Stem Cell Reports, Volume *9*

Supplemental Information

THAP1: Role in Mouse Embryonic Stem Cell Survival and Differentiation

Francesca Aguilo, Zuchra Zakirova, Katie Nolan, Ryan Wagner, Rajal Sharma, Megan Hogan, Chengguo Wei, Yifei Sun, Martin J. Walsh, Kevin Kelley, Weijia Zhang, Laurie J. Ozelius, Pedro Gonzalez-Alegre, Thomas P. Zwaka, and Michelle E. Ehrlich

FIGURE S1, related to FIGURE 3

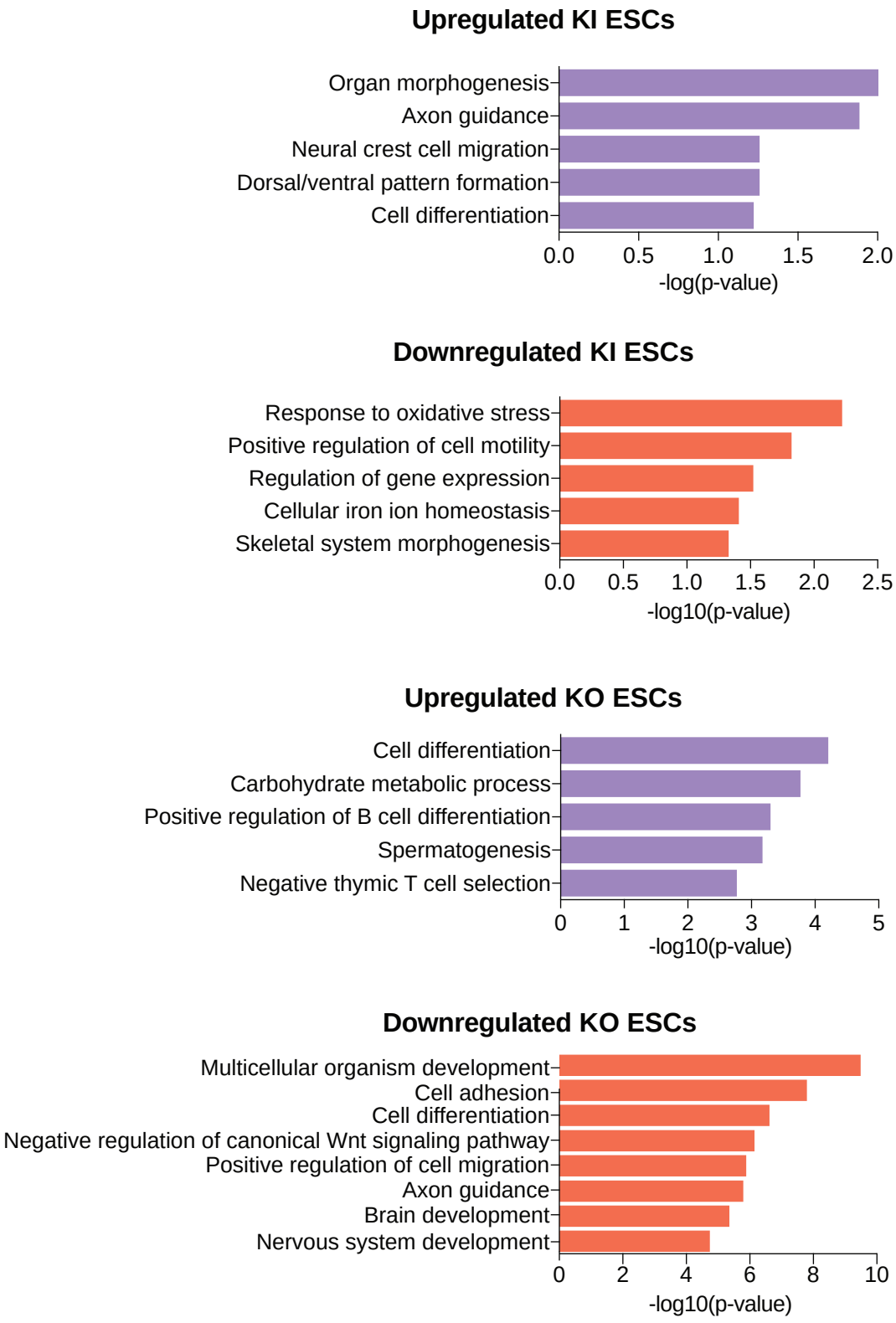
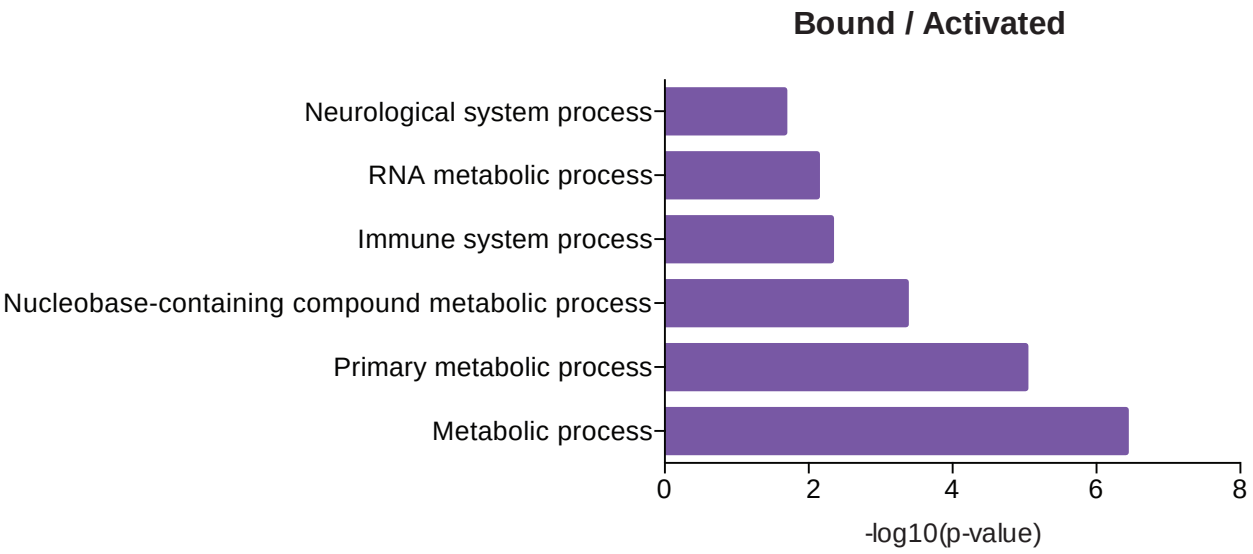
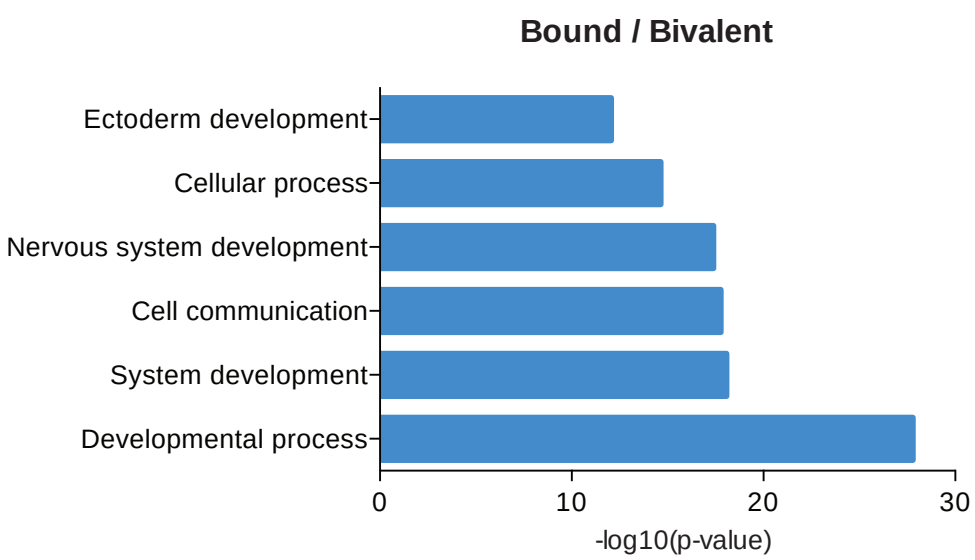


