Supplementary Information

Materials and Methods

Cell culture

Supplementary Table S2 lists the primary fibroblasts cultured from donors of different ages used in this study. These fibroblasts were grown in Eagles minimal essential media (EMEM) (ATCC Cat # 30-2003) supplemented with either 10% fetal bovine serum (for the growth and propagation of fetal (MRC5) and new born (BJ) fibroblasts) or with 20% fetal bovine serum (in case of older donor fibroblasts). Cells at the time of procurement were of differing passage numbers and were used for reprogramming experiments within 2-3 passages from the time they were acquired. H1 (WA01) and H9 (WA09) human ES cells were obtained from the WiCell Research Institute.

Measurement of senescence associated beta-galactisidase (SAβ-gal) and other markers in young and old fibroblasts and in iPS and ES cells

Fibroblasts of differing ages (Supplementary table S2) were characterized for the expression SAβ-gal and other markers- H3K27me3, H3K9me3 and LAP2α, that are typically associated with cells derived from older individuals (Miller et al., 2013). As controls, we used 'younger' fibroblasts (MRC5 and BJ) and established hES and iPSC lines grown under feeder-free conditions. Cells were plated at a density of about 2.5×10^4 cells per 12 well in quadruplicate. Two days later, cells were fixed and stained with the SA β-gal kit (Cell Signaling Cat # 9860) as per the manufacturer's instructions. The development of characteristic intracellular blue color was typically observed four or five days later. Upon this observation, the staining solution from each well was aspirated and overlaid with 70% glycerol. Each well was scored for SA β-gal expressing cells under 20X and 40X magnifications on a Nikon Eclipse TS100 microscope. An average from 4 wells was taken as an estimate of SAβ-gal levels in each cell type.

To quantify nuclear levels of H3K27me3, H3K9me3 and LAP2α in fibroblasts of all ages, hES and iPS cells, cells were grown in 12 well plates as stated above and fixed with 4% paraformaldehyde in PBS. Fixed cells were washed and permeabilized with 0.2% TritonX in PBS for about 10-15 minutes in an elliptical shaker. Following couple of washes with PBS, a blocking solution containing 3% goat serum and 3% BSA was added. Plates were allowed to incubate in blocking solution for about 30 minutes on the shaker followed by a single wash with PBS. Each cell type was stained separately using the antibodies specific for 'heterochromatin' markers- H3K27me3 (Millipore Cat # 17-622, 2 μg/ml) and H3K9me3 (Abcam Cat#ab8898, 0.5 μg/ml) as well as the nuclear lamina-associated polypeptide LAP2α (Abcam Cat # ab5162, 2 ul/ml). Cells were also incubated, separately, with polyclonal rabbit serum (Millipore Cat # PP64B, 2 μg/ml). Primary antibodies and control serum were diluted in blocking solution. All primary antibody incubations were performed overnight at 4C on a rocking shaker. The following day, plates were washed thrice and incubated with the secondary antibody- donkey anti-rabbit alexa 647 IgG (Thermofisher cat # A-31573, 2 μg/ml) in PBS for about 45 minutes to 1 hour at room temperature on a shaker. Plates were wrapped to minimize exposure to ambient light during this incubation. All plates were washed twice with PBS at the end of this incubation and incubated with PBS containing DAPI (2μg/ml) for about 15 minutes in the dark. This solution was replaced with PBS and wells were imaged with a Zeiss Observer Z1 microscope (Zeiss, Germany) using 20x/0.8NA objective and Hamamatsu Orca-Flash4.0LT sCMOS camera. Using the ZEN2 acquisition software, we automated tile- scanning of the each well using DAPI and Cy5 filters. Levels of nuclear staining in each cell within acquired frames in the Cy5 channel were quantified with a custom script written on Metamorph, analyzed in Excel and plotted using SPSS.

Constructs and generation of lentivirus

Lentivirus was prepared and concentrated as previously described previously (Hammachi et al., 2012; Papapetrou et al., 2009). Titers were estimated by FACS analysis by determining the fraction of cells expressing fluorescent markers representing one or other factor (see Supplemetary Table S1). The FloJo 9.7 software was used to compensate and

analyze data from samples measured on a BD Fortessa flow cytometer. A titer value for one or other lentivirus (Supplementary Table S1) was determined for every fibroblast type used in this study and on average, post-concentrated lentiviral supernatants ranged between 10^7 to 10^8 titering units/ml (TU/ml). In general, we found older human fibroblasts (37-96y) were more resistant to lentiviral infection compared to fetal fibroblasts (MRC5). Thus, the amount of lentivirus required to infect older donors were generally 2-4 times greater than that used to infect fetal cells. Optimal ratios of MOIs for each of the lentivirus encoding WT O,S,K and M used in reprogramming MRC5 cells was determined empirically (as shown in Table 1 and Supplementary Tables S3).

Reprogramming human fibroblasts

The general strategy for reprogramming human fibroblasts, adapted from Hammachi et al., 2012 was as follows- cells were plated at a density of about 1-1.3 x 10^4 cells/cm2 (120,000 cells per 6-well plate or 700,000 cells per 10 cm plate) on day 0. The following day (day 1), they were infected at MOIs (Table 1, Supplementary Tables S3 and S4) that were determined to be optimal for reprogramming (see above section on constructs and generation of lentivirus). Mock infected (EMEM plus 10 or 20% FBS depending on the age of the donor from whom the fibroblast was derived from (see above) with 4μg/ml polybrene) and cells infected with different combinations of lentiviruses were allowed to incubate for about 16-18 hours following which media was replaced with fresh EMEM plus FBS. From days 2 through 5, media was replaced each day with cells monitored on the EVOS-FL fluorescence microscrope (ThermoFisher) to determine morphological changes and to make qualitative comparisons of fluorescence levels in different conditions. On Day 5, cells were harvested by trypsinization and counted. A small aliquot (about 20-30,000 cells) was used for FACS analysis to quantify the fraction of cells expressing all three (when reprogramming with three factor combinations) or four factors used in the reprogramming mix. In most experiments to determine reprogramming efficiencies with feeders (360,000 cells/6 cm plate of mitomycinC treated mouse embryonic fibroblasts, GlobalStem Cat #GSC-6001M), about 60,000 cells (per 6 cm plate i.e ~ 2500-3000 cells/ cm2) were re-plated. Valproic acid, (VPA, 1mM) (Sigma

Cat # P4543) was added to WT OSK or OSKM conditions on days 6 and 7 after cells were re-plated. VPA was not required when cells were reprogrammed with OSvK or OSvKM. On day 6, cells re-plated with fibroblast media (EMEM plus FBS) on both feeder-free (vitronectin- coated) or feeder-containing plates was washed once with PBS and replaced with either Essential 8 media (ThermoFisher Cat # A1517001) on feeder-free plates or with human ES media (DMEM/F12 media containing 20% knockout serum (ThermoFisher Cat # 10828-028), L-glutamine, non-essential amino acids, β-mercaptoethanol and 6 ng/ml FGF2 (R &D Systems Cat # 223-FB-025). hES media or Essential 8 media was re-placed everyday except weekends (where double the typical volume of media/plate was added) until day 25 when plates were fixed with 4% formaldehyde and scored for TRA 1-81 positive colonies.

Scoring TRA1-81 positive colonies

Briefly, fixed plates were washed, permeabilized and incubated with blocking solution as mentioned above (section on Measurement of senescence associated beta-galactosidase (SA β-gal) and other markers in young and old fibroblasts and in iPS and ES cells). All plates were then incubated with mouse anti-human TRA 1-81 IgG (Millipore cat # MAB-4381, 5 μg/ml) in blocking solution overnight on a rocking shaker at 4C. The following day, plates were washed twice with PBS and incubated with PBS containing the secondary antibody- goat anti-mouse IgG conjugated to HRP (Santacruz Cat #sc-2005, 0.4 μg/ml) for about 1 hour at room temperature on a shaker. Colonies were stained with the DAB kit (Vector Labs, Cat # SK-4100) as per the manufacturer's instructions. Total number of TRA 1-81 positive colonies was counted in each plate independently by two researchers. Additionally, since these colonies typically varied in size for any given condition, we categorized colonies as being either large ($\geq$ 50 cells/colony) or small to medium-sized colonies (25-50 cells/colony). Typically at least two large or medium colonies at day 25 were manually picked and passaged for further characterization of pluripotency. iPS- like colonies were typically expanded in hES media containing 10 μM Y27632 in the presence of mitomycin-C inactivated MEFs.

Characterization of derived iPS colonies

At least two large or medium- sized iPS colonies derived using either WT OSKM, OS(v)KM or OS(v)K from all age groups were manually picked and passaged over 10 times in feeder-containing plates. Levels of OCT4 (Abcam ab19857, 2 μg/ml) and NANOG (Abcam ab21624, 2 μg/ml) expression were measured by immunofluorescence in some of the hiPS lines derived. Additionally, mRNA expression of pluripotency associated genes- *NANOG*, *TRIM71* and *LIN28A* from few of the established iPSC clones were compared by QPCR to those found in established hESCs- H1 and H9. The primers used to detect these transcripts are listed in Supplementary Table S6. Tri-lineage (ectoderm, mesoderm, endoderm) differentiation potential for each established iPS line (passage 10-15) was determined as previously described (Chambers et al., 2011; D'Amour et al., 2005; Nakanishi et al., 2009; Watanabe et al., 2007). We used the following primary antibodies to determine differentiation potentials of all iPSC clones- Rabbit anti-PAX6 (ectoderm), Covance Cat #PRB-278, 1:500), goat anti-BRACYURY (mesoderm) (R & D systems Cat # AF2085, 1:40) and goat anti-SOX17 (endoderm) (R & D systems Cat # AF1924,1:200). Alexa647 conjugated secondary antibodies with DAPI were used to stain and visualize plates in the Cy5 and DAPI channels on an AxioVert 200M widefield microscope (Zeiss, Germany) using a 20x/0.8 NA objective and Coolsnap ES CCD camera (Photometrics, Tucson, AZ). Images were acquired using the metamorph (Molecular Devices, Sunnywale, CA) software. Karyotype analysis was performed on iPSC clones that were derived with either the OSvKM or OSvK mix, by the MSKCC molecular cyotogenetics core.

Gene expression measurements by QPCR and microarray analyses

Fibroblasts derived from MRC5 cells grown on 6 well plates as mentioned above, infected with different combinations of lentivirus (Supplementary Table S2), were harvested by the addition of trizol (Thermofisher Cat #15596-018) to each well. Total RNA was extracted as per manufacturer's instructions. cDNA was synthesized from 2 μg of RNA using the Maxima first strand cDNA synthesis kit with DNase (Thermofisher Cat

K1671) and expression of genes using primers listed in the table below (Supplementary Table S6) was determined using SYBR green. QPCR plates were made using a liquid handler with each sample in quadruplicate in a 384-well plate and run using the Roche Lightcycler 480. QPCR data was analyzed as previously mentioned (Bryant et al., 2008). Whole genome expression analysis was performed at the MSKCC Intergrated Genomics core using the Illumina human HT-12 V4.0 oligonucleotide bead chip array. Total RNA (20 ng) from mock-infected MRC5 cells (in duplicate), MRC5 expressing OSKM or OS(v)KM at day 5 (each condition in triplicate), a hESC line (H1 hESC), and 2 hiPSC lines (S1 and R4) was submitted to MSKCC's Integrated Genomics core for cRNA preparation. Labeling of cRNA with Cy3, hybridization as well as scanning of the array were performed as previous described as previously described on an Illumina microarray scanner. Data were quantile normalized using Illumina's Genome Studio and analyzed using Partek's Genomics Suite.

Chromatin-immunoprecipitation, endonuclease and sequencing (ChIP-Endo-Seq)

About 7 x 10^5 MRC5 fibroblasts were seeded on 10 cm plates on day 0 and reprogramming using either OSKM, OSvKM, OSK or OSvK was carried out as described above. On day 5, cells on plates were washed once with PBS (room temperature) and cross-linked by adding EMEM + 10% FBS containing 1% formaldehyde. Plates were returned to the incubator for 10-15 minutes. Cells were washed twice with ice cold PBS containing a cocktail of protease inhibitors (Sigma Cat # P8340; 10μl/ml PBS) and 1mM phenylmethylsulfonyl fluoride (PMSF; Roche Cat # 11359061001). Cross-linked cells from each Day 5 mix (OSKM, OSvKM, OSK and OSvK, all at passages between 7-9) as well as mock infected fibroblasts were scraped, pooled in 15 ml tubes and pelleted. We typically pooled cross-linked cells from at least 4 to 5, 10cm plates per condition (~ 5 million cells in total) for each ChIP-Seq assay. As controls, we harvested approximately 5 million cross-linked human ES (H9 or WA09hESCs, p. 35 from WiCell) or human iPS cells (a characterized iPSC line derived using OSvKM (3S2 hiPSC, p.20) from 6 cm plates and grown in feeder-free conditions.

