Control of Embryonic Stem Cell Lineage Commitment by Core Promoter Factor, TAF3

Zhe Liu[1,4], **Devin R. Scannell**[1,4], **Michael B. Eisen**[1], and **Robert Tjian**[1,2,3]

[1]Howard Hughes Medical Institute, Molecular and Cell Biology Department, University of California, Berkeley, Berkeley, CA 94720, USA

[2]LKS Bio-medical and Health Sciences Center, CIRM Center of Excellence, University of California, Berkeley, Berkeley, California 94720, USA

Summary

Deciphering the molecular basis of pluripotency is fundamental to our understanding of development and embryonic stem cell function. Here we report that TAF3, a TBP-associated core promoter factor, is highly enriched in ES cells. TAF3 is required for endoderm lineage differentiation and prevents premature specification of neuroectoderm and mesoderm. In addition to its role in the core promoter recognition complex TFIID, genome-wide binding studies reveal that TAF3 localizes to chromosomal regions bound by CTCF and cohesin. Enrichment for TAF3/ CTCF/cohesin bound regions distinguishes TAF3-activated from TAF3-repressed genes. Notably, CTCF directly interacts with and recruits TAF3 to promoter distal sites and TAF3-dependent DNA looping is observed between the promoter distal sites and core promoters occupied by TAF3/CTCF/cohesin. Thus, our findings support a new role of TAF3 in mediating long-range chromatin regulatory interactions to safeguard the finely-balanced transcriptional programs that give rise to pluripotency.

Introduction

A hallmark of embryonic stem (ES) cells is their ability to generate all somatic cell types that make up an animal (Bradley et al., 1984). This differentiation potential of ES cells, or pluripotency, is thought to hold great promise for the future of regenerative medicine (Daley and Scadden, 2008). However, to fully develop the emerging field of stem cell-based therapies, a deeper understanding of the molecular basis underlying ES cell pluripotency and the mechanisms controlling cellular differentiation is required. The regulatory pathways that govern ES cell self-renewal and pluripotency include a subset of sequence specific DNA binding transcription factors (Oct4, Nanog, Sox2, Klf4, etc) (Jaenisch and Young, 2008) consistent with the importance of enhancer- and promoter- binding transcription factors in regulating lineage specification during early embryogenesis (Arnold and Robertson, 2009; Tam and Loebel, 2007).

In eukaryotic cells, a key feature of transcriptional regulation is the complex and still poorly understood interplay between gene specific transcription factors and components of the

[3]Corresponding author, jmlim@berkeley.edu.
[4]Contributed to this work equally

multi-subunit core promoter recognition machinery (Naar et al., 2001). Until recently, it was believed that proper gene and cell-type specific transcriptional read-outs were exclusively controlled by combinatorial arrays of classic sequence-specific enhancer binding activators and repressors (Farnham, 2009; Tjian and Maniatis, 1994). By contrast, the so called general or ubiquitous transcription machinery responsible for core promoter recognition was thought to serve mainly as a passive integrator or processor of upstream regulatory signals. However, an increasing number of cell type- and tissue-specific components of the core promoter recognition apparatus have been identified in metazoan organisms and shown to play a role in directing and regulating programs of transcription during the development of specific cell types (Goodrich and Tjian, 2010).

In this report, we focus on one such component of the core promoter recognition complex-the TATA binding protein associated factor 3, TAF3, that was originally identified as a subunit of the TFIID complex in HeLa cells (Gangloff et al., 2001). It was later found that, while other TFIID subunits are destroyed during myogenesis, TAF3 is selectively retained in myotubes in a specialized complex with TBP-related factor 3, TRF3 (Deato and Tjian, 2007). A similar TRF3/TAF3 complex functions during Zebrafish hematopoiesis (Hart et al., 2009). A recent study implicates sub-nuclear localization of TAF3 as another potential mechanism to regulate transcription during myogenesis (Yao et al., 2011). Intriguingly, TAF3 recognizes trimethylated histone H3 lysine 4 (H3K4me3) (Vermeulen et al., 2007), which is associated not only with actively transcribed genes but also with silent developmental genes that are poised for activation upon ES cell differentiation (Bernstein et al., 2006; Mikkelsen et al., 2007). Thus, these studies establish that TAF3, either as a subunit of TFIID or in association with other potential partners (*e.g.* TRF3) may regulate transcription by targeting cell-type specific complexes to core promoters including those that are marked by H3K4me3. Here we report a novel mode of TAF3 action: TAF3 binds the architectural protein CTCF via its vertebrate-specific domain to mediate regulatory interactions between distal CTCF/cohesin bound regions and proximal promoters. Remarkably, we show that this TAF3 activity is critical for early lineage segregation during stem cell differentiation. Thus, our findings unmask new mechanisms that directly link dynamic organization of chromatin structure and transcriptional control of stem cell plasticity.

Results

High Levels of TAF3 in ES Cells

To explore the possibility that TAF3 and/or TRF3/TAF3 complexes may be utilized in different developmental pathways, we analyzed TAF3 protein levels across different tissue types and cell lines by western blot. Unexpectedly, we found the highest TAF3 protein levels (~10× relative to C2C12's) in mouse ES cells (Figure 1A). Even more interestingly, when we induced ES cells to form embryoid bodies (EBs), TAF3 protein levels became selectively reduced while the prototypic subunits (TAF4a and TBP) of TFIID remained mostly unchanged (Figure 1B). Indeed, the levels of TAF3 protein remaining upon EB formation resemble the lower levels observed in differentiated cell-types. Thus, the situation in ES cells is quite distinct from what we previously observed in myoblasts wherein induction to form myotubes leads to a dramatic decrease in canonical TFIID subunits (TAF4a and TBP) while TAF3 levels remained largely unchanged (Deato and Tjian, 2007). We next used qRT-PCR to investigate whether high levels of TAF3 in ES cells are accompanied by a concomitant enrichment of TRF3. Surprisingly, *Trf3* mRNA was not detectable in either ES cells or EBs (Figure S1A). Consistent with this, we found that in ES cell nuclear extract, TAF3 protein migrated at a molecular weight >1MD in a superose 6 column (Figure S1B), while the TAF3/TRF3 complex from myoblast/myotubes migrates as a native species of ~180kd (Deato and Tjian, 2007). These findings indicate that although

TAF3 is highly expressed in ES cells, one of its potential partners (TRF3) is absent, suggesting another distinct functional mode for TAF3.

TAF3 *vs* ES cell Self-renewal and Proliferation

To assess the significance of high levels of TAF3 in ES cells, we generated two independent lentivirus mediated shRNAs that specifically and stably knocked down TAF3 in ES cells (Figure 2A). By 72 hours post-infection greater than 90% reduction of TAF3 protein was observed when compared to control ES cells treated with non-target shRNA. We found that TAF3 knockdown (K/D) in ES cells did not affect the expression of the known ES cell self-renewal genes, Oct4 and Nanog (Figure 2A, 2B, 2F and 2G). TAF3 K/D cells were also able to form alkaline phosphatase positive colonies as efficiently as control ES cells (Figure 2C) and the percentage of SSEA1[+]Oct4[+] cells upon efficient TAF3 K/D remained largely unaltered (Figure 2D and S2A). Furthermore, there was little or no correlation between TAF3[+] and Nanog[+] cells as determined by immuno-fluorescence of individual ES cells (Figure 2G and 2H). These findings suggest that high levels of TAF3 are not required for the proper expression of canonical ES cell self-renewal genes or markers (Oct4, Nanog and SSEA1).

Another important hallmark of ES cells and their self-renewal properties is their elevated proliferative rates compared to somatic cells, which results in a higher proportion of S phase ES cells in the population (White and Dalton, 2005). To gain a semi-quantitative assessment of whether TAF3 K/D impedes ES cell proliferation, we pulse-labeled S phase control and TAF3 K/D cells with 5-ethynyl-2′-deoxyuridine (EdU) and analyzed changes in the cell cycle by flow cytometry. We observed only modest reductions (5% - 8%) in the S phase population of TAF3 K/D cells. Indeed, greater than 70% of TAF3 K/D cells can be found in S phase and no significant G1 or G2/M cell cycle arrest was detected (Figure 2E and S2B). These results indicate that ES cell proliferation is likely not critically dependent on high levels of TAF3. It is also worth noting that the protein levels of canonical TFIID subunits (TAF1, TAF4 and TBP) remained stable after TAF3 depletion (Figure 2A), suggesting that the integrity of TFIID also does not rely on high levels of TAF3 in ES cells. This finding is consistent with the increasingly accepted view that the prototypic holo-TFIID is most critical for Pol II mediated transcription of genes encoding products implicated in DNA replication and cell division. Indeed, ablation of essential TFIID subunits invariably induces cell cycle arrest and often cell death (Shen et al., 2003). The striking absence of any notable self-renewal or proliferation phenotypes in TAF3 K/D cells suggests that, unlike the canonical TFIID TAF(s), TAF3 may provide a more gene and cell-type specific function, perhaps contributing to the proper transcription of a subset of genes involved in pluripotency of ES cells.

