Stem Cell Reports, Volume *17*

Supplemental Information

DOT1L inhibition enhances pluripotency beyond acquisition of epithelial identity and without immediate suppression of the somatic transcriptome

Coral K. Wille and Rupa Sridharan

Figure S1 related to Figure 1

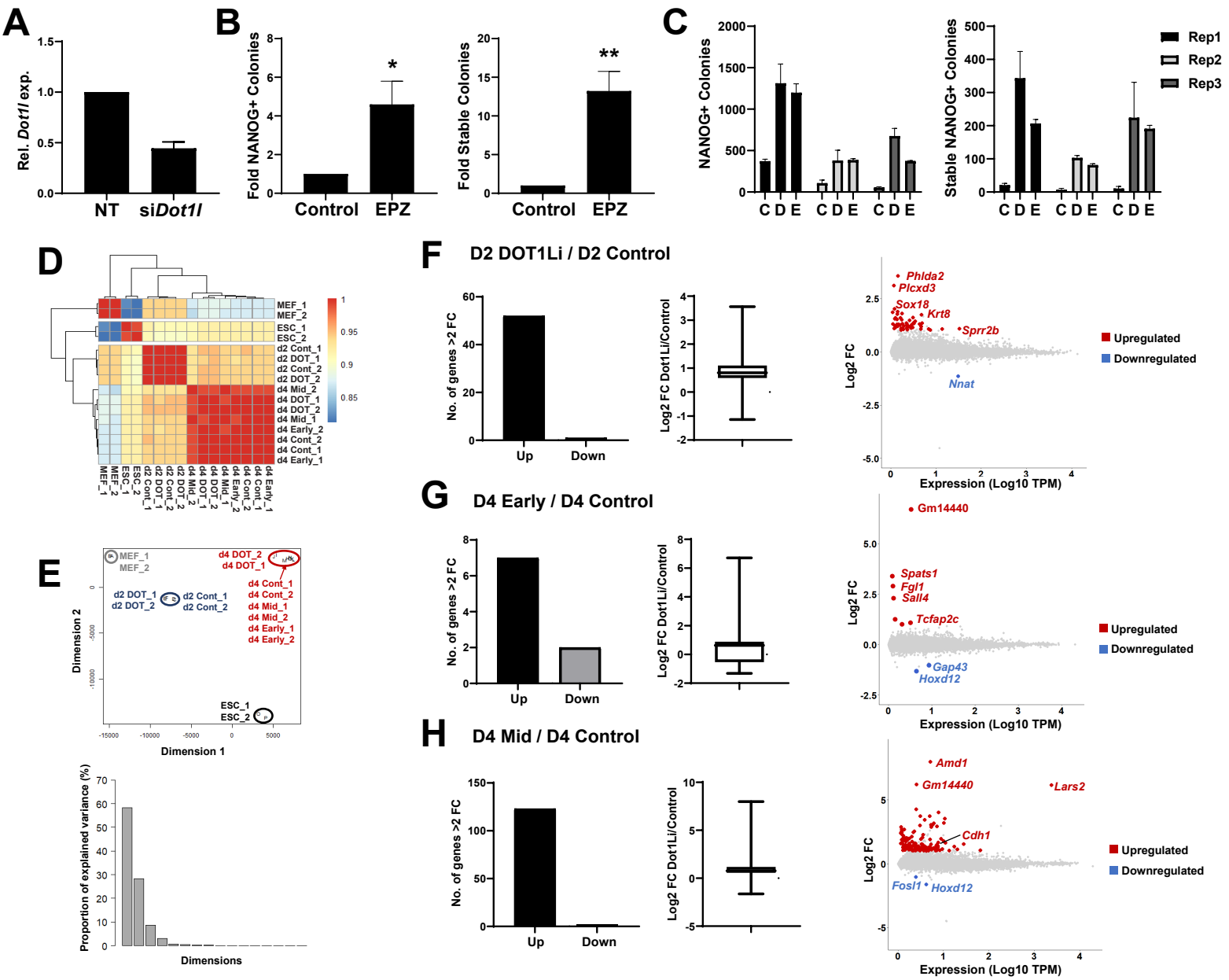


Figure S1. Loss of H3K79me during reprogramming results in few steady-state transcriptional changes.

A. Relative *Dot1l* expression measured on day 2 of reprogramming. Nontargeting (NT) siRNA treated cells set to 1.

B. Reprogramming efficiency (Left) and stable NANOG+ colonies post dox withdrawal (Right) of MEFs treated with control and EPZ5676 (EPZ). Colonies obtained in control treatment set to 1. Error bars represent the SEM of three independent experiment replicates, each comprised of 2 technical replicates. **P<0.01 and *P<0.05 by unpaired t-test.

C. Average colonies obtained (Left) and stable NANOG+ colonies post dox withdrawal (Right) in each independent experiment replicate (Rep), of MEFs treated with control (C), DOT1Li (D), and EZP5676 (E). Error bars represent the SEM of 2 technical replicates.

D. Pearson correlation of all timecourse RNA-Seq independent experiment replicates. Notations: Day (d), control treatment (Cont), and DOT1Li (DOT).

E. Top: PCA of all timecourse RNA-Seq independent experiment replicates. MEFs notated in gray, reprogramming day 2 in blue, reprogramming day 4 in red, and ESCs in black. Bottom: Percent of variance explained by each component.