Cells were resuspended in 500 ul of FA-lysis (50 mM HEPES, 150 mM NaCI, 2 mM EDTA (pH 8.0), 1% Triton X-100, 0.1% Sodium deoxycholate) with 0.2% added SDS and sonicated twice at power of 4 for 10 seconds using a micro-tip on a Branson Sonifer 250. The sonicate was then spun down at 13,000 rpm for 5 min and the supernatant was transferred to a new tube. FA-lysis buffer (1.5 ml) without SDS was added to the sonicated lysate and then divided evenly for each 1-3 ul antibody used (RPB1 CTD, Cell Signaling 2629S Lot: 3; H3K4me3, Abcam ab8580 Lot: GR240214-2; H3K27ac, Abcam ab4729 Lot: GR243994-1; H3K27me3, Abcam ab6002 Lot: GR218433-1; H3K4me1, Abcam ab8895 Lot: GR243233-1-1; H3K9me3, Abcam ab8898 Lot: GR188519-1; OCT4, Abcam ab19857 Lot: GR187600-1; SOX2, R&D Systems AF2018; KLF4, R&D Systems AF3640; cMYC, R&D Systems AF3696). 20 ul of protein A/G magnetic beads (Pierce Protein A/G Magnetic Beads Prod #88803) was added to microtubes (QIAGEN Collection Microtubes Cat No. 19560) and then the mix was incubated overnight at 4 C using a rocker.

Before each reaction beads were precipitated with a magnetic block and then washed and resuspended with 200 ul of the following solutions: FA lysis, High Salt (50 mM HEPES, 1 M NaCI, 2 mM EDTA (pH 8.0), 1% Triton X-100, 0.1% NaDeoxycholate), Wash 2 (50 mM Hepes/KOH, 0.5 M NaCI, 2 mM EDTA (pH 8.0), 1% TritonX-100, 0.1% NaDeoxycholate), Wash 3 (25 mM LiCI, 1% NP4O-Nonidet (IPEGAL), 1% NaDeoxycholate, 10 mM Tris-CI (pH 8.0) and Tris buffer (10 mM Tris-HCI (pH depends on following reaction), adapted from (Rhee and Pugh, 2012)). After the first set of washes, the chromatin bound beads are digested with either micrococcal nuclease (MNase) digestion buffer (previous Tris wash pH 8.5; 18.84 ul of 10 mM Tris pH 8.5, 0.16 ul of 500 mM CaCl2 and 1 ul of MNase (0.00002 U/ul) for H3K4me3, H3K27ac, H3K27me3 and H3K4me1 or DNaseI digestion buffer (previous Tris wash pH 7.5; 18 ul of 10 mM Tris pH 7.5, 0.2 of 500 mM MgCl2, 0.8 ul of 500 mM CaCl2 and 1 ul of DNaseI (0.02 U/ul for PolII and 0.006 U/ul for transcription factors) for PolII, OCT4, SOX2, KLF4 and cMYC and incubated in an Eppendorf Thermomixer R with the MTP block for 60 minutes at 37C at 1400 rpm. Beads were again washed with the last wash at pH 7.5 then incubated in end repair mix (2 ul of H2O, 4 ul of 10x T4 PNK buffer (NEB),

2 ul of DTT 100 mM, 8 ul of 2 mM dNTP, 2 ul of 20 mM rATP, 1 ul of T4 DNA polymerase (3 U/ul NEB M0203S) and 1 ul T4 polynucleotide kinase (10 U/ul NEB M0201L)) for 30 minutes at 20C at 1400 rpm. Beads were again washed in which the last wash was at pH 7.5 then incubated in adaptor ligation mix (5.3 ul of H2O, 4 ul of 10x T4 PNK buffer (NEB), 8 ul of 30% PEG, 2 ul of 20 mM rATP, 0.4 ul of 50 uM P5 and P7 (Illumina TruSeq adaptors, P5 is IS1: 5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3' and IS3 5'-AGATCGGAAGAGC-3'; P7 is IS2 5'-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3' and IS3) and 0.3 ul T4 DNA ligase (30 Weiss U/ul Thermo Scientific EL0013)) for 15 minutes at 20C at 1400 rpm. Beads were again washed in which the last wash was at pH 8.8 then incubated in the fill in mix (10 ul of H2O, 4 ul 10x ThermoPol buffer (NEB B9004S), 5 ul of 2 mM dNTP and 1 ul of Bst Polymerase, Large Fragment (8 U/ul NEB M0275S)) for 20 minutes at 37C at 1400 rpm. Beads were again washed in which the last wash was at pH 7.5 then the chromatin was eluted with 200 ul of elution buffer (25 mM Trizma, 2 mM EDTA (pH 8.0), 200 mM NaCI and 0.5% SDS) and incubated for 15 min at 65C at 1400 rpm followed by magnetic precipitation and the supernatant was transferred to a new microtubes. 0.5 ul of Protease K (Protease K, recombinant PCR Grade Roche Diagnostics 11733400) was added and then incubated overnight at 65C.

1 ul of glycogen (Glycogen for molecular biology Roche Diagnostics 14397127) and 400 ul 100% ethanol was added to the eluted chromatin and then incubated 1 hour at -80C. DNA was then precipitated for 30 minutes at 13000 rpm and then rinsed with 70% ethanol and resuspended in 30 ul H2O. DNA concentration for each sample was initially measured by QPCR using oligos IS4 (5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT-3') and Indexing Oligo (5'-CAAGCAGAAGACGGCATACGAGATcctgcgaGTGACTGGAGTTCAGACGTGT-3') 2 ul of DNA was added to 10.5 ul H2O and then mixed with 12.5 ul of 2x QPCR mix (QPCR 2x reaction buffer: 20 mM Tris-HCl (pH 8.3), 13 mM $MgCl_2$, 100 mM KCl, 400 uM dNTPs, 4% DMSO, 2x SYBR Green I (Molecular Probes), 0.01% Tween 20, 0.01%

NP40, 10 ng/ul of each oligo primer, and 0.025–0.1 U/ul of Taq polymerase (Roche)). This mix was then measured in four separate wells with 5 ul each in a 384 well plate in a Roche Light Cycler 480 QPCR machine (15 sec at 95C, 15 sec at 62C and 60 sec at 72C for 40 cycles) (Bryant et al., 2008). The measurement was then analyzed using Absolute Quantification/2nd Derivative Max and then the average Cp for the quadruplicate measurement noted. A second PCR reaction (regular non-quantitative 10 mM Tris HCl pH 8.3, 1.5 mM MgCl2, 50 mM KCl, 0.2 mM dNTP, 10 ng/ul of each oligo primer, and 0.025–0.1 U/ul of Taq polymerase (Roche)) with 2 ul of DNA in a total of 50 ul reaction mix using oligos IS4 and Indexing Oligo for 5 cycles more then the Cp value measured previously by QPCR. 20 ul from this PCR reaction was run on a 3% agarose gel (1.5% Agarose Genetic Analysis Grade Fisher Scientific BP1356-500 and 1.5% NuSieve GTG Agarose Lonza) and imaged using ChemiDoc MP Imaging System from Bio Rad and the lanes were quantitated using Image Lab 5.0 software. DNA abundance was measured for each lane from 150 – 550 bp range. To make the sequencing library, the previously determined DNA measurements were then used to calculate the least number of PCR rounds needed to amplify a sufficient amount of DNA to visualize on an agarose gel (assuming a PCR growth rate of 1.8 fold per round and 20 ul of DNA added to the PCR mix instead of 2 ul). For each ChIP a different Indexing Oligo, each with a unique index was used. After amplification 10 ul of the 50 ul was again run on a 3% agarose gel and DNA measured as described above. This measurement was then used to calculate the amount of DNA needed to mix together in a pool at its desired ratio for sequencing. 10 million hits per ChIP is a good target. Pooled DNA was then ethanol precipitated by adding 1 ul glycogen and 2 volumes of 100% ethanol. This was then resuspended in 20 ul H2O and all of it run on a 10% acrylamide. The lane from 150 – 550 bp in size was cut out, the gel was mechanically crushed and the DNA was eluted overnight in 600 ul of oligo elution buffer (10 mM TrisCl pH 7.5, 50 mM NaCl, 1 mM EDTA pH 8.0). The eluted DNA was then ethanol precipitated as described above. This pool was then resuspended in 30 ul H2O and submitted for paired end sequencing with 50 bp read on each side.

Genome Mapping

All data from ChIP-Endo-Seq experiments were mapped to hg19 using the following steps. Data from the sequencing facility was received as fastq files in which the bar codes were already removed and sequence data from each bar code was put into separate pairs of fastq files. Each of the paired fastq files was first modified to eliminate the adaptor sequence (since many sequences are less 50 bp in size) using the program cutadapt-1.9.1 with the command 'cutadapt -a AGATCGGAAGAGC'. The pair of fastq files was then examined and the paired sequence info was eliminated if either sequence in the pair had all of its sequence removed by cutadapt. The sequence was then mapped to hg19 using the program bwa-0.7.13 and the commands 'bwa aln' followed by 'bwa sampe' which generated a SAM file for each pair of fastq files. All SAM files from the same immunoprecipitation were then examined to eliminate any duplicates (i.e. sequences that have the same chromosome, start and end position). Finally, a new file was generated that contained chromosome number, start and end position for each immunoprecipitation.

Measuring Values and Error

The above position data was then used to calculate multiple different sets of features. Features could be of many different categories, for example, every 2000 bp of the entire genome, all of the genes transcribed regions (GRCh37), possible OCT4 binding sites in 100 bp bins etc. Feature values were calculated from position data by counting the total number of midpoints that are within a given feature (count). This count is then divided by the size, in bp, of the feature and this quotient is then divided by total number of positions mapped divided by the genome size. Thus, a value of 1 for a feature would indicate that the ChIP-Endo-Seq neither enriched nor depleted immunoprecipitated sequences for that feature. This value will be referred to as the unnormalized value.

To more reliable compare feature values between experiments, quantile normalization was performed for each feature category and immunoprecipitation antibody. A standard was chosen from an immunoprecipitation that had a large number unique positions mapped for each antibody. The unnormalized feature values were then quantile

normalized to the standard for each antibody and feature category. This method of normalization assumes that the distribution of values is similar for each cell type and antibody being compared. For most cases this assumption is valid. For example, the distribution of values for Pol II in transcribed genes is likely similar for many different cell types, even though individual genes have different expression levels in different cell types. There are some cases though were the assumption of similar distributions, does not hold. For example, the distribution of OCT4 signal in fibroblasts and iPS cells is most likely quite different, since OCT4 is not expressed in fibroblasts but is expressed in iPS cells. For cases were there is good reason to suspect that a cell type-specific antigen would have a dissimilar distribution (OCT4, SOX2 and KLF4 in fibroblasts and H3K27me3 in iPS cells) a different standard is chosen for that cell type and antibody. The above quantile normalized values will be referred to as normalized values.

Measurement error, for each feature category, was estimated by comparing replicates from the same cell type (fibroblasts, OSKM, OSvKM, OSK, OSvK and iPS) and antibody. A commonly used method for estimating error would be to measure the variance for each individual feature by comparing replicates and then using those values to estimate an expected error for each feature. Unfortunately error estimates for individual features are often unreliable, since the variance calculation is limited by the number of replicates (Yu et al., 2013). To get an error estimation, which utilizes many more measurements then the number of replicates, we first calculated the percent standard deviation (i.e. sd / avg) for each feature and then calculate the average percent standard deviation for many different features that fall within a range counts and feature sizes. This will generate a percent standard deviation table, for each replicate and feature category, covering the entire range of counts and feature sizes. For the special case in which the feature count is 0 the average standard deviation is calculated, for all features within a size range that also have an average normalized value greater then 0 (i.e. at least one of the other replicates has a count of 1 or more). The standard deviation for each feature is then calculated by looking up the corresponding average percent standard deviation from the table and multiplying that value by the features normalized value. For the cases in which the feature's normalized value is zero, the standard deviation is taken

directly from the table. For cases in which no replicates are available, measurement error was estimated by using a percent standard deviation table from a different cell type but same antibody.

