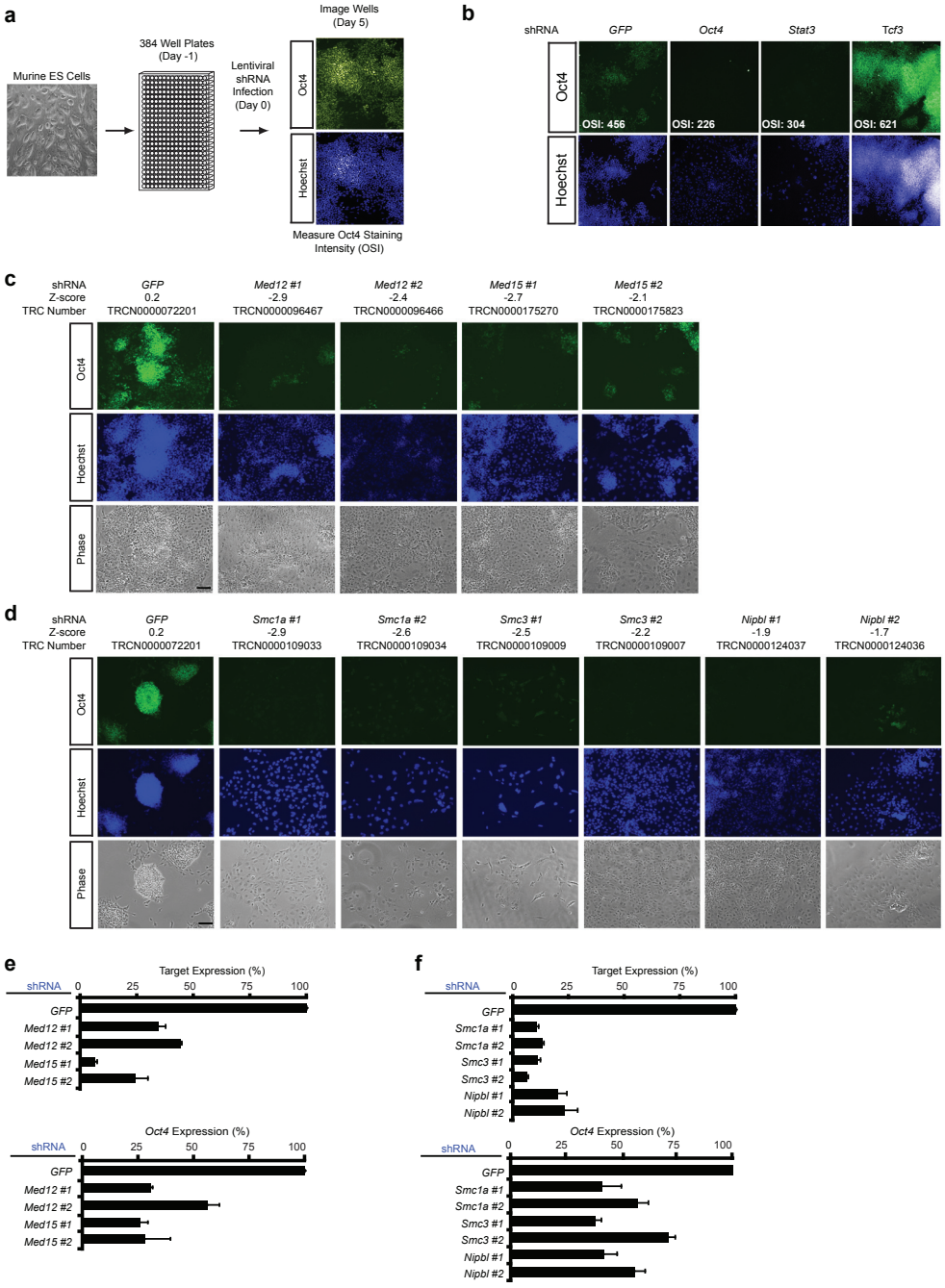
SUPPLEMENTARY INFORMATION

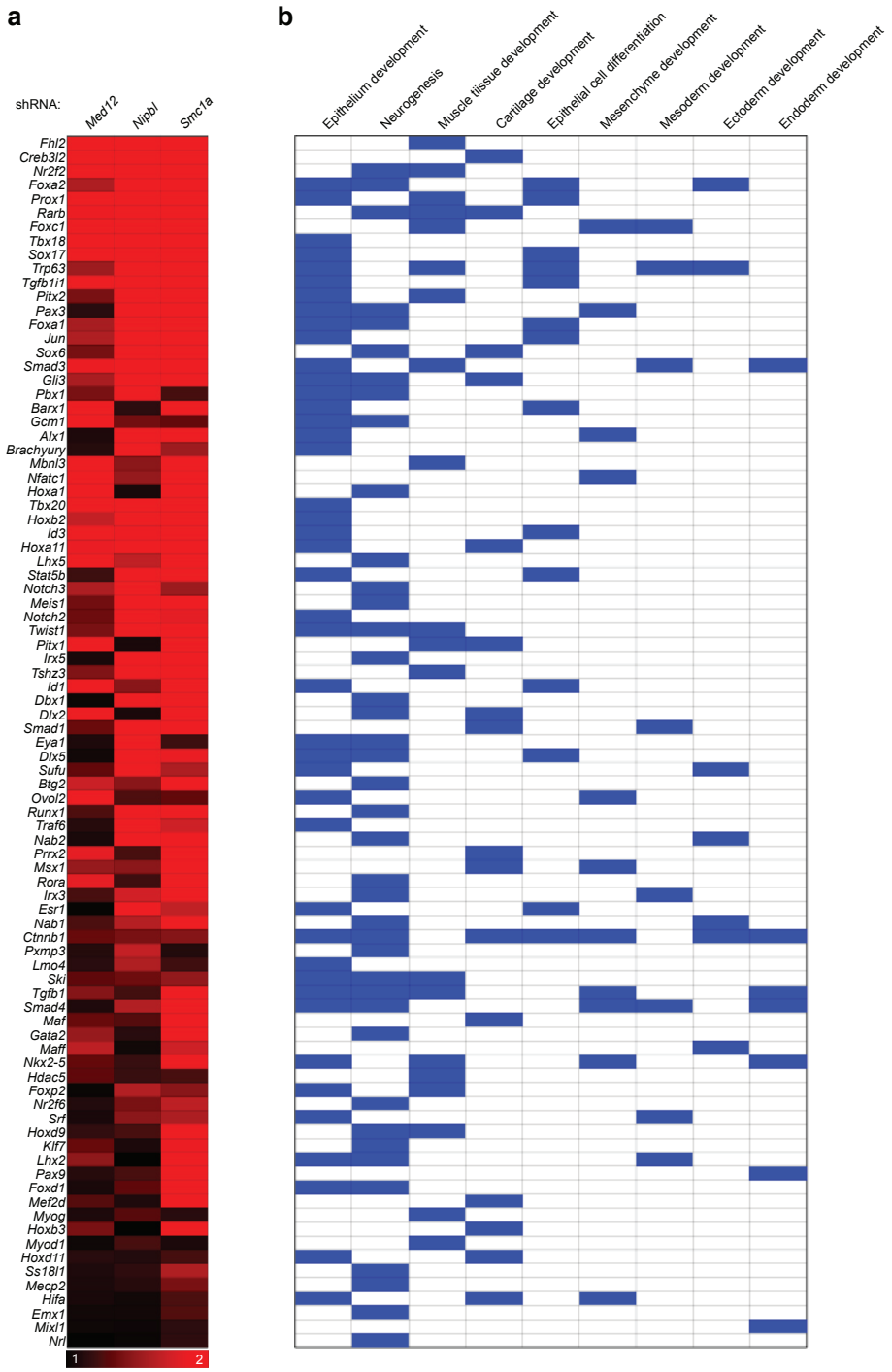
CONTENTS

Supplementary Tables **See separate Supplementary Table files**

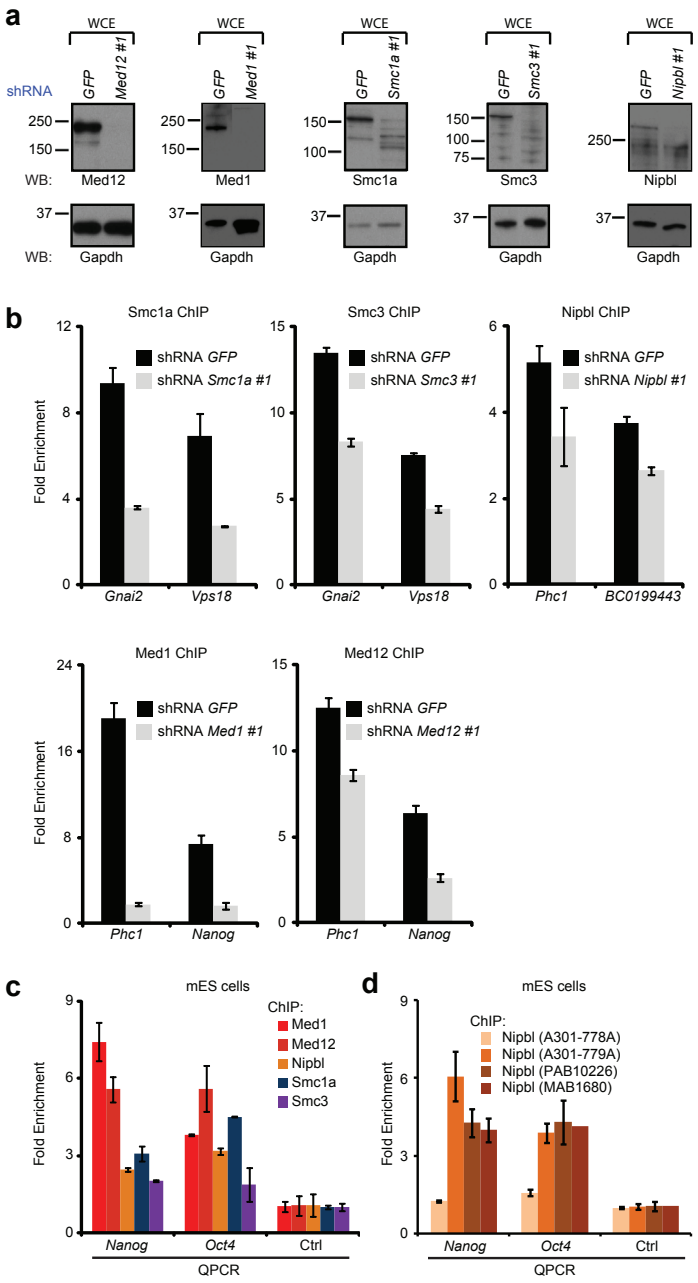
Supplementary Table 1 – Z-scores of shRNAs Used in the Screen
Supplementary Table 2 – Classification of Screen Hits
Supplementary Table 3 – *Med12*, *Smc1a* and *Nipbl* Knockdown Expression Data
Supplementary Table 4 – Bound Genomic Regions
Supplementary Table 5 – Summary of Occupied Genes
Supplementary Table 6 – Summary of ChIP-Seq Data Used
Supplementary Table 7 – Chromosome Conformation Capture (3C) Primers



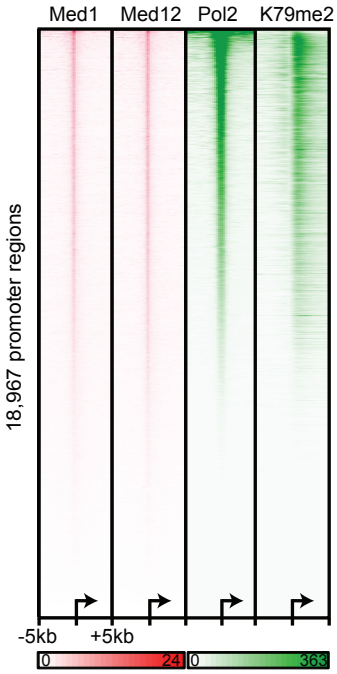
Supplementary Figure 1: Screening protocol and validation of mediator and cohesin shRNAs. a, Outline of the screening protocol. Murine embryonic stem cells were seeded without a MEF feeder layer into 384-well plates. The following day cells were infected with individual lentiviral shRNAs targeting chromatin regulators and transcription factors. Infections were done in quadruplicate (chromatin regulator set) or duplicate (transcription factor set) on separate plates (Supplementary Table 1). Five days post-infection cells were fixed and stained with Hoechst and for Oct4. Cells were identified based on the Hoechst staining and the average Oct4 staining intensity was quantified using Cellomics software. **b,** Representative images from control wells on a 384-well plate infected with shRNAs targeting positive regulators of pluripotency (*Oct4* and *Stat3*) and a negative regulator of pluripotency (*Tcf3*) [1-5]. OSI indicates the average Oct4 staining intensity of the cells in the well. **c, d,** Multiple shRNAs targeting mediator (c) and cohesin (d) components reduce Oct4 protein levels and result in changes in colony morphology. Murine ES cells were infected with the indicated shRNA and stained with Hoechst and for Oct4. Scale bar = 100μM. **e, f,** Effect of multiple mediator and cohesin shRNAs on transcript levels for *Med12*, *Med15*, *Smc1a*, *Smc3*, *Nipbl* and *Oct4*. Murine ES cells were infected with the indicated shRNA and transcript levels were evaluated by real-time qPCR. The error bars represent the standard deviation of the average of 3 independent PCR reactions.



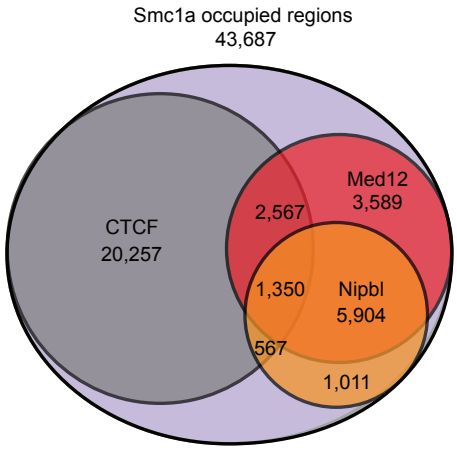
