

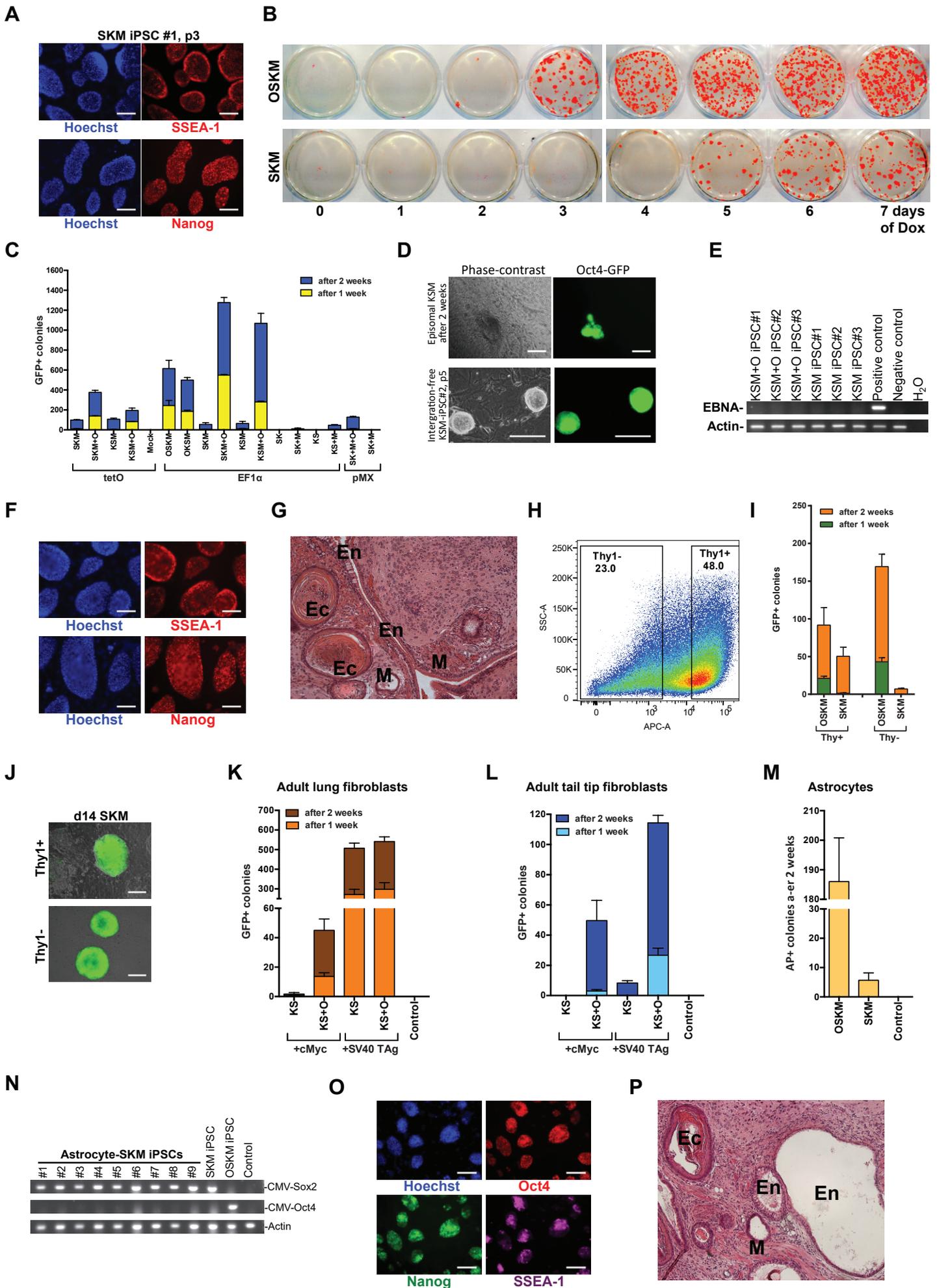
**Cell Stem Cell, Volume 25**

**Supplemental Information**

**Excluding Oct4 from Yamanaka Cocktail Unleashes  
the Developmental Potential of iPSCs**

**Sergiy Velychko, Kenjiro Adachi, Kee-Pyo Kim, Yanlin Hou, Caitlin M.  
MacCarthy, Guangming Wu, and Hans R. Schöler**

Figure S1

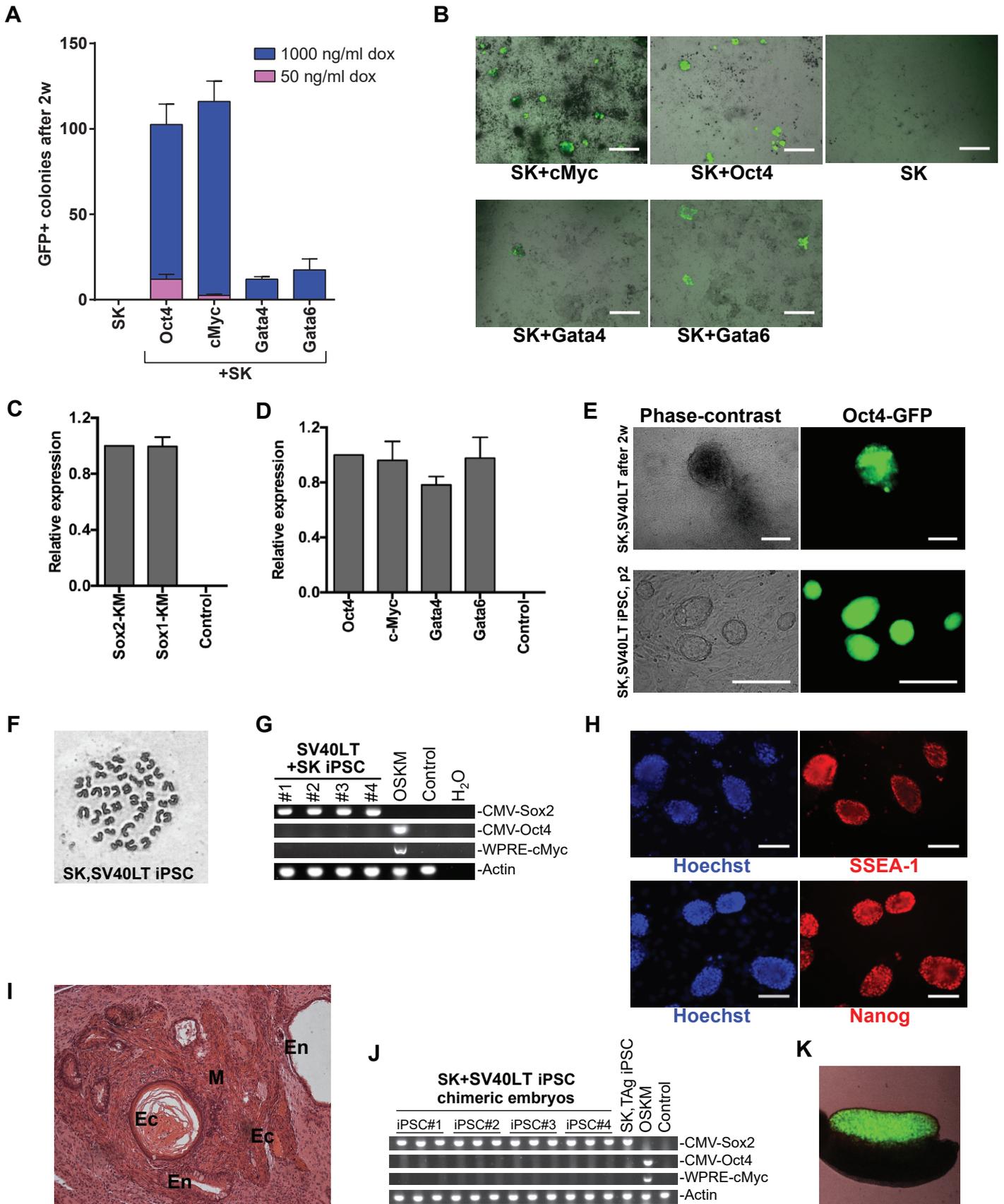


**Figure S1. Related to Figure 1**

**Sox2, Klf4, and cMyc can reprogram to pluripotency without exogenous POU factor expression**

- (A) Nanog and SSEA1 immunofluorescence staining of SKM iPSC line at p8. Nuclei were stained with Hoechst 33342.
- (B) Alkaline phosphatase (AP) staining of iPSC colonies generated from GOF18-Rosa26-rtTA MEFs by time-restricted expression of SKM or OSKM. The reprogramming experiment started with  $10^3$  transduced MEFs plated in each well. The colonies were imaged 10 days after induction.
- (C) Reprogramming of Oct4-GFP MEFs by SKM expressed under either tet-inducible or constitutive EF1alpha promoter.
- (D) Phase-contrast and Oct4-GFP images of primary colonies and established clonal lines of integration-free KSM iPSCs generated with episomal vector (scale=250  $\mu$ m).
- (E) PCR genotyping confirming the absence of genomic integration or episomal expression of pCXLE reprogramming vector in iPSC lines at p5.
- (F) Nanog and SSEA1 immunofluorescence staining of integration-free KSM iPSC line, p8. Nuclei were stained with Hoechst 33342.
- (G) Hematoxylin and eosin staining of teratoma sections with representation of all three germ layers (ectoderm: keratinizing epithelium; mesoderm: smooth muscles, connective tissue; endoderm: cuboidal and gut epithelium).
- (H) FACS to sorting of Thy- and Thy+ MEF sub-populations.
- (I) Reprogramming of Thy- and Thy+ Oct4-GFP MEFs (C) by OSKM or SKM. The colonies were counted 1 and 2 weeks after infection.
- (J) Representative phase-contrast and Oct4-GFP images of primary colonies from (D) (scale=250  $\mu$ m).
- (K), (L) Reprogramming of Oct4-GFP adult lung fibroblasts (K) and adult tail tip fibroblasts (L) by KS in combination with cMyc, Oct4, or SV40 large T Antigen (SV40LT).
- (M) Reprogramming of mouse cortical astrocytes with OSKM vs SKM. The cells were fixed and stained for AP activity after 2 weeks of infection.
