue culture dishes in ESC culture media (high-glucose DMEM supplemented with 15% fetal bovine serum (Hyclone, Cat No: SH30070.03), 100 mM MEM non-essential amino acids, 1 mM L-glutamine, 0.1 mM β-mercaptoethanol, 100 units/mL penicillin/streptomycin, 1 mM sodium pyruvate and 1000 U/mL LIF (Millipore).

iPSC Differentiation Assays

Prior to differentiation assays of iPS clones (Figure S3H), Piggybac (PB) transgenes were excised by transient co-transfection with PBase. For LIF withdrawal, 83 x10^3 cells were plated on a gelatin-coated 6-well plates with high-glucose DMEM supplemented

with 15% fetal bovine serum. For retinoic acid (RA) induced differentiation, 83 x10^3 cells were treated with 0.2 μM of RA. For embryoid body (EB) formation, 4.5 x 10^4 cells were plated in 6-well ultra-low attachment plates (Corning) in the same medium used for LIF withdrawal conditions.

Colony Formation Assay

For colony formation assays a single cell suspension of ESCs was seeded at a density of 800 cells per 6-well plate in ES media in the presence or absence of LIF (1000 U/ml). After 6 days, cells were stained for alkaline phosphatase activity using a commercial kit (SIGMA).

Microscopy

Images were acquired using an ImageXpress Ultra high-content confocal microscope (Molecular Devices). For each well, 4 non-overlapping images were collected with 2x binning with 20x objective as 1000x1000 pixel 16-bit files. Each image corresponded to an 800 μm x 800 μm area.

Immunofluorescence

Cells were cultured and fixed with formaldehyde as described above. Cells were washed once with PBS and permeabilized at RT for 1 hour with 1% Triton-X-100, 1% BSA and 5% donkey serum in PBS. Cells were then incubated overnight at 4C with 1:150 RCAB0002P-F rabbit anti-mouse Nanog (CosmoBio) in 1% BSA and 5% donkey serum in PBS and washed 3 times for 5 minutes with PBS. Cells were incubated with the secondary antibodies (1:1000 Alexa Fluor® 488 donkey anti-rabbit IgG (Invitrogen)) in 1% BSA and 5% donkey serum in PBS. Cells were then stained with Hoechst 33342 5 μg/mL for 15 minutes at RT and washed with PBS 3 times prior to confocal fluorescence imaging.

For immunofluorescence analysis of iPS clones (Figure S3), cells were cultured on glass slides and fixed in 4% paraformaldehyde (PFA), followed by permeabilization with 0.25% Triton X-100. The cells were stained with antibodies against Nanog (Abcam, cat#ab70482), Oct4 (Santa Cruz Biotechnology, cat#sc-5279), Sox2 (Santa Cruz Biotechnology, cat#sc-17320), SSEA1 (R&D, cat#MAB2155), and H3K27me3 (Millipore, cat#07-449).

Alkaline Phosphatase (AP) Staining

AP staining was measured using an Alkaline Phosphatase Staining Kit (Stemgent) following the manufacturer's recommendations.

Cell-level Data Extraction

Confocal images were taken in blue (Hoechst 33342 nuclear staining) and green (cytoplasmic GFP) channels. Cell segmentation is used to extract cell-level parameters (i.e. fluorescence). There are several solutions (Berlage, 2005; Carpenter et al., 2006; Gonzalez et al., 2009; Wang et al., 2010), most of which rely on nuclear staining to first distinguish individual cells and then use local variations in cytoplasmic staining (potentially from the reporter) to distinguish the cell margins. Images were processed for cell segmentation analysis with the MetaXpress software package (Molecular Devices, http://www.moleculardevices.com/products/software/high-content-analysis/metaxpress.html), with parameters tuned to describe the range of cell and nuclear diameters. Nuclei were segmented using the blue channel with a minimum width of 8 μm (10 pixels) and a maximum width of 26 μm (33 pixels). Cytoplasmic regions were segmented using the GFP channel with a minimum width of 10 μm (13 pixels) and a maximum width of 30 μm (38 pixels). For each image, we extracted average fluorescence per cell over the entire cytoplasmic area (that is sum of GFP intensities over all pixels representing the cytoplasmic area divided by number of pixels). Values were exported as text files, which were pre-processed with Perl to extract well-specific values into separate files. Cell-level data were then imported into R (Development Core Team, 2005).

