THAP1: Role in Mouse Embryonic Stem Cell Survival and Differentiation

Francesca Aguilo,1,2,10* Zuchra Zakirova,3 Katie Nolan,3 Ryan Wagner,3,5 Rajal Sharma,1 Megan Hogan,4,5 Chengguo Wei,6 Yifei Sun,1,2 Martin J. Walsh,1,7 Kevin Kelley,4 Weijia Zhang,6 Laurie J. Ozelius,8 Pedro Gonzalez-Alegre,9 Thomas P. Zwaka,4,5 and Michelle E. Ehrlich2,3,7,*

1Department of Pharmacological Sciences 2Department of Pediatrics 3Department of Neurology 4Department of Cell, Developmental and Regenerative Biology 5Black Family Stem Cell Institute 6Department of Medicine Bioinformatics Core 7Department of Genetics and Genomic Sciences 8Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA 9Department of Neurology, Harvard Medical School, Massachusetts General Hospital, Charlestown, MA 02129, USA 10Present Address: Department of Medical Biosciences, Wallenberg Center for Molecular Medicine, Umeå University, 901 85 Umeå, Sweden

*Correspondence: francesca.aguilo@umu.se (F.A.), michelle.ehrlich@mssm.edu (M.E.E.)
http://dx.doi.org/10.1016/j.stemcr.2017.04.032

SUMMARY

THAP1 (THAP [Thanatos-associated protein] domain-containing, apoptosis-associated protein 1) is a ubiquitously expressed member of a family of transcription factors with highly conserved DNA-binding and protein-interacting regions. Mutations in THAP1 cause dystonia, DYT6, a neurologic movement disorder. THAP1 downstream targets and the mechanism via which it causes dystonia are largely unknown. Here, we show that wild-type THAP1 regulates embryonic stem cell (ESC) potential, survival, and proliferation. Our findings identify THAP1 as an essential factor underlying mouse ESC survival and to some extent, differentiation, particularly neuroectodermal. Loss of THAP1 or replacement with a disease-causing mutation results in an enhanced rate of cell death, prolongs Nanog, Prdm14, and/or Rex1 expression upon differentiation, and results in failure to upregulate ectodermal genes. ChIP-Seq reveals that these activities are likely due in part to indirect regulation of gene expression.

INTRODUCTION

THAP1 (THAP [Thanatos-associated protein] domain-containing, apoptosis-associated protein 1) is a member of a large family of proteins which are primarily transcription factors (Gervais et al., 2013; Roussigne et al., 2003). The THAP domain, an atypical zinc finger (CysX2–4CysX35–53 CysX-His), is highly conserved and is part of the DNA-binding domain (DBD) with homology to P-transposable elements (Majumdar and Rio, 2015). THAP1 mutations cause DYT6 dystonia (Fuchs et al., 2009), and mutations are located throughout the protein, with about 50% in the DNA-binding domain. Importantly, recessive mutations have been identified (Houlden et al., 2010; Schneider et al., 2011; Xiromerisiou et al., 2012). THAP1 functions, targets, and the mechanisms by which its mutations lead to dystonia are largely unknown, including the effects of mutations on DNA binding (Campagne et al., 2012).

Functional studies of THAP1 in human umbilical vein endothelial cells (HUVECs) show a role in the S phase of mitosis via modulation of pRb-E2F cell-cycle target genes, including RRM1 (Clouaire et al., 2005). In vitro, a coiled-coil domain is required for dimerization (Sengel et al., 2011). Other interactors include prostate apoptosis response-4 protein (Par-4), an effector of cell death linked to prostate cancer and neurodegenerative diseases (Roussigne et al., 2003); HCF-1, a transcriptional co-activator involved in cell-cycle regulation; and O-GlcNAc transferase (OGT), which catalyzes the addition of O-GlcNAc and thereby also participates in epigenetic regulation of gene expression with an essential function in dividing cells (Mazars et al., 2010).

In mouse models of DYT6 which harbor either a disease-causing C54Y mutation in the DBD or a null allele (ΔExon2) (Ruiz et al., 2015), rare homozygous embryos survived to day 14. They were small with defects in peripheral organs and brain, which showed deficits in the number and morphology of neurons. To study the impact of the mutant alleles on stem cell maintenance and differentiation, we generated mouse embryonic stem cells (mESCs) homozygous for either the C54Y (Thap1C54YExon2) or ΔExon2 (Thap1ΔExon2) alleles. Herein, we characterize both ESCs, which are viable with intact stem cell characteristics, but with abnormalities in cell death, cell cycle, and proliferation rate, that are more severe in the ΔExon2 than in the C54Y homozygote. Furthermore, we show that during differentiation of embryoid bodies (EBs), wild-type THAP1 is required for repression of a cohort of core pluripotency-associated genes and survival in ΔExon2 cells, and...
Figure 1. Generation of Thap1-Recombinant ESCs
(A) Structure of Thap1 gene and its encoded protein.
(B) Genotyping of wild-type (WT), Thap1<sup>C54Y/C54Y</sup> (KI), and Thap1<sup>-/-Exon2</sup> (KO) ESCs by PCR and comparison with the pattern of WT, Thap1<sup>C54Y/+</sup> (WT/KI), and Thap1<sup>-/-Exon2/+</sup> (WT/KO) heterozygote mice. 
(C) Thap1 exon 2 (Thap1 Ex 2) transcript level measured by qRT-PCR in wild-type (WT), Thap1<sup>C54Y/C54Y</sup> (KI), and Thap1<sup>-/-Exon2</sup> (KO) ESCs. An ANOVA was performed which revealed a significant difference among the genotypes (F(2,23) = 91.69, p < 0.001). The Holm-Sidak multiple comparisons test was performed post hoc, revealing significant differences between the genotypes. Data are presented as mean ± SEM of three independent experiments. *p < 0.05; ****p < 0.001.
(D) Distribution of THAP1 protein in nuclear and cytoplasmic fraction of wild-type (WT), Thap1<sup>C54Y/C54Y</sup> (KI) and Thap1<sup>-/-Exon2</sup> (KO) ESCs. Histone deacetylase 1 (HDAC1) and phosphoglycerate kinase 1/2 (PGK1/2) were used as control of nuclear and cytoplasmic extract purity, respectively.
(E) Level of Thap1 transcript spanning exon 1 and exon 3 (Thap1 Ex 1–3) measured by qRT-PCR in wild-type (WT), Thap1<sup>C54Y/C54Y</sup> (KI), and Thap1<sup>-/-Exon2</sup> (KO) ESCs. ANOVA revealed a significant difference among the genotypes (F(2,21) = 73.85, p < 0.0001). The Holm-Sidak
(legend continued on next page)
apparently for terminal neuronal differentiation in C54Y homozygote surviving cells.

