Title
Molecular Mechanisms for Regulation of Embryonic Stem Cell Lineage Commitment by Core Promoter Factor, TAF3

Author
Liu, Zhe

Publication Date
2012

Peer reviewed|Thesis/dissertation

Molecular Mechanisms for Regulation of Embryonic Stem Cell Lineage Commitment by Core Promoter Factor, TAF3

By

Zhe J. Liu

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Molecular and Cell Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Robert Tjian, Chair
Professor Michael B. Eisen
Professor Richard Harland
Professor Danica Chen

Spring 2012

Abstract

Molecular Mechanisms for Regulation of Embryonic Stem Cell Lineage Commitment by Core Promoter Factor, TAF3

by

Zhe J. Liu

Doctor of Philosophy in Molecular and Cell Biology

University of California, Berkeley

Professor Robert Tjian, Chair

In eukaryotic cells, a key regulatory step in target gene expression is the recognition of the core promoter by the TFIID complex. TFIID is composed of the TATA-box binding protein (TBP) and TBP-Associated Factors (TAFs) (Goodrich and Tjian, 2010; Hochheimer and Tjian, 2003; Naar et al., 2001). To understand gene regulation at the core promoter, I have been focusing on one component of the core promoter recognition complex-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) and the sub-nuclear localization of TAF3 serves as another potential mechanism to influence transcription during myogenesis (Yao et al., 2011). Intriguingly, TAF3 recognizes trimethylated histone H3 lysine 4 (H3K4me3) (Vermeulen et al., 2007), which is associated with actively transcribed genes and also with silent developmental genes that are poised for activation upon ES cell differentiation (Mikkelsen et al., 2007).

In our recent studies, surprisingly it was found that TAF3 is highly enriched in ES cells. More importantly, during *in vitro* EB formation, which mimics early embryogenesis, TAF3 protein levels were selectively reduced. However, the levels of other canonical TAFs and TBP remained stable. Functional and gene expression analysis revealed that high levels of TAF3 are dispensable for ES cell self-renewal but are required for endoderm lineage differentiation and to prevent premature specification of neuroectoderm and mesoderm. Genome-wide binding studies by ChIP-seq found that unlike other canonical TFIID subunits such as TBP and TAF1 that bind only promoters, TAF3 targets both promoters and distal enhancer-like sites in ES cells.

Importantly, TAF3 sites were further clustered into four distinct classes based on the context of co-localization with other factors. Briefly, as well as TAF3, class 1 regions are promoter regions enriched for TFIID, Pol II and H3K4me3. Their proximity (< 1kb) to the TSS of known genes confirms that they are predominantly TFIID-bound promoters. By contrast, Class 2 regions have low levels of TFIID and H3K4me3 but are enriched for Oct4/Nanog/Sox2 and mediator components. Class 3 regions are specifically enriched for TAF3, CTCF and cohesin subunits. Finally, class 4 regions are not enriched for any of the factors we considered besides TAF3. Regions from the last three Classes are generally far away (~10's kb) from the TSS.

Further studies revealed that genes down-regulated upon TAF3 depletion are selectively associated with regions enriched for TAF3, CTCF and cohesin (Class 3) and TAF3 only (Class 4), while no such association was seen for genes up-regulated upon TAF3 depletion. Thus, it's likely that class 3 and class 4 regions are required for the efficient expression of genes in their vicinity and depletion of TAF3 interferes with this function. Indeed, our parallel biochemical experiments confirmed a direct protein-protein interaction between that TAF3 and CTCF. This novel interaction was further shown to mediate DNA looping between promoter distal sites and core promoters to regulate proper transcription activation. Notably, we also found that the class 2 binding regions are enriched around up-regulated genes upon TAF3 depletion, consistent with a mechanism of transcription repression by TAF3 in association with Oct4/Nanog/Sox2. Together, our data support the model that TAF3 orchestrates a complex long-distance chromatin interaction network that safeguards the finely-balanced transcriptional programs underlying ES pluripotency.

This dissertation is dedicated to my parents

Manli Qiu and *Shilin Liu*

For their endless support and encouragement through my life

Table of Contents

List of Abbreviations

ANTP	*drosophila* Hox gene *Antennapedia*
BAB1/2	*drosophila* Bric à brac 1/2
CTCF	CCCTC-binding factor
ChIP	chromatin imunoprecipitation
DAPI	4',6-diamidino-2-phenylindole
EB	Embryoid Body
ES Cell	Embryonic Stem Cell
GO	Gene Ontology
HFD	Histone Fold Domain
H3K4me3	Histone 3 lysine 4 Trimethylation
H3K27me3	Histone 3 lysine 27 Trimethylation
IP	Imunoprecipitation
K/D	Knockdown
MEF	Mouse embryonic fibroblast
MD	Mega Dalton
Pol II	RNA polymerase II
qPCR	quantitative Polymerase Chain Reaction
SAGE	Serial Analysis of Gene Expression
TAF3	TBP Associated Factor 3
TBP	TATA-Box Bind Protein
TFIIA	Transcription factor II A
TFIID	Transcription factor II D
TSS	Transcription Start Site
TRF2	TBP-Related Factor 2

TRF3	TBP-Related Factor 3
PHD	Plant Homeo Domain
PRC	Polycomb Repressive Complex

Acknowledgements

I would like to thank my Ph.D. advisor-Dr. Tjian for his great encouragement and tremendous support through my graduate years. His mind-opening guidance and impeccable taste in science always inspired me to think deeper and reach higher. He taught me to be a positive forward thinker. He taught me to be daring and creative but still stay critical and objective. He taught me that an independent scientist should be persistent in the pursuit of his own scientific ideas. With this opportunity, I also would like to express my gratitude to every Tjian lab members for their helpful advice and assistance.

I would like to thank my collaborators-Drs. Devin Scannell and Michael Eisen for their bioinformatics support in the TAF3 project. Thanks to their help, not only the project has been greatly advanced but also I have learned many useful genomic data analysis techniques. I also would like to thank my other qualifying exam/thesis committee members, Drs. Richard Harland, Danica Chen, Randy Schekman and Qiang Zhou, for their insightful comments on my research project and Dr. Janet Rossant (Sickkid, CA) for her valuable experimental suggestions.

Curriculum Vitae

Zhe Liu
Ph.D. Candidate

MCB/Tjian lab
Berkeley, CA 94720
Tel: 510-642-4442
zhejliu@gmail.com

EDUCATION

Ph.D. Molecular & Cell Biology, University of California, Berkeley, 2012
Focus area: Transcriptional regulation of stem cell pluripotency; Advisor: R. Tjian
M.S. School of Medicine, Tsinghua University, China, 2007
Focus area: Structural biology & biochemistry Advisor: X. J. Liu and Z. Rao
B. S. School of Life Sciences, Tsinghua University, China, 2005

HONORS

Predoctoral CIRM Fellow, 2010-2012
Outstanding Graduate Student Instructor Award, UC Berkeley, 2010
Tang Fellow, UC Berkeley, 2007-2012
Outstanding Research Thesis in Biomedical Sciences, Tsinghua University, 2007
First-class Scholarship for academic excellence (Nationwide competition), 2006
Outstanding Research Thesis in Biological Sciences, Tsinghua University, 2005
First-class Scholarship for academic excellence (Nationwide competition), 2004
GuojiuMaotai Fellowship, Tsinghua University, 2003
Weilun Fellowship, Tsinghua University, 2002
1st Prize of the 19th Beijing College Physics Contest, 2002

PUBLICATIONS (Peer-reviewed)

- **Liu, Z.***, Scannell, D.R.*, Eisen, M.B., and Tjian, R. (2011). Control of embryonic stem cell lineage commitment by core promoter factor, TAF3. **Cell** *146*, 720-731. **(Cover story)**
- Meng, Z., **Liu, Z.**, Lou, Z., Gong, X., Cao, Y., Bartlam, M., Zhang, K., and Rao, Z. (2009). Purification, characterization and crystallization of pyrroline-5-carboxylate reductase from the hyperthermophilic archeon Sulfolobus Solfataricus. **Protein Expr Purif** *64*, 125-130.
- Meng, Z., Lou, Z., **Liu, Z.**, Li, M., Zhao, X., Bartlam, M., and Rao, Z. (2006b). Crystal structure of human pyrroline-5-carboxylate reductase**. J Mol Biol** *359*, 1364-1377.
- Meng, Z., Lou, Z., **Liu, Z.**, Hui, D., Bartlam, M., and Rao, Z. (2006a). Purification, characterization, and crystallization of human pyrroline-5-carboxylate reductase. **Protein Expr Purif** *49*, 83-87.
- **Liu, Z.***, Lin, H.*, Ye, S., Liu, Q.Y., Meng, Z., Zhang, C.M., Xia, Y., Margoliash, E., Rao, Z., and Liu, X.J. (2006). Remarkably high activities of testicular cytochrome c in destroying reactive oxygen species and in triggering apoptosis. **Proc Natl Acad Sci U S A** *103*, 8965-8970.

(*indicate equal contribution)

PRESENTATIONS

03/16/2012 Junior Fellow Interview Symposium, Janelia Farm Research Campus (Oral Presentation)

03/06/2012 Berkeley Academic Exchange Forum, UC Berkeley (Invited presentation)

10/13/2011 Illumina Sequencing Seminar, UCSF Mission Bay (Invited presentation)

09/15/2011 CIRM Grantee Meeting, PARC 55, San Francisco (Oral presentation & Poster)

09/07/2011 Transcription Imaging Consortium Meeting, Janelia Farm Research Campus (Oral presentation)

09/01/2011 Mechanisms of Eukaryotic Transcription Meeting, Cold Spring Harbor Laboratory (Poster)

ACADEMIC RESPONSIBILITIES

2011-2012 UC Berkeley Stem Cell Faculty Search Committee (Student representative)

Chapter 1 Introduction

1.1 Transcriptional Regulation in Eukaryotic Cells

Regulated gene expression at the right time and at the right promoter is a prerequisite to executing the correct programs of development. A key regulatory step in gene expression is transcription initiation. In eukaryotic cells, the transcription initiation from the Pol II transcribed genes is orchestrated by a set of general transcription machineries, which include the core promoter recognition complexes and a number of co-regulator complexes (Figure 1.1.1) (Goodrich and Tjian, 2010; Hochheimer and Tjian, 2003; Naar et al., 2001; Tjian and Maniatis, 1994).

Before the start of transcription, the Transcription Factor II D (TFIID) complex, binds to the TATA box in the core promoter of the gene. The recruitment of TFIID to the core promoter promotes the binding of other general transcription factors (TFIIA-F), the formation of open promoter complex and eventually the RNA polymerase loading and firing. The TFIID complex comprises of the TATA-box Binding Protein (TBP) and 12-15 TBP Associated Factors (TAFs) with distinct functions (Figure 1.1.2) (the exact number of TAFs is species-dependent) (Goodrich and Tjian, 2010; Smale and Kadonaga, 2003). Some TAFs have been shown to be important for the structural assembly and stability of the TFIID complex (Tatarakis et al., 2008; Wright et al., 2006). Several other TAFs were reported to act as dock sites for site-specific transcription factors and thus function to integrate and process upstream transcriptional activation signals (Levine and Tjian, 2003; Liu et al., 2009).

Mediator is a multiprotein complex that functions as a transcriptional coactivator and is conserved in all eukaryotes. The human Mediator complex has 26 subunits and is 1.2 MD in size. Its large surface area provides great potential for protein–protein interaction, even though its sequences do not contain many predicted functional domains. Mediator has been shown to associate with general transcription factors, as well as RNA polymerase II, and is essential for activator-dependent transcription, implying that it may provide a fundamental control of the formation of the initiation complex (Naar et al., 2001; Taatjes et al., 2004). It has also been demonstrated that the Mediator complex is involved in transcription elongation (Takahashi et al., 2011).

In eukaryotic cells, chromosomal DNA is wrapped around histones and thus packaged to form nucleosomes. The packaging of DNA into chromatin fiber condenses and organizes the genome, but obscure many regulatory DNA elements. For transcription initiation to occur, general transcription factors need to gain access to the core promoter region. To enable such accessibility and to exchange nucleosome composition in chromosomal regions, cells have evolved a set of specialized chromatin remodeling complexes (remodelers), which harness the energy of ATP hydrolysis to move and restructure nucleosomes (Clapier and Cairns, 2009).

Epigenetic modifications on the histone tails and DNA comprise of another important layer of transcriptional regulation in eukaryotes. Histone modifications such as acetylation, methylation, phosphorylation, ubiquitination and proline isomerization have been implied in positive and negative regulation of transcription in a context dependent manner (Zhou et al., 2011). DNA methylation however has been generally implied in long-term gene silence and epigenetic inheritance (Robertson, 2005).

Interestingly, in metazoan animals, an adding complexity of transcriptional regulation is that many cell-type specific enhancer-binding factors influences the transcriptional activity of the core promoter machinery from short distances (<1kb) and also from very distant genomic loci (1 to 10Mb). Although it has been suggested that these site-specific transcription factors function by recruiting co-regulators to their target sites and thus form functionally diverse enhancesomes, the molecular basis of the long-range enhancer and core promoter communications remains elusive.

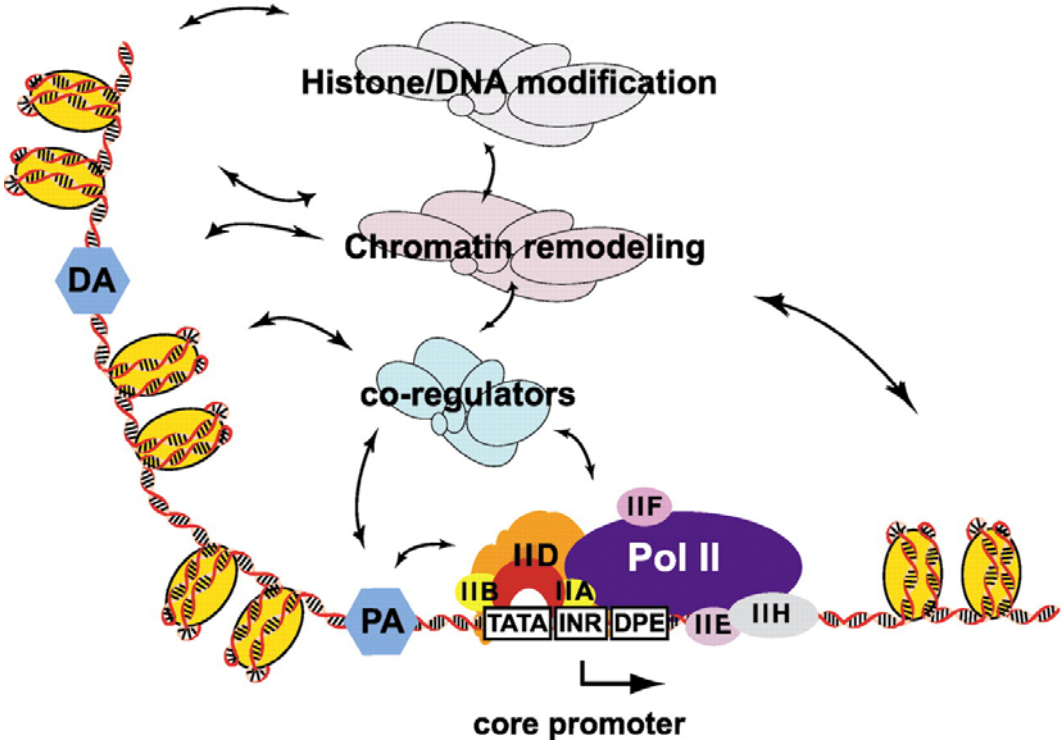


Figure 1.1.1 Eukaryotic transcription initiation is regulated by a multilayered ensemble of multi-subunit complexes. Activated transcription is controlled by co-regulatory coactivator complexes that mediate activation of transcription by activators located in a promoter-proximal position (PA) or many kilobases away in a promoter-distal (DA) position. Chromatin remodelers regulate the assessbility of DNA elements for transcription factor binding. Histone/DNA modification complexes further contribute to epigenetic transcriptional activation and silencing (adapted from(Hochheimer and Tjian, 2003)).

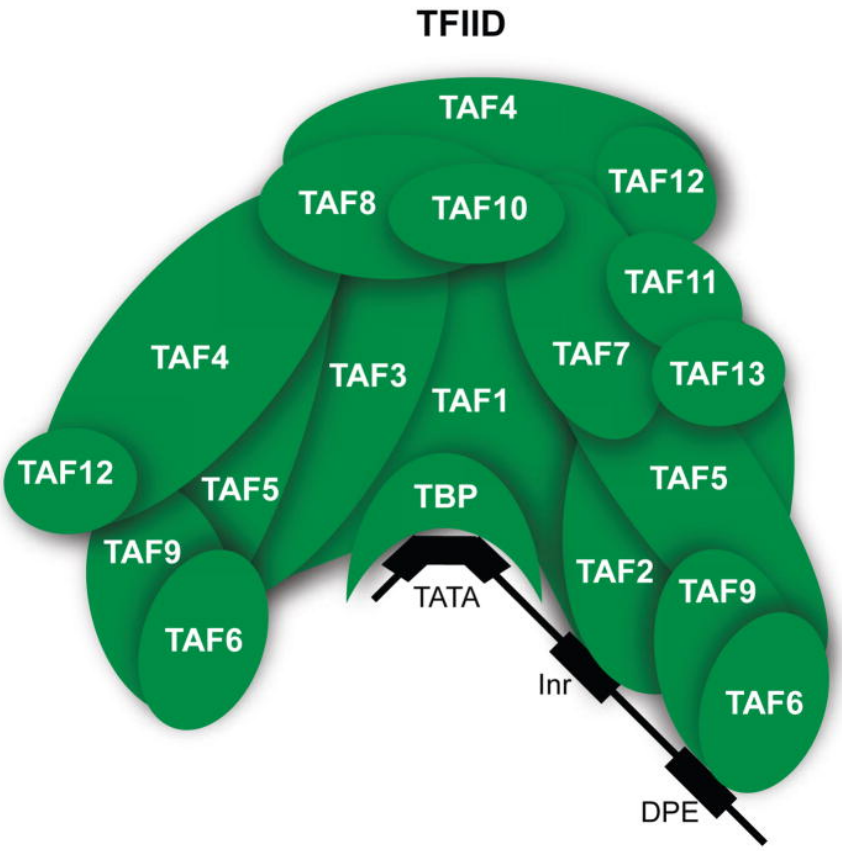


Figure 1.1.2 Schematic geometrical organization of TFIID subunits. TBP, TAF2 and TAF6 bind to TATA-Box (TATA); initiator (Inr) and downstream promoter element (DPE) separately; (adapted from (Goodrich and Tjian, 2010)).

1.2 Diversified Transcription Initiation Complexes

Although until recently the components of the general transcription machineries were thought to be largely invariant among different cell types within an organism, an increasing number of cell type-specific and gene-selective homologs of basal transcription factors have been identified in metazoan organisms, including additional members of the TBP family such as TBP-related factors (TRFs) as well as numerous tissue specific homologs of TAFs (Figure 1.2.1) (D'Alessio et al., 2009; Goodrich and Tjian, 2010; Hochheimer and Tjian, 2003). Functional and developmental analyses suggested that those gene selective general transcription factors are necessary for proper animal development and are crucial for the expression of cell-type specific genes. For example, in mammals, TBP-related factor 2 (TRF2) forms a stable complex with TFIIA and is involved in the transcription of spermiogenesis genes (Zhang et al., 2001). TBP-related factor 3 (TRF3) is shown to form a complex with TAF3 and the TRF3/TAF3 complex is required for myogenesis (Deato et al., 2008). The TRF3/TAF3 complex is also involved in haematopoiesis in zebrafish (Hart et al., 2009). TAF4b, a homolog of TAF4, is highly up-regulated in oocytes and is necessary for the ovarian development (Freiman et al., 2001). Therefore, to understand animal development, it is crucial for us to study how general transcription factors confer gene expression selectivity in different cell types.