Data Acquisition Set Normalization

Reporter fluorescence values were transformed to a logarithmic scale.

Because large-scale screening is necessarily staggered across multiple data acquisition sets (for example, into 384-well screening plates, each of which has its own set of controls), we normalized each dataset (plate) to a reference plate dataset (Figures S2A and S2B). Cell-level parameter values, including fluorescence, are expected to fall into a particular range, with values at the extremes typically captured by the controls, making such a transformation appropriate.

Because cell populations across datasets can be considered comparable (typically, because most screened conditions have no effect), we perform a quantile-based non-parametric normalization to reduce non-linear cross-dataset effects. The cumulative distribution function of the parameter (*i.e.* log fluorescence) was calculated over all cells in the plate to be normalized and for the reference plate. The fluorescence of each cell in the dataset is normalized to the fluorescence level of the closest ranked cell in the reference dataset. For example, if there are 1.2 million cells in the 384-well reference plate and 1 million cells in the plate to be normalized, the fluorescence of the cell with the 750,000th highest fluorescence will be reassigned to the fluorescence of the 1,200,000 * 750,000/1,000,000 = 900,000th highest ranked cell in the reference plate.

Excluded Wells

To avoid constructing unreliable estimates of the fluorescence distribution, we disregard conditions where fewer than 100 cells can be identified following image processing. In practice, this is not a significant challenge because a 384-well screening plate can accommodate several thousand cells per well. Over 95% of all conditions in our screen met these requirements. Those conditions that did not were mostly cell death-specific controls such as si*Wee1* (data not shown).

Outlier Categorization

The above procedure assigned each cell in each plate a particular normalized log-transformed fluorescence value. For all cells in each non-excluded well, we calculated the median log-transformed normalized cell value. These median fluorescence values thus represent the effect of each particular well on the population of cells. The median fluorescences are rescaled as Z-scores, with a cutoff for significance at |Z-score| >2. All condition effects with |Z-score| ≤ 2 are considered to be non-significant. Candidate hit genes are those whose median over technical replicates has a |Z-score| >2. Because all siRNA conditions were tested in technical triplicate, we thus identify all genes whose depletion is significant and concordant in our screen in at least 2 of 3 technical replicates.

Secondary Screen Analysis

For each candidate gene, a deconvoluted pool of 4 independent siRNAs was screened in technical triplicate. For reasons of visual clarity, we display the median over 3 technical replicates rather than the individual values or the error bars for each of those 4 siRNAs. Each of our triplicate plates had over 200 empty wells that were also read in the GFP and DAPI channels and were used as negative controls.

We assessed significance using a 2-tailed *t*-test with 2-sample equal variance comparing each siRNA sequence in its 3 technical replicates to all technical replicates of the empty control. We applied the Benjamini-Hochberg FDR correction to p-values and set a significance threshold of $p<0.05$. Based on this analysis, 13 of 24 genes had repeatable, significantly distinct effects from the empty control in ≥2 of the 4 deconvoluted siRNAs and 8 of 28 genes were significantly distinct in ≥3 of the 4 repeats.

The secondary screen was designed to demonstrate that an effect from a particular siRNA was not due to an off-target effect from an individual siRNA in the pool. By this analysis, the probability of 2 independent siRNAs from the pool having no true effect but producing these statistically significant effects is $<0.05^2=0.0025$. Since there are 6 ways that any 2 of 4 siRNAs can be chosen, hence we expect to have fewer than $6*0.0025*24=0.36$, false positives in our set of 24 genes.

shRNA Design, Lentivirus Generation, and Mouse ESC Transduction

shRNAs utilized in this study are listed in Table S3. Oligonucleotides encoding each shRNA duplex were synthesized by Integrated DNA Technologies and cloned into the AgeI/EcoRI sites of the lentiviral-based shRNA expression vector pLKO.pim (pLKO.1 PuroR-IRES-mCherry) following the supplier's protocol (http://www.addgene.org/plko) (Moffat et al., 2006). All shRNA constructs were confirmed by sequencing. Lentiviruses were generated in HEK-293T cells by Superfect-mediated cotransfection of lentiviral-based shRNA plasmids and the pCMV-dR8.2 (packaging) and pCMV-VSVG (envelope) plasmids. Viral supernatants were concentrated using Amicon Ultra centrifugal filter units (Millipore) at 1600g for 20 min. and stored at -80 °C. For infection, mESCs were infected in media supplemented with polybrene (8 μg/ml; Sigma). Cells were incubated overnight with virus and subsequently cultured in fresh media for 4 days. Infected cells were cultured in media supplemented with 2 μg/mL puromycin for an additional 4 days, after which mRNA was extracted.

