EXTENDED EXPERIMENTAL PROCEDURES

Protein expression and purification

The bacterial expression plasmids pET-28B-huOct4, pET-28B-huSox2, pET-28B-huKlf4, and pET-28B-huMyc encode the human O, S, K, and M, respectively, fused to an N-terminal 6X histidine tag. The O, S, K, and M cDNA sequences were generated by PCR from the respective lentiviral constructs used for generating human iPS cells (Hockemeyer et al., 2008), introducing a NotI and EcoRI restriction sites for inserting into the pET-28B plasmid. The histidine-tagged O, S, K, and M proteins were expressed in *E. Coli* Rosetta (DE3) pLysS (Novagen # 70956-3). Transformed cells were grown at 37 °C to a density of 0.5 at $A_{600\ nm}$ and protein expression was induced by 0.5 mM IPTG for 4 hr for Oct4, 2 hr for Sox2, and overnight for Klf4 and c-Myc at 30 °C. The proteins were purified over Hi-trap HP nickel-charged columns (GE healthcare # 17-5248-01) under denaturing conditions. The purified proteins Oct4 and Sox2 were refolded by initially dialyzing to 2 M Urea in 2 M increment gradients and then to 0 M Urea using a desalting column (GE healthcare # 17-1408-01). The purified denatured Klf4 was refolded by dialyzing to 2 M Urea in 2 M increment gradients and refolded by diluting directly to 1 µM concentration in DNA binding buffer (10 mM Tris-HCl pH 7.5, 1 mM MgCl2, 10 uM ZnCl2, 10 mM KCl, 1 mM DTT, 5% glycerol, 0.5 mg/ml BSA). The c-Myc:Max heterodimer complex was reconstituted as described previously (Farina et al., 2004). The recombinant human histones were expressed and purified as described previously (Tanaka et al., 2004).

The mammalian expressed human O, S, K, and M recombinant proteins were obtained from OriGene (Oct4 # TP311998, Sox2 # TP300757, Klf4 # TP306691, c-Myc #

TP301611). The DDK-tagged mammalian proteins were expressed in HEK293 cells (human embryonic kidney cells) and purified under native conditions using anti-DDK affinity column followed by conventional chromatography steps.

Due to the presence of contaminants (Figure 1A), the mammalian protein concentrations were calculated by quantifying the intensity of each of the O,S,K,M bands running at the expected sizes in SDS-PAGE (without including the contaminants) and comparing it with their respective bacterial counterparts; the latter having been quantitated by direct protein measurements using absorbance at 280 nM. To reduce error the band intensities were quantified under variable concentrations.

Nucleosome preparation

The 162 bp *LIN28B* DNA fragment corresponds to the genomic location:

hg18-chr6: 105,638,004-105,638,165

AGTGGTATTAACATATCCTCAGTGGTGAGTATTAACATGGAACTTACTCCAACAATA

CAGATGCTGAATAAATGTAGTCTAAGTGAAGGAAGAAGGAAAGGTGGGAGCTGCC

ATCACTCAGAATTGTCCAGCAGGGATTGTGCAAGCTTGTGAATAAAGACA

The DNA sequence was created by PCR with end-labeled primers (see below for sequences). The 162 bp fluorescent-tagged DNA fragments were gel extracted and further purified using ion-exchange liquid chromatography on a Mono-Q column and 2 M salt step gradient. The nucleosomes were reconstituted as described previously (Tanaka et al., 2004). Briefly, 10 µg of Cy5 or FAM end labelled PCR fragment of *LIN28B* DNA was mixed with purified and refolded H2A/H2B dimers and H3/H4 tetramers at a 1:1 DNA:Histone-octamer molar ratio in 10 mM Tris-HCl pH8, 5 M Urea,

2 M NaCl, 1 mM EDTA, 0.1 mg/ml BSA. The nucleosomes were assembled using salt-urea gradient by dialyzing against a solution containing 2, 1.5, 1, 0.8, and then 0.6 M NaCl and 10 mM Tris–HCl pH 8.0, 5 M Urea, 1 mM EDTA, 10 mM 2-mercaptoethanol for 4 hr in each buffer at 4 °C. The nucleosomes were then dialyzed against a no Urea buffer containing 0.6 M NaCl and 10 mM Tris–HCl pH 8.0, 1 mM EDTA, 1 mM 2-mercaptoethanol and then the same buffer containing 0.1 M NaCl for 8 hrs at 4 °C. The reconstituted nucleosomes were heat shifted by incubating at 37 °C for 6 hr.