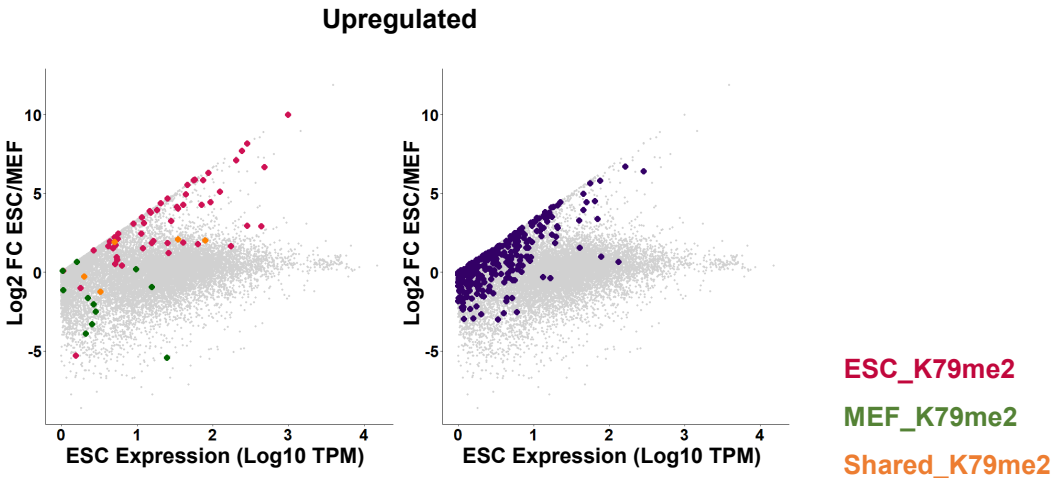
F-H. Left: Number of genes upregulated or downregulated more than 2-fold change (FC) with a posterior probability of differential expression greater than 0.95 determined by EBSeq. Middle: Box plot of Log2 FC of all genes with a posterior probability of differential expression greater than 0.95. Right: Expression measured as Log10 of the averaged transcripts per million (TPM) of the two samples versus Log2 FC of all genes. More than 2-fold upregulated indicated in red and downregulated indicated in blue. F. Day 2 DOT1Li versus day 2 Control, G. Day 4 Early (days 0-2) DOT1Li versus day 4 Control, H. Day 4 Mid (days 2-4) DOT1Li versus day 4 Control.

Figure S2 related to Figure 2

A

GO Term	Significance	Number	Examples
Cellular macromolecule metabolic process	6.4E-91	3900	*Arl2bp*, *Ddx49*
Nucleic acid metabolic process	4.5E-73	2605	*Polr2e*, *Ets1*
RNA binding	1.8E-62	581	*Oas1h*, *Esf1*
Gene expression	2.5E-57	1315	*Eif4a1*, *Med10*

B



C

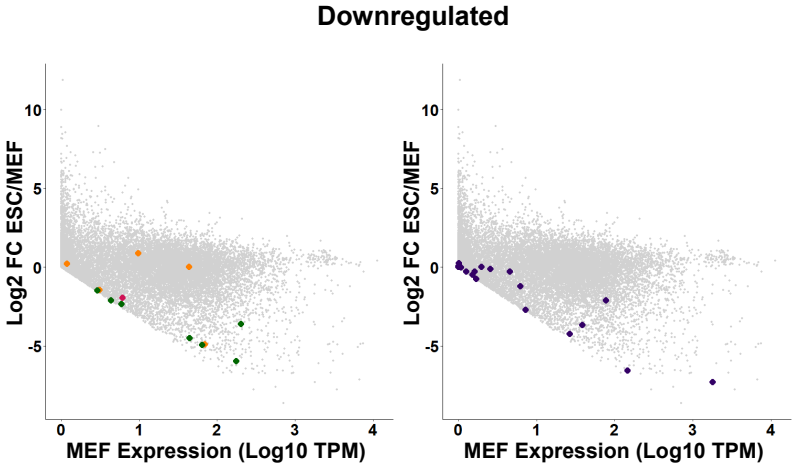


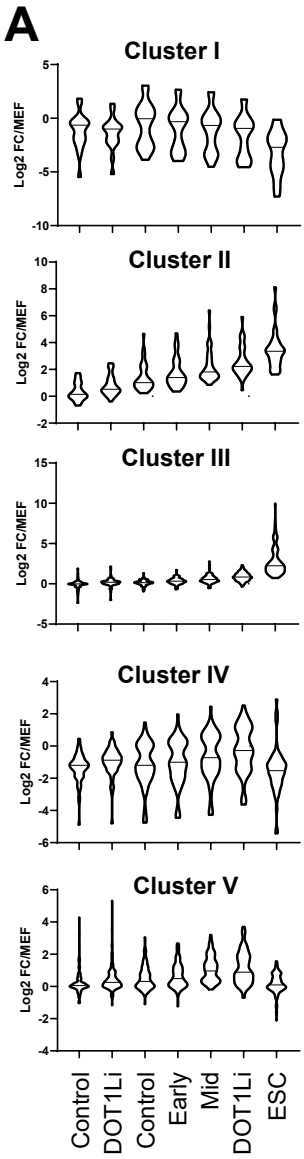
Figure S2. H3K79me2 is enriched on numerous genes yet few change transcriptionally in steady state mRNA levels.

A. Most significant gene ontology (GO) terms of genes that have an H3K79me2 peak shared in both MEFs and ESCs (Fig 2B, orange).

B. H3K79me2 peak status of DOT1Li upregulated genes designated by color, plotted on ESC expression measured as Log10 TPM versus Log2 fold change (FC) in ESCs relative to MEFs of all genes.

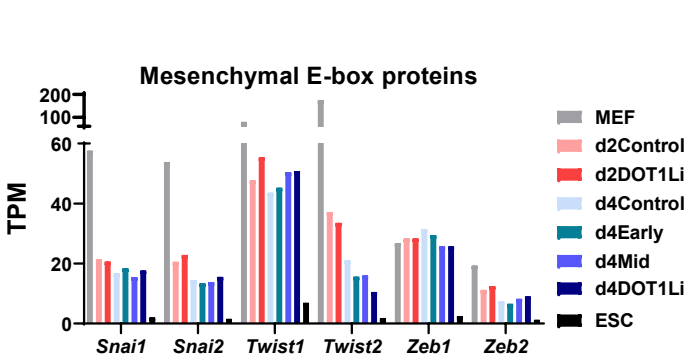
C. H3K79me2 peak status of DOT1Li downregulated genes designated by color, plotted on MEF expression measured as Log10 TPM versus Log2 fold change (FC) in ESCs relative to MEFs of all genes.