Supplementary Figure 2: Annotation of upregulated transcription factor genes in the *Med12*, *Nipbl*, and *Smc1a* knockdown expression datasets. **a,** Heat map demonstrating that the decreased expression of *Med12*, *Nipbl*, and *Smc1a* result in the upregulation of a similar set of developmental transcription factor genes. Genes that are displayed are upregulated following *Med12*, *Nipbl*, and *Smc1a* knockdowns, and were annotated in at least one of the Gene Ontology categories shown in b. Genes were rank ordered based on the mean expression changes for the *Med12* and *Nipbl* knockdowns. This was done because mediator-Nipbl occupy one set of sites whereas cohesin can occupy two sets of sites, cohesin-CTCF or cohesin-mediator-Nipbl. Expression data was generated from ES cells that were infected with *GFP* control, *Med12*, *Nipbl*, or *Smc1a* shRNAs. Five days post-infection, gene expression levels relative to the control *GFP* infection were determined with Agilent whole genome expression arrays. A relative signal scale is shown at the bottom of the panel. **b**, The decreased expression of *Med12*, *Nipbl*, and *Smc1a* result in the upregulation of transcription factor genes associated with developmental processes. Developmental categories from Gene Ontology (GO) are indicated at the top of the display. The annotation of a gene in the GO category is denoted by a blue box.



Supplementary Figure 3: Validation of mediator, cohesin and nipbl antibodies used for ChIP-Seq. **a,** Antibodies against Med12, Med1, Smc1a, Smc3 and Nipbl are specific and shRNAs targeting *Med12*, *Med1*, *Smc1a*, *Smc3* and *Nipbl* result in reduced levels of the target protein. Murine ES cells were infected with the indicated shRNA and protein levels were determined by western blot analysis. **b,** Gene specific ChIPs demonstrating that a reduction in Smc1a, Smc3, Nipbl, Med1 and Med12 protein levels by shRNA result in a decreased ChIP signal at the indicated gene. Murine ES cells were infected with the indicated shRNA; gene specific ChIP experiments were performed and analyzed by real-time qPCR. Fold enrichment is relative to a negative control region. The error bars represent the standard deviation of the average of 3 independent PCR reactions. **c,** Gene specific ChIPs verifying that mediator, cohesin and Nipbl occupy the promoter regions of *Oct4* and *Nanog* in ES cells. Fold enrichment is relative to a negative control region. The error bars represent the standard deviation of the average of 3 independent experiments. **d,** Gene specific ChIPs indicating that the Nipbl antibodies PAB10226 and MAB1680 also enrich for *Nanog* and *Oct4* promoter occupied Nipbl to similar levels as the A301-779A antibody utilized to generate the ChIP-Seq dataset. Fold enrichment is relative to a negative control region. The error bars represent the standard deviation of the average of 3 independent.