- (N) PCR genotyping analysis of 9 clonal astrocyte-derived iPSCs confirming the integration of SKM transgene and the absence of Oct4.
- (O) Oct4, Nanog, and SSEA1 immunofluorescence staining of astrocyte-derived iPSC line #4, p4. Nuclei were stained with Hoechst 33342 (blue).
- (P) Hematoxylin and eosin staining of teratoma sections generated with astrocyte-derived SKM-iPSCs with representation of all three germ layers (ectoderm: keratinizing epithelium; mesoderm: smooth muscles; endoderm: cuboidal and gut epithelium).

Figure S2



**Figure S2. Related to Figure 2**

**Sox2 and Klf4 are sufficient for reprogramming of highly-proliferative cells to iPSCs**

- (A) Reprogramming of Oct4-GFP MEFs by bicistronic SK in combination with Oct4, cMyc, Gata4, or Gata6. Different level of transgene expression was achieved by induction with either 1 µg/ml, 50 ng/ml of dox for 12 days. GFP<sup>+</sup> colonies were counted after 14 dpi.
- (B) Brightfield and Oct4-GFP–merged representative overview images of reprogrammed colonies from Figure S2A (scale=1 mm).
- (C), (D) qPCR titration of polycistronic lentiviral vectors using universal lentivirus 3'LTR (WPRE region) primers. *Rpl37a* and *Hprt1* were used as reference genes, the error bars represent SD between the two.
- (E) Phase-contrast and Oct4-GFP images of primary colonies and passaged SK iPSCs generated from immortalized Oct4-GFP MEF (scale=250 µm).
- (F) Giemsa-stained metaphase spreads of clonal SV40LT, SK iPSC#2 line.
- (G) PCR genotyping verifying tet-miniCMV-SK transgene in four SV40LT, SK-generated iPSC lines while confirming the absence of Oct4 or cMyc integration.
- (H) Nanog and SSEA1 immunofluorescence staining of SV40LT, SK iPSC#2 line, p7. Nuclei were stained with Hoechst 33342.
- (I) Hematoxylin and eosin staining of teratoma sections with representation of all three germ layers (ectoderm: keratinizing epithelium, thyroid; mesoderm: smooth muscles; endoderm: cuboidal and gut epithelium).
- (J) PCR genotyping verifying tetO-SK transgene in all the tested chimeric embryos derived from 4 lines of SV40LT, SK iPSC and absence of Oct4 or cMyc integration.
- (K) Brightfield and GFP merged images of embryonic day a 13.5 gonad dissected from SV40LT, SK iPSC#4 chimeric embryo.

