Stem Cell Reports, Volume *15*

Supplemental Information

Context-Dependent Requirement of Euchromatic Histone Methyltransferase Activity during Reprogramming to Pluripotency

Simon E. Vidal, Alexander Polyzos, Kaushiki Chatterjee, Ly-sha Ee, Emily Swanzey, Jorge Morales-Valencia, Hongsu Wang, Chaitanya N. Parikh, Bhishma Amlani, Shengjiang Tu, Yixiao Gong, Valentina Snetkova, Jane A. Skok, Aristotelis Tsirigos, Sangyong Kim, Effie Apostolou, and Matthias Stadtfeld

SUPPLEMENTAL FIGURES

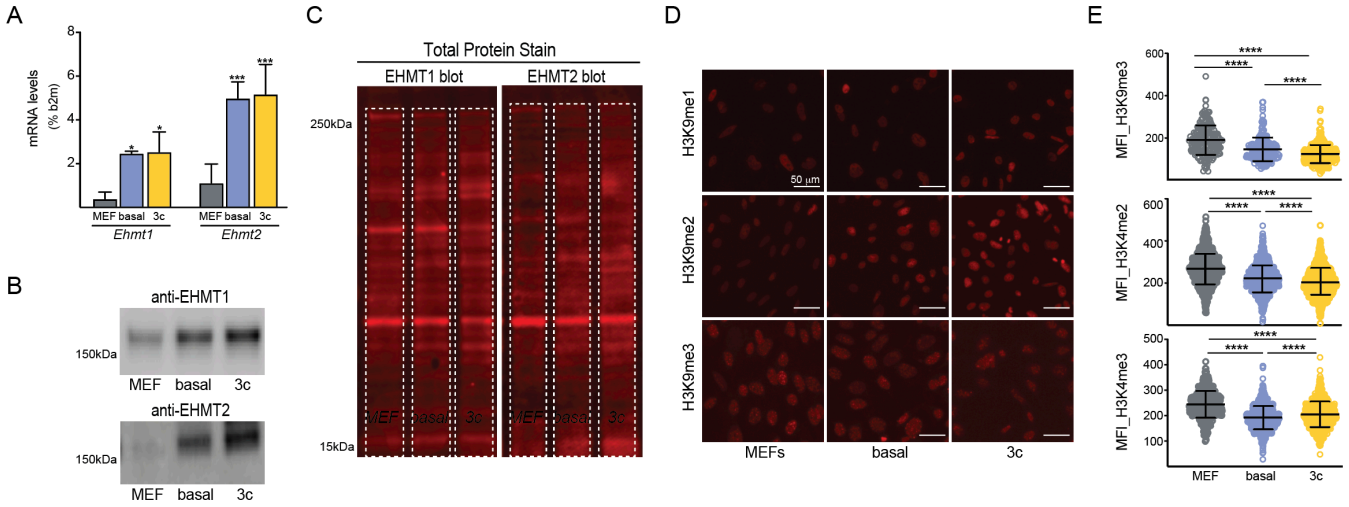

Figure S1. Increased EHMT activity during early stages of OKSM-driven reprogramming. Related to Figure 1. (A) Quantification of *Ehmt1* and *Ehmt2* by quantitative real-time PCR using RNA isolated from indicated conditions 24 h after initiation of reprogramming. N = 2 independent experiments. **(B)**. Representative immunoblots of EHMT1 and EHMT2 protein 24 h after initiation of reprogramming. **(C)** Total protein staining of corresponding blot in (C). White dotted lines outline the areas used for total protein quantification, as used in Figures 1C, S4D and S5E. **(D)** Fluorescence images of MEFs and cells 24 h after initiation of reprogramming in indicated conditions after staining with antibodies against H3K9me1, H3K9me2 or H3K9me3, respectively. **(E)** Representative quantification of H3K9me3, H3K4me2 and H3K4me3 levels in indicated conditions. More than 100 nuclei of similar size were measured for each cell population in at least two independent experiments. Significance in (A,E) with one-way ANOVA with Tukey post-test with *p<0.05, ***p<0.001 and ****p<0.0001.