Data Tables

The normalized values and standard deviations for each immunoprecipitation and feature category were then used to generate two final tables (in postgres SQL): a gene table and a 2 Kb table. These two tables were then used to generate all of the ChIP-Endo-Seq calculations in this paper. For each feature category and antibody, replicate normalized values were averaged together and either the standard deviation of those values or the estimated standard deviation (see above), which ever was larger, was used as the standard deviation for each feature (from the replicate average table). The gene table was then generated by combining the replicate average tables for the feature category of genes from the Pol II immunoprecipitation for each of the cell types in this study (fibroblasts, OSKM, OSvKM, OSK, OSvK and iPS). Each row in the table represents a different gene with columns containing a gene (or feature) ID, chromosome number, transcription start site, end site, stand ID in addition to four arrays with each element in the array representing a different cell type. The first array contains the average normalized value, the second array contains the standard deviation and the third array contains the difference between the normalized value of fibroblasts and the normalized value of the elements cell type. The final array is the significant difference array. Each element contains the value of the same element in the difference array divided by the sum of the fibroblasts standard deviation and the standard deviation from the same element. The element order is: fibroblasts, OSKM, OSvKM, iPS, OSK and OSvK. The 2 Kb table is similar to the gene table but each row represents a unique non-overlapping 2 Kb segment. In addition there are four arrays, like in the gene table, for each of the different antibody used in this study (Pol II, H3K27ac, H3K4me1, H3K27me3, H3K9me3, OCT4, SOX2 or SOX2-VP16, KLF4 and cMYC). Each row contains a column with a unique feature ID, chromosome number, feature start position and end position. There are four arrays for Pol II and each histone mark, which were generated as described for the gene table.

There are also four arrays for the enhancer mark, which is the sum of Pol II, H3K27ac and H3K4me1. Since the normalized value for the transcription factors (OCT4, SOX2 or SOX2-VP16, KLF4 and cMYC) are calculated using features 100 bp in size, it is possible that there might be more then one feature, for each transcription factor, per 2 Kb segment. For such a case, the feature with the highest normalized value within the array is used for each transcription factor. The gene and 2 Kb tables populated a SQL database which was then queried for the results.

Transcription Factor Binding Site Calculations

A general method of how the data is queried is as follow. First, rows where selected that surpassed a particular cutoff (often a significance cutoff). For example, rows can be selected from the gene table, in which the OSKM element of the significant difference array is above 2. This would select all of the genes that were significantly up regulated between fibroblasts and OSKM Day 5 cells. The second step could then count how many of the selected rows surpass another cut off for a different element and/or array or generate a distribution plot of the values in any array's element. For example, the query could count how many of the above select rows have the iPS element of significant difference array above 2. This would determine the percentage of genes that are up-regulated in OSKM that are still up-regulated in iPS cells.

To determine the number of bound sites for OCT4, SOX2 or SOX2-VP16 and KLF4 in each cell type, the 2 Kb table was queried to select the rows in which the significant difference array is greater then 0.75 at the cell types element and the normalized value array is less then 8 at the fibroblasts element. Since the above transcription factors are not expressed in fibroblasts, the selection for a low normalized value in fibroblasts will reduce false positives arising from mapping errors at repetitive sequences. Since cMYC is expressed and presumable bound in fibroblast cells, a different selection was performed. This query selected for a normalized value divided by standard deviation greater then 0.7 and normalized value > 0 for the given cell type.

To determine the percentage of cells bound by a transcription factor in one cell type that is also bound by the same factor in another cell type (overlap), first the selection for bound sites were performed (as described above), followed by a count of the number of those features that have a normalized value above 8 in a different cell type. This query could have been performed by selecting for both cell types that are above a significance cutoff, but that would tend to underestimate the percent overlap of bound sites between cell types, since the measurement error strongly influences the number of rows above a cutoff. The same method to determine the percent overlap of a transcription factor between cell types (described above) was also used to determine the percent bound by one transcription factor that is also bound nearby by a different factor in same cell type.

To determine the percent of singly bound (e.g. OCT4 bound with no SOX2 nearby) or doubly bound sites (e.g. OCT4 and SOX2 are both bound near each other), in Day 5 cells, that were also bound in iPS cells, a more stringent initial selection was performed. For doubly bound sites, both transcription factors were select together using the significance cutoff. For singly bound sites, the bound factor was select for using the significance cutoff (described above) and the non-bound factor was select against by requiring its normalized value to be less then 1. The percent of these singly and doubly bound sites in Day 5 cells also bound in iPS cells was determined by counting the number of these sites that have a normalized value above 8 for either of the transcription factors in iPS cells.

Distribution Plots

Distribution plots were made by first performing a selection (e.g. OCT4 bound sites in OSKM) and then using all of the values from a particular array and element (e.g. enhancer normalized values in OSKM) to generate a distribution function. First, Log2 is calculated for each of the values (note: a small number, e.g. 0.6, is added to the value prior to taking the Log2 to prevent the undefined value of Log2 0). Second, the number of Log2 values that lies within successive non-overlapping ranges is counted (e.g. how many Log2 values are between 1.0 and 1.5, how many between 1.5 and 2.0 etc.). The p value is then calculated by first dividing this count by the total number of values, giving

the fraction of values within each range. This fraction is then divided by its range size to give the p value. For example, if there are 1000 values being examined and 100 have a Log2 value between 1.0 and 1.5 then the p value will be (100 / 1000) / (1.5 – 1.0) = 0.2. Each of the p values will then be plotted at the midpoint of its range (i.e. (1.5 + 1.0) / 2 = 1.25).

Enhancer Calculations

Putative enhancers were select using the following criteria: enhancer normalized value greater then 6.0 AND (H3K4me1 norm value + 0.6) / (H3K4me3 norm value + 0.6) greater then 1.0 AND (H3K27ac norm value + 0.6) / (Pol II norm value + 0.6) greater then 1.0 where all norm values were taken from the same cell type (note: the 0.6 was added to prevent division by zero). The first criterion selects for high enhancer signal, the second criterion eliminates active or poised promoters and the third criterion eliminates transcribed genes without an embedded active enhancer. In this paper, destroyed enhancers were defined, as enhancers that were active in fibroblasts but not active in Day 5 or iPS cells. So destroyed enhancers were selected selecting for an enhancer in fibroblasts (as defined above) AND an enhancer normalized value less then 6.0 AND a significant difference value less then -2.0 in the cell type being examined. The first criterion selects for an active enhancer in fibroblasts, the second criterion ensures that the enhancers is longer active and the third criterion ensures that the enhancer strength reduced a significant amount. Created enhancers are enhancers that were not active in fibroblasts but are active one of the other cell types. Created enhancers were selected by selecting for an enhancers in a given cell type (other then fibroblasts) AND an enhancer normalized value in fibroblasts less then 6.0 AND significant difference value in the given cell type greater then 2.0.

To determine how many enhancers created in one cell type were also created in another cell type, first the selection for created enhancers were performed (as described above) followed by a count of how many of these sites have an enhancer signal above 6.0 in a different non-fibroblast cell type. To determine how many enhancers destroyed in one

cell type were also destroyed in another cell type, first the selection for destroyed enhancers were performed (as described above) followed by a count of how many of these sites have an enhancer signal below 6.0 in a different non-fibroblast cell type.

To determine how many created or destroyed enhancers have either OCT4 or SOX2 bound within, the size of the enhancer must be taken into account. Most of the previous analysis used 2 Kb fragments for determining its values, but enhancers can be much larger in size then 2 Kb. If OCT4 and SOX2 bound within a single 2 Kb fragment and formed an enhancer in which the enhancer marks extending over 20 Kb only one of the ten 2 Kb fragments would show OCT4 and SOX2 binding. So to prevent enhancers larger then 2 Kb from misrepresenting the results, the following analysis was performed. First, created or destroyed enhancers were selected (as described above). Second, all adjacent 2 Kb fragments are grouped together as a single enhancer and the largest normalized value for OCT4 and SOX2 is assigned to the grouped enhancer. Finally the total number of grouped enhancers is counted and the number of grouped enhancers with a OCT4 or SOX2 normalized values is greater then 8.0 is counted. These two numbers were then used to calculate the percent of created or destroyed enhancers bound by either transcription factor.

References

Bryant, G.O., Prabhu, V., Floer, M., Wang, X., Spagna, D., Schreiber, D., and Ptashne, M. (2008). Activator control of nucleosome occupancy in activation and repression of transcription. PLoS Biol *6*, 2928-2939.

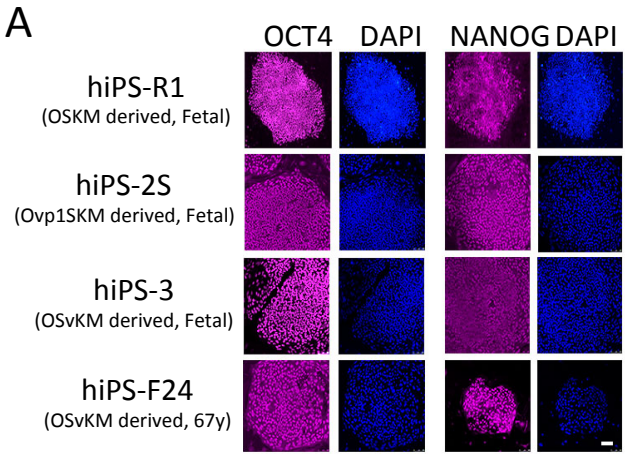
Chambers, S.M., Mica, Y., Studer, L., and Tomishima, M.J. (2011). Converting human pluripotent stem cells to neural tissue and neurons to model neurodegeneration. Methods Mol Biol *793*, 87-97.

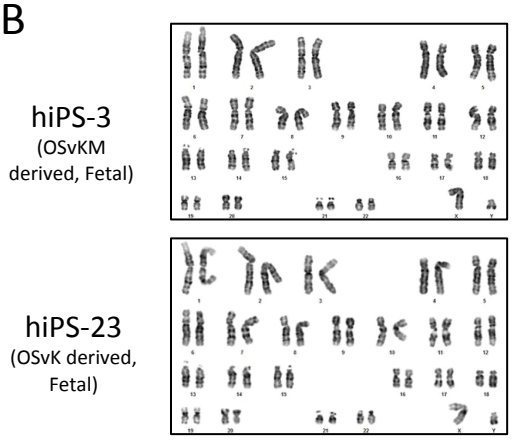
D'Amour, K.A., Agulnick, A.D., Eliazer, S., Kelly, O.G., Kroon, E., and Baetge, E.E. (2005). Efficient differentiation of human embryonic stem cells to definitive endoderm. Nat Biotechnol *23*, 1534-1541.

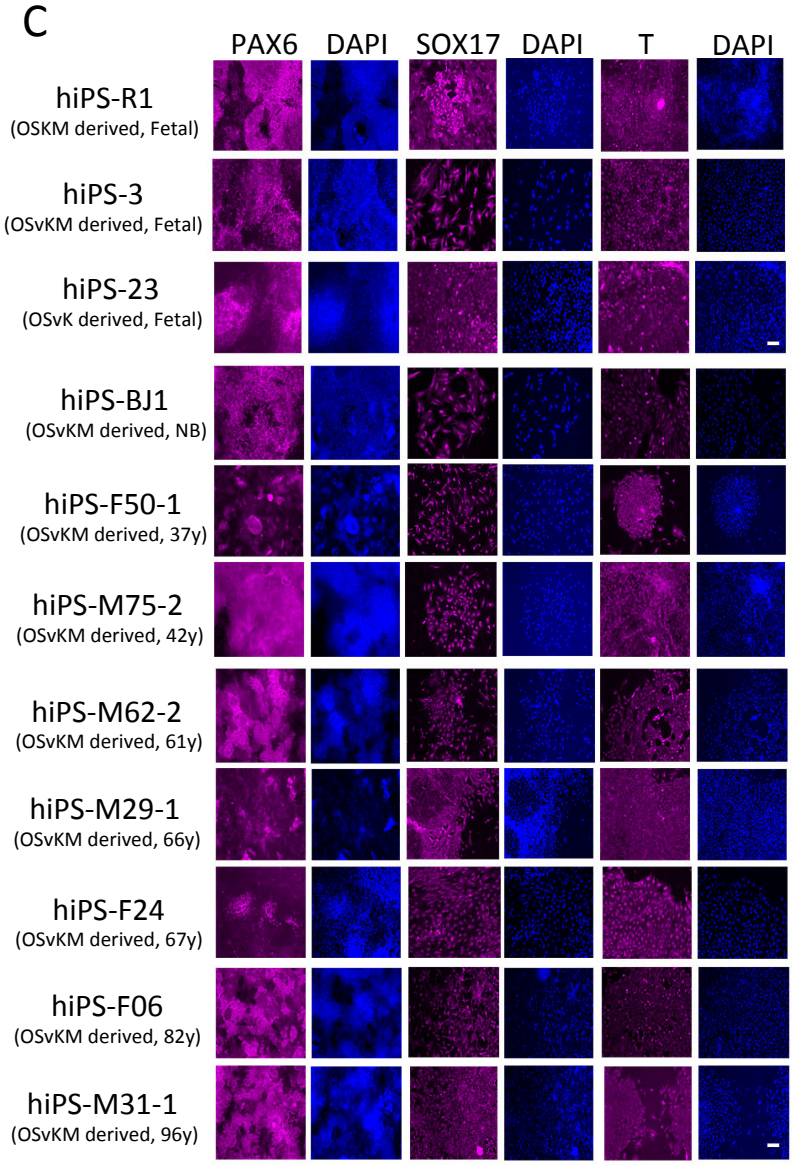
Hammachi, F., Morrison, G.M., Sharov, A.A., Livigni, A., Narayan, S., Papapetrou, E.P., O'Malley, J., Kaji, K., Ko, M.S., Ptashne, M.*, et al.* (2012). Transcriptional activation by Oct4 is sufficient for the maintenance and induction of pluripotency. Cell Rep *1*, 99-109.