Figure S3

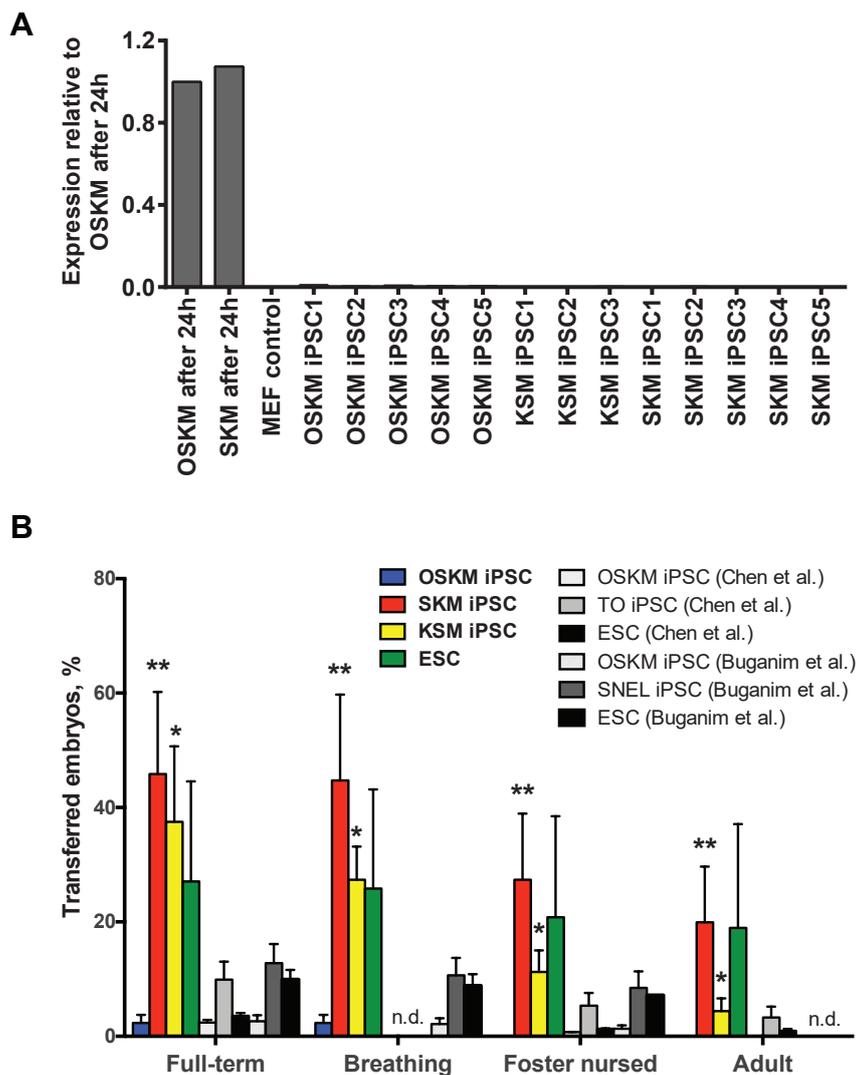


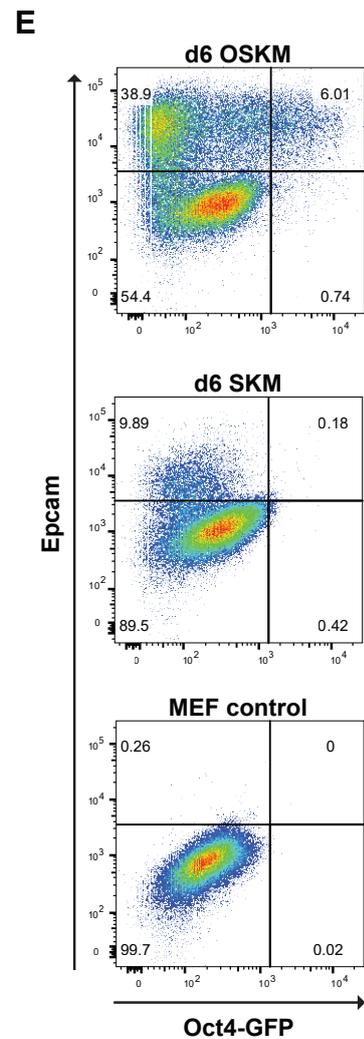
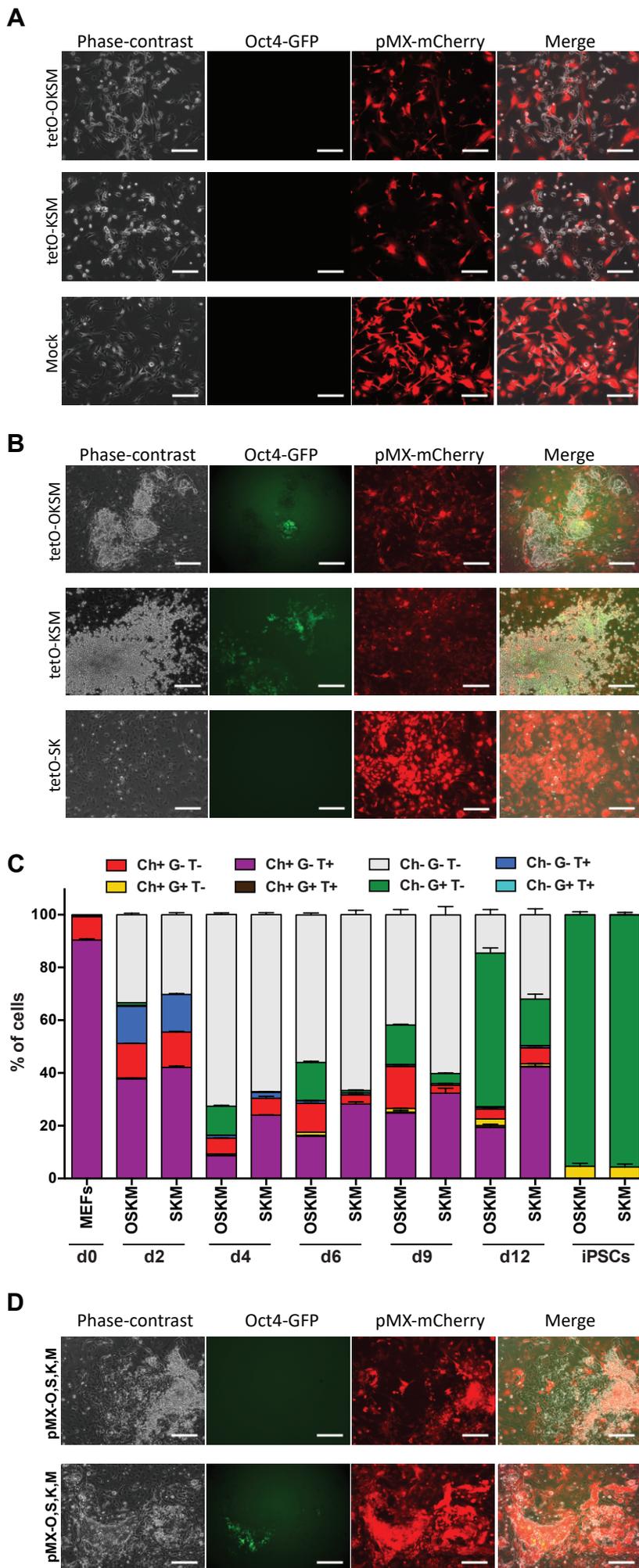
Figure S3. Related to Figure 3