RESULTS

Generation of Thap1-Recombinant Embryonic Stem Cells

To explore the function of THAP1 in ESCs, we derived ESCs from two mouse alleles, (1) Thap1C54Y, a constitutive knockin (KI) of the C54Y causative mutation in the DBD of THAP1, and (2) Thap1ΔExon2, a constitutive knockout (KO), allele lacking exon 2 (Ruiz et al., 2015; Figure 1A). Genotypes of wild-type (WT), Thap1C54Y/Thap1C54Y, and Thap1ΔExon2/ΔExon2 ESCs were analyzed by PCR and compared with WT, Thap1C54Y+/+, and Thap1ΔExon2+/+ heterozygote mice (Figure 1B). Consistent with THAP1 autorepression (Erogullari et al., 2014) and failure of THAP1C54Y to bind at the promoter (Gavarini et al., 2010), Thap1C54Y cells exhibited higher levels of Thap1 mRNA than WT ESCs, whereas full-length Thap1 mRNA was undetectable in Thap1ΔExon2 ESCs (Figure 1C). THAP1 antibodies recognize several THAP1-like immunoreactive (THAP1-LIR) species (Ortiz-Virumbrales et al., 2014). Subcellular fractionation followed by western blot analysis revealed the presence of three distinct THAP1-LIR species in the nuclear fraction at ~29 kDa, ~50 kDa, and ~75 kDa (Figure 1D, upper panel). Only the ~75-kDa species was induced or drastically reduced in Thap1C54Y or Thap1ΔExon2 ESCs, respectively, following the same pattern of the corresponding mRNA as assessed by qRT-PCR (Figure 1C). Thus, the ~29- and ~50-kDa species appear to be largely composed of cross-reacting proteins in ESCs. In murine brain, the ~29-kDa species was also non-specific, whereas the 50- and 75-kDa THAP1-LIR species were nuclear and neuron specific, and virtually undetectable in ΔExon2 embryos (Ortiz-Virumbrales et al., 2014). Primers spanning exon 1 and exon 3 of Thap1 mRNA amplified a de-repressed transcript, i.e., its expression was induced, in Thap1C54Y/Thap1C54Y and Thap1ΔExon2 ESCs (Figure 1E). This naturally occurring THAP1ΔExon2 represents less than 1% relative to the major isoform (containing exon 2) (Figure 1F) and in vivo does not substitute for the loss of the full-length isoform.

Loss of THAP1 Alters ESC Viability

To investigate THAP1 function in self-renewal, we analyzed the expression of stage-specific embryonic antigen 1 (SSEA1) and POU5F1 (Palmqvist et al., 2005) in all three genotypes (Figure 2A, top and middle panels). There were no detectable differences. The level of alkaline phosphatase (AP) staining (Figure 2A, bottom panels), characteristic of undifferentiated ESCs, was also genotype independent. These data suggest that THAP1 is not required to maintain the pluripotency state of ESCs.

To assess the proliferation rates of WT, Thap1C54Y, and Thap1ΔExon2 ESCs, we counted the cells at predetermined times after plating. Thap1C54Y ESC proliferation rate was slightly slower than in WT cells, whereas deletion of THAP1 Exon2 severely compromised proliferation (Figure 2B). Viability was not affected in Thap1C54Y ESCs as indicated by annexin V, whereas the number of viable cells was reduced ~25% in Thap1ΔExon2 ESCs compared with WT (Figure 2C). Propidium iodide (PI) staining followed by DNA flow cytometry revealed a significant reduction in G2/M-phase cell populations and a corresponding increase in the number of cells in S phase (Figure 2D) in the Thap1ΔExon2 ESCs. Again, Thap1C54Y showed no significant differences compared with WT ESCs. Thus, loss of full-length THAP1, but not the C54Y mutation, affects cell viability by increasing the rate of cell death and arresting the cell cycle at the S phase.

Global Expression Profile of Thap1C54Y and Thap1ΔExon2 ESCs

We studied the global transcriptional response in both ESC lines via high-throughput RNA sequencing (RNA-Seq), and identified a larger number of differentially regulated genes in Thap1ΔExon2 than in Thap1C54Y ESCs (Figure 3A and Table S1). Specifically, comparing Thap1C54Y with WT ESCs, there were 391 upregulated and 163 downregulated genes, whereas comparing Thap1ΔExon2 with WT we identified 1,245 upregulated and 1,237 downregulated genes (Figure 3A). Notably, 93% of the upregulated and 89.5% of the downregulated Thap1C54Y genes were common to both genotypes (Figure 3B), but a heatmap illustrated that Thap1C54Y and Thap1ΔExon2 alleles have distinct effects on the transcriptional ESC profile (Figure 3C). These differences were also reflected by gene ontology (GO) analyses of biological processes of uniquely differentially expressed genes (Figure S1), and could not be attributed to metabolic or apoptotic gene expression dysregulation. Interestingly, GO analyses of biological processes of upregulated genes (log2 fold change = +1 of the Thap1ΔExon2 ESC dataset) revealed categories related to embryonic pattern specification, synapsis, chromosome organization, meiosis, and negative regulation of cell differentiation (Figure 3D,

multiple comparisons test was performed post hoc, revealing significant differences between the genotypes. Data are presented as mean ± SEM of at least three independent experiments. ****p < 0.001.

(F) Comparison of Thap1 isoforms by qRT-PCR in wild-type (WT), Thap1C54Y/Thap1C54Y (KI), and Thap1ΔExon2/ΔExon2 (KO) ESCs. Data are presented as mean ± SEM.
upper panel). Analysis of downregulated genes (log2 fold change = -1 of the Thap1\textsuperscript{dExon2} ESC dataset) demonstrated enrichment for processes involved in neuron development, including neuron differentiation, axon guidance, axonogenesis, and cell projection organization (Figure 3D, lower panel). Known pluripotency genes, including Rex1, Prdm14, and Nanog were increased in both Thap1\textsuperscript{C54Y} and Thap1\textsuperscript{dExon2} ESCs, whereas Klf4, Esrrb, and Pou5F1 were increased only in Thap1\textsuperscript{dExon2} ESCs, all compared with WT ESCs (Figure 3E). In contrast, ectodermal markers, e.g., Gbx2 and Sox11, were decreased in Thap1\textsuperscript{C54Y} ESCs, and Pax6, Sox11, Olig2, and Fgf5 were decreased in Thap1\textsuperscript{dExon2} ESCs.
Figure 3. THAP1 Negatively Regulates the ESC Transcriptome

(A) Bar graph illustrating the numbers of upregulated and downregulated genes in Thap1^{C54Y/C54Y} (KI) and Thap1^{dExon2} (KO) compared with wild-type (WT) ESCs by RNA sequencing (RNA-Seq).

(B) Venn diagram showing the overlap of upregulated (left panel) and downregulated (right panel) genes in Thap1^{C54Y/C54Y} (KI) and Thap1^{dExon2} (KO) ESCs.

(C) Heatmap illustrating the global gene expression performed in duplicates of wild-type 1 and 2 (WT_1 and WT_2), Thap1^{C54Y/C54Y} 1 and 2 (KI_1 and KI_2), and Thap1^{dExon2} 1 and 2 (KO_1 and KO_2) ESCs.