TAF3 is a Regulator of Pluripotency

To test whether the high levels of TAF3 in ES cells are required for pluripotency, we induced stable pools of TAF3 K/D and control ES cells to form EBs. Control ES cells formed EBs with the expected heterogeneous cell lineages (Figure 3D). By contrast, TAF3 K/D EBs appeared abnormal and lacked well defined structures (*e.g.* multiple cell layers, cavitations in the inner cells), suggesting that one or more differentiation programs may have been compromised. We next used qRT-PCR to survey the expression levels of lineage-specific markers in both control and TAF3 K/D EBs (Figure 3A-C). As expected, the expression of lineage-specific markers was generally up-regulated in control EBs. However, in TAF3 K/D EBs the expression of primitive endoderm markers (*Gata6* and *Gata4*) was largely abolished, while mesoderm and ectoderm markers (*T, Pax3, nestin* and *Fgf5*) were induced at earlier time points and expressed at higher levels.

To better understand the TAF3 K/D EB phenotype, we stained control and TAF3 K/D EBs with lineage-specific antibodies. As expected, GATA4 stained the outer layer and some internal cells in control EBs, while no significant GATA4 signal was detected in TAF3 K/D EBs (Figure 3E). Likewise, Afp, a late endoderm marker, stained control EBs but not TAF3 K/D EBs (Figure 3F and S3C). These results are consistent with our qRT-PCR results that indicated impaired endoderm development in TAF3 K/D EBs. To address ectoderm differentiation, we used an ES cell line (46C) with GFP knocked into the *Sox1* locus (Ying et al., 2003) to analyze control and TAF3 K/D samples (Figure 3G). At EB day 8 more than half of the TAF3 K/D EBs (21/30 for shRNA A and 16/30 for shRNA B) developed strong internal GFP signals. In contrast, none of the control EBs (0/30) were GFP positive. In light of the finding that early neuroectoderm markers (*Sox1* and *nestin*) become dramatically up-regulated in TAF3 K/D EBs, we plated control and TAF3 K/D EBs (day 4) onto laminin coated slides. We observed extensive axon network out-growth from TAF3 K/D EBs but not from control EBs (Figure 3H, S3D and S3E). Apparently, even without chemical induction, TAF3 K/D can divert a significant proportion of ES cells to differentiate into neurons. We further confirmed the TAF3 K/D phenotype of EBs by western blot analysis (Figure S3B). In conclusion, these data strongly suggest that high levels of TAF3 in ES cells may be essential for proper cell lineage specification during differentiation.

To test whether TAF3 expression is lineage-specific in EBs, we co-stained EBs (day10) with anti-GATA4 and anti-TAF3 antibodies. Whereas GATA4 strongly stained the outer layer cells of EBs, TAF3 staining (though at low levels compared to ES cells; Figure 1B) was quite homogenous and apparently not lineage-specific (Figure 3I and 3J). Together, these results suggest a critical role for TAF3 in directing cell fate choices at very early stages during ES cell differentiation.

TAF3 Regulates Lineage-Specific Gene Expression Programs

To identify the full range of genes regulated by TAF3 in ES cells we combined shRNA-based TAF3 K/D with mRNA-seq. We also measured expression changes at EB day 3 and EB day 6 to characterize the temporal dynamics and downstream consequences of TAF3 depletion. On average we detected 2119 genes (~10% of those assayed) up-regulated at each timepoint (Table S1) and by EB day 6 these were massively biased towards neuroectoderm associated Gene Ontology categories such as "nervous system development" (P < 1E-20), "axon guidance" (P < 1E-9) and "synaptic transmission" (P < 1E-4; Table S2). Notably however, this bias was evident even in undifferentiated TAF3 K/D cells ("nervous system development", P < 1E-7) and could be traced to early neuroectoderm (*Sox1, Pax6*), neural crest (*Zic1, Zic2*) and neuronal stem cell markers (*Nes*; Figure 4A; Column 1). Subsequently, by EB day 3 and EB day 6, many markers of more differentiated cell types such as neurons (*Tubb3, Grm2, Kcnc1, Foxp2*), glia (*Fabp7, Gli1*) and oligodendrocyte (*Olig3*) were significantly up-regulated.

In contrast to neuroectodermal genes, endoderm markers were uniformly down-regulated by TAF3 K/D (Figure 4A; Column 2). Essentially all showed defects by EB day 3, suggesting that TAF3 depletion rapidly limits endoderm differentiation potential. It was surprising then that Gene Ontology analysis failed to identify a strong unifying theme for down-regulated genes, though consistent with our cell cycle assays (Figure 2E), we detected modest down-regulation of some housekeeping genes (Table S2). Reasoning that this was due to the smaller number of significantly down-regulated genes at each timepoint (1165 vs 2119 on average) and the comparative lack of information regarding endoderm development (Grapin-Botton and Constam, 2007), we directly compared our data to tissue- and stage-specific SAGE libraries (Khattra et al., 2007). Briefly, we ordered genes by the change in expression following TAF3 K/D and assessed the numbers of genes found exclusively in either of two SAGE libraries in windows along this axis. Our measure of tissue bias (Figure

4B, red line) shows that up-regulated genes were biased towards the brain library while down-regulated genes were biased towards the endoderm library. By contrast, tissue bias computed using random orderings of genes (grey line) or random pairings of libraries (Figure S4A) showed no relationship to the underlying gene expression (black points). Therefore, our data demonstrate that TAF3 is required both for repression of a neural expression program and for activation of many endodermal genes.

During early embryogenesis, the correct execution of patterning gene expression programs is essential for proper cell migration and fate allocation (Arnold and Robertson, 2009). In this regard, our genome-wide expression data also revealed an unexpected overlap between up-regulated genes and processes related to Wnt-β-catenin signaling such as "Wnt receptor signaling pathway" (P < 1E-6; Table S2). In addition to several members of the core β-catenin pathway, at least 8 Wnts were strongly affected as were the majority of frizzled and sFRP genes (Figure 4A; Columns 3-5). We also observed up-regulation of several members of the nodal pathway (Nodal, Lefty1, Cer1; Figure 4A; Column 3). In this case the pathway is too small to achieve statistical significance in genome-wide tests but the observation may nevertheless be biologically significant as discussed below. In contrast to the number of changes in the nodal pathway, only one subunit each of the Mediator and TFIID complexes (discounting TAF3 itself) displayed strong TAF3 K/D defects (Figure 4A; Columns 6-7).

Lineage Restriction of TAF3 K/D cells in a Teratoma Model

As a more stringent pluripotency test, a teratoma model was used to evaluate the *in vivo* differentiation capacity of TAF3 K/D cells. Control teratomas contained the expected tissue types from all three germ layers (Figure S5A; panel a - i). In contrast, teratomas generated from TAF3 K/D cells were devoid of endoderm tissues and were mainly composed of muscle (mesoderm), neural tissue and epidermis (ectoderm) (Figure S5A; panel j - o). To gain a quantitative view of how TAF3 K/D affected teratoma formation, qRT-PCR and western blots were used to assess the expression of different tissue-specific genes (Figure S5B and S5C). The expression of pan-endoderm markers (*Foxa2, Hnf4a, Afp* and *Gata4*) was significantly down-regulated (10 to 20 fold) while the expression of skeletal muscle specific genes *(Myog, MyoD, Myf5 & Myl2)* was dramatically up-regulated (5 to 10 fold) in TAF3 K/D teratomas. However, in contrast to the results of the *in vitro* EB formation experiments, the expression of neuronal markers was reduced (~2 fold) in TAF3 K/D samples. Despite this, differentiation towards neural lineages was largely unimpeded as significantly high levels of neural tissue can be observed in teratoma sections (Figure S5A; panel k and n). This finding using the teratoma assay thus deviates somewhat from the results we observed during EB formation wherein loss of TAF3 induced a high level of up-regulation of neuronal markers. We speculate that either the dramatic up-regulation of mesoderm differentiation ultimately overwhelms the neuroectoderm program during the long periods of teratoma formation *in vivo* (~6 weeks vs 6-10 days in EBs) or that environmental and physiological cues in the teratoma niche can reset the differentiation bias of TAF3 K/D cells.

Interestingly, *Trf3* was turned on after teratoma formation and *Taf3* mRNA was still expressed at modest levels in TAF3 K/D teratomas (~three fold less than control; Figure S5D), supporting the notion that residual TAF3 was sufficient to form a complex with TRF3 and promote myogenesis. This presents the intriguing possibility that enhanced early mesoderm differentiation might override the later negative effects of TAF3 K/D on myogenesis. Thus, TAF3 K/D cells likely retain some differentiation plasticity which depends on the environment and influences differentiation into distinct sets of mesoderm and ectoderm lineages.