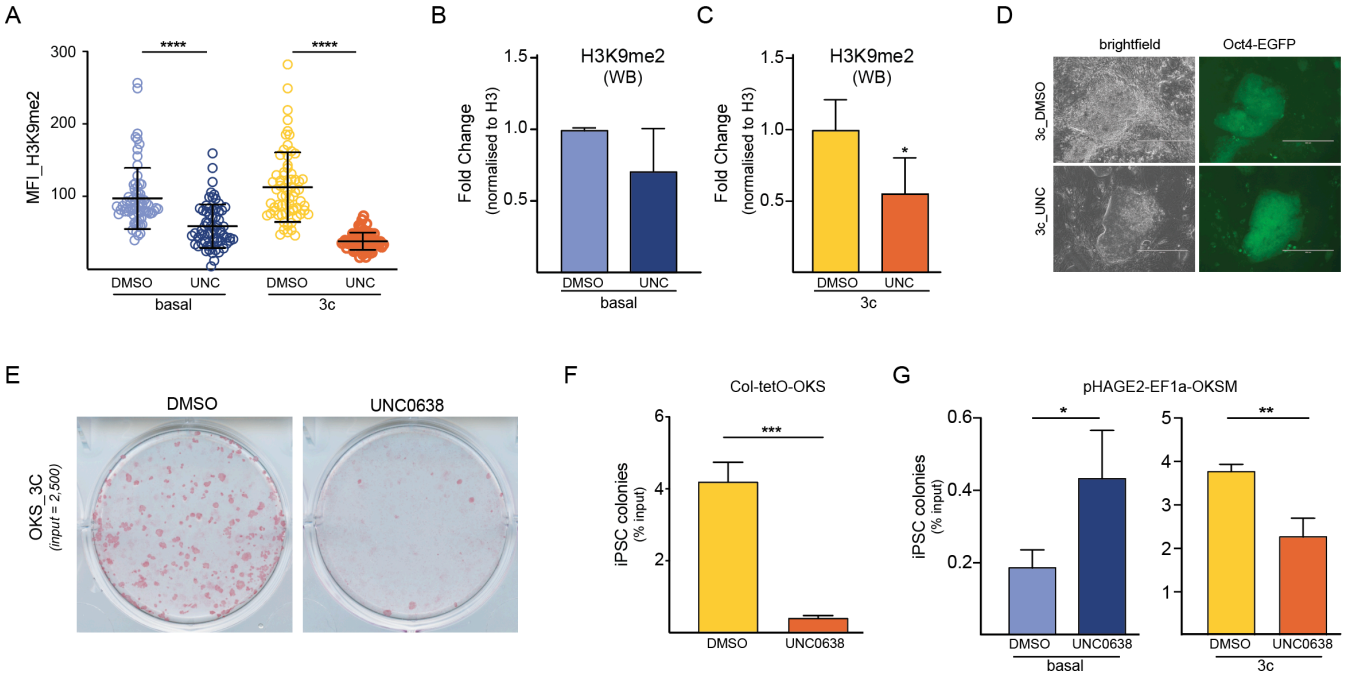

Figure S2. EHMT activity supports 3c enhanced reprogramming. Related to Figure 2. (A) Quantification by IF of H3K9me2 levels in cells either exposed to UNC0638 (UNC) or DMSO for the first 24 h of OKSM expression. Significance with one-way ANOVA with Sidak post-test, ****p<0.0001. At least 100 size-matched nuclei were quantified for each condition in three independent experiments. **(B)** Representative fold change (normalized to H3) of H3K9me2 protein levels analyzed after 24 h of reprogramming initiation by WB in basal conditions with UNC treatment (n = 4 independent experiments). **(C)** Like (B) but in 3c enhanced conditions. **(D)** Representative images of colonies scored as iPSCs based on morphology and Oct4-EGFP expression. Scale bars indicate 400µm. **(E)** Representative AP staining of transgene-independent iPSC colonies obtained upon expression of OKS in MEFs for nine days in 3c enhanced reprogramming conditions in absence and presence of UNC0638. **(F)** Quantification of iPSC colony formation upon OKS expression in indicated conditions. N = 3 independent experiments **(G)** Number of iPSC colonies formed upon transducing MEFs with a constitutive lentiviral vector expressing OKSM and reprogramming cells under basal or 3c enhanced conditions in absence of presence of EHMT inhibitor. N = 3 independent experiments. Significance in (F,G) with two-tailed t-test with *p<0.05, **p<0.01 and ***p<0.001.