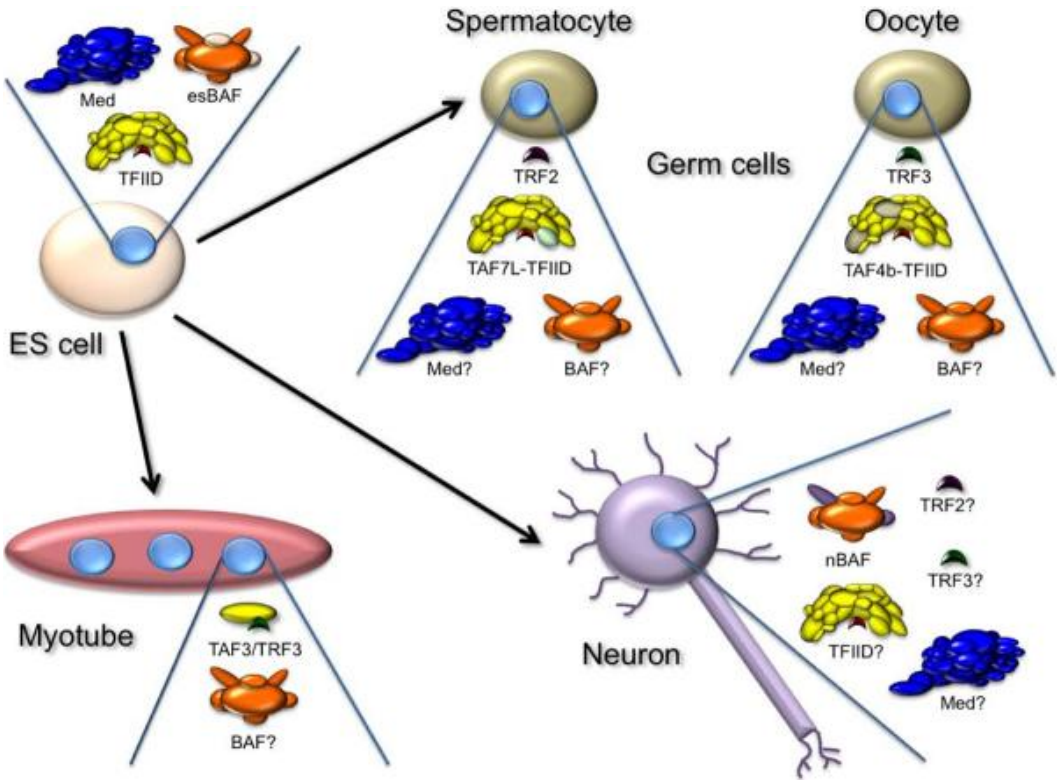


Figure 1.2.1 Summary of tissue-specific functions of core promoter recognition and co-regulatory complexes (adapted from (D'Alessio et al., 2009)).

1.3 TATA binding protein associated factor 3

TATA binding protein associated factor 3-TAF3 was originally identified as a subunit of the TFIID complex in Hela cells (Gangloff et al., 2001). TAF3 contains a N-terminal Histone Fold Domain (HFD) and a C-terminal PHD finger domain. The HFD of TAF3 binds TAF10, a TFIID core structural component. Recently, it was found that, upon progressive destruction of the other subunits in the TFIID complex during myogenesis, TAF3 is selectively retained in myotubes and can be found in a specialized complex with TBP-related factor 3, TRF3 (Deato and Tjian, 2007). A similar TRF3/TAF3 complex has been found to play a role during Zebrafish hematopoiesis (Hart et al., 2009).

In vertebrate, the selective gene expression is tightly regulated by two antagonizing family of histone modification complexes, the *Trithorax* complexes and the *Polycomb* complexes (Schuettengruber et al., 2007). The vertebrate *Trithorax* family is a set of Hisone 3 Lysine 4 trimethylase complexes that tri-methylate the histone 3 lysine 4 around TSS and positively regulate target gene expression. Polycomb Repression Complex 2 (PRC2) is a multi-subunit complex harboring the Hisone 3 Lysine 27 (H3K27) trimethylase activity. The trimethylation of H3K27 recruits Polycomb Repression Complex 1 (PRC1) to the target genes and repress their expression (Boyer et al., 2006; Morey and Helin, 2010). Most interestingly, those two antagonizing activities are not mutually exclusive even at the same gene locus. In ES cells, a large number of key developmental genes are marked by both H3K4me3 and H3K27me3, a “bivalent” mark (Meshorer and Misteli, 2006). The PHD finger of TAF3 has been shown to recognize trimethylated Histone H3 Lysine 4 (H3K4me3) (Vermeulen et al., 2007). Thus, it seems possible that TAF3, either as a subunit of TFIID or in association with other potential partners (TRF3) may play an active role in regulating transcription of developmental genes by targeting cell-type specific complexes to gene promoters including those that are marked by H3K4me3.

1.4 Evolutionary Aspects of TAF3

In eukaryotes, TFIID orchestrates Pol II Pre-Initiation Complex (PIC) assembly. Most of the TAFs are evolutionarily conserved across eukaryotic kingdoms. In exception, there are no TAF3 homologs identified in plants (Table 1.4.1). In single cell eukaryotes, TAF3 proteins are comparably small (*S. pombe*, 150aa, *S. cerevisiae*, 355aa) and only contain an identifiable N-terminal Histone Fold Domain (HFD). Although the HFD is conserved across different species, a Plant Homeo Domain (PHD) finger is uniquely present at the C-terminus of drosophila and vertebrate TAF3(s) (Figure 1.4.1). Intriguingly, except the HFD and the PHD, no significant sequence homologies were detectable in the other regions between drosophila TAF3 and vertebrate TAF3(s), suggesting these regions of TAF3 might be important for species-specific molecular regulations. Indeed, biochemical and genetic analyses have shown that regions in middle part of *drosophila* TAF3 interact with *drosophila*-specific transcription factors (BAB1/2, GAGA and ANTP) and regulate proper *drosophila* development(Chopra et al., 2008; Pointud et al., 2001; Prince et al., 2008). However, it is still unknown whether the similar pairs of interaction exist in vertebrate and whether TAF3 is crucial for the regulation of vertebrate development programs. The middle part of vertebrate TAF3(s) is quite conserved (Figure 1.4.1). Thus, it is interesting to further investigate whether and how these vertebrate-specific regions might contribute to transcription.

New name	*H. sapiens*	*D. melanogaster*	*S. cerevisiae*	Arabidopsis *thaliana*
TAF1	$TAF_{II}250$	$TAF_{II}230$	TAF145/130	AtTAF1
				AtTAF1b
TAF2	$TAF_{II}150$	$TAF_{II}150$	TAF150	AtTAF2
TAF3	$TAF_{II}140$	$TAF_{II}155$	TAF47	NI
TAF4	$TAF_{II}130/135/105$	$TAF_{II}110$	TAF48	AtTAF4
				AtTAF4b
TAF5	$TAF_{II}100$	$TAF_{II}80$	TAF90	AtTAF5
TAF5L	PCAF65ö	Cannobal		NI
TAF6	$TAF_{II}80$	$TAF_{II}60$	TAF60	AtTAF6
				AtTAF6b
TAF6L	TAF65ő	TAF60-2		NI
TAF7	$TAF_{II}55$	$TAF_{II}55$	TAF67	AtTAF7
TAF7L	TAF2Q			NI
TAF8	$TAF_{II}45/50$	TAF8	TAF65	AtTAF8
TAF9	$TAF_{II}32/31$	$TAF_{II}40$	TAF17	AtTAF9
TAF9L	$TAF_{II}31L$			
TAF10	$TAF_{II}30$	$TAF_{II}24$	TAF25	AtTAF10
TAF10b		$TAF_{II}16$		NI
TAF11	$TAF_{II}28$	$TAF_{II}30ö$	TAF40	AtTAF11
				AtTAF11b
TAF12	$TAF_{II}20/15$	$TAF_{II}30ő$	TAF61/68	AtTAF12
				AtTAF12b
TAF13	$TAF_{II}18$	TAF18	TAF19	AtTAF13
TAF14			TAF30	AtTAF14*
				AtTAF14b*
TAF15	$TAF_{II}68$			AtTAF15*
				AtTAF15b*

Table 1.4.1 this shows the homolog of TAFs across different species. NI denotes “Not Identified” (Adapted from (Lago et al., 2004))

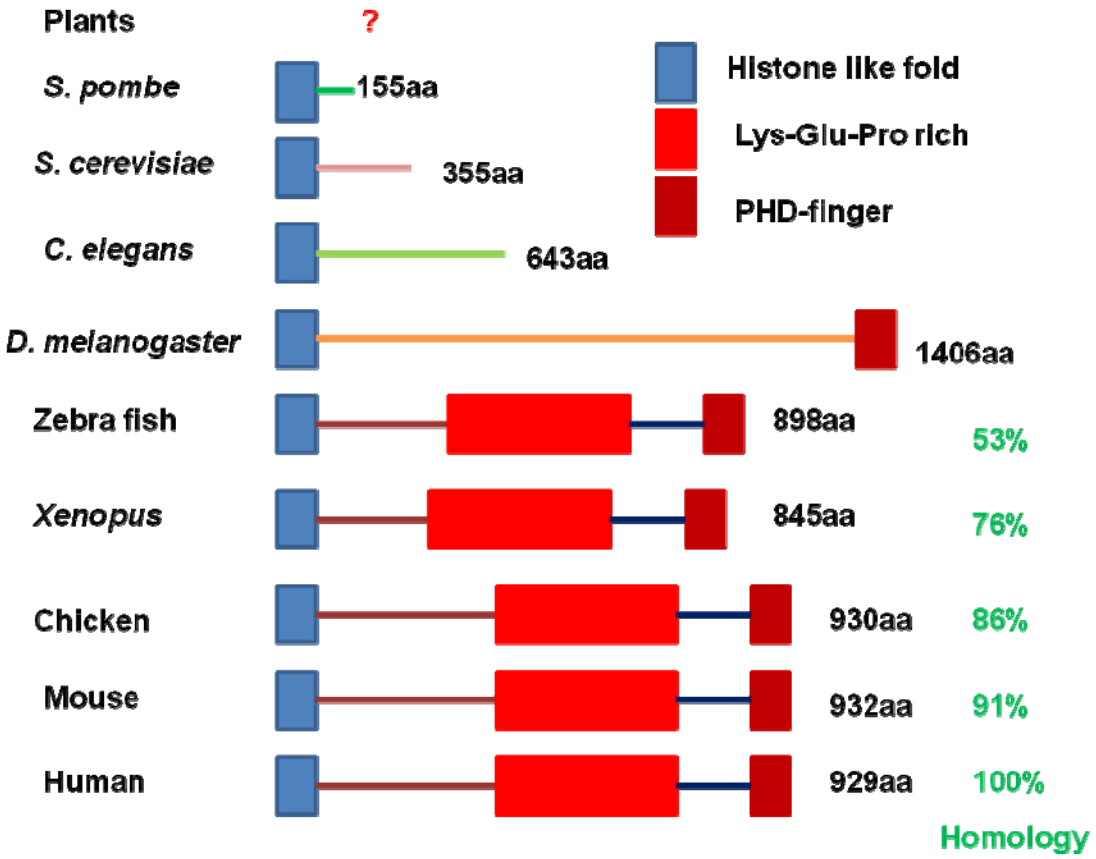


Figure 1.4.1, TAF3 domain structure across different species. Protein sequence homologies were calculated against human TAF3

Chapter 2 Initial Characterization of TAF3 in ES cells

2.1 TAF3 Antibody Generation and Validation

3 X Flag tagged Full length of TAF3 protein was produced in the baculovirus expression system using SF-9 cells and was purified by anti-Flag resin to homogeneity (Figure 2.1.1). Three guinea pigs were immunized with this protein. After initial screening, it was found that one anti-TAF3 serum was able to detect over-expressed TAF3 in 293T cells (Figure 2.1.2, lane 1 and 7). Correspondingly, a band at the same molecular weight was observed in other cell lines (with highest levels in ES cell line D3) (Figure 2.1.2).

To further validate the specificity of this antibody, TAF3 protein were first immune-precipitated from either 293T cells over-expressing 3 X Flag mouse TAF3 or ES cells by the T7 rabbit antibody against TAF3. Then, the IP products were resolved on gel and were blotted with the guinea pig anti-TAF3 antibody (Figure 2.1.3). As we can see, TAF3 protein was selectively enriched by the anti-TAF3 IPs and can be detected by the guinea pig antibody. These experiments which applied a double selection on TAF3 detection further confirmed the specificity of this guinea pig antibody. The anti TAF3 serum was further purified through affinity column conjugated with 3 X Flag tagged full-length TAF3 protein. Better specificity was achieved as evident from the comparison between Figure 2.1.2 and Figure 2.1.4.

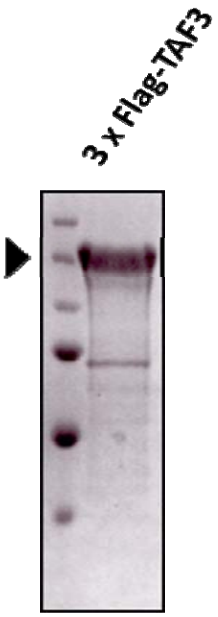


Figure 2.1.1, 3 X Flag-tagged full length TAF3 protein for immunization (Page-blue Protein Staining)

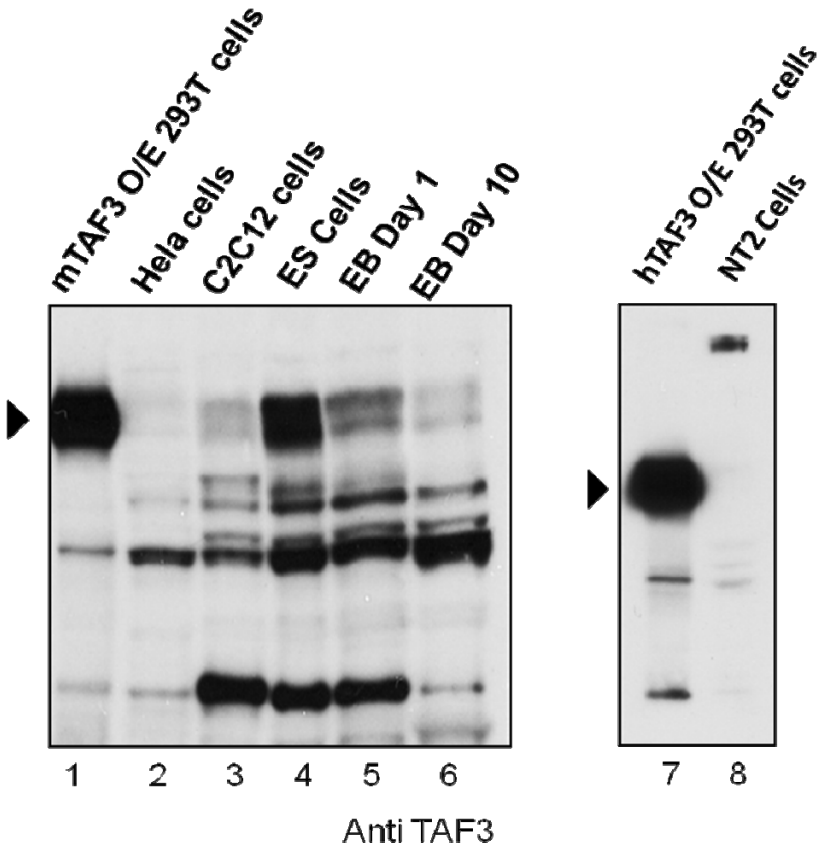


Figure 2.1.2 Western blots with the Guinea pig anti-TAF3 antibody. Lane 1, 3 X Flag tagged mouse TAF3 over-expression in 293T cells; Lane 2, HeLa cells; Lane 3, C2C12 cells; Lane 4, ES cells (D3); Lane 5, embryoid body day 1; Lane 6, embryoid body day 10; Lane7, 293T cells over-expressing 3 X Flag tagged human TAF3; Lane 8, NT2 cells.

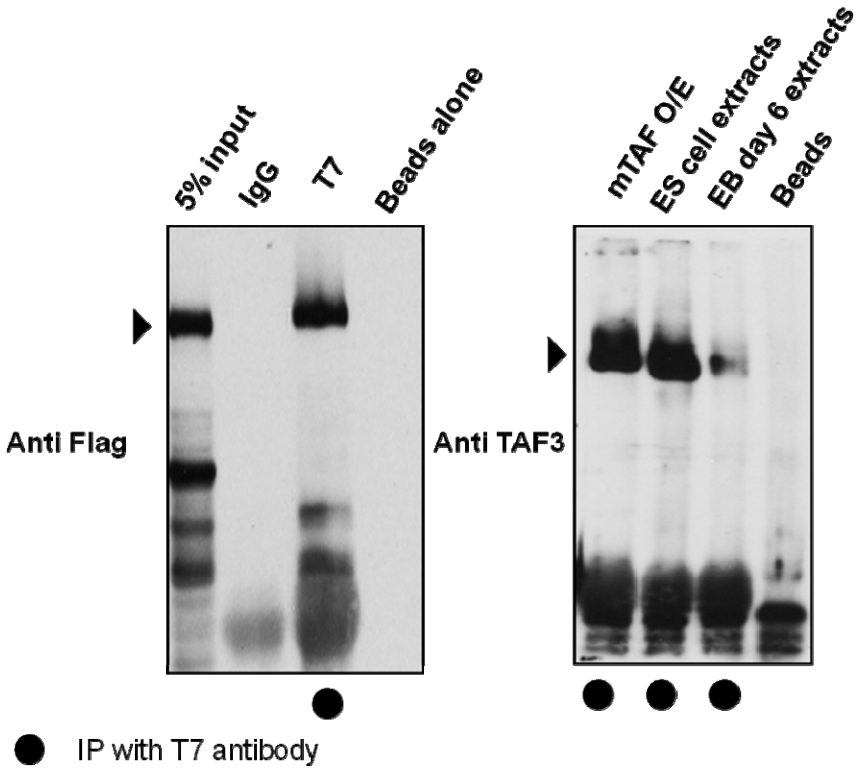


Figure 2.1.3 Validation of the Antibody specificity. IP experiments were performed with T7 TAF3 antibody (rabbit) using different cell lines as indicated above. IP products were blotted with the Guinea pig anti-TAF3 antibody

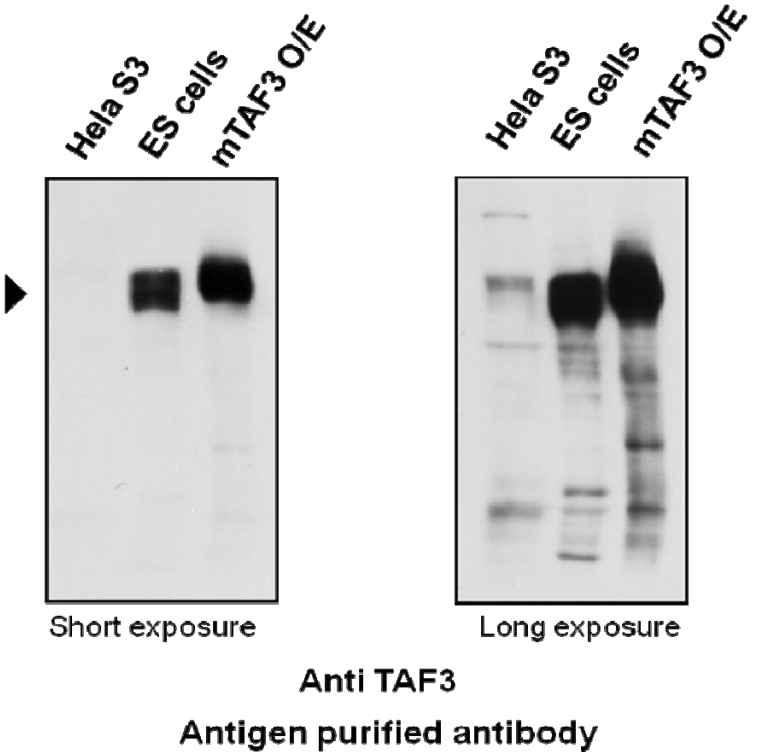


Figure 2.1.4 Western Blot Analysis with antigen purified anti-TAF3 antibody. Left panel, short exposure; Right panel, long exposure.

2.2 High Levels of TAF3 in ES cells

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). 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 (~10X relative to C2C12's) in mouse ES cells (Figure 2.2.1).

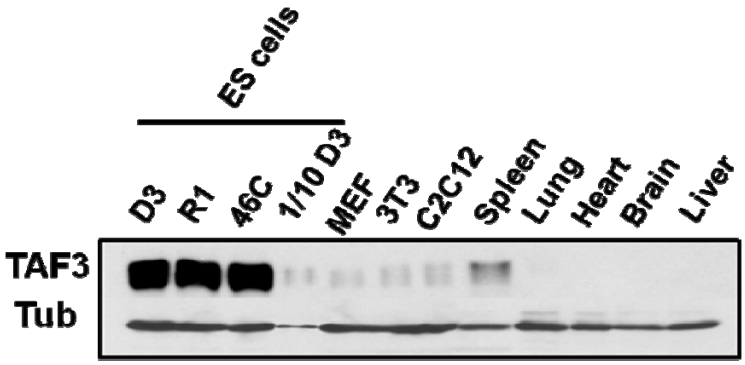


Figure 2.2.1, Western blots to examine TAF3 levels in different cell types and tissues.