Binding reactions

The end-labelled oligonucleotides containing specific or non-specific sites (see below for sequences), *LIN28B*-DNA, and *LIN28B*-nucleosomes were incubated with recombinant proteins in DNA-binding buffer (10 mM Tris-HCl (pH7.5), 1 mM MgCl2, 10 µM $ZnCl_2$, 1 mM DTT, 10 mM KCl, 0.5 mg/ml BSA, 5% Glycerol) at room temperature for 60 min. Free and bound DNA were separated on 4% non-denaturing polyacrylamide gels run in 0.5X Tris–borate–EDTA and visualized using a PhosphorImager using Cy5 fluorescence setting (excitation at 633 nm and emission filter 670 BP 30) and high sensitivity setting. The apparent dissociation constant (K_d) was calculated in two ways:

1. Total K_d was calculated to quantify the total affinity of each protein to DNA accounting for both specific and non-specific binding by assuming that the amount of nonspecific binding is linearly proportional to the concentration of protein used. Total amount of DNA was quantified from Cy5 fluorescence of the free DNA band at 0 nM protein concentration. The amount of free DNA at each protein concentration was determined from the intensity of Cy5 fluorescence of

the free DNA bands using Multi-Gauge software (Fujifilm Science lab). The fraction of bound DNA was calculated using the equation below:

$$\text{fraction } DNA_{bound} = 1 - (\frac{DNA_{free}}{DNA_{total}})$$

Binding curves describing the fraction of bound DNA as a function of protein concentration [TF] from two separate experiments were fitted to the data using nonlinear regression in GraphPad Prism software (version 6.04 for windows). The goodness of the fit was assessed using an R^2 greater than 0.97. The K_d for each protein was calculated using the equation below and fixing Bmax to a maximum of 1.

$$\text{fraction } DNA_{bound} = \frac{B_{max} \times [TF]}{K_d + [TF]} + (NS \times [TF])$$

where NS is the slope of nonspecific binding.

2. Specific K_d was calculated to quantify the specific affinity of each protein to DNA not accounting for non-specific binding. The amount of free DNA and bound DNA at each protein concentration was determined from the intensity of Cy5 fluorescence of the free DNA and the first DNA-TF complex bands using Multi-Gauge software (Fujifilm Science lab). The fraction of bound DNA was calculated using the equation below:

$$\text{fraction } DNA_{bound} = \frac{DNA_{bound}}{DNA_{free} + DNA_{bound}}$$

Binding curves describing the fraction of bound DNA as a function of protein concentration [TF] from two separate experiments were fitted to the data using nonlinear regression in GraphPad Prism software (version 6.04 for windows). The goodness of the fit was assessed using an R^2 greater than 0.97. The K_d for each protein, which is determined as the protein concentration at half-maximum DNA binding was calculated using the equation below and fixing Bmax to a maximum of 1.

$$\text{fraction } DNA_{bound} = \frac{B_{max} \times [TF]}{K_d + [TF]}$$

For competition assays excessive amounts (from 5 to 40 fold) of non-labelled probes containing specific and non-specific sites were added to the binding reaction and incubated for 60 min at room temperature to reach equilibrium. The binding reactions were loaded on the 4% EMSA gels as described above. EMSA gels were run at 80 volts at room temperature. As specific competitors, the following DNA probes were used: *FGF4* promoter for Oct4 and Sox2, LEFTY1 promoter for Klf4, and *CDKN2D* promoter for c-Myc. As non-specific competitors the following DNA probes were used: NS for Oct4 and Sox2, *NANOG* promoter for Klf4 and c-Myc. See below for the DNA probes sequence.

DNase footprinting

DNase footprinting reactions were carried out by incubating the 6-FAM end labelled *LIN28B* free (50 ng) in the presence or absence of the purified TFs or histone octamers with 0.006 (DNA or DNA+TFs) or 0.06 (nuc. or nuc.+TFs) unit of DNase-I (Worthington) in 50 µl DNA-binding buffer (10 mM Tris-HCl (pH7.5), 1 mM $MgCl_2$, 10 µM $ZnCl_2$,

1 mM DTT, 10 mM KCl, 0.5 mg/ml BSA, 5% Glycerol) supplemented with additional 50 µl 10 mM $MgCl_2$ and 5 mM $CaCl_2$ at 25 °C for 1 min. The reaction was stopped by adding 90 µl (200 mM NaCl, 30 mM EDTA, 1% SDS) and chilling on ice for 10 min. One tenth of reaction volume (~20 µl) of 3 M NaOAc (pH 5.2) was added to the reaction before the DNA fragments were extracted with Phenol–chloroform extraction. The DNA fragments were further purified using MinElute PCR purification kit (Qiagen) and eluted in 10 µl dH_2O. The digested DNA fragments were separated by capillary electrophoresis as described previously (Zianni et al., 2006). Briefly, the digested DNA fragments (5 µl) were added to 4.9 µl HiDi formide (Applied Biosciences) and 0.1 µl GeneScan-500 LIZ size standards (Applied Biosciences). After denaturing at 95 °C for 10 min, the samples were run on an ABI 96-capillary 3730XL Sequencer, using G5 dye setting, running a genotyping module with an increased injection time of 30 sec and injection voltage of 3 kV. The generated electropherograms were analyzed using the peak scanner software (Applied Biosciences) and PeakStudio V 2.2.