FIGURE S2, related to FIGURE 4

A



B



C

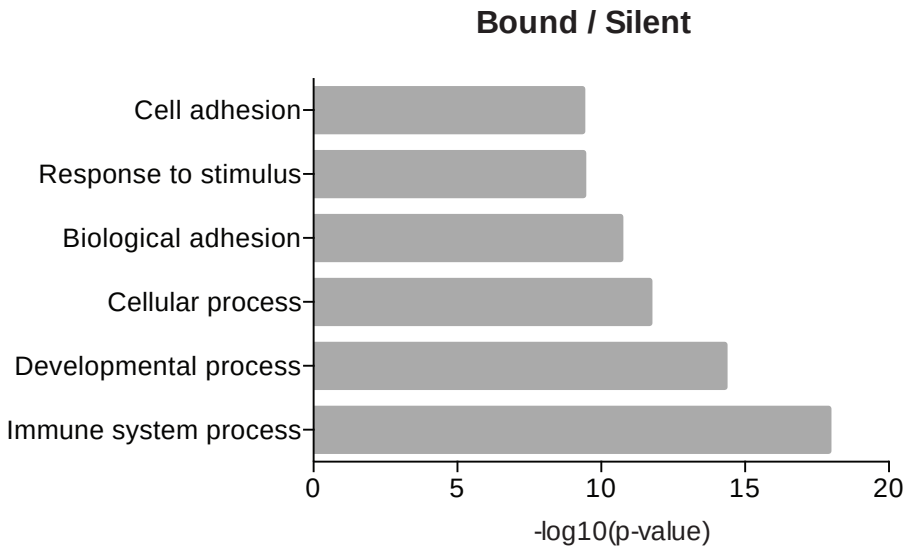


FIGURE S3, related to FIGURE 6