TAF3 K/D Blocks Both Primitive and Definitive Endoderm Differentiation

Our previous results strongly suggest that TAF3 is required for the expression of endodermal genes (Figure 3A, 4A and 4B) but do not clearly distinguish between the primitive and definitive endoderm lineages. As these have very similar molecular signatures (Grapin-Botton and Constam, 2007), we sought to exclude the possibility that down-regulation of one could mask up-regulation of the other. Specifically, although defects in expression of the well characterized primitive endoderm specific genes (Sox7, *Lamb1, Col4a2* and *Sparc*) (Figure 4A) provide strong evidence that specification of primitive endoderm was impeded, most definitive endoderm markers (*Foxa2, Sox17* and *Hnf4a*) are pan-endodermal and require additional context. We therefore directed control and TAF3 K/D cells towards definitive endoderm using Activin A (Gadue et al., 2006) and analyzed the resulting definitive endoderm (CXCR$^+$c-Kit$^+$) and mesoderm (CXCR4$^+$Flk-1$^+$) cell populations by flow cytometry (Figure 5A). Definitive endoderm differentiation was much less efficient in the TAF3 K/D samples (~16-19%) than control samples (~43-45%). By contrast, mesoderm differentiation was enhanced from ~22% in control samples to ~39-46% in TAF3 K/D samples. Since both definitive endoderm and mesoderm are derived from a common precursor in cell culture, mesendoderm (Tada et al., 2005), it is likely that depletion of TAF3 results in an imbalance in mesendoderm lineage specification by disfavoring the expression of endodermal genes. This hypothesis is further supported by our qRT-PCR results (Figure 5B).

As an additional independent test of the role of TAF3 in definitive endoderm differentiation, we took advantage of human ES cells that were programmed to form definitive endoderm (Sox17 O/E lines; (Seguin et al., 2008). Figure 4C shows clearly that orthologous genes that increase during differentiation to definitive endoderm in human ES cells are down-regulated following TAF3 K/D in mouse ES cells and vice versa. Thus, multiple lines of evidence show that TAF3 is required for both primitive and definitive endoderm development.

TAF3 as a TFIID subunit Targets Core Promoters in ES cells

Given the wide-ranging consequences of TAF3 depletion we sought to identify direct targets of TAF3 regulation. Are the high levels of TAF3 in ES cells mainly a component of TFIID or are there other TAF3 containing complexes in play akin to the situation found during myotube formation? Can we survey the diversity of genes bound by TAF3 and perhaps discern some differential function associated with activated versus repressed genes? To address these questions we performed ChIP-seq experiments on TAF3, two canonical TFIID subunits (TBP, TAF1) and Pol II in mouse ES cells. TAF3 was robustly detected at 80% of promoters (Figure 6A; Figure S6A; P = 0 by permutation) and its enrichment at promoters was strongly correlated with that of TAF1 and TBP (Figure 6C). TAF3 has been shown to anchor TFIID to H3K4me3 in human cell lines (Vermeulen et al., 2007) and our data verify that this relationship is likely to persist in mouse ES cells: Enrichments of TAF1, TBP and TAF3 all correlate strongly with H3K4me3 levels (Figure 6C). Together with previous evidence of co-purification (Gangloff et al., 2001), these data leave little doubt that TAF3 binds core promoters in ES cells as a component of TFIID.

TAF3 binding at promoters was also positively correlated with Pol II binding (R = 0.78, P < 1E-10; Figure 6C; Figure S6B) and the level of expression as assayed by mRNA-seq (R = 0.51, P < 1E-10). Surprisingly however, we were unable to detect significant differences in TAF3 enrichment between promoters of TAF3 dependent genes and other genes (Figure S6G). Similarly, a simple linear model of expression as a function of TAF3, TAF1 and TBP promoter binding identified a large shared contribution (presumably corresponding to TFIID), but no residual relationship between promoter-bound TAF3 and expression changes following TAF3 K/D (data not shown). Thus, although TAF3 is recruited to core promoters

in ES cells and contributes to gene expression via TFIID, this function of TAF3 does not appear to account for the defects in lineage-specific expression observed upon depletion of TAF3 from ES cells.

Binding of TAF3 is Correlated with Binding of CTCF and cohesin

Although the vast majority of regions enriched for TAF1, TBP or Pol II were also enriched for TAF3, the opposite was not true. At a false discovery rate of 1% our ChIP-seq data indicated that 19K (of 38K total) regions enriched for TAF3 binding were not enriched for any of TAF1, TBP or Pol II (range 12-19K; Figure 6A solid box; Figure S6C). These regions were generally further from core promoters (Figure S6D) and less enriched (Figure S6E) than TFIID-associated regions but often overlapped with regions enriched for other factors active in ES cells (Figure S6F). For example, the number of CTCF peaks coincident with TFIID-independent TAF3 peaks was ten times that expected by chance. To get a more quantitative understanding of TAF3 binding in ES cells, we computed correlations among 17 factors at 125K locations across the genome (See *Experimental Procedures*). These correlations correctly reconstructed the known relationships among all 17 factors such as the concerted binding of TFIID components to promoters and the association between TFIID/ Pol II and H3K4me3 in ES cells (Figure 6C). Notably however, we also identified a striking correlation between the binding of TAF3 and CTCF that is not shared by any other member of TFIID or Pol II. A similar relationship was observed for cohesin (Smc1A, Smc3) and both relationships were found to be more robust when located distal to the core promoter. The observation that high levels of TAF3 at promoter distal sites are often accompanied by high levels of CTCF and cohesin while low levels of TAF3 signal low levels of CTCF and cohesin, suggests that these molecules operate together at these promoter distal sites to perform a linked function. Moreover, as there was no correlation between binding of TAF1, TBP or Pol II with CTCF, this TAF3 activity does not appear to depend on TFIID (Figure 6C). Examples of regions enriched for TAF3 and CTCF are shown in Figure 6B.

Regions Enriched for TAF3, CTCF and cohesin Distinguish TAF3-Activated from TAF3-Repressed Genes

Given the existence of distinct classes of TAF3 binding, one plausible explanation for the opposing effects of TAF3 depletion on neuroectodermal and endodermal genes is that they are subject to different types of TAF3 regulation. To discriminate regions bound by TAF3 in the context of different partners, we performed principal components analysis on our dataset and clustered regions significantly enriched for TAF3 into four classes (Figure 6D; See *Experimental Procedures*). This procedure groups together regions that have similar enrichment profiles (Figure 6E) across all 17 factors and histone modifications examined in Figure 6C. Briefly, as well as TAF3, class 1 regions are enriched for TFIID, Pol II and H3K4me3. Their proximity to the TSS of known genes confirms that they are predominantly TFIID-bound core promoters (Figure 6E). By contrast, Class 2 regions have low levels of TFIID and H3K4me3 but are enriched for Oct4/Nanog/Sox2 and mediator components. They are considered further in the *Discussion*. Class 3 regions are specifically enriched for TAF3, CTCF and cohesin subunits, and correspond to the novel function proposed above. Finally, class 4 regions are not enriched for any of the factors we considered besides TAF3 (Figure 6A).

We tested whether particular TAF3 binding classes were associated with TAF3-dependent genes by comparing the density of each bound region type (within 100Kb of the gene) between TAF3-dependent genes and sets of matched control genes (See *Experimental Procedures*). Genes whose expression was TAF3-dependent in ES cells exhibited genome-wide associations with TAF3 binding classes that were not observed among control genes (Figure 6F). First, both up- and down-regulated genes were enriched for at least one class of

TAF3 bound region compared to controls, suggesting that TAF3 may directly regulate both sets of genes. Second, up-and down-regulated genes exhibited radically different associations with TAF3 bound regions. Most strikingly, whereas genes down-regulated upon TAF3 K/D were surrounded by more regions enriched for TAF3, CTCF and cohesin (Class 3 regions; Figure 6 D-F) than expected by chance, no such association was seen for genes up-regulated upon TAF3 K/D (Figure 6F). The simplest interpretation of these data is that class 3 regions are required by certain genes for efficient expression and are enriched in the vicinity of these genes. Depletion of TAF3 interferes with this function. By contrast, genes that are up-regulated after TAF3 K/D rely on other mechanisms to achieve high levels of expression. Indeed, our data suggest that activation by Sox2, Oct4 and Nanog may comprise such an alternative mechanism that can apparently be opposed by TAF3 (Figure 6F). Our data support two other observations. First, regions bound by TAF3 only (Class 4 regions; Figure 6 D-F) are also associated with down-regulated genes upon TAF3 K/D. This suggests that TAF3 may also contribute to lineage commitment at distal sites by mechanisms independent of CTCF/cohesin. Second, TAF3 binding in the context of TFIID (Class 1 regions; Figure 6 D-F) is associated neither with up- nor down-regulated genes. This is confirms our previous observations (Figure S6G) and is consistent with the critical role played by TFIID at the basal promoters of many genes.

CTCF Directly Interacts with and Recruits TAF3 to Promoter Distal Sites

The strongly correlated binding of CTCF and TAF3 to promoter distal sites (Figure 6E) suggested that these two proteins may be tightly associated perhaps even forming a protein complex. Consistent with this hypothesis, TAF3 and CTCF co-eluted when ES cell nuclear extracts were chromatographed on a Superose 6 column (Figure S1B). Importantly, CTCF but not cohesin (Smc1a and Smc3) or mediator subunit (Med12) were selectively enriched by TAF3 immunoprecipitation (Figure 7A) and the interaction between TAF3 and CTCF appears to be direct and independent of DNA (Figure 7A; lane 4). Co-immunoprecipitations using 293T cells expressing full length or truncated Flag-HA-tagged TAF3 and CTCF proteins confirmed that TAF3 directly interacts with CTCF through its vertebrate-specific region (501-730aa) without assistance of the Histone Fold (1-79aa) or the PHD finger (869-914aa) (Figure 7B and 7C).