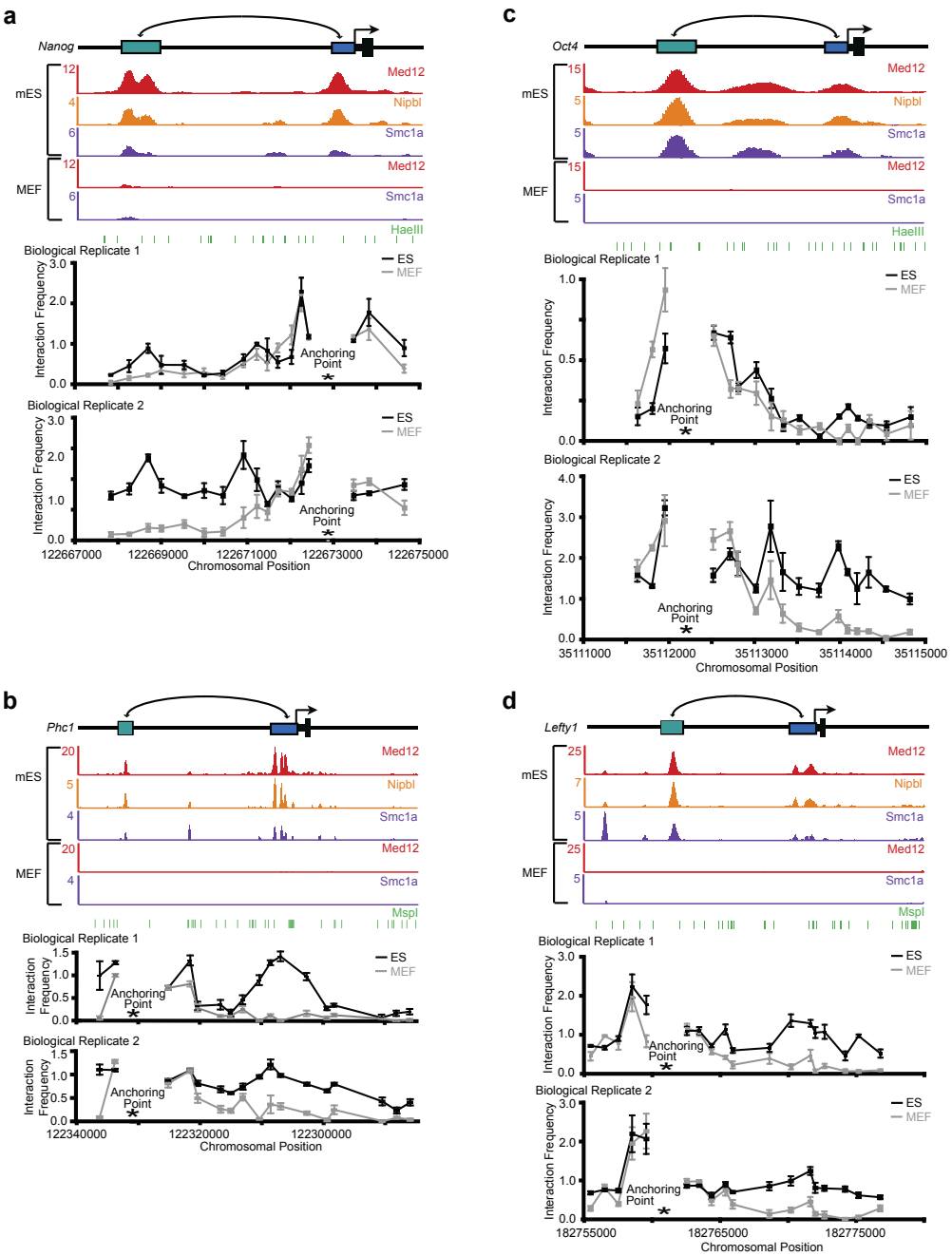
Supplementary Figure 4: Mediator occupies the promoters of actively transcribed genes. Density map of ChIP-Seq results for mediator (Med1, Med12), RNA polymerase II (Pol2) and di-methylated histone H3 lysine 79 (K79me2) demonstrates mediator occupancy at genes that are actively transcribed in ES cells. Normalized read counts are shown for 10kb surrounding 18,967 Refseq promoters (from -5kb to +5kb) sorted by maximum level of Pol2 enrichment. A relative signal scale (reads/million) and the position of the transcription start site are shown at the bottom of the panel.



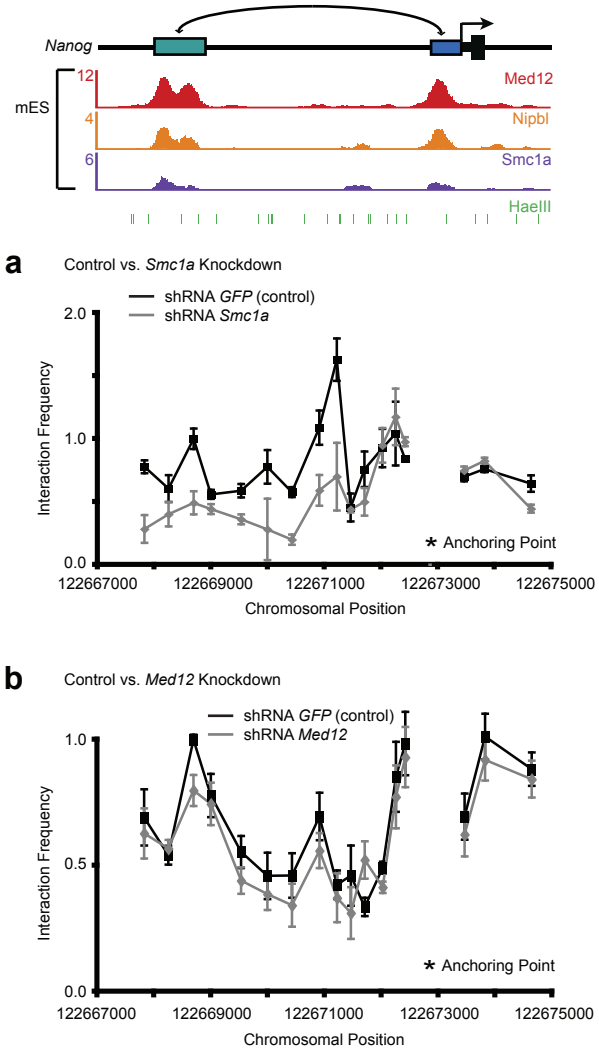
Supplementary Figure 5: Nipbl occupies regions co-occupied by mediator and cohesin. Venn diagram demonstrating the overlap of high confidence (P-val $<10^{-9}$) CTCF, mediator (Med12) and Nipbl occupied sites with cohesin (Smc1a). The overlap of Smc1a, Med12 and Nipbl sites is highly significant (P-val $< 10^{-300}$), whereas the overlap of Smc1a, CTCF and Nipbl is no greater than expected by chance (P-val = 1).

		shRNA		
		Med12	*Smc1a*	*Nipbl*
shRNA	*Med12*	1.00	0.80	0.76
	Smc1a	0.80	1.00	0.83
	Nipbl	0.76	0.83	1.00

Supplementary Figure 6: Mediator, cohesin and *Nipbl* knockdown expression datasets are similar. Pearson correlations indicate that the expression changes are similar at genes co-occupied by mediator (Med12), cohesin (Smc1a) and Nipbl in response to a *Med12*, *Smc1a* or *Nipbl* knockdown. Genes used for the analysis have evidence of a co-occupied Smc1a-Med12-Nipbl region within the gene body or within 10kb upstream of the transcriptional start site, evidence of Pol2 occupancy within the gene body and significant (P-val <0.01) expression changes for a *Smc1a*, *Med12* and *Nipbl* knockdown in independent experiments. Gene expression levels relative to the control *GFP* infection were determined with Agilent whole genome expression arrays.



Supplementary Figure 7: Mediator and cohesin binding profiles predict enhancer-promoter looping events. a-d, A looping event between the upstream enhancer and the core promoter of *Nanog*, *Phc1*, *Oct4 (Pou5f1)* and *Lefty1* was detected by Chromosome Conformation Capture (3C) in ES cells, but not in MEFs. Biological replicates are shown for each locus. ES cell and MEF crosslinked chromatin was digested by the indicated restriction enzyme and religated under conditions that favor intramolecular ligation events. The interaction frequency between the anchoring point and distal fragments was determined by PCR and normalized to BAC templates and control regions. The restriction enzyme sites are indicated above the 3C graph. The error bars represent the standard error of the average of 3 independent PCR reactions. The genomic coordinates are NCBI build 36/mm8. The ChIP-Seq binding profiles for Med12, Nipbl and Smc1a are shown in reads/million with the base of the y-axis set to 0.5 reads/million.



Supplementary Figure 8: Enhancer-promoter looping at *Nanog* decreases with a mediator or cohesin knockdown. Chromosome Conformation Capture (3C) data demonstrating that the interaction frequency between the promoter and enhancer of *Nanog* decreases for a cohesin (*Smc1a*) or a mediator (*Med12*) knockdown. ES cells were infected with a control shRNA (*GFP*) or shRNAs targeting *Smc1a* or *Med12*. Crosslinked chromatin was digested by the HaeIII restriction enzyme and religated under conditions that favor intramolecular ligation events. The interaction frequency between the anchoring point and distal fragments was determined by PCR and normalized to BAC templates and control regions. For both graphs the interaction frequency between primer Nanog 4 (within the enhancer, Supplementary Table 7) and primer Nanog 20 (anchoring primer, Supplementary Table 7) was normalized to 1 for the control shRNA (*GFP*) infected cells. All other interaction frequencies were scaled accordingly. The restriction enzyme sites are indicated above the 3C graph. The error bars represent the standard error of the average of 3 independent PCR reactions. The genomic coordinates are NCBI build 36/mm8. The ChIP-Seq binding profiles for Med12, Nipbl and Smc1a are shown in reads/million with the base of the y-axis set to 0.5 reads/million.