2.3 TRF3/TAF3 Complex is Absent in ES cells.

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 2.3.1). Consistent with this, published ES cell genome-wide ChIP-seq studies found no Pol II or H3K4me3 enrichment around the promoter region of Trf3 (Mikkelsen et al., 2007), suggesting that *Trf3* is largely silent in ES cells. These findings indicate that although TAF3 is highly expressed in ES cells, one of its potential partners (TRF3) is absent, suggesting an independent function of TAF3. In support of this notion, we found that in ES cell nuclear extract, TAF3 protein migrated at a molecular weight >1MD in a superose 6 column (Figure 2.3.2), while the TAF3/TRF3 complex from myoblast/myotubes migrates as a native species of ~180kd (Deato and Tjian, 2007).

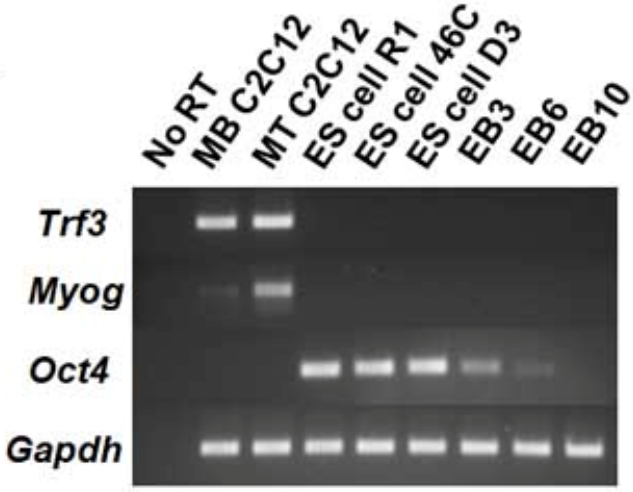


Figure 2.3.1, qRT-PCR to measure mRNA levels of Trf3, Myog, Oct4 and Gapdh in C2C12 MB(myoblast), C2C12 MT(myotube), ES cells (D3, R1 46C) and EBs (D3).

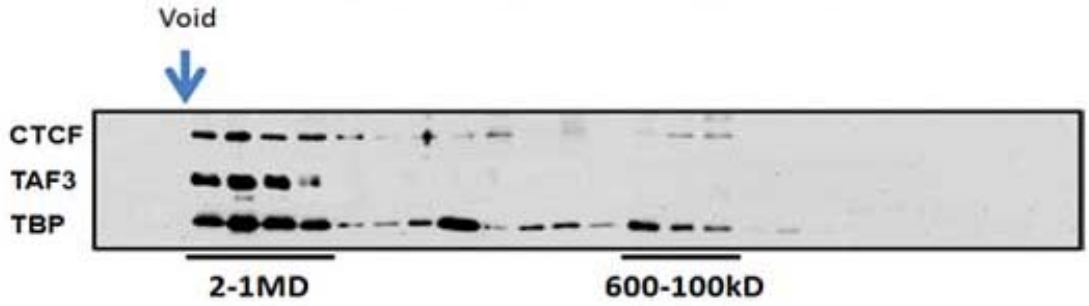
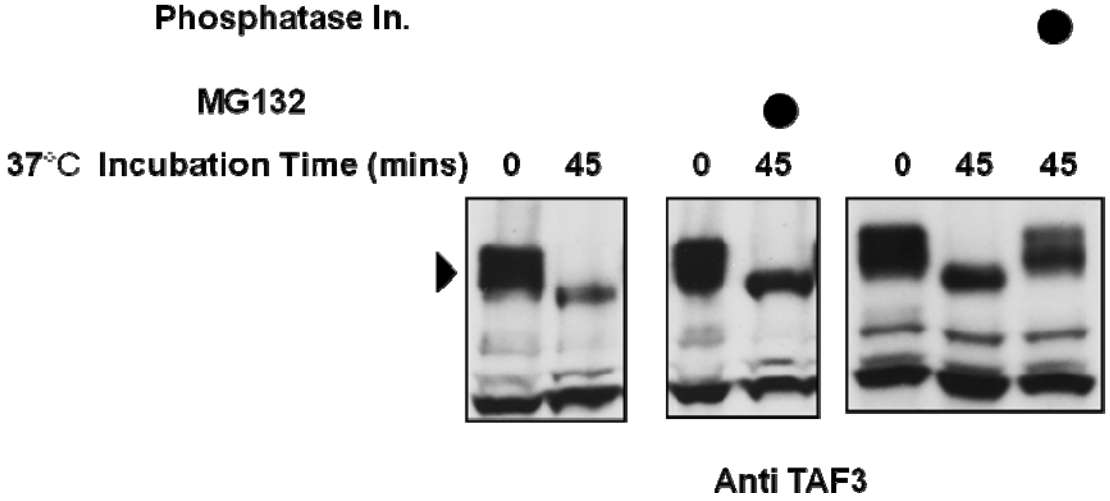


Figure 2.3.2, TAF3 migrated at around 1-2 MD on Superose 6 column in ES cell nuclear extract.

2.4 TAF3 is phosphorylated in the cells.

When ES cell samples were bolted with anti-TAF3 antibody, several migration species of TAF3 protein were observed (Figure 2.1.2), suggesting certain forms of covalent modification. Interestingly, after the ES cell extracts were incubated at 37°C for 45 mins, these higher molecular weight (MW) species became disappeared (Figure 2.4.1, Left panel). Reasoning that those higher MW forms were likely susceptible to either degradation or dephosphorylation, ES cell extracts were incubated with either proteasome inhibitor (MG132) or phophatase inhibitor cocktail. Interestingly, only the phophatase inhibitor cocktail but not the MG132 treatment retained these higher molecular weight species (Figure 2.4.1, middle and right panels). This result suggests that TAF3 protein is phosphorylated in the cells.

Figure 2.4.1, Western blot analysis of ES cell extracts incubated at 37°C for 45mins



with/without MG132 or phosphatase inhibitors. Left panel, control experiment; middle panel, experiment with the MG132; right panel, experiment with the phosphatase inhibitor cocktail.

2.5 TAF3 Phosphorylation States are Cell-cycle Regulated.

It was reported that several components of TFIID complex, such as TBP, TAF1, TAF4, and TAF7, are subject to hyper-phosphorylation during mitosis (Pijnappel et al., 2009). The induction signaling events and biological consequences of such phosphorylation still remain elusive. As a first step towards understanding these phosphorylation modifications, it would be important for us to find out whether the phosphorylation states of TAF3 are cell-cycle regulated and, if so, at which stage of the cell cycle these phosphorylation events occur.

To address these questions, ES cells were first synchronized to mitosis by a 12 hour Nocodazole (50nM) treatment. Then, ES cells were released from mitotic arrest by washing cells with regular culture medium. Subsequently, the phosphorylation states of TAF3 were examined in a time course. As we can see in Figure 2.5.1, TAF3 protein was hyper-phosphorylated after mitotic arrest. However, as the releasing proceeded, hypo-phosphorylated species of TAF3 was quickly increased. These results suggest that TAF3 protein phosphorylation mainly occurs in mitosis while TAF3 might be subjected to dephoshorylation in G1 and S phase.

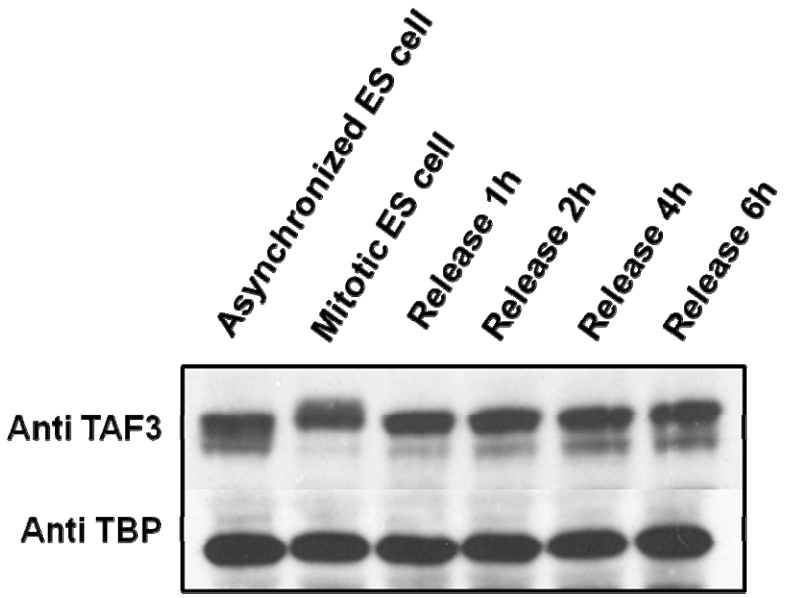


Figure 2.5.1, Western blots to probe the temporal changes in TAF3 phosphorylation after ES cells were released from mitosis

2.6 TAF3 is Associated With Chromatin Tightly.

TAF3 was thought to bind chromatin through its interactions with TFIID and H3K4me3. To test whether TAF3 was differentially associated with chromatin compared to other components of TFIID (such as TBP and TAF4), we extracted total protein from ES cell nuclei with buffers that contain a range of salt and detergent concentrations. We found that, with RIPA buffer, TAF3 was extracted efficiently out of chromatin with other components of the TFIID (TBP and TAF4) (Figure 2.6.1). However, with a buffer that contains lower concentrations of NaCl and NP40, TAF3 is completely retained in the chromatin (precipitant) while decent mount of TBP and TAF4 (around 20-30%) was extracted (Figure 2.6.1). This observation rise possibilities that 1) TAF3 binds to chromatin more tightly than TBP and TAF4 and 2) TAF3 might function distinctly from other TFIID components.

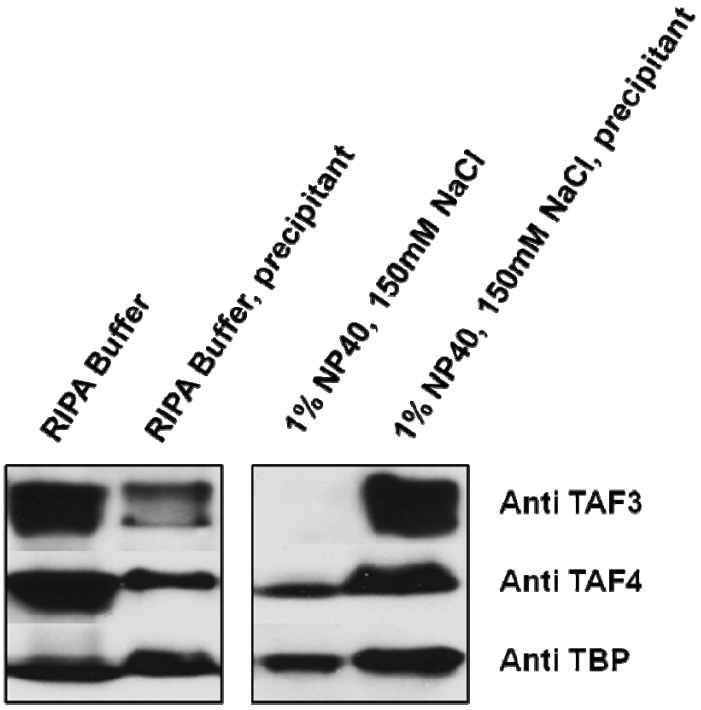


Figure 2.6.1, ES cell nuclei were extracted with different buffers. Specifically, RIPA buffer and the buffer that contains 1% NP40 and 150mM NaCl were used separately in left and right panels. After the extracts were cleared with ultra-centrifugation, the supernatant and precipitant obtained were resolved on the gels and were then blotted with antibodies against TAF3, TAF4 and TBP.

2.7 Selective Reduction of TAF3 Protein Levels during ES cell Differentiation

The remarkably higher levels of TAF3 in ES cells (Figure 2.1.2) suggest a model that TAF3 might be selectively enriched in ES cells and performs a novel function. To further validate whether high levels of TAF3 are ES cell-specific, ES cells were induced to form embryoid bodies by culturing on low-attachment plates without LIF. The levels of TAF3 protein were monitored by western blots in a time course. Interestingly, TAF3 is rapidly depleted upon ES cell differentiation while the levels of other TFIID components (TBP and TAF4) remained largely unchanged (Figure 2.7.1). This tight and differential regulation of TAF3 protein levels during ES cell differentiation indicates that TAF3 might be important for the maintenance the ES cell specific properties, such as self-renewal and pluripotency.

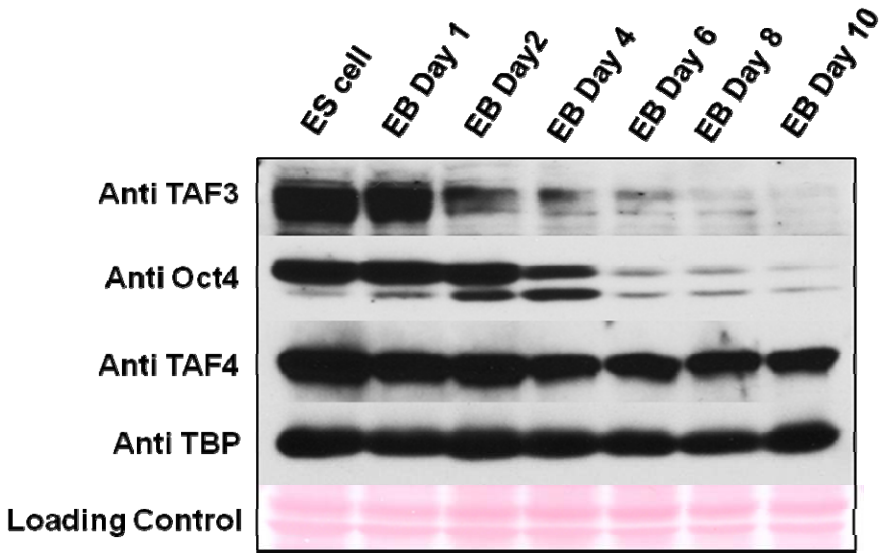


Figure 2.7.1, ES cells (D3) were induced to form embryoid bodies. Samples were collected at different time points as indicated and were further lysed by 1 X SDS-PAGE buffer. Western blots were performed using antibodies against TAF3, Oct4, TAF4 and TBP. Ponceaus S staining served as the loading control.

Chapter 3 Balanced early lineage segregation Requires TAF3

3.1 Generation of Lentivirus-mediated shRNAs Against TAF3

To study whether high levels of TAF3 are required for ES cell self-renewal and pluripotency, we decided to perform loss-of-function experiment. To do this, 8 shRNAs against mouse TAF3 were designed and cloned into PLKO.1 vector. Then, to test the knockdown efficiency, each shRNA was transfected into 293T cells with plasmids expressing 3 X Flag-TAF3 and Flag-Oct4. We found that two shRNAs out of the 8 worked well in this co-transfection assay (Figure 3.1.1). To further examine these two shRNAs in vivo, lentivirus particles were generated and concentrated by a previously described protocol (Moffat et al., 2006). At 24 hours post infection, ES cells were refreshed with medium supplemented with 1ug/ml of puromycin. By 48 hours post selection (72 hours post infection), greater than 90% reduction of TAF3 protein was observed when compared to control ES cells treated with non-target shRNA (Figure 3.1.2).

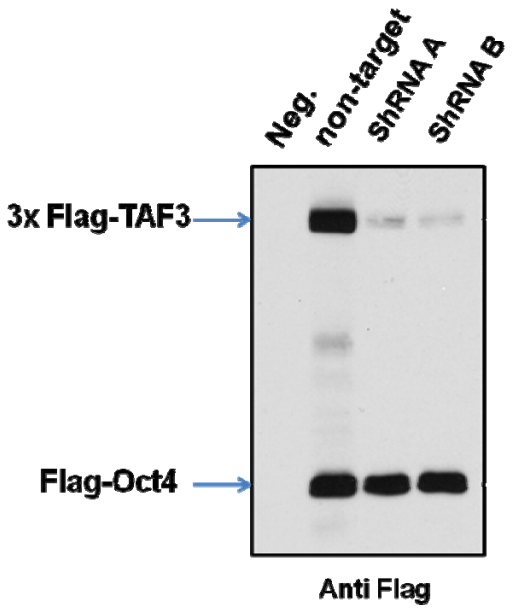


Figure 3.1.1, Individual shRNA vector was transfected into 293T cells with plasmids expressing 3 X Flag-TAF3 and Flag-Oct4. 48hours after transfection, samples were resolved on the gel and anti-Flag western blot was performed to evaluate the knockdown efficiencies by shRNA A and shRNA B. Flag-Oct4 served as the transfection efficiency control.

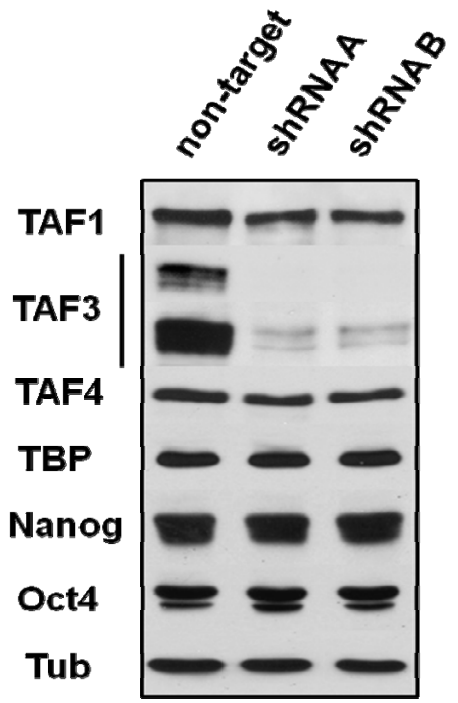


Figure 3.1.2, ES cell samples were harvested 72-hours post infection and were immune-blotted with antibodies against TAF1, TAF3, TAF4, TBP, Nanong, Oct4 and tubulin.

3.2 TAF3 Depletion does not Affect the Expression of Canonical Self-renewal Genes

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 3.2.1, 3.2.3 and 3.2.4). TAF3 K/D cells were also able to form alkaline phosphatase positive colonies as efficiently as control ES cells (Figure 3.2.2) and the percentage of $SSEA1^{+}Oct4^{+}$ cells upon efficient TAF3 K/D remained largely unaltered (Figure 3.2.5). Furthermore, there was little or no correlation between $TAF3^{+}$ and $Nanog^{+}$ cells as determined by immuno-fluorescence of individual ES cells (Figure 3.2.6). 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).

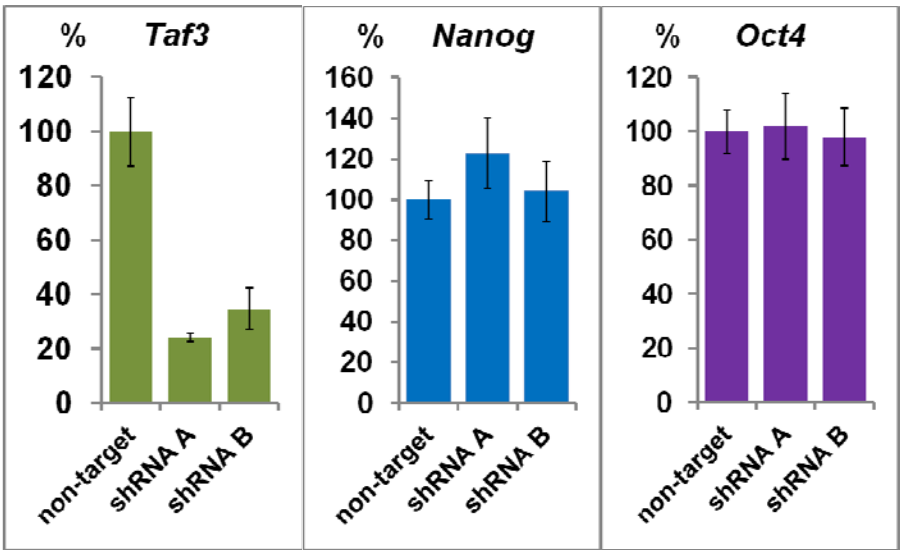


Figure 3.2.1, qRT-PCR to measure the mRNA levels of *Taf3*, *Nanog* and *Oct4* in control and TAF3 K/D cells (D3, P2 post infection).