Figure S3 related to Figure 3



B

Motif	Cluster	p-value	Percent of targets	Function
E-Box (SNAI2) (ZEB1)	II III	1E-02 1E-10	50.0% 39.8%	Repressor
TCF7	IV	1E-7	16.7%	Activator
HOXD12	V	1E-05	2.1%	Activator
HIC1	II/III V	1E-05 1E-02	4.0% 77.8%	Repressor



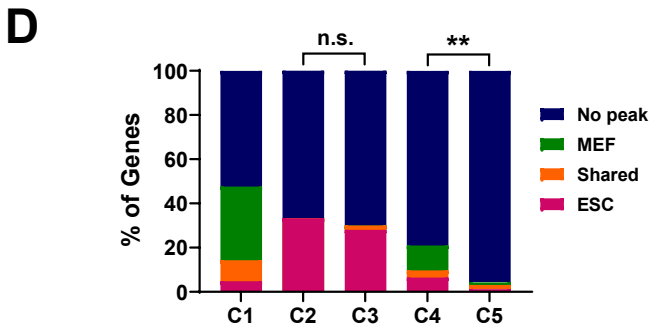


Figure S3. DOT1L inhibition leads to transcriptional changes not observed in MEFs or ESCs.

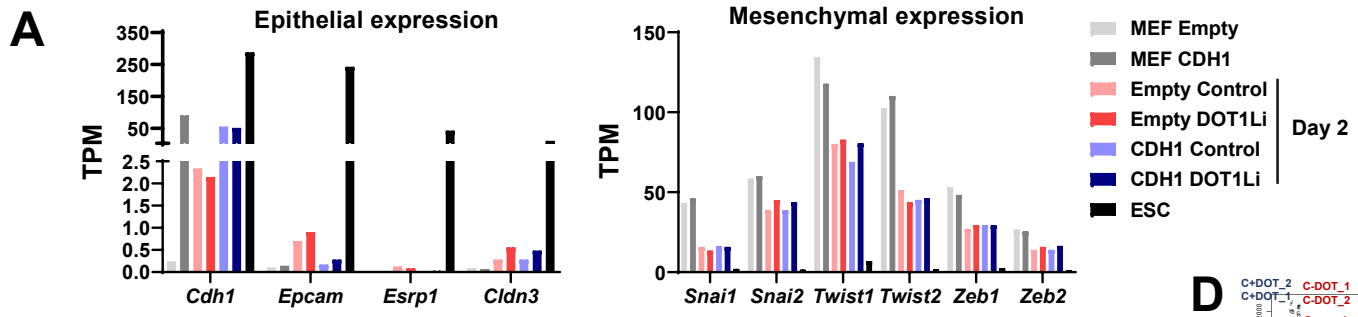
A. Violin plot of Log2 fold change (FC) TPM relative to MEFs for each cluster.

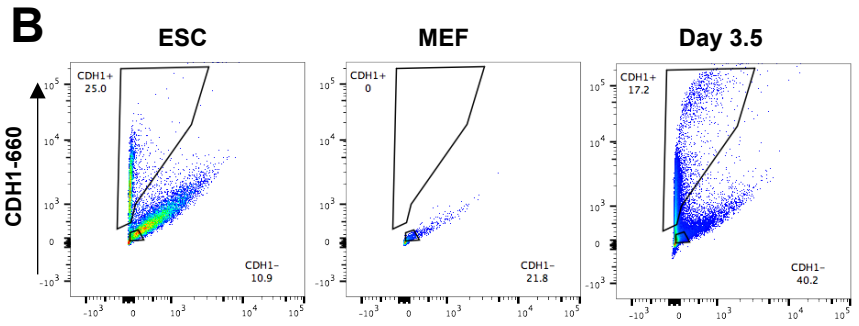
B. Table of motifs in each cluster identified with HOMER. The p-value, percent of targets within the cluster, and function of the binding protein for motifs bound by DOT1Li-DE genes are displayed.

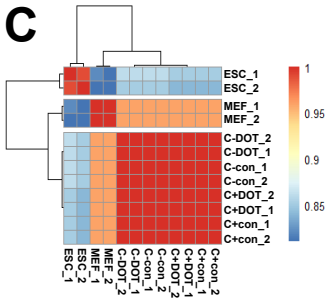
C. Expression (TPM) bar graph of mesenchymal E-box binding proteins.

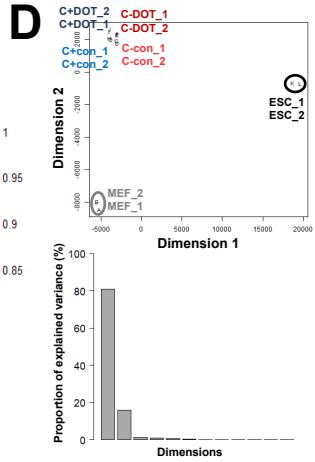
D. Bar graph of the percentage of genes with an H3K79me2 called peak in each cluster. Significance determined by Fisher's exact test. All pairwise comparison are ****P<0.0001, except those noted as **P<0.01 or not significant (n.s.) P>0.05.