Supplementary Data Files

Supplementary Data File 1

Formatted (.WIG) files for Med1_mES, Med12_mES, Nipbl_mES, Smc1a_mES, Smc3_mES, TBP_mES, Oct4_mES, Sox2_mES, Nanog_mES, Pol2_mES, H3K79me2_mES, CTCF_mES, Med1_MEFs, Med12_MEFs, Smc1a_MEFs and CTCF_MEFs.

Supplementary Data File 1 contains data zipped, formatted (WIG.GZ) for upload into the UCSC genome browser [6]. To upload the file, first unzip the files onto a computer with Internet access. Then use a web browser to go to http://genome.ucsc.edu/cgi-bin/hgCustom?hgsid=105256378. Select genome (Mouse) and assembly (Feb. 2006 (NCBI36/mm8)). In the "Paste URLs or Data" section, select "Browse…" on the right of the screen. Use the pop-up window to select the unzipped files, and then select "Submit". The upload process may take some time.

These files present ChIP-Seq data. The first track for each data set contains the ChIP-Seq density across the genome in 25bp bins. The minimum ChIP-Seq density shown in these files is 0.5 reads per million. Subsequent tracks identify genomic regions identified as enriched (P-val < 10^{-9}).

This data is contained in 3 separate zipped files - see Supplementary Data 1 - parts 1, 2 and 3

Supplementary Experimental Procedures

Cell Culture Conditions

Embryonic Stem Cells
V6.5 murine embryonic stem (mES) cells were grown on irradiated murine embryonic fibroblasts (MEFs) unless otherwise stated. Cells were grown under standard mES cell conditions as described previously [7]. Briefly, cells were grown on 0.2% gelatinized (Sigma, G1890) tissue culture plates in ESC media; DMEM-KO (Invitrogen, 10829-018) supplemented with 15% fetal bovine serum (Hyclone, characterized SH3007103), 1000 U/mL LIF (ESGRO, ESG1106), 100 μM nonessential amino acids (Invitrogen, 11140-050), 2 mM L-glutamine (Invitrogen, 25030-081), 100 U/mL penicillin, 100 μg/mL streptomycin (Invitrogen, 15140-122), and 8 nL/mL of 2-mercaptoethanol (Sigma, M7522).

Mouse Embryonic Fibroblasts (MEFs)
Low passage MEFs were grown on tissue culture plates in DMEM (Invitrogen, 11965) supplemented with 10% fetal bovine serum (Hyclone, characterized SH3007103), 100 μM nonessential amino acids (Invitrogen, 11140-050), 2 mM L-glutamine (Invitrogen, 25030-081), 100 U/mL penicillin, 100 μg/mL streptomycin (Invitrogen, 15140-122), and 8 nL/mL of 2-mercaptoethanol (Sigma, M7522).

High-Throughput shRNA Screening

Library Design and Lentiviral Production
Small hairpins targeting 197 chromatin regulators and 2021 transcription factors were designed and cloned into pLKO.1 lentiviral vectors (Open Biosystems) as previously described [8]. On average 5 different shRNAs targeting each chromatin regulator or transcription factor were used. Lentiviral supernatants were arrayed in 384-well plates with negative control lentivirus (shRNAs targeting *GFP*, *RFP*, *Luciferase* and *LacZ*) [8].

Lentiviral Infections
Murine ES cells were split off MEFs and placed in a tissue culture dish for 45 minutes to selectively remove the MEFs. Murine ES cells were counted with a Coulter Counter (Beckman, #1499) and seeded using a μFill (Bioteck) at a density of 1500 cells/well in 384-well plates (Costar 3712) treated with 0.2% gelatin (Sigma, G1890). An initial cell plating density of 1500 cells/well was established so that an adequate amount of cells would survive puromycin selection for analysis. However, the initial cell plating density was kept low enough to avoid wells reaching confluency during the timeframe of the assay. One day following cell plating the media was removed, replaced with ESC media containing 8 μg/ml of polybrene (Sigma, H9268-10G) and cells were infected with 2 μl of shRNA lentiviral supernatant. Infections were performed in duplicate (transcription factor set) or quadruplicate (chromatin regulator set) on separate plates. Supplementary Table 1 denotes which screening set the shRNAs were in. Control wells on each plate were mock infected and designated as “Empty”.

Positive control wells on each plate were infected with 3 µl of validated control shRNA lentiviral supernatant targeting *Oct4* (TRCN0000009613), *Tcf3* (TRCN0000095454) and *Stat3* (TRCN0000071454) that was generated independently of the screening sets (Lentiviral Production and Infection). Sequence and shRNAs are available from Open Biosystems. Plates were spun for 30 minutes at 2150 rpm following infection. Twenty-four hours post infection cells were treated with 3.5 μg/ml of puromycin (Sigma, P8833) in ESC media to select for stable integration of the shRNA construct. ESC media with puromycin was changed daily. Five days post infection cells were crosslinked for 15 minutes with 4% paraformaldehyde (EMS Diasum, 15710).

Immunofluorescence
Following crosslinking, the cells were washed once with PBS, twice with blocking buffer (PBS with 0.25% BSA, Sigma, A3059-10G) and then permeabilized for 15 minutes with 0.2% Triton X-100 (Sigma, T8797-100ml). After two washes with blocking buffer cells were stained overnight at 4ºC for Oct4 (Santa Cruz Biotechnology, sc-5279; 1:100 dilution) and washed twice with blocking buffer. Cells were incubated for 4 hours at room temperature with goat anti-mouse-conjugated Alexa Fluor 488 (Invitrogen; 1:200 dilution) and Hoechst 33342 (Invitrogen; 1:1000 dilution). Finally, cells were washed twice with blocking buffer and twice with PBS before imaging.

Image Acquisition and Analysis
Image acquisition and data analysis were performed essentially as described [8]. Stained cells were imaged on an Arrayscan HCS Reader (Cellomics) using the standard acquisition camera mode (10X objective, 9 fields). Hoechst was used as the focus channel. Objects selected for analysis were identified based on the Hoechst staining intensity using the Target Activation Protocol and the Fixed Threshold Method. Parameters were established requiring that individual objects pass an intensity and size threshold. The Object Segmentation Assay Parameter was adjusted for maximal resolution between individual cells. Following object selection, the average Oct4 pixel staining intensity was determined per object and then a mean value for each well was calculated. Image acquisition for a well continued until at least 2500 objects were identified, the entire well (9 fields) was imaged or less than 20 objects were identified for three fields imaged in a row. To account for viability defects or low titer lentivirus for the chromatin regulator screening set an shRNA was excluded from subsequent analysis if less than 250 objects were identified for any one of the 4 replicates. The 250 identified objects threshold was determined based on the average number of identified objects for the “Empty” (no virus) wells (mean: 53.4, standard deviation: 49.3). To account for viability defects or low titer lentivirus for the transcription factor screening set a shRNA was excluded from subsequent analysis if less than 300 objects were identified for any one of the 2 replicates. The 300 identified objects threshold was determined based on the average number of identified objects for the “Empty” (no virus) wells (mean: 39.2, standard deviation: 147.5).

To normalize for plate effects, a Z-score based on the Oct4 staining intensity was calculated for each well using the following negative control infections, 24 different shRNAs targeting *GFP*, 16 different shRNAs targeting *RFP*, 25 different shRNAs targeting *Luciferase* and 20 different shRNAs targeting *LacZ*. There

were a total of between 16 and 22 wells infected with various negative control shRNAs on each 384-well plate, with the exception of one plate within the transcription factor set that contained 99 wells with control infections. The average Oct4 staining intensity for the negative control infected wells was calculated along with a standard deviation to give an estimation of the amount of the signal variability. The average Oct4 staining intensity for all the negative control infected wells on a plate and the standard deviation were utilized to calculated a Z-score for every well on the plate. The Z-scores for the four quadruplicate infections (chromatin regulator set) or two duplicate infections (transcription factor set) were averaged for a final Z-score for every shRNA. The Z-score data for both sets were combined (Supplementary Table 1). Representative control 384-well plate images (shRNAs targeting *Oct4*, *Stat3*, *Tcf3* and *GFP*) were exported (Cellomics Software), converted from DIBs to TIFs (CellProfiler, http://www.cellprofiler.org), and manipulated with Photoshop CS3 Extended (Supplementary Fig. 1a, b).