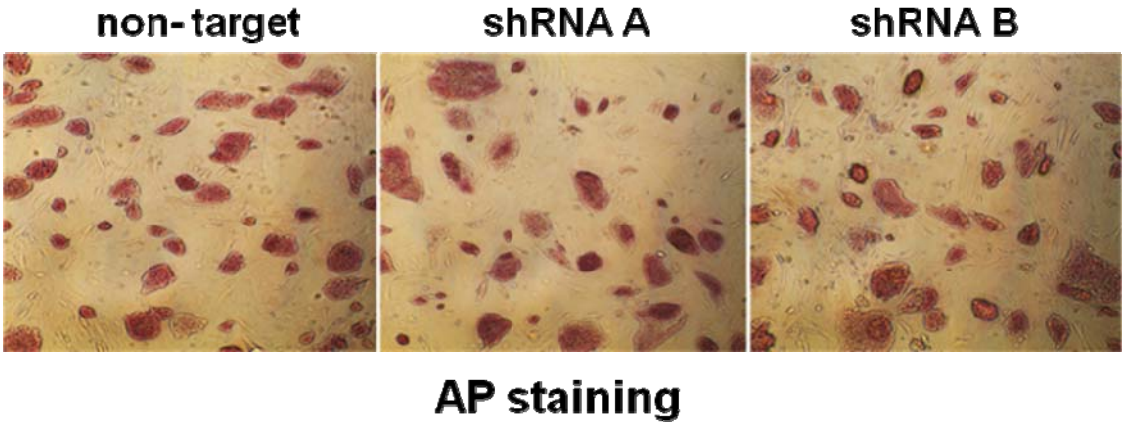


Figure 3.2.2, Alkaline phosphatase staining of control and TAF3 K/D cell colonies (R1, P4 post infection) on feeder cells.

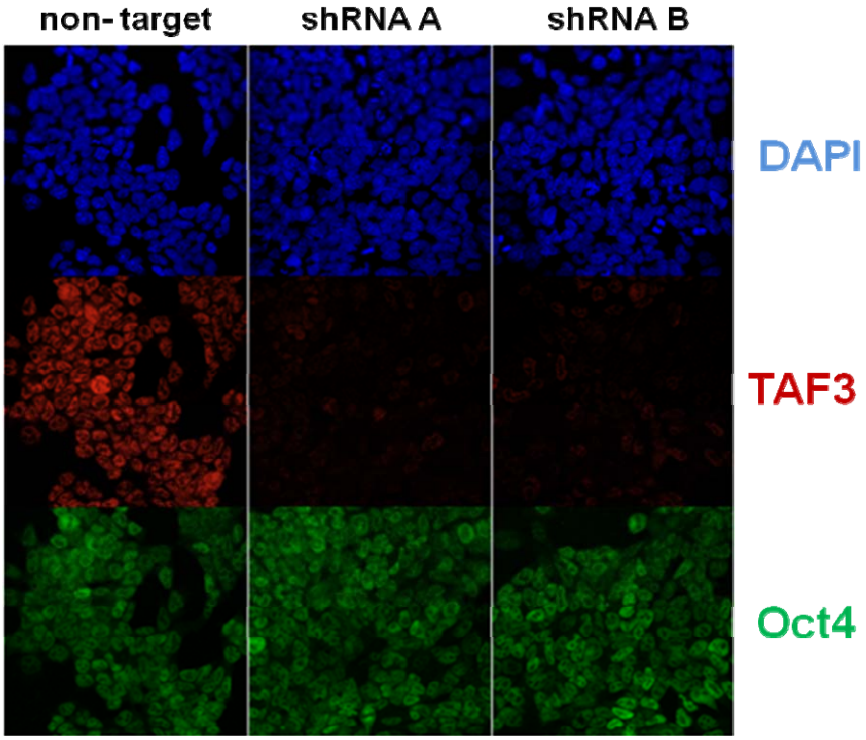


Figure 3.2.3, Confocal fluorescence images of control and TAF3 K/D cells (R1, Passage 4 post infection) co-stained with antibodies against TAF3 (Red) and Oct4 (Green). Nuclei were counterstained with DAPI.

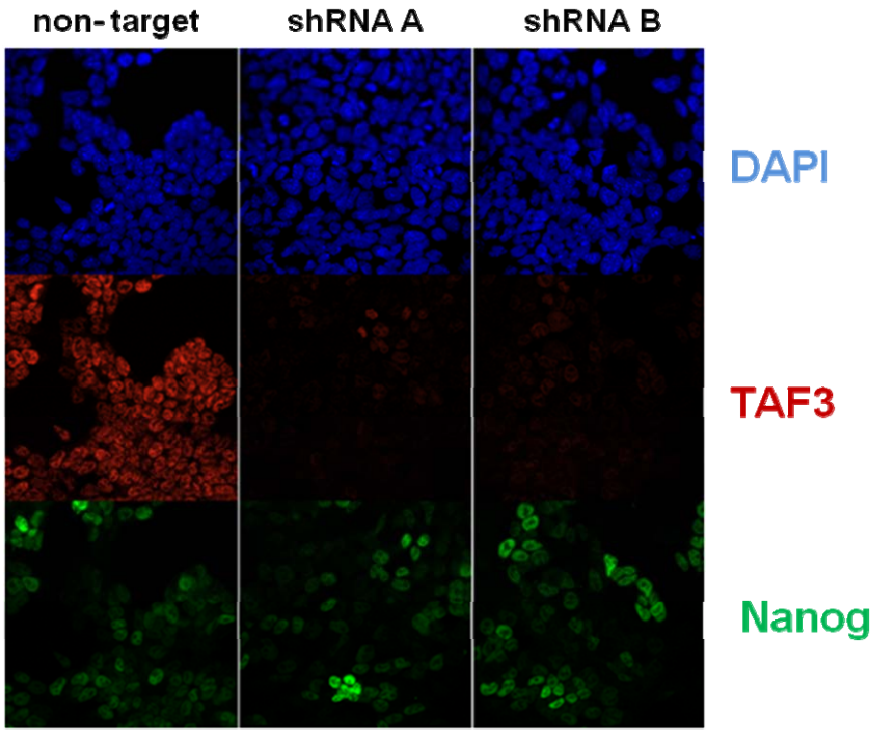


Figure 3.2.4, Confocal fluorescence images of control and TAF3 K/D cells (R1, Passage 4 post infection) co-stained with antibodies against TAF3 (Red) and Nanog (Green). Nuclei were counterstained with DAPI.

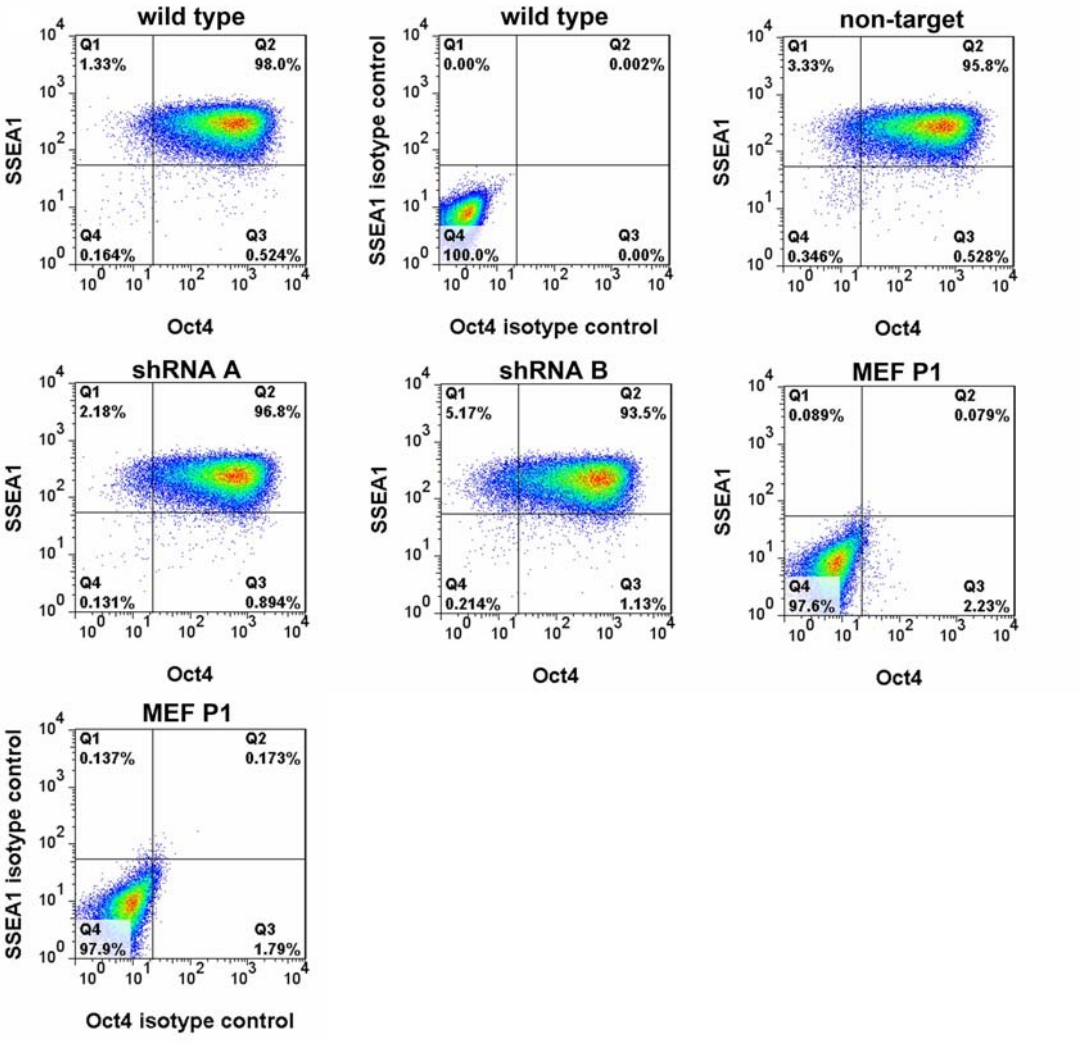


Figure 3.2.5, Flow cytometric analysis of SSEA1^{+}Oct4^{+} cells in wild type, control and TAF3 K/D cell and MEF cell populations. Cells treated with non-specific antibodies of the same isotype served as the negative control.

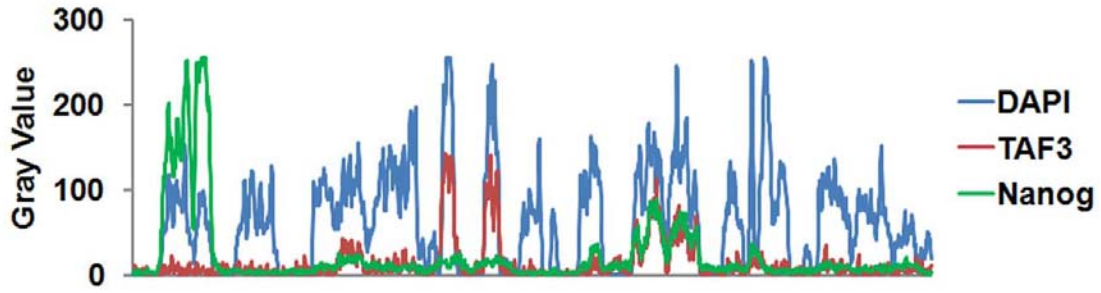


Figure 3.2.6, Single cell expression correlation analysis of TAF3 and Nanog in shRNA A treated cells (R1, P4 post infection). Gray values along a random chosen path in DAPI, TAF3 (shRNA A), and Nanog channels of (G) were plotted in the diagram.

3.3 High Levels of TAF3 are Dispensable for ES cell proliferation

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 3.3.1). 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 3.1.2), 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.

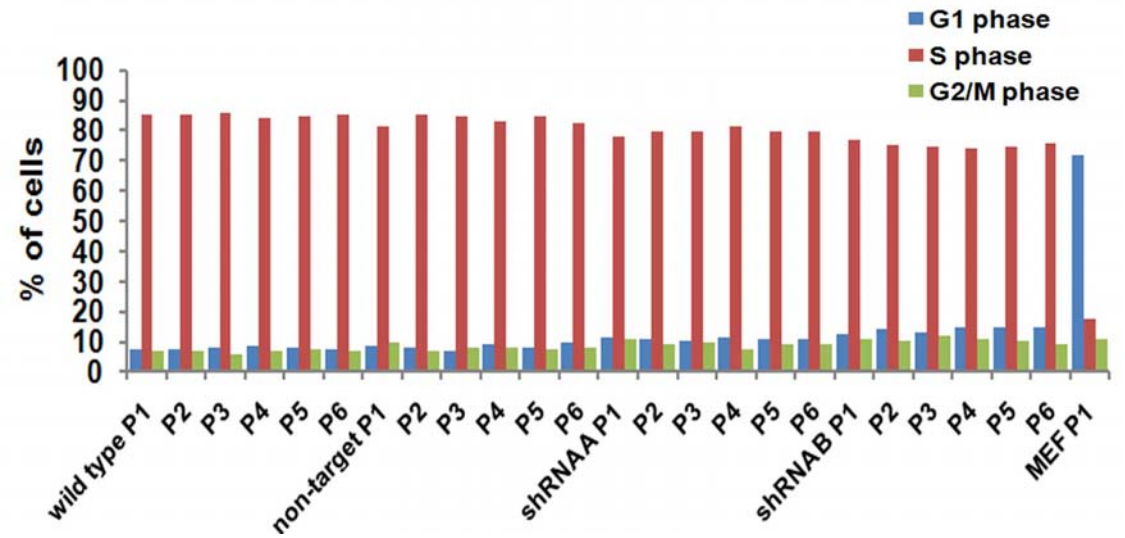


Figure 3.3.1 Percentages of G1, S and G2/M cells in wild type, control and TAF3 K/D cell populations (analyzed for 6 passages post infection).

3.4 Misregulation of Lineage-specific Gene Expression Programs After TAF3 Depletion

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 3.4.3). 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 3.4.1). 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 3.4.4). Likewise, Afp, a late endoderm marker, stained control EBs but not TAF3 K/D EBs (Figure 3.4.5). 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 3.4.6). 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 3.4.7). 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 3.4.8). In conclusion, these data strongly suggest that high levels of TAF3 in ES cells may be essential for proper cell lineage specification during differentiation.

Tight temporal and spatial regulation of transcription is critical for proper embryo development and the expression of developmental genes usually becomes cell type- or germ layer- restricted during ES cell differentiation (Tam and Loebel, 2007). 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 2.7.1) was quite homogenous and apparently not lineage-specific (Figure 3.4.9 and 3.4.10). Together, these results suggest a critical role for TAF3 in directing cell fate choices at very early stages during ES cell differentiation.

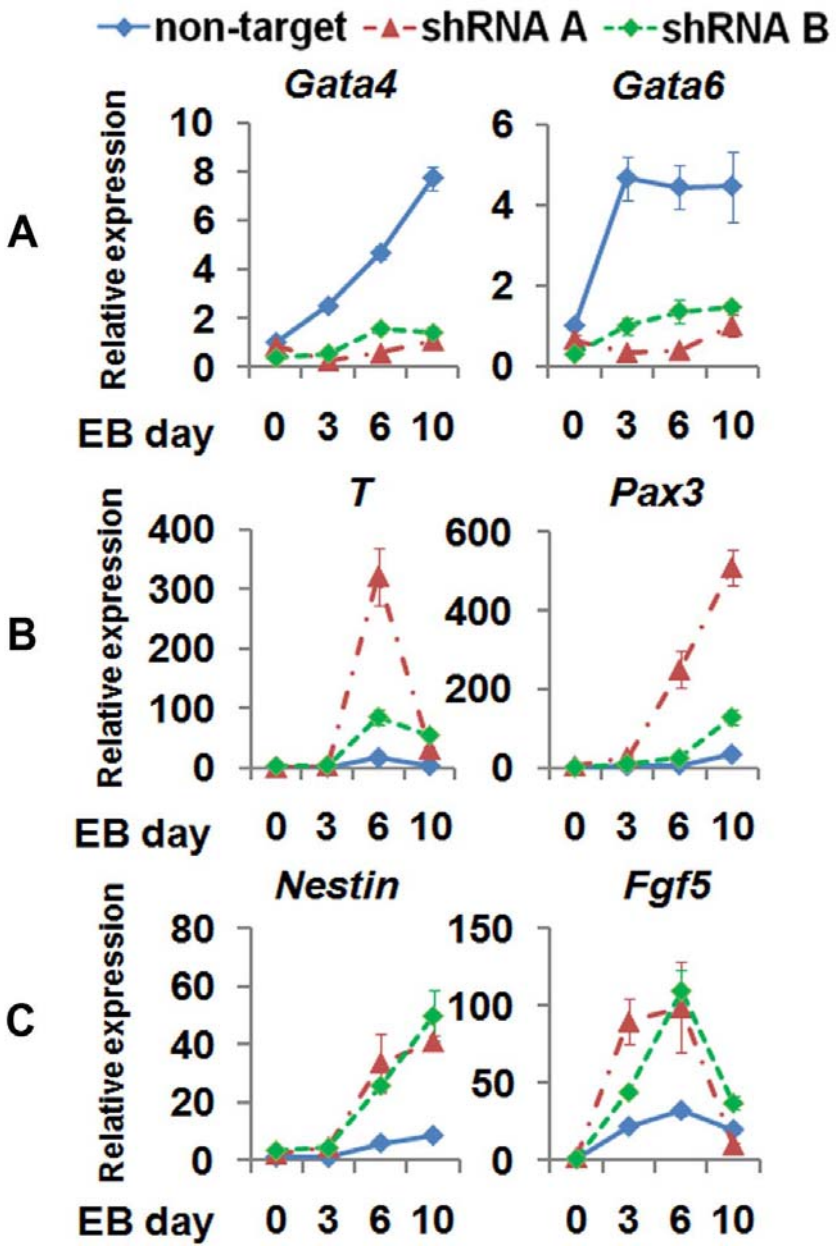


Figure 3.4.1, qRT-PCR to measure the expression of lineage-specific markers in control and TAF3 K/D samples.

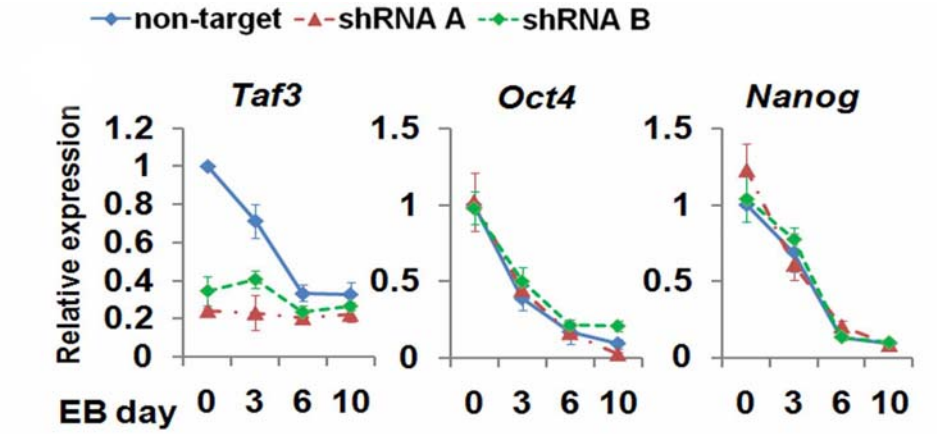


Figure 3.4.2, qRT-PCR analyses of the expression of Taf3, Oct4 and Nanog in control and TAF3 K/D samples.

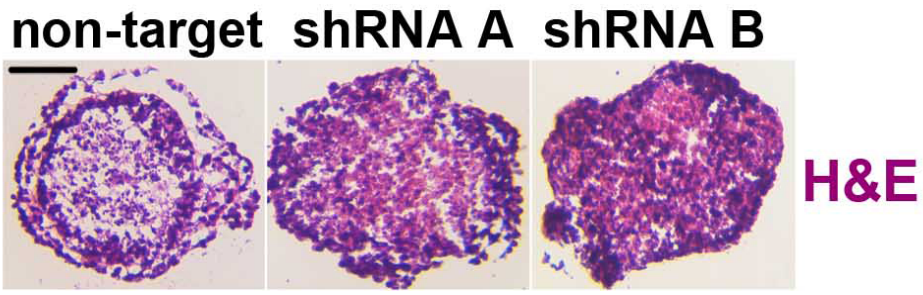


Figure 3.4.3, H&E staining of EB sections from control and TAF3 K/D EBs (day 10). Scale Bars, 50 μm.

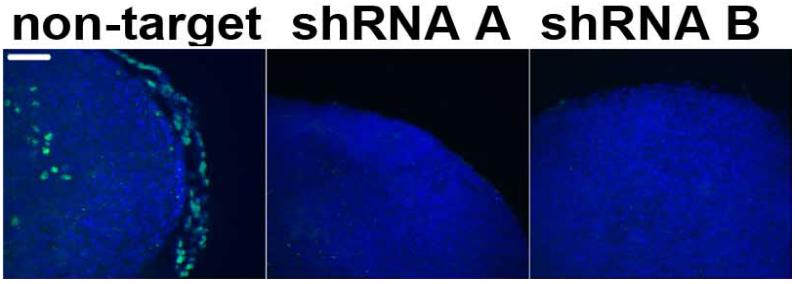


Figure 3.4.4, Confocal fluorescence images of control and TAF3 K/D EBs (day 10) stained with antibody against GATA4 (E) or Afp (F). Scale Bars, 50 μm.