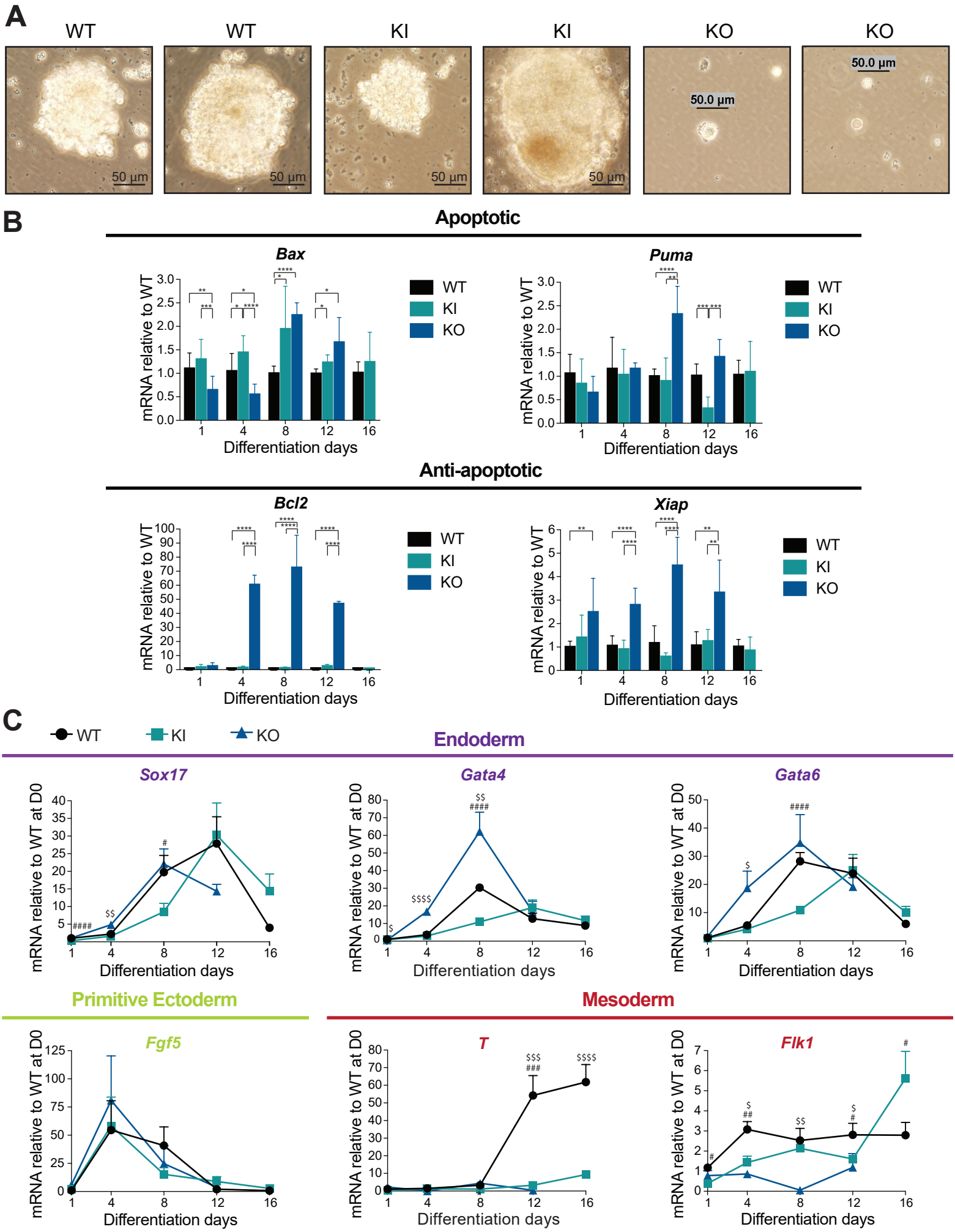


FIGURE S4, related to FIGURE 6

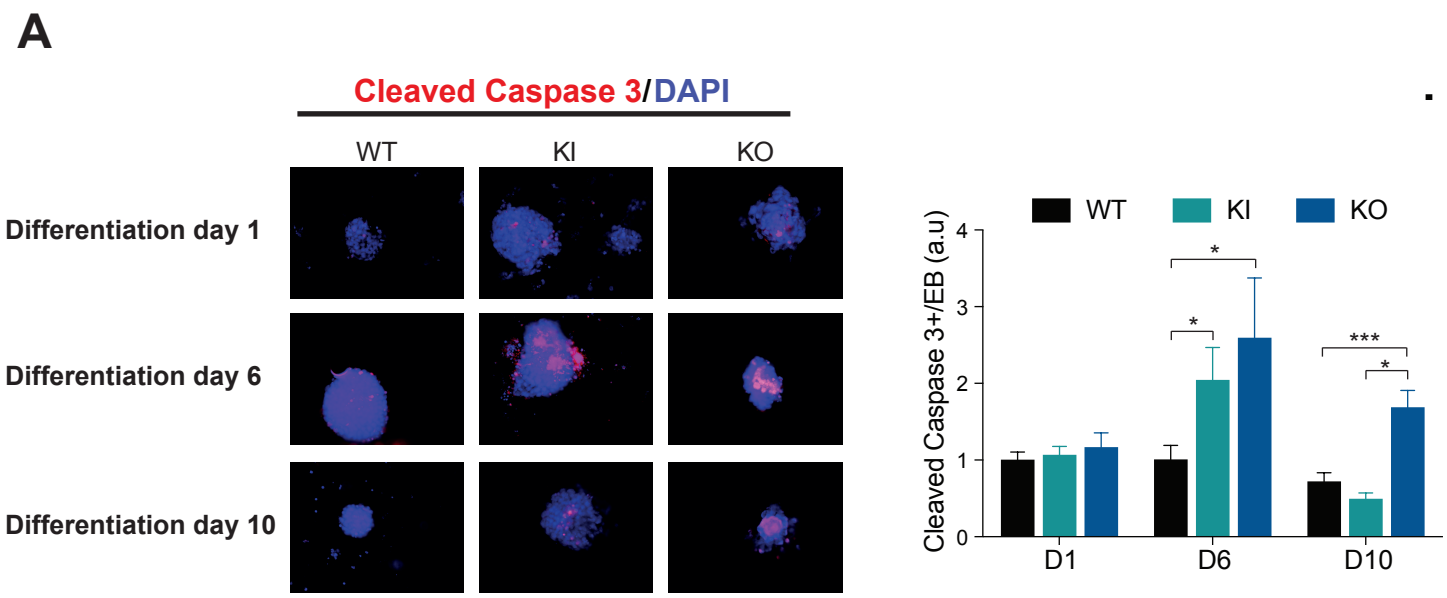
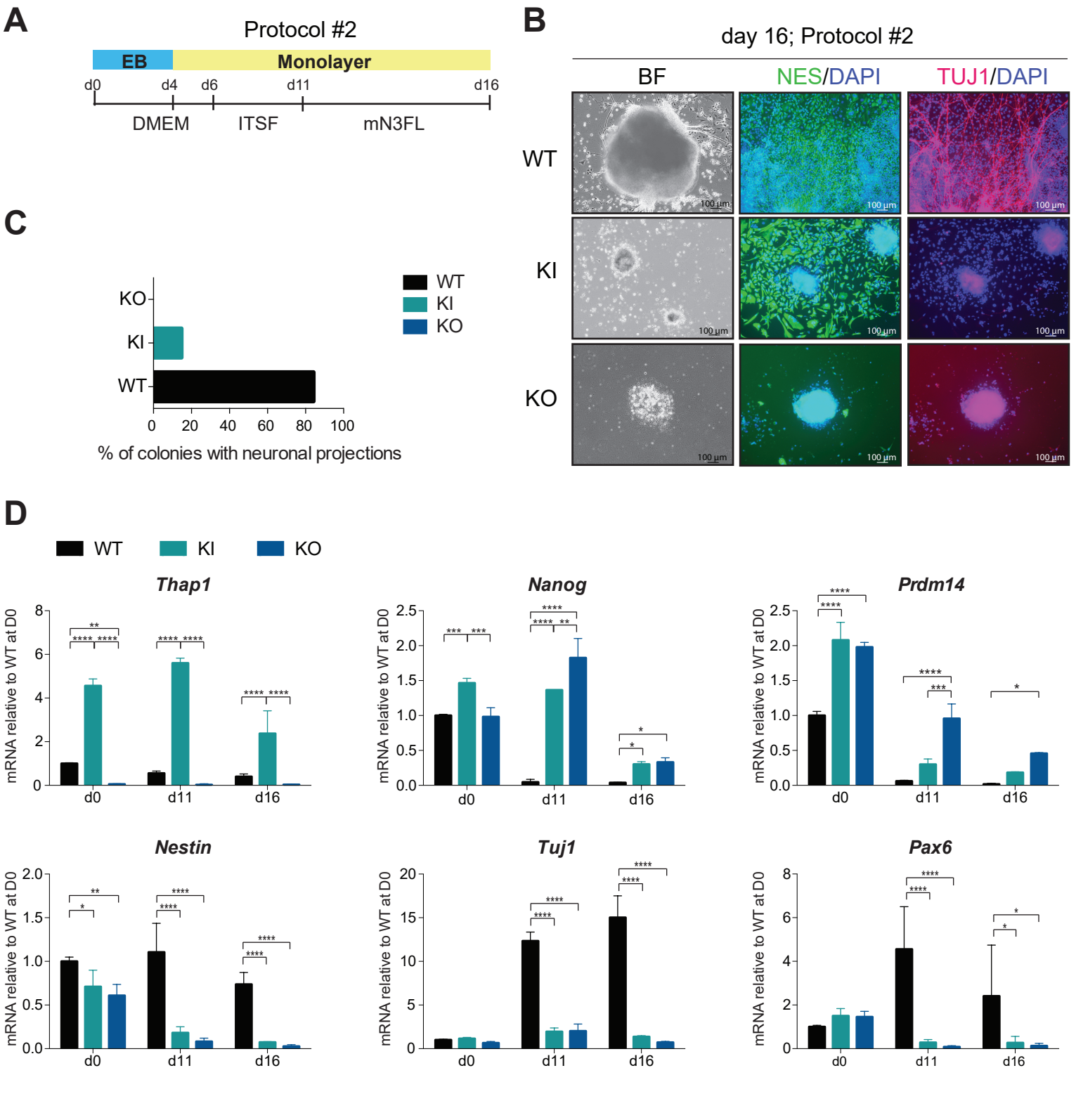