Western blotting after EMSA (WEMSA)

The EMSA was carried out as described above with 10 fold more protein and nucleosomes and run on a 1.5 mm thick mini-gel cassette (life technologies # NC2015) containing 5% polyacrylamide gel. To avoid Cy5 fluorescence saturation, 90% of the nucleosome used in binding reaction was not labelled. The gel was then visualized using Cy5 fluorescence as described above. To charge the proteins, the gel was incubated for 2 hr in denaturing buffer (1% SDS, 375 mM Tris-HCl pH 7.5) at 20 ºC. The proteins were transferred to a 0.22 µm Sequi-Blot™ PVDF membrane (Bio-Rad) using

NuPAGE transfer buffer (life technologies # NP0006), supplemented with 0.1% SDS and 20 % methanol for 1 hr at 100 Volts at 4 ºC.

The proteins were fixed to the membrane by incubating in 10% Glacial Acetic Acid for 15 min at room temperature. The membranes were blocked with PBS-0.1% Tween containing 10% non-fat dry milk overnight at 4 ºC. The primary antibody incubations with anti-human Oct4 antibody (0.5 µg/ml; Abcam # ab19857), human Sox2 antibody (1 µg/ml; R&D systems # AF2018), human KLF4 antibody (0.5 µg/ml; R&D systems # AF3640), human c-Myc antibody (1 µg/ml; R&D systems # AF3696), anti-human H3 (0.5 µg/ml; abcam # ab1791), and anti-human H2B (0.8 µg/ml; abcam # ab1790) were performed for 2 hr at room temperature. The secondary antibody incubations with goat anti-rabbit IgG-HRP (1:5000 dilution; Santa Cruz # sc-2004) and donkey anti- goat IgG-HRP (1:2000; Santa Cruz # sc-2020) were performed for 1 hr at room temperature. Blots were visualized by using SuperSignal West Pico chemiluminescent substrate (Thermo-Scientific # 34080) in Fujifilm LAS-4000 imaging system. The membranes were stripped by incubating with Restore Western-Blot Plus Stripping Buffer (Thermo-Scientifc # 46430) for 30 min at RT and re-blocked after blotting with each antibody. The same membrane was serially blotted and stripped with all antibodies shown.

Genomic data analysis

The O, S, K, and M ChIP-seq aligned data along with the called peaks (FDR-controlled at 0.005) were obtained from the GEO database (GSE36570) (Soufi et al., 2012). The MNase-seq data (GSM543311) (Kelly et al., 2012) were aligned to build version NCBI36/HG18 of the human genome and seven replicates were pooled together

generating 145,546,004.00 unique reads. The MNase-seq reads were extended to 150 bp to cover one nucleosome and thus resulting in 6.6 fold genome coverage.

To identify regions bound by single factors, we separated peaks if their centers were at least 500 bp apart from each other. Sites bound by all possible OSKM combinations were merged if their peak centers were within 100 bp or less from each other. Regions spanning 1 kb upstream and downstream from the center of the curated peaks were divided into 10 bp bins (n = 200). Tag counts from O, S, K, and M ChIP-seq and MNase-seq were assigned to each corresponding bin and used as a measure for enrichment. The curated genomic locations were organized in ascending rank-order according to the tag counts from the central 20 bins (200 bp) as described in the text. Sites were considered to be nucleosome-depleted if their central 200 bp tag counts were smaller than that of the average 200 bp flanking regions (ratio < 1).

bHLH factors ChIP-seq data were obtained from GEO with the accession code GSE43916 for Ascl1 (Wapinski et al., 2013), GSM1167583 for Tal1, GSM1167584 for Mitf (Calero-Nieto et al., 2014), GSM1306365 and GSM1306367 for Olig2 (Suvà et al., 2014), GSM751036 for NeuroD (Tennant et al., 2013), and GSE50415 for MyoD (MacQuarrie et al., 2013). MNase-seq data for MEFs were obtained from GSM1004654 (Teif et al., 2012). The ChIP-seq and MNase-seq data for the above factors were processed as described for OSKM in human fibroblasts. The mouse sequencing data were aligned to the Mouse genome built mm9, accordingly.

Motif analysis

For *de novo* motif discovery, we used Discriminative DNA Motif Discovery algorithm (DREME) (Bailey, 2011). We focused on motifs occurring at the central

200 bp of O, S, or K peaks, using central motif enrichment analysis (CentriMo) (Bailey and Machanick, 2012). We quantified the occurrences of the first hits that returned with the most statistical significance within the O, S, K, and M sites using Find Individual Motif Occurrences (FIMO) (Grant et al., 2011). Motifs that showed most central enrichment were considered. Moreover, the newly discovered motifs were compared to the JASPAR and UniPROBE motif databases using the Motif comparison tool (TOMTOM) (Gupta et al., 2007; Mathelier et al., 2014; Newburger and Bulyk, 2009). The above tools are part MEME-ChIP suit v.4.9.1 (Machanick and Bailey, 2011), available at http://meme.nbcr.net.