Figure 3.4.5, Confocal fluorescence images of control and TAF3 K/D EBs (day 10) stained with antibody against GATA4 (E) or Afp (F). Scale Bars, 50 μm.

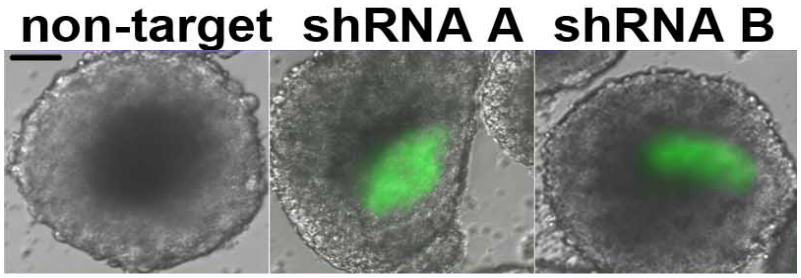


Figure 3.4.6, Live-cell fluorescence images of control and TAF3 K/D EBs (day 8, 46C). Scale Bars, 50 μm.

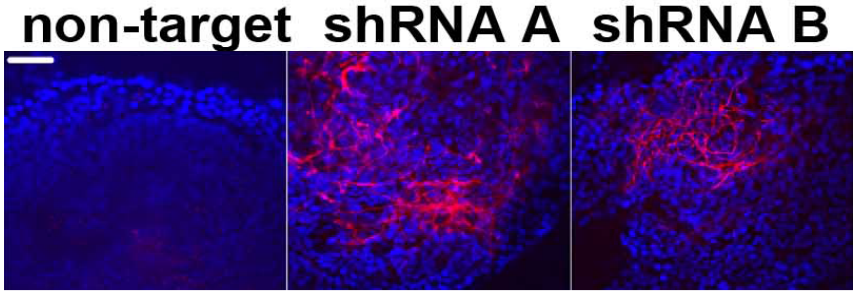


Figure 3.4.7, Confocal fluorescence images of control and TAF3 K/D EBs stained with antibody against Tubb3. EBs (day 4) were cultured on Laminin coated slides for 6 additional days. Scale Bars, 50 μm.

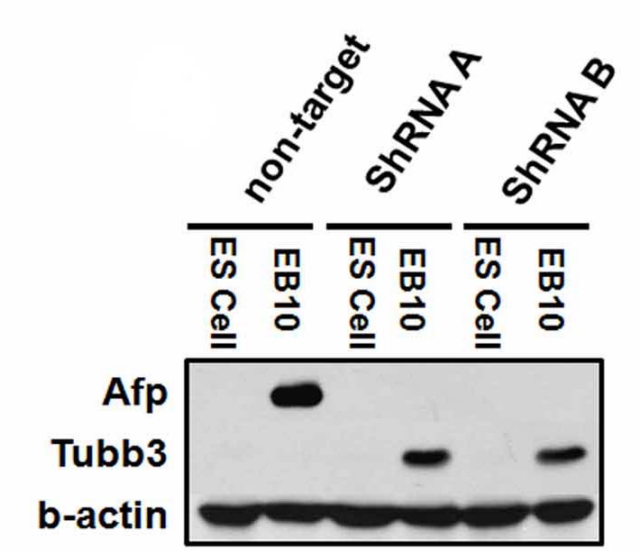


Figure 3.4.8, Western blots to probe Afp and Tubb3 levels in control and TAF3 K/D samples.

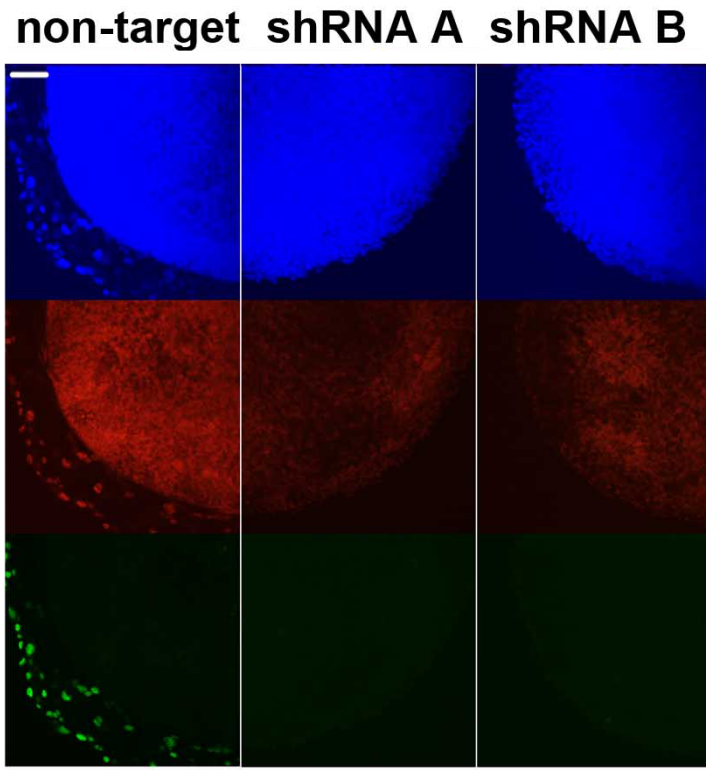


Figure 3.4.9, Confocal fluorescence images of control and TAF3 K/D EBs (day 10) co-stained with antibodies against TAF3 and GATA4. Scale Bars, 50 μm.

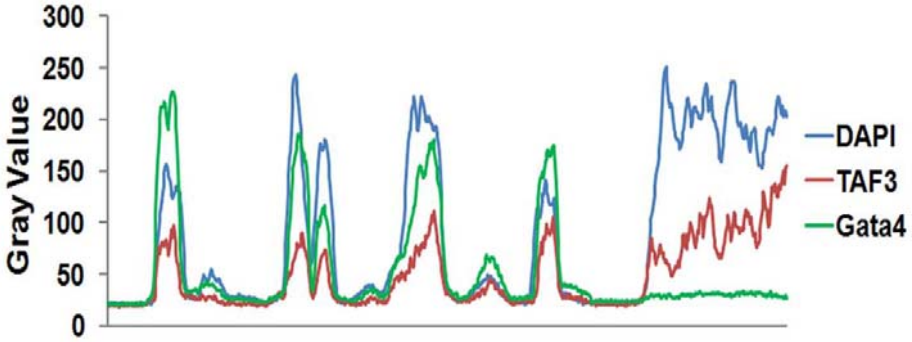


Figure 3.4.10, Gray values along a path from outside to inside of the EB (non-target) in DAPI, TAF3, and GATA4 channels of (I) were plotted in the diagram.

3.5 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 3.5.1; 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 3.5.1; 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 3.5.2 and 3.5.3). 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 3.5.1; 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 3.5.4), 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.

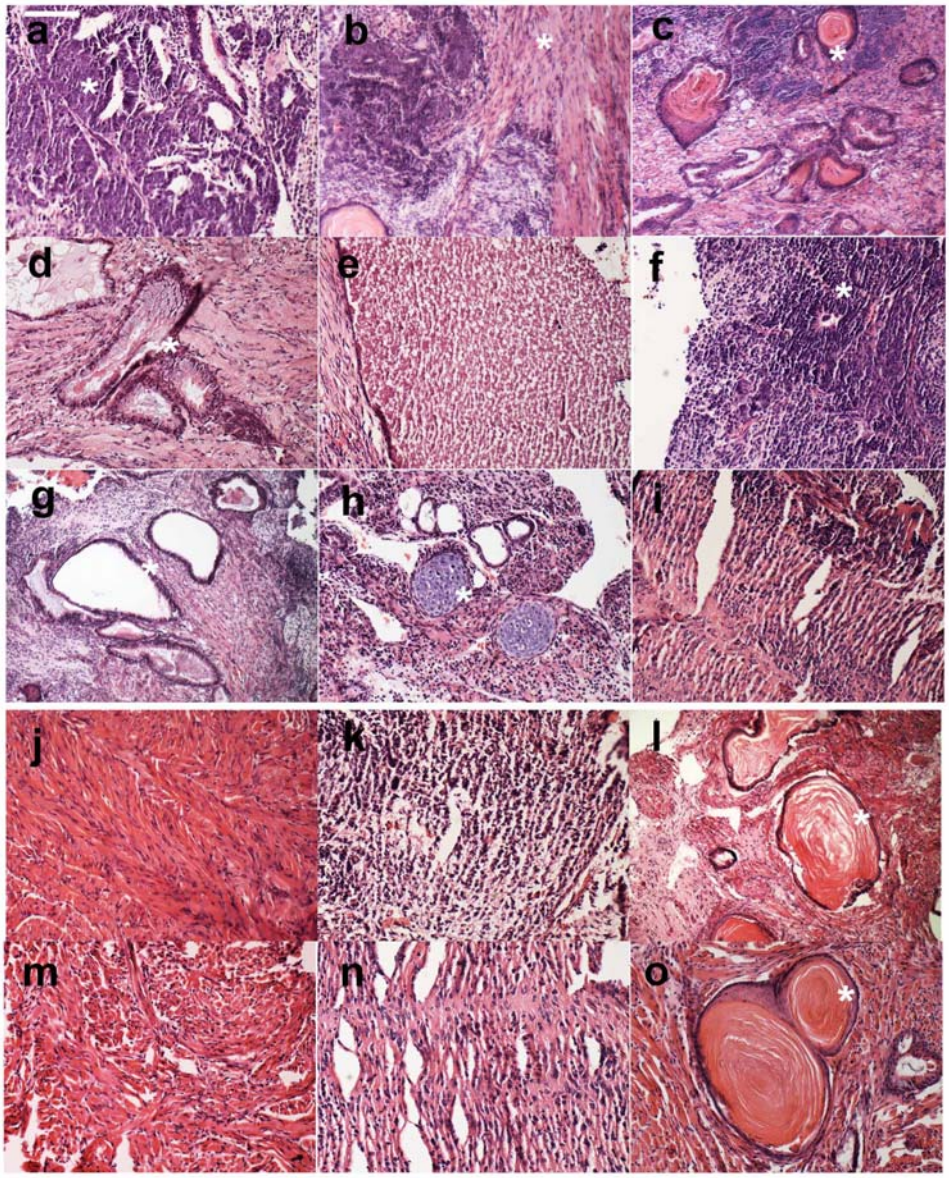


Figure 3.5.1, H&E staining of control (non-target) and TAF3 K/D teratomas. Control teratomas contained tissues from endoderm (a, gut-like epithelium; d, intestinal gland; g, respiratory epithelium), mesoderm (b, muscle; e, fat cells; h, cartilage) and ectoderm (c, epidermis; f, neuroepithelium; i, bipolar neurons). TAF3 K/D teratomas (shRNA A2 j, k, l; shRNA B1 m, n, o) mainly composed of muscle (j, m) and ectodermal tissues (k, l, n, o). Scale Bar, 100 μm

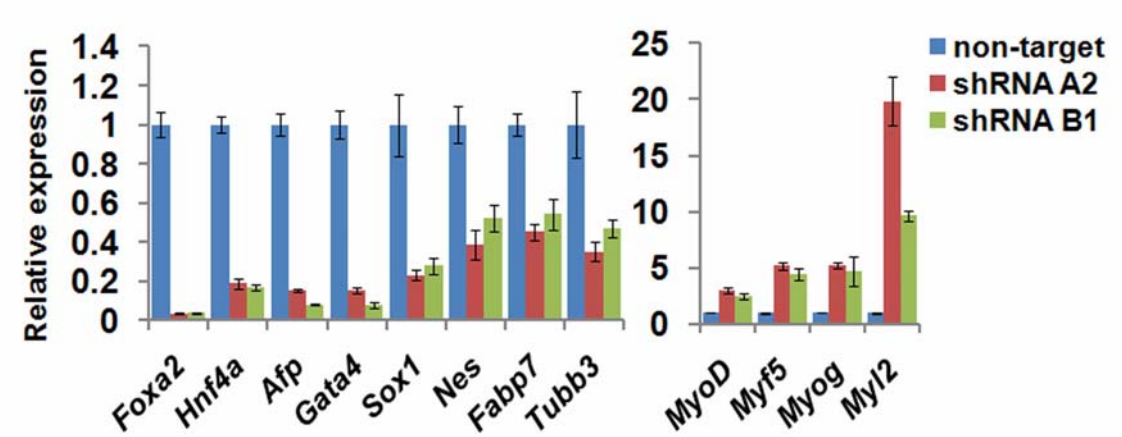


Figure 3.5.2, qRT-PCR to evaluate the relative expression of tissue-specific genes in control and TAF3 K/D teratomas.

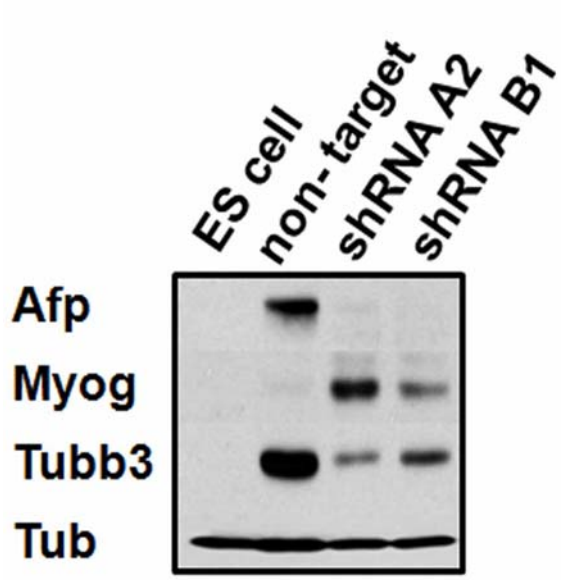


Figure 3.5.3, Western blots to compare the expression of tissue-specific genes between control (non-target) and TAF3 K/D teratomas.

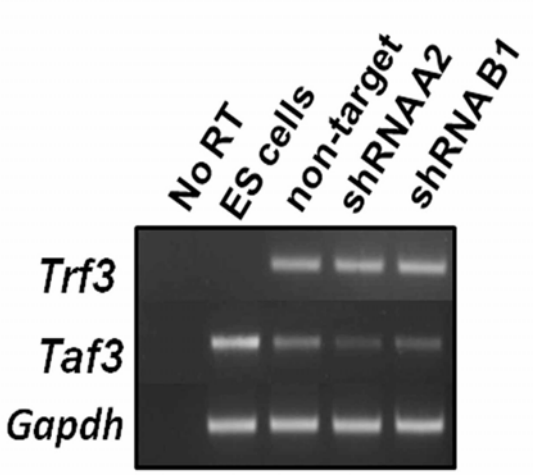


Figure 3.5.4, Semi-quantitative PCR to measure the expression of Taf3 and Trf3 after teratoma formation.

3.6 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 3.4.1, 4.1.1 and 4.1.2) 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 4.1.1) 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 ($CXCR^+c\text{-}Kit^+$) and mesoderm ($CXCR4^+Flk\text{-}1^+$) cell populations by flow cytometry (Figure 3.6.1). 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 3.6.2).

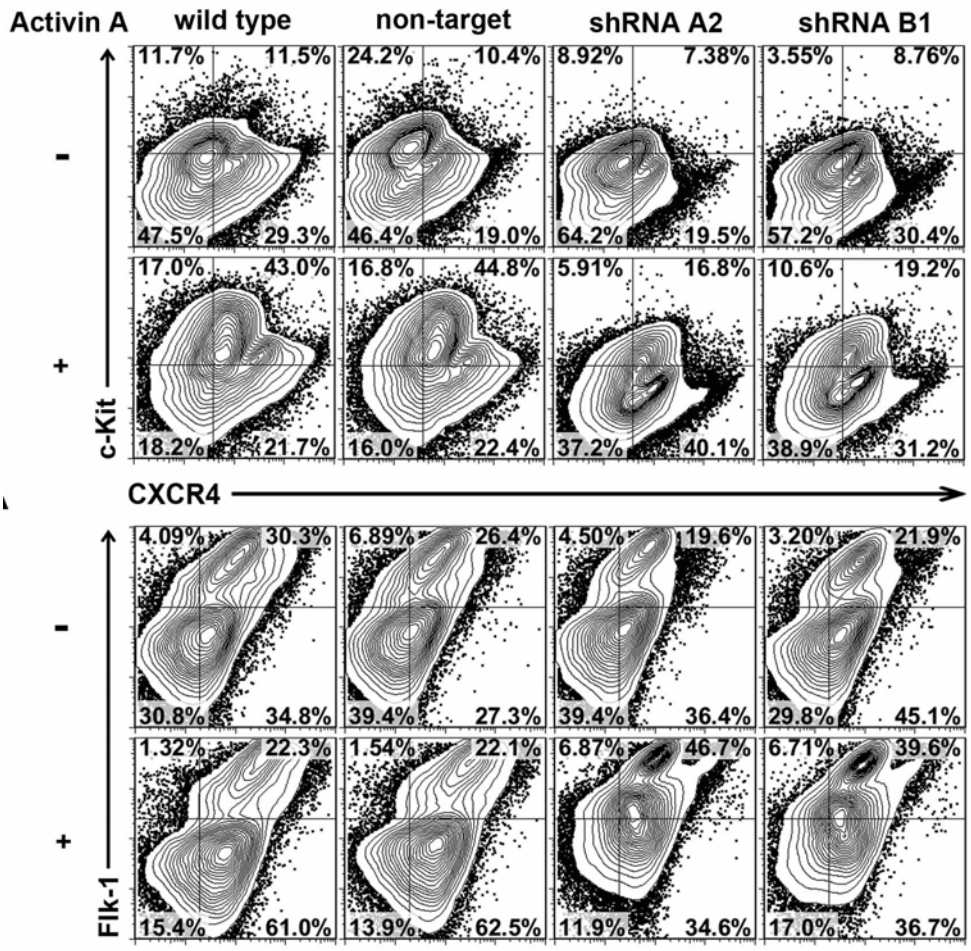


Figure 3.6.1, Flow cytometric analysis of definitive endoderm ($CXCR4^+c\text{-}kit^+$) and mesoderm ($CXCR4^+Flk\text{-}1^+$) cells after the directed in vitro differentiation. Experiments were conducted with or without Activin A.

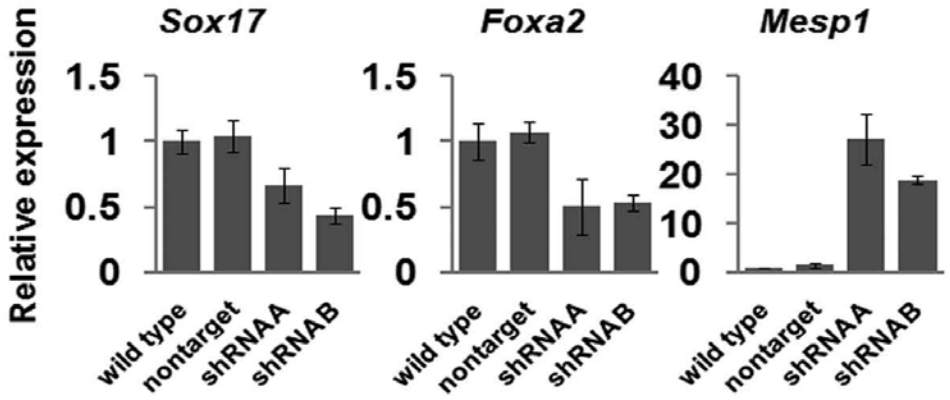


Figure 3.6.2, qRT-PCR to examine the relative expression of endoderm (*Sox17*, *Foxa2*) and mesoderm (*Mesp1*) markers in Activin A treated samples.

Chapter 4 TAF3-mediated Transcriptional Regulation

4.1 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 4.1.1; 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 4.1.1; 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 3.3.1), 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 4.1.2, 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 4.1.4) 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 4.1.1; Columns 3-5). We also observed up-regulation of several members of the nodal pathway (*Nodal*, *Lefty1*, *Cer1*; Figure 4.1.1; 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 4.1.1; Columns 6-7).

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 4.1.3 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.