(legend continued on next page)
ESCs, all compared with WT. Endoderm specification genes, including *Gata4*, *Gata6*, *Sox17*, and *Foxa2*, were also significantly decreased in both genotypes compared with WT ESCs (Figure 3E), whereas no differences in mesoderm markers were detected (data not shown). The RNA-Seq results were confirmed by qRT-PCR for selected genes (Figure 3F). Altogether, these results provide initial evidence that pluripotency factors and developmental regulators are differentially, and oppositely, regulated after C54Y mutation of Thap1, and that such changes in the ESC transcriptome are even more apparent after loss of full-length THAP1.

**THAP1 Genomic Binding Analysis**

To gain an overview of the global role of THAP1 in ESCs, we performed chromatin immunoprecipitation followed by massively parallel sequencing (ChIP-Seq). We identified 1,731 high-confidence peaks bound by THAP1, corresponding to 1,052 target genes (Table S2), 65 of which were upregulated and 43 were downregulated in *Thap1* 

ESCs (Figure 4A). THAP1 binding was enriched at the transcription start site (TSS), extending over a 2-kb interval (Figure 4B). THAP1 bound only slightly more at promoters (27.04%) than in intergenic regions (24.10%) (Figure 4C). We used public ESC ChIP-Seq datasets for p300, Med1, and monomethylated histone H3 Lys4 (H3K4me1), acetyl histone H3 Lys27 (H3K27ac), and trimethylated histone H3 Lys36 (H3K36me3) (Arnold et al., 2013; Heintzman et al., 2007, 2009; Rada-Iglesias et al., 2011; Visel et al., 2009) to determine that THAP1 occupancy was detected at activated enhancers but not with the H3K27me3 silent mark (Figure 4D). By clustering the ChIP-Seq enrichment profiles of RNA polymerase II Ser-5 phosphorylated (Pol II Ser-5P) and the histone modification trimethylated histone H3 Lys4 (H3K4me3), with high-confidence THAP1 peaks, we found that THAP1 occupancy occurred at promoter regions of activated genes (Figure 4E) associated with GO categories related to metabolic processes, including nucleobase-containing compound and RNA metabolic, and neurologic system process (Figures 4F and S2A). THAP1 also occupied bivalent or poised genes that are associated with developmental processes, system development, cell communication, and nervous system and ectoderm development, among others (Figures 4F and S2B). Occupancy of THAP1 at silent genes was related to diverse functions, including developmental and immune system process (Figures 4F and S2C). We identified the top DNA motifs enriched at the THAP-domain-binding sequence (THABS) and de novo motifs using the top 1,000 THAP1 peaks and the algorithm MEME (Figure 4G, left panel). Motifs #1 and #3 overlapped with the THABS identified by ENCODE in the human K562 cell line (Kheradpour and Kellis, 2014). Motif #3 contained the core GGCA sequence, essential for recognition by the THAP domain (Clouaire et al., 2005). Secondary motifs overrepresented in THABS overlapped with the binding sequence for RE-silencing transcription factor (REST), which represses the expression of neuronal genes in differentiated non-neuronal cells (Ballas et al., 2005), and Yin Yang 1 (YY1), which plays multiple roles in the development of the central and peripheral nervous systems (He and Casaccia-Bonnefil, 2008; Figure 4G, right panel). Consistent with previous reports, inspection of individual gene tracks showed THAP1 binding at the core promoter of *Thap1* and *Tor1A* (Erogullari et al., 2014; Gavolini et al., 2010; Kaiser et al., 2010), thus validating our approach (Figures 5A and 5B, left panel). ChIP-qPCR using anti-THAP1 antibody confirmed THAP1 occupancy at these sites in WT but not in *Thap1* 

ESCs (Figures 5A and 5B, right panel), and at promoters of the cell-cycle genes Aurora A kinase (*Aurka*) and tumor protein p53 (*Trp53*), at the apoptosis genes Bcl-2-associated death promoter (*Bad*) and DNA fragmentation factor subunit beta (*Dffb*), and at the splicing factor 3a subunit 2 (*Sfas2). Despite this occupancy, the expression of these genes is not altered in *Thap1* 

and *Thap1* 

ESCs.
**Figure 4. THAP1 ChIP-Seq Analysis**

(A) Pie chart depicting the number of THAP1 only bound genes, activated and bound, and repressed and bound in *Thap1ΔExon2* (KO) ESCs.

(B) Histogram showing the distribution of THAP1-binding peaks relative to the nearest TSS.

(C) Pie chart of the genomic distribution of THAP1-binding peaks including promoters (within 5 kb upstream of TSS), downstream (within 10 kb downstream of the gene), introns, exons, upstream (within 10 kb upstream of the gene), and intergenic regions.

(D) p300, MED1, H3K4me1 (characteristic of predicted enhancer), H3K27ac (active state), and H3K27me3 (repressed state) occupancy around the summit of THAP1 peaks.

**Legend continued on next page**

Please cite this article in press as: Aguilo et al., THAP1: Role in Mouse Embryonic Stem Cell Survival and Differentiation, *Stem Cell Reports* (2017), http://dx.doi.org/10.1016/j.stemcr.2017.04.032
compared with WT (Table S1), suggesting that THAP1 is present on specific promoters as part of an inactive complex poised to respond to defined developmental stimuli yet to be elucidated. Binding of THAP1 at the non-specific region of chromosome 8 was not detected (Figure S5C).

**Thap1C54Y and Thap1ΔExon2 EBs Show Increased Cell Death and Fail to Properly Differentiate**

To study lineage specification, we assayed EB formation and differentiation (Höpf et al., 2004). The absence of leukemia inhibitory factor (LIF), replacement of fetal calf serum (FCS) with KnockOut serum replacement, and growth in suspension triggered cell differentiation. WT, Thap1C54Y, and Thap1ΔExon2 ESCs formed EB aggregates (Figure S3A). It was immediately apparent that Thap1ΔExon2 EBs were smaller and all died between days 12 and 16. We assayed mRNA levels of markers that either promote or inhibit cell death (Youle and Strasser, 2008). Compared with WT, the expression of Bax was higher in Thap1C54Y and Thap1ΔExon2 EBs, and Puma also increased in the latter, with a peak at day 8 observed in both genotypes (Figure S3B, upper panel). The expression of the anti-apoptotic markers Xiap and Bcl-2 was also higher in Thap1ΔExon2 EBs compared with Thap1C54Y and WT EBs, with a striking 50- to 75-fold upregulation of Bcl-2 in Thap1ΔExon2 EBs (Figure S3B, lower panel). Cell viability in EBs was actually decreased in both Thap1C54Y and Thap1ΔExon2 ESCs compared with WT ESCs as indicated by annexin V (Figure 6A) and TUNEL staining (Figure 6B), and cleaved caspase-3 staining (Figure S4A), more so in the ΔExon2 EBs. Cell death was not diminished in either genotype when EBs were grown in 40 μM Z-VAD-FMK, a pan-caspase inhibitor (Figure S4B). Thap1ΔExon2 EBs demonstrated the highest percentage of cell death whereas an intermediate level was observed in Thap1C54Y EBs, compared with WT, over the 4-day period (Figure S4B). Failure of Z-VAD-FMK to preserve survival is consistent with literature showing that cellular commitment to apoptosis occurs at the level of BCL-2 family-regulated mitochondrial outer membrane permeabilization (i.e., cytochrome c release), and that subsequent death phenotypes occur whether or not caspases activate (Chipuk and Green, 2005).