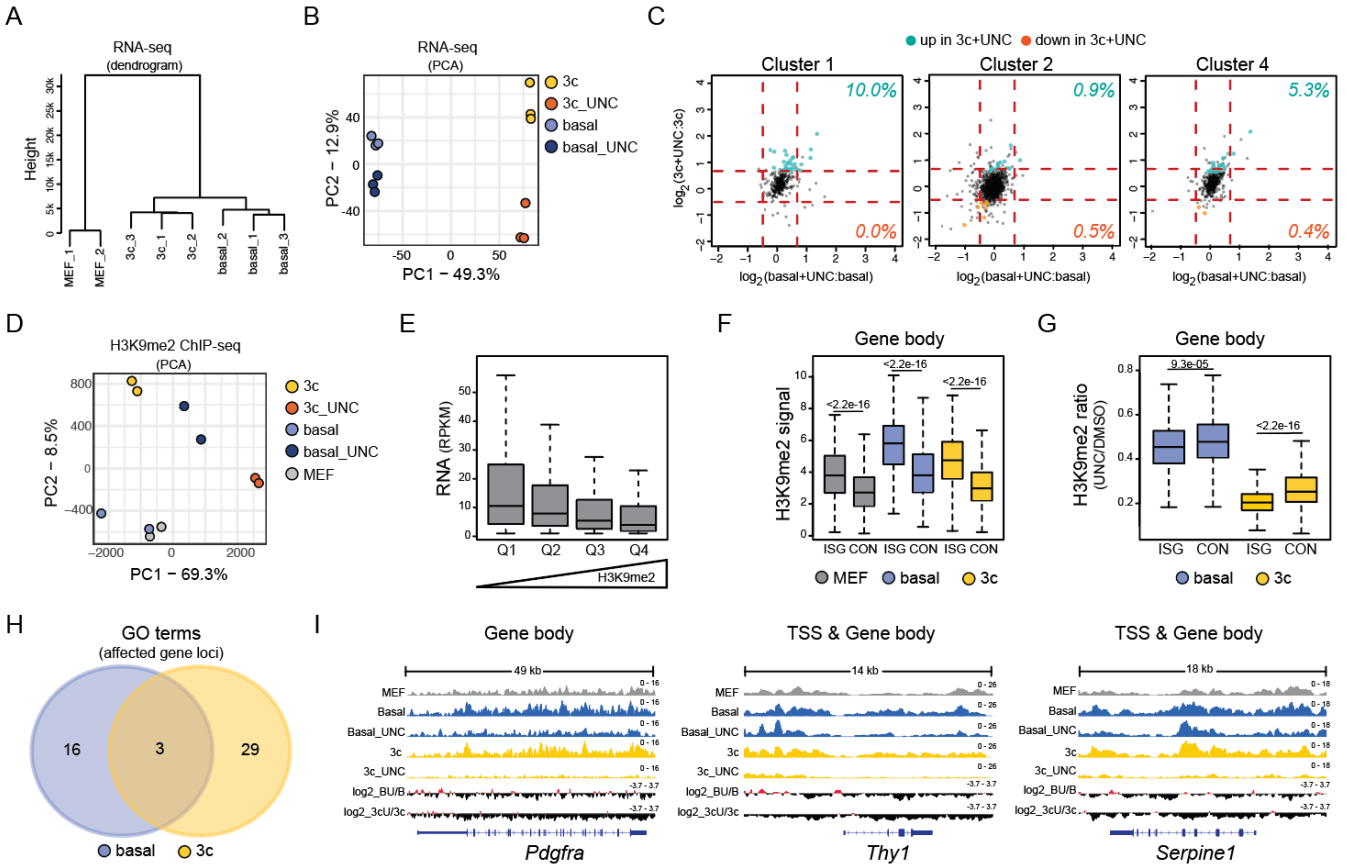

Figure S3. Molecular consequences of EHMT inhibition during early reprogramming stages. Related to Figure 3. (A) Hierarchical tree after unsupervised clustering of indicated RNA-seq samples. **(B)** Principal component analysis (PCA) of RNA-seq data obtained from cells undergoing reprogramming for two days in indicated conditions. **(C)** Effect of EHMT inhibition (UNC) on the expression levels of genes associated with clusters 1, 3 and 5 during basal (X-axis) and 3c enhanced (Y-axis) reprogramming. Transcripts with significantly changed abundance during enhanced reprogramming (p-adj<0.05; FC>1.5) are highlighted in green (failed downregulation) and orange (failed upregulation), respectively. Numbers indicate percentage of cluster-specific genes affected by UNC0638 (UNC). **(D)** PCA analysis of H3K9me2 ChIP-seq data in indicated conditions and cell types. **(E)** Average RNA expression levels (RPKM) of gene loci ranged based on H3K9me2 levels into four quartiles (Q1 lowest H3K9me2 and Q4 highest H3K9me2). N = 2 independent experiments. **(F)** H3K9me2 ChIP-signal over the gene body in 328 MEF-associated genes inefficiently silenced (ISG) in presence of UNC0638 during 3c reprogramming and 2664 control MEF-associated genes that are efficiently silenced (CON). N = 2 independent experiments. **(G)** Ratio of H3K9me2 signal in presence and absence of UNC0638 over the gene body of ISGs and control genes during basal and 3c reprogramming. N = 2 independent experiments. **(H)** Pie diagram showing limited overlap of GO terms associated with gene loci whose H3K9me2 levels are most strongly affected during basal or during 3c enhanced reprogramming, respectively (see also Figure 3H). **(I)** Representative tracks showing H3K9me2 levels at three MEF-associated genes in MEFs and in indicated reprogramming conditions. Noteworthy features are increased H3K9me2 levels at gene bodies and/or TSS regions during reprogramming and dramatic reduction of this chromatin mark upon UNC0638 treatment in presence of 3c compounds.