FIGURE S5, related to FIGURE 7



SUPPLEMENTAL FIGURE LEGENDS

Figure S1, related to Figure 3. Gene ontology analysis of uniquely regulated genes.

Gene ontology (GO) analyses of biological processes of uniquely upregulated and downregulated genes in *Thap1*$^{C54Y/C54Y}$ (KI) and *Thap1*$^{\Delta Exon2}$ (KO) ESCs from the RNA-Seq datasets. The p-value for the enrichment of biological process GO-term is shown.

Figure S2, related to Figure 4. Gene ontology analysis of THAP1 bound genes.

Gene ontology (GO) analyses of biological processes of THAP1 bound and activated genes (A), THAP1 bound and bivalent genes (B), and THAP1 bound and silent genes (C) as described under the methods section. The p-value for the enrichment of biological process GO-term is shown.

Figure S3, related to Figure 6. THAP1 is required for normal embryoid body formation.

(A) Bright field showing WT, *Thap1*$^{C54Y/C54Y}$ (KI) and *Thap1*$^{\Delta Exon2}$ (KO) EB morphology. The scale bar represents 50μm.

(B) RT-qPCR analysis of pro-apoptotic genes (*Bax* and *Puma*) and anti-apoptotic genes (*Bcl2* and *Xiap)* during EB differentiation of WT, *Thap1*$^{C54Y/C54Y}$ (KI) and *Thap1*$^{\Delta Exon2}$ (KO) ESCs at the indicated time points. Data are normalized to WT, relative for each day of differentiation. Significant differences were observed by two-way ANOVA among the genotypes when examining expression of *Bax* [F (2,63) = 7.609, p= 0.0011], *Puma* [F (2, 85) = 1481, p = 0.0009], *Bcl2* [F (2,9) = 74.64, p < 0.0001] and *Xiap* [F (2, 67) = 54.39, p < 0.0001]. Bonferroni multiple comparisons test was performed *post hoc* revealing significant differences between the genotypes. Data are represented as mean ± SEM of three independent experiments. *p < 0.05; **p < 0.01, ***p < 0.005, ****p < 0.001.

(C) RT-qPCR analysis of endodermal (*Sox17*, *Gata4* and *Gata6*), mesodermal (*T* and *Flk1*) and primitive ectodermal markers (*Fgf5*) during EB differentiation of WT, *Thap1*$^{C54Y/C54Y}$ (KI) and *Thap1*$^{\Delta Exon2}$ (KO) ESCs at the indicated time points. Data are normalized to WT, relative to D0. The two-way ANOVA results can be found in Table S3. Bonferroni multiple comparisons test was performed *post hoc* revealing significant differences between the genotypes. Data are represented as mean ± SEM of three independent experiments. *p < 0.05; **p < 0.01, ***p < 0.005, ****p < 0.001 relative to WT. Symbol '#' denotes differences between KI vs WT, and symbol '$' denotes differences between KO vs WT.

Figure S4, related to Figure 6. THAP1 is required for ES cell survival

(A) Wild-type (WT), *Thap1*$^{C54Y/C54Y}$ (KI) and *Thap1*$^{\Delta Exon2}$ (KO) EBs were immunostained for cleaved caspase-3 (CC3) on differentiation days 1, 6 and 10. Quantification of CC3+ immunopositive EBs (right panel) was performed by scoring CC3+ immunopositive cells as a function of total EB area, normalized to WT for each differentiation day. Two-way ANOVA was performed, followed by Bonferroni post testing corrections. Data are represented as mean ± SEM; n = 10 per group for each day, data pooled from three independent experiments. *p < 0.05; **p < 0.01, *** p < 0.005, ****p < 0.001.