Combining Screening Data (Supplementary Table 1)
We recently published the results of an ES screen where 197 chromatin regulators were selectively targeted for knockdown [9]. For the present study we screened an additional 2021 genes primary encoding transcription factors. In order to generate a more complete picture of factors required for maintaining ES cell state we included the set of chromatin regulator results from the previous study. The shRNAs from each set are denoted in Supplementary Table 1.

The same methodology was followed for screening with both the chromatin regulator and transcription factor sets with the following exception, infections for the chromatin regulator set were done in quadruplicate and infections for the transcription factor set were carried out in duplicate, due to the large size of the transcription factor screening set (30 x 384-well plates, 2021 genes). Because the average Z-scores of the added controls (Oct4 and Stat3) were within close proximity for both screening sets (Chromatin Regulator Set: -3.3 and –2.4 for Oct4 and Stat3 respectively; Transcription Factor Set: -3.0 and –2.1 for Oct4 and Stat3 respectively) we reasoned that Z-scores between the two screening sets were comparable.

Criteria for Identifying Screening Hits (Supplementary Table 2)
We used multiple Z-score level thresholds to select chromatin regulators and transcription factors that had significantly reduced Oct4 levels for inclusion in Supplementary Table 2. First, a chromatin regulator or transcription factor had to have at least two shRNA with a Z-score less than –1.5 and it was possible to classify the gene based on the literature. Second, a chromatin regulator or transcription factor with a single shRNA hit and a Z-score of less than –1.5 was also included if it could be classified with one of the multiple shRNA hits. Third, the following chromatin regulators (Cbx7, Cbx8/Pc3 and Ezh2) were included even though each was only a single shRNA hit, because all had strong negative Z-scores, all are polycomb proteins, and polycomb has been previously demonstrated to be important for regulating ES cell state [10]. The -1.5 cut-off was chosen because it was within close proximity to the Z-score of the Stat3 controls (-2.4 and –2.1 for the chromatin regulator and the transcription factor sets respectively).

Validation of shRNAs

Lentiviral Production and Infection
Lentivirus was produced according to Open Biosystems *Trans*-lentiviral shRNA Packaging System (TLP4614). The shRNA constructs targeting *Med1*, *Med12*, *Med15*, *Smc1a*, *Smc3*, *Nipbl*, *Oct4*, *Stat3* and *Tcf3* are listed below. All are available, including sequences from Open Biosystems. The shRNA targeting *GFP* (TRCN0000072201, Hairpin Sequence: gtcgagctggacggcgacgta) was one of the negative controls for the screen.

Smc1a #1	TRCN0000109033
Smc1a #2	TRCN0000109034
Smc3 #1	TRCN0000109009
Smc3 #2	TRCN0000109007
Nipbl #1	TRCN0000124037
Nipbl #2	TRCN0000124036
Med12 #1	TRCN0000096467
Med12 #2	TRCN0000096466
Med15 #1	TRCN0000175270
Med15 #2	TRCN0000175823
Med1 #1	TRCN0000099578
Oct4	TRCN0000009613
Stat3	TRCN0000071454
Tcf3	TRCN0000095454

For validation of the mediator and cohesin shRNAs, mES cells were split off MEFs, placed in a tissue culture dish for 45 minutes to selectively remove the MEFs and then plated in 6-well plates (200,000 cells/well). The following day cells were infected in ESC media containing 8 μg/ml polybrene (Sigma, H9268-10G) and plates were spun for 30 minutes at 2150 rpm. After 24 hours the media was removed and replaced with ESC media containing 3.5 μg/mL puromycin (Sigma, P8833). ESC media with puromycin was changed daily. Five days post infection RNA or proteins were extracted or the cells were crosslinked for immunofluorescence.

Immunofluorescence
Cells were crosslinked, permeabilized and stained as described for high-throughput screening. Images were acquired on a Nikon Inverted TE300 with a Hamamatsu Orca camera. Openlab (http://www.improvision.com/products/openlab/) was used for image acquisition. Openlab and Photoshop CS3 Extended were used for image manipulation.

RNA Extraction, cDNA, and TaqMan Expression Analysis
RNA utilized for real-time qPCR was extracted with TRIzol according to the manufacturer protocol (Invitrogen, 15596-026). Purified RNA was reverse

transcribed using Superscript III (Invitrogen) with oligo dT primed first-strand synthesis following the manufacturer protocol.

Real-time qPCR were carried out on the 7000 ABI Detection System using the following TaqMan probes according to the manufacturer protocol (Applied Biosystems).

Gapdh	Mm99999915_g1
Med12	Mm00804032_m1
Med15	Mm01171155_m1
Smc1a	Mm01253647_m1
Smc3	Mm00484012_m1
Nipbl	Mm01297461_m1
Oct4	Mm00658129_gH

Expression levels were normalized to *Gapdh* levels. All knockdowns are relative to control shRNA *GFP* infections.

Chromatin Immunoprecipitation

Biological replicates of all ChIP-Seq datasets with the exception of mediator (Med12 and Med1) in MEFs were generated and combined for analysis. A summary of the ChIP-Seq data is contained within Supplementary Table 6.

For Med1 (CRSP1/TRAP220) occupied genomic regions, we performed ChIP-Seq experiments using Bethyl Laboratories (A300-793A) antibody. The affinity purified antibody was raised in rabbit against an epitope corresponding to amino acids 1523-1581 mapping at the C-terminus of human Med1.

For Med12 occupied genomic regions, we performed ChIP-Seq experiments using Bethyl Laboratories (A300-774A) antibody. The affinity purified antibody was raised in rabbit against an epitope corresponding to amino acids 2150-2212 mapping at the C-terminus of human Med12.

For Smc1a occupied genomic regions, we performed ChIP-Seq experiments using Bethyl Laboratories (A300-055A) affinity purified rabbit polyclonal antibody. The epitope recognized by A300-055A maps to a region between residue 1175 and the C-terminus of human Smc1a.

For Smc3 occupied genomic regions, we performed ChIP-Seq experiments using Abcam (ab9263) antibody. The affinity purified antibody was raised in rabbit against an epitope corresponding to the last 100 amino acids of the human Smc3 protein.

For TBP occupied genomic regions, we performed ChIP-Seq experiments using Abcam (ab818) antibody. The antibody was raised with a synthetic peptide, which represents amino acid residues 1-20 of human TBP.

For Pol2 occupied genomic regions, we performed ChIP-Seq experiments using Covance 8WG16 antibody. This mouse monoclonal antibody was raised against the C-terminal heptapeptide repeat region on the largest subunit of Pol2, purified from wheat germ extract.

For H3K79me2 occupied genomic regions, we performed ChIP-Seq experiments using Abcam ab3594 rabbit polyclonal antibody. The antibody was raised with a synthetic peptide that is within residues 50 to the C-terminus of Human Histone H3, dimethylated at K79.

For CTCF occupied genomic regions, we performed ChIP-Seq experiments using an Upstate 07-729 rabbit polyclonal antibody.

For Nipbl occupied genomic regions, we performed ChIP-Seq experiments using a Bethyl A301-779A rabbit polyclonal antibody. The affinity purified antibody was raised in rabbit to a region between amino acid residues 1025 and 1075 of human Nipbl.

Protocols describing chromatin immunoprecipitation materials and methods have been previously described [10]. Embryonic stem cells or MEFs were grown to a final count of 5-10 x 10^7 cells for each ChIP experiment. Cells were chemically crosslinked by the addition of one-tenth volume of fresh 11% formaldehyde solution for 15 minutes (ES cells) or 10 minutes (MEFs) at room temperature. Cells were rinsed twice with 1X PBS and harvested using a silicon scraper and flash frozen in liquid nitrogen. Cells were stored at –80^{o}C prior to use. Cells were resuspended, lysed in lysis buffers and sonicated to solubilize and shear crosslinked DNA. Sonication conditions vary depending on cells, culture conditions, crosslinking and equipment.

For Nipbl, Smc1a, Smc3, Pol2, H3K79me2 and Med1 the sonication buffer was 20mM Tris-HCl pH8, 150mM NaCl, 2mM EDTA, 0.1% SDS, 1% Triton X-100. We used a Misonix Sonicator 3000 and sonicated at approximately 24 watts for 10 x 30 second pulses (60 second pause between pulses). Samples were kept on ice at all times. The resulting whole cell extract was incubated overnight at 4°C with 100 μl of Dynal Protein G magnetic beads that had been pre-incubated with approximately 10 μg of the appropriate antibody. Beads were washed 1X with the sonication buffer, 1X with 20mM Tris-HCl pH8, 500mM NaCl, 2mM EDTA, 0.1% SDS, 1%Triton X-100, 1X with 10mM Tris-HCl pH8, 250nM LiCl, 2mM EDTA, 1% NP40 and 1X with TE containing 50 mM NaCl.