We also used qRTP-PCR to assay the expression of differentiation markers in EBs, representing the three germ layers, many of which were already dysregulated in the ESCs. Thap1 mRNA levels were constant in WT EBs up to day 16, the last day examined. In Thap1C54Y ESCs, Thap1 mRNA was higher than in WT, and increased steadily up to day 16. Thap1 mRNA was undetectable in Thap1ΔExon2 EBs (Figure 6C, top left panel). The expression of ESC genes Pdm114 and Rex1 increased progressively between day 1 and day 12 in Thap1ΔExon EBs, and the latter also progressively increased in Thap1C54Y EBs (Figure 6C, top right panel), suggesting that failure to silence pluripotency genes during differentiation may contribute to restriction of developmental potential. Thap1C54Y and Thap1ΔExon2 EBs expressed markers of endodermal lineage (Figure S3C, upper panel) following a pattern similar to that of WT EBs. However, the expression of the mesodermal markers T (i.e., Brachyury) and Fik1 was abnormal in the Thap1C54Y and Thap1ΔExon2 EBs (Figure S3C, lower right panel). Consistent with the GO analysis, the expression of the ectodermal markers Sox11, Nestin, and TuJ1 were decreased in Thap1C54Y and Thap1ΔExon2 EBs, whereas the marker of primitive ectoderm Fg5 was normally expressed in Thap1C54Y/C54Y and Thap1ΔExon2 EBs compared with WT EBs (Figure S3C, lower panel). Failure of Thap1C54Y to increase expression of TuJ1 suggests a stall at early ectodermal differentiation.

**THAP1 Is Essential for Neural Differentiation of Mouse ESCs**

To further investigate the role of THAP1 during neural differentiation of mESCs from EBs, we used one protocol based on retinoic acid (RA)-dependent neural differentiation (Figure 7A, protocol #1; Okada et al., 2004), and a second protocol without RA but with the addition of N2 supplemented with epidermal growth factor β and insulin (Figure S5A, protocol #2; Okabe et al., 1996). Survival of mutant cells, particularly ΔExon2, was improved in protocol #2 relative to #1. Using protocol #1, undifferentiated, neural progenitor Nestin-positive cells were detected in Thap1C54Y EBs, but at a lower expression level than in protocol #2, which has a longer culture period. Using protocol #2, Nestin was detected in both WT and Thap1C54Y differentiated cells, but TuJ1-positive neuronal projections, indicative of more mature neurons, were detected almost exclusively in WT ESCs and were abnormally restricted to the internal portion of cell mass in Thap1C54Y ESCs, again consistent with a stall at early ectodermal differentiation (Figure S5B). Thap1ΔExon2 EBs failed to develop Nestin- or TuJ1-immunopositive neuronal projections with either

(E) Pol II Ser-5P and H3K4me3 (characteristic of promoters) occupancy around the summit of THAP1 peaks.

(f) Pie chart depicting THAP1 binding sites (black lines) at active (purple), bivalent (blue), and silent (gray) genes (defined in Experimental Procedures).

(G) The top 1,000 TF ChIP-Seq peaks were used to identify THAP1 de novo motif (left panel showing the five most significant motifs) and overlapping of THAP1 binding with the indicated transcription factors (right panel). Number (N) of sites and the corresponding p values are indicated.
Figure 5. ChIP and qRT-PCR Validation of THAP1 Target Genes

(A) ChIP-Seq binding profiles for THAP1 replicates (Rep) 1 and 2 at the Thap1 locus (left panel). ChIP qRT-PCR analysis of THAP1 at Thap1 TSS in WT, Thap1<sup>C34Y/C34Y</sup> (KI), and Thap1<sup>ΔE2</sup> (KO) ESCs (right panel). Data are normalized to percent input and relative to WT. An ANOVA was performed, which revealed a significant difference among the genotypes when examining THAP1 binding (F(2,8) = 1544, p < 0.0001). Holm-Sidak’s multiple comparisons test was performed post hoc, revealing significant differences between the genotypes. Data are presented as mean ± SEM of four independent experiments. ****p < 0.001.

(B) ChIP-Seq binding profiles for THAP1 replicates (Rep) 1 and 2 at the Tor1A locus (left panel). ChIP qRT-PCR analysis of THAP1 at Tor1A TSS in WT, Tor1A<sup>C34Y/C34Y</sup> (KI), and Tor1A<sup>ΔE2</sup> (KO) ESCs (right panel). Data are normalized to percent input and relative to WT. An ANOVA was performed, which revealed a significant difference among the genotypes when examining THAP1 binding (F(2,0) = 1854, p < 0.0001). The Holm-Sidak multiple comparisons test was performed post hoc, revealing significant differences between the genotypes. Data are presented as mean ± SEM of four independent experiments. ****p < 0.001.

(C) ChIP-qPCR analysis of THAP1 at cell-cycle genes (Aurka and Trp53), apoptosis (Bad and Dffb), and Sf3a2 TSS in WT, Thap1<sup>C34Y/C34Y</sup> (KI), and Thap1<sup>ΔE2</sup> (KO) ESCs. Chromosome 8 was used as a negative control. Data are normalized to percent input and relative to WT. Significant differences were observed among the genotypes when examining the binding of THAP1 at the cell-cycle genes Aurka (F(2,7) = 118.2, p = 0.0001) and Trp53 (F(2,9) = 37.43, p < 0.0001). Significant differences were observed among the genotypes when examining the THAP1 binding at apoptosis related genes Bad (F(2,9) = 74.64, p < 0.0001) and Dffb (F(2,9) = 39.21, p < 0.0001). Significant differences were observed among the genotypes when examining the THAP1 binding at the RNA splicing related gene, Sf3a2 (F(2,9) = 39.21, p < 0.0001). No differences were observed among the genotypes with ANOVA when examining the THAP1 binding at Chrs8. The Holm-Sidak multiple comparisons test was performed post hoc, revealing significant differences between the genotypes. Data are presented as mean ± SEM of four independent experiments. ***p < 0.005, ****p < 0.001.
A

**Differentiation day 1**

- **WT**
- **KI**
- **KO**

% gated cells (relative to WT)

|        | WT | KI | KO |
|--------|----|----|----|
| Annexin V - | 0.1 | 0.2 | 0.3 |
| Annexin V + | 2.5 | 4.8 | 6.7 |

**Differentiation day 6**

- **WT**
- **KI**
- **KO**

% gated cells (relative to WT)

|        | WT | KI | KO |
|--------|----|----|----|
| Annexin V - | 0.4 | 0.7 | 0.9 |
| Annexin V + | 3.2 | 5.4 | 7.6 |