To test the order of the recruitment between TAF3 and CTCF, we examined the occupancy of CTCF and TAF3 at distinct sets of genomic loci in control TAF3 K/D and CTCF K/D cells. We found that CTCF continues to bind its target sites in the absence of TAF3 (Figure S7A) whereas CTCF is required for efficient recruitment of TAF3 to distal (TAF3/CTCF) sites but not required for TAF3 occupancy at core promoters (Figure 7D). These findings further validate the functional relationship between TAF3 and CTCF. Moreover, they confirm that TAF3 localization to CTCF-bound sites (Class 3 in Figure 6) is mechanistically distinct from TAF3 localization to TFIID-bound regions (Class 1 in Figure 6), corroborating our computational inference of distinct TAF3-binding categories. These observations together provide strong evidence for an unexpected promoter distal co-activator mechanism involving TAF3 in association with CTCF.

TAF3 Mediates DNA Looping Between the Promoter Distal Sites and Core Promoters Occupied by TAF3/CTCF/cohesin

The over-representation of TAF3/CTCF/cohesin bound regions associated with TAF3-activated genes suggests that TAF3 might provide a novel function at these locations to activate gene transcription possibly by facilitating long distance DNA looping. To address this possibility, we focused on two TAF3-activated genes, *Mapk3* and *Psmd1* (Table S1), which could be confidently associated with specific TAF3/CTCF/cohesin sites (most are 10's of kbs from the nearest TSS's and thus cannot be unambiguously assigned to a specific

target gene). Each of these genes is located at ~5 kb downstream of a TAF3/CTCF/cohesin bound region with no other TSS nearby. Chromatin Conformation Capture (3C) experiments were performed to scan the interaction frequency between the promoter distal site and regions within the gene locus. In each case, DNA looping between the distal TAF3/CTCF/ cohesin site and the core promoter was observed and, notably, TAF3 is required for efficient DNA looping (Figure 7E and S7B). These findings strongly support a molecular mechanism in which TAF3/CTCF mediates long-range chromatin transactions that likely regulate proper transcription activation (Figure 7H). Consistent with this model, shRNA mediated depletion of CTCF also reduced expression of these genes and the simultaneous K/D of TAF3 and CTCF did so more effectively (Figure 7F, 7G and S7C). Since activation of Ras/ Erk/Mapk is sufficient to induce primitive endoderm differentiation (Li et al., 2010; Verheijen et al., 1999), the down-regulation of *Mapk3* is a likely contributor to the endoderm defects we observed in TAF3 K/D samples (Figure 4A).

Together, these results strongly suggest that TAF3 and CTCF act together at a subset of CTCF sites to perform a linked function important for specifying endoderm lineages (Figure 7H). To further verify this model, CTCF K/D cells were induced to form EBs. As expected, we observed significantly compromised endoderm differentiation as demonstrated by loss of marker (*Gata4, Afp* and *Apoa1*) expression (Figure S7D and S7E). Interestingly, these defects were less dramatic than those seen upon TAF3 depletion, indicating that other CTCF independent functions mediated by TAF3 might also play a role. One possibility is that the Class 4 binding regions (Figure 6D-E) that are also associated with TAF3-activated genes (Figure 6F) contribute to endodermal gene expression. TAF3 may also regulate lineage commitment via one of its core promoter functions such as H3K4me3 binding.

Discussion

Misregulation of Lineage Commitment by TAF3 Depletion

A notable feature of our mRNA-seq data is that, although a similar number of genes was affected by TAF3 K/D at each time point, many more lineage-specific genes were over-represented at later time points (Table S1 and S2). Thus, an interesting question is what underlies the progressive bias towards lineage-specific genes. The most parsimonious explanation is that TAF3 K/D first induces a broad effect on gene expression. However, some of the genes that TAF3 regulates are likely critical for appropriate lineage specification. The altered expression of these genes upon TAF3 depletion 1) would directly limit the differentiation potential of cells or 2) would lead to altered cellular responses to environmental and physiological signaling events. Consistent with mechanism 1, key developmental regulators of neuroectoderm and endoderm become profoundly affected by depletion of TAF3 in ES cells. Indeed, our findings suggest that TAF3 directly influences the balance between neuroectoderm and endoderm differentiation of ES cells at the transcriptional level. This initial misregulation may in turn trigger a temporal cascade of changes that disrupt lineage-specific programs of gene expression (Figure 4A and S4A). Several lines of evidence suggest that TAF3 K/D may also indirectly affected mesoderm differentiation at later times by disrupting some key signaling pathways (mechanism 2). For example, many genes involved in mesoderm formation (*T, Nodal, Gsc, Wnt3, Wnta, Foxh1 and Fgf8*) were not significantly altered until EB day 3 or day 6 (Figure 4A), more consistent with an indirect regulation by TAF3. Significantly, some core mesoderm patterning programs malfunctioned and became unresponsive to the over-expression of nodal antagonist Cer1/Lefty1 (Figure 4A), that normally represses mesoderm formation (Tam and Loebel, 2007). Likewise, TGF-p signaling became deregulated upon TAF3 depletion. Specifically, when cells lacking TAF3 were exposed to activin A (TGF-p activator) concentrations that would normally induce definitive endoderm differentiation, TAF3 K/D cells instead differentiated more into mesoderm (Figure 5). Taken together, these

results are most consistent with TAF3 mediating two distinct mechanisms that regulate proper lineage specification.

Multifaceted roles of TAF3 in Transcriptional Regulation

Given the well established role of TAF3 in TFIID, it's quite surprising that, with the exception of the modest down-regulation of some housekeeping genes (Table S1 and S2), we were unable to detect an obvious association between levels of TAF3 at proximal core promoters *(i.e.* in TFIID) and changes in gene expression following TAF3 K/D. One possible explanation is that for the large class of genes where TAF3 functions as part of the prototypic TFIID complex the residual low levels of TAF3 after K/D mimic the low levels typically seen in differentiated cells *(i.e.* muscle) and are sufficient to maintain TFIID function in ES cells. In such a model, the very high levels of TAF3 seen in ES cells may engage in additional ES cell specific functions.

Intriguingly, the striking enrichment of TAF3 at regions bound by CTCF and cohesin presents a likely candidate for an ES cell specific function. Although the data shown here do not exclude the possibility that TAF3 and CTCF also bind together in differentiated cell types, they strongly suggest that TFIID is absent from the majority of these regions in ES cells (Figure 6B-C). Indeed, conventional "peak-calling" methods (Zhang et al., 2008) determined that 5463 regions were enriched for TAF3, CTCF and some combination of cohesin and mediator subunits (but excluding all other factors) compared to 211 such regions enriched for TBP but not TAF3. By contrast, essentially every region bound by TAF1 and Pol II is enriched for both TAF3 (98%) and TBP (94%). Even more importantly, these regions are over-represented around TAF3-activated genes (Figure 6F), thus establishing a potential mechanism for the differential regulation of genes from different lineages.

CTCF has been implicated in multiple regulatory functions, including transcriptional activation/repression, insulator activity and imprinting (Phillips and Corces, 2009). The molecular basis of these diverse activities remains unclear. Here we show that by interacting with CTCF, TAF3 can directly mediate linkages between distal TAF3/CTCF/cohesin bound regions and proximal core promoters thus providing another means to influence transcription activation at target genes (Figure 7H). Our genome-wide correlation analysis (Figure 6F) suggests that this mechanism likely governs the proper transcription of many genes. However, it's worth noting that our data do not exclude the possibility that TAF3 could also perform core promoter independent functions with CTCF at those locations.

In addition to regions that are bound by TAF3 in the context of TFIID and those that are bound in the context of CTCF/cohesin, our data suggest that two other modes of TAF3 binding may exist. Specifically, we observe a class of sites that appear to be enriched for TAF3 only (Figure 6A) and another class that is enriched for Oct4/Nanog/Sox2 as well as mediator subunits (Figure 6B). Although it has been previously shown that not all binding is functionally significant (Li et al., 2008; MacArthur et al., 2009), these cases are interesting as they exhibit biased representation with respect to TAF3-dependent genes. Notably, the class 2 binding regions are enriched around up-regulated genes upon TAF3 depletion (Figure 6F), consistent with a mechanism of transcription repression by TAF3 in association with Oct4/Nanog/Sox2. If indeed TAF3 is involved with new functions at these regions, it could represent an interesting point of convergence between pathways responsible for self-renewal and pluripotency.

The present work demonstrates a novel ES cell specific role of TAF3 in maintaining pluripotency. In conjunction with our previous work on TAF3 and other examples such as TAF4b (Goodrich and Tjian, 2010), these studies collectively reinforces cell-type specific

regulatory functions of components of core promoter complexes as a general paradigm. Indeed, it now seems likely that this idea may have very wide applicability. For example, during the course of demonstrating a specialized role for TAF3 in myotubes, we found that mediator components were specifically down-regulated in these cells (Deato et al., 2008). This parallels independent work showing that multiple mediator subunits are required along with cohesin for Oct4 expression and ES cell self-renewal (Kagey et al., 2010). Given these commonalities (especially the link to cohesin), it was striking to find a negligible self-renewal phenotype in TAF3 K/D cells (Figure 2). Is it possible that these core promoter factors have been charged with independently guaranteeing the two defining characteristics of stem cells? If this hypothesis survives additional testing it may suggest that along with site-specific transcription factors and chromatin modifiers (Jaenisch and Young, 2008), core promoter complexes and their associated functions comprise another important layer of transcriptional regulation for safeguarding the integrity of the stem cell state.