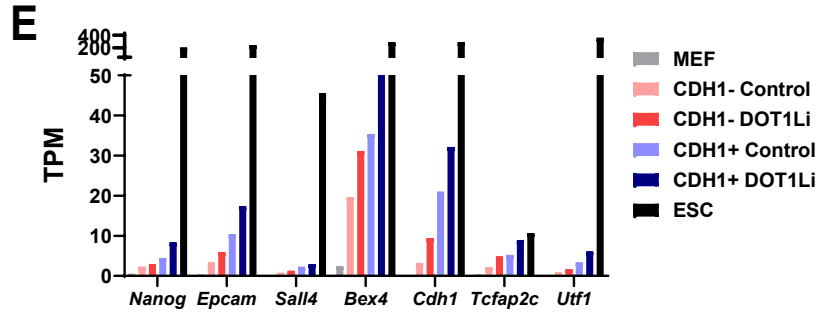
Figure S4 related to Figure 4

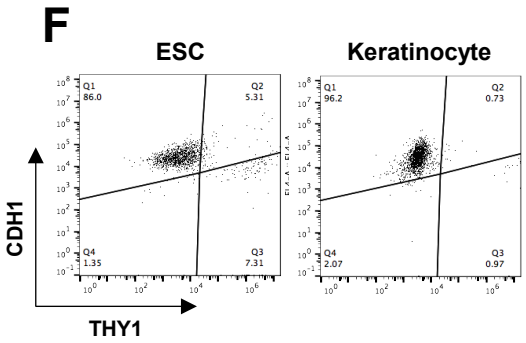


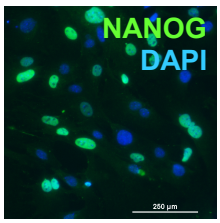


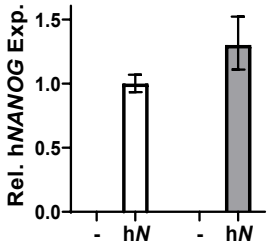


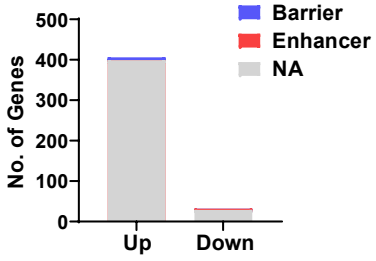




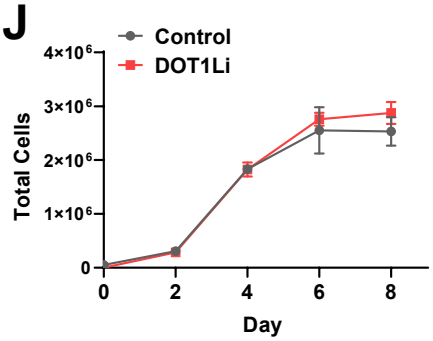


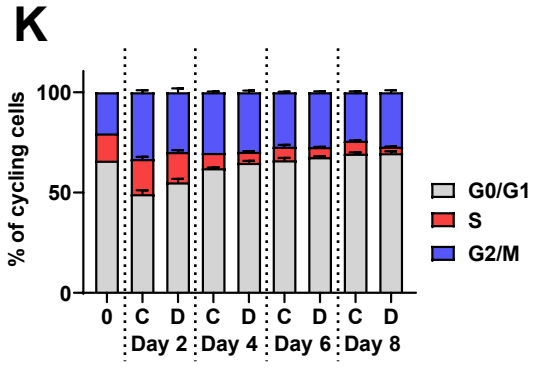






Borkent et al., Stem Cell Reports, 2016





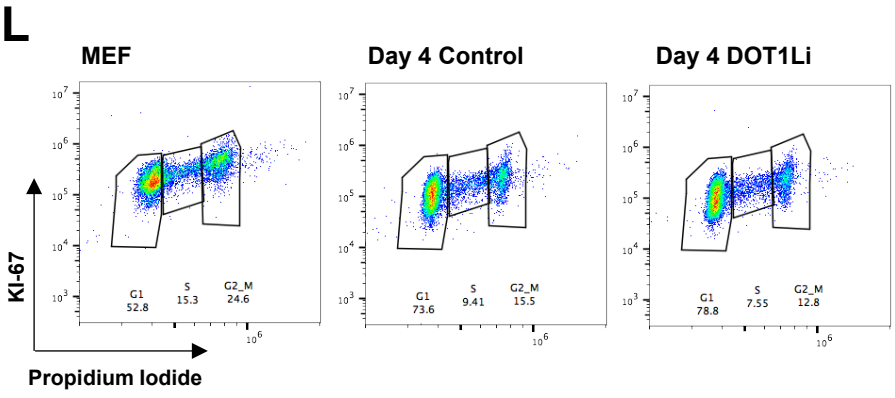


Figure S4. Inhibition of DOT1L enhances reprogramming of epithelial cells.

A. Expression (TPM) bar graphs of MEFs and day 2 reprogramming cells transduced with empty vector control or *Cdh1*, treated with and without DOT1Li.

B. Flow cytometry sorting of CDH1 positive and negative cells for Fig 4C with ESC and MEF controls. Gates indicate collected cells.

C. Pearson correlation of all CDH1 sort RNA-Seq independent experiment replicates. Notations: CDH1- (C-), CDH1+ (C+), control treatment (con), and DOT1Li (DOT).

D. Top: PCA of all CDH1 sort RNA-Seq independent experiment replicates. MEFs notated in gray, CDH1- (C-) reprogramming in red, CDH1+ (C+) reprogramming in blue, and ESCs in black. Bottom: Percent of variance explained by each component.

E. Expression (TPM) bar graph of representative upregulated genes in CDH1- and CDH1+ that overlapped with genes upregulated in ESCs relative to MEFs (Fig 4E).

F. Flow cytometry analysis of THY1 and CDH1 on ESCs and keratinocytes.

G. Immunofluorescence of human NANOG transduced in MEFs. Scale bar = 250 μM.

H. Relative human *NANOG* expression measured on day 4 of reprogramming in cells treated with control (white bars) or DOT1Li (gray bars). Control treated cells transduced with h*NANOG* set to one.

I. Overlap of DOT1Li-DE genes and Borket et al., 2016 screen. Genes chosen for overlap affected reprogramming more than 2-fold. “Barriers” indicate genes targeted by shRNAs enriched in reprogrammed cells and “enhancers” are genes targeted by shRNAs depleted in reprogrammed cells.

J. Cells treated with control (gray) or DOT1Li (red) were counted every two days during reprogramming. Error bars represent the SD of three technical replicates.