**Differentiation day 10**

- **WT**
- **KI**
- **KO**

% gated cells (relative to WT)

|        | WT | KI | KO |
|--------|----|----|----|
| Annexin V - | 0.7 | 1.4 | 2.1 |
| Annexin V + | 2.9 | 4.1 | 5.3 |

B

**TUNEL/DAPI**

- **WT**
- **KI**
- **KO**

Differentiation day 1

Differentiation day 6

Differentiation day 10

C

**Pluripotent**

- **WT**
- **KI**
- **KO**

- **Thap1**

mRNA relative to WT at D0

|        | D0 | 1  | 4  | 8  | 12 | 16 |
|--------|----|----|----|----|----|----|
| mRNA   | 1  | 2.5| 5  | 7.5| 10 | 15 |

- **Prdm14**

mRNA relative to WT at D0

|        | D0 | 1  | 4  | 8  | 12 | 16 |
|--------|----|----|----|----|----|----|
| mRNA   | 1  | 2  | 4  | 7  | 10 | 15 |

- **Rex1**

mRNA relative to WT at D0

|        | D0 | 1  | 4  | 8  | 12 | 16 |
|--------|----|----|----|----|----|----|
| mRNA   | 1  | 2.5| 5  | 7.5| 10 | 15 |

**Ectoderm**

- **Sox11**

mRNA relative to WT at D0

|        | D0 | 1  | 4  | 8  | 12 | 16 |
|--------|----|----|----|----|----|----|
| mRNA   | 1  | 2.5| 5  | 7.5| 10 | 15 |

- **Nestin**

mRNA relative to WT at D0

|        | D0 | 1  | 4  | 8  | 12 | 16 |
|--------|----|----|----|----|----|----|
| mRNA   | 1  | 2  | 4  | 7  | 10 | 15 |

- **Tuj1**

mRNA relative to WT at D0

|        | D0 | 1  | 4  | 8  | 12 | 16 |
|--------|----|----|----|----|----|----|
| mRNA   | 1  | 2.5| 5  | 7.5| 10 | 15 |

(legend on next page)
protocol (Figures 7B and S5B). The quantification of neuronal projections in the three genotypes is illustrated in the bar graphs (Figures 7C and S5C). Expression of the ESC genes Prdm14, Nanog, and Rex1 in samples from protocol #1 (Figure 6) was abnormally enhanced in Thap1Exon2 and Thap1Exon2 EBs, even at the later stages of neuronal differentiation (Figure 7D, upper panel). We also monitored progression of ESC differentiation toward neural fates by analysis of Pax6, Sox11, Nestin, and Tuj1 expression with qRT-PCR. Thap1Exon2 and Thap1Exon2 failed to activate the expression of these ectodermal markers, suggesting a role for THAP1 in multiple stages of neuronal differentiation (Figure 7D, lower panel). Similar results were obtained by expression analysis in protocol #2 (Figure S5D). Collectively, our results show that THAP1 is required to silence the expression of the pluripotency factors and to activate the expression of the neural factors (Figure 7E).

DISCUSSION

We have identified THAP1 as a major regulator of ESC survival and, to a lesser degree, proliferation and differentiation. As THAP1 mutations result in a neurologic disease, we were particularly interested in the effects of a causative mutation, C54Y, on neuronal differentiation. We had previously posited, based on observations of homozygote mutant embryos, that THAP1 is required in multiple stages and layers during embryogenesis, organogenesis, and maturation of the nervous system, including neurogenesis and neuritogenesis, and that the C54Y variant is frequently unable to substitute for the WT (Ruiz et al., 2015).

In all assays, the homozygous Exon2 phenotype was more severe than that of the homozygous C54Y phenotype, confirming that C54Y does not confer a complete loss of function. Both genotypes showed a decreased rate of proliferation, more so in the Exon2, consistent with regulation of the cell cycle by THAP1 in HUVECs and lymphoblasts (Cayrol et al., 2007; Vemula et al., 2014). Only the Exon2 homozygote showed a decrease in viability in the ESC stage. Thus, a loss of WT THAP1 promotes cell death even though THAP1 is reportedly a pro-apoptotic factor (Roussigné et al., 2003). The lack of a requirement for THAP1 for pluripotency is in marked contrast to RONIN/THAP11, another member of the THAP family to have been characterized in this manner (Dejosez et al., 2008).

We compared this study of the global transcriptome in ESCs harboring a mutation of THAP1 with previous microarray assays of HUVECs with up- or downregulation of THAP1 and of human lymphoblasts harboring a disease-causing intrinsic variant of THAP1, both of which revealed that the main pathway regulated by THAP1 is the cell cycle. The cell-cycle abnormalities and cell death described herein bear a strong resemblance to these HUVECs and lymphoblasts. Similar to the Thap1Exon2 ESCs, the mutant human lymphoblastoid cells had decreased viability and showed a reduced number of cells in the G2 phase (Vemula et al., 2014). In HUVECs, knockdown of THAP1 resulted in an impaired G1/S phase (Cayrol et al., 2007). Despite the presence of cell-cycle abnormalities in the THAP1 null cells, cell-cycle genes were not a highly ranked GO category, but genes pertaining to meiosis and X chromosome inactivation categories were also differentially expressed in Thap1Exon2 and Thap1Exon2 compared with WT ESCs.

We examined the role of THAP1 in ESC differentiation by performing EB formation assays. EB formation and growth were highly compromised in Thap1Exon2 cells. Indeed, we...
A) Protocol #1

B) day 10; Protocol #1

C) KO

D) Pluripotent

E) Ectoderm

(legend on next page)
showed that cell viability was compromised in both \(Thap1^{C54Y}\) and \(Thap1^{ΔExon2}\) EBs, with a greater cell death phenotype after loss of full-length THAP1. The extent of cell death in the \(ΔExon2\) EBs precludes a conclusion regarding an effect on differentiation. Our investigations into the mechanism of cell death implicate the mitochondrial pathway of apoptosis as responsible for cell death commitment and caspase activation. Cell death execution, however, apparently can proceed independently of caspase activation. Delineation of a definitive sequence of events leading to cell death requires further characterization. Although there is increased cell death in the \(Thap1^{C54Y}\) EBs, their survival is sufficient to enable some conclusions regarding differentiation. Specifically, mesoderm specification may be delayed, but, more strikingly for a mutation that leads to a neurologic disease, neuronal differentiation appears to stall at an early stage.