Molecular modelling

We have modeled the macromolecular motions that take place during the initial recognition of O, S, K, and M DBDs to their binding sites using the MORPH server as described previously (Krebs, 2000). Briefly, we used the DNA-free structures of O, S, K, and Myc:Max DBDs as the initial state and the DNA-DBDs complexes structures as the final state (see below for PDB ids used). Based on adiabatic mapping, the possible states accommodating the conformational space between free and bound states were calculated within the energy barriers constraints. By defining a set of hinges, the protein motion describing the rigid-body rotation of a small part “core” in relation to a larger part was directly linked protein flexibility. DNA flexibility was not accounted for in our molecular dynamics. The DNA-free states of Oct4-POU_S, Oct4-POU_{HD}, Klf4-3ZFs, and c-Myc-bHLH were built based on their sequence homology (92%, 85%, 93%, and 89%) to the experimental NMR structures of Oct1-POU_S, Oct1-POU_{HD}, Klf5-3ZFs and Max-bHLH, respectively (PDBs: 1POU, 1POG, 2EBT, and 1R05) (Assa-Munt et al., 1993;

Cox et al., 1995; Sauvé et al., 2004) using Modeller program (Sali and Blundell, 1993). The DNA-free structure of Sox2-HMG has been solved using NMR and submitted to the protein databank under the PDB id 2LE4. The structures of DNA in complex with Oct4-POU, Sox2-HMG, Klf4-2ZFs, Klf4-3ZFs and Myc:Max-bHLH (PDBs: 1GT0, 2WBS, 2WBU, and 1NKP) were solved using X-ray crystallography (Esch et al., 2013; Nair and Burley, 2003; Reményi et al., 2003; Schuetz et al., 2011). The Sox2-HMG N46Q mutant was modelled based on its sequence homology (94%) to hSRY-HMG mutant (PDB: 1J47) (Murphy et al., 2001) as described above. The DNA bound to Sox2 wt or Sox2 N46Q mt was superimposed on the nucleosomal DNA obtained from PDB-3LZ0 (Vasudevan et al., 2010) using the super command from Pymol (Version 1.5.0.1 Schrödinger, LLC) and the RMSD was calculated using the rms_curr command between the phosphate backbone carbon atoms. The DNA accessible surface area (ASA) exposed to solvent was calculated from free-DNA or bound to Oct4-POU_S, Oct4 POU_{HD}, Oct4-$POU_{S\text{-}HD}$, Klf4-2ZFs, and Klf4-3ZFs from the corresponding crystal structures using areaimol from the CCP4 package (Lee and Richards, 1971).

The X-ray crystal structures of Mitf (4ATK, 4ATI), NeuroD (2QL2), MyoD (1MDY), and Tal1 (2YPB) were obtained from the RCSB protein data bank (Longo et al., 2008; Ma et al., 1994; El Omari et al., 2013; Pogenberg et al., 2012). The Ascl1 and Olig2 structures were obtained from the SWISS-Model server based on their sequence homology to NeuroD (2QL2) (Kiefer et al., 2009; Kopp and Schwede, 2006). The images used in the figures were ray-traced and created using the PyMOL molecular graphics system (Version 1.5.0.1 Schrödinger, LLC).

DNA-binding sites. The DNA oligonucleotides used as binding sites (top and lower strands) are shown below. The expected sites are highlighted in blue, red, orange and green for O, S, K, and M, respectively. The Cy5 5'-end-labelled oligonucleotides were obtained from IDT (Integrated DNA Technologies). The double stranded probes were generated by annealing the single strands using the following reaction: 1 nano-moles of each strand (10 µl of 100 µM) were mixed in 50 µl final volume annealing buffer (20 mM Tris-HCl pH7.6, 50 mM NaCl, 0.1 mM DTT, 1 mM EDTA). The reaction was incubated at 70 ºC for 10 min, and slowly cooled at room temperature overnight.

name_TF	Upper strand	Lower strand
FGF4	TTTAAGTATCCCATTAGCATCCAAACAAAGAGTTTTC	GAAAACTCTTTGTTTGGATGCTAATGGGATACTTAAA
NANOG	CTTACAGCTTCTTTTGCATTACAATGTCCATGGTGGA	TCCACCATGGACATTGTAATGCAAAAGAAGCTGTAAG
NS	CTGCAGGTGGGATTAACTGTGAATTCA	TGAATTCACAGTTAATCCCACCTGCAG
lEFTY	GAGCTCCCAGGAGGTCCCAGGGGTGTGACCTCTCT	AGAGAGGTCACACCCCTGGGACCTCCTGGGAGCTC
CDKN2D	AGGAGCCTGCAGCTGCCACGTGGGAAGGCCTGAGA GGACATAGT	ACTATGTCCTCTCAGGCCTTCCCACGTGGCAGCTGCA GGCTCCT

PCR primers: The DNA oligonucleotides used as primers for PCR to generate the *LIN28B* sequence (162 bp) from human genomic DNA are shown. The chemical modifications at the end of each oligonucleotide are also shown. All the DNA oligonucleotides were obtained from IDT (Integrated DNA Technologies).