Experimental Procedures

ES Cell Maintenance

Mouse D3 (ATCC) and 46C ES cells were cultured on 0.1% gelatin coated plates in the absence of feeder cells. The ES cell medium was prepared by supplementing knockout DMEM (Invitrogen) with 15% FBS, 1mM glutamax, 0.1mM nonessential amino acids, 1mM sodium pyruvate, 0.1mM 2-mercaptoethanol and 1000 units of LIF (Millipore). Mouse R1 (ATCC) was maintained in the same medium with a feeder layer of irradiated MEFs (Passage 3).

Validation of mRNA-seq and ChIP-seq Data by Q-PCR

Relative mRNA abundance of 40 genes that showed differential expression at EB day 6 (shRNA A) were confirmed with qRT-PCR by comparing shRNA B treated samples with controls (Figure S4B and S4C). Enrichment of distinct sets of factors as indicated at 21 genomic regions was validated by ChIP-qPCR using a rabbit antibody and a guinea pig antibody against TAF3 (Figure S6H). The primer information for data validation is in Table S3.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Arnold SJ, Robertson EJ. Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. Nat Rev Mol Cell Biol. 2009; 10:91–103. [PubMed: 19129791]

Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell. 2006; 125:315–326. [PubMed: 16630819]

Bradley A, Evans M, Kaufman MH, Robertson E. Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. Nature. 1984; 309:255–256. [PubMed: 6717601]

Daley GQ, Scadden DT. Prospects for stem cell-based therapy. Cell. 2008; 132:544–548. [PubMed: 18295571]

Deato MD, Marr MT, Sottero T, Inouye C, Hu P, Tjian R. MyoD targets TAF3/TRF3 to activate myogenin transcription. Mol Cell. 2008; 32:96–105. [PubMed: 18851836]

Deato MD, Tjian R. Switching of the core transcription machinery during myogenesis. Genes Dev. 2007; 21:2137–2149. [PubMed: 17704303]

Farnham PJ. Insights from genomic profiling of transcription factors. Nat Rev Genet. 2009; 10:605–616. [PubMed: 19668247]

Gadue P, Huber TL, Paddison PJ, Keller GM. Wnt and TGF-beta signaling are required for the induction of an in vitro model of primitive streak formation using embryonic stem cells. Proc Natl Acad Sci U S A. 2006; 103:16806–16811. [PubMed: 17077151]

Gangloff YG, Pointud JC, Thuault S, Carre L, Romier C, Muratoglu S, Brand M, Tora L, Couderc JL, Davidson I. The TFIID components human TAF(II)140 and Drosophila BIP2 (TAF(II)155) are novel metazoan homologues of yeast TAF(II)47 containing a histone fold and a PHD finger. Mol Cell Biol. 2001; 21:5109–5121. [PubMed: 11438666]

Goodrich JA, Tjian R. Unexpected roles for core promoter recognition factors in cell-type-specific transcription and gene regulation. Nat Rev Genet. 2010; 11:549–558. [PubMed: 20628347]

Grapin-Botton A, Constam D. Evolution of the mechanisms and molecular control of endoderm formation. Mech Dev. 2007; 124:253–278. [PubMed: 17307341]

Hart DO, Santra MK, Raha T, Green MR. Selective Interaction Between Trf3 and Taf3 Required for Early Development and Hematopoiesis. Dev Dynam. 2009; 238:2540–2549.

Jaenisch R, Young R. Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. Cell. 2008; 132:567–582. [PubMed: 18295576]

Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, Ebmeier CC, Goossens J, Rahl PB, Levine SS, et al. Mediator and cohesin connect gene expression and chromatin architecture. Nature. 2010; 467:430–435. [PubMed: 20720539]

Khattra J, Delaney AD, Zhao Y, Siddiqui A, Asano J, McDonald H, Pandoh P, Dhalla N, Prabhu AL, Ma K, et al. Large-scale production of SAGE libraries from microdissected tissues, flow-sorted cells, and cell lines. Genome Res. 2007; 17:108–116. [PubMed: 17135571]

Li L, Sun L, Gao F, Jiang J, Yang Y, Li C, Gu J, Wei Z, Yang A, Lu R, et al. Stk40 links the pluripotency factor Oct4 to the Erk/MAPK pathway and controls extraembryonic endoderm differentiation. Proc Natl Acad Sci U S A. 2010; 107:1402–1407. [PubMed: 20080709]

Li XY, MacArthur S, Bourgon R, Nix D, Pollard DA, Iyer VN, Hechmer A, Simirenko L, Stapleton M, Luengo Hendriks CL, et al. Transcription factors bind thousands of active and inactive regions in the Drosophila blastoderm. PLoS Biol. 2008; 6:e27. [PubMed: 18271625]

MacArthur S, Li XY, Li J, Brown JB, Chu HC, Zeng L, Grondona BP, Hechmer A, Simirenko L, Keranen SV, et al. Developmental roles of 21 Drosophila transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions. Genome Biol. 2009; 10:R80. [PubMed: 19627575]

Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature. 2007; 448:553–560. [PubMed: 17603471]

Naar AM, Lemon BD, Tjian R. Transcriptional coactivator complexes. Annu Rev Biochem. 2001; 70:475–501. [PubMed: 11395415]

Phillips JE, Corces VG. CTCF: master weaver of the genome. Cell. 2009; 137:1194–1211. [PubMed: 19563753]

Seguin CA, Draper JS, Nagy A, Rossant J. Establishment of endoderm progenitors by SOX transcription factor expression in human embryonic stem cells. Cell Stem Cell. 2008; 3:182–195. [PubMed: 18682240]

Shen WC, Bhaumik SR, Causton HC, Simon I, Zhu X, Jennings EG, Wang TH, Young RA, Green MR. Systematic analysis of essential yeast TAFs in genome-wide transcription and preinitiation complex assembly. EMBO J. 2003; 22:3395–3402. [PubMed: 12840001]

Tada S, Era T, Furusawa C, Sakurai H, Nishikawa S, Kinoshita M, Nakao K, Chiba T. Characterization of mesendoderm: a diverging point of the definitive endoderm and mesoderm in

embryonic stem cell differentiation culture. Development. 2005; 132:4363–4374. [PubMed: 16141227]

Tam PP, Loebel DA. Gene function in mouse embryogenesis: get set for gastrulation. Nat Rev Genet. 2007; 8:368–381. [PubMed: 17387317]

Tjian R, Maniatis T. Transcriptional activation: a complex puzzle with few easy pieces. Cell. 1994; 77:5–8. [PubMed: 8156597]

Verheijen MH, Wolthuis RM, Bos JL, Defize LH. The Ras/Erk pathway induces primitive endoderm but prevents parietal endoderm differentiation of F9 embryonal carcinoma cells. J Biol Chem. 1999; 274:1487–1494. [PubMed: 9880524]

Vermeulen M, Mulder KW, Denissov S, Pijnappel WW, van Schaik FM, Varier RA, Baltissen MP, Stunnenberg HG, Mann M, Timmers HT. Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. Cell. 2007; 131:58–69. [PubMed: 17884155]

Yao J, Fetter RD, Hu P, Betzig E, Tjian R. Subnuclear segregation of genes and core promoter factors in myogenesis. Genes Dev. 2011; 25:569–580. [PubMed: 21357673]

Ying QL, Stavridis M, Griffiths D, Li M, Smith A. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. Nat Biotechnol. 2003; 21:183–186. [PubMed: 12524553]