The mechanism(s) via which the \(Thap1\) mutations prevent repression of pluripotency markers is not entirely obvious, but there are leads, particularly in the categories of dysregulated genes. An in silico search for THABS motifs in mouse \(Nanog, Prdm14, Pou5F1, Esrb,\) and \(Klf4\) was negative, and ChIP-Seq did not uncover binding sites in these genes. Regulation of the pluripotency network requires cooperation between multiple transcription factors, including the pluripotency factors themselves, signal transduction pathways, and epigenetic modifications. Notably, of the DEGs detected by RNA-Seq, many genes in these categories are altered, e.g., \(Bmp4, Dnmt3a\) and \(Dnmt3b, Satb1, Klf10, Sirt6,\) and \(Sox17,\) with a greater degree of dysregulation in the \(Thap1^{ΔExon2}\) relative to the \(Thap1^{C54Y}\) ESCs. Interestingly, \(Dnmt3a\) and \(3b, Satb1,\) and \(Sox17\) KO ESCs display phenotypes that are similar in many ways to those of the \(Thap1\) mESCs. Specifically, upon exposure to differentiation factors, they simultaneously fail to downregulate pluripotency genes and upregulate lineage-specific genes, some of which are already decreased prior to differentiation (Bergsland et al., 2011; Chen et al., 2003; Savarese et al., 2009). Thus, the mechanism via which \(Thap1\) regulates these genes requires further examination but may represent a pathway for inverse regulation of pluripotency and differentiation genes.

The low percentage (10%) of overlap between RNA-Seq and ChIP-Seq datasets suggests an alternative role of \(Thap1\) in regulating gene expression other than direct binding at DNA, as exists with other zinc-finger factors (Aguilo et al., 2015). This conclusion may also be supported by the relatively low fold changes observed in the \(Thap1^{C54Y}\) ESCs, as much greater changes are seen in \(Thap1^{ΔExon2}\) ESCs, where the full-length transcription factor is eliminated. It is possible that the mutant \(Thap1\) proteins are able to bind to some, but not all, targets of WT \(Thap1,\) and in addition, may bind to sequences that are normally not bound by WT \(Thap1\) (Campagne et al., 2012). Also, these mutant proteins may participate in \(Thap1\) protein interactions. For example, \(Thap1\) interacts with HCF-1 via a motif that does not include amino acid 54 (Mazars et al., 2010). Therefore, the C54Y form may still be able to interact with HCF-1, Par-4, and partners as yet to be identified, thereby recruiting such transcription factors to the target DNA. This hypothesis would be consistent with the fact that there are individuals with dystonia with homozygote \(Thap1\) mutations. In the case where \(Thap1\) binding is decreased or eliminated, there may be increased binding of transcription factors with an affinity for similar motifs. As noted, these include YY1 and REST, thereby altering levels of their direct and indirect targets, which include many of the RNA-Seq DEGs. In fact, there is a very strong super-shift with anti-YY1 in electrophoretic mobility shift assays using the \(Tor1a\) THABS and embryonic brain nuclear extract (data not shown).

In summary, we have identified \(Thap1\) as an essential regulator of mESC potential, including viability and likely neuronal differentiation, the latter particularly for the naturally occurring C54Y mutation. These activities are likely due in part to direct regulation of gene expression but to a large extent to indirect regulation, the mechanisms of which are important to identify. Notably, dysregulated pathways in ESCs with mutations of \(Thap1\) overlap with those seen in mouse models of DYT6, and the ESC system

---

**Figure 7. Characterization of Neural Differentiation**

(A) Schematic representation of cell-culture strategy of neural differentiation using protocol #1.
(B) Immunofluorescence analysis of Nestin (green; middle panel) and TuJ1 (red; right panel) of WT, \(Thap1^{C54Y/C54Y}\) (KI), and \(Thap1^{ΔExon2}\) (KO) differentiated cells at day 10 of differentiation. Bright field (left panel) and DAPI stain (blue) are shown. Scale bars, 100 μm.
(C) Bar graph depicting percentage of colonies with neuronal projection.
(D) qRT-PCR analysis of \(Thap1,\) pluripotent, and ectoderm marker genes during neural differentiation (protocol #1) of \(Thap1^{C54Y/C54Y}\) (KI) and \(Thap1^{ΔExon2}\) (KO) ESCs at the indicated time points. Data are presented relative to WT at day 0. The two-way ANOVA results for \(Thap1,\) pluripotent, and ectoderm marker genes during RA-induced differentiation can be found in Table S3. The Holm-Sidak multiple comparisons test was performed post hoc for each gene of interest by day, revealing significant differences between the genotypes. Data are presented as mean ± SEM of at least three independent experiments. *\(p < 0.05\), **\(p < 0.005\), ****\(p < 0.001\).
(E) Schema illustrating that \(Thap1\) is required to silence the expression of the pluripotency factors and to activate the expression of the neuronal factors, as well as mediate ESC survival.
may potentially be exploited to discover how following differentiation, mutations of the ubiquitously expressed THAP1 protein can lead to a strictly neurologic disease.

**EXPERIMENTAL PROCEDURES**

Preparation of nuclear extracts and qRT-PCR gene expression analysis were performed as previously described (Ruiz et al., 2015).

Further experimental details and lists of antibodies, public datasets, and primer sets utilized in this study may be found in Supplemental Experimental Procedures and other Supplemental Information.

**Cell Culture and EB Formation**

Wild-type (WT), Thap1<sup>C54Y</sup>, and Thap1<sup>DExon2</sup> ESCs were maintained on 0.1% gelatin-coated tissue culture plates on irradiated mouse embryonic fibroblasts (MEFs) in DMEM containing 15% FCS, 0.1 mM 2-mercaptoethanol, L-glutamine, non-essential amino acids, and LIF (1,000 units/mL) and penicillin-streptomycin (Thermo Fisher). Cells were passaged once onto gelatin-coated dishes and then aggregated to form EBs in 10-cm Petri dishes at 5 × 10⁶/plate in the absence of LIF. Cells were harvested for extraction of total RNA at the indicated time points. All cell cultures were maintained at 37°C with 5% CO₂.

**Differentiation Assays**

Induction of neural differentiation was performed based on generation of EBs, using two different approaches (Okabe et al., 1996; Okada et al., 2004), with and without addition of RA as indicated in Supplemental Experimental Procedures.

**Cellular Proliferation, Apoptosis Assay, and Cell-Cycle Analyses**

Cellular proliferation, apoptosis assay, and cell-cycle analysis were performed using a Muse Cell Analyzer (Millipore) following the manufacturer’s instructions.

**Alkaline Phosphatase Activity**

AP activity was measured using the Stemgent Alkaline Phosphatase Staining kit (Stemgent) following the manufacturer’s recommendations.

**Chromatin Immunoprecipitation with High-Throughput Sequencing**

THAP1 ChiP-Seq was performed as described in Supplemental Experimental Procedures.

**RNA-Seq and Analysis**

RNA-Seq library preparation was performed at the Weill-Cornell Medical College Genomic Core Facility using the TrueSeq RNA sample preparation kit (Illumina RS-122-2001) and sequenced by the Illumina HiSeq 2,500 platform as 100-bp pair-ended reads.

**Gene Ontology Analysis**

For functional profiling of THAP1 binding regions identified by ChiP-Seq, GO analysis was performed using Panther software (http://pantherdb.org) (Mi et al., 2016). GO analysis from RNA-Seq datasets was performed using the web tool The Database for Annotation, Visualization and Integrated Discovery (DAVID) (http://david.abcc.ncifcrf.gov/) (Huang et al., 2009a, 2009b).