Quantitative RT-PCR Analysis

RNA was extracted using Trizol and the RNeasy Mini Kit (Qiagen). 1 μg of total RNA was converted into double-stranded cDNA using the High Capacity reverse transcription kit (Applied Biosystems). Quantitative PCR was performed using the Fast SYBR® Green Master Mix (Applied Biosystems) on the LightCycler480 Real-Time PCR System (Roche). Gene-specific primers used for this study are provided in Table S3.

Reprogramming Assays from Neural Stem (NS) Cell Reprogrammed Intermediates

Pre-iPS cell reprogramming assays were performed as described previously (Costa et al., 2013, Silva et al., 2008). Briefly, clonal lines of reprogrammed cell intermediates (*i.e.*, pre-iPSCs) (rOKM in Figure S2A) were established from neural stem cells containing an *Oct4*-GFP reporter transgene and infected with pMX retroviruses expressing the reprogramming factors Oct4, Klf4 and c-Myc (rOKM) and maintained in serum plus LIF (serum/LIF) medium. Reprogramming was initiated by a switch to 2i plus LIF (2i/LIF) medium.

To investigate the consequences of depletion of Snai1 and Snai2 in iPSC generation, a clonal line of proliferative *Oct4*-GFP negative cells (reprogramming intermediates or pre-iPSCs) sorted in ES media (serum/LIF) was transfected using Lipofectamine 2000 transfection reagent (Invitrogen) with 1 μg of PB-flox-3xFlag-Nanog-Pgk-Hygromycin, or PB-flox-3xFlag-Empty-Pgk-Hygromycin plus 2 μg of the PBase expression vector pCAGPBase, and drug was applied for a minimum of 10 days. Stable *Oct4*-GFP negative reprogramming intermediates with trangenic Nanog or empty vector (EV) were infected with lentiviruses expressing puromycin-IRES-mCherry (pLKO.pim) and shRNAs against Snai1, Snai2, luciferase or EV. Dual hygromycin and puromycin selection was applied to these reprogramming intermediates for minimum of 6 days, and almost 100%

of surviving cells were mCherry positive. After selection, $1x10^4$ pre-iPSCs were seeded in gelatin-coated 12-well plates in serum/LIF. Medium was switched to N2B27/2i/LIF the following day and GFP-positive colonies or AP-positive colonies were score at day 10.
A similar strategy was used to analyze the consequences of ectopic expression of Snai1 and Snai2 alone or in combination with ectopic Nanog in iPSC generation. Reprogramming intermediates were transfected with various combinations of 1 µg of PB-flox-Nanog-IRES-Blast, PB-flox-Empty-IRES-Blast, PB-flox-3xFlag-Snai1-Pgk-Hygro, PB-flox-3xFlag-Snai2-Pgk-Hygro, or PB-flox-3xFlag-Empty-Pgk-Hygro plus 2 µg of the PBase expression vector, pCAG-PBase. Dual hygromycin and blasticidin selection was applied to transfectants for a minimum of 12 days and maintained until medium was switched to 2i/LIF.

To analyze the consequences of Snai1 and Snai2 competition on E-box motifs, reprogramming intermediates overexpressing ^{3xFLAG}Snai1, alone or in combination with Nanog, were transfected with 10 µg of pcDNA3-Snai2-Myc. G418 selection was applied for a minimum of 10 days.