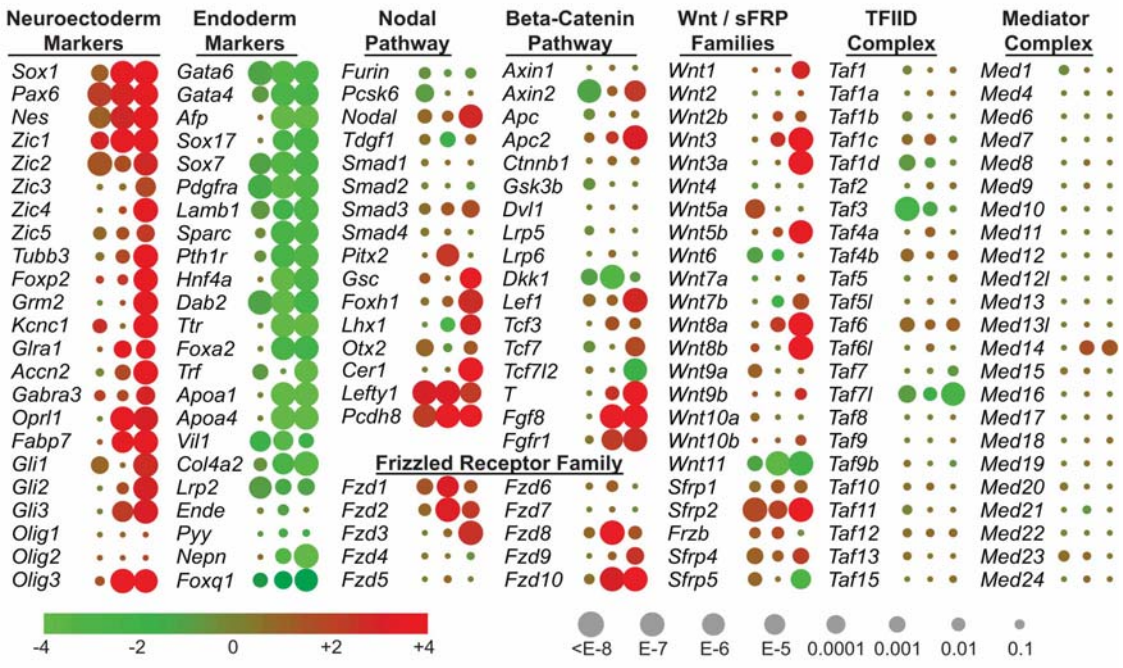


Figure 4.1.1, Difference in expression (log2(anti-TAF3 shRNA A/non-target shRNA)) and statistical significance between TAF3 depleted ES cells and controls for selected genes at three timepoints, ES Cell, EB day 3 and EB day 6.

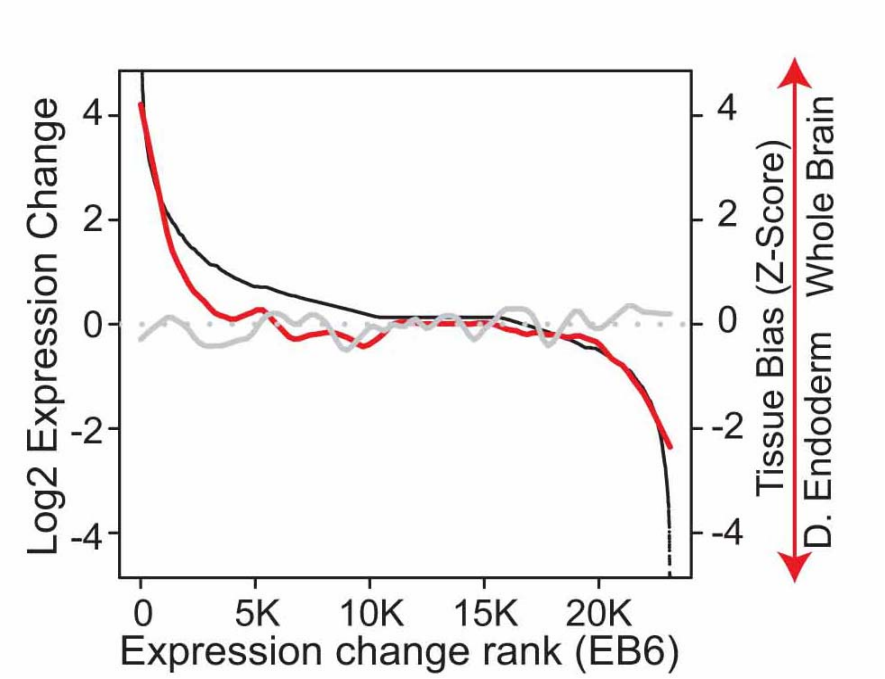


Figure 4.1.2, Tissue bias (red line; *Experimental Procedures*) calculated between definitive endoderm (E8.0; SM143) and whole brain (E12.5; SM108) SAGE libraries plotted on the difference in expression (defined as in (A); black points) between TAF3 K/D cells and controls. Tissue bias recalculated using randomized gene expression data was plotted in grey.

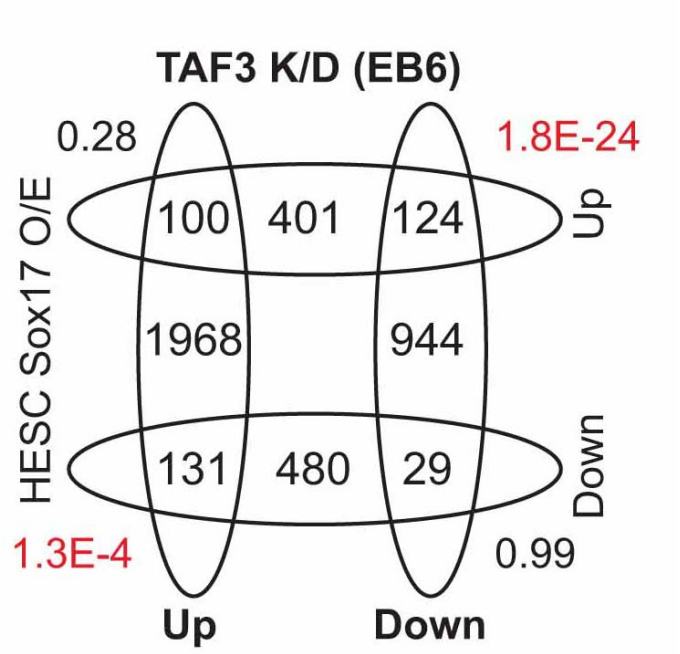


Figure 4.1.3, Overlaps between genes that were differentially expressed in TAF3 K/D EBs at day 6 (EB6) and genes that were differentially expressed in Sox17 over-expressing human embryonic stem cells. P-values (italics) for each intersection were calculated assuming a hypergeometric distribution and a total ortholog count of 14,400. Values significant at the 5% level were plotted in red.

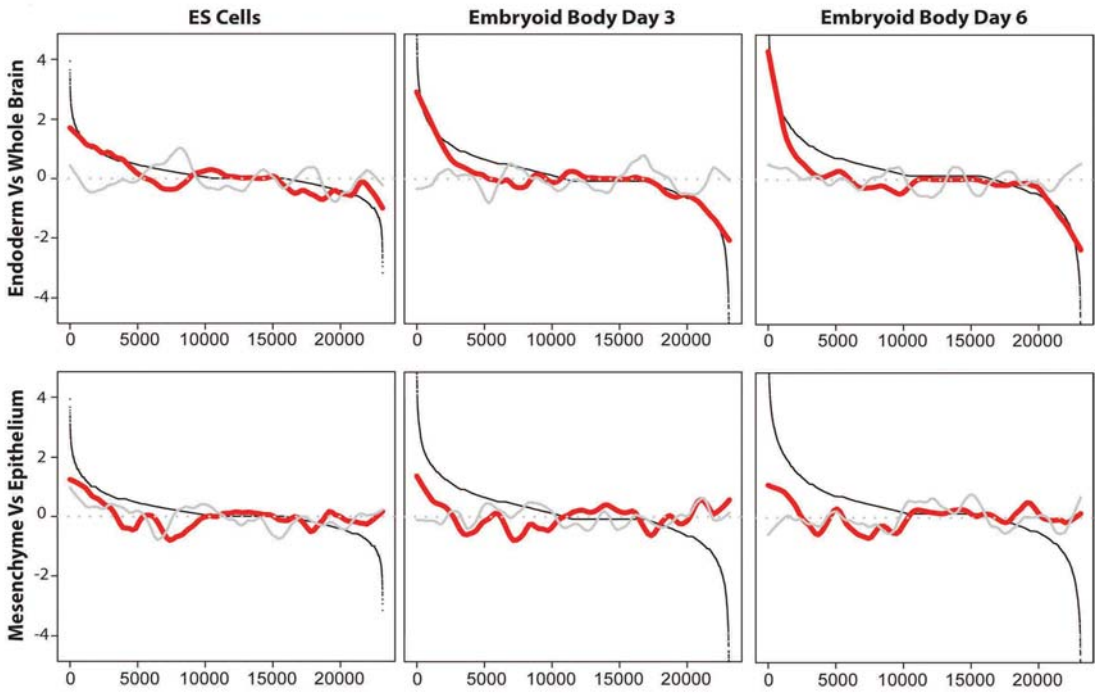


Figure 4.1.4, Tissue bias (red line) calculated between definitive endoderm and whole brain SAGE libraries (top three panels) or mesenchyme and epithelium libraries (bottom three panels) plotted on the difference in expression (log2(TAF3 shRNA A/non-target shRNA); black points) between samples derived from TAF3 K/D and control samples (ES cell, EB3 and EB6). Tissue bias recalculated using randomized gene expression data was plotted in grey.

4.2 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 4.2.1A; Figure 4.2.2; P = 0 by permutation) and its enrichment at promoters was strongly correlated with that of TAF1 and TBP (Figure 4.3.4). 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 4.3.4). 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 4.3.4; Figure 4.2.3) 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 4.2.5). 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.

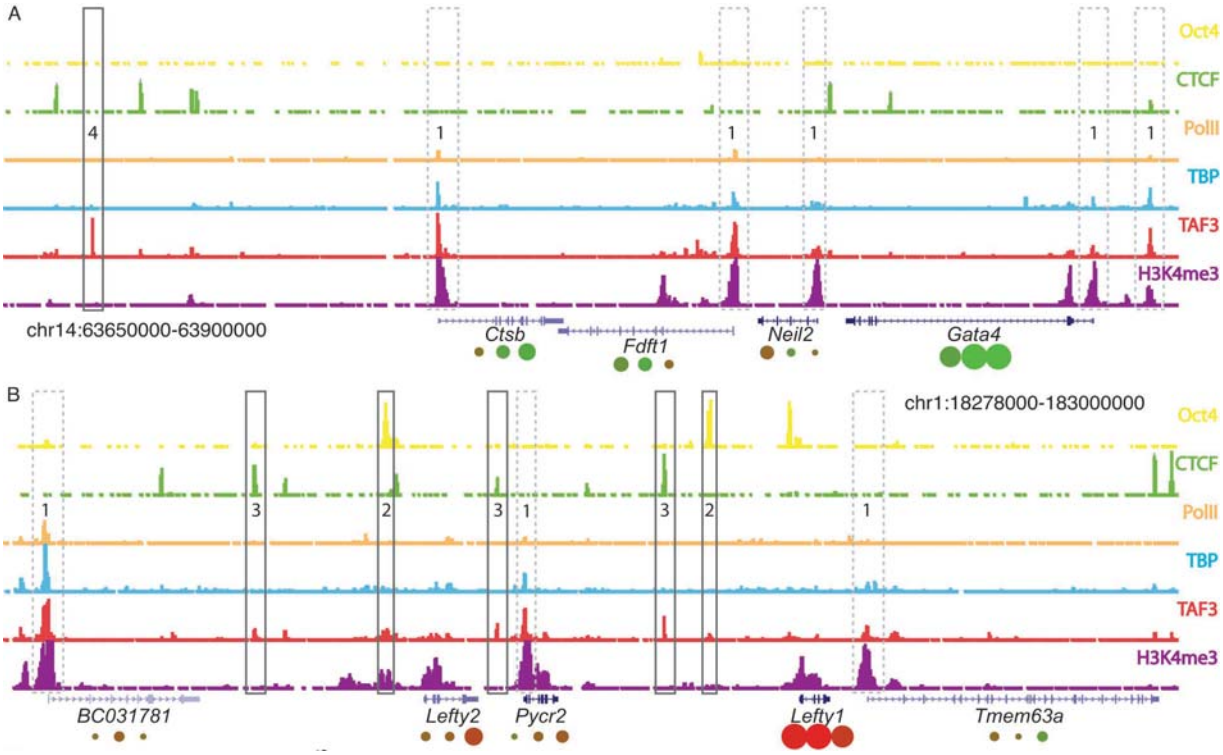


Figure 4.2.1, (A) Read accumulation for six ChIP-seq datasets (TAF3, TBP, Pol II, H3K4me3, CTCF and Oct4) at the *Gata4* locus. Vertical axis was from 3 to 50 reads for all factors. TAF3 dependent expression changes were plotted below genes as for Figure 4.1.1. Active promoters were denoted with dashed boxes. (B) As for (A) at the *Lefty1* locus.

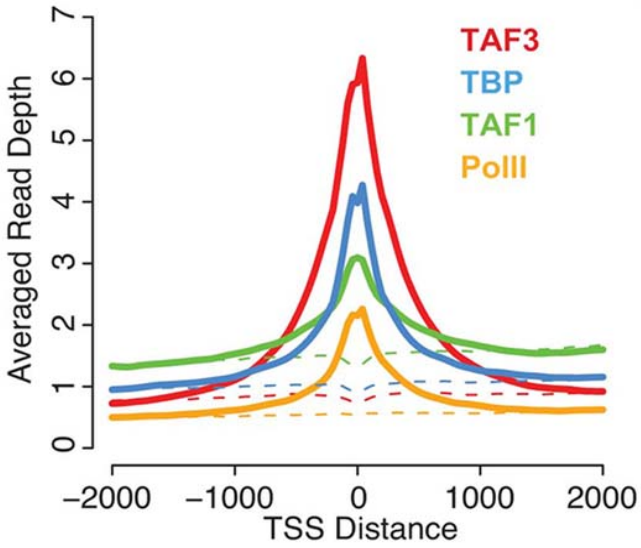


Figure 4.2.2, TAF3 is enriched at transcription start sites and enrichment is correlated with that of other TFIID members. Number of mapped reads per base centered on the annotated (Ensembl Mouse, Release 56) transcription start (solid lines) or end (dashed lines) site averaged over all protein-coding genes.

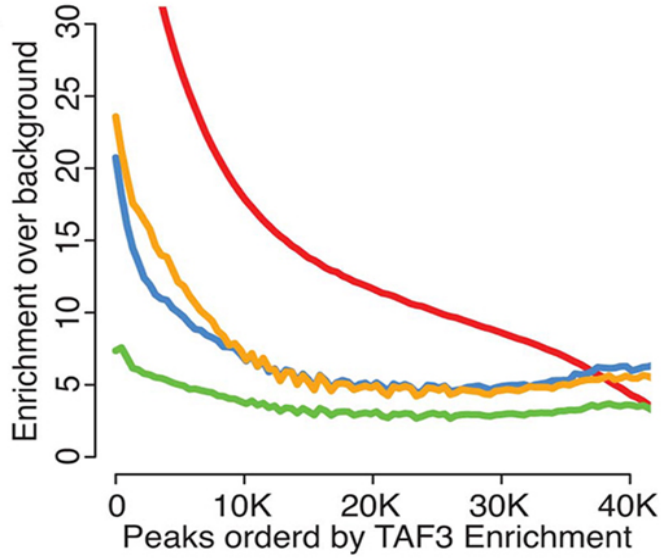


Figure 4.2.3, Enrichment over background as calculated by MACS (macs14 1.4.0beta) for TAF3, TAF1, TBP and Pol II at all locations in the "44K dataset". Peaks were ordered by TAF3 enrichment. All plots were smoothed using a lowess smoothing fraction of 0.2.

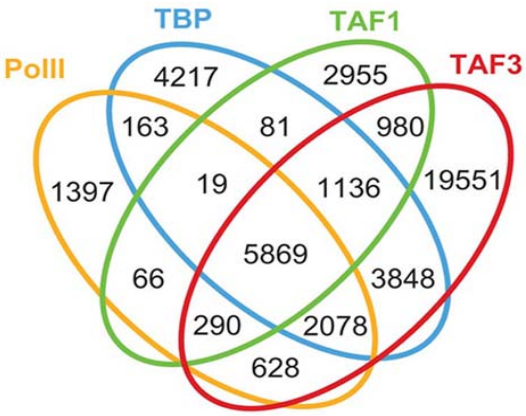


Figure 4.2.4, Overlaps among bound regions called by MACS (macs14 1.4.0beta) at a false discovery rate of 1% for TAF3, TAF1, TBP and Pol II.

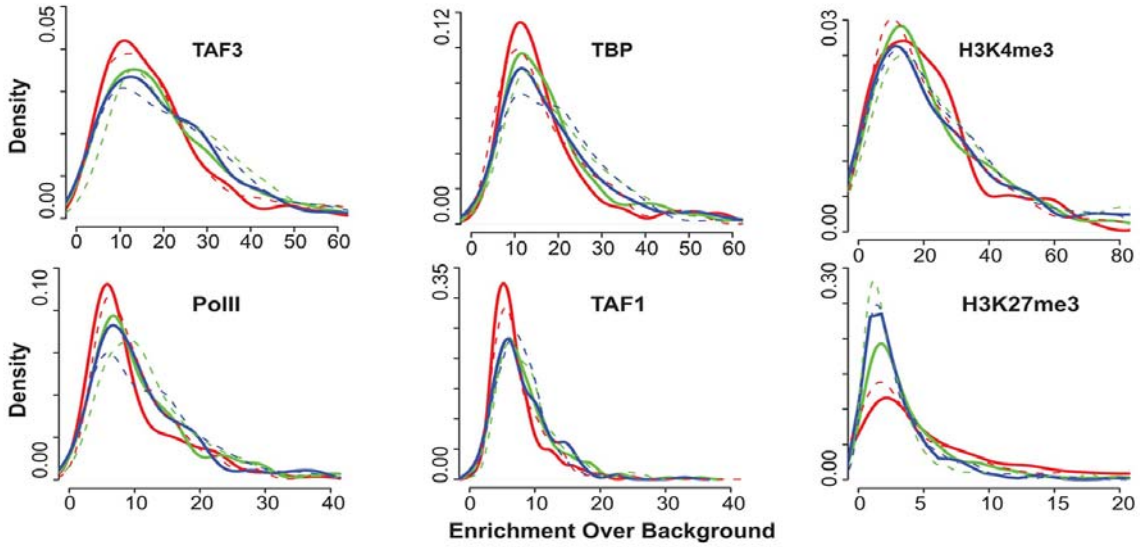


Figure 4.2.5, Binding of TFIID subunits at the core promoter does not distinguish TAF3-dependent genes from unaffected genes or controls. Enrichment over background as calculated by MACS (macs14 1.4.0beta) at promoters of genes up-regulated (red), down-regulated (green) or unaffected by TAF3 K/D (blue) as well as control gene sets (dashed lines). Data refer to 8581 genes for which the core promoter could be confidently identified on the basis of a single bound region within 2Kb of the annotated TSS and no other bound regions detected within 5Kb of the TSS. Control gene sets were sampled to have the same distribution of mapped read counts in wild type ES cells as the target set.

4.3 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 4.2.1A solid box; Figure 4.2.4). These regions were generally further from core promoters (Figure 4.3.1) and less enriched (Figure 4.3.2) than TFIID-associated regions but often overlapped with regions enriched for other factors active in ES cells (Figure 4.3.3). 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 4.3.4). 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 4.3.4). Examples of regions enriched for TAF3 and CTCF are shown in Figure 4.2.1B.

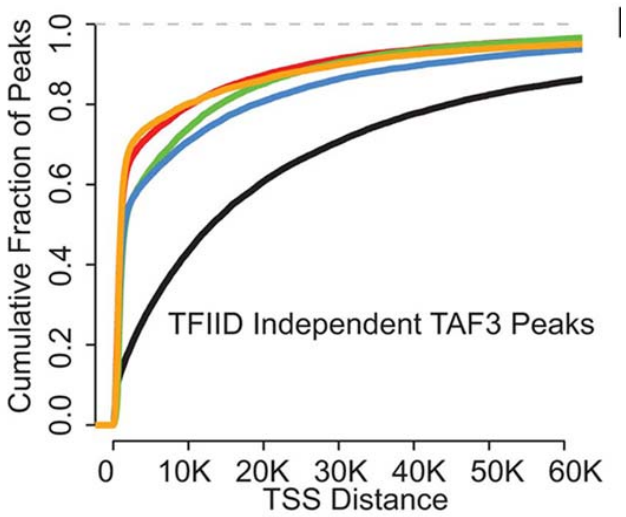


Figure 4.3.1, TAF3 bound regions that are not enriched for by TAF1, TBP or Pol II (as defined for Figure 4.2.2) are distributed further from the TSS.