**Statistical Analysis**

All values are expressed as mean ± SD. Statistical analysis was performed by two-way ANOVA followed by Bonferroni’s multiple testing corrections using GraphPad Prism 5 Software. A probability value of p < 0.05 was considered statistically significant.

**ACCESSION NUMBERS**

All next-generation sequencing data are deposited in NCBI GEO database under accession numbers GEO: GSE86947 and GSE86911.

**SUPPLEMENTAL INFORMATION**

Supplemental Information includes Supplemental Experimental Procedures, five figures, and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.stemcr.2017.04.032.

**AUTHOR CONTRIBUTIONS**

Conceptualization, F.A., Z.Z., M.J.W., L.J.O., P.G.-A., T.P.Z., and M.E.E.; Methodology, K.K.; Formal analysis, F.A., Z.Z., R.S., C.W, and W.Z.; Investigation, F.A., Z.Z., K.N., R.W., M.H., and Y.S.; Writing – Original Draft, F.A. and M.E.E.; Writing – Review & Editing, Z.Z., P.G.-A., L.J.O., M.J.W., T.P.Z., and M.E.E.; Funding Acquisition, M.E.E.; Supervision and Project Administration, F.A. and M.E.E.

**ACKNOWLEDGMENTS**

We thank Dr. A. Alonso of the Weill-Cornell College of Medicine’s Epigenomic Sequencing Core for expert advice for sequencing and library preparation. This work was funded by NIH R01NS081282 (to M.E.E. and L.J.O.) New York State Stem Cell Science (NYSTEM C029553 to M.E.E.), and the Huffington Foundation (T.P.Z.).

Received: October 11, 2016
Revised: April 26, 2017
Accepted: April 27, 2017
Published: June 1, 2017

**REFERENCES**

Aguilo, F., Zhang, F., Sancho, A., Fidalgo, M., Di Cecilia, S., Vashisht, A., Lee, D.-F., Chen, C.-H., Rengasamy, M., Andino, B., et al. (2015). Coordination of m(6)A mRNA methylation and gene transcription by ZFP217 regulates pluripotency and reprogramming. Cell Stem Cell 17, 689–704.

Arnold, C.D., Gerlach, D., Stelzer, C., Boryn, Ł.M., Rath, M., and Stark, A. (2013). Genome-wide quantitative enhancer activity maps identified by STARR-seq. Science 339, 1074–1077.
Ballas, N., Grunseich, C., Lu, D.D., Speh, J.C., and Mandel, G. (2005). REST and its corepressors mediate plasticity of neuronal gene chromatin throughout neurogenesis. Cell 121, 645–657.

Bergslund, M., Ramsköld, D., Zauter, C., Klum, S., Sandberg, R., and Muhr, J. (2011). Sequentially acting Sox transcription factors in neural lineage development. Genes Dev. 25, 2453–2464.

Campagne, S., Muller, I., Milon, A., and Gervais, V. (2012). Towards the classification of DYT6 dystonia mutants in the DNA-binding domain of THAP1. Nucleic Acids Res. 40, 9927–9940.

Cayrol, C., Lacroix, C., Mathe, C., Ecchard, V., Ceribelli, M., Lor- eau, E., Lazat, V., Dessen, P., Mantovani, R., Aguilar, L., et al. (2007). The THAP-zinc finger protein THAP1 regulates endothelial cell proliferation through modulation of pRB/E2F cell-cycle target genes. Blood 109, 584–594.

Chen, Y., He, Z.X., Liu, A., Wang, K., Mao, W.W., Chu, J.X., Lu, Y., Fang, Z.F., Shi, Y.T., Yang, Q.Z., et al. (2003). Embryonic stem cells generated by nuclear transfer of human somatic nuclei into rabbit oocytes. Cell Res. 13, 251–263.

Chipuk, J.E., and Green, D.R. (2005). Do inducers of apoptosis trigger caspase-independent cell death? Nat. Rev. Mol. Cell Biol. 6, 268–275.

Clouaire, T., Roussigne, M., Ecchard, V., Mathe, C., Amalric, F., and Girard, J.-P. (2005). The THAP domain of THAP1 is a large C2CH module with zinc-dependent sequence-specific DNA-binding activity. Proc. Natl. Acad. Sci. USA 102, 6907–6912.

Dejosez, M., Krumenacker, J.S., Zitur, L.J., Passeri, M., Chu, L.-F., Songyang, Z., Thomson, J.A., and Zwaka, T.P. (2008). Ronin is essential for embryogenesis and the pluripotency of mouse embryonic stem cells. Cell 133, 1162–1174.

Erogullari, A., Hollstein, R., Seibler, P., Braunnholz, D., Koschmider, E., Depping, R., Eckhold, J., Lohnau, T., Gillessen-Kaesbach, G., Grünwald, A., et al. (2014). THAP1, the gene mutated in DYT6 dystonia, autoregulates its own expression. Biochim. Biophys. Acta 1839, 1196–1204.

Fuchs, T., Gavarini, S., Saunders-Pullman, R., Raymond, D., Ehrlich, M.E., Bressman, S.B., and Ozelius, L.J. (2009). Mutations in the THAP1 gene are responsible for DYT6 primary torsion dystonia. Nat. Genet. 41, 286–288.

Gavarini, S., Cayrol, C., Fuchs, T., Lyons, N., Ehrlich, M.E., Girard, J.-P., and Ozelius, L.J. (2010). Direct interaction between causative genes of DYT1 and DYT6 primary dystonia. Ann. Neurol. 68, 549–553.

Gervais, V., Campagne, S., Durand, J., Muller, I., and Milon, A. (2013). NMR studies of a new family of DNA binding proteins: the THAP proteins. J. Biomol. NMR 56, 3–15.

He, Y., and Casaccia-Bonnefil, P. (2008). The Yin and Yang of YY1 in Drosophila and other eukaryotic organisms. Microbiol. Spectr. 3, MDNA3-0004-2014.

Mazars, R., Gonzalez-de-Peredo, A., Cayrol, C., Lavigne, A.-C., Votruba, J.L., Ortega, N., Lacroix, C., Gautier, V., Huet, G., Ray, A., et al. (2010). The THAP-zinc finger protein THAP1 associates with coactivator HCF-1 and O-GlcNAc transferase: a link between DYT6 and DYT1 dystonias. J. Biol. Chem. 285, 13364–13371.

Mi, H., Poudel, S., Muruganujan, A., Casagrande, J.T., and Thomas, P.D. (2016). PANTHER version 10: expanded protein families and functions, and analysis tools. Nucleic Acids Res. 44, D336–D342.

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. Dev. Biol. 179, 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.

Ortiz-Virumbrales, M., Ruiz, M., Hone, E., Dolios, G., Wang, R., Morant, A., Kottwitz, J., Ozelius, L.J., Gandy, S., and Ehrlich, M.E. (2014). Dystonia type 6 gene product THAP1: identification of a 50 kDa DNA-binding species in neuronal nuclear fractions. Acta Neuropathol. Commun. 2, 139.