Reprogramming Assays in Mouse Embryonic Fibroblast (MEF) Derived Reprogrammed Intermediates

Clonal lines of reprogrammed cell intermediates (*i.e.*, pre-iPSCs, *Nanog*-GFP-negative cells) (rOKMS in Figure S2G, and Figure S3I) were established from MEFs containing a *Nanog*-GFP reporter transgene and infected with pMX retroviruses expressing the reprogramming factors Oct4, Klf4, c-Myc and Sox2 (rOKMS) and maintained in serum plus LIF (serum+LIF) medium. Reprogramming was initiated by a switch to 2i plus LIF (2i+LIF) medium. Reprogrammed intermediates were transfected and selected as described above for NSC reprogrammed intermediates. GFP-positive colonies were scored at day 10 after 2i+LIF medium switch.

Cell Cycle Analysis

For cell cycle analysis, an equal number of proliferating reprogramming intermediates, $5x10^5$ cells, cultured in serum+LIF under specific drug selection for depletion or ectopic expression as indicated in Supplemental Figures, were washed with DPBS (Dulbecco's phosphate buffered saline), permeabilized with 0.1% Triton X-100 in DPBS, stained with 10 µM dye4'-6-diamidino-2-phenylindole (DAPI) at room temperature for 10 min, and analyzed by flow cytometry. Analysis was performed in FlowJo software using the Dean-Jett-Fox cell cycle model.

Chromatin Immunoprecipitation Coupled with Quantitative Real-Time PCR (ChIP-qPCR)

ChIP assays were performed as described (Lee et al., 2006). Briefly, cells were cross-linked with 1% (w/v) formaldehyde for 10 min at room temperature, and formaldehyde

was inactivated by the addition of 125 mM glycine. Chromatin extracts containing DNA fragments were immunoprecipitated using anti-Nanog (Bethyl Labotatories, cat #A300-397A) or anti-Flag (Sigma, cat #F1804) antibodies. Primer sequences were designed according to overlapping peaks of Nanog and Med1 in *Snai1*, *miR-290-295*, and *Lin28* loci and are provided in Table S3. ChIP-Seq datasets for Nanog, Med1, H3K4me1, and H3K27ac were download from GEO: Nanog (Accession number: GSE11724) (Marson et al., 2008), Med1 (Accession number: GSE22562) (Kagey et al., 2010), H3K4me1 and H3K27ac (Accession number:GSE27841) (Whyte et al., 2012). Reads were uniquely (-m 1) aligned to the mouse (mm9) genome by Bowtie software (version 1.0.0) and Chip-Seq peaks were determined by MACS (version 1.4.2) using the default settings. Aligned reads were sorted and converted to a binary tiled file (tdf), and visualized using IGV software (http://www.broadinstitute.org/software/igv/home) (Robinson et al., 2011; Thorvaldsdottir et al., 2013). The immunoprecipitated DNA was analyzed by real-time PCR with a LightCycler 480 (Roche) instrument with LightCycler DNA master SYBR Green I reagents. Differences between samples and controls were calculated based on the $2^{-\Delta\Delta CT}$ method and normalized by percentage of input DNA recovery. Measurements were performed in triplicate.

Flow Cytometry Analysis

For flow cytometry analysis, single-cell suspensions were evaluated on an LSRII Flow Cytometer System (BD Biosciences). The fluorescence activities of GFP reporters (*Oct4*-GFP, *Nanog*-GFP, and *Rex1*-GFP) and DAPI were detected in the FITC channel and Indo-Violet-A channel respectively. Data were analyzed with FlowJo software.

Overlap of RA and EB Differentiation Transcriptome Data with RNAi Screen Candidates

Microarray datasets for all probes in either LIF withdrawal differentiation (Hailesellasse Sene et al., 2007) or 2 μM RA differentiation (Ivanova et al., 2006) were extracted from their respective supplementary tables. Probes were mapped to gene aliases using the ChipDB package in Bioconductor 2.14.0 and aliases mapped to updated gene names using Genome wide annotation for Mouse (org.Mm.eg.db) in Bioconductor. Analysis was restricted to the set of 10,011 genes assayed in both differentiation datasets as well as in the genome-wide RNAi screen (Figure S1E, center region shaded with lines on left and entire boxed region on right). The set of genes differentially regulated (>2-fold either up or down) relative to the day 0 time-point following 2 days of respective differentiation conditions or noted as a candidate in the RNAi screen was computed. Intersection between gene sets (Figure S1E, right) and contingency tables (Figure S1F) was calculated using R.