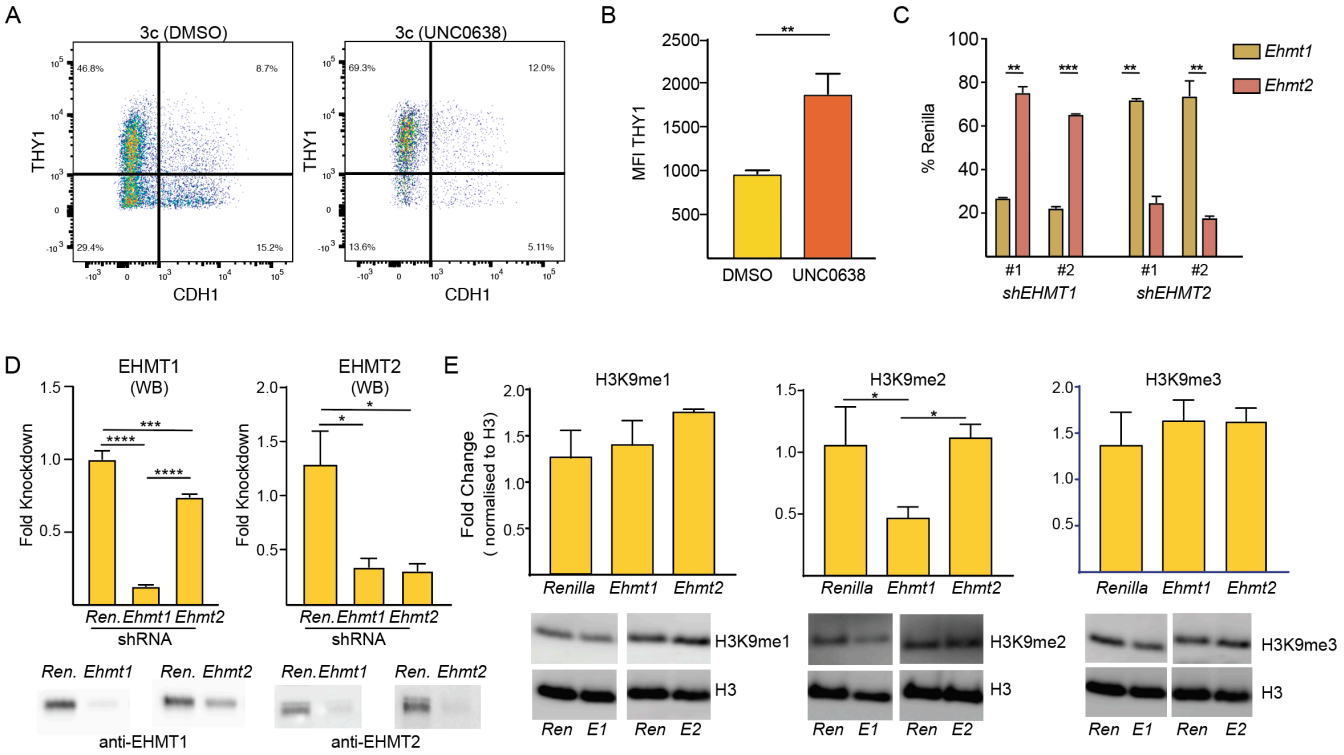

Figure S4. Consequences of EHMT inhibition for later stages of iPSC reprogramming. Related to Figure 4. (A) Flow cytometry histograms showing expression levels of the fibroblast marker THY1 on CDH1^{+} cells at Day 4 of 3c enhanced reprogramming in absence and presence of EHMT inhibitor. **(B)** Quantification of THY1 levels on CDH1^{+} intermediates at Day 4 of 3c enhanced reprogramming, MFI = mean fluorescence intensity (geometric mean; as measured by flow cytometry). Significance with t-test, **p<0.01. N = 3 independent experiments. **(C)** mRNA levels of *Ehmt1* and *Ehmt2* in MEFs transduced with indicated shRNAs and undergoing reprogramming in 3c conditions for 24 h relative to cells expressing shRNA targeting *Renilla*. Significance with multiple two-tailed t-tests with **p<0.01 and ***p<0.001. N = 2 independent experiments. **(D)** Representative immunoblots (bottom) and quantification (top) of mean fold knockdown of EHMT1 and EHMT2 protein levels analyzed after 24 h of reprogramming initiation in shRNA transduced MEFs (Renilla controls). N = 3 independent experiments. **(E)** Representative immunoblots and quantification of mean fold change of H3K9me1, H3K9me2 and H3K9me3 protein levels analyzed after 24 h of reprogramming initiation in shRNA transduced MEFs. Ren = *Renilla* shRNA; E1 = *Ehmt1* shRNA; E2 = *Ehmt2* shRNA. N = 3 independent experiments. Significance in (D,E) calculated with one-way ANOVA with Tukey post-test with *p<0.05, **p<0.01 and ***p<0.001.