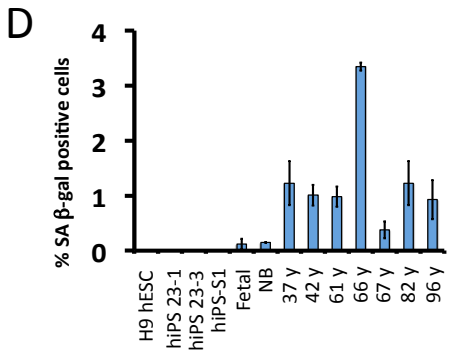
Miller, J.D., Ganat, Y.M., Kishinevsky, S., Bowman, R.L., Liu, B., Tu, E.Y., Mandal, P.K., Vera, E., Shim, J.W., Kriks, S.*, et al.* (2013). Human iPSC-based modeling of late-onset disease via progerin-induced aging. Cell Stem Cell *13*, 691-705.
Nakanishi, M., Kurisaki, A., Hayashi, Y., Warashina, M., Ishiura, S., Kusuda-Furue, M., and Asashima, M. (2009). Directed induction of anterior and posterior primitive streak by Wnt from embryonic stem cells cultured in a chemically defined serum-free medium. FASEB J *23*, 114-122.
Papapetrou, E.P., Tomishima, M.J., Chambers, S.M., Mica, Y., Reed, E., Menon, J., Tabar, V., Mo, Q., Studer, L., and Sadelain, M. (2009). Stoichiometric and temporal requirements of Oct4, Sox2, Klf4, and c-Myc expression for efficient human iPSC induction and differentiation. Proc Natl Acad Sci U S A *106*, 12759-12764.
Rhee, H.S., and Pugh, B.F. (2012). ChIP-exo method for identifying genomic location of DNA-binding proteins with near-single-nucleotide accuracy. Curr Protoc Mol Biol *Chapter 21*, Unit 21 24.
Watanabe, K., Ueno, M., Kamiya, D., Nishiyama, A., Matsumura, M., Wataya, T., Takahashi, J.B., Nishikawa, S., Nishikawa, S., Muguruma, K.*, et al.* (2007). A ROCK inhibitor permits survival of dissociated human embryonic stem cells. Nat Biotechnol *25*, 681-686.
Yu, D., Huber, W., and Vitek, O. (2013). Shrinkage estimation of dispersion in Negative Binomial models for RNA-seq experiments with small sample size. Bioinformatics *29*, 1275-1282.

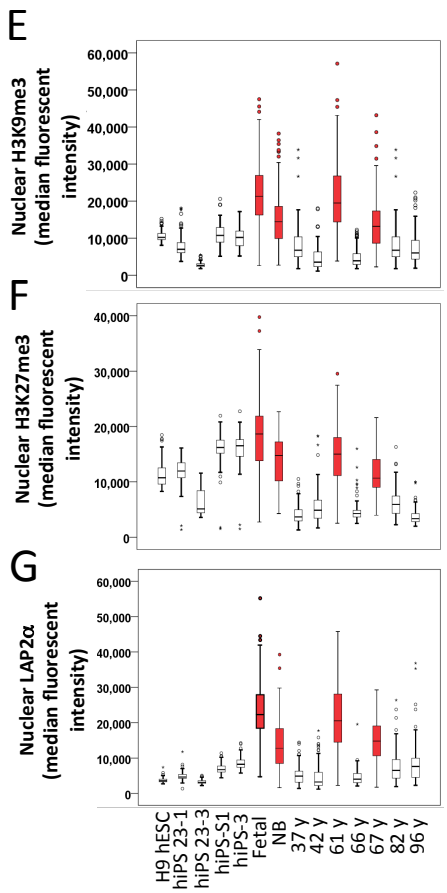
Supplementary Figure S1- related to Figure 1





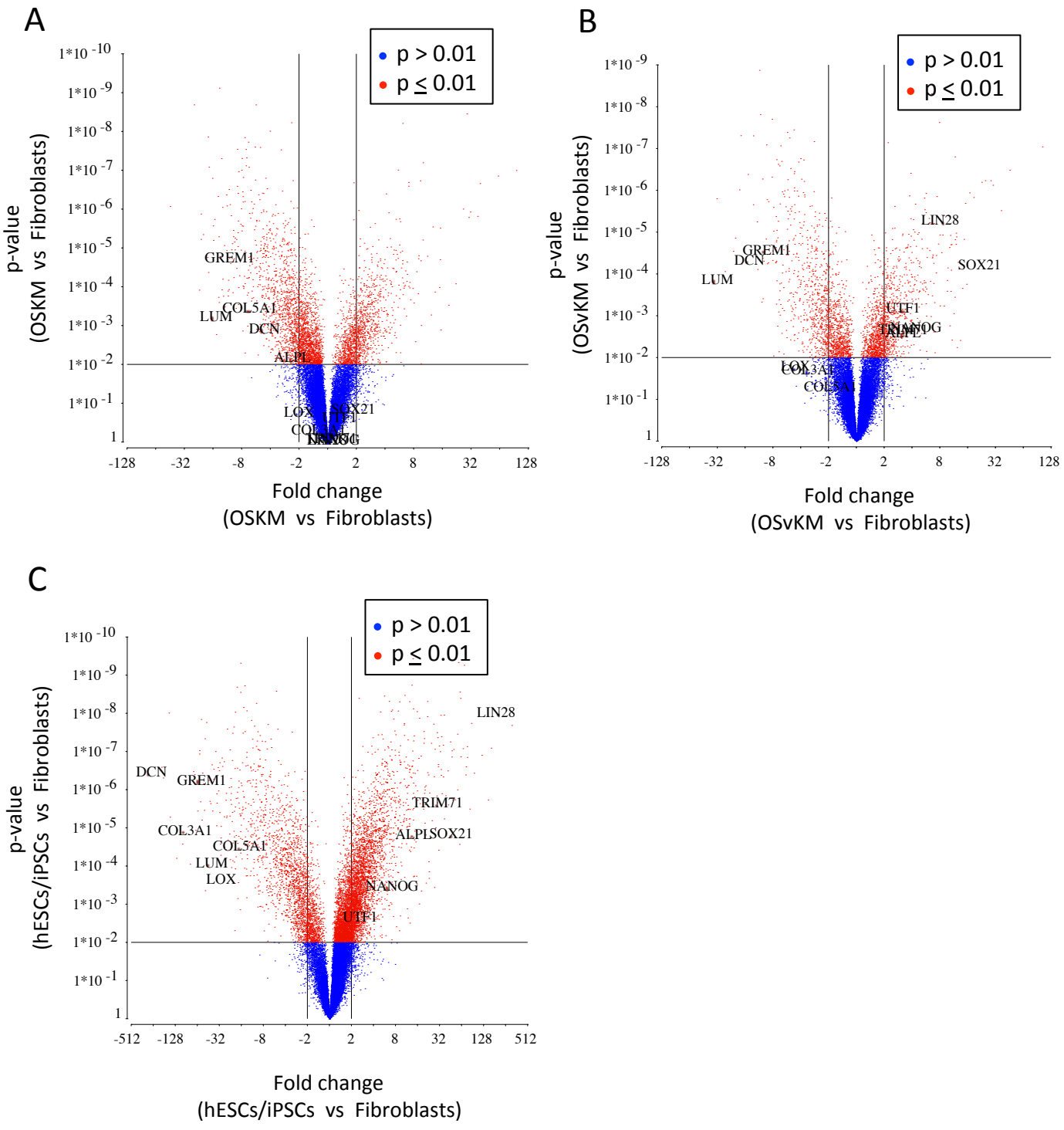






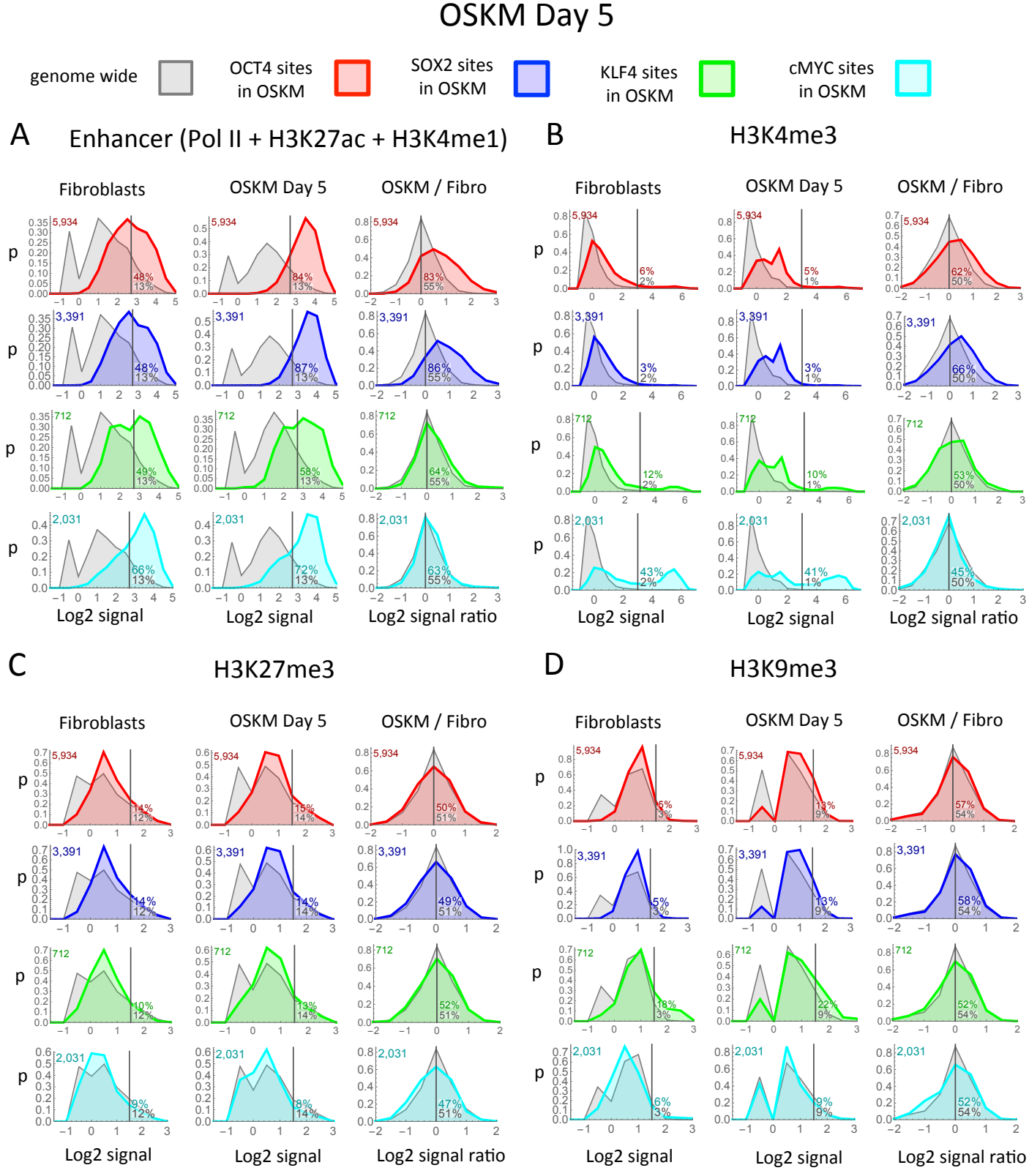
Supplementary Figure S1. Characterization of derived iPS colonies and analysis of age-associated markers in donor fibroblasts of different ages, hESC (H9) and hiPS cells.
(A) Expression of OCT4 and NANOG in iPS colonies (> passage 10) derived from fetal and 67 year old human fibroblasts. The first three rows show iPS colonies derived from fetal fibroblasts using either OSKM (hiPS-R1), Ovp1SKM (hiPS-2S, also see Table S1) or OSvKM (hiPS-3). The last row shows OCT4 and NANOG expression in an iPS line derived from a 67 year old donor using OSvKM. Scale bar = 100 mM. **(B)** Karyotype analysis of iPS lines derived from fetal fibroblasts using OSvKM or OSvK. **(C)** Tri-lineage differentiation of iPS lines derived from donor fibroblasts of different ages (top to bottom: first 3 rows (hiPS-R1, hiPS-3 and hiPS-23) - fetal, newborn (hiPS-BJ1), 37y (hiPS-F50-1), 42y (hiPS-M75-2), 61y (hiPS-M62-2), 66y (hiPS-M29-1), 67y (hiPS-F24), 82y (hiPS-F06) and 96y (hiPS-M31-1)) using OSKM (top row) or with OS(v)KM (all other rows). PAX6, SOX17 and T (BRACYURY) were used as markers for neural (ectodermal), endodermal and mesodermal differentiation respectively. Scale bar = 100 mM. **(D)** Cellular levels of senescence-activated of b-galactosidase (SA-bgal), **(E)** nuclear H3K9me3, **(F)** nuclear H3K27me3 and **(G)** nuclear LAP2a. Red boxes highlighted in (E), (F), and (G) show levels of these markers in fetal, new born (NB) and two older donor fibroblasts (61y and 67y) that were on average higher than other fibroblasts used in this study. Two standard deviations (SDs) above or below the mean are shown in circles, 3 SDs are shown with an asterisk.