K. Quantification of cell cycle analysis. Cells on day 0 (0), and cells treated with DOT1Li (D) or control (C) were assessed every two days during reprogramming.

L. Representative propidium iodide and KI-67 flow cytometry cell cycle analysis (Fig S4K).

Figure S5 related to Figure 5

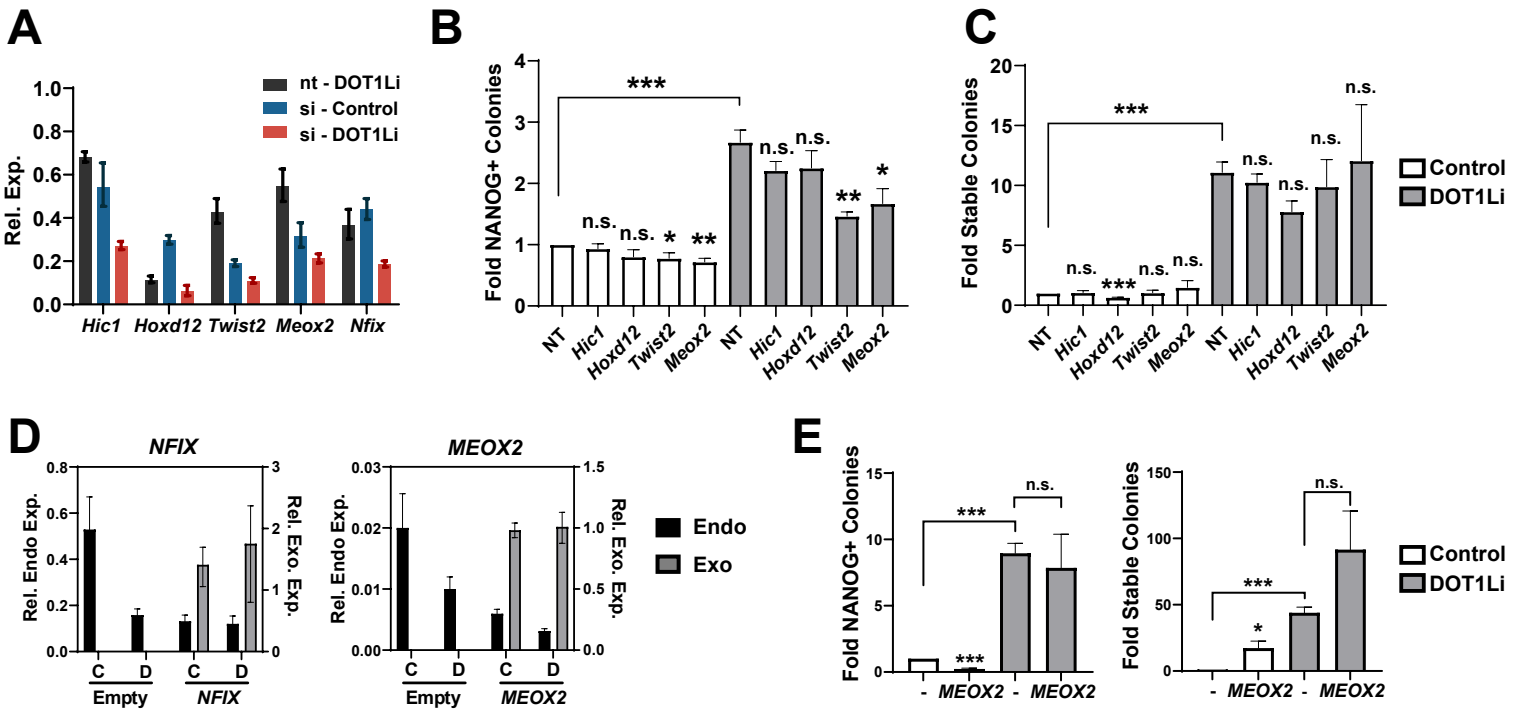


Figure S5. Inhibition of DOT1L enhances reprogramming beyond modulation of single genes.

A. Relative expression of si-depleted genes on day 4 of reprogramming. Cells transduced with nontargeting (NT) siRNA treated with control set to 1.

B. NANOG+ colonies on days 6-7 of reprogramming of cells transfected with nontargeting (NT) and siRNA against the indicated DOT1L direct target gene, treated with control (white bars) or DOT1Li (gray bars). Control treated NT set to 1. Error bars represent the SEM of 3-4 independent experiment replicates, each consisting of 2-3 technical replicates. ***P<0.001, **P<0.01, *P<0.05, and not significant (n.s.) P>0.05 by unpaired t test.

C. Transgene independent stable colonies post dox removal of cells transfected with nontargeting (NT) and siRNA against the indicated DOT1L direct target gene, treated with control (white bars) or DOT1Li (gray bars). Control treated NT set to 1. Error bars represent the SEM of 3-4 independent experiment replicates, each consisting of 2-3 technical replicates. ***P<0.001 and not significant (n.s.) P>0.05 by unpaired t test.

D. Relative expression of endogenous (Endo) in black and exogenous (Exo) in gray of *NFIX* (Left) and *MEOX2* (Right) on day 4 of reprogramming. Cells were treated with Control (C) or DOT1Li (D).

E. NANOG+ colonies at day 6 of reprogramming (Left) and stable colonies post dox removal (Right) of cells transduced with empty vector control (-) or *MEOX2*, treated with control (white bars) or DOT1Li (gray bars). Error bars represent the SEM of 3 independent experiment replicates, each consisting of 2-3 technical replicates. ***P<0.001, *P<0.05, and not significant (n.s.) P>0.05 by unpaired t test.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Reprogramming experiments

MEFs were seeded at a density of 30,000 to 50,000 cells/12-well or 20,000 cells/24-well onto gelatinized coverslips. Keratinocytes were seeded at a density of 10,000 cells/24 well onto gelatinized coverslips. Feeder MEFs were added at 1/2x confluency and reprogramming was initiated with 2 µg/ml of doxycycline (dox) with either vehicle (DMSO) control, 5 µM SGC0946/DOT1Li (ApexBio A4167), or 3 µM EPZ5676 (MedChem Express HY-15593). ESC media made with FBS or knock-out serum replacement (KSR), as indicated, with fresh dox and chemicals was replaced every two days. Keratinocytes were maintained in keratinocyte media two days post-OSKM induction before changing to ESC media to avoid FBS-induced differentiation. In the case of exogenous gene expression, MEFs were transduced with lentivirus and selected (if possible) before seeding.