GO Analysis

Gene ontology molecular function analysis was performed using the DAVID gene ontology functional annotation tool (http://david.abcc.ncifcrf.gov/tools.jsp) (Huang da et al., 2009a, b) with all NCBI *Mus musculus* genes as a reference list.

Supplementary References

Berlage, T. (2005). Analyzing and mining image databases. Drug discovery today *10*, 795-802.

Carpenter, A.E., Jones, T.R., Lamprecht, M.R., Clarke, C., Kang, I.H., Friman, O., Guertin, D.A., Chang, J.H., Lindquist, R.A., Moffat, J.*, et al.* (2006). CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biol *7*, R100.

Development Core Team (2005). R: A language and environment for statistical computing (R Foundation for Statistical Computing, Vienna, Austria).

Gonzalez, R.C., Woods, R.E., and Eddins, S.L. (2009). Digital Image Processing Using MATLAB. In (Gatesmark Publishing), p. 827.

Hailesellasse Sene, K., Porter, C.J., Palidwor, G., Perez-Iratxeta, C., Muro, E.M., Campbell, P.A., Rudnicki, M.A., and Andrade-Navarro, M.A. (2007). Gene function in early mouse embryonic stem cell differentiation. BMC genomics *8*, 85.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res *37*, 1-13.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc *4*, 44-57.

Ivanova, N., Dobrin, R., Lu, R., Kotenko, I., Levorse, J., DeCoste, C., Schafer, X., Lun, Y., and Lemischka, I.R. (2006). Dissecting self-renewal in stem cells with RNA interference. Nature *442*, 533-538.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S.*, et al.* (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430-435.

Lee, D.F., Su, J., Ang, Y.S., Carvajal-Vergara, X., Mulero-Navarro, S., Pereira, C.F., Gingold, J., Wang, H.L., Zhao, R., Sevilla, A.*, et al.* (2012a). Regulation of embryonic and induced pluripotency by aurora kinase-p53 signaling. Cell Stem Cell *11*, 179-194.

Lee, D.F., Su, J., Sevilla, A., Gingold, J., Schaniel, C., and Lemischka, I.R. (2012b). Combining competition assays with genetic complementation strategies to dissect mouse embryonic stem cell self-renewal and pluripotency. Nat Protoc *7*, 729-748.

Lee, T.I., Johnstone, S.E., and Young, R.A. (2006). Chromatin immunoprecipitation and microarray-based analysis of protein location. Nat Protoc *1*, 729-748.

Marson, A., Levine, S.S., Cole, M.F., Frampton, G.M., Brambrink, T., Johnstone, S., Guenther, M.G., Johnston, W.K., Wernig, M., Newman, J.*, et al.* (2008). Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell *134*, 521-533.

Moffat, J., Grueneberg, D.A., Yang, X., Kim, S.Y., Kloepfer, A.M., Hinkle, G., Piqani, B., Eisenhaure, T.M., Luo, B., Grenier, J.K.*, et al.* (2006). A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. Cell *124*, 1283-1298.

Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat Biotechnol *29*, 24-26.

Schaniel, C., Ang, Y.S., Ratnakumar, K., Cormier, C., James, T., Bernstein, E., Lemischka, I.R., and Paddison, P.J. (2009). Smarcc1/Baf155 couples self-renewal gene repression with changes in chromatin structure in mouse embryonic stem cells. Stem Cells *27*, 2979-2991.

Sridharan, R., Tchieu, J., Mason, M.J., Yachechko, R., Kuoy, E., Horvath, S., Zhou, Q., and Plath, K. (2009). Role of the murine reprogramming factors in the induction of pluripotency. Cell *136*, 364-377.

Thorvaldsdottir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Briefings in bioinformatics *14*, 178-192.

Toyooka, Y., Shimosato, D., Murakami, K., Takahashi, K., and Niwa, H. (2008). Identification and characterization of subpopulations in undifferentiated ES cell culture. Development *135*, 909-918.

Wang, Q., Niemi, J., Tan, C.M., You, L., and West, M. (2010). Image segmentation and dynamic lineage analysis in single-cell fluorescence microscopy. Cytometry Part A : the journal of the International Society for Analytical Cytology *77*, 101-110.

Whyte, W., Bilodeau, S., Orlando, D., Hoke, H., Frampton, G., Foster, C., Cowley, S., and Young, R. (2012). Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. Nature *482*, 221-225.