For Med12 and CTCF, the sonication buffer was 10mM Tris-HCl pH8, 100mM NaCl, 1mM EDTA, 0.5mM EGTA, 0.1% Na-Deoxycholate, 0.5% N-

lauroylsarcosine. We used the same sonication and wash conditions as described above.

For TBP, the sonication buffer was 10mM Tris-HCl pH8, 100mM NaCl, 1mM EDTA, 0.5 mM EGTA, 0.1% Na-Deoxycholate and 0.5% N-lauroylsarcosine. We used a Misonix Sonicator 3000 and sonicated at approximately 24 watts for 10 x 30 second pulses (60 second pause between pulses). After Sonication, 10% Triton-X was added. After immunoprecipitation, beads were washed 4X with the RIPA buffer (50 mM Hepes-KOH pH 7.6, 500 mM LiCl, 1 mM EDTA, 1% NP40 and 0.7% Na-Deoxycholate) and 1X with TE containing 50 mM NaCl.

Bound complexes were eluted from the beads (50 mM Tris-HCl, pH 8.0, 10 mM EDTA and 1% SDS) by heating at 65°C for 1 hour with occasional vortexing and crosslinking was reversed by overnight incubation at 65°C. Whole cell extract DNA reserved from the sonication step was also treated for crosslink reversal. Immunoprecipitated DNA and whole cell extract DNA were treated with RNaseA and Proteinase K. DNA was purified by phenol:chloroform:isoamyl alcohol extraction.

ChIP-Seq Sample Preparation and Analysis

All protocols for Illumina/Solexa sequence preparation, sequencing and quality control are provided by Illumina (http://www.illumina.com/pages.ilmn?ID=203). A brief summary of the technique and minor protocol modifications are described below.

Sample Preparation
DNA was prepared for sequencing according to a modified version of the Illumina/Solexa Genomic DNA protocol. Fragmented DNA was prepared for ligation of Solexa linkers by repairing the ends and adding a single adenine nucleotide overhang to allow for directional ligation. A 1:100 dilution of the Adaptor Oligo Mix (Illumina) was used in the ligation step. A subsequent PCR step with limited (18) amplification cycles added additional linker sequence to the fragments to prepare them for annealing to the Genome Analyzer flow-cell. After amplification, a narrow range of fragment sizes was selected by separation on a 2% agarose gel and excision of a band between 150-350 bp (representing shear fragments between 50 and 250nt in length and ~100bp of primer sequence). The DNA was purified from the agarose and diluted to 10 nM for loading on the flow cell.

Polony Generation and Sequencing
The DNA library (2-4 pM) was applied to the flow-cell (8 samples per flow-cell) using the Cluster Station device from Illumina. The concentration of library applied to the flow-cell was calibrated such that polonies generated in the bridge amplification step originate from single strands of DNA. Multiple rounds of amplification reagents were flowed across the cell in the bridge amplification step

to generate polonies of approximately 1,000 strands in 1μm diameter spots. Double stranded polonies were visually checked for density and morphology by staining with a 1:5000 dilution of SYBR Green I (Invitrogen) and visualizing with a microscope under fluorescent illumination. Validated flow-cells were stored at 4^{o}C until sequencing.

Flow-cells were removed from storage and subjected to linearization and annealing of sequencing primer on the Cluster Station. Primed flow-cells were loaded into the Illumina Genome Analyzer 1G. After the first base was incorporated in the Sequencing-by-Synthesis reaction the process was paused for a key quality control checkpoint. A small section of each lane was imaged and the average intensity value for all four bases was compared to minimum thresholds. Flow-cells with low first base intensities were re-primed and if signal was not recovered the flow-cell was aborted. Flow-cells with signal intensities meeting the minimum thresholds were resumed and sequenced for 26 or 32 cycles.

ChIP-Seq Data Analysis

Images acquired from the Illumina/Solexa sequencer were processed through the bundled Solexa image extraction pipeline, which identified polony positions, performed base-calling and generated QC statistics. Sequences were aligned using ELAND software to NCBI Build 36 (UCSC mm8) of the mouse genome. Only sequences that mapped uniquely to the genome with zero or one mismatch were used for further analysis. When multiple reads mapped to the same genomic position, a maximum of two reads mapping to the same position were used. A summary of the total number of ChIP-Seq reads that were used in each experiment is provided (Supplementary Table 6). ChIP-Seq datasets profiling the genomic occupancy of H3K79me2 [11], Oct4 [11], Sox2 [11], Nanog [11], RNA polymerase II [12] and CTCF [13] in mES cells were obtained from previous publications and reanalyzed using the methods described below.

Analysis methods were derived from previously published methods [11,14-16]. Sequence reads from multiple flow cells for each IP target and/or biological replicates were combined. For all datasets, excluding Pol2 and H3K79me2, each read was extended 200bp, towards the interior of the sequenced fragment, based on the strand of the alignment. For Pol2 and H3K79me2 datasets, each read was extended 600bp towards the interior and 400bp towards the exterior of the sequenced fragment, based on the strand of the alignment. Across the genome, in 25 bp bins, the number of extended ChIP-Seq reads was tabulated. The 25bp genomic bins that contained statistically significant ChIP-Seq enrichment were identified by comparison to a Poissonian background model. Assuming background reads are spread randomly throughout the genome, the probability of observing a given number of reads in a 1kb window can be modeled as a Poisson process in which the expectation can be estimated as the number of mapped reads multiplied by the number of bins (40) into which each

read maps, divided by the total number of bins available (we estimated 70%). Enriched bins within 200bp of one another were combined into regions.

The Poissonian background model assumes a random distribution of background reads, however we have observed significant deviations from this expectation. Some of these non-random events can be detected as sites of apparent enrichment in negative control DNA samples and can create many false positives in ChIP-Seq experiments. To remove these regions, we compared genomic bins and regions that meet the statistical threshold for enrichment to a set of reads obtained from Solexa sequencing of DNA from whole cell extract (WCE) in matched cell samples. We required that enriched bins and enriched regions have five-fold greater ChIP-Seq density in the specific IP sample, compared with the control sample, normalized to the total number of reads in each dataset. This served to filter out genomic regions that are biased to having a greater than expected background density of ChIP-Seq reads. A summary of the enriched genomic regions (P-val $<10^{-9}$) and genes (P-val $<10^{-9}$) for each antibody is provided (Supplementary Table 4 and 5). Genomic coordinates for Supplementary Tables 4 and 5 are build NCBI36/mm8.

ChIP-Seq Density Map (Supplementary Fig. 4)
Genes were aligned with each other according to the position and direction of their transcription start site. For each experiment, the ChIP-Seq density profiles were normalized to the density per million total reads. Genes were sorted as by maximum level of Pol2 enrichment.

ChIP-Seq Enriched Region Maps (Fig. 2c and Fig. 5a, b)
The visualization shows the location of enriched regions (P-val $<10^{-9}$, Supplementary Table 4) in a collection of datasets (query datasets, indicated on the top) in relation to the enriched regions of another dataset (base dataset, indicated on the y-axis). For each of the enriched regions in the base dataset, corresponding genomic regions were calculated as +/- 5kb from the center of that enriched region (one genomic region per enriched region, row). For each of these genomic regions, the location and length of any enriched regions in the query datasets were drawn.

Assigning ChIP-Seq Enriched Regions to Genes (Supplementary Table 5)
The complete set of RefSeq genes was downloaded from the UCSC table browser (http://genome.ucsc.edu/cgi-bin/hgTables?command=start) on December 20, 2008. For all datasets, excluding Pol2 and H3K79me2, genes with enriched regions (P-val $<10^{-9}$) within 10kb of their transcription start site, or within the gene body were called bound. For Pol2 and H3K79me2 datasets, genes with enriched regions (P-val $<10^{-9}$) within the gene body were called bound. See Supplementary Table 4 for the enriched genomic regions (P-val $<10^{-9}$).

Note Regarding Summary of Occupied Genes Table (Supplementary Table 5)
Supplementary Table 5 provides binding information on every entry in the

RefSeq table downloaded on December 20, 2008 (See ChIP-Seq analysis above) and the bound gene numbers reflect counts of these entries. It should be noted however, that some of the gene names are not unique and thus the density map in Supplementary Fig. 4 may have fewer rows than there are entries in Supplementary Table 5.

Note Regarding Calculation of Co-occupied Regions (Supplementary Table 4) Supplementary Table 4 contains the genomic coordinates of enriched regions (P-val $<10^{-9}$) co-occupied by the indicated pair of factors. These coordinates are the union of all overlapping enriched regions of the two factors. It is possible for an enriched region of one factor to span, or bridge a gap between, two separate enriched regions of the other factor, in those cases, only one enriched region would be reported and it would be the union of all three enriched regions. This will cause the number of reported co-occupied regions to be less than the number of strictly overlapping sites reported in the Venn diagrams of Fig. 2b and Supplementary Fig. 5. The Venn diagrams are strictly the number of Smc1a sites that are partially overlapped by either CTCF, mediator (Med12) or Nipbl.