Supplementary Figure S2- related to Figure 2



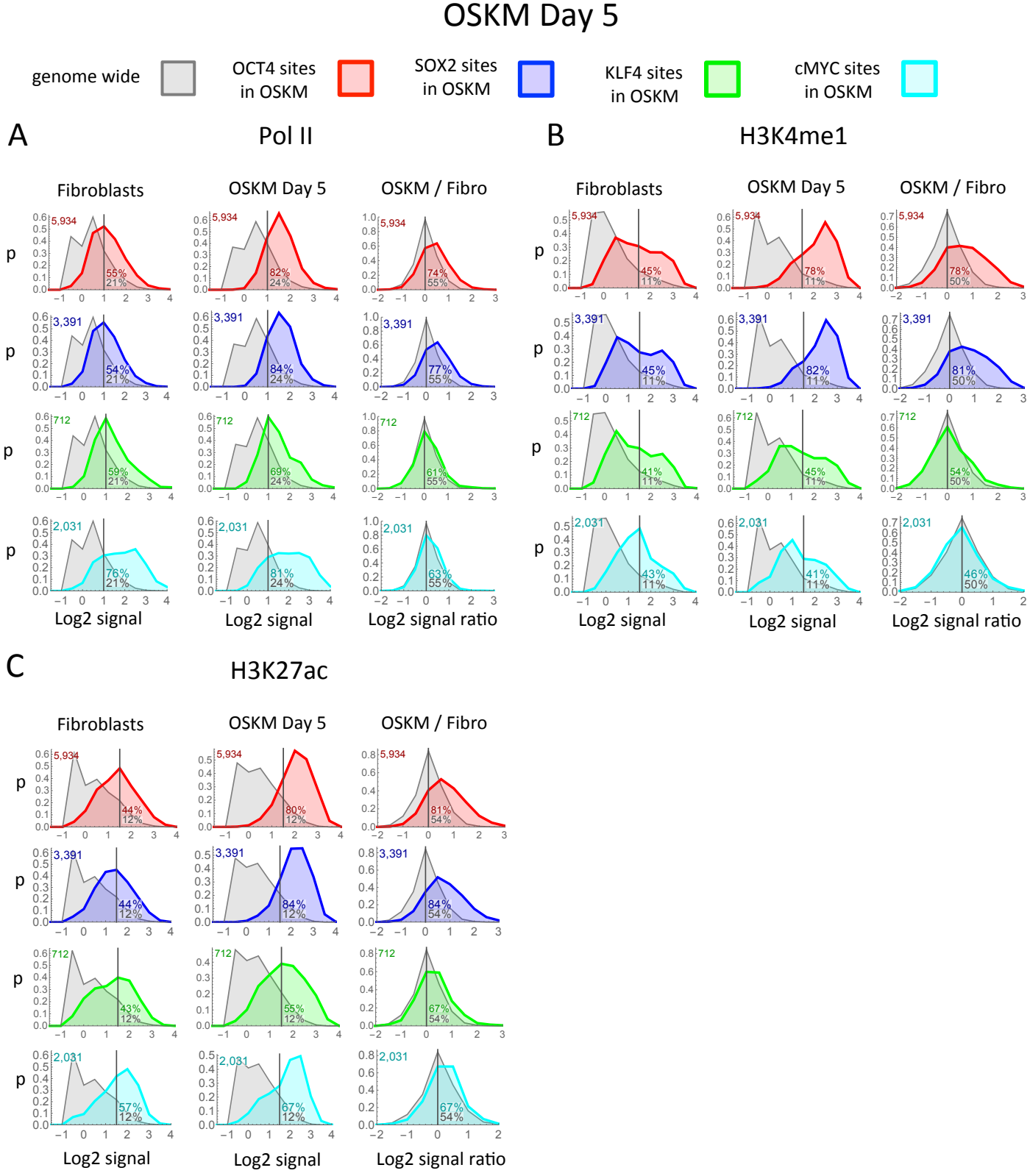
Supplementary Figure S2. Analysis of differentially expressed genes.
(A) Volcano plots showing genes that are differentially expressed (by microarray analysis) in day 5 OSKM versus parental fetal fibroblasts, **(B)** day 5 OSvKM versus parental fibroblasts and **(C)** hESC/iPSCs versus parental fibroblasts. In each plot a dot represents a probe for a gene in the Illumina HT12 array, vertical cut off lines represent 2 $\log_2$ fold up or down-regulated level and the horizontal cutoff line represents $p = 0.01$. Expression level of genes below this line ($p > 0.01$) are shown in blue and those above this line ($p < 0.01$) are shown in red. Expression level of genes associated with pluripotency and those specifically expressed in fibroblasts are labeled.

Supplementary Figure S3- related to Figure 5



Supplementary Figure S3. Distribution of signals at the site of TF binding in OSKM Day 5 cells.
The distribution of signals at the site of TF (OCT4, SOX2, KLF4 and cMYC) binding in OSKM Day 5 cells is shown (colored curves) along with the distribution of the signal genome wide (gray curves) as described in detail in Figure 6. Plots labeled 'Fibroblasts' show the distribution of the signal measured in fibroblasts at the sites where the TF bound in OSKM day 5 cells (colored curves) or genome wide (gray curves). Plots labeled 'OSKM Day 5' show the distribution of the signal measured in OSKM day 5 cells at the site of TF binding in the same cells where the TF bound (colored curves) or genome wide (gray curves). Plots labeled 'OSKM / Fibro' show the distribution of the signal measured in OSKM day 5 cells divided by the signal measured in fibroblasts, prior to taking the Log2, at the TF bound in OSKM Day 5 cells (colored curves) or the genome wide ratio of the signals (gray curve). Distribution plots show (**A**) the enhancer signal, which is the sum of the Pol II, H3K27ac and H3K4me1 signal, (**B**) the promoter signal H3K4me3, (**C**) the repressive signal H3K27me3 and (**D**) the repressive signal H3K9me3.

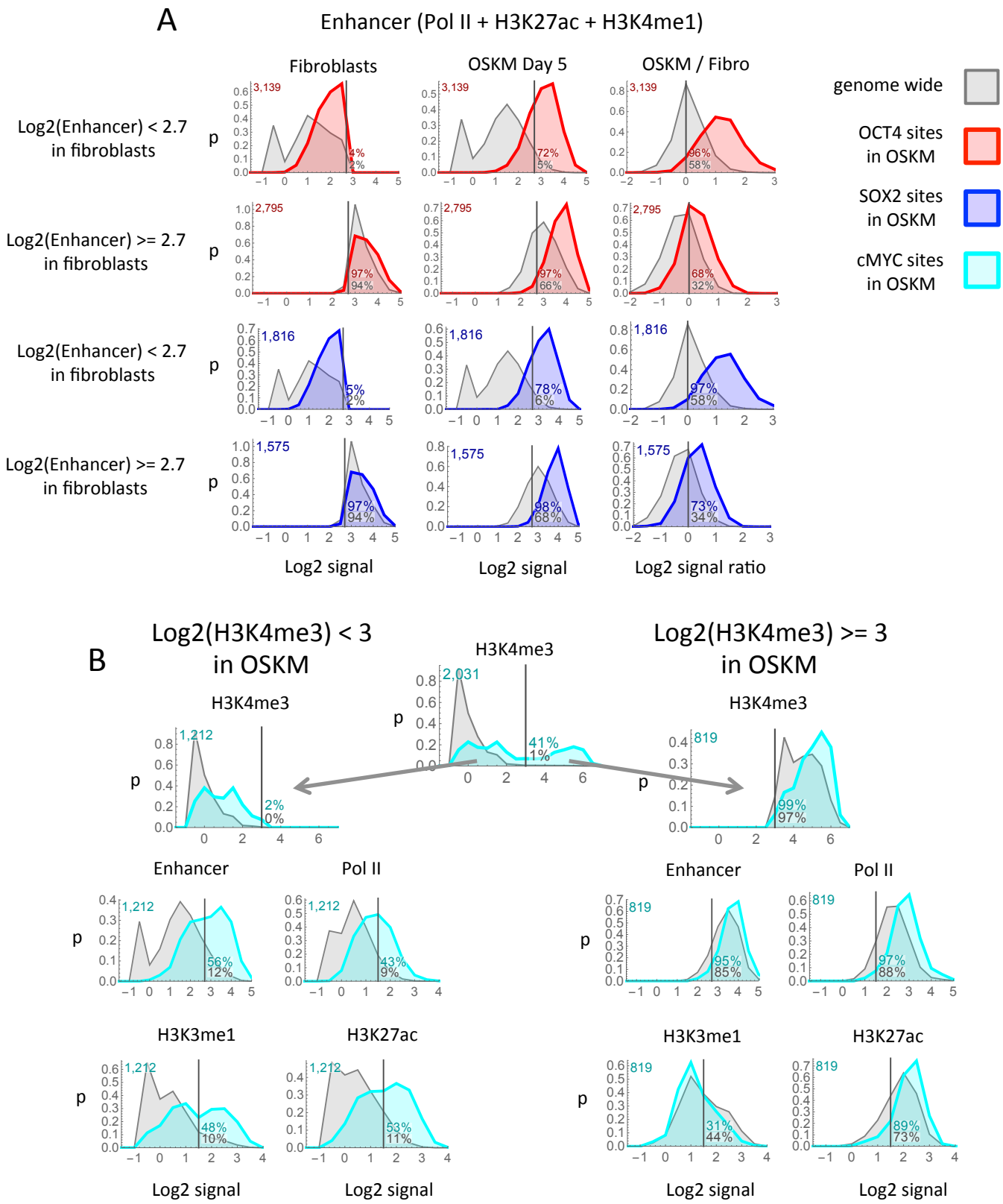
Supplementary Figure S4- related to Figure 5



Supplementary Figure S4. Distribution of signals at the site of TF binding in OSKM Day 5 cells.
The distribution plots are as described in Figure 5 and S3. Distribution plots show (**A**) the Pol II signal, (**B**) the H3K4me1 signal and (**C**) the H3K27ac signal.