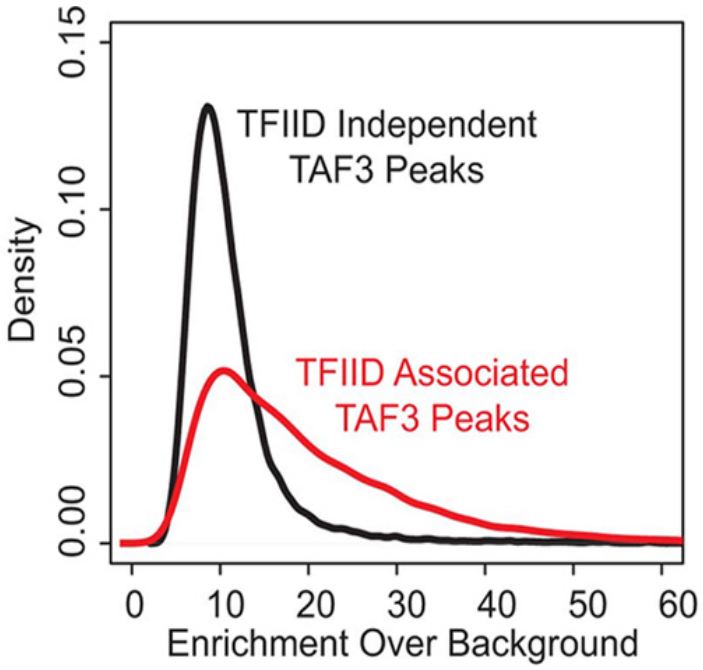


Figure 4.3.2, TAF3 bound regions that are not enriched for by TAF1, TBP or Pol II (as defined for Figure 4.2.2) are less enriched over background.

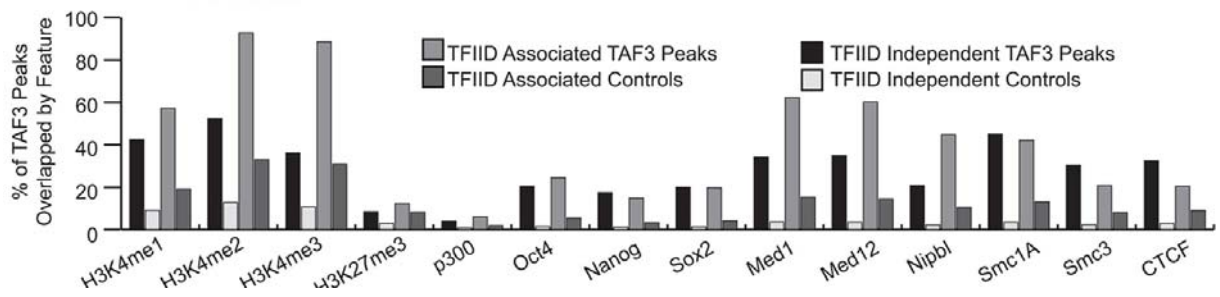


Figure 4.3.3, TAF3 TFIID-associated and TFIID-independent bound regions are concentrated in regions enriched for other factors. The vertical axis shows the percentage of TAF3 bound regions or the mean of matched control samples (See *Experimental Procedures*), overlapped by significantly enriched regions from a panel of previous ChIP-seq experiments. The standard deviation of the sampled control sets was not plotted as it was less than 1% in all cases.

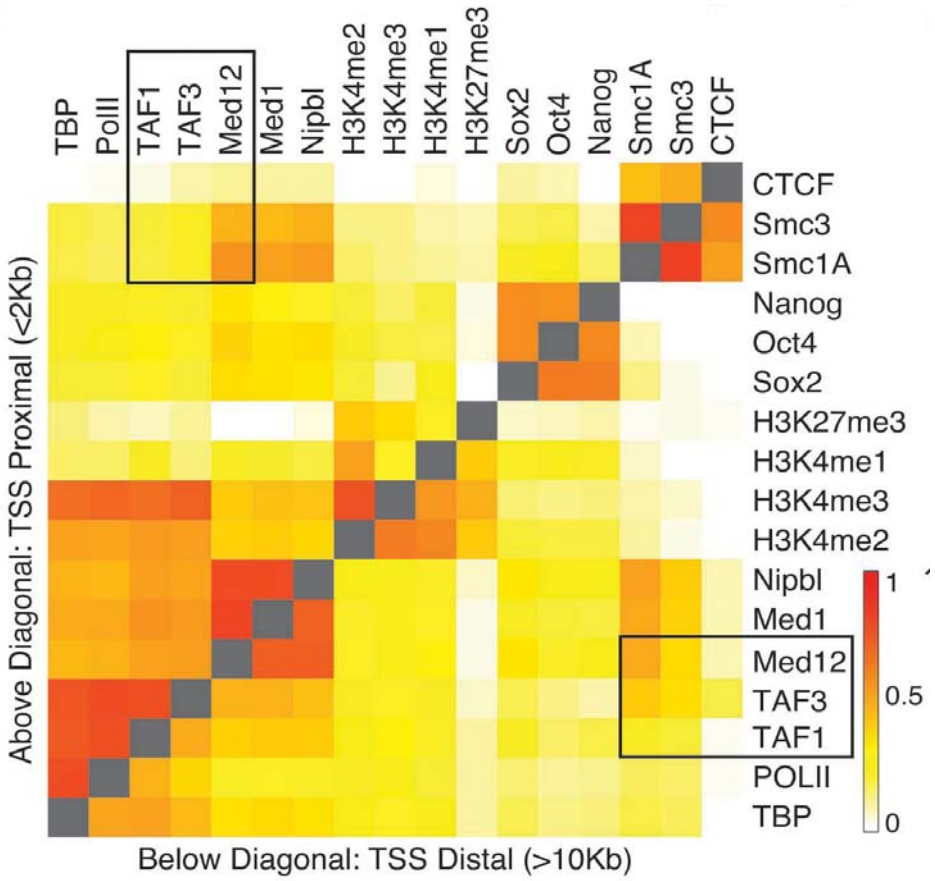


Figure 4.3.4, Binding correlations (Spearman coefficient) between all pairs of factors considered in this study based on enrichment values calculated using the MACS model at 125K genomic loci. Negative correlations were set to zero for display.

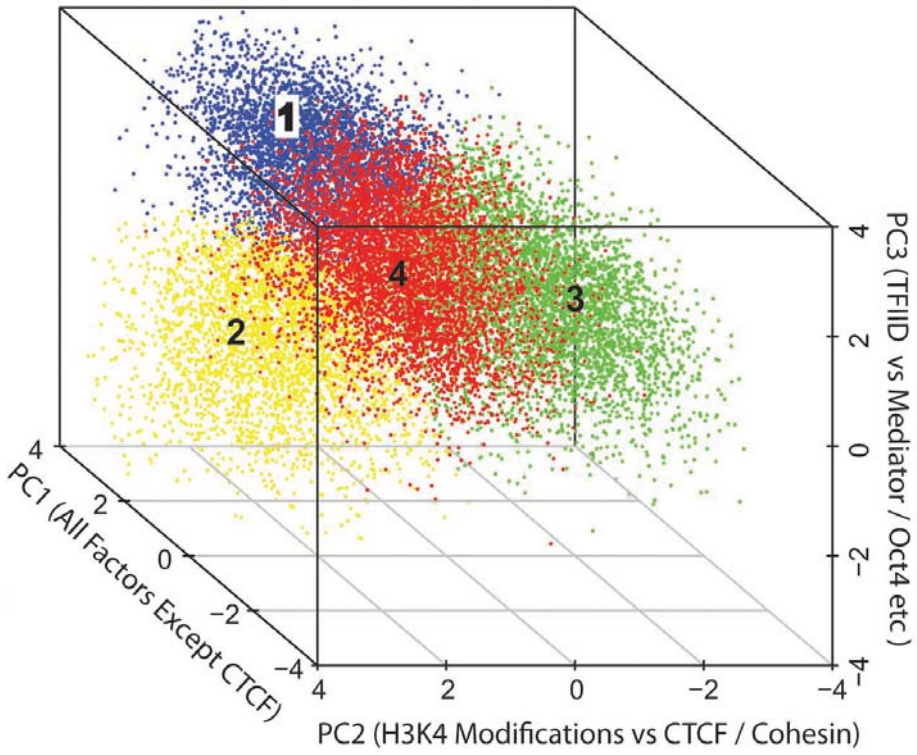


Figure 4.3.5, Clustering of TAF3 bound regions into four classes using the first three principal components of binding variation computed from all factors in Figure 4.3.4.

	Regions	TSS (Kb)	Enrichment Over Background								
			TAF3	PolII	TAF1	TBP	K4me3	Oct4	Med12	Smc1A	CTCF
1.	12,570	<1	19	**10**	**4**	**9**	**20**	8	6	4	5
2.	5,599	17	10	4	2	4	3	**30**	**10**	7	5
3.	7,211	12	11	3	2	3	1	7	5	**15**	**85**
4.	9,387	9	10	3	2	3	2	6	3	3	5

Figure 4.3.6, Summary statistics for the four TAF3 binding classes in Figure 4.3.5. All values except the number of regions in each class (first column) are median values for the whole class. Enrichment was calculated as for Figure 4.3.4.

4.4 TAF3/CTCF/cohesin Bound Regions 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 4.3.5; See Experimental Procedures). This procedure groups together regions that have similar enrichment profiles (Figure 4.3.6) across all 17 factors and histone modifications examined in Figure 4.3.4. 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 4.3.6). 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 4.2.1A).

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 4.4.1). 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 4.3.5, 4.3.6 & 4.4.1) than expected by chance, no such association was seen for genes up-regulated upon TAF3 K/D (Figure 4.4.1). 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, Oct2 and Nanog may comprise such an alternative mechanism that can apparently be opposed by TAF3 (Figure 4.4.1). Our data support two other observations. First, regions bound by TAF3 only (Class 4 regions; Figure 4.3.5, 4.3.6 and 4.4.1) 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 4.3.5, 4.3.6 and 4.4.1) is associated neither with up- nor down-regulated genes. This is confirms our previous observations (Figure 4.2.5) and is consistent with the critical role played by TFIID at the basal promoters of many genes.

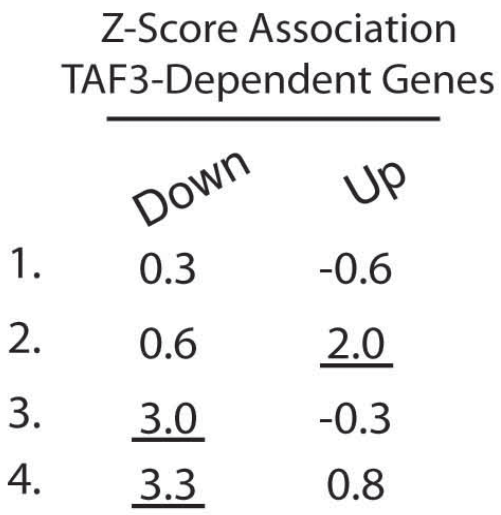


Figure 4.4.1, Z-score association between TAF3 binding classes from Figure 4.3.5 and genes whose expression was significantly affected by TAF3 depletion. Significant values were underlined; Z-score between +/-1 (P-value > 0.6); Z-scores > 2 (P-value < 0.05); Z-scores > 3 (P-value < 0.001). See *Experimental Procedures* for details of Z-score calculation.

Chapter 5 TAF3 Binds CTCF and Mediates DNA Looping

5.1 The vertebrate domain of TAF3 interacts with CTCF

The strongly correlated binding of CTCF and TAF3 to promoter distal sites (Figure 4.3.6) 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 2.3.2). Importantly, CTCF but not cohesin (Smc1a and Smc3) or mediator subunit (Med12) were selectively enriched by TAF3 immunoprecipitation (Figure 5.1.1) and the interaction between TAF3 and CTCF appears to be direct and independent of DNA (Figure 5.1.1; 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 5.1.2 and 5.1.3).

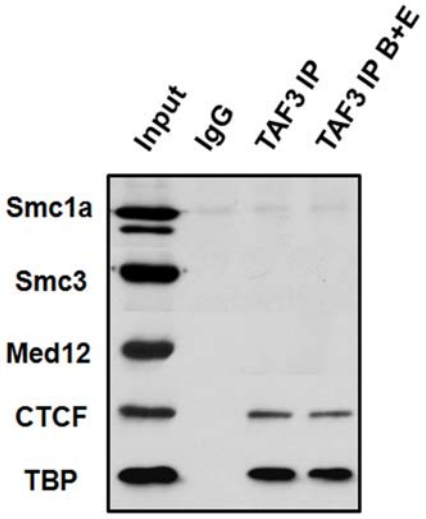


Figure 5.1.1, Western Blots to examine the relative enrichment of cohesin (Smc1a and Smc3), mediator subunit (med12) and CTCF by TAF3 immunoprecipitation using ES cell nuclear extract. B + E, Benzonase and Ethidium bromide treatment that eliminates DNA-mediated interactions.

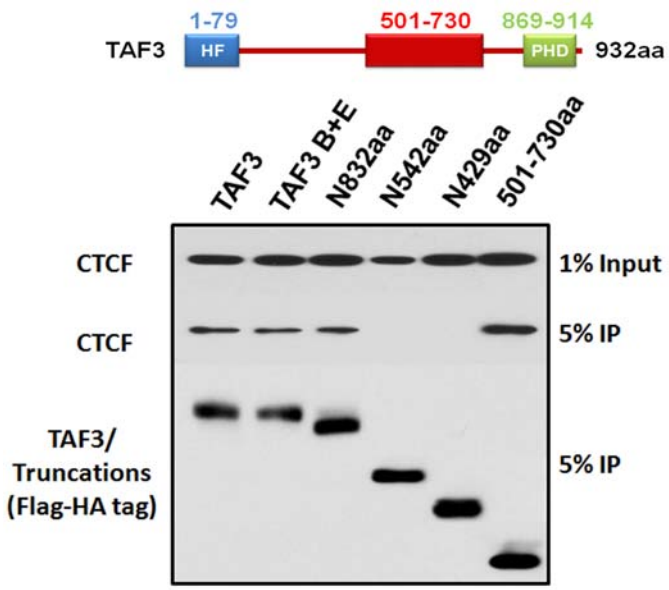


Figure 5.1.2, Mapping the CTCF-interacting domain of TAF3. Flag-HA tagged TAF3, RFP or truncated TAF3 proteins were over-expressed in 293T cells with CTCF and were then purified with anti-Flag resin. TAF3 and CTCF enrichments were analyzed by western blots.

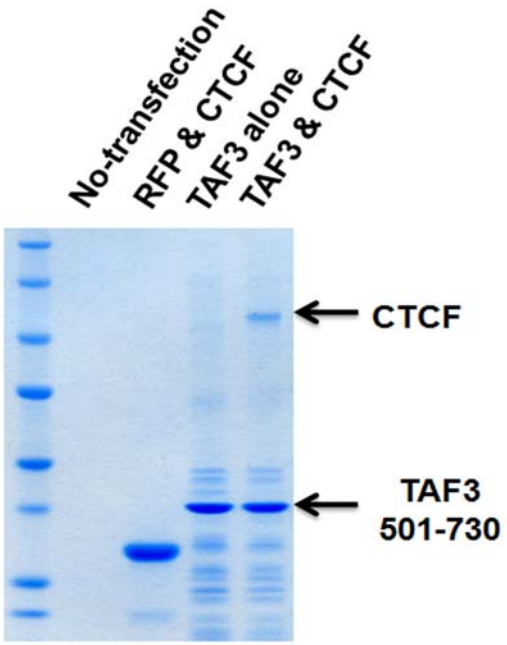


Figure 5.1.3, Coomassie Staining of the purified Flag-HA-TAF3(501-730aa)/CTCF complex.

5.2 CTCF Recruits TAF3 to Its Target Sites

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 5.2.2) whereas CTCF is required for efficient recruitment of TAF3 to distal (TAF3/CTCF) sites but not required for TAF3 occupancy at core promoters (Figure 5.2.1). 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 4.3.5) is mechanistically distinct from TAF3 localization to TFIID-bound regions (Class 1 in Figure 4.3.5), 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.

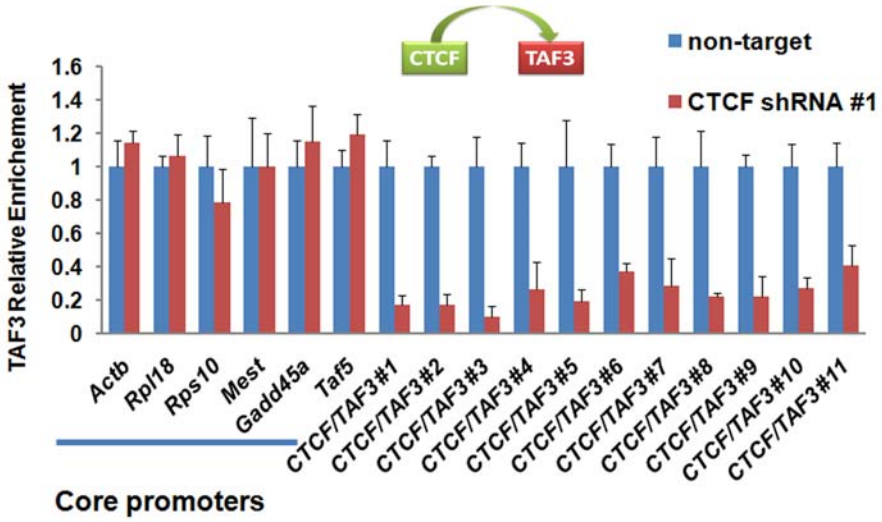


Figure 5.2.1, ChIP analysis to examine the relative TAF3 enrichment at 6 core promoters bound by TAF3 and 11 distal sites bound by TAF3/CTCF in control and CTCF K/D cells.

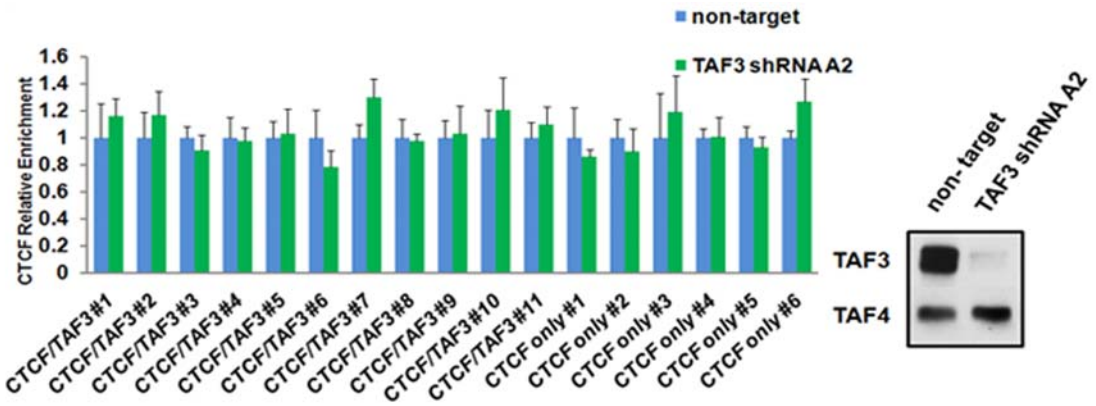


Figure 5.2.2, ChIP-qPCR analysis of the relative CTCF enrichments at 11 distal sites bound by TAF3/CTCF and 6 distal sites bound by CTCF but not TAF3 in control and TAF3 K/D cells.

5.3 TAF3 Mediates Long-distance Chromatin Interactions

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 5.3.1 and 5.3.4). These findings strongly support a molecular mechanism in which TAF3/CTCF mediates long-range chromatin transactions that likely regulate proper transcription activation (Figure 6.3.1). 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 5.3.2, 5.3.3 and 5.3.5). 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 4.1.1).

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 6.3.1). 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 5.3.6A and 5.3.6B). 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 4.3.5 and 4.3.6) that are also associated with TAF3-activated genes (Figure 4.4.1) contribute to endodermal gene expression. TAF3 may also regulate lineage commitment via one of its core promoter functions such as H3K4me3 binding.

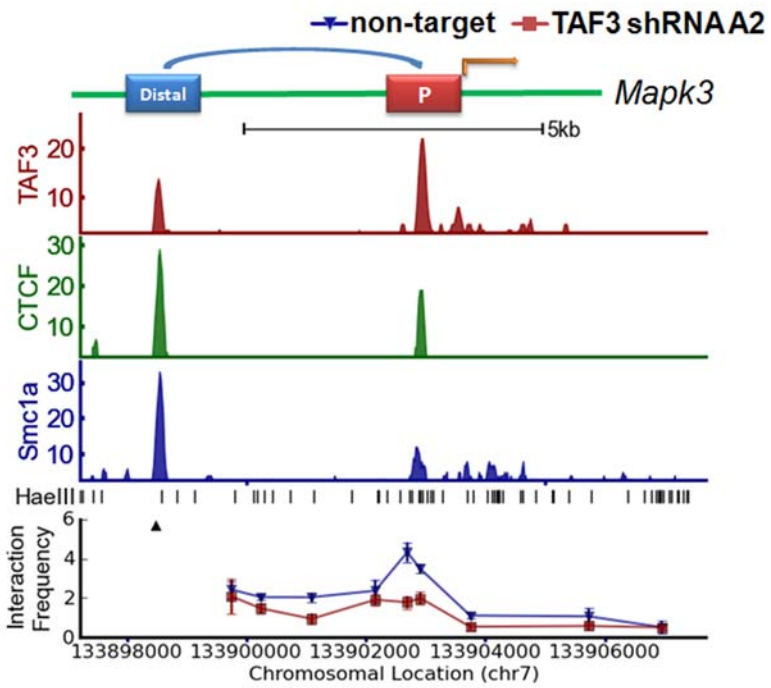


Figure 5.3.1, 3C experiments to assess long-range interactions between the distal TAF3/CTCF/cohesin site and regions around the core promoter of *Mapk3* in control (blue) and TAF3 K/D (red) cells. ▲, anchor point.