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008; 9:R137. [PubMed: 18798982]
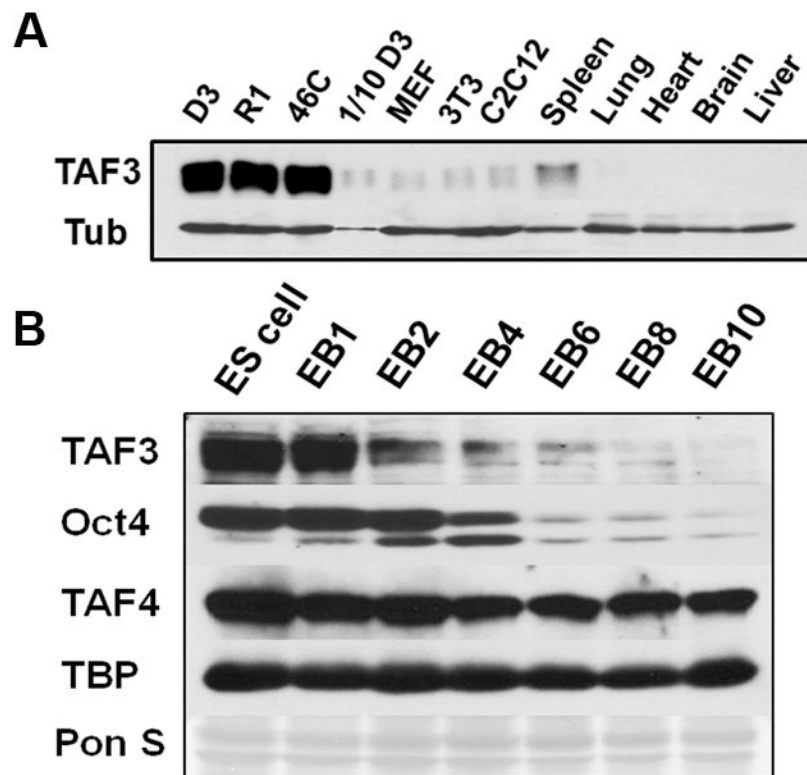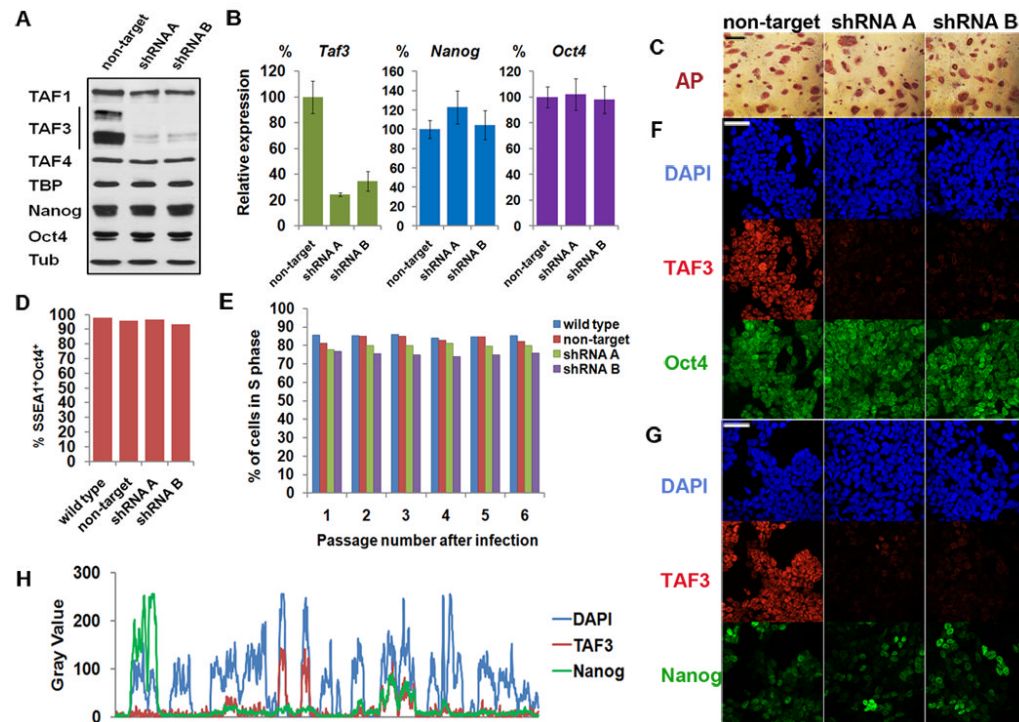
Figure 1. High Levels of TAF3 in ES Cells
(A) Western blots to examine TAF3 levels in different cell types and tissues.
(B) TAF3 is selectively reduced upon ES cell differentiation (EB day 1, 2, 4, 6, 8 and 10).
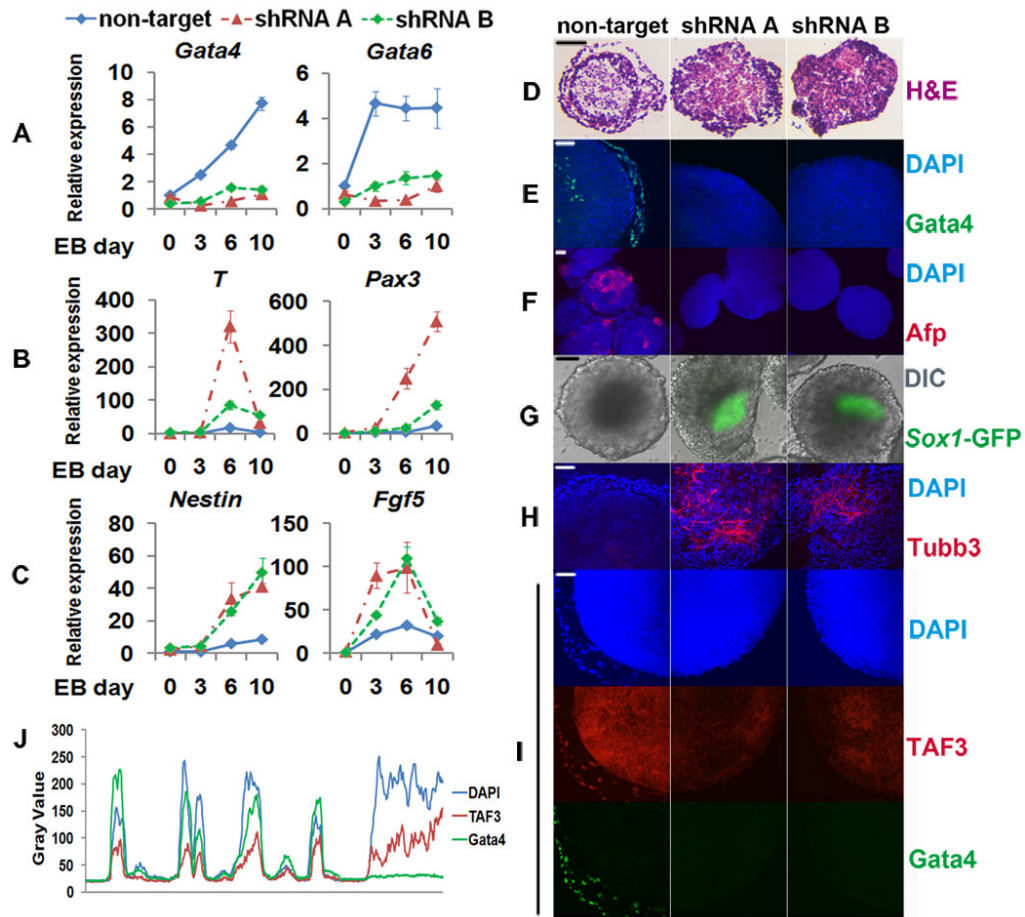
Figure 2. TAF3 Depletion does not Affect ES cell Self-renewal Genes

(A) Western blots to analyze the protein levels of *Taf1, Taf3, Taf4, Tbp, Nanog* and *Oct4* in control and TAF3 K/D cells (D3, P2 post infection).

(B) qRT-PCR to measure the mRNA levels of *Taf3, Nanog* and *Oct4* in control and TAF3 K/D cells (D3, P2 post infection).

(C) Alkaline phosphatase staining of control and TAF3 K/D cell colonies (R1, P4 post infection) on feeder cells.

(D) Percentages of SSEA1$^+$Oct4$^+$ cells in wild type, control and TAF3 K/D cell populations (D3, P6 post infection).

(E) Percentages of S phase cells in wild type, control and TAF3 K/D cell populations (D3, analyzed for 6 passages post infection).

(F-G) Confocal fluorescence images of control and TAF3 K/D cells (R1, Passage 4 post infection) co-stained with antibodies against TAF3 and Oct4 (F) or TAF3 and Nanog (G). Nuclei were counterstained with DAPI.

(H) Single cell expression correlation analysis of TAF3 and Nanog in shRNA A treated cells (R1, P4 post infection). Gray values along a random chosen path in DAPI, TAF3 (shRNA A), and Nanog channels of (G) were plotted in the diagram.

Scale Bars, (C) 150 μm, others 50μm.

Figure 3. TAF3 Regulates ES Cell Pluripotency

(A-C) qRT-PCR to measure the expression of lineage-specific markers in control and TAF3 K/D samples.

(D) H&E staining of EB sections from control and TAF3 K/D EBs (day 10). (E-F) Confocal fluorescence images of control and TAF3 K/D EBs (day 10) stained with antibody against GATA4 (E) or Afp (F).

(G) Live-cell fluorescence images of control and TAF3 K/D EBs (day 8, 46C).

(H) Confocal fluorescence images of control and TAF3 K/D EBs stained with antibody against Tubb3. EBs (day 4) were cultured on Laminin coated slides for 6 additional days.

(I) Confocal fluorescence images of control and TAF3 K/D EBs (day 10) co-stained with antibodies against TAF3 and GATA4.