**SKM iPSCs hold dramatically enhanced developmental potential**

(A) qPCR analysis of lentiviral transgene expression of OSKM and SKM transgenes after 24h of induction in MEFs and in the clonal iPSC lines. *Rpl37a* was used as a housekeeping gene.

(B) Percentage of 4N-aggregated embryos that gave rise to full-term pups, pups that initiated breathing, pups that survived foster-nursing (at least 48 hours), and pups that survived until adulthood (at least 3 months). The grey and black bars depict previously published data for comparison. Data are represented as the mean of all tested lines. The statistical significance was determined by Mann-Whitney test. Error bars represent SEM.

Figure S4



**Figure S4. Related to Figure 4**  
**Reprogramming factor induction causes immediate retrovirus silencing**

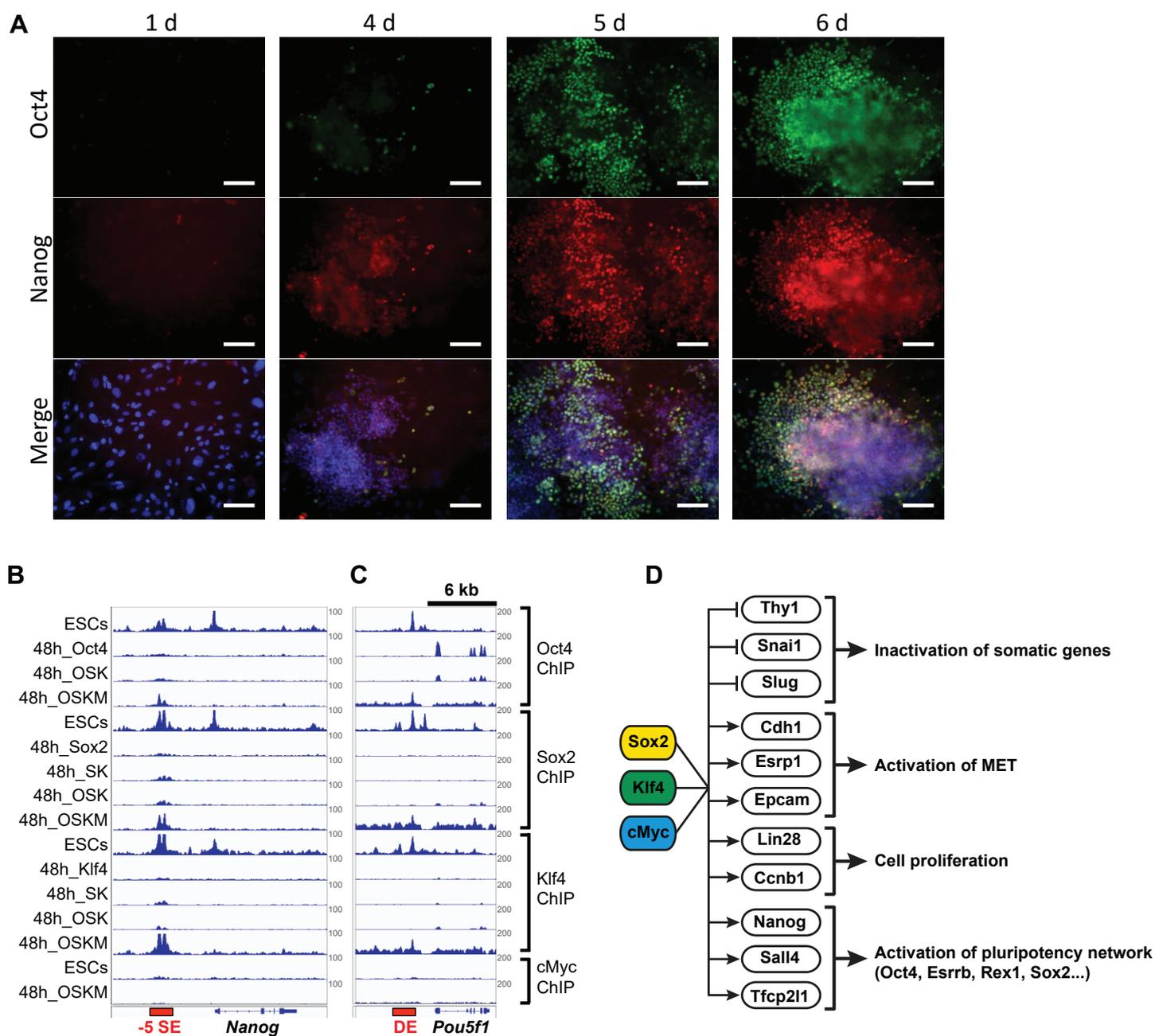
(A,B) Representative phase-contrast Oct4-GFP and pMX-mCherry images of presorted mCherry<sup>+</sup> MEFs infected with tet-inducible KSM or OKSM vectors after 3 days of induction (A), and 5 or 6 days of induction for OKSM and KSM, respectively (B).