Sequence Name	Bases	Sequence	Modifications and Services
lin28B- FWD	27	AGT GGT ATT AAC ATA TCC TCA GTG GTG	Standard Desalting
Cy5-lin28B-FWD	27	/5Cy5/AGT GGT ATT AAC ATA TCC TCA GTG GTG	5' Cy5 HPLC Purification
6-FAM-lin28B-FWD	27	/56-FAM/AGT GGT ATT AAC ATA TCC TCA GTG GTG	5' 6-FAM Standard Desalting

lin28B-RVS	25	TGT CTT TAT TCA CAA GCT TGC ACA A	Standard Desalting
Cy5-lin28B-RVS	25	/5Cy5/TGT CTT TAT TCA CAA GCT TGC ACA A	5' Cy5 HPLC Purification
6-FAM-lin28B-RVS	25	/56-FAM/TGT CTT TAT TCA CAA GCT TGC ACA A	5' 6-FAM Standard Desalting

SUPPLEMENTAL FIGURE LEGENDS

Figure S1. Recombinant Oct4, Sox2, Klf4, and c-Myc show specific DNA-binding activities in vitro. Related to Figure 1.

(A) Schematic diagram showing the DNA-binding domains organization of the full-length Oct4, Sox2, Klf4, c-Myc and Max (O, S, K, M and X) proteins. The number of amino acids is indicated.

(B) Representative EMSAs showing the affinity of increasing amounts of recombinant bacterial (bact.) and mammalian (mamm.) –expressed O, S, K, and M proteins to Cy5-labelled DNA probes containing their respective canonical binding sites. The concentrations used for each protein (nM) are indicated above each lane. Lanes are numbered underneath each gel. Black arrows indicate the migration of Free DNA and DNA-protein complexes. DNA sequences of the Cy5-labelled probes are shown in the Extended Experimental Procedures.

(C) Same as in (B) but showing EMSAs with DNA probes containing non-specific sequences for each protein.

(D) SDS-PAGE and coomassie staining showing the homogeneity of the recombinant human histones (H2A, H2B, H3, and H4) purified from bacteria under denaturing conditions (left panel) and then refolded to H2A/H2B dimers and H3/H4 tetramers (middle-panel). EMSA (right panel) showing free Cy5-labelled *LIN28B* DNA and Cy5-labelled *LIN28B* DNA assembled to nucleosomes in vitro by salt gradient dilution with the refolded H2A/H2B dimers and H3/H4 tetramers. *LIN28B* DNA was generated by PCR using the primers shown in Extended Experimental Procedures. DNA was visualized using Ethidium-Bromide staining (Et-Br) and Cy5 fluorescence (Cy5) as

indicated underneath each gel in the right panel. The sizes of protein standards in kDa and DNA standards in bp are shown.

Figure S2. Recombinant Oct4, Sox2, Klf4, and c-Myc show a range of specificity to free DNA versus nucleosomal DNA in vitro. Related to Figure 2.

(A) Representative EMSA showing the affinity of recombinant bact. O, S, K, M-X and X proteins (1 nM) to Cy5-labelled probes (2 nM) containing canonical sites in the presence of 40 fold molar excess of specific non-labelled competitor (s) or non-specific non-labelled competitor (n) or absence of competitor (-).

(B—D) Representative EMSAs showing the affinity of recombinant O, S, K, and M proteins (bact. top panels and mamm. bottom panels) (1 nM) to Cy5-labelled *LIN28B* free DNA (lanes 1—8) and Cy5-labelled *LIN28B* nucleosomal (nuc.) DNA (lanes 9—16) (2 nM) in the presence of 5, 10 and 20 fold molar excess of non-labelled specific competitor (s) (lanes 3—5 and 11—13) or non-labelled non-specific competitor (n) (lanes 6—8 and 14—16) or absence of competitor (-) (lanes 2 and 10). Concentrations of competitors in nM are indicated above each lane. Lanes are numbered underneath each gel. Full black arrow heads indicate free and TF-bound *LIN28B*-DNA, and white arrow heads indicate free and TF-bound *LIN28B* nucleosomal DNA. Brown boxes show Klf4-*LIN28B*-nucleosome complexes under prolonged exposure.

Figure S3. Oct4, Sox2, Klf4, and c-Myc show a range of affinity and specificity to nucleosomes in vivo. Related to Figure 3.