Gene Specific ChIPs

Gene specific ChIPs were performed in the indicated cell type following the protocol outlined in ChIP-Seq Sample Preparation. For the Gene specific ChIPs carried out in the knockdown cells, approximately $8x10^6$ ES cells (total) in 5 x 10cm tissue culture plates were infected with the indicated shRNA as described (Validation of shRNAs) except that the plates were not spun post infection. Syber Green real-time qPCR was carried out on the 7000 ABI Detection System according to the manufacturer protocol (Applied Biosystems). Data was normalized to the whole cell extract and control regions. Primers to the genes tested and control regions are listed below.

Gnai2
5'- ACAGAGCGATACGGCTCAGCAA-3'
5'-AAGTGGTAGCCGAAGGCAAGTGAA-3'

Vps18
5'-TCCTAGCGCCAACATGAGGAACT-3'
5'-TTTCAGCCGCGAGTGTTAACTGGA-3'

Phc1
5'-TTTGCTCTGCGTGACACTGAAGGT-3'
5'-AAATCCCAGCGCTTCTAGACGTAG-3'

BC0199443
5'-TGCCCACGTCGTAACAAGGTTT-3'
5'-AAGGCCGATCCTTTCTGGTTCA-3'

Nanog
5'-ATAGGGGGTGGGTAGGGTAG-3'
5'-CCCACAGAAAGAGCAAGACA-3'

Oct4
5'-TTGAACTGTGGTGGAGAGTGCT-3'
5'-TGCACCTTTGTTATGCATCTGCCG-3'

Ctrl
5'-TGGGTGCCGTATGCCACATTAT-3'
5'-TTTCTGGCCATCCGCACCTTAT-3'

ChIP-Western and Co-Immunoprecipitation (Fig. 3a, b)

For ChIP-Western, same conditions as for ChIP-Seq were used. For co-immunoprecipitation, murine ES cells were harvested in cold PBS and extracted for 30 min at 4°C in TNEN250 (50 mM Tris pH 7.5, 5 mM EDTA, 250 mM NaCl, 0.1% NP-40) with protease inhibitors. After centrifugation, supernatant was mixed to 2 volumes of TNENG (50 mM Tris pH 7.5, 5 mM EDTA, 100 mM NaCl, 0.1% NP-40, 10% glycerol). Protein complexes were immunoprecipitated overnight at 4°C using 5ug of Nipbl (Bethyl, A301-779A or Rabbit IgG (Upstate, 12-370) bound to 50ul of Dynabeads®. Immunoprecipitates were washed three times with TNEN125 (50 mM Tris pH 7.5, 5 mM EDTA, 125 mM NaCl, 0.1% NP-40). For both ChIP-Western and co-immunoprecipitation, beads were boiled for 10 minutes in XT buffer (Bio-Rad) containing 100mM DTT to elute proteins. After SDS-PAGE, Western blots were revealed with antibodies against Med23 (Bethyl, A300-425A), Smc1a (Bethyl, A300-055A), Smc3 (Abcam, Ab9236) and Nipbl (Bethyl, A301-779A).

Protein Extraction and Western Blot Analysis (Fig. 5c and Supplementary Fig. 3a)

ES cells were lysed with CelLytic Reagent (Sigma, C2978-50ml) containing protease inhibitors (Roche). After SDS-PAGE, Western blots were revealed with antibodies against Med1 (Bethyl, A300-793A), Med12 (Bethyl, A300-774A), Smc1a (Bethyl, A300-055A), Smc3 (Abcam, ab9263), Nipbl (Bethyl, A301-779A) or Gapdh (Abcam, ab9484).

Mediator Affinity Purification

The mediator complex was purified from murine ES cell nuclear extracts using immobilized GST-SREBP-1a (residues 1-50)[17]. Bound material washed 4x with 20 column volumes of 0.5M KCl HEGN (20mM Hepes, 0.1mM EDTA, 10% Glycerol, 0.1% NP-40 & 0.5M KCl) buffer, 2x with 0.15M KCl HEGN buffer, and eluted. The eluted sample was further purified with a CDK8 antibody. After

binding, this resin was washed 4x with 50 column volumes of 0.5M KCl HEGN buffer, 2x with 0.1M KCl HEGN buffer and eluted with 0.1M Glycine, pH 2.75. Western blot analysis was conducted with Smc3 (Abcam ab9263-50), Med15 (Taatjes Lab stock), Med12 (Bethyl A300-774A) or Nipbl (Bethyl A301-779A) antibodies.

Chromosome Conformation Capture (3C)

3C analysis was performed essentially as described by Miele *et al.* [18] with a few modifications. 10^8 mES or MEF cells were crosslinked as described (ChIP-Seq Sample Preparation and Analysis). For 3C analysis performed in *GFP* control, *Smc1a* or *Med12* shRNA knockdown cells, the cells were infected as described (Validation of shRNAs), except that the plates were not spun post infection. 10 x 10 cm tissue culture plates with approximately 1.5 x 10^6 ES cells/plate were infected for each shRNA and five days post infection cells were crosslinked for 15 minutes (ChIP-Seq Sample Preparation and Analysis).

Crosslinked cells were lysed and chromatin was digested with 1000 units HaeIII (NEB) for the *Nanog* and *Oct4* loci or 2000 units MspI (NEB) for the *Phc1* and *Lefty1* loci. Crosslinked fragments were subsequently ligated with 50 units T4 DNA ligase (Invitrogen) for 4 hours at 16°C. A control template was generated using a BAC clone (RP23-474F18) covering the *Nanog* locus, a BAC clone (RP24-352O13) covering the *Phc1* locus, a BAC clone (RP23-438H19) covering the *Oct4* locus and a BAC clone (RP23-230B21) covering the *Lefty1* locus. Ten μg of BAC DNA was digested with 2000 units HaeIII or 1800 units MspI. Random ligation of the fragments was done with 5 units T4 DNA ligase in a total volume of 60 μL. 3C primers were designed for fragments both upstream and downstream of the transcription start site within HaeIII or MspI fragments. Primers Nanog 20, Phc1 48, Oct4 346 and Lefty1 5 were used as the anchor points (Supplementary Table 7). 3C analysis was done, in which every PCR for a primer pair was done in triplicate and quantified. Each data point was corrected for PCR bias by dividing the average of three PCR signals by the average signal in the BAC control template.

Data from ES cells and MEFs were normalized to each other using the interaction frequencies between fragments in control regions (see below for primer pairs and Supplementary Table 7 for sequences). A normalization factor was determined by calculating the log ratio of each interaction frequency within the control region in ES over MEFs, followed by calculating the average of all log ratios. The raw interaction frequencies in ES were subsequently normalized to MEFs using this factor. The same normalization strategy was utilized for normalizing data from *GFP* control shRNA infected cells to *Smc1a* or *Med12* knockdown ES cells. Genomic coordinates for Supplementary Table 7 are build NCBI36/mm8.

The following primer pairs were used for normalization between ES cells and MEFs for the *Nanog* locus (Biological Replicate 1 and 2); Acta2 11 and Acta2 16, Acta2 48 and Acta2 52, Gapdh 17 and Gapdh 19, Gapdh 17 and Gapdh 21, Gapdh 17 and Gapdh 32, Gapdh 21 and Gapdh 39, Gene Desert 5 and Gene Desert 6, Gene Desert 12 and Gene Desert 14, Gene Desert 25 and Gene Desert 26, Gene Desert 12 and Gene Desert 26.

The following primer pairs were used for normalization between ES cells and MEFs for the *Phc1* locus (Biological Replicate 1); Gene Desert 0 and Gene Desert 1, Gene Desert 0 and Gene Desert 2, Gene Desert 27 and Gene Desert 28, Phc1 47 and Phc1 48, Phc1 48 and Phc1 49. The following primer pairs were used for normalization between ES cells and MEFs for the *Phc1* locus (Biological Replicate 2); Gene Desert 0 and Gene Desert 1, Gene Desert 0 and Gene Desert 2, Gene Desert 27 and Gene Desert 28, Acta2 0 and Acta2 1, Acta2 2 and Acta2 7, Acta2 8 and Acta2 9, Acta2 0 and Acta2 13, Gapdh 0 and Gapdh 2, Gapdh 7 and Gapdh 8, Gapdh 9 and Gapdh 12, Gapdh 4 and Gapdh 12.