(J) Gray values along a path from outside to inside of the EB (non-target) in DAPI, TAF3, and GATA4 channels of (I) were plotted in the diagram.
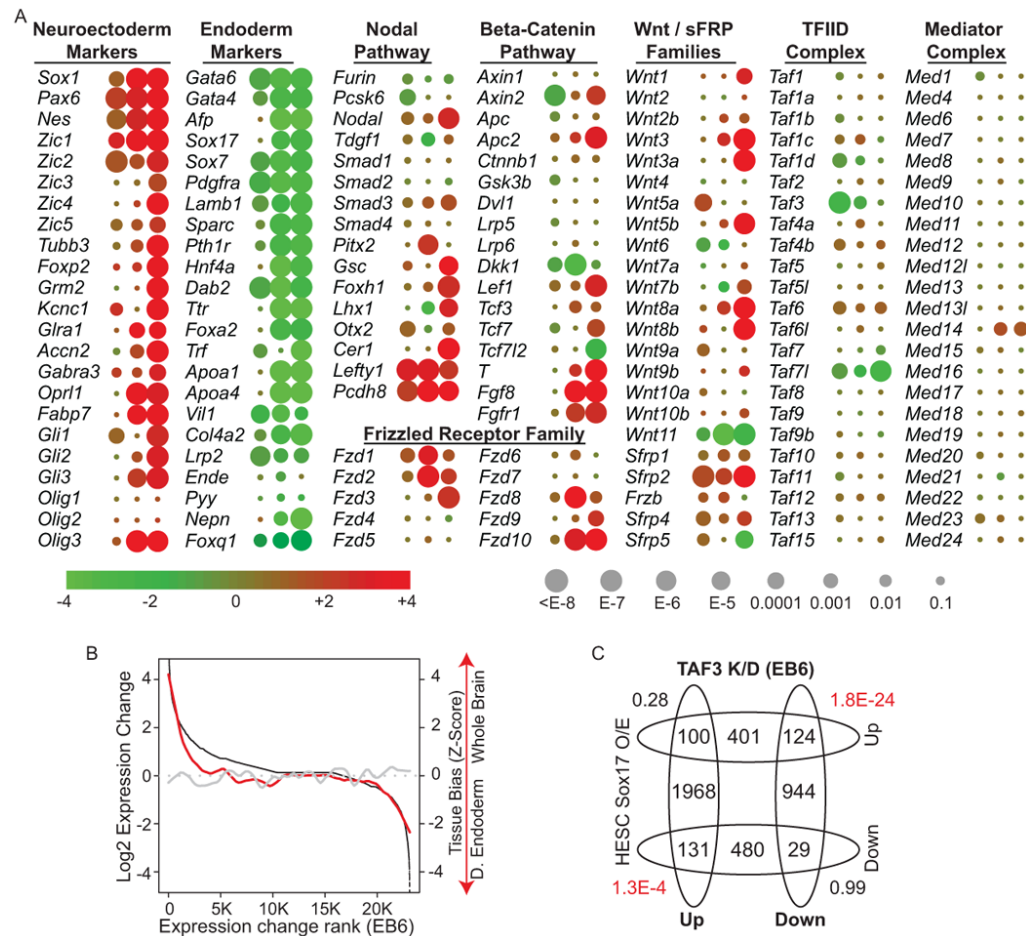
Scale Bars, 50 μm.

Figure 4. TAF3 Represses Neuroectodermal Genes and is Required for Expression of Endodermal Genes

(A) Difference in expression (log2(anti-TAF3 shRNA A/non-target shRNA)) and statistical significance between TAF3 depleted ES cells and controls for selected genes at three timepoints, ES Cell, EB day 3 and EB day 6.

(B) Tissue bias (red line; *Experimental Procedures*) calculated between definitive endoderm (E8.0; SM143) and whole brain (E12.5; SM108) SAGE libraries plotted on the difference in expression (defined as in (A); black points) between TAF3 K/D cells and controls. Tissue bias recalculated using randomized gene expression data was plotted in grey.

(C) Overlaps between genes that were differentially expressed in TAF3 K/D EBs at day 6 (EB6) and genes that were differentially expressed in Sox17 over-expressing human embryonic stem cells. P-values (italics) for each intersection were calculated assuming a hypergeometric distribution and a total ortholog count of 14,400. Values significant at the 5% level were plotted in red.
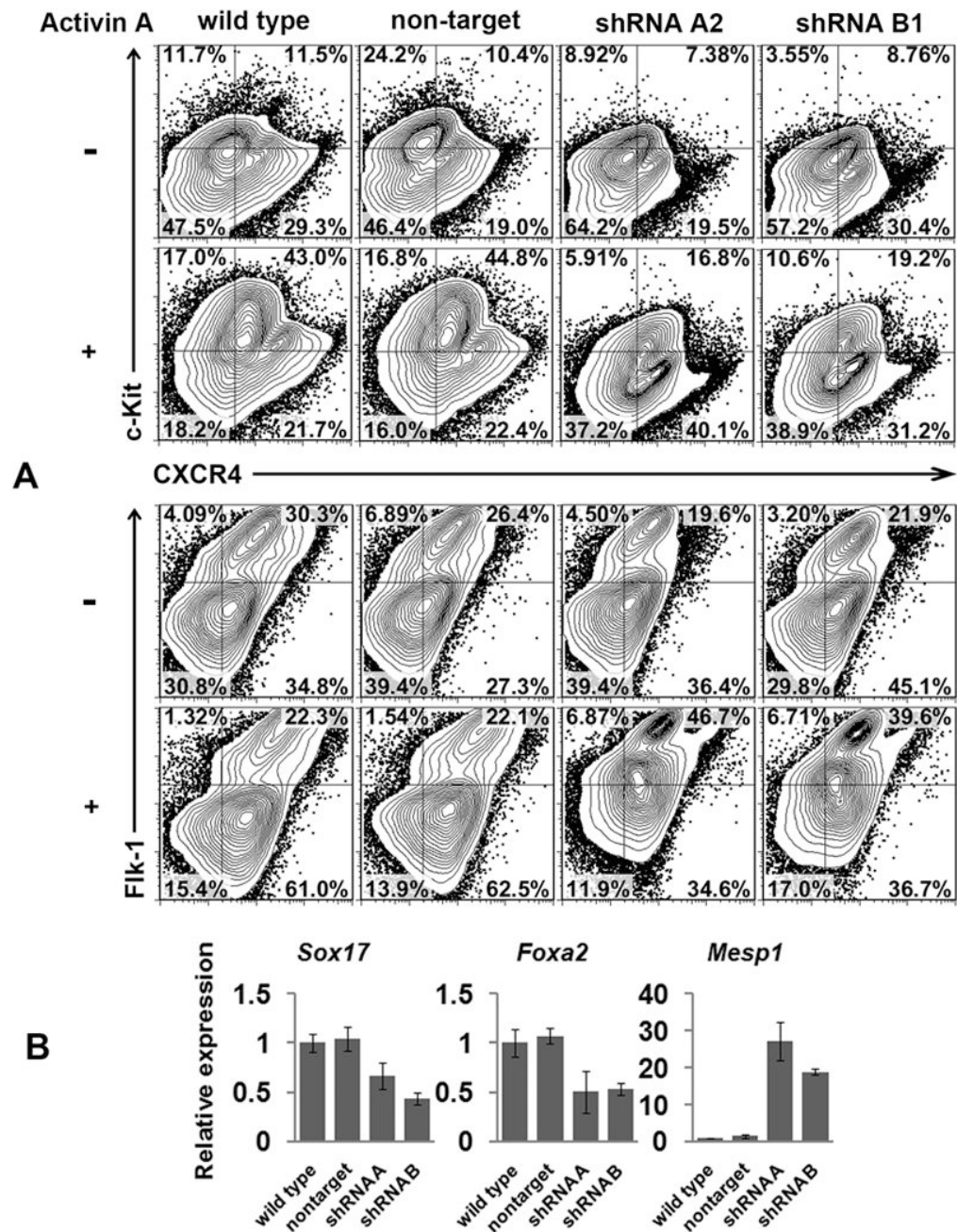
Figure 5. TAF3 is Required for Efficient ES cell Differentiation into Definitive Endoderm
(A) Flow cytometric analysis of definitive endoderm (CXCR4+c-kit+) and mesoderm (CXCR4+Flk-1+) cells after the directed *in vitro* differentiation. Experiments were conducted with or without Activin A.
(B) qRT-PCR to examine the relative expression of endoderm (*Sox17, Foxa2*) and mesoderm (*Mesp1*) markers in Activin A treated samples.
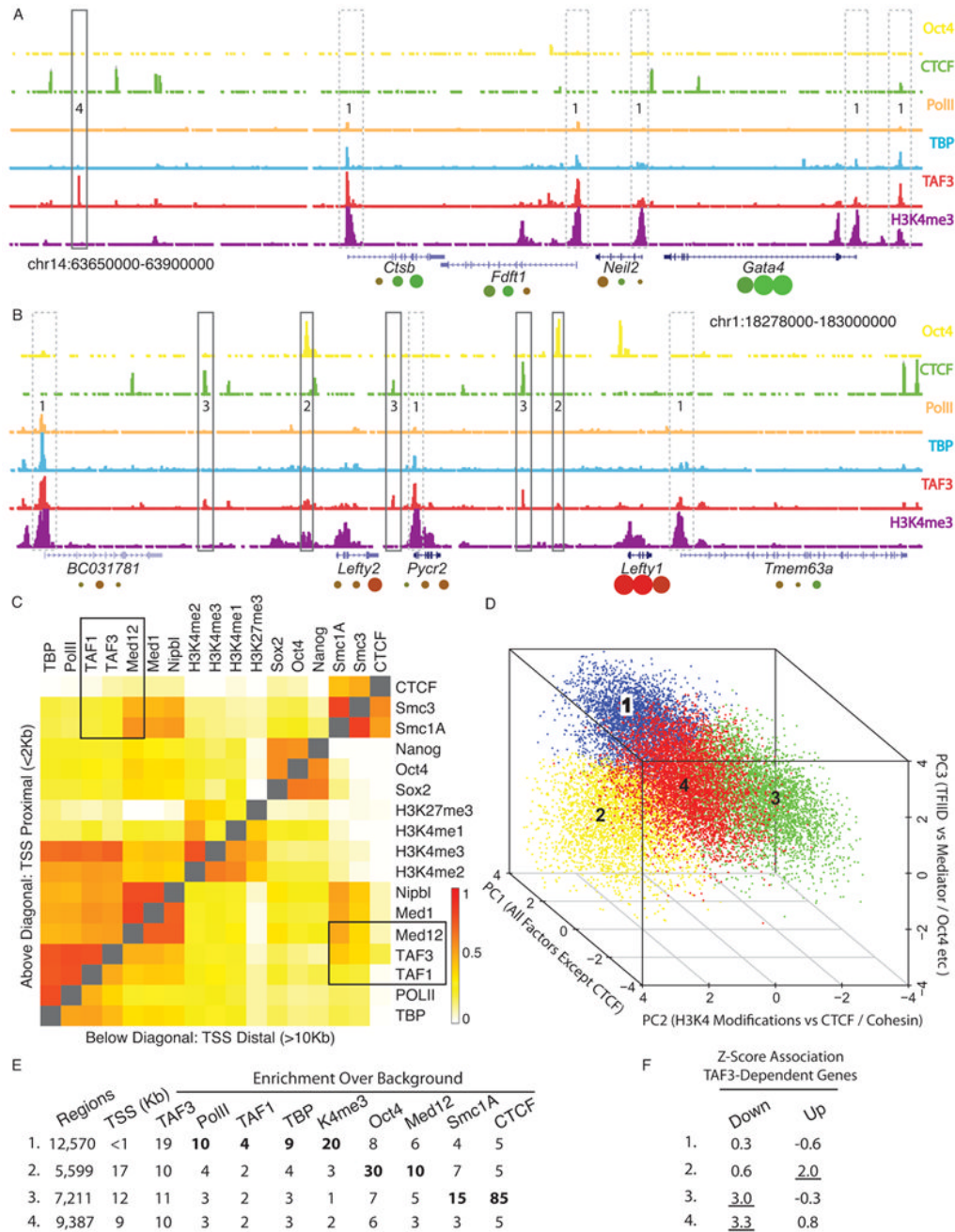
Figure 6. TAF3 Binds Core Promoters with TFIID and Distal Sites in a TFIID Independent Manner