(A) The O, S, K, and M ChIP-seq peaks at 48 hr post induction in human fibroblasts were called using different FDR thresholds to show the extent of non-specific DNA binding for each factor in vivo. The plots are color coded as indicated.

(B) Nucleosome enrichment as measured by MNase-seq in human fibroblasts within regions bound by O, S, K, and/or M combinations at 48 hr post induction in fibroblast. The bottom and top of the box represent the 25th and 75th percentile and the middle band is the 50th percentile of the MNase-seq value; whisker ends represent the min and max values. Outlier values are eliminated.

(C—M) Read density heatmaps (top panels) in red color scale (0—20) showing the intensity of MNase-seq tags, spanning ± 1 kb from the center of the O, S, K, and/or M peaks where the factors bind within 100 bp or less from each other. The number of targeted sites is indicated above. Metaplots (bottom panels) showing the average nucleosome enrichment (MNase-seq tags) within the same O, S, K, and/or M sites shown above but separated into TSS-proximal (red) and TSS-distal groups (blue). Sites that were within 1 kb to the nearest TSS were considered proximal, while sites that were more than 1 kb away from the nearest TSS were considered distal.

Figure S4. Oct4, Sox2, Klf4, and c-Myc bind to nucleosome containing H2A and H3 histones. Related to Figures 3.

~~(A)~~ Representative EMSAs showing the binding of Oct4 (1 nM) on its own and in combination with Sox2, Klf4, and c-Myc-Max (3 nM) (left panels). The proteins from EMSA were transferred onto a PVDF membrane (WEMSA) and blotted for H3, H2B, Oct4, and/or Sox2, Klf4 and Myc as indicated (the three panels on the right). Black arrow heads indicate the observed TF-nucleosome complexes.

Figure S5. The apparent flexibility of Oct4, Sox2, Klf4, and c-Myc DBDs correlate with their nucleosome binding compatibility. Related to Figure 4.

(A) DNase-I footprinting showing the protection of *LIN28B*-DNA in the absence (blue lines) or presence (red lines) of Klf4. Electropherograms of 5'-6FAM end-labeled *LIN28B* oligonucleotides generated by DNase-I digestion of DNA (0.006 U) of the top-strand (top panel) or the bottom-strand (bottom panel). Dashed boxes represent specific sites protected by Klf4 and the corresponding sequence is indicated underneath.

(B, C, and D) Cartoon representation showing the three dimensional structures of O, S, and K free of DNA as determined by NMR. All the NMR-determined 3D states are aligned and shown by transparent colors to indicate the measured flexibility of the free DBDs. The PDB ids of each structure are indicated.

(E, F, and G) The gradual transition of O, S, and K DBDs from DNA-free to DNA-bound was measured by morphing (Extended Experimental Procedures). Arrows and color transparency indicate the extent of the apparent flexibility of each DBD. The used color scheme is shown at the bottom.

Figure S6. The HLH factors compatibility with nucleosomes correlates with central degenerate E-box motif. Related to Figures 5 and 6.

(A) Same as Figure S5 (B, C, and D) for Myc-Max bHLH hetero dimer not bound to DNA.

(B) Same as Figure S5 (E, F, and G) for the transition of Myc-Max bHLH from DNA-free to DNA-bound states.

(C) Pair-wise sequence alignment (left panel) of the basic region of Mitf and c-Myc showing identical amino-acids (*) and highly similar amino acids (**:**, **.**). Cartoon

representations of Mitf bHLH in complexes with DNA containing the canonical E-box motif (middle panel) and centrally degenerate E-box (right panel). The motifs bound and PDB ids are indicated above. The cyan and pink arrows represent the position of the exposed nucleotides within the central E-box motif not making base-contacts with the relative bHLH conformation. The central two nucleotides (CANNTG) are colored in purple in the DNA cartoon. The color scheme of the bHLH along with leucine zipper (LZ) is shown at the bottom.

(D) Read density heatmaps (in color scales) showing the intensity of Ascl1 ChIP-seq signal (blue) 48 hr post induction in MEFs and MNase-seq (red) in non-induced MEFs spanning ± 1 kb from the center of the Ascl1 peaks. The analyzed sequences were organized in a descending rank order according to the MNase-seq tags within the central 200 bp (double arrows). The number of targeted sites is indicated. The nucleosome enriched sites were separated from the nucleosome depleted sites (dashed line).

(E) Logo representations of *de novo* motifs identified in Asl1 nucleosome-enriched targets (top) and nucleosome-depleted targets (bottom). The motifs were aligned to canonical motifs (middle). The number of targets analyzed and percentage of motif enrichments are indicated.

Figure S7. Nucleosome binding compatibility of bipartite DBDs. Related to Figure 6.

(A) Cartoon representations of the 3-D crystal structures of the Pou domains (blue) of Oct4 and Brn5 in complexes with DNA (red) containing canonical motifs. Side and top

views are shown and dashed curved arrows are shown to represent the extent of exposed DNA surface. The PDB ids are indicated

(B) Same as (A) for the Paired domains of Pax5 and Pax6.