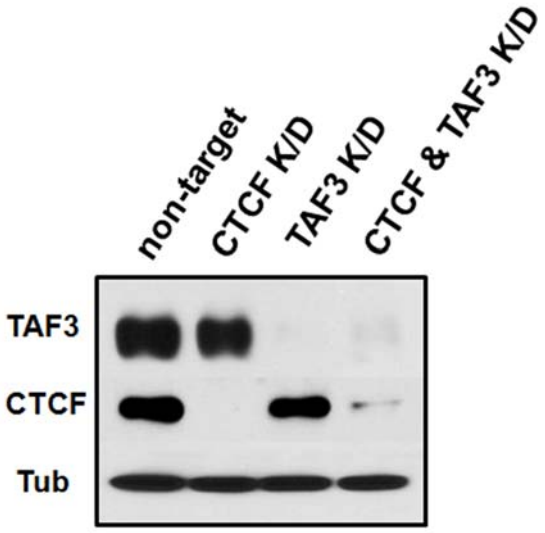


Figure 5.3.2, Western blots to examine TAF3 and CTCF protein levels in control, TAF3 K/D (shRNA A), CTCF K/D (shRNA #1) and TAF3 & CTCF K/D cells (72 hours post infection).

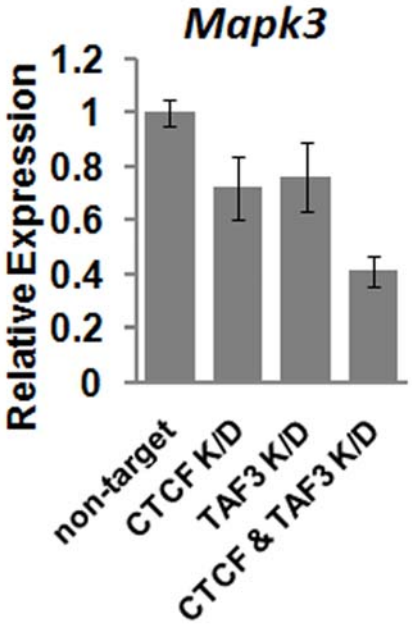


Figure 5.3.3, qRT-PCR analysis to measure *Mapk3* expression in control, TAF3 K/D, CTCF K/D and TAF3 & CTCF K/D cells.

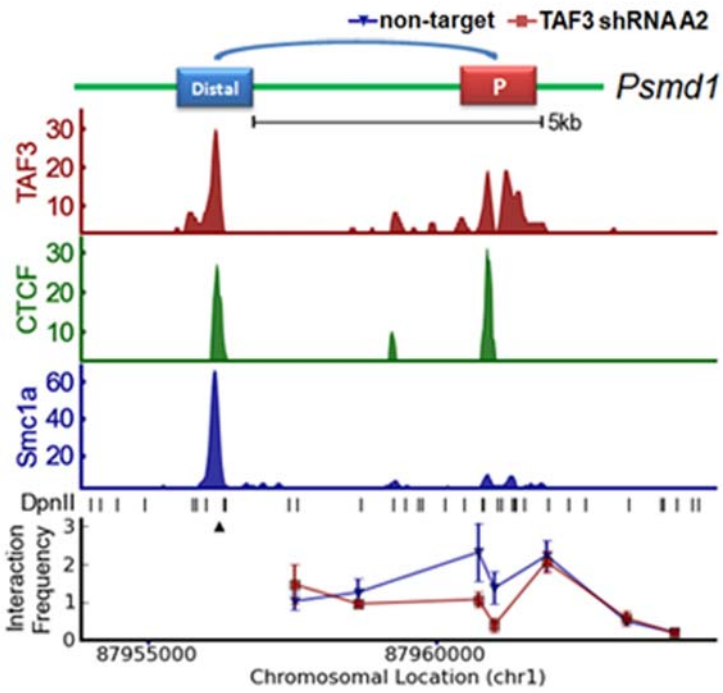


Figure 5.3.4, 3C experiments at *Psmd1* locus as presented in Figure 5.3.1.

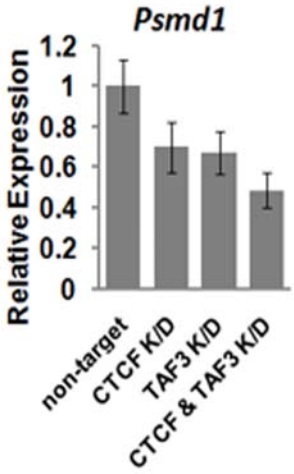


Figure 5.3.5, qRT-PCR analysis to measure *Psmd1* expression in control, TAF3 K/D , CTCF K/D and TAF3 & CTCF K/D cells.

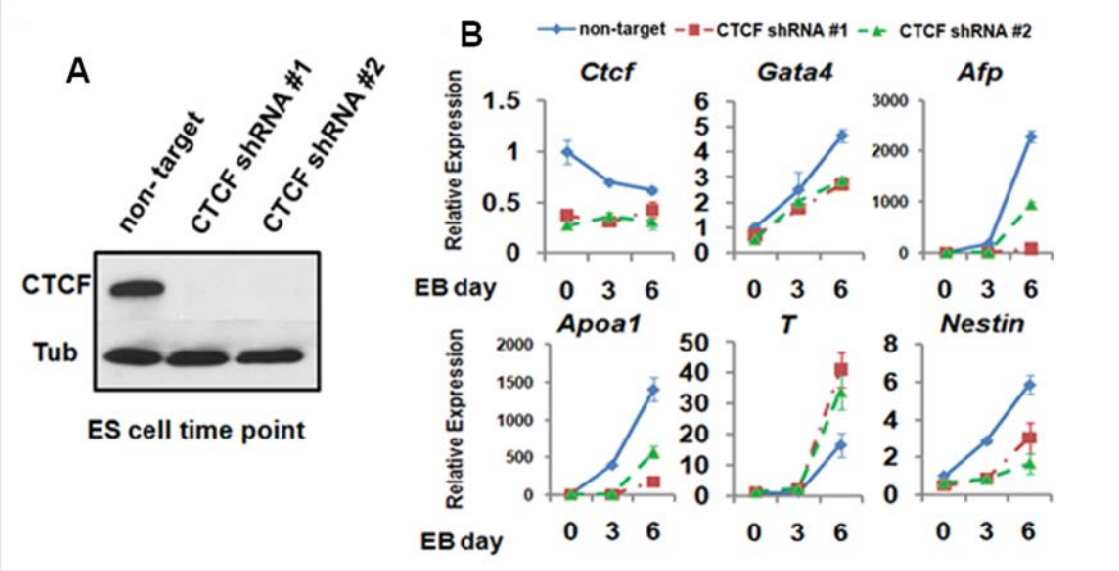


Figure 5.3.6, (A) Western blots to assess CTCF protein levels in control and CTCF K/D ES cells (90 hours post infection).
(B) qRT-PCR analysis of the expression of germ layer markers (Endoderm; *Gata4*, *Afp* and *Apoa1*, Mesoderm; *T* and ectoderm; *nestin*) during EB differentiation of control and CTCF K/D cells.

Chapter 6 Discussion

6.1 Core Promoter Complexes as Guardians of the Stem Cell State

The present work demonstrates a novel ES cell specific role of TAF3 in maintaining ES cell pluripotency. In conjunction with our previous work on TAF3 and examples such as TAF4b (Goodrich and Tjian, 2010), it 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 showed that mediator is specifically down-regulated in these cells also (Deato et al., 2008). This parallels independent works showing that Med12 functions with Nanog in ES cells (Tutter et al., 2009) and that Med1 functions as a co-activator of GATA1 during erythropoiesis (Stumpf et al., 2006). In addition, our mRNA-seq data point to other cell-type specific candidates for both TFIID and mediator. Specifically, Taf7l and Med14 displayed significant defects in TAF3 K/D cells indicative of developmental regulation (Figure 4.1.1), which agrees well with previous findings that Med14 has a specialized function in adipogenesis (Grontved et al., 2010) and Taf7l is required for efficient spermatogenesis (Cheng et al., 2007). Moreover, since Taf7l dysregulation is evident in ES cells, it is possible that Taf7l is directly regulated by TAF3. Thus, it is of immediate interest to determine whether those factors might also function in other developmental processes. Whatever the outcome of these cases, we predict that additional instances of cell-type specific regulatory functions of core promoter factors will be forthcoming.

6.2 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 4.1.1 and 4.1.4). 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 4.1.1), 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 4.1.1), that normally represses mesoderm formation (Tam and Loebel, 2007). Likewise, TGF-β signaling became deregulated upon TAF3 depletion. Specifically, when cells lacking TAF3 were exposed to activin A (TGF-β activator) concentrations that would normally induce definitive endoderm differentiation, TAF3 K/D cells instead differentiated more into mesoderm. Taken together, these results are most consistent with TAF3 mediating two distinct mechanisms that regulate proper lineage specification.

6.3 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 4.3.4). 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 4.4.1), 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 6.3.1). Our genome-wide correlation analysis (Figure 4.4.1) 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 4.2.1A) and another class that is enriched for Oct4/Nanog/Sox2 as well as mediator subunits (Figure 4.2.1B). 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 4.4.1), 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. 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.

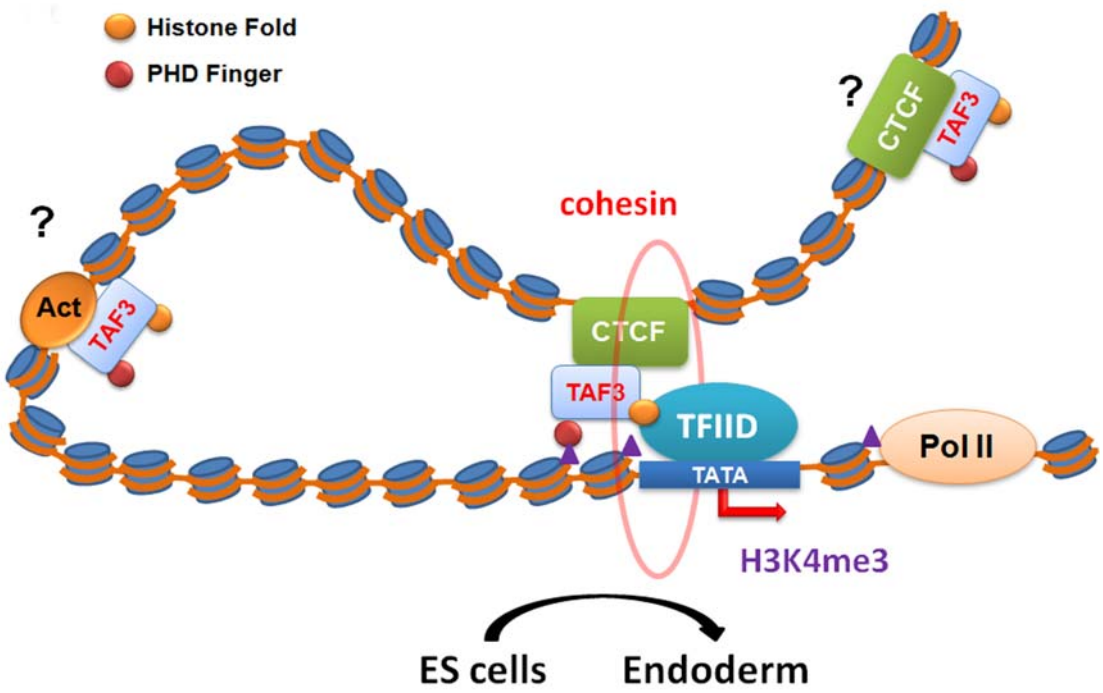


Figure 6.3.1, TAF3 interacts with CTCF and mediates regulatory DNA looping to specify endoderm differentiation.

Chapter 7 Other Works

7.1 The physiological roles of Brachyury in Animal Development and Cancer Progression

The *in vivo* function of Brachyury, a classical enhancer binding transcription factor, was originally characterized by early mouse genetic studies. Specifically, heterozygous mutant animals have a short tail phenotype, while homozygous ones die in *utero* and lack a notochord and mesoderm posterior (Naiche et al., 2005). The *Brachyury* gene was later cloned by Herrmann *et al.* (Herrmann et al., 1990) and was shown to express initially near the epiblast-ExE region, then in the entire nascent primitive streak and subsequently to be permanently silenced during later stages of development (Wilkinson et al., 1990). Interestingly, consistent with its crucial role in mesoderm induction, ectopic expression of Brachyury is sufficient to reprogram prospective ectodermal tissue to mesoderm in *Xenopus* (Cunliffe and Smith, 1992) and leads to the increased extension of the anteroposterior axis in transgenic mice (Stott et al., 1993). Re-expression of *Brachyury* was observed in various human tumors of epithelial origin and in human tumor cell lines derived from lung, colon and prostate carcinomas (Palena et al., 2007). Notably, multiple studies demonstrated that Brachyury promotes the epithelial-mesenchymal transition and metastatic behavior of human tumor cells (Fernando et al., 2011; Fernando et al., 2010; Sarkar et al., 2011).

In contrast to the well-illustrated physiological and pathogenic function of *Brachyury*, relatively little is known about the transcription mechanisms utilized by Brachyury to act on its target genes. More critically, until now only a few mammalian genes have been proposed as Brachyury targets but none has been confirmed *in vivo*. Thus, an uncovering of the gene regulatory network governed by Brachyury will be a key step towards understanding the role of Brachyury in mesoderm formation and cancer progression.

7.2 Brachyury Selectively Binds Cis-Regulatory Elements of Genes that Promote Mesoderm Cell Migration and Differentiation

Reasoning that the absence of knowledge about the molecular function of *Brachyury* is probably due to its highly-restricted spatiotemporal expression during embryonic development and lack of well-characterized antibodies against the protein, we have used the recently-developed Activin-A-mediated differentiation method that efficiently drives ES cells to form primitive streak cells *in vitro* (Gadue et al., 2006). We have also generated two polyclonal antibodies specific to Brachyury. After the induction of mesoderm formation by Activin A, a more than 300-fold increase in *Brachyury* expression (mRNA) was observed (Figure 7.2.1-2). Next, ChIP-seq experiments were performed with both anti-Brachyury antibodies. With a False Discovery Rate (FDR) of 1% and by only preserving peaks detected by both antibodies, we identified 3160 high-confidence Brachyury binding sites in the genome. These sites display genome-wide biased association with transcription start sites (Figure 7.2.3; 61% are within 5kb from TSS *v.s.* 9% of random control regions) and are significantly enriched around genes involved in the formation of primary germ layers (P-value < $5X10^{-10}$), dorsal/ventral pattern formation (P-value < $6X10^{-8}$), mesoderm cell migration (P-value < $3X10^{-7}$) and mesoderm formation (P-value < $2X10^{-6}$) (Table 7.2.1-2). In agreement with its role in cancer progression, we also observed a strong correlation of Brachyury binding with genes up-regulated in prostate cancer (P-value < $8X10^{-9}$) and endometerial cancer (P-value < $3X10^{-7}$) (Table 7.2.3).

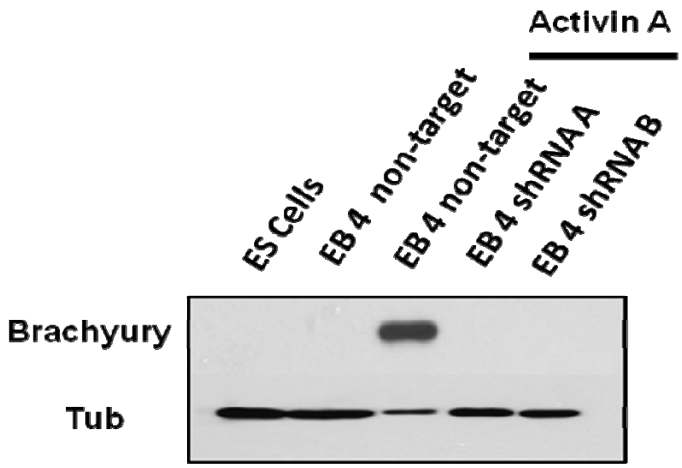


Figure 7.2.1, Western Blot analysis of Brachyury protein levels in ES cells (D3), control EBs (without Activin A induction) and control, shRNA A & shRNA B EBs that were treated with Activin A.

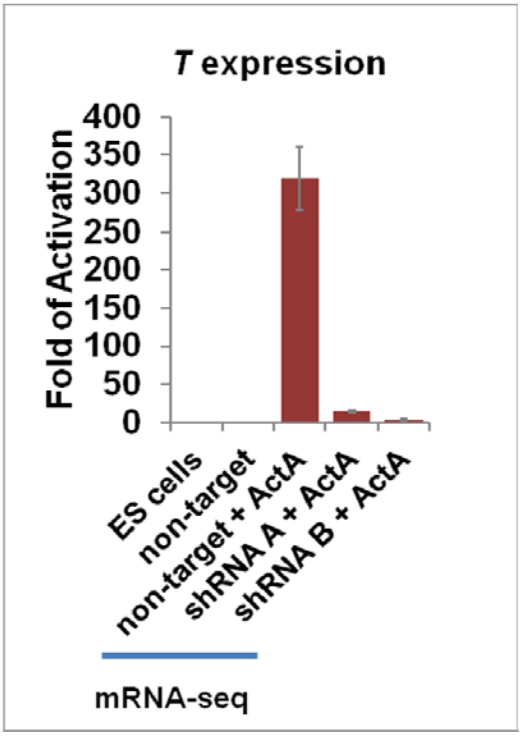


Figure 7.2.2, qPCR analysis of Brachyury mRNA levels in ES cells (D3), control EBs (without Activin A induction) and control, shRNA A & shRNA B EBs that were treated with Activin A.

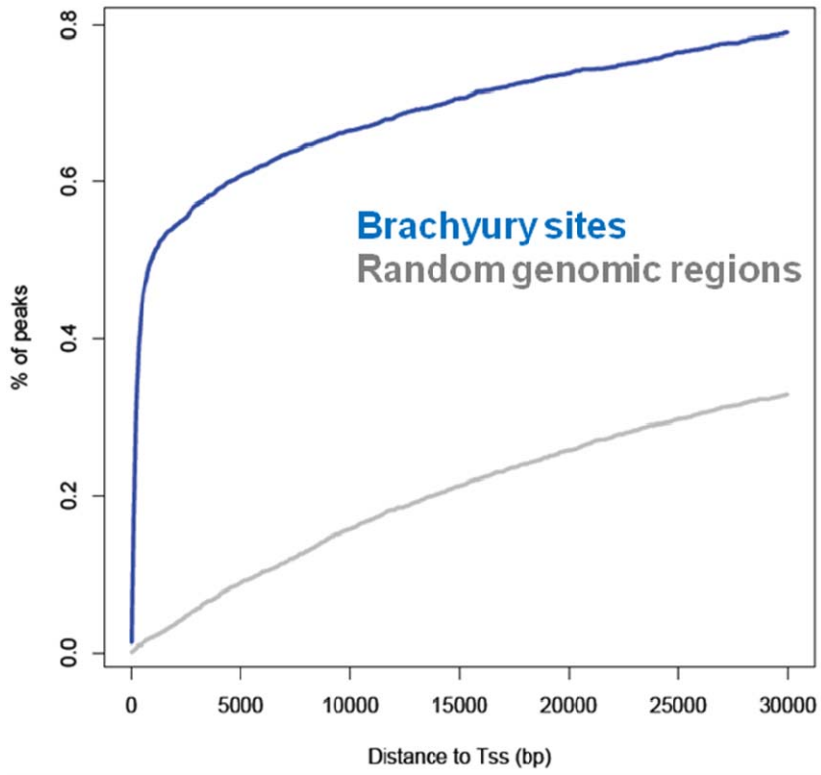


Figure 7.2.3, Association of Brachyury-bound (Blue) or random (grey) genomic regions with gene Transcription Start Site (TSS)-Percentages of peaks within the indicated distance to TSS were plotted.

GO:Biological Process	Binom raw P-value	Binom FDR q-value
formation of primary germ layer	4.13E-10	9.85E-08
dorsal/ventral pattern formation	5.23