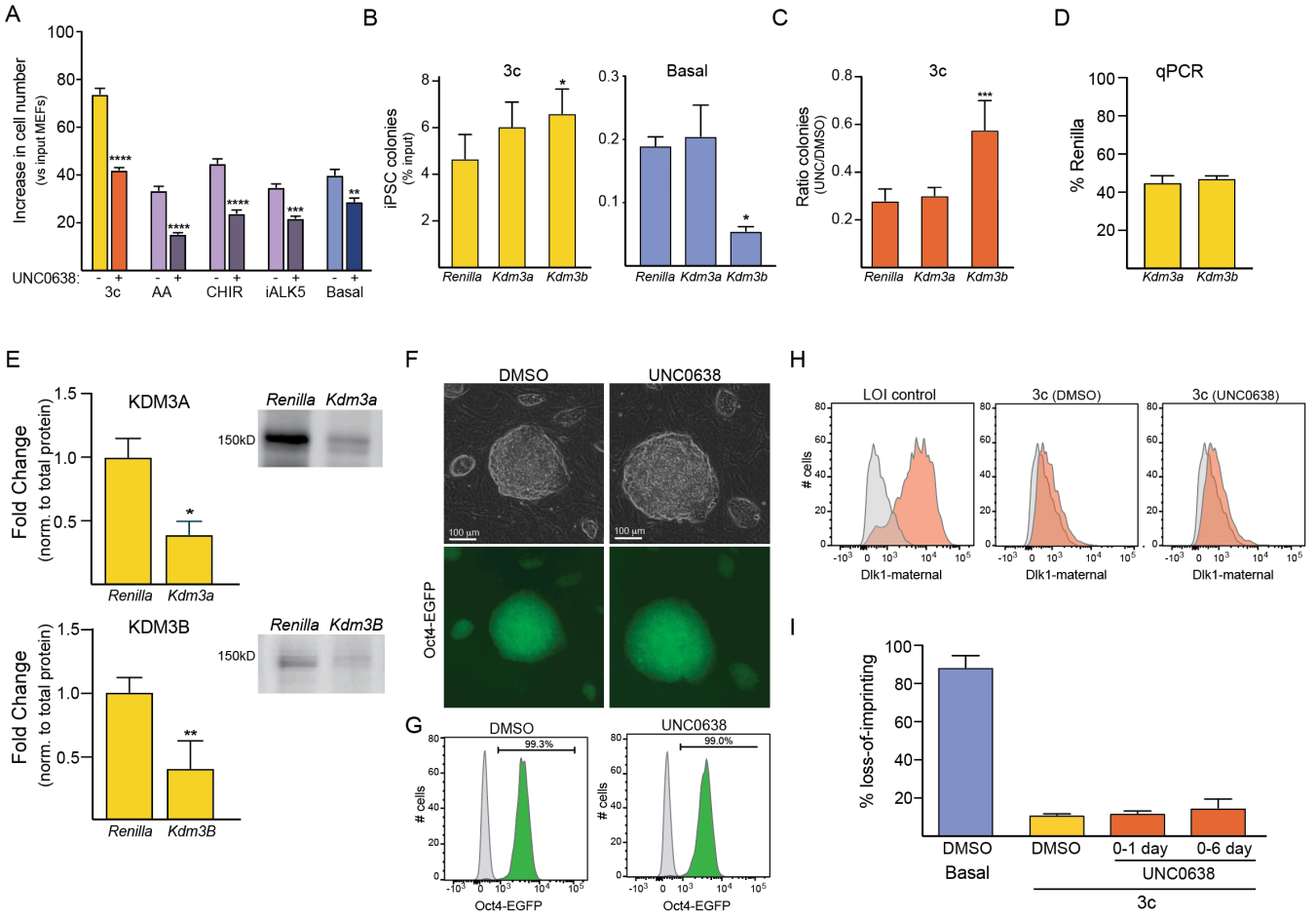

Figure S5. AA establishes a requirement for EHMT activity during enhanced iPSC reprogramming. Related to Figure 5. (A) Quantification of total cell numbers after expression of OKSM factors for 48 h in indicated conditions in absence or presence of UNC0638. N = 3 independent experiments. **(B)** Percentage of iPSC colonies (per input MEFs) formed after shRNA-mediated KD of *Kdm3a* or *Kdm3b* during 3c enhanced (left panel) and basal reprogramming (right panel), respectively. N = 2 (for basal) or 4 (for 3c) independent experiments. **(C)** Ratio of iPSC colonies formed during 3c enhanced reprogramming in presence and absence of EHMT inhibitor upon KD of indicated H3K9 demethylases. Significance in (B, C) with one-way ANOVA with Dunnett post-test with *p<0.05 and ***p<0.001. N = 4 independent experiments. **(D)** mRNA levels of *Kdm3a* and *Kdm3b* in cells undergoing reprogramming in 3c conditions for 24 h upon expression of respective shRNAs relative to cells expressing shRNA targeting *Renilla*. N = 2 independent experiments. **(E)** Representative immunoblots and quantification of KDM3A and KDM3B protein levels analyzed after 24 h of reprogramming initiation in shRNA transduced MEFs normalized to total protein levels in Renilla controls. N = 3 independent experiments. **(F)** Representative images of P1 3c-iPSCs derived in either absence or presence of UNC0638. **(G)** Quantification of Oct4-EGFP expression in 3c-iPSCs derived in either absence or presence of UNC0638. **(H)** Expression of the normally silenced maternal *Dlk1* allele as measured by flow cytometry in 3c-iPSCs derived either in absence or presence of EHMT inhibitor, compared to control cells in which imprint dysregulation (LOI) has resulted in upregulation of maternal *Dlk1*. Grey-shaded histogram indicates background fluorescence levels. **(I)** Quantification of the percentage cells expressing maternal *Dlk1* in iPSCs derived under indicated conditions. N = 3 independent experiments.

Table S5. Oligos used

Name	Purpose	Sequence (5'-3')
miRE-Xho-fwd	cloning	TGAACTCGAGAAGGTATATTGCTGTTGACAGTGAGCG
miRE-EcoOligorev	cloning	TCTCGAATTCTAGCCCCTTGAAGTCCGAGGCAGTAGGC
Ehmt1_1_97mer	shRNA	TGCTGTTGACAGTGAGCGCCGCTATGATGATGATGAATAATAGTGAAGCCACAGATGTATTATTCATCATCATCATAGCGTTGCCTACTGCCTCGGA
Ehmt1_2_97mer	shRNA	TGCTGTTGACAGTGAGCGAGACGGTGATTGAGATGTTTAATAGTGAAGCCACAGATGTATTAAACATCTCAATCACCGTCCTGCCTACTGCCTCGGA
Ehmt2_1_97mer	shRNA	TGCTGTTGACAGTGAGCGACCCCCTGATCTTTGAGTGTAATAGTGAAGCCACAGATGTATTACACTCAAAGATCAGGGGGGTGCCTACTGCCTCGGA
Ehmt2_2_97mer	shRNA	TGCTGTTGACAGTGAGCGATGCAGCTCAATCGAAAGCTTATAGTGAAGCCACAGATGTATAAGCTTTCGATTGAGCTGCAGTGCCTACTGCCTCGGA
Kdm3a_97mer	shRNA	TGCTGTTGACAGTGAGCGCCAGGAGATTACAATTCAACAATAGTGAAGCCACAGATGTATTGTTGAATTGTAATCTCCTGATGCCTACTGCCTCGGA
Kdm3b_97mer	shRNA	TGCTGTTGACAGTGAGCGCCACGATGAAGAAGTACTCAAATAGTGAAGCCACAGATGTATTTGAGTACTTCTTCATCGTGATGCCTACTGCCTCGGA
Kdm3a_fwd	qPCR	CAGCAACTCCATCTAGCAAGG
Kdm3a_rev	qPCR	TGTTCTCGGTACTTCAGGTTTTG
Kdm3b_fwd	qPCR	TGTGGTGTTGTGAGCCGTC
Kdm3b_rev	qPCR	TCTGGGATCTACTGACTTGACC
b2m_fwd	qPCR	TTCTGGTGCTTGTCTCACTGA
b2m_rev	qPCR	CAGTATGTTCGGCTTCCCATTC

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Mice

Derivation, handling, and genotyping of reprogrammable mice (JAX011001) with the *Oct4-GFP* allele were described previously (Stadtfeld et al., 2010). All animal experiments were in accordance with the guidelines of the NYU School of Medicine Institutional Animal Care and Use Committee.

Basic cell culture and cell culture-based assays

MEF cultures were established by trypsin digestion of midgestation (embryonic day (E) 13.5–E15.5) embryos and maintained in DMEM supplemented with 10% FBS, L-glutamine, penicillin/streptomycin, nonessential amino acids and β-mercaptoethanol. Reprogrammable MEFs were heterozygous for *Rosa26-rtTA* and for *Oct4-GFP* and either heterozygous for an inducible OKSM allele (Stadtfeld et al., 2010) or homozygous for an inducible OKS allele (Borkent et al., 2016). Established iPSCs were cultured on growth-arrested feeder cells in KO-DMEM (Invitrogen) supplemented with L-glutamine, penicillin/streptomycin, nonessential amino acids, β-mercaptoethanol, 1,000 U/ml LIF and 15% FBS ("ESC medium"). Reprogramming was carried out as previously described (Vidal et al., 2014). Briefly, were seeded on a layer of growth-arrested feeder cells in ESC medium in the presence of 1 μg/ml Dox and, if applicable, L-ascorbic acid (50 μg/ml), CHIR99021 (3 μM) and TGF-b RI Kinase Inhibitor II (250 nM) ("3c"). The number of input MEFs ranged between 50-500 cells/cm^2 and was adjusted to prevent overcrowding in the more efficient conditions and allow the formation of sufficient numbers of colonies for reliable quantification in the less efficient conditions. If applicable, UNC0638 (1 μM) was added. Media was changed every other day. Colonies were scored visually based on ESC-like morphology and expression of the Oct4-GFP reporter allele after several days in absence of dox, which represents a stringent assessment of pluripotency (Amlani et al., 2018). Colonies were visualized for low magnification image capture after alkaline phosphatase staining. To determine the effect of UNC0638 treatment on cellular proliferation 25,000 reprogrammable MEFs were seeded on each well of gelatinized 12-well plates in MEF media. After 12 hours, media was replaced with mESC media containing either DMSO or UNC0638, Dox and the applicable reprogramming-enhancing chemicals. After 72 hours, cells were harvested, adjusted to the same volume and counted using a Beckman Coulter Z1 Particle Counter.