(C) Cartoon representation (left panel) showing the three dimensional structures of the bipartite paired (PRD and HD) domain of Pax8 free of DNA as determined by NMR. All the NMR-determined 3D states are aligned and shown by transparent colors to indicate the measured flexibility of the free DBD. The gradual transition of the paired domain from DNA-free to DNA-bound was measured by morphing (Extended Experimental Procedures). Color transparency indicates the extent of the apparent flexibility. The motif recognized by the paired domains is shown above. The color scheme is shown at the bottom.

REFERENCES FOR SUPPLEMENTAL MATERIAL

Assa-Munt, N., Mortishire-Smith, R.J., Aurora, R., Herr, W., and Wright, P.E. (1993). The solution structure of the Oct-1 POU-specific domain reveals a striking similarity to the bacteriophage lambda repressor DNA-binding domain. Cell *73*, 193–205.

Bailey, T.L. (2011). DREME: motif discovery in transcription factor ChIP-seq data. Bioinformatics *27*, 1653–1659.

Bailey, T.L., and Machanick, P. (2012). Inferring direct DNA binding from ChIP-seq. Nucleic Acids Res. *40*, e128.

Calero-Nieto, F.J., Ng, F.S., Wilson, N.K., Hannah, R., Moignard, V., Leal-Cervantes, A.I., Jimenez-Madrid, I., Diamanti, E., Wernisch, L., and Göttgens, B. (2014). Key regulators control distinct transcriptional programmes in blood progenitor and mast cells. EMBO J.

Cox, M., van Tilborg, P.J., de Laat, W., Boelens, R., van Leeuwen, H.C., van der Vliet, P.C., and Kaptein, R. (1995). Solution structure of the Oct-1 POU homeodomain determined by NMR and restrained molecular dynamics. J. Biomol. NMR *6*, 23–32.

Esch, D., Vahokoski, J., Groves, M.R., Pogenberg, V., Cojocaru, V., Vom Bruch, H., Han, D., Drexler, H.C.A., Araúzo-Bravo, M.J., Ng, C.K.L., et al. (2013). A unique Oct4 interface is crucial for reprogramming to pluripotency. Nat. Cell Biol. *15*, 295–301.

Farina, A., Faiola, F., and Martinez, E. (2004). Reconstitution of an E box-binding Myc:Max complex with recombinant full-length proteins expressed in Escherichia coli. Protein Expr. Purif. *34*, 215–222.

Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics *27*, 1017–1018.

Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L., and Noble, W.S. (2007). Quantifying similarity between motifs. Genome Biol *8*, R24.

Hockemeyer, D., Soldner, F., Cook, E.G., Gao, Q., Mitalipova, M., and Jaenisch, R. (2008). A drug-inducible system for direct reprogramming of human somatic cells to pluripotency. Cell Stem Cell *3*, 346–353.

Kelly, T.K., Liu, Y., Lay, F.D., Liang, G., Berman, B.P., and Jones, P.A. (2012). Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. Genome Res. *22*, 2497–2506.

Kiefer, F., Arnold, K., Künzli, M., Bordoli, L., and Schwede, T. (2009). The SWISS-MODEL Repository and associated resources. Nucleic Acids Res. *37*, D387–D392.

Kopp, J., and Schwede, T. (2006). The SWISS-MODEL Repository: new features and functionalities. Nucleic Acids Res. *34*, D315–D318.

Krebs, W.G. (2000). SURVEY AND SUMMARY: The morph server: a standardized system for analyzing and visualizing macromolecular motions in a database framework. Nucleic Acids Res. *28*, 1665–1675.

Lee, B., and Richards, F.M. (1971). The interpretation of protein structures: estimation of static accessibility. J. Mol. Biol. *55*, 379–400.

Longo, A., Guanga, G.P., and Rose, R.B. (2008). Crystal structure of E47-NeuroD1/beta2 bHLH domain-DNA complex: heterodimer selectivity and DNA recognition. Biochemistry *47*, 218–229.

Ma, P.C., Rould, M.A., Weintraub, H., and Pabo, C.O. (1994). Crystal structure of MyoD bHLH domain-DNA complex: perspectives on DNA recognition and implications for transcriptional activation. Cell *77*, 451–459.

Machanick, P., and Bailey, T.L. (2011). MEME-ChIP: motif analysis of large DNA datasets. Bioinformatics *27*, 1696–1697.

MacQuarrie, K.L., Yao, Z., Fong, A.P., Diede, S.J., Rudzinski, E.R., Hawkins, D.S., and Tapscott, S.J. (2013). Comparison of genome-wide binding of MyoD in normal human myogenic cells and rhabdomyosarcomas identifies regional and local suppression of promyogenic transcription factors. Mol. Cell. Biol. *33*, 773–784.