To assess transgene independence, reprogramming cells were washed once with ESC media, and ESC media free of doxycycline and drugs was replaced in wells. Sustained NANOG expression was measured 2-4 days post doxycycline removal by immunofluorescence. Experiments were timed for individual MEFs so that cells were exposed sufficiently to OSKM to produce *bona fide* colonies, but not so long that NANOG positive colonies were overcrowded and uncountable in the presence of doxycycline.

Reprogramming statistical analysis

A colony was considered a grouping of 4 or more NANOG+ cells. Each independent experiment replicate was derived from an independent MEF isolation, and was the average of 2-3 technical replicates. Technical replicates consisted of coverslips in individual wells of a single experimental replicate. Fold was calculated relative to the control condition unless otherwise indicated, and error bars depict the standard error of the mean (SEM) of independent experiment replicates. Experimental replicate (n) information is listed in each figure legend. Significance was calculated as specified in legends using Graphpad Prism 9 for figures aggregated from three or more independent experiment replicates. To compare two conditions, significance was calculated using the unpaired two-tailed t-test function. Comparison among 3 samples was performed with one-way ANOVA repeated measure, matched based independent experiment replicate. The mean of every sample group was compared post-hoc using the Tukey test to calculate significance for multiple comparisons.

Lentiviral vectors, packaging, and transduction

NANOG (Addgene 16578), *MEOX2* (Addgene 116761), and *Cdh1* (Addgene 71366) were acquired from Addgene. The CDS of *Cdh1* was amplified from Addgene 71366 with 5'-TGTTTCGAAATGGGAGCCCGGTGCCGC-3' and 5'-TGTGCGGCCGCTTAGTCGTCCTCGCCACCGCC-3', and moved into the BstBI and NotI sites of pCDH-CMV-MCS-EF1α-Neo. The CDS of human *NFIX* (NM_002501.4) was cloned in to the EcoRI site of Tet-O-FUW lentiviral vector, and was a kind from Dr. Jason Tchieu (Cincinnati Children's Hospital). Lentiviral transfer vectors were transfected into 293T cells with packaging vector pspax2 (Addgene 12260) and envelop vector vsvg using linear polyethylenimine. Media was changed to MEF media with 20 mM HEPES 4 hours post transfection. Virus-containing media was harvested at 48h and 72h, combined, and filtered through a 0.45 µm PVDF filter. Virus containing media was combined with fresh media at a ratio of 1:1 and 10 µg/ml Hexadimethrine Bromide (polybrene) to transduce target cells.

RNA isolation and library preparation

Reprogramming was initiated in two independent experiment replicate samples, and RNA was isolated from each using TRIzol at the indicated timepoint. One-fifth the volume of chloroform was added and phases were separated by max centrifugation. The upper layer was isolated, RNA was precipitated with 0.53 volumes of ethanol, and applied to a RNeasy column (Qiagen 74104). RNA was washed with 500 µl RW1 and DNA was digested on the column with DNase (Qiagen 79254) for 30 min. RNA was washed according to the RNeasy protocol and eluted in 30 ul of H_2O. RNA was quantitated and 1 µg of each sample was combined with 20 ng of RNA from 293T (human) cells, prepared as above, as a spike-in control for sequencing normalization (timecourse and CDH1 sort studies). The cDNA library was constructed using TruSeq RNA Sample Preparation kit V2 (RS-122-2002) according to the manufacturer's instructions. Libraries were assessed with Qubit and Bioanalyzer3.0.

RNA-Seq computational and statistical analysis

Greater than 40 million reads of timecourse DOT1Li libraries (Fig 1) and CDH1 sort libraries (Fig 4) were sequenced PE150 by Novogene on an Illumina HiSeq 4000. Greater than 25 million reads of CDH1 expression libraries (Fig S4) were sequenced SE100 by the University of Wisconsin Biotechnology Center on a HiSeq 2500. Sequencing quality was assessed with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Sequenced reads were processed using Trimmomatic (Bolger et al., 2014) with the following parameters: LEADING:3 TRAILING:3 CROP:100 HEADCROP:10 SLIDINGWINDOW:4:15 MINLEN:36. Reads were aligned to the mm9 and hg18 genome using RSEM-1.2.4 (Li and Dewey, 2011) with a mismatch per seed set to 2 and seed length set to 28 (-bowtie-m 200 --bowtie-n 2 --forward-prob 0.5 --seed-length 28 --paired-end). RSEM-1.2.4 alignment yielded transcripts per million (TPM) for each gene. Matrices of TPM of all genes were generated in R and used to assess RNA-Seq replicates. Pearson correlation of replicate samples was performed in R and plotted with the package pheatmap. Classical multidimensional scaling was performed to determine the proportion of variance explained by dimensions with the following parameters:
mds <- cmdscale(dist(PCA_data), k=3, eig=TRUE), mds$eig,
eig_pc <- mds$eig * 100 / sum(mds$eig).
Two dimensions were plotted as over 80% of the data was explained for all analyses using: mds <- cmdscale(dist(PCA_data)). Replicate RNA-Seq sample TPM were averaged for all main figures since they clustered together in both statistical tests.