Supplementary Figure S5- related to Figure 5

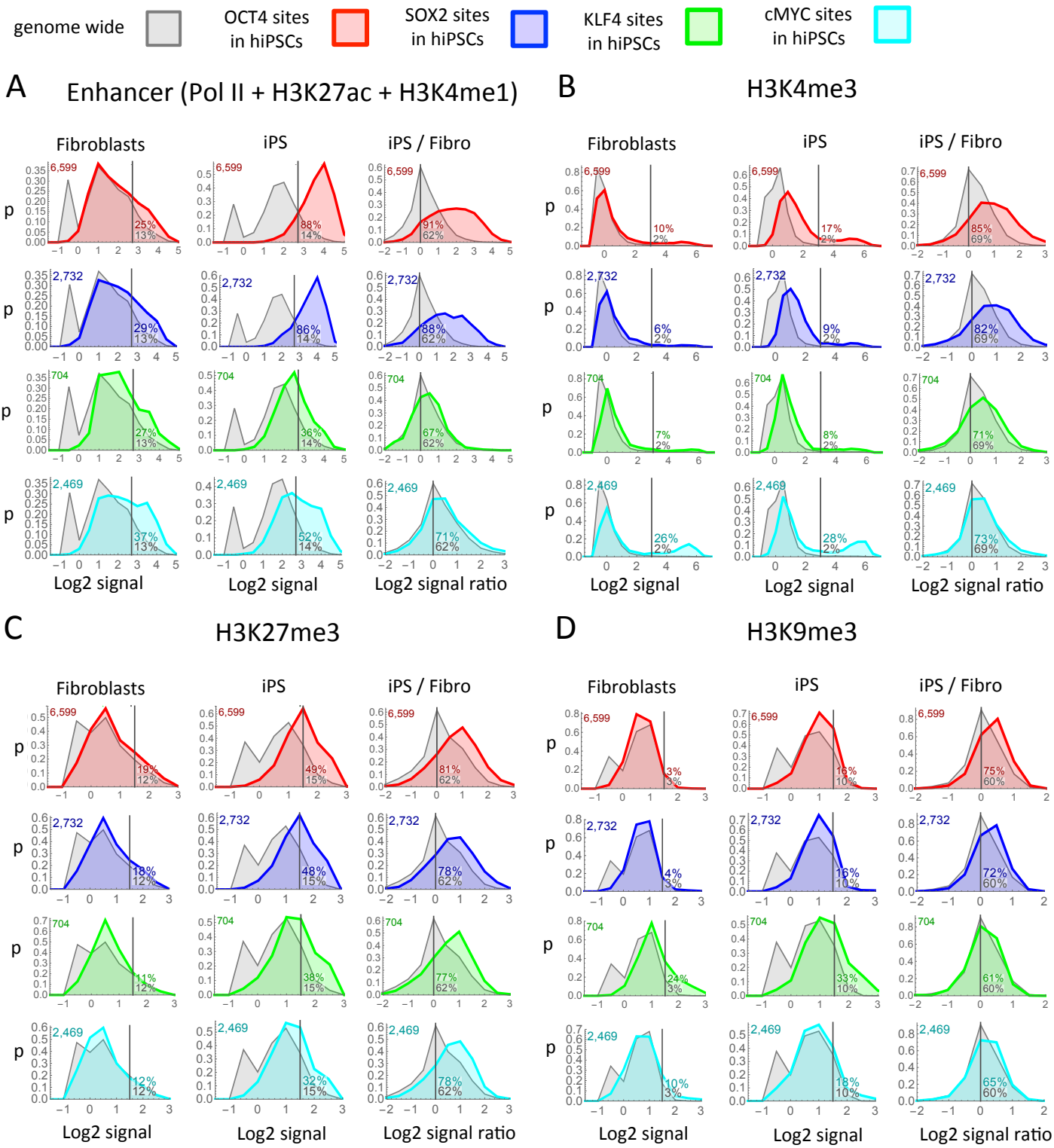
OSKM Day 5



Supplementary Figure S5. Distribution of signals at the site of TF binding in OSKM Day 5 cells. The distribution plots are as described in Figure 5 and S3. (**A**) Distribution plots show the enhancer signal, which is the sum of the Pol II, H3K27ac and H3K4me1 signal. Plots labeled 'Log2(Enhancer) < 2.7 in fibroblasts' show the distribution of enhancers signal at the sites of TF binding in OSKM that also have a Log2(enhancer signal) < 2.7 in fibroblasts *i.e.* at the sites of TF binding that did not have an active enhancer in the parental cells prior to binding (colored curves) or that have a Log2(enhancer signal) < 2.7 in fibroblasts genome wide (gray curve). Plots labeled 'Log2(Enhancer) >= 2.7 in fibroblasts' show the distribution of enhancers signal at the sites of TF binding in OSKM that also have a Log2(enhancer signal) >= 2.7 in fibroblasts *i.e.* at the sites of TF binding that did not have an active enhancer in the parental cells prior to binding (colored curves) or that have a Log2(enhancer signal) >= 2.7 in fibroblasts genome wide (gray curves). (**B**) Distribution plots show the signal labeled above each plot from cMYC bound sites in OSKM day 5 cells. The plots on the left side labeled 'Log2(H3K4me3) < 3 in OSKM' show the distribution of the labeled signal at the sites of cMYC binding and a Log2(H3K4me3 signal) < 3 in OSKM day 5 cells. The signals measured in the distribution plots were also from OSKM day 5 cells. The plots on the right side labeled 'Log2(H3K4me3) >= 3 in OSKM' show the distribution of the labeled signal at the sites of cMYC binding and a Log2(H3K4me3 signal) >= 3 in OSKM day 5 cells. The signals measured in the distribution plots were also from OSKM day 5 cells.

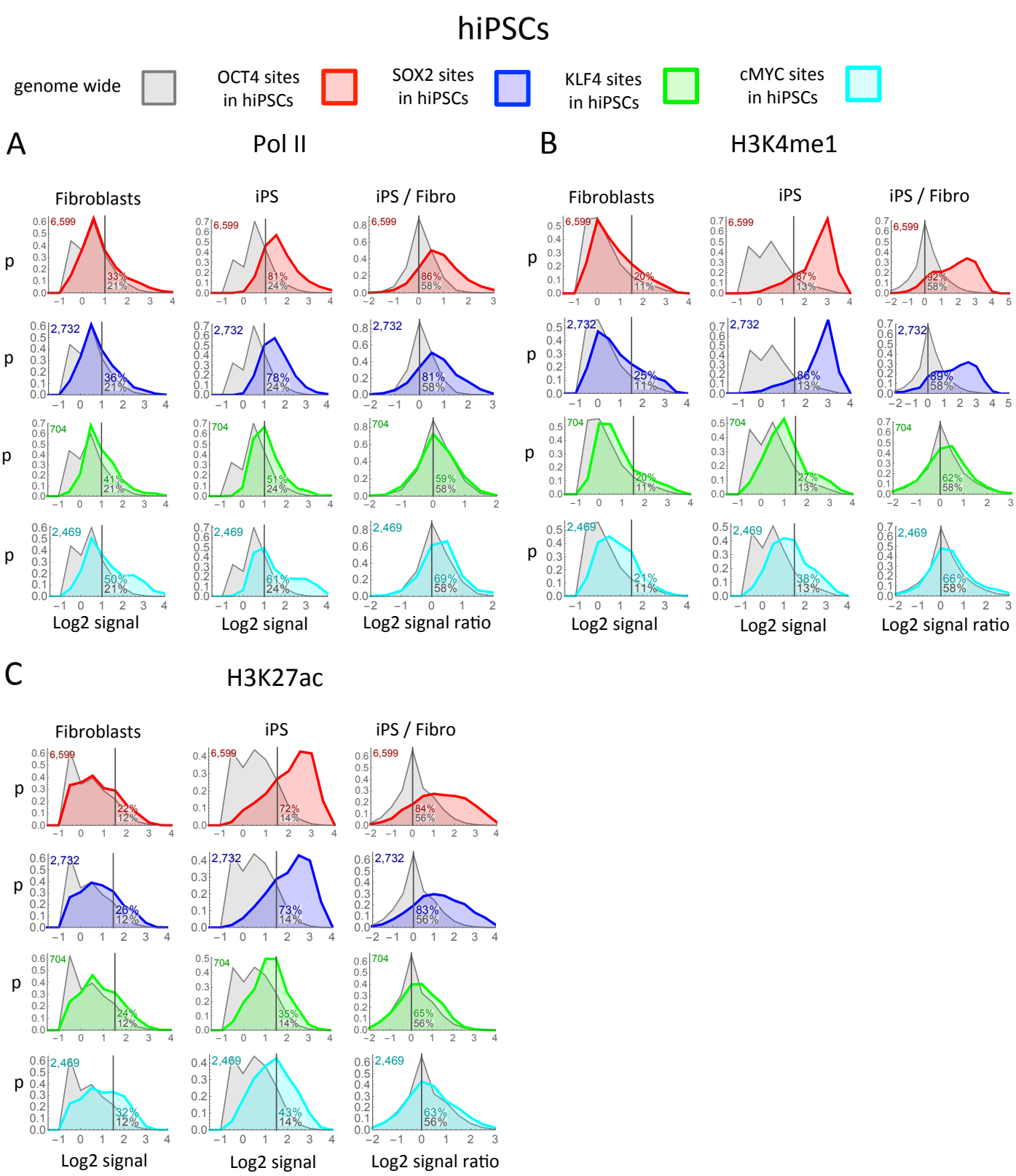
Supplementary Figure S6 - related to Figure 5

hiPSCs



Supplementary Figure S6. Distribution of signals at the site of TF binding in human iPS cells (hiPSCs).
The distribution plots are as described in Figure S3 except TF binding sites and signals were measured in hiPS cells instead of OSKM day 5 cells. Distribution plots show (**A**) the enhancer signal, which is the sum of the Pol II, H3K27ac and H3K4me1 signal, (**B**) the promoter signal H3K4me3, (**C**) the repressive signal H3K27me3 and (**D**) the repressive signal H3K9me3.

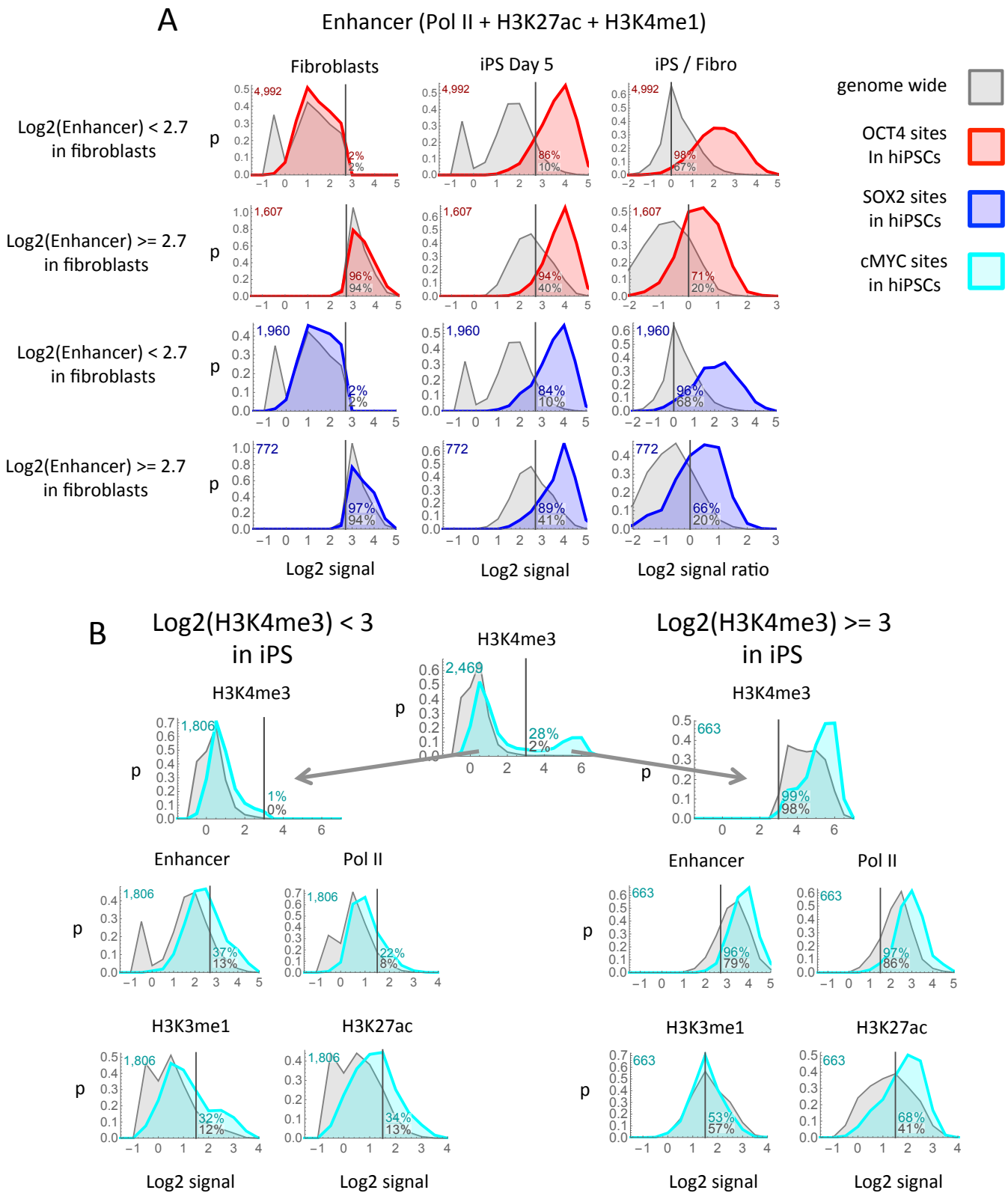
Supplementary Figure S7- related to Figure 5



Supplementary Figure S7. Distribution of signals at the site of TF binding in hiPSCs.
The distribution plots are as described in Figure S3 except TF binding sites and signals were measured from hiPS cells instead of OSKM Day 5 cells. Distribution plots show (**A**) the Pol II signal, (**B**) the H3K4me1 signal and (**C**) the H3K27ac signal.