Immunofluorescence

Reprogrammable MEFs grown on gelatinized 24-well plates in absence of feeders were fixed with paraformaldehyde (4%), permeabilized with Triton X-100 (0.5%), and stained in blocking buffer (5% goat serum, 2 mg/ml fish skin gelatin and 0.2% Tween20 in PBS) with primary antibodies against H3K9me1 (ab9045; 1:200), H3K9me2 (ab1220; 1:200), H3K9me3 (ab8898; 1:200), H3K4me2 (ab7766; 1:200), H3K4me3 (ab8580; 1:200), EHMT1 (ab41969; 1:200) or EHMT2 (C6H3, 1:50) for one hour at room temperature, following by staining with appropriate AlexaFluor555-conjugated secondary antibodies (1:1000). After counterstaining nuclei with DAPI, cells were imaged on a Nikon Eclipse TiE scope equipped with a Lumencor Light Engine and a Neo 5.5 cSMOS camera (Andor). Images belonging to the same experimental series were imaged on the same day using identical instrument and software settings. Fluorescence intensities were determined in NIS elements after background subtraction and after using the autodetect function to identify nuclei based on DAPI signal. Measured fluorescence values were exported to

Prism (GraphPad) for selection of similar-sized nuclei (n>100 for each experiment) and statistical analysis.

Tetraploid blastocyst injections

Zygotes were isolated from BDF1 females as previously described (Stadtfeld et al., 2012) and cultured overnight until they reached the 2-cell stage. One hour after electro-fusion, 1-cell embryos were separated from embryos that had failed to fuse, cultured for another two days and then injected with iPSCs, using 5-10 cells per blastocyst. Viable pups were defined as those that survived for at least three days following birth. Statistical analysis was conducted in Prism (GraphPad).

shRNA mediated knockdown

97-mer oligonucleotides against specific target genes were designed using the splashRNA algorithm (Pelossof et al., 2017), PCR amplified using the primers miRE-Xho-fwd and miRE-EcoOligorev and cloned into the miRE plasmid backbones (Fellman et al. 2013). Viruses were produced using the packaging vectors psPax2, MD2G and Pasha/DCr8 and used on reprogrammable MEFs seeded at about 30% confluency at titers that achieved 20-30% transduction efficiency (determined using an EGFP cassette present miRE) before selection. After virus washout, selection was carried out in MEF media with 500 μg/ml G418 for a total of four days, with one passage after two days. After washout of G418 and one day of culture in MEF media for recovery, cells were harvested, counted and seeded onto growth-arrested feeder cells for reprogramming or onto gelatinized plates for RNA and/or protein extraction. Transduction efficiencies after selection ranged between 80-90% based on EGFP fluorescence. Knockdown was confirmed using quantitative PCR using gene-specific primers. All oligonucleotides are listed in **Table S8**.

Lentiviral reprogramming

Reprogramming was carried out by incubation of MEFs carrying the Oct4-EGFP reporter (but no transgenic OKSM cassette) with doxycycline-inducible mouse OKSM (pHAGE2-tetO-STEMCCA)(Sommer et al., 2009) and rtTA (FUdeltaGW-rtTA, Addgene 19780) (Maherali et al., 2008) lentiviruses to achieve between 20-30% of cells expressing OKSM in a dox-dependent manner (corresponding to on average a single pHAGE2-tetO-STEMCCA insertion per transduced cell). Without selection, MEFs were then seeded at a density of approximately 2,500 cells/cm^2 (for basal reprogramming) and 250 cells/cm^2 (for 3c reprogramming) onto a layer of Mitomycin-C-treated feeder cells in ESC media supplemented with doxycycline and, if applicable, UNC0638 and 3c compounds. Media was changed every two days and on day 7 (for 3c reprogramming) or day 12 (for basal reprogramming) dox and all other compounds were washed out, followed by four more days of culture in ESC medium before colonies were scored based on Oct4 expression. Reprogramming efficiencies were calculated based on the total number of input MEFs.

Flow cytometry

Reprogramming cultures were harvested by incubation with pre-warmed 0.25% trypsin/1mM EDTA solution for 5 minutes at 37°C. Single-cell suspensions were obtained by repetitive pipetting and transfer through a 40 μm cell strainer. Cells were incubated with eFluor 450-conjugated anti-THY1 (53-2.1), biotin-conjugated anti-CDH1 (DECMA-1) and PE/Cy7-conjugated anti-EpCAM (G8.8), followed by incubation with Streptavidin-APC (all eBiosciences)

and data acquired on a FACS LSR2 (BD Biosciences), using DAPI to identify dead cells. Data analysis was conducted with FlowJo (Tree Star) and with Prism (GraphPad). For the relative quantification of different intermediate cell populations in absence and presence of UNC0638, cultures were harvested completely at the indicated day of reprogramming, stained with the appropriate antibodies, volume adjusted and run through the flow cytometer for the same amount of time at identical acquisition speeds. Total numbers of intermediates were determined after gating in FlowJo and corresponding ratios (UNC0628 sample/DMSO sample) calculated. For the assessment of *Dlk1-Dio3* imprint stability, 3c iPSCs were obtained in either absence of presence of UNC0638 from reprogrammable MEFs carrying a dual fluorescent reporter system for quantification of allele-specific expression of the *Dlk1* gene, which normally is only expressed from the paternal allele (Swanzey and Stadtfeld, 2016). Established iPSCs cells were differentiated at early passage to trigger Dlk1 expression and the percentage of Dlk1-expressing cells with an (abnormal active) maternal allele quantified. Statistical analysis of flow cytometry data was conducted in Prism, using at least 500 cells of each population of interest for each biological repeat.