(B) The pan caspase inhibitor, Z-VAD-FMK does not inhibit cell death in *Thap1*$^{C54Y/C54Y}$ (KI) and *Thap1*$^{\Delta Exon2}$ (KO) EBs over four days of EB differentiation. Two-way ANOVA was performed for each day of treatment, followed by Bonferroni post testing corrections. Data are represented as mean ± SEM of three independent experiments. *p < 0.05; **p < 0.01, ***p < 0.005, ****p < 0.001.

Figure S5, related to Figure 7. Characterization of neural differentiation cell cultures

(A) Schematic diagram of Protocol #2 for neural differentiation.

(B) Immunofluorescence analysis of NESTIN (green; middle panel) and TUJ1 (red; right panel) of WT, *Thap1*$^{C54Y/C54Y}$ (KI) and *Thap1*$^{\Delta Exon2}$ (KO) differentiated cells at the indicated day using protocol #2. Bright field (left panel) and DAPI stain (blue) are shown. The scale bar represents 100 μm.

(C) Bar graph depicting % of colonies with neuronal projection.

(D) RT-qPCR analysis of *Thap1*, pluripotency markers (*Nanog* and *Prdm14*), and ectoderm marker genes (*Nestin*, *Tuj1*, and *Pax6*) during neural differentiation of WT, *Thap1*$^{C54Y/C54Y}$ (KI) and *Thap1*$^{\Delta Exon2}$ (KO) ESCs at the indicated time points. Data are represented relative to WT at day 0. Two-way RM ANOVA was performed for *Thap1*, revealing an interaction effect [F (4,27) = 24.72, p < 0.0001], in addition, significant differences were observed by days [F (2,27) = 31.16, p < 0.0001], and by genotype [F (2,27) = 430.7, p < 0.0001]. Two-way RM ANOVA was performed for *Nanog*, revealing an interaction effect [F (4,16) = 55.52, p < 0.0001], in addition, significant differences were observed by days [F (2,16) = 171.5, p < 0.0001], and by genotype [F (2,16) = 99.32, p < 0.0001]. Two-way RM ANOVA was performed for *Prdm14*, revealing an interaction effect [F (4,9) = 11.84, p = 0.0012], in addition, significant differences were observed by days [F (2,9) = 275.1, p < 0.0001], and by genotype [F (2,9) = 67.04, p < 0.0001]. Two-way RM ANOVA was performed for *Nestin*, revealing an interaction effect [F (4,27) = 6.423, p = 0.0009], in addition, significant differences were observed by days [F (2,27) = 35.89, p < 0.0001], and by genotype [F (2,27) = 85.86, p < 0.0001]. Two-way RM ANOVA was performed for *Tuj1*, revealing an interaction effect [F (4,27) = 76.88, p < 0.0001], significant differences were observed by days [F (2,27) = 95.69, p < 0.0001], and by genotype [F (2,27) = 294.6, p < 0.0001]. Two-way RM ANOVA was performed for *Pax6*, revealing an interaction effect [F (4,26) = 7.176, p = 0.0005], significant differences were observed by genotype [F (2,26) = 14.86, p < 0.0001], but not

by day. Holm-Sidak's multiple comparisons test was performed *post hoc* revealing significant differences between the genotypes. Data are represented as mean ± SEM of at least three independent experiments. *p < 0.05; **p < 0.01, ***p < 0.005, ****p < 0.001.

SUPPLEMENTAL TABLES

Table S1, related to Figure 3: RNA-Seq. Gene expression in KI and KO ESCs compared to WT ESCs. Uniquely regulated genes are also depicted. Fold-change is shown. See separate excel File, Table S1.xlsx

Table S2, related to Figure 4: ChIP-Seq for THAP1 in WT ESCs. See separate excel File, Table S2.xlsx

Table S3, related to Figure 6, Figure 7 and Figure S3: Statistical analyses tables for the indicated figures. See separate excel File, Table S3.xlsx

Table S4, related to experimental procedures: RT-qPCR primers used for gene expression analysis and ChIP validation. Primers used for ESCs and mice genotyping. See separate excel File, Table S4.xlsx

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Antibodies

The following commercially available antibodies were used at the indicated concentrations for western blot: anti-THAP1 (Proteintech, 12584-1-AP, 1:1000), anti-HDAC1 (Abcam, ab7028, 1:1000), and anti-PGK1/2 (SantaCruz, sc-23805, 1:1000). ChIP analysis was performed with the indicated amount of anti-THAP1 antibody and number of cells: (Proteintech, 12584-1-AP; 5 μg antibody for 1 x 10^7 cells). For immunocytochemistry (ICC) analysis SSEA1 (BD Pharmigen, 561560, 1:400), anti-OCT4 (Abcam, ab19857, 1:1000), anti-NESTIN (NovusBio, NB100-1604, 1:1000), anti-β3 Tubulin antibody (TUJ-1) (SantaCruz, sc-58888, 1:250) and anti-cleaved caspase 3 (CC3) (CellSignaling, 9661, 1:400) were used. All cells were counterstained with diamidino-2-phenylindole (DAPI) solution (ThermoFisher, 62248, 1:1000).

Derivation of mouse ESCs

Animal use was approved by the IACUC committee at the Icahn School of Medicine at Mount Sinai. Heterozygous *THAP1*^{C54Y} female mice were super-ovulated by intraperitoneal injection of 5 IU pregnant mare's serum (PMS; National Hormone and Peptide Program) followed 48 hours later by 5 IU of human chorionic gonadotropin (HCG; National Hormone and Peptide *THAP1*^{C54Y} male mice (1 female per male). Three days later, 8-cell stage embryos were isolated from the oviduct and uterus of the females using FHM media (Millipore Specialty Media, MR-024-D), and cultured overnight in KSOM + amino acids media (Millipore Specialty Media, MR-121-D). The next day, 24 blastocysts were plated onto irradiated mouse embryo fibroblasts in a gelatin-coated 24-well plate (one blastocyst per well), and cultured in a 6% CO2 incubator undisturbed for 4 days. All culture steps were performed using ES media with LIF (see below).