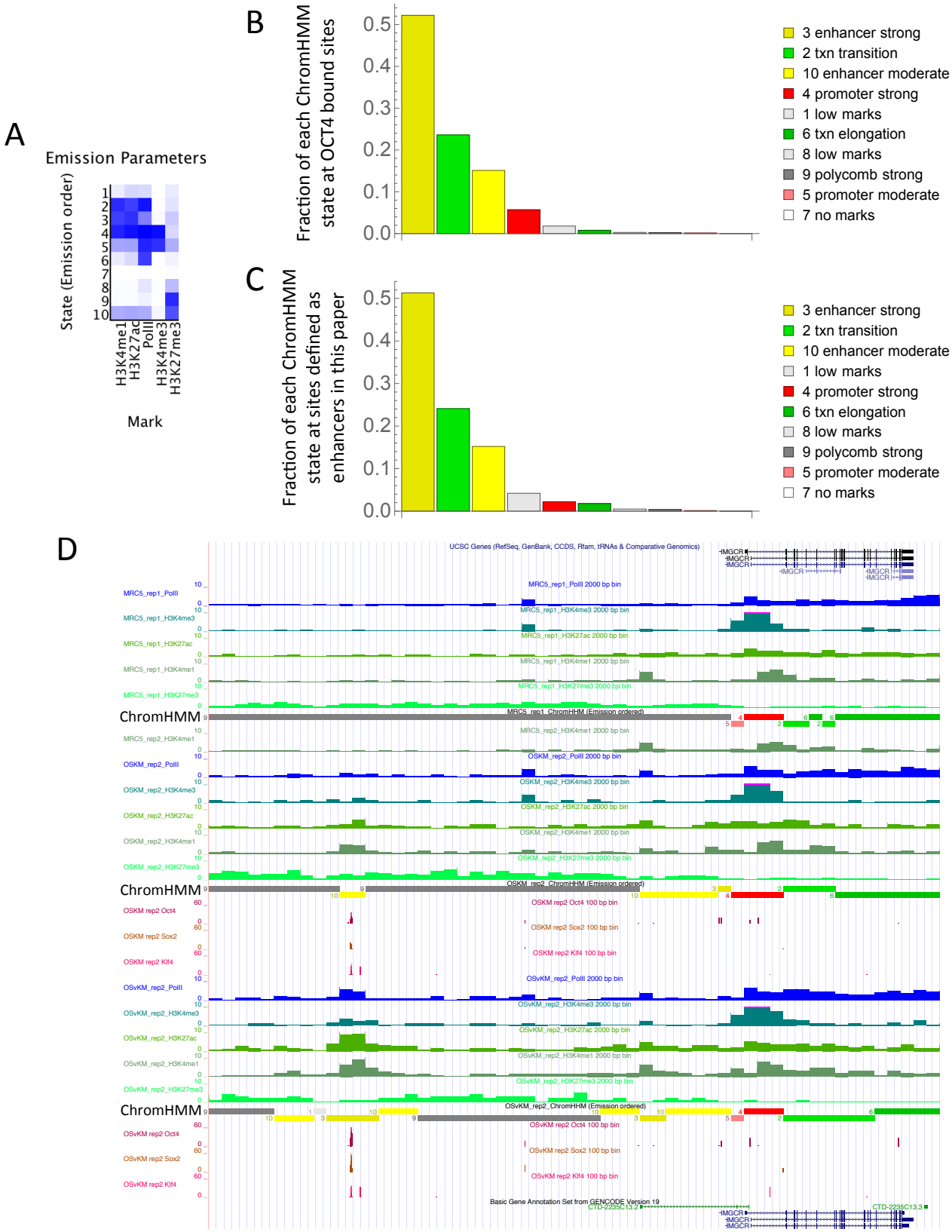
Supplementary Figure S8- related to Figure 5

hiPSCs



Supplementary Figure S8. Distribution of signals at the site of TF binding in hiPSCs.
The distribution plots are as described in Figure S5 except TF binding sites and signals were measured from iPS cells instead of OSKM day 5 cells.

Supplementary Figure S9- related to Figures 5, 6 and 3



Supplemental Figure S9. ChromHMMM analysis of ChIP-Endo-Seq data.
(**A**) The emission parameters from a ChromHMM analysis on the ChIP-Endo-Seq data from OSvKM day 5 cells. The analysis was performed seeking 10 states. (**B**) The fraction of each ChromHMM state at OCT4 bound sites OSvKM day 5 cells. (**C**) The fraction of each ChromHMM state at sites defined as enhancers in this paper in OSvKM day 5 cells. (**D**) The ChIP-Endo-Seq data tracks along with the ChromHMM tracks from fibroblasts (MRC5), OSKM and OSvKM day 5 cells from the same genomic region displayed in Figure 3.

Supplementary Table S1

WT or Fusion Protein	Tf (full length, FL) or DBD	Hinge	Transcriptional Regulatory Region	Repeats (for hinge and/or regulatory region)	Tandem fluorescent marker expressed	C-terminal tag
OCT4 WT and fusions						
OCT4 (WT) (O)	hOCT4 (FL)	-	-	-	Vex-GFP	-
OCT4-VP16 Weak (Ow)	hOCT4 (FL)	-	VP16 (425-455)	1	Vex-GFP	-
OCT4-VP16 Medium (Ov)	hOCT4 (FL)	-	VP16 (425-490)	1	Vex-GFP	-
OCT4-VP16 Strong (Os)	hOCT4 (FL)	-	VP16 (PADALDDFDLDML)	3	Vex-GFP	-
OCT4λVP2 (Ovp2)	hOCT4 (FL)	λ C1 (91-132)	VP16 (413-460)	2	Vex-GFP	-
OCT4λVP1 (Ovp1)	hOCT4 (FL)	λ C1 (91-132)	VP16 (413-460)	1	Vex-GFP	-
OCT4-HP1α (Oh)	hOCT4 (FL)	-	HP1α (FL)	1	Vex-GFP	-
OCT4-HP1β	hOCT4 (FL)	-	HP1b (FL)	1	Vex-GFP	-
OCT4-HP1γ	hOCT4 (FL)	-	HP1g (FL)	1	Vex-GFP	-
OCT4λEnR (Oe)	hOCT4 (FL)	λ C1 (91-132)	EnR (168-281)	1	Vex-GFP	-
OCT4-DBD	hOCT4 DBD (138-289)	-	-	-	Vex-GFP	-
OCT4-DBD (O4DBD)	hOCT4 DBD (138-289)	-	-	-	Vex-GFP	T7
SOX2 WT and fusions						
SOX2 (WT) (S)	hSOX2 (FL)	-	-	-	mCitrine	-
SOX2-VP16 Weak (Sw)	hSOX2 (FL)	-	VP16 (425-455)	1	mCitrine	-
SOX2-VP16 Medium (Sv)	hSOX2 (FL)	-	VP16 (425-490)	1	mCitrine	-
SOX2-VP16 Strong (Ss)	hSOX2 (FL)	-	VP16 (PADALDDFDLDML)	3	mCitrine	-
SOX2-HP1α (Sh)	hSOX2 (FL)	-	HP1α (FL)	1	mCitrine	-
SOX2-DBD	hSOX2 DBD (41-109)	-	-	-	mCitrine	-
SOX2-DBD (S2DBD)	hSOX2 DBD (41-109)	-	-	-	mCitrine	V5
KLF4 WT and fusions						
KLF4 (WT) (K)	hKLF4(FL)	-	-	-	mCherry	-
KLF4-VP16 Weak (Kw)	hKLF4 (FL)	-	VP16 (425-455)	1	mCherry	-
KLF4-VP16 Medium (Kv)	hKLF4 (FL)	-	VP16 (425-490)	1	mCherry	-
KLF4-VP16 Strong (Ks)	hKLF4 (FL)	-	VP16 (PADALDDFDLDML)	3	mCherry	-
KLF4-HP1α (Kh)	hKLF4 (FL)	-	HP1α (FL)	1	mCherry	-
KLF4-DBD	hKLF4 DBD (387-471)	-	-	-	mCherry	-
KLF4-DBD (K4DBD)	hKLF4 DBD (387-471)	-	-	-	mCherry	6xHis
cMYC WT and fusions						
cMYC (WT) (M)	hcMYC (FL)	-	-	-	mCerulean	
cMYC-VP16 Weak (Mw)	hcMYC (FL)	-	VP16 (425-455)	1	mCerulean	-
cMYC-VP16 Medium (Mv)	hcMYC (FL)	-	VP16 (425-490)	1	mCerulean	-
cMYC-VP16 Strong (Ms)	hcMYC (FL)	-	VP16 (PADALDDFDLDML)	3	mCerulean	-
cMYC-HP1α	hcMYC (FL)	-	HP1α (FL)	1	mCerulean	-
cMYC-DBD	hcMYC DBD (354-440)	-	-	-	mCerulean	-
cMYC-DBD (MDBD)	hcMYC DBD (354-440)	-	-	-	mCerulean	HA

Supplementary Table S1. Lentiviral constructs used in this study, Related to Figures 1- 6.
Lentiviral constructs encoding the WT or one or another fusion derivative of OCT4, SOX2, KLF4, and cMYC are listed. Each construct also expresses a fluorescent protein as indicated (vex-GFP, mCitrine, mCherry and mCerulean). The three VP16 fusions (Hammachi et al., 2012), are referred to as 'weak' (w), 'medium' (v), and 'strong' (s) depending on how efficiently they activate transcription in reconstruction experiments. The medium strength 'VP16' activating region is a 66 amino acid derivative of the *ca.* 80 amino acid VP16 carboxyl terminal domain (CTD) (Hammachi et al., 2012). The number of copies of the activation region in the corresponding fusion is indicated in the 'Repeats' column. The 'hinge' region (λC1), present in some constructs as indicated, encodes the linker region of the c1 lambda repressor. A C-terminal tag (T7, V5, 6xHis or HA tag) was fused to each of the four DBDs. FL: full length; EnR: Engrailed; HP1: Heterochromatin Protein 1 isoforms α, β, or γ.

Supplementary Table S2

Fibroblast ID	Age	Gender	Race	Source	Catalog #
MRC5	14 weeks	Male	Caucasian	ATCC	CCL-171
BJ	New born (NB)	Male	Unknown	ATCC	CRL-2522
GM01650	37y	Female	Caucasian	Coriell Cell Repositories	GM01650
GM00275	42y	Male	Caucasian	Coriell Cell Repositories	GM00275
AG02262	61y	Male	Caucasian	Coriell Cell Repositories	AG02262
GM03529	66y	Male	Black	Coriell Cell Repositories	GM03529
GM03524	67y	Female	Black	Coriell Cell Repositories	GM03524
GM01706	82y	Female	Caucasian	Coriell Cell Repositories	GM01706
GM00731	96y	Male	Caucasian	Coriell Cell Repositories	GM00731

Supplementary Table S2. Primary human fibroblasts used for reprogramming in this study, Related to Figure 1.

Supplementary Table S3

Cell type and Reprogramming Mix	MOI for each lentivirus in reprogramming mix	Reprogramming Efficiency (TRA1-81 positive colonies on day 25 per ~ 2500 cells/cm2 plated at day 5)		
		Total # colonies	Large colonies	Medium/Small colonies
cMYC variants				
OSKM	4,4,4,0.4	8	2	6
OSKMs	4,4,4,0.4	0.5 +/- 0.7	0	0.5 +/- 0.7
OSKMv	4,4,4,0.4	0.5	0	0.5
OSKMw	4,4,4,0.4	2.8 +/- 1	0	2.8 +/- 1
OSKMh	4,4,4,0.4	0	0	0
OSKM	4,4,4,1	7.5	0	7.5
OSKMs	4,4,4,1	0.75 +/- 1	0	0.75 +/- 1
OSKMv	4,4,4,1	2.3 +/- 1.8	0	2.3 +/- 1.8
OSKMw	4,4,4,1	3.5 +/- 1.4	0.5 +/- 0.7	3 +/- 0.7
OSKMh	4,4,4,1	0	0	0
OSKMs	4,4,4,4	0	0	0
OSKMv	4,4,4,4	0	0	0
OSKMw	4,4,4,4	0.3 +/- 0.4	0	0.3 +/- 0.4
OSKMh	4,4,4,4	0	0	0
OSKM	4,1,1,0.4	13 +/- 4.2	1 +/- 1.4	12 +/- 2.8
OSKMs	4,1,1,0.4	0	0	0
OSKMv	4,1,1,0.4	0	0	0
OSKMw	4,1,1,0.4	0	0	0
OSKMh	4,1,1,0.4	0	0	0
OSKM	4,1,1,1	9 +/- 5.7	2 +/- 2.8	7 +/- 2.8
OSKMs	4,1,1,1	0	0	0
OSKMv	4,1,1,1	0	0	0
OSKMw	4,1,1,1	0	0	0
OSKMh	4,1,1,1	0	0	0
OSKMs	4,1,1,4	0	0	0
OSKMv	4,1,1,4	0	0	0
OSKMw	4,1,1,4	0.5 +/- 0.7	0	0.5 +/- 0.7
OSKMh	4,1,1,4	0	0	0
OCT4 variants				
OSKM	0.4,1,1,0.4	1 +/- 1.4	0	1 +/- 1.4
OsSKM	0.4,1,1,0.4	0.5 +/- 0.7	0	0.5 +/- 0.7
OvSKM	0.4,1,1,0.4	0	0	0
OwSKM	0.4,1,1,0.4	0	0	0
Ovp2SKM	0.4,1,1,0.4	4 +/- 0	2 +/- 0	2 +/- 0
Ovp1SKM	0.4,1,1,0.4	8 +/- 1.4	5.5 +/- 0.7	2.5 +/- 2.1
OhSKM	0.4,1,1,0.4	0	0	0
OeSKM	0.4,1,1,0.4	0	0	0
OSKM	1,1,1,0.4	13.5 +/- 4.9	5.5 +/- 2.1	8 +/- 2.8
OsSKM	1,1,1,0.4	2 +/- 1.4	0.5 +/- 0.7	1.5 +/- 0.7
OvSKM	1,1,1,0.4	1 +/- 0	0	1 +/- 0
OwSKM	1,1,1,0.4	0	0	0
Ovp2SKM	1,1,1,0.4	4.5 +/- 2.1	2 +/- 0	2.5 +/- 2.1
Ovp1SKM	1,1,1,0.4	14 +/- 0	5 +/- 2.8	9 +/- 2.8
OhSKM	1,1,1,0.4	0	0	0
OeSKM	1,1,1,0.4	0	0	0
OSKM	4,1,1,0.4	33 +/- 21.2	21 +/- 18.4	12 +/- 2.8
OsSKM	4,1,1,0.4	1 +/- 0	0	1 +/- 0
OvSKM	4,1,1,0.4	0	0	0
OwSKM	4,1,1,0.4	5.5 +/- 2.1	2.5 +/- 0.7	3 +/- 2.8
Ovp2SKM	4,1,1,0.4	1.5 +/- 0.7	0	1.5 +/- 0.7
Ovp1SKM	4,1,1,0.4	16.5 +/- 7.8	10.5 +/- 4.9	6 +/- 2.8
OhSKM	4,1,1,0.4	0	0	0
OeSKM	4,1,1,0.4	0	0	0
SOX2 variants				
OSKM	4,0.4,4,0.4	4	0	4
OSsKM	4,0.4,4,0.4	10 +/- 0.7	5.5 +/- 0.7	4.5 +/- 0
OSvKM	4,0.4,4,0.4	4.3 +/- 1	1.5 +/- 0.7	2.8 +/- 0.4
OSwKM	4,0.4,4,0.4	6 +/- 0.7	2 +/- 0	4 +/- 0.7
OShKM	4,0.4,4,0.4	0	0	0
OSKM	4,1,4,0.4	2.5	0	2.5