(C) Quantification of time-course FACS data of pMX-mCherry (Ch), Oct4-GFP (G) and Thy1 (T) expression of MEFs reprogrammed with OSKM or SKM after 2, 4, 6, 9, and 12 days of induction.

(D) Representative phase-contrast images pMX-mCherry<sup>+</sup> Oct4-GFP MEFs infected with pMX-Oct4, pMX-Sox2, pMX-Klf4, and pMX-cMyc vectors after 5 days of infection.

(E) FACS analysis of Oct4-GFP and Epcam expression in MEFs reprogrammed with OSKM or SKM on 6 dpi.

Figure S5



**Figure S5. Related to Figure 5**

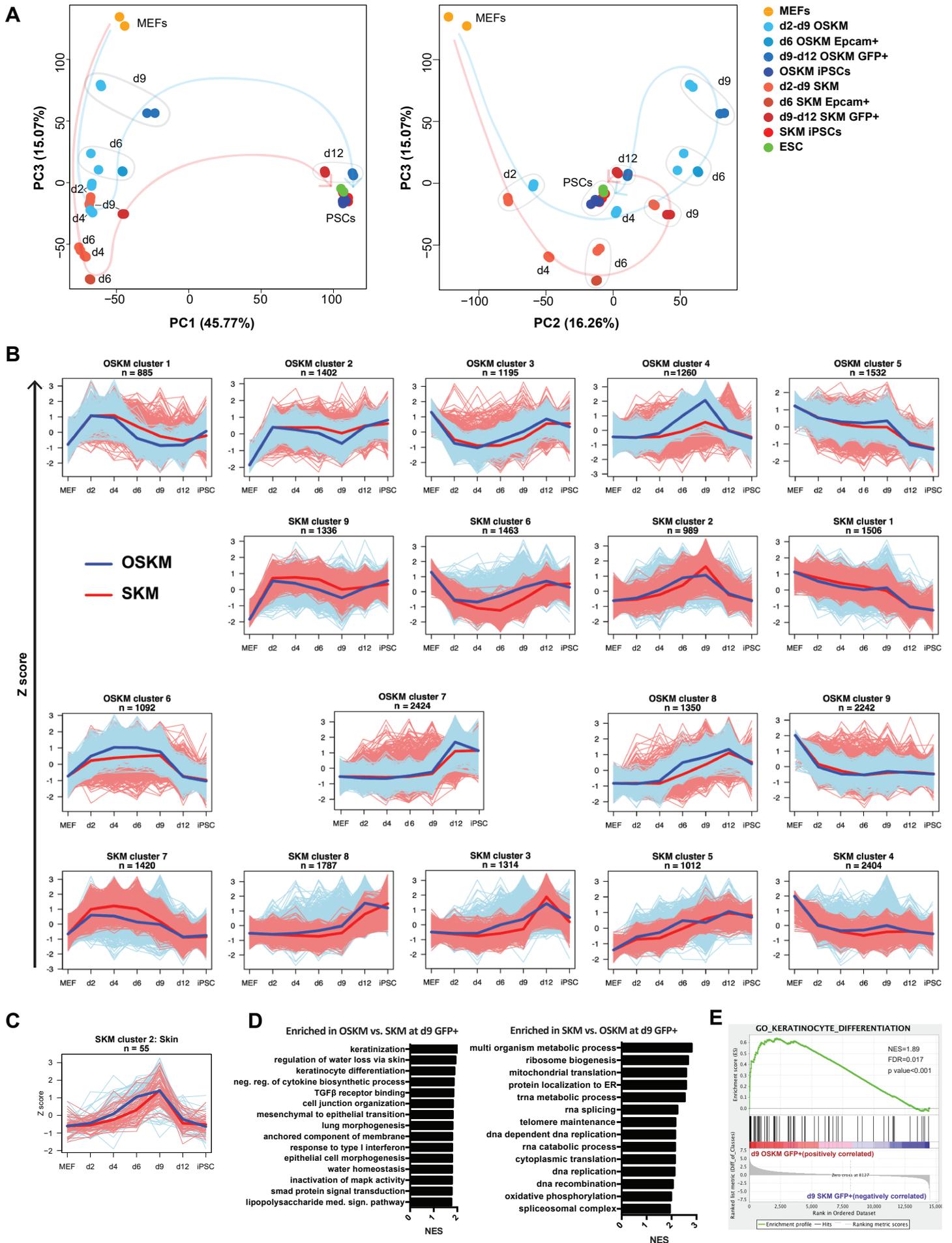
**Nanog is a direct target of Sox2 and Klf4**

(A) Time-course immunostaining for Oct4 and Nanog of GOF18-Rosa26-rtTA MEFs induced with SKM. Nuclei were stained with Hoechst 33342.

(B,C) Genome plots of Oct4, Sox2, Klf4, and cMyc binding at *Nanog* (B) and *Pou5f1* (C) loci.

(D) Model summarizing time-course RNA-seq data showing the molecular roadmap of SKM reprogramming.

Figure S6



**Figure S6. Related to Figure 6**

**Comparison of global gene expression changes of during OSKM and SKM reprogramming**

(A) Principle component analysis (PCA) of global gene expression in time-course samples during OSKM and SKM reprogramming. PC1 vs. PC2 plot is shown in Figure 6A.