Outgrowths that were visible after 4 days of culture were isolated using a micropipettor, and transferred to a small volume of 0.25% trypsin, 2.21 mM EDTA (Corning, 25-053-Cl). After disruption of the outgrowth, the cells were re-plated into 24-well plates with irradiated mouse embryonic fibroblasts. Cell growth was monitored, and clones that contained colonies with ES-like morphology were passaged through 6-well plates to 60 mm dishes. DNA samples were prepared from parallel cultures that were used for genotyping of the resulting ES clones. Of the 24 plated blastocysts, 6 ES clones were derived. Genotyping revealed that clones 5, 14 and 18 were wild-type, clone 9 was heterozygous for *THAP1*^{C54Y}, and clones 6 and 17 were homozygous for *THAP1*^{C54Y}.

For generation of *Thap1*$^{\Delta Exon2}$ ESCs, cre-mediated excision of exon 2 was performed by transfection of 1.5 x 10^6 *Thap1*$^{C54Y/C54Y}$ ESCs with pSalk-Cre (pTZ1111), a mammalian expression vector in which pCMV promoter drives Cre recombinase expression as previously described (Dejosez et al., 2008).

Differentiation assays

Induction of neural differentiation was performed based on generation of EBs, using two different approaches, with and without addition of RA. Protocol #1 was adapted from (Okada et al., 2004). Two-day-old EBs were cultured in N2B27 medium [DMEM/F-12, 0.1 mM 2-mercaptoethanol, N2 supplement (100X) (ThermoFisher), B27 supplement (50X) (ThermoFisher)], and penicillin/streptomycin (ThermoFisher) supplemented with 1 μM RA (Sigma-Aldrich) for 4 days. At day 6, EBs were plated in tissue culture dishes pre-coated with 0.1% gelatin and propagated in the same medium for additional 4 days. Medium was replenished every 2 days. Protocol #2 was adapted from (Okabe et al., 1996) with some modifications. Four-day-old EBs were plated in tissue culture dishes pre-coated with 0.1% gelatin and propagated for 7 days in ITSFn medium [DMEM/F-12 (Dulbecco's Modified Eagle Medium/Nutrient Mixture F-12)] containing insulin (5 μg/mL), transferrin (50 μg/mL), 30 nM selenium chloride, and fibronectin (5 μg/mL). At day 11, cells were cultured in mN3FL medium [DMEM/F12 supplemented with insulin (25 μg/mL), transferrin (50 μg/mL), progesterone (20 nM), putrescine (100 μM), selenium chloride (30 nM), bFGF (5 ng/mL), and laminin (1 μg/mL)] for an additional 5 days. Cells were harvested for extraction of total RNA and ICC at the indicated time points.

Immunofluorescence staining

Cells were fixed in 4% paraformaldehyde (PFA) at room temperature (RT) for 20 min and then permeabilized at room temperature (RT) for 30 min with 0.25% Triton-X-100 in TBS. Cells were then washed twice with PBS and blocking was performed for 30 min with 3% goat serum in TBS. The cells were stained with the specified antibodies overnight at 4°C in 1% goat serum in TBS. After overnight incubation, the cover slips were washed thoroughly with TBS and incubated in either goat anti-rabbit IgG H&L (Alexa Fluor® 488, ab150077, 1:400), goat anti-mouse IgG H&L (Alexa Fluor® 594, ab150116, 1:400) or goat anti-chicken IgY H&L (Alexa Fluor® 488, ab150169, 1:400) secondary antibody (Abcam) for one hour. For nuclear staining, 4', 6-diamidino-2-phenylindole (DAPI Solution (1 mg/mL) diluted 1:1000 in DPBS) was added to each well and incubated for 10 minutes, after which the fixed cells were again washed with TBS three times. Cells were examined under an Olympus IX51 inverted fluorescent microscope.

Terminal deoxynucleotidyl transferase dUTP nick end labeling (TUNEL) immunocytochemistry

On differentiation days 1, 6 and 10 embryoid bodies (EBs) were fixed with 4% paraformaldehyde (PFA). EBs were assayed with Roche *In Situ* Cell Death Detection Kit (Catalog No. 11684795910; Sigma-Aldridge) per manufacturer's instructions followed by nuclear counterstaining using 4', 6-diamidino-2-phenylindole (DAPI Solution (Catalog No. 62248, ThermoFisher Scientific). EBs were examined under an Olympus BX61 fluorescent microscope. For quantification of TUNEL + immunopositive EBs, a minimum of ten high-power fields from each genotype per day were chosen and quantified using a CellProfiler (Carpenter et al., 2006; Kamentsky et al., 2011) pipeline which scored TUNEL immunopositive cells as a function of total EB area.

Cell proliferation assay with Z-VAD-FMK

Cell viability was assessed on the first day of EB differentiation with and without Z-VAD-FMK pan caspase inhibitor (Selleckchem) as previously described (Lubitz et al., 2007). Briefly, mass culture of EBs was started with 1.2×10^6 ES cells per bacterial dish by using standard EB medium without LIF. Resulting EBs were grown in the presence of 40 μM Z-VAD-FMK prepared in dimethyl sulfoxide (DMSO) or the same vol/vol pure DMSO and percentage of cell viability was assessed on four consecutive days (in triplicates) using Muse™ Cell Analyzer (Millipore; Billerica, MA) following the manufacturer's instructions. Z-VAD-FMK was replenished daily. Experiment was performed in quadruplicate.

Preparation of nuclear extracts

Cells were expanded in 100 mm diameter tissue culture plates, washed with cold PBS, and scraped off, followed by nuclear extract preparation as previously described with some modifications (Dignam et al., 1983). Briefly, the cellular pellet was resuspended in at least 5 volumes of buffer A (10 mM HEPES pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl, 1 mM DTT) in the presence of protease inhibitors (3 μg/ml aprotinin, 750 μg/ml benzamidine, 1 mM phenylmethylsulfonyl fluoride, 5 mM NaF and 2 mM sodium orthovanadate) and incubated for 10 minutes in ice. Following centrifugation, the pellet was resuspended in 2 volumes of buffer A, homogenized with a Dounce homogenizer, and spun down at maximum speed for 5 minutes. Nuclei were then resuspended in 2 volumes of buffer B (20 mM HEPES pH 7.9, 1.5 mM $MgCl_2$, 500 mM NaCl, 25% Glycerol, 0.5 mM EDTA, 1 mM DTT) supplemented with protease inhibitors and incubated in a rotator at 4 degrees for 30 minutes. The resulting samples were spun down at maximum speed for 15 minutes and frozen at -80 ^oC for further analysis.