Mathelier, A., Zhao, X., Zhang, A.W., Parcy, F., Worsley-Hunt, R., Arenillas, D.J., Buchman, S., Chen, C., Chou, A., Ienasescu, H., et al. (2014). JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. Nucleic Acids Res. *42*, D142–D147.

Murphy, E.C., Zhurkin, V.B., Louis, J.M., Cornilescu, G., and Clore, G.M. (2001). Structural basis for SRY-dependent 46-X,Y sex reversal: modulation of DNA bending by a naturally occurring point mutation. J. Mol. Biol. *312*, 481–499.

Nair, S.K., and Burley, S.K. (2003). X-ray structures of Myc-Max and Mad-Max recognizing DNA. Molecular bases of regulation by proto-oncogenic transcription factors. Cell *112*, 193–205.

Newburger, D.E., and Bulyk, M.L. (2009). UniPROBE: an online database of protein binding microarray data on protein-DNA interactions. Nucleic Acids Res. *37*, D77–D82.

El Omari, K., Hoosdally, S.J., Tuladhar, K., Karia, D., Hall-Ponselé, E., Platonova, O., Vyas, P., Patient, R., Porcher, C., and Mancini, E.J. (2013). Structural basis for LMO2-driven recruitment of the SCL:E47bHLH heterodimer to hematopoietic-specific transcriptional targets. Cell Rep. *4*, 135–147.

Pogenberg, V., Ogmundsdóttir, M.H., Bergsteinsdóttir, K., Schepsky, A., Phung, B., Deineko, V., Milewski, M., Steingrímsson, E., and Wilmanns, M. (2012). Restricted leucine zipper dimerization and specificity of DNA recognition of the melanocyte master regulator MITF. Genes Dev. *26*, 2647–2658.

Reményi, A., Lins, K., Nissen, L.J., Reinbold, R., Schöler, H.R., Wilmanns, M., Remenyi, A., and Scholer, H.R. (2003). Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. Genes Dev *17*, 2048–2059.

Sali, A., and Blundell, T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. *234*, 779–815.

Sauvé, S., Tremblay, L., and Lavigne, P. (2004). The NMR solution structure of a mutant of the Max b/HLH/LZ free of DNA: insights into the specific and reversible DNA binding mechanism of dimeric transcription factors. J. Mol. Biol. *342*, 813–832.

Schuetz, A., Nana, D., Rose, C., Zocher, G., Milanovic, M., Koenigsmann, J., Blasig, R., Heinemann, U., and Carstanjen, D. (2011). The structure of the Klf4 DNA-binding domain links to self-renewal and macrophage differentiation. Cell Mol Life Sci *68*, 3121–3131.

Soufi, A., Donahue, G., and Zaret, K.S. (2012). Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. Cell *151*, 994–1004.

Suvà, M.L., Rheinbay, E., Gillespie, S.M., Patel, A.P., Wakimoto, H., Rabkin, S.D., Riggi, N., Chi, A.S., Cahill, D.P., Nahed, B. V, et al. (2014). Reconstructing and Reprogramming the Tumor-Propagating Potential of Glioblastoma Stem-like Cells. Cell *157*, 580–594.

Tanaka, Y., Tawaramoto-Sasanuma, M., Kawaguchi, S., Ohta, T., Yoda, K., Kurumizaka, H., and Yokoyama, S. (2004). Expression and purification of recombinant human histones. Methods *33*, 3–11.

Teif, V.B., Vainshtein, Y., Caudron-Herger, M., Mallm, J.-P., Marth, C., Höfer, T., and Rippe, K. (2012). Genome-wide nucleosome positioning during embryonic stem cell development. Nat. Struct. Mol. Biol. *19*, 1185–1192.

Tennant, B.R., Robertson, A.G., Kramer, M., Li, L., Zhang, X., Beach, M., Thiessen, N., Chiu, R., Mungall, K., Whiting, C.J., et al. (2013). Identification and analysis of murine pancreatic islet enhancers. Diabetologia *56*, 542–552.

Vasudevan, D., Chua, E.Y.D., and Davey, C.A. (2010). Crystal structures of nucleosome core particles containing the "601" strong positioning sequence. J. Mol. Biol. *403*, 1–10.

Wapinski, O.L., Vierbuchen, T., Qu, K., Lee, Q.Y., Chanda, S., Fuentes, D.R., Giresi, P.G., Ng, Y.H., Marro, S., Neff, N.F., et al. (2013). Hierarchical Mechanisms for Direct Reprogramming of Fibroblasts to Neurons. Cell *155*, 621–635.

Zianni, M., Tessanne, K., Merighi, M., Laguna, R., and Tabita, F.R. (2006). Identification of the DNA bases of a DNase I footprint by the use of dye primer sequencing on an automated capillary DNA analysis instrument. J. Biomol. Tech. *17*, 103–113.