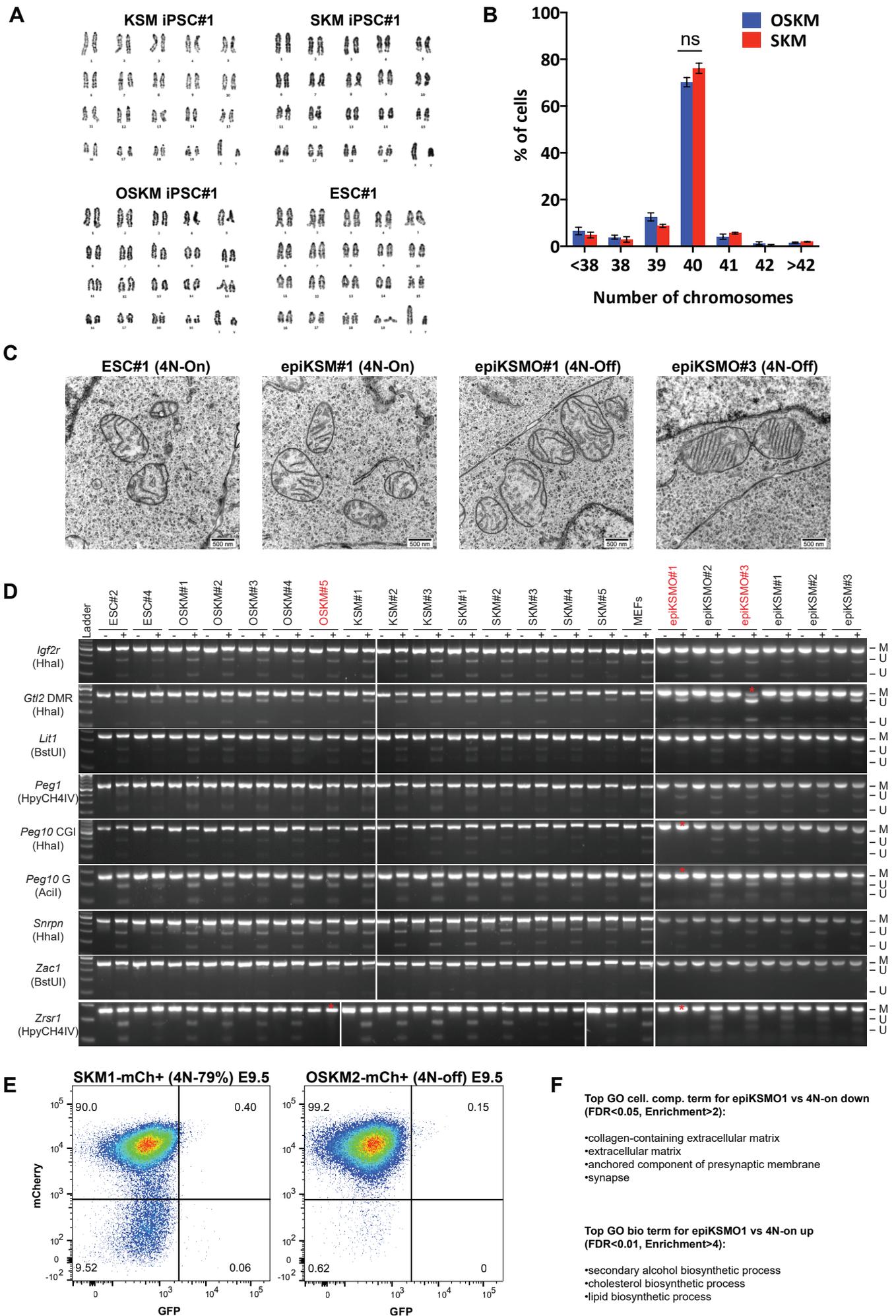
(B) Time-course gene expression of 9 OSKM and 9 SKM gene clusters exhibiting similar kinetics (downregulation, upregulation, transient upregulation, etc.), determined by mFuzz package. Z-scores of all the genes within each cluster were plotted and compared between OSKM and SKM samples. The plots of SKM clusters were placed below OSKM clusters with similar behavior for convenience. Only the genes with TPM at least 1 and FC>2 across the samples were considered.

(C) Time-course plots of 'Skin' GO term genes from SKM mFuzz cluster 2 (for OSKM mFuzz cluster 4 see Figure 6D).

(D) Gene set enrichment analysis of d9 GFP<sup>+</sup> OSKM vs. SKM samples performed by GSEA3.0 package. The data are represented as normalized enrichment score (NES). Only nonredundant categories are shown ( $p < 0.05$  and FDR < 0.1).

(E) A representative gene set enrichment plot of d9 GFP<sup>+</sup> OSKM vs. SKM samples generated by GSEA3.0 for gene ontology term 'Keratinocyte differentiation'.

Figure S7



**Figure S7. Related to Figure 7**

**SKM iPSCs are characterized by more ESC-line gene expression**

(A) Representative images of Giemsa-stained metaphase spreads of established OSKM, SKM, and KSM iPSC and ESC lines.

(B) Karyotyping analysis of bulk OSKM or SKM iPSCs on the second passage. Three bulk lines were generated with each reprogramming vector by sorting the Oct4-GFP<sup>+</sup> cells 2 weeks after reprogramming. At least 100 chromosomal spreads were analyzed for each line. Error bars represent SD, n=3. Statistical significance was calculated using t-test.

(C) Representative electron microscopy images of mitochondria for 4N-on and 4N-off iPSC and ESC lines.

(D) Combined bisulfite restriction analysis (COBRA) assay of the imprinted genes that could be misregulated during reprogramming. U, non-methylated band; M, methylated band. Red font and red asterisk mark the lines exhibiting LOI.

(E) Representative FACS plots of E9.5 embryos derived from 4N aggregates with 4N-on and 4N-off iPSCs marked with mCherry. The mCherry<sup>+</sup> cells were subsequently used for RNA-seq.

(F) Representative and the top gene ontology (GO) term enriched in epiKSMO#1 iPSC 4N-off line vs. 4N-on iPSCs ( $\geq 5$  reads in any sample, FC>1.2, p<0.05 by DESeq2).