The following primer pairs were used for normalization between ES cells and MEFs for the *Oct4* locus (Biological Replicate 1); Acta2 11 and Acta2 16, Gapdh 17 and Gapdh 19, Gapdh 17 and Gapdh 21, Gapdh 21 and Gapdh 39, Gene Desert 5 and Gene Desert 6, Gene Desert 12 and Gene Desert 14, Gene Desert 25 and Gene Desert 26, Oct4 346 and Oct4 344, Oct4 346 and Oct4 348. The following primer pairs were used for normalization between ES cells and MEFs for the *Oct4* locus (Biological Replicate 2); Gapdh 17 and Gapdh 19, Gapdh 17 and Gapdh 21, Gapdh 21 and Gapdh 39, Gene Desert 5 and Gene Desert 6, Gene Desert 12 and Gene Desert 14, Gene Desert 25 and Gene Desert 26, Oct4 346 and Oct4 344, Oct4 346 and Oct4 348.

The following primer pairs were used for normalization between ES cells and MEFs for the *Lefty1* locus (Biological Replicate 1 and 2); Gene Desert 0 and Gene Desert 1, Gene Desert 0 and Gene Desert 2, Gene Desert 27 and Gene Desert 28, Acta2 0 and Acta2 1, Acta2 8 and Acta2 9, Acta2 0 and Acta2 13, Gapdh 0 and Gapdh 2, Gapdh 7 and Gapdh 8, Gapdh 9 and Gapdh 12, Gapdh 4 and Gapdh 12.

The following primer pairs were used for normalization between *GFP* control shRNA knockdown cells and *Smc1a* #1 shRNA (See Validation of shRNAs) knockdown cells; Gene Desert 5 and Gene Desert 6, Gene Desert 12 and Gene Desert 14, Gene Desert 25 and Gene Desert 26, Acta2 11 and Acta2 16, Acta2 48 and Acta2 52, Gapdh 17 and Gapdh 19, Gapdh 17 and Gapdh 21, Gapdh 17 and Gapdh 32, Gapdh 21 and Gapdh 39.

The following primer pairs were used for normalization between *GFP* control shRNA knockdown cells and shRNA *Med12* #1 (See Validation of shRNAs) knockdown cells; Gene Desert 5 and Gene Desert 6, Gene Desert 12 and Gene

Desert 14, Gene Desert 25 and Gene Desert 26, Gene Desert 12 and Gene Desert 26, Acta2 11 and Acta2 16, Acta2 48 and Acta2 52, Gapdh 17 and Gapdh 19, Gapdh 17 and Gapdh 21, Gapdh 17 and Gapdh 32, Gapdh 21 and Gapdh 39.

Microarray Analysis

Information regarding the expression levels of mediator and cohesin subunits across a variety of cell types can be found at http://biogps.gnf.org [19].

Cell Culture and RNA Isolation
For ES cell knockdown expression analysis, ES cells were split off MEFs, placed in a tissue culture dish for 45 minutes to selectively remove the MEFs and plated in 6-well plates. The following day cells were infected with lentiviral shRNAs targeting *GFP*, *Smc1a* #1, *Med12* #1 or *Nipbl* #1 (See Validation of shRNAs) in ESC media containing 8 μg/ml polybrene (Sigma, H9268-10G). After 24 hours the media was removed and replaced with ESC media containing 3.5 μg/mL puromycin (Sigma, P8833). Five days post infection RNA was isolated with TRIzol (Invitrogen, 15596-026), further purified with RNeasy columns (Qiagen, 74104) and DNase treated on column (Qiagen, 79254) following the manufacturer's protocols. RNA from two biological replicates was used for duplicate microarray expression analysis with the exception of the Nipbl knockdown expression data.

Microarray Hybridization and Analysis
For microarray analysis, Cy3 and Cy5 labeled cRNA samples were prepared using Agilent's QuickAmp sample labeling kit starting with 1 μg total RNA. Briefly, double-stranded cDNA was generated using MMLV-RT enzyme and an oligo-dT based primer. *In vitro* transcription was performed using T7 RNA polymerase and either Cy3-CTP or Cy5-CTP, directly incorporating dye into the cRNA.
Agilent mouse 4x44k expression arrays were hybridized according to our laboratory's standard method, which differs slightly from the standard protocol provided by Agilent. The hybridization cocktail consisted of 825 ng cy-dye labeled cRNA for each sample, Agilent hybridization blocking components, and fragmentation buffer. The hybridization cocktails were fragmented at 60°C for 30 minutes, and then Agilent 2X hybridization buffer was added to the cocktail prior to application to the array. The arrays were hybridized for 16 hours at 60°C in an Agilent rotor oven set to maximum speed. The arrays were treated with Wash Buffer #1 (6X SSPE / 0.005% n-laurylsarcosine) on a shaking platform at room temperature for 2 minutes, and then Wash Buffer #2 (0.06X SSPE) for 2 minutes at room temperature. The arrays were then dipped briefly in acetonitrile before a final 30 second wash in Agilent Wash 3 Stabilization and Drying Solution, using a stir plate and stir bar at room temperature.

Arrays were scanned using an Agilent DNA microarray scanner. Array images were quantified and statistical significance of differential expression for each

hybridization was calculated using Agilent's Feature Extraction Image Analysis software with the default two-color gene expression protocol. To calculate an average dataset from the biological replicates (*Smc1a* and *Med12* knockdowns) the $\log_{10}$ ratio values for each Agilent Feature were averaged and the log ratio p-values were multiplied (Supplementary Table 3). For each gene in our RefSeq set (see ChIP-Seq analysis section), we selected the Agilent Feature with the best average p-value that was annotated to that gene. Genes with no annotated features were reported as NA. Heatmaps were generated using $\log_2$ ratio values according to the provided color scale.

Determining Genes Co-occupied by Smc1a, Med12 and Nipbl with Expression Changes (Fig. 2d)

Smc1a, Med12 and Nipbl co-occupied regions were initially mapped to a gene if the following criteria were met. The gene had evidence for Smc1a (P-val $<10^{-9}$), Med12 (P-val $<10^{-9}$) and Nipbl (P-val $<10^{-9}$) co-occupancy within the gene body or within 10kb upstream of the transcriptional start site, evidence of Pol2 occupancy (P-val $<10^{-9}$) within the gene body and significant (P-val<0.01) expression changes for a *Smc1a*, *Med12* and *Nipbl* knockdown in independent experiments. Expression data following a *Smc1a*, *Med12* or *Nipbl* knockdown are shown for these genes in Fig. 2d.

Supplementary References

1 Cole, M.F., Johnstone, S.E., Newman, J.J., Kagey, M.H., & Young, R.A., Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells. *Genes Dev* 22 (6), 746-755 (2008).

2 Niwa, H., Miyazaki, J., & Smith, A.G., Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat Genet* 24 (4), 372-376 (2000).

3 Hay, D.C., Sutherland, L., Clark, J., & Burdon, T., Oct-4 knockdown induces similar patterns of endoderm and trophoblast differentiation markers in human and mouse embryonic stem cells. *Stem Cells* 22 (2), 225-235 (2004).

4 Nichols, J. *et al.*, Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* 95 (3), 379-391 (1998).

5 Pereira, L., Yi, F., & Merrill, B.J., Repression of Nanog gene transcription by Tcf3 limits embryonic stem cell self-renewal. *Mol Cell Biol* 26 (20), 7479-7491 (2006).

6 Kent, W.J. *et al.*, The human genome browser at UCSC. *Genome Res* 12 (6), 996-1006 (2002).

7 Boyer, L.A. *et al.*, Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* 122 (6), 947-956 (2005).

8 Moffat, J. *et al.*, A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell* 124 (6), 1283-1298 (2006).

9 Bilodeau, S., Kagey, M.H., Frampton, G.M., Rahl, P.B., & Young, R.A., SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state. *Genes Dev* 23 (21), 2484-2489 (2009).

10 Boyer, L.A. *et al.*, Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* 441 (7091), 349-353 (2006).

11 Marson, A. *et al.*, Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* 134 (3), 521-533 (2008).

12 Seila, A.C. *et al.*, Divergent transcription from active promoters. *Science* 322 (5909), 1849-1851 (2008).

13 Chen, X. *et al.*, Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* 133 (6), 1106-1117 (2008).

14 Mikkelsen, T.S. *et al.*, Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448 (7153), 553-560 (2007).

15 Johnson, D.S., Mortazavi, A., Myers, R.M., & Wold, B., Genome-wide mapping of in vivo protein-DNA interactions. *Science* 316 (5830), 1497-1502 (2007).

16 Guenther, M.G. *et al.*, Aberrant chromatin at genes encoding stem cell regulators in human mixed-lineage leukemia. *Genes Dev* 22 (24), 3403-3408 (2008).

17 Ebmeier, C.C. & Taatjes, D.J., Activator-Mediator binding regulates Mediator-cofactor interactions. *Proc Natl Acad Sci U S A* (2010).

18 Miele, A. & Dekker, J., Mapping cis- and trans- chromatin interaction networks using chromosome conformation capture (3C). *Methods Mol Biol* 464, 105-121 (2009).

19 Wu, C. *et al.*, BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biol* 10 (11), R130 (2009).