(A) Read accumulation for six ChIP-seq datasets (TAF3, TBP, Pol II, H3K4me3, CTCF and Oct4) at the *Gata4* locus. Vertical axis was from 3 to 50 reads for all factors. TAF3 dependent expression changes were plotted below genes as for Figure 4. Peaks referenced in the text were boxed and numbered as in panels D-F. Active promoters were denoted with dashed boxes.

(B) As for (A) at the *Lefty1* locus.

(C) Binding correlations (Spearman coefficient) between all pairs of factors considered in this study based on enrichment values calculated using the MACS model at 125K genomic loci. Negative correlations were set to zero for display.

(D) Clustering of TAF3 bound regions into four classes using the first three principal components of binding variation computed from all factors in (C).

(E) Summary statistics for the four TAF3 binding classes in (D). All values except the number of regions in each class (first column) are median values for the whole class. Enrichment was calculated as for (C).

(F) Z-score association between TAF3 binding classes from (D) and genes whose expression was significantly affected by TAF3 depletion. Significant values were underlined; Z-score between +/-1 (P-value > 0.6); Z-scores > 2 (P-value < 0.05); Z-scores > 3 (P-value < 0.001). See *Experimental Procedures* for details of Z-score calculation.
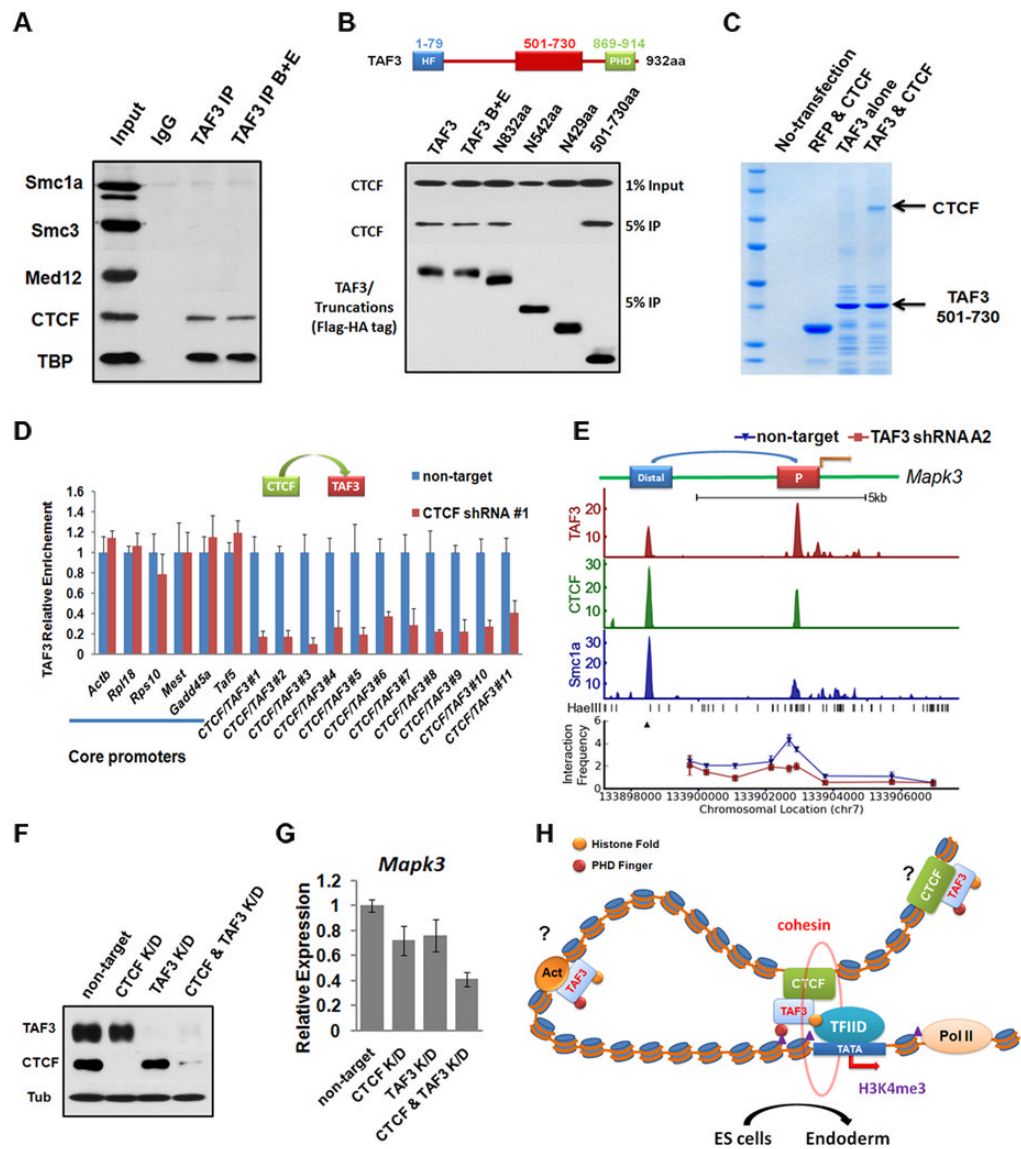
Figure 7. TAF3 Interacts with CTCF and Mediates DNA Looping

(A) Western Blots to examine the relative enrichment of cohesin (Smc1a and Smc3), mediator subunit (med12) and CTCF by TAF3 immunoprecipitation using ES cell nuclear extract. B + E, Benzonase and Ethidium bromide treatment that eliminates DNA-mediated interactions.

(B) and (C) Mapping the CTCF-interacting domain of TAF3. Flag-HA tagged TAF3, RFP or truncated TAF3 proteins were over-expressed in 293T cells with CTCF and were then purified with anti-Flag resin. TAF3 and CTCF enrichments were analyzed by western blots.

(C) Coomassie Staining of the purified Flag-HA-TAF3(501-730aa)/CTCF complex.

(D) ChIP analysis to examine the relative TAF3 enrichment at 6 core promoters bound by TAF3 and 11 distal sites bound by TAF3/CTCF in control and CTCF K/D cells.

(E) 3C experiments to assess long-range interactions between the distal TAF3/CTCF/ cohesin site and regions around the core promoter of *Mapk3* in control (blue) and TAF3 K/D (red) cells. ▲, anchor point.

(F) Western blots to examine TAF3 and CTCF protein levels in control, TAF3 K/D (shRNA A), CTCF K/D (shRNA #1) and TAF3 & CTCF K/D cells (72 hours post infection).

(G) qRT-PCR analysis to measure *Mapk3* expression in control, TAF3 K/D, CTCF K/D and TAF3 & CTCF K/D cells.

(H) TAF3 interacts with CTCF and mediates regulatory DNA looping to specify endoderm differentiation.