**Supplementary Table 1. Related to Figure 1**  
**Overview of the studies attempting to replace Oct4 in Yamanaka cocktail**

Reprogramming cocktail	Somatic cell type	Vectors used	Reprogramming efficiency compared to Oct4	Negative control	Reference
KSM KSM+Oct1 KSM+Oct6	Mouse embryonic fibroblasts	Moloney murine leukemia virus (MMLV)-based, monocistronic	0%	KSM	(Nakagawa et al., 2008)
KSM + BIX-01294	Mouse fetal neural progenitor cells	Murine stem cell virus (MSCV)-based, monocistronic	~13%	KSM	(Shi et al., 2008)
KSM + Nr5a1 KSM + Nr5a2 KS + Nr5a2	Mouse embryonic fibroblasts	MMLV-based, monocistronic	n.d., >0%	KS KSM KSM + Nanog/Sall4 / Esrrb/Klf2/ Klf5/N-Myc/ Stat3/Zfx	(Heng et al., 2010)
KSM + E-cadherin	Mouse embryonic fibroblasts	MMLV-based, monocistronic	~2%	KSM	(Redmer et al., 2011)
KSM + Sall4 + Nanog	Mouse embryonic fibroblasts	Lentiviral TetO, monocistronic	n.d., >0%	n.d.	(Buganim et al., 2012)
KSM + human Oct4 KSM + Xenopus Oct91 KSM + medaka Pou2 KSM + axolotl Oct4 KSM + axolotl Pou2 KSM + zebrafish Pou2	Mouse embryonic fibroblasts, human skin fibroblast	MMLV-based, monocistronic	~100% for human Oct4 ~70% for Xenopus Oct91 ~10% for medaka Pou2 ~10% for axolotl Pou2 ~2% for axolotl Oct4 0% for zebrafish Pou2	KSM	(Tapia et al., 2012)
KSM + Gata3 KSM + Gata4 KSM + Gata6 KSM + Sox7 KSM + Pax1 KSM + CEBPa KSM + HNF4a KSM + GRB2 KM+ Sox1/Sox3 + Gata3/Gata6/Pax1	Mouse embryonic fibroblasts; for KSM + Gata3 also mouse gastric epithelial cells, mouse adult dermal fibroblast, and mouse keratinocytes	Lentiviral TetO, monocistronic	~140% for Gata6 ~115% for Gata3 ~80% for Sox7 ~50% for Pax1 ~40% for Gata4	KSM	(Shu et al., 2013)
KS <sup>VP16M</sup> + Gata3 <sup>VP16</sup> KS <sup>VP16M</sup> + Gata4 <sup>VP16</sup> KS <sup>VP16M</sup> + Gata6 <sup>VP16</sup>	Human foreskin fibroblast	MMLV-based, MSCV-based, monocistronic	~0.1% for Gata3 <sup>VP16</sup> 0% for Gata4 <sup>VP16</sup> 0% for Gata6 <sup>VP16</sup>	KS <sup>VP16M</sup>	(Montserrat et al., 2013)
KSM + Tet1 KS + Tet1	Mouse embryonic fibroblasts, mouse trophoblast stem cells	Lentiviral TetO, monocistronic	~70%	KSM	(Gao et al., 2013b)
KSM + forskolin KSM + 2-Me-5HT KSM + D4476	Mouse embryonic fibroblasts	Lentiviral CMV, monocistronic	~100 and 14% for forskolin	KSM KS	(Hou et al., 2013)

KS + FSK KS + 2-Me-5HT KS + D4476			~30 and 0% for 2-Me-5HT ~30 and 0% for D4476 with and without cMyc, respectively		
KSM + TALE-based transcriptional activator of Oct4	Mouse embryonic fibroblasts	Piggyback- transposed dox- inducible polycistronic KSM + monocistronic Oct4/TALE	~160%	n.d.	(Gao et al., 2013a)
KSM + SB43 +VPA	Mouse embryonic fibroblasts	Lentiviral TetO, monocistronic	~70%	KSM	(Tan et al., 2015)
KSM + Brn4	Mouse embryonic fibroblasts	Lentiviral TetO- STEMCCA-BKSM, polycistronic	n.d., >0%	n.d.	(Bar-Nur et al., 2015)
KSM + Gata1 KSM + Gata2 KSM + Gata3 KSM + Gata4 KSM + Gata5 KSM + Gata6	Mouse embryonic fibroblasts	Lentiviral TetO, monocistronic	~190% for Gata2 ~140% for Gata6 ~130% for Gata1 ~115% for Gata3 ~80% for Gata5 ~40% for Gata4	KSM	(Shu et al., 2015)
KSM + GNAS KSM + forskolin	Mouse embryonic fibroblasts	Lentiviral TetO- STEMCCA-KSM, polycistronic + lentiviral GNAS	~10% for forskolin ~5% for GNAS	KSM*	(Fritz et al., 2015)
KSM + library of artificial TFs	Mouse embryonic fibroblasts	MMLV-based, monocistronic KSM + monocistronic lentiviral artificial TFs	n.d., >0%	KSM	(Eguchi et al., 2016)
KSM + Oct6 mutant	Mouse embryonic fibroblasts	MMLV-based, monocistronic	~3%	KSM + Oct6	(Jerabek et al., 2016)

\* a few Oct4<sup>+</sup> colonies were generated, but not characterized

**Supplementary Table 2. Related to Figure 3**  
**Results of tetraploid complementation experiments**

Lines tested	Karyotype	Aggregates transferred	Full-Term pups	Breathing	Survived after 48h	Survived after 3 months
OSKM#1	40, XY	46	1	1	0	0
OSKM#2	40, XY	42	0	0	0	0
OSKM#3	40, XY	55	1	1	0	0
OSKM#4	40, XY	52	4	4	0	0
OSKM#5	41, XY, +14	63	0	0	0	0
ESC line#1	40, XY	30	22	22	22	22
ESC line#2	40, XY	40	14	12	4	1
ESC line#3	40, X0 +19	30	0	0	0	0
ESC line#4	39, X0	30	0	0	0	0
KSM#1	40, XY	43	19	15	3	3
KSM#2	40, XY	32	18	10	6	2
KSM#3	41, XY +1	25	3	4	2	0
SKM#1	40, XY	33	26	26	23	19
SKM#2	40, XY	22	13	13	6	3
SKM#3	39, X0	22	15	15	6	4
SKM#4	40, XY	43	4	4	3	2
SKM#5	39, X0	36	5	3	2	2
epi-KSM#1	40, XY	42	19	13	12	9
epi-KSM#2	40, XY	50	15	12	4	3
epi-KSM#3	40, XY	43	24	17	11	10
epi-KSM+O#1	40, XY	31	0	0	0	0
epi-KSM+O#2	40, XY	30	19	16	8	5
epi-KSM+O#3	40, XY	29	0	0	0	0