Palmqvist, L., Glover, C.H., Hsu, L., Lu, M., Bossen, B., Piret, J.M., Humphries, R.K., and Helgason, C.D. (2005). Correlation of murine embryonic stem cell gene expression profiles with functional measures of pluripotency. Stem Cells 23, 663–680.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature 470, 279–283.
Roussigne, M., Cayrol, C., Clouaire, T., Almalric, F., and Girard, J.-P. (2003). THAP1 is a nuclear proapoptotic factor that links prostate-apoptosis-response-4 (Par-4) to PML nuclear bodies. Oncogene 22, 2432–2442.

Ruiz, M., Perez-Garcia, G., Ortiz-Virumbrales, M., Méneret, A., Morant, A., Kottwitz, J., Fuchs, T., Bonet, J., Gonzalez-Alegre, P., Hof, P.R., et al. (2015). Abnormalities of motor function, transcription and cerebellar structure in mouse models of THAP1 dystonia. Hum. Mol. Genet. 24, 7159–7170.

Savarese, F., Dávila, A., Nechanitzky, R., De La Rosa-Velazquez, I., Pereira, C.F., Engelke, R., Takahashi, K., Jenuwein, T., Kohwi-Shigematsu, T., Fisher, A.G., et al. (2009). Satb1 and Satb2 regulate embryonic stem cell differentiation and Nanog expression. Genes Dev. 23, 2625–2638.

Schneider, S.A., Ramirez, A., Shafiee, K., Kaiser, F.J., Erogullari, A., Brüggemann, N., Winkler, S., Bahman, I., Osmanovic, A., Shafa, M.A., et al. (2011). Homozygous THAP1 mutations as cause of early-onset generalized dystonia. Mov. Disord. 26, 858–861.

Sengel, C., Gavarini, S., Sharma, N., Ozelius, L.J., and Bragg, D.C. (2011). Dimerization of the DYT6 dystonia protein, THAP1, requires residues within the coiled-coil domain. J. Neurochem. 118, 1087–1100.

Vemula, S.R., Xiao, J., Zhao, Y., Bastian, R.W., Perlmutter, J.S., Racette, B.A., Paniello, R.C., Wszolek, Z.K., Uitti, R.J., Van Gerpen, J.A., et al. (2014). A rare sequence variant in intron 1 of THAP1 is associated with primary dystonia. Mol. Genet. Genomic Med. 2, 261–272.

Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., et al. (2009). ChiP-seq accurately predicts tissue-specific activity of enhancers. Nature 457, 854–858.

Xiromerisiou, G., Houlden, H., Scarmeas, N., Stamou, M., Kara, E., Hardy, J., Lees, A.J., Korlipara, P., Limousin, P., Paudel, R., et al. (2012). THAP1 mutations and dystonia phenotypes: genotype phenotype correlations. Mov. Disord. 27, 1290–1294.

Youle, R.J., and Strasser, A. (2008). The BCL-2 protein family: opposing activities that mediate cell death. Nat. Rev. Mol. Cell Biol. 9, 47–59.
Supplemental Information

THAP1: Role in Mouse Embryonic Stem Cell Survival and Differentiation

Francesca Aguilo, Zuchra Zakirova, Katie Nolan, Ryan Wagner, Rajal Sharma, Megan Hogan, Chenguoo 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

Upregulated KI ESCs
- Organ morphogenesis
- Axon guidance
- Neural crest cell migration
- Dorsal/ventral pattern formation
- Cell differentiation

Downregulated KI ESCs
- Response to oxidative stress
- Positive regulation of cell motility
- Regulation of gene expression
- Cellular iron ion homeostasis
- Skeletal system morphogenesis

Upregulated KO ESCs
- Cell differentiation
- Carbohydrate metabolic process
- Positive regulation of B cell differentiation
- Spermatogenesis
- Negative thymic T cell selection

Downregulated KO ESCs
- Multicellular organism development
- Cell adhesion
- Cell differentiation
- Negative regulation of canonical Wnt signaling pathway
- Positive regulation of cell migration
- Axon guidance
- Brain development
- Nervous system development
FIGURE S3, related to FIGURE 6

A

WT  WT  KI  KI  KO  KO

B

Apoptotic

$\text{Bax}$

Anti-apoptotic

$\text{Bcl2}$

$\text{Puma}$

$\text{Xiap}$

C

$\text{Sox17}$

$\text{Gata4}$

$\text{Gata6}$

$\text{Fgf5}$

$\text{T}$

$\text{Fik1}$

$\text{WT}$  $\text{KI}$  $\text{KO}$
A

**Cleaved Caspase 3/DAPI**

Differentiation day 1

Differentiation day 6

Differentiation day 10

B

**Day 1**

**% cell viability**

Control  Z-VAD-FMK

WT  KI  KO

**Day 2**

**% cell viability**

Control  Z-VAD-FMK

WT  KI  KO

**Day 3**

**% cell viability**

Control  Z-VAD-FMK

WT  KI  KO

**Day 4**

**% cell viability**

Control  Z-VAD-FMK

WT  KI  KO
FIGURE S5, related to FIGURE 7

A) Protocol #2

B) day 16; Protocol #2

C) % of colonies with neuronal projections

D) mRNA relative to WT at D0

**Thap1**

**Nanog**

**Prdm14**

**Nestin**

**TuJ1**

**Pax6**
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 Thap1ΔExon2 (KO) and Thap1ΔExon2 (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, Thap1ΔExon2 (KI) and Thap1ΔExon2 (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, Thap1ΔExon2 (KI) and Thap1ΔExon2 (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 (Six1, 7, Gata4 and Gata6), mesodermal (T and Flik1) and primitive ectodermal markers (Fgf5) during EB differentiation of WT, Thap1ΔExon2 (KI) and Thap1ΔExon2 (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), Thap1ΔExon2 (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-VA-D-FMK does not inhibit cell death in Thap1ΔExon2 (KI) and Thap1ΔExon2 (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, Thap1ΔExon2 (KO) and Thap1ΔExon2 (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, Thap1ΔExon2 (KI) and Thap1ΔExon2 (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 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 embryob 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-CI). After disruption of the outgrowth, the cells were released 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(C54Y) ESCs, cre-mediated excision of exon 2 was performed by transfection of 1.5 x 10^6 Thap1(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 nN3FL 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 × 10⁶ 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 MgCl₂, 10 mM KCl, 1 mM DTT) in the presence of protease inhibitors (3 μg/ml aprotinin, 750 μg/ml benzamidene, 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 MgCl₂, 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 °C 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).
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% IGE- PAL 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/CGCGATATCC/BiotiT/TATCTGACT; R: 5Phos/GTCAGATAAGATATCGGT) 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 HiSeq 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=ACCGGATATCTTATCTGACT –a reverse=AGTCAGATAAGATATCGGT -- 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 (PETs) 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/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  | GSM76909             | (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.

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.