RT-qPCR analysis

Total RNA purification was performed with the miRNeasy mini kit (Qiagen), and was carried out according to the manufacturer's instructions. Five hundred nanograms of RNA were reversed-transcribed using the High Capacity RNA-to-cDNA Kit (Applied Biosystems, Foster City, CA, USA). The cDNA solution was subjected to real-time qPCR in a Step-One Plus system (Applied Biosystems) using the PerfeCTa SYBR Green FastMix ROX (Quanta BioSciences). Quantitative PCR consisted of 40 cycles, 15 s at 95 °C and 30 s at 60 °C each, followed by dissociation curve analysis. Gene expression specific primers used for this study are listed in **Table S4**. *β-actin* mRNA was assayed as loading control.

Chromatin immunoprecipitation with high-throughput sequencing (ChIP-Seq)

THAP1 ChIP-Seq was carried out using methods described in (Hnisz et al., 2016). THAP1 mouse embryonic stem cells (500 million cells) were dual-crosslinked at room temperature with 1.5 mM ethylene glycol bis (sulfosuccinimidyl succinate) (EGS) for 30 minutes, followed by 1% formaldehyde for 10 min, and then neutralized with 125 mM glycine. Crosslinked cells were washed three times with ice-cold PBS, snap-frozen in liquid nitrogen, and stored at -80°C before further processing. 250 μl of Protein G Dynabeads (Life Technologies) were blocked with 0.5% BSA (w/v) in PBS, and pre-incubated with 25 μg of anti-THAP1 antibody (validated for ChIP in Gavarini et al., 2010). Nuclei were isolated as previously described (Lee et al., 2006), and sonicated in lysis buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM EDTA pH 8.0, 0.1% SDS, and 1% Triton X-100) on a Fisher Sonic Dismembrator for 14 cycles at 30s each on ice (power 5) with 60 s between cycles. Sonicated lysates were cleared once by centrifugation and incubated overnight at 4°C with magnetic beads bound with antibody. Protein-DNA complexes were sequentially washed with buffer A (50 mM HEPES-KOH pH7.9, 140 mM NaCl, 1 mM EDTA pH 8.0, 0.1% Na-Deoxycholate, 1% Triton X-100, 0.1% SDS), buffer B (50 mM HEPES-KOH pH7.9, 500 mM NaCl, 1 mM EDTA pH 8.0, 0.1% Na-Deoxycholate, 1% Triton X-100,

0.1% SDS), buffer C (20 mM Tris-HCl pH8.0, 250 mM LiCl, 1mM EDTA pH 8.0, 0.5% Na-Deoxycholate, 0.5% IGEPAL C-630 0.1% SDS) and buffer D (TE with 50 mM NaCl).

Library preparation

ChIP-Seq libraries were prepared using an adaptation of the Nextera Library Preparation protocol (Illumina), as described in (Hnisz et al., 2016). ChIP DNA fragments were end-repaired using T4 DNA polymerase (NEB) followed by A-tailing with Klenow (NEB). A biotinylated bridge linker (F: /5Phos/CGCGATATC/iBiodT/TATCTGACT; R: /5Phos/GTCAGATAAGATATCGCGT) with T- overhangs was added and the proximity ligation was performed overnight at 16°C in 1.5 mL volume. Un-ligated DNA was digested with exonuclease and lambda nuclease (NEB). DNA was eluted off the beads in elution buffer (50 mM Tris-HCL pH 8.0, 10 mM EDTA, 1% SDS) followed by overnight crosslink reversal, RNAse A treatment, and proteinase K digestion. A phenol:chloroform:isoamyl alcohol extraction was performed followed by an ethanol precipitation. Precipitated DNA was resuspended in Nextera DNA resuspension buffer (Illumina). The DNA was then tagmented with the Nextera Tagmentation kit (Illumina). The tagmented library was purified with a Zymo column and was bound to Streptavidin beads to enrich for ligation junctions (containing the biotinylated bridge linker). 15 cycles of the polymerase chain reaction were performed to amplify the library. The amplified library was size selected (200-500 bp) with a Pippin prep machine and sequenced with 100x100 paired-end sequencing on an Illumina Hi-Seq 2500 platform.

ChIP-Seq analysis

All ChIP-Seq datasets were processed with custom script ("Dowen et al. pipeline") as previously described (Dowen et al., 2014; Hnisz et al., 2016). Image analysis and base calling was done using the Solexa pipeline. Reads were examined for the presence of at least 10 base pairs of linker sequence. Reads that did not contain linker were not processed further. Reads containing linker were trimmed using cutadapt (cutadapt -m 17 –a forward=ACGCGATATCTTATCTGACT –a reverse=AGTCAGATAAGATATCGCGT -- overlap 10) (http://code.google.com/p/cutadapt/). Trimmed mate pairs were mapped independently to mm9 using Bowtie version 1.1.1 (bowtie -e 70 -k 1 -m 1 -v 2 -p 4 --best --strata --S) (Langmead et al., 2009). To remove PCR bias artifacts, reads were filtered for redundancy: PETs with identical genomic coordinates and strand information at both ends were collapsed into a single PET. Regions of local enrichment (PET peaks) were called using MACS 1.4.2 (Zhang et al., 2008) with the parameters "-p 1e-09-no-lambda –no-model", and an FDR of 0.01. Genes within 1 Kb of THAP1 peaks were identified using BETA-minus (Wang et al., 2013).

ChIP-Seq signal tracks were presented by Integrative Genomics Viewer (IGV) software (http://www.broadinstitute.org/igv/). THAP1 *de-novo* motif analysis and transcription factor binding sequence motifs enriched in THAP1 peak sequences were obtained with MEME-ChIP software (http://meme.nbcr.net/meme/cgi-bin/meme-chip.cgi) (Machanick and Bailey, 2011).