Western blot analysis

Reprogrammable MEFs were grown on gelatinized plates in absence of feeders. Cells were collected 24 hours after initiation of reprogramming for each treatment group. Nuclear protein was extracted using the Thermo Scientific NE-PER kit (Catalog Number 78833). For analysis of histone methylation marks, protein extraction was carried out using the Abcam histone extraction kit (ab113476). Protein was quantified using the Bio Rad protein Assay Dye (#5000006). Equal amount of protein was loaded for each set of experiment (range 5 to 15 µg) with NuPAGE™ LDS Sample Buffer (4X) in NuPAGE 4-12% Bis Tris protein gels. Gels were run using MES running buffer or MOPS running buffer in XCell SureLock Mini-Cell. Protein gels were blotted using the XCell II™ Blot Module and PVDF membrane, followed by straining of transfer membranes with Revert™ 700 Total Protein Stain. Fluorescent images for total protein quantification were taken immediately after staining using the Azure Biosystem C series. Membranes were transferred to a blocking buffer (5% nonfat milk in TBST) and incubated with gentle shaking for 2 h. Membranes were then probed with primary antibody and incubated overnight with gentle agitation at 4°C. After incubation blots were washed with TBST (Tris-Buffered Saline with Tween-20) and probed with respective secondary antibodies in blocking buffer. Finally, blots were washed with TBST and incubated in Clarity western ECL substrate chemiluminescent detection reagent (Bio-Rad #1705062). The following antibodies were used:
EHMT1(Ab41969) 1:1000, EHMT2(C6H3) 1:800, H3K9me2(Ab8896) 1:800, H3K9me2(Ab1220) 1:800, H3K9me3(Ab8898) 1:800, H3(Ab1791) 1:1000, KDM3A(12835-1-AP) 1:500, KDM3B(19915-1-AP) 1:500. For densitometric analysis, chemiluminescent blots were imaged with Azure Biosystem. AzureSpot software was used to measure individual band and total protein lane intensities. For each case background was subtracted for quantification. All blots were normalized over total protein or H3, as indicated in the figure legend. All data was analyzed using GraphPad Prism software. One-way ANOVA with Tukey's positive test or student's t test were used for statistical analyses.

RNA-seq library preparations and analysis

Total RNA was extracted from either 2 million reprogrammable MEFs or established iPSCs grown on gelatinized plates in absence of feeders with the RNeasy Plus Kit (Qiagen) and samples with

RIN values > 8 were subjected to Automated TruSeq stranded total RNA with RiboZero Gold library preparation (Illumina). Single-end 50 bp reads were generated with HiSeq2500. RNA-seq raw sequencing data were aligned to mouse genome version mm10 with the tophat algorithm (version 2.1.0)(Kim et al., 2013) and the use of «--b2-very-sensitive» parameter. Samtools (version 1.8) (Li et al., 2009) was used for data filtering and file format conversion. Aligned reads were assigned to exons with the use of the HT-seq count (version 0.5.4p3) algorithm (Anders et al., 2015) and the following parameters «-m intersection -nonempty». Differentially expressed genes were identified with the use of DESeq R package (R.3.4.4) (Anders and Huber, 2010), excluding genes with RPKM<1. R was used for PCA analysis of all genes whose expression was above 1 RPKM in at least one condition and for k-mean clustering on Z-transformed normalized expression levels of DEGs between MEFs and early reprogramming intermediates with default settings and k=5. Gene ontology analysis was conducted using Gorilla (Eden et al., 2009) and REVIGO (Supek et al., 2011). Raw sequencing data are submitted in GEO under accession number GSE130490.

ChIP-seq

MEFs were reprogrammed in basal or 3c conditions with and without 1μM UNC0638 for 48 hours on gelatinized plates in absence of feeders. Native ChIP-seq was performed as previously described (Chen et al., 2018). 10 million cells were trypsinized, washed in ice-cold PBS, and flash frozen. Cell pellets were resuspended in Dounce Buffer (10mM Tris-HCl pH 7.5, 4mM $MgCl_2$, 1mM $CaCl_2$, protease inhibitors) with a 1:1 ratio of *Drosophila* Kc cells (for spike-in controls) and homogenized using 20 repetitions of a 25G syringe. Cells were treated with 150U/ml micrococcal nuclease (Worthington Biochemical) for 20 minutes at 37˚C and reactions were quenched with 0.5M EDTA and incubated on ice for 5 minutes. MNase-digested cells were incubated on ice for 1 hour with vortexing every 10 minutes with 1ml Hypotonic Lysis Buffer (0.2mM EDTA, 0.1mM benzamidine, 0.1mM PMSF, 1.5mM DTT, PIC) followed by a 3000g spin at 4˚C for 5 minutes. Supernatants were precleared by rotation for 2 hours at 4˚C with Protein G Dynabeads (Life Technologies) that had been washed 3x in IP Buffer (10mM Tris-HCl pH 8.0, 1% Triton X-100, 0.1% Deoxycholate, 0.1% SDS, 90mM NaCl, 2mM EDTA, PIC). Precleared chromatin was combined in 2 million cell equivalents with 5ul H3K9me2 antibody (Abcam, ab1220) and 20ul washed Protein G Dynabeads in IP Buffer and rotated at 4˚C overnight. Beads were then washed twice with ChIP Wash Buffer (20mM Tris-HCl pH 8.0, 0.1% SDS, 1% Triton X-100, 2mM EDTA, 150mM NaCl, PIC), once with Final ChIP Wash Buffer (20mM Tris-HCl pH 8.0, 0.1% SDS, 1% Triton X-100, 2mM EDTA, 500mM NaCl, PIC), and eluted twice 100ul Elution Buffer (100mM $NaHCO_3$, 1% SDS) and RNase A at 68˚C. Input chromatin for each sample was diluted to 200ul in Elution Buffer and DNA from all samples and inputs was purified using the QIAquick PCR Purification Kit (Qiagen). Libraries were prepared using the KAPA HyperPrep Kit (Roche) and sequenced (paired-end 50) on an Illumina HiSeq 4000.