A matrix of unnormalized reads per replicate mapping to the mouse or human genome was generated with R. Differentially expressed (DE) genes were called with EBSeq (Leng et al., 2013) using the human matrix for normalization (MedianNorm(human_matrix)). Differentially expressed genes were filtered to have a posterior probability DE greater than 0.95 and a 2 or greater posterior fold change (PostFC). Expression changes reported are PostFC values determined by EBSeq unless otherwise noted to be TPM. Cluster3.0 was used to perform k-means clustering of DE genes (de Hoon et al., 2004). Clusters were visualized as Log2 fold change relative to MEF TPM with Java TreeView (Saldanha, 2004). Motif discovery was performed using HOMER Motif Analysis (Heinz et al., 2010) with the following parameters: findMotifs.pl -start -1000 -end 100 -len 6,10. Gene Ontology was performed using HOMER and DAVID (Huang et al., 2009). DAVID parameters included: GOTERM_BP_4 and GOTERM_MF_4, viewed by functional annotation clustering. GO Terms within functional annotation clusters with an enrichment score more than 2 were considered significant. DE genes were compared to a published shRNA iterative screen dataset (Borkent et al., 2016). Fisher's exact test was performed in R.

DOT1L target gene scRNA-seq analysis

Expression of potential DOT1L targets was assessed with single cell RNA-seq data (GEO: GSE108222). Briefly, cells expressing the gene of interest were displayed on a t-distributed Stochastic Neighbor Embedding (t-SNE) cluster plot of MEFs, ESCs, and reprogramming cells in FBS constructed using Monocle2 v2.6.3 on R version 3.4.3 with cells that passed quality control as previously described (Tran et al., 2019).

ChIP-Seq analysis

H3K79me2 ChIP-Seq data with the accession number GSE90895 (Chronis et al., 2017) was downloaded from Gene Expression Omnibus and aligned to mm9 using Bowtie2 (Langmead and Salzberg, 2012) with the default parameters. Sam files were converted into Bam files and sorted with samtools-1.2 (Li et al., 2009) with the default parameters. Peaks were called relative to the input with MACS2 (Zhang et al., 2008) using the following parameters: --broad -p 0.0001. Peak files were annotated using EASeq (Lerdrup et al., 2016) from the center of the peak to the nearest gene center. Genes with peaks were visually mapped back on to the Log10 TPM versus Log2 posterior fold change using R ggplot2 (geom_point graph). Overlaps of DE genes with H3K79me2-modified genes performed with Venny2.1 (https://bioinfogp.cnb.csic.es/tools/venny/). ChIP enrichment was visualized using Integrative Genomics Viewer (IGV) (Robinson et al., 2011).

Immunofluorescence

Coverslips were fixed in 4% paraformaldehyde-PBS, permeabilized in 0.5% Trition-X-PBS, and washed in 0.2% Tween-20-PBS. Coverslips were blocked for 30 min in blocking buffer (1x PBS, 5% donkey serum, 0.2% Tween-20, and 0.2% fish skin gelatin). Cells were stained for 1 hr with primary

antibody in blocking buffer, rinsed 2x in wash buffer, and stained for 1 hr with secondary (1:1000) in blocking buffer. Coverslips were rinsed with wash buffer, stained with DAPI (0.1 µg/ml) in wash buffer, and rinsed with wash buffer. The following antibodies were used for immunofluorescence: anti-murine NANOG (Cosmo Bio RCAB0002P-F, 1:100 and Cell Signaling Technology 8822S, 1:1000), anti-human NANOG (R&D Systems AF1997, 1:100), anti-CDH1 (eBiosciences 14-3249-80, 1:100), and anti-DPPA4 (Thermofisher PA5-47530, 1:100). Colony counts and imaging were performed on Nikon Eclipse Ti using NIS Elements software.

Immunoblot

Whole cells were lysed in RIPA buffer (150 mM NaCl, 1% NP-40, 0.5% Na deoxycholate, 0.1% SDS, 25 mM Tris pH 7.4) with 1x protease inhibitor (Roche 04693116001), sonicated for 5 secs at 20% amplitude, and quantified by cell count or with the DC Protein Assay (BioRad 5000112) according to the manufacturer's instructions. Equal amounts of protein were loaded onto an SDS-Page gel and transferred to a nitrocellulose membrane. Membranes were blocked in blocking buffer (5% milk, 0.1% Tween-20, 1x PBS) followed by incubation with primary antibody in blocking buffer. Membranes were washed 0.1% Tween-20-PBS and incubated with secondary antibody in blocking buffer. Membranes were washed and visualized with ECL reagent. Images were quantified using Image Studio Lite software. Primary antibodies included: anti-H3K79me2 (Abcam ab3594, 1:1000) and anti-α-TUBULIN (Cell Signaling 3873, 1:5000).

Flow cytometry and sorting

For cell cycle determination, cells were fixed in 1 volume of PBS with 9 volume of cold 70% ethanol and permeabilized by freezing at -20 for at least 2 hours. Cells were washed with FACS buffer (1xPBS, 2% FBS, 1mM EDTA), and stained for 30 min in 100 µL FACS buffer containing 1 µL anti-KI-67-Alexa Fluor 488 (BioLegend 151204) per million cells, as previously described (Kim and Sederstrom, 2015). Cells were washed with FACS buffer and DNA was stained for 20 min with 50 µg/ml of propidium iodide (in 1xPBS and 2 mM $MgCl_2$) supplemented with 100 µg/ml RNaseA.