OSsKM	4,1,4,0.4	29 +/- 7	16.5 +/- 7.8	12.5 +/- 0.7
OSvKM	4,1,4,0.4	20.3 +/- 5.3	11.3 +/- 4.6	9 +/- 0.7
OSwKM	4,1,4,0.4	13.5 +/- 0	3 +/- 2.8	10.5 +/- 2.8
OShKM	4,1,4,0.4	0	0	0
OSKM	4,4,4,0.4	1	0	1
OSsKM	4,4,4,0.4	14.6 +/- 0.7	5.3 +/- 1	9.3 +/- 1.8
OSvKM	4,4,4,0.4	129.5 +/- 30	77.8 +/- 15.5	51.7 +/- 30.6
OSwKM	4,4,4,0.4	26.3 +/- 4.6	13.8 +/- 6.7	12.5 +/- 2.1
OShKM	4,4,4,0.4	0	0	0
KLF4 variants				
OSKM	4,1,0.4,0.4	12.5 +/- 3.5	1.5 +/- 0.7	11 +/- 2.8
OSKsM	4,1,0.4,0.4	5.3 +/- 0.4	0	5.3 +/- 1.8
OSKvM	4,1,0.4,0.4	5.3 +/- 1.8	0.5 +/- 0.7	4 +/- 7.8
OSKwM	4,1,0.4,0.4	1.5 +/- 2	0	1.5 +/- 2
OSKhM	4,1,0.4,0.4	0.8 +/- 1	0	0.8 +/- 1
OSKM	4,1,1,0.4	20.1 +/- 5.7	2.5 +/- 2.2	17.6 +/- 4
OSKsM	4,1,1,0.4	3.5 +/- 1.1	0	3.5 +/- 1.1
OSKvM	4,1,1,0.4	2.3 +/- 0.3	0	2.3 +/- 0.4
OSKwM	4,1,1,0.4	3.5 +/- 0.7	0	3.5 +/- 0.7
OSKhM	4,1,1,0.4	0	0	0
OSKM	4,1,4,0.4	23.8 +/- 3.9	0.3 +/- 0.4	23.5 +/- 3.5
OSKsM	4,1,4,0.4	1.8 +/- 0.4	0	1.8 +/- 0.4
OSKvM	4,1,4,0.4	2 +/- 2.8	0	2
OSKwM	4,1,4,0.4	1.8 +/- 1	0	1.8 +/- 1
OSKhM	4,1,4,0.4	0.8 +/- 0.4	0	0.8 +/- 0.4
Other combinations				
OS	4,1	0	0	0
OSK	4,1,1	0	0	0
Ovp1Sv	1,0.4	0	0	0
Ovp1Sv	4,1	0	0	0
Ovp1SvK	1,0.4,1	38.5 +/- 4.9	0	38.5 +/- 4.9
Ovp1SvK	4,1,1	32.5 +/- 9.2	0.5 +/- 0.7	32 +/- 9.9
Ovp1SvKM	1,0.4,1,0.4	16.8 +/- 3.9	12.5 +/- 4.9	4.3 +/- 1
Ovp1SvKM	4,1,1,0.4	23.8 +/- 6	16 +/- 5.7	7.8 +/- 0.4
Mock viruses or DBDs				
4 mock viruses (vexGFP + mCit + mCherry + mCeru)	4,4,4,4	0	0	0
4 mock viruses (vexGFP + mCit + mCherry + mCeru)	10,10,10,10	0	0	0
O4DBD, S2DBD, K4DBD, MDBD	4,4,4,4	0	0	0
O4DBD, S2DBD, K4DBD, MDBD	10,10,10,10	0	0	0

Supplementary Table S3. Reprogramming Efficiencies with combinations of WT and fusion derivatives in fetal fibroblasts, Related to Figure 1.

Combinations of OSKM WT and fusion derivatives, at different multiplicities of infection (MOIs) as indicated, were used to reprogram fetal fibroblasts. To determine optimal MOIs for reprogramming fetal fibroblasts, three different MOIs (0.4, 1 and 4, each set alternatively highlighted in grey) of one factor were used while keeping MOIs of the other three factors constant. Plates were fixed and stained on day 25 to determine the number of colonies that were TRA1-81 positive (see Supplementary materials and methods). In the case of OCT4, VP16 and engrailed fusions separated by a hinge region (vp1 and vp2) were also used (see Supplementary Table S1). As controls, cells infected with all four mock lentiviruses (expressing each of the fluorescent protein alone) or lentiviruses expressing the DNA binding domains (DBD) of each of the factors (Chambers and Tomlinson, 2009; McConnell and Yang, 2010; Murre et al., 1989; Scholer et al., 1990) were used. Sizes of TRA1-81 positive colonies (see Supplementary materials and methods) were estimated using a phase contrast microscope as either large (*ca.* $\geq$ 50 cells / colony) or small to medium-sized colonies (*ca.* 25-50 / colony). ND = not determined.

Supplementary Table S4

#	Cell type and Reprogramming Mix	MOI for each lentivirus in reprogramming mix	Reprogramming Efficiency (TRA1-81 positive colonies on day 25-30 per ~ 2500 cells/cm2 plated on day 5)		
			Total # colonies	Large colonies	Medium/Small colonies
	Donor fibroblasts of differing ages (with cMYC)				
1	Fetal, OSKM	4,1,1,0.4	52.3 +/- 20.9	25.8 +/- 13.8	26.5 +/- 10
2	Fetal, OSvKM	4,4,4,0.4	129.5 +/- 30	77.8 +/- 15.5	51.7 +/- 30.6
3	New born (NB), OSKM	4,1,1,0.4	6 +/- 1.4	0.9 +/- 0.2	5.1 +/- 1.6
4	New born (NB), OSvKM	4,4,4,0.4	64 +/- 9	13 +/- 2.8	51.1 +/- 11.5
5	37y, OSKM	4,1,1,0.4	0	0	0
6	37y, OSvKM	4,4,4,0.4	79.5 +/- 11.3	24.5 +/- 0	55 +/- 11.3
7	42y, OSKM	4,1,1,0.4	0	0	0
8	42y, OSvKM	4,4,4,0.4	47 +/- 8.5	22.5 +/- 6.4	24.5 +/- 2.1
9	61y, OSKM	4,1,1,0.4	19.3 +/- 0.4	1.3 +/- 1.8	18 +/- 1.4
10	61y, OSvKM	4,4,4,0.4	64.8 +/- 11	21.3 +/- 1	43.5 +/- 10
11	66y, OSKM	4,1,1,0.4	0	0	0
12	66y, OSvKM	4,4,4,0.4	23	9	14
13	67y, OSKM	4,1,1,0.4	4.8 +/- 1	0	4.8 +/- 1
14	67y, OSvKM	4,4,4,0.4	6 +/- 1.4	0.5 +/- 0.7	5.5 +/- 0.7
15	82y, OSKM	4,1,1,0.4	0.5 +/- 0.7	0	0.5 +/- 0.7
16	82y, OSvKM	4,4,4,0.4	40.5 +/- 2.1	10.5 +/- 3.5	30 +/- 5.7
17	96y, OSKM	4,1,1,0.4	0	0	0
18	96y, OSvKM	4,4,4,0.4	45 +/- 1.4	19 +/- 1.4	26 +/- 2.8
	Donors fibroblasts of differing ages (without cMYC)				
19	Fetal, OSK	4,1,1	0	0	0
20	Fetal, OSvK	4,4,4	132.6 +/- 16.8	43.4 +/- 39.8	89.3 +/- 23
21	New born (NB), OSK	4,1,1	0	0	0
22	New born (NB), OSvK	4,4,4	281 +/- 62.2	16 +/- 1.4	265 +/- 63.6
23	37y, OSK	4,1,1	0	0	0
24	37y, OSvK	4,4,4	59 +/- 15.6	7 +/- 1.4	52 +/- 14
25	42y, OSK	4,1,1	1 +/- 1.4	0.5 +/- 0.7	0.5 +/- 0.7
26	42y, OSvK	4,4,4	325 +/- 4.2	28 +/- 8.5	297 +/- 4.2
27	61y, OSK	4,1,1	0.3 +/- 0.6	0	0
28	61y, OSvK	4,4,4	23 +/- 3.6	7.3 +/- 1.2	15.7 +/- 2.5
29	66y, OSK	4,1,1	43	0	43
30	66y, OSvK	4,4,4	164 +/- 29.7	24 +/- 1.4	140 +/- 31
31	67y, OSK	4,1,1	0	0	0
32	67y, OSvK	4,4,4	2.3 +/- 2.5	0	2.3 +/- 2.5
33	82y, OSK	4,1,1	0	0	0
34	82y, OSvK	4,4,4	139 +/- 28.3	27.5 +/- 4.9	111.5 +/- 23.3
35	96y, OSK	4,1,1	0.5 +/- 0.7	0	0.5 +/- 0.7
36	96y, OSvK	4,4,4	142.5 +/- 13.4	30.5 +/- 7.8	112 +/- 5.7

Supplementary Table S4. Reprogramming efficiencies of fibroblasts of different ages using OSKM, OSvKM, OSK and OSvK, Related to Figure 1.

The left column identifies the parental cell type and the reprogramming mix used. The other columns show the relative MOI's for each mix, and the total number and sizes of the resultant TRA1-81 positive colonies are shown in the last three columns. Some of the data shown in this table is taken from Figure 1 E-H. Sizes of TRA1-81 positive colonies were estimated using a phase contrast microscope as either large (*ca.* $\geq$ 50 cells / colony) or small to medium-sized colonies (*ca.* 25-50 cells / colony). ND = not determined. (also see Supplementary materials and methods).

Supplementary Table S5

Gene	Forward primer (5’ to 3’)	Reverse primer (5’ to 3’)
NANOG-5' ORF	ccgactgtaaagaatcttcacctat	atggatagttttcttcaggcccaca
NANOG-3'A ORF	ttccagccaaattctcctgccagtg	attaaggccttccccagcagcttccaa
NANOG-3'B ORF	taatacagcagaccactaggtattt	tagtttaggaataaatccatggttt
LIN28A ORF	ggtggagtattctgtattgggagtg	gcgcttctgcatgctctttcctttt
TRIM71 ORF	aagtttggtgctcaaggcagcggct	gatggcgatgccggaagggcggtcc
RPL7 ORF	cccttaagaaaaagcgaaggaattt	acttctttctcaggcgcttgatctt

Supplementary Table S5. Primers used for gene expression analysis by RT-QPCR, Related to Figure 2.