To examine chromatin patterns, enrichment profiles around THAP1 binding peaks with previously reported ESC ChIP-Seq datasets were generated. Each gene in the mouse genome was also classified into (i) active, containing H3K4me3 at the TSS and di-methylated histone H3 Lys 79 (H3K79me2) within the first 5 Kb of the gene body; (ii) bivalent or poised, associated with H3K4me3 and H3K27me3 at the TSS and (iii) silent, associated with the absence of H3K4me3 and H3K79me2 and the possibility to contain H3K27me3, and overlapped the THAP1-bound peaks. Below is the list of ChIP-Seq datasets used, corresponding GEO Accession numbers and publication:

Datasets	GEO Accession Number	Publication
p300	GSM918750	(Consortium, 2012; Yue et al., 2014)
MED1	GSM560347, GSM560348	(Kagey et al., 2010)
H3K4me1	GSM769009	(Yue et al., 2014)
H3K27ac	GSM594578, GSM594579	(Creyghton et al., 2010)
Pol2 Ser-5	GSM515662	(Rahl et al., 2010)
H3K4me3	GSM769008, GSE11724	(Marson et al., 2008; Yue et al., 2014)
H3K79me2	GSE11724	(Marson et al., 2008)
H3K27me3	GSM1000089, GSM307619	(Mikkelsen et al., 2007; Yue et al., 2014)

ChIP qRT-PCR validation

ChIP experiments were performed as described (Aguilo et al., 2016). Fold enrichment over 10% input was calculated using the 2DeltaCt method. The primer sets used for ChIP analysis are listed in **Table S4**.

RNA-Seq library preparation

RNA-Seq library preparation was performed at the Weill Cornell Medical College Genomic Core facility (New York) using the TrueSeq RNA sample preparation kit (Illumina RS-122-2001) as per manufacturer's recommendations. Samples were sequenced by the Illumina HiSeq 2500 platform (Illumina) as 100 bp pair-ended reads.

RNA-Seq analysis

RNA-Seq reads were filtered and trimmed using Trimmomatic (Bolger et al., 2014) and then aligned to the mm9 mouse reference genome and indexes based on UCSC annotations using TopHat (Trapnell et al., 2009). HT-seq (Anders et al., 2015) was used to find the read counts across the UCSC reference genome. Differentially expressed genes were identified by the R package DESeq2 (Love et al., 2014) using a false discovery rate (FDR) < 0.1 and fold-change >1.5.

Heat maps

Heat maps were derived using counts transformed by the varianceStabilizingTransformation function in DESeq2. Counts were then normalized by Z-score across samples and plotted.

SUPPLEMENTAL REFERENCES

Aguilo, F., Li, S., Balasubramaniyan, N., Sancho, A., Benko, S., Zhang, F., Vashisht, A., Rengasamy, M., Andino, B., Chen, C.H., *et al.* (2016). Deposition of 5-Methylcytosine on Enhancer RNAs Enables the Coactivator Function of PGC-1alpha. Cell Rep *14*, 479-492.

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

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

Carpenter, A.E., Jones, T.R., Lamprecht, M.R., Clarke, C., Kang, I.H., Friman, O., Guertin, D.A., Chang, J.H., Lindquist, R.A., Moffat, J., *et al.* (2006). CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biol *7*, R100.

Consortium, E.P. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57-74.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., *et al.* (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A *107*, 21931-21936.

Dignam, J.D., Lebovitz, R.M., and Roeder, R.G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res *11*, 1475-1489.

Dowen, J.M., Fan, Z.P., Hnisz, D., Ren, G., Abraham, B.J., Zhang, L.N., Weintraub, A.S., Schuijers, J., Lee, T.I., Zhao, K., *et al.* (2014). Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. Cell *159*, 374-387.

Hnisz, D., Weintraub, A.S., Day, D.S., Valton, A.L., Bak, R.O., Li, C.H., Goldmann, J., Lajoie, B.R., Fan, Z.P., Sigova, A.A., *et al.* (2016). Activation of proto-oncogenes by disruption of chromosome neighborhoods. Science *351*, 1454-1458.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., *et al.* (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430-435.

Kamentsky, L., Jones, T.R., Fraser, A., Bray, M.A., Logan, D.J., Madden, K.L., Ljosa, V., Rueden, C., Eliceiri, K.W., and Carpenter, A.E. (2011). Improved structure, function and compatibility for CellProfiler: modular high-throughput image analysis software. Bioinformatics *27*, 1179-1180.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550.

Lubitz, S., Glaser, S., Schaft, J., Stewart, A.F., and Anastassiadis, K. (2007). Increased apoptosis and skewed differentiation in mouse embryonic stem cells lacking the histone methyltransferase Mll2. Mol Biol Cell *18*, 2356-2366.

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

Marson, A., Levine, S.S., Cole, M.F., Frampton, G.M., Brambrink, T., Johnstone, S., Guenther, M.G., Johnston, W.K., Wernig, M., Newman, J., *et al.* (2008). Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell *134*, 521-533.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P., *et al.* (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553-560.

Okabe, S., Forsberg-Nilsson, K., Spiro, A.C., Segal, M., and McKay, R.D. (1996). Development of neuronal precursor cells and functional postmitotic neurons from embryonic stem cells in vitro. Mech Dev *59*, 89-102.

Okada, Y., Shimazaki, T., Sobue, G., and Okano, H. (2004). Retinoic-acid-concentration-dependent acquisition of neural cell identity during in vitro differentiation of mouse embryonic stem cells. Dev Biol *275*, 124-142.

Rahl, P.B., Lin, C.Y., Seila, A.C., Flynn, R.A., McCuine, S., Burge, C.B., Sharp, P.A., and Young, R.A. (2010). c-Myc regulates transcriptional pause release. Cell *141*, 432-445.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105-1111.

Wang, S., Sun, H., Ma, J., Zang, C., Wang, C., Wang, J., Tang, Q., Meyer, C.A., Zhang, Y., and Liu, X.S. (2013). Target analysis by integration of transcriptome and ChIP-seq data with BETA. Nat Protoc *8*, 2502-2515.

Yue, F., Cheng, Y., Breschi, A., Vierstra, J., Wu, W., Ryba, T., Sandstrom, R., Ma, Z., Davis, C., Pope, B.D., *et al.* (2014). A comparative encyclopedia of DNA elements in the mouse genome. Nature *515*, 355-364.

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