**Supplementary Table 3. Related to Key Resource Table  
Primers used in the study**

**Primers for PCR genotyping**

<b>Gene Name</b>	<b>Sequence</b>	<b>Annealing temperature</b>
<i>Viral Oct4</i>	5'-GAGACGCCATCCACGCTGT-3'	60°C
	5'-GGTGAGAAGGCCAAGTCTGAAG-3'	
<i>Viral Brn4</i>	5'-GAGACGCCATCCACGCTGT-3'	60°C
	5'-ATGGACAAGGGAGCTGGAAC-3'	
<i>Viral SKM</i>	5'-GAGACGCCATCCACGCTGT-3'	60°C
	5'-GGCTTCAGCTCCGTCTCCAT-3'	
<i>Viral KSM</i>	5'-GAGACGCCATCCACGCTGT-3'	60°C
	5'-GTGGAGAAGGACGGGAGCAG-3'	
<i>Viral cMyc</i>	5'-ACAGCTTCGAAACTCTGGTGCAT-3'	60°C
	5'-AGGAAGGTCCGCTGGATTGA-3'	
<i>ActB</i>	5'-ACTGCCGCATCCTCTTCCTC-3'	60°C
	5'-AGGAAGGTCCGCTGGATTGA-3'	

**Primers bisulfate sequencing and COBRA**

<b>Gene Name</b>	<b>Sequence</b>	<b>Annealing temperature</b>
<i>Col1a1</i>	5'-TGGTATAAAAGGGGTTTAGGTTAGT-3'	60°C
	5'-ACAATAACCCCTAAAAAACAAAAA-3'	
<i>Nanog</i>	5'-TTTGTAGGTGGGATTAATTGTGAA-3'	60°C
	5'-AAAAAACAAAACACCAACCAAAT-3'	
<i>Pou5f1 (Oct4)</i>	5'-GGGTTAGAGGTTAAGGTTAGAGGG-3'	60°C
	5'-CCCCACCTAATAAAAAATAAAAAA-3'	
<i>Viral KSM</i>	5'-GAGACGCCATCCACGCTGT-3'	60°C
	5'-GTGGAGAAGGACGGGAGCAG-3'	
<i>Viral cMyc</i>	5'-ACAGCTTCGAAACTCTGGTGCAT-3'	60°C
	5'-AGGAAGGTCCGCTGGATTGA-3'	
<i>ActB</i>	5'-ACTGCCGCATCCTCTTCCTC-3'	60°C
	5'-AGGAAGGTCCGCTGGATTGA-3'	

**Q-PCR primers**

<b>Gene Name</b>	<b>Sequence</b>	<b>Annealing temperature</b>
<i>WPRE</i>	5'-TGTTGCCACCTGGATTCTGC-3' 5'-AGGAAGGTCCGCTGGATTGA-3'	60°C
<i>Viral cMyc</i>	5'-ACAGCTTCGAAACTCTGGTGCAT-3' 5'-AGGAAGGTCCGCTGGATTGA-3'	60°C
<i>Rlp37a</i>	5'- ACTTGCTCCTTCTGTGGCAAGAC -3' 5'- TTCATGCAGGAACCACAGTGC -3'	60°C
<i>Hprt1</i>	5'-CTGGTGAAAAGGACCTCTCGA-3' 5'-CTGAAGTACTCATTATAGTCAAGGGCAT-3'	60°C
<i>Nr5a2</i>	5'-AAGGGTGAGCTGCAAAGGGGA-3' 5'-CCCAGGTTTTGGGGACGCTCG-3'	60°C
<i>Nanog</i>	5'-CTTTCACCTATTAAGGTGCTTGC-3' 5'-ATGGCATCGGTTTCATCATGGTAC-3'	60°C
<i>Sall4</i>	5'-TGGCAGACGAGAAGTTCTTTC-3' 5'-TCCAACATTTATCCGAGCACAG-3'	60°C
<i>Thy1</i>	5'-TTACCCTAGCCAACCTCACCACCA-3' 5'-AAATGAAGTCCAGGGCTTGGAGGA-3'	60°C
<i>Slug (Snai1)</i>	5'-CACATTCGAACCCACACATTGCCT-3' 5'-TGTGCCCTCAGGTTTGATCTGTCT-3'	60°C
<i>Epcam (Cd326)</i>	5'-GCTGGCAACAAGTTGCTCTCTGAA-3' 5'-CGTTGCACTGCTTGGCTTTGAAGA-3'	60°C
<i>Cdh1(E-cad)</i>	5'-AACAACTGCATGAAGGCGGGAATC-3' 5'-CCTGTGCAGCTGGCTCAAATCAAA-3'	60°C
<i>Oct4-CDS</i>	5'-GGCTAGAGAAGGATGTGGTTCGAG-3' 5'-CCTGGGAAAGGTGTCCCTGTAG-3'	60°C
<i>Esrrb</i>	5'-GCCTTTACTATCTGTGCCTGGT-3' 5'-TAGTGCTTCTCTTTGGTGCTGT-3'	60°C
<i>Tfcp2l1</i>	5'-AACCCGCCCAGGTAGAGGCT-3' 5'-AGGGCAGCCACGTGGGAAGA-3'	60°C
<i>Rex1 (Zfp42)</i>	5'-GGCTGCGAGAAGAGCTTTATTCA-3' 5'-AGCATTTCTTCCCGGCCTTT-3'	60°C
<i>Lin28a</i>	5'-CCGCAGTTGTAGCACCTGTCT-3' 5'-GAAGAACATGCAGAAGCGAAGA-3'	60°C