ChIP-seq analysis

Raw sequenced reads were aligned to both mouse (assembly mm10) and drosophila genome (assembly dm6) with Bowtie 2 (Langmead et al., 2009)(version 2.2.6) and --very-sensitive option, while filtering of poor quality and multi-mapped aligned reads was performed with the use of samtools (version 1.8). Picard tools (version 2.12.2) and MarkDuplicates option was used to remove duplicates, while enrichment of ChIP signal of the uniquely mapped reads was performed with the use of bedtools(Quinlan and Hall, 2010) coverageBed command (version 2.26.0) with --mean option. All reads from chrM and reads marked by blacklist regions (Amemiya et al., 2019)

were removed. Enrichment of mouse H3K9me2 signal was normalized to the number of drosophila aligned and filtered reads (per million). Representation of the enrichment signal with bigwig files was performed with genomeCoverageBed and bedGraphToBigWig(Kent et al., 2010). Genome wide or region specific (gene body, promoter) comparison of the median H3K9me2 normalized ChIP signal was performed with Wilcoxon rank sum test. PCA analysis of the genome wide distribution of H3K9me2 was performed with R. Genome was split into 2.5 kb bins and bins with no reads in all the experiments were discarded from the PCA analysis

SUPPLEMENTAL REFERENCES

Amemiya, H.M., Kundaje, A., and Boyle, A.P. (2019). The ENCODE Blacklist: Identification of Problematic Regions of the Genome. Sci Rep *9*, 9354.

Amlani, B., Liu, Y., Chen, T., Ee, L.S., Lopez, P., Heguy, A., Apostolou, E., Kim, S.Y., and Stadtfeld, M. (2018). Nascent Induced Pluripotent Stem Cells Efficiently Generate Entirely iPSC-Derived Mice while Expressing Differentiation-Associated Genes. Cell Rep *22*, 876-884.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol *11*, R106.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166-169.

Borkent, M., Bennett, B.D., Lackford, B., Bar-Nur, O., Brumbaugh, J., Wang, L., Du, Y., Fargo, D.C., Apostolou, E., Cheloufi, S*., et al.* (2016). A Serial shRNA Screen for Roadblocks to Reprogramming Identifies the Protein Modifier SUMO2. Stem Cell Reports *6*, 704-716.

Chen, C.C.L., Goyal, P., Karimi, M.M., Abildgaard, M.H., Kimura, H., and Lorincz, M.C. (2018). H3S10ph broadly marks early-replicating domains in interphase ESCs and shows reciprocal antagonism with H3K9me2. Genome Res *28*, 37-51.

Eden, E., Navon, R., Steinfeld, I., Lipson, D., and Yakhini, Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics *10*, 48.

Kent, W.J., Zweig, A.S., Barber, G., Hinrichs, A.S., and Karolchik, D. (2010). BigWig and BigBed: enabling browsing of large distributed datasets. Bioinformatics *26*, 2204-2207.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol *14*, R36. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.

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

Maherali, N., Ahfeldt, T., Rigamonti, A., Utikal, J., Cowan, C., and Hochedlinger, K. (2008). A high-efficiency system for the generation and study of human induced pluripotent stem cells. Cell Stem Cell *3*, 340-345.

Pelossof, R., Fairchild, L., Huang, C.H., Widmer, C., Sreedharan, V.T., Sinha, N., Lai, D.Y., Guan, Y., Premsrirut, P.K., Tschaharganeh, D.F., *et al.* (2017). Prediction of potent shRNAs with a sequential classification algorithm. Nat Biotechnol *35*, 350-353.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842.

Sommer, C.A., Stadtfeld, M., Murphy, G.J., Hochedlinger, K., Kotton, D.N., and Mostoslavsky, G. (2009). Induced pluripotent stem cell generation using a single lentiviral stem cell cassette. Stem Cells *27*, 543-549.

Stadtfeld, M., Apostolou, E., Ferrari, F., Choi, J., Walsh, R.M., Chen, T., Ooi, S.S., Kim, S.Y., Bestor, T.H., Shioda, T., *et al.* (2012). Ascorbic acid prevents loss of Dlk1-Dio3 imprinting and facilitates generation of all-iPS cell mice from terminally differentiated B cells. Nat Genet *44*, 398-405, S391-392.

Stadtfeld, M., Maherali, N., Borkent, M., and Hochedlinger, K. (2010). A reprogrammable mouse strain from gene-targeted embryonic stem cells. Nat Methods *7*, 53-55.

Supek, F., Bosnjak, M., Skunca, N., and Smuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS One *6*, e21800.

Swanzey, E., and Stadtfeld, M. (2016). A reporter model to visualize imprinting stability at the Dlk1 locus during mouse development and in pluripotent cells. Development *143*, 4161-4166.

Vidal, S.E., Amlani, B., Chen, T., Tsirigos, A., and Stadtfeld, M. (2014). Combinatorial modulation of signaling pathways reveals cell-type-specific requirements for highly efficient and synchronous iPSC reprogramming. Stem Cell Reports *3*, 574-584.