Live cells were used for surface marker staining. Cells were washed in 1x PBS and incubated with antibodies in 1xPBS with 1% FBS for 1 hour. Sorting was performed on a BD FACS AriaII (UW Carbone Cancer Center, Grant #: 1S10RR025483-01) with appropriate controls for gating. FACS quantitation was performed on a BD Accuri C6 Flow Cytometer. Primary antibodies used were anti-THY1-PE (eBioscience 12-0903-81, 1:100) and anti-CDH1-eFluor660 (eBioscience 50-3249-80, 1:100).

siRNA transfection

MEFs were plated at a confluency of 20,000 cells per 24 well on coverslips. Cells were transfected 24 hours after plating with 20 nM siRNA using DharmaFECT1 (Dharmacon T200104) according to the manufacturer's instructions. Reprogramming was initiated immediately using 2 µg/ml of doxycycline with either vehicle (DMSO) control, 5 µM SGC0946/DOT1Li (ApexBio A4167). Cells were transfected every two days during reprogramming, and siRNA was increased to 40 nM at day 4 to account for increased cell number. Knockdown efficiencies were determined 48h after transfection. si*Dot1l* (J-057964-12), si*Hoxd12* (M-046274-01), si*Nfix* (MQ-045912-01), si*Fosl1*(MQ-040704-01) were purchased from Dharmacon Horizon. si*Hic1* (mm.Ri.Hic1.13.1), si*Meox2* (mm.Ri.Meox2.13.1), si*Twist2* (mm.Ri.Twist2.13.1,) were purchased from IDT.

RT-qPCR

RNA was isolated using the Isolate II RNA Mini Kit (Bioline BIO-52072), and 1 µg was converted to cDNA using the qScript cDNA Synthesis Kit (VWR 101414). Technical replicates of 20 ng (based on RNA concentration) were used to measure Ct on a BioRad CFX96 thermocycler with iTaq UniverSYBR Green SMX (BioRad 1725125) in 10 µl reactions. Relative expression was calculated against the geometric mean of two housekeeping genes. Hierarchical clustering of relative expression values performed with Cluster3.0 (de Hoon et al., 2004) using the uncentered correlation similarity metric and the average linkage clustering method. Primers used in this study:

Gene	Forward 5'-3'	Reverse 5'-3'
Bex1	TGGTGGTGAGCATCTCTAGAAAGAG	TAGAAGCTGGTAACAGGGAG

Cd5l	AGGTGATCTGCACAGACTTTGAT	TTGAGGTGATTATGGGGGCG
Cdh1	GCCACCAGATGATGATACCC	GGAGCCACATCATTTCGAGT
Dot1l	GCGGCTGTGTGACAAATACA	CCATACACCTCAGGGGAGAA
Dsc2	AGGGCCCAGTAGAGGTACTA	GTGGGAAGGGACCCAATGAA
Egfl6	CTGCAGTGTTCTGTGATCCCT	CTTCATGGTGCGCTTGTGAG
Epcam	CATTTGCTCCAAACTGGCGT	TTGTTCTGGATCGCCCCTTC
Fabp4	CATAACCCTAGATGGCGGGG	CGCCTTTCATAACACATTCCACC
Fosl1	CGGCCAGGAGTCATACGAG	CCTTTCTTCGGTTTCTGCAC
Hic1	GAGCTTTGGTGACAACCTGTA	CTGCCATATAACGCCTCTTCTT
Hoxd12	CTCTTGCCTGCGATCTTCACT	GAATTCATTGACCAGGAATTCGTT
Klhl30	CGCCCAAGTATGTCAGCAAC	TGACACTCCACGCATCTGTC
Krt18	AATCGAGGCACTCAAGGAAGAA	GGCATCCACTTCCACAGTCA
Meox2	TGCGCACCAGGGGATTATG	TGGGAATCTGAGCTGTCGC
hMEOX2	AGACTGAGGCGATACGAGATAG	CCTTTACCCTCTTCCACTTCATC
Nanog	CATCCCGAGAACTATTCTTGCT	GAGGCAGGTCTTCAGAGGAA
hNANOG	CAGCTGTGTGTACTCAATGATAGATTT	CAACTGGCCGAAGAATAGCAATGGTGT
Nfix	CATCAGGGCCCAACTTCTCA	CTTGGTGCTGCTGGTGGAA
hNFIX	GAGTCCCAGCTACTACAACATC	GCTCTCCATCTCACTGTCATC
Sall4	AACATATGCGGGCGGGCCTTCA	CCAGGAGGCGGGGTCCACACTC
Tcfap2c	GGGCTTTTCTCTCTTGGCTGGT	TCCACACGTCACCCACACAA
Twist2	CGCTACAGCAAGAAATCGAGC	GCTGAGCTTGTCAGAGGGG
Gapdh	TTCACCACCATGGAGAAGGC	CCCTTTTGGCTCCACCCT
β-actin	TGTTACCAACTGGGACGACA	TCTCAGCTGTGGTGGTGAAG

SUPPLEMENTAL REFERENCES

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

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.

Chronis, C., Fiziev, P., Papp, B., Butz, S., Bonora, G., Sabri, S., Ernst, J., and Plath, K. (2017). Cooperative Binding of Transcription Factors Orchestrates Reprogramming. Cell *168*, 442-459.e20.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

de Hoon, M.J.L., Imoto, S., Nolan, J., and Miyano, S. (2004). Open source clustering software. Bioinformatics *20*, 1453–1454.

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

Kim, K.H., and Sederstrom, J.M. (2015). Assaying Cell Cycle Status Using Flow Cytometry. Curr Protoc Mol Biol *111*, 28.6.1-11.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357–359.

Leng, N., Dawson, J.A., Thomson, J.A., Ruotti, V., Rissman, A.I., Smits, B.M.G., Haag, J.D., Gould, M.N., Stewart, R.M., and Kendziorski, C. (2013). EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. Bioinformatics *29*, 1035–1043.

Lerdrup, M., Johansen, J.V., Agrawal-Singh, S., and Hansen, K. (2016). An interactive environment for agile analysis and visualization of ChIP-sequencing data. Nat. Struct. Mol. Biol. *23*, 349–357.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12*, 323.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

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

Saldanha, A.J. (2004). Java Treeview--extensible visualization of microarray data. Bioinformatics *20*, 3246–3248.

Tran, K.A., Pietrzak, S.J., Zaidan, N.Z., Siahpirani, A.F., McCalla, S.G., Zhou, A.S., Iyer, G., Roy, S., and Sridharan, R. (2019). Defining Reprogramming Checkpoints from Single-Cell Analyses of Induced Pluripotency. Cell Rep *27*, 1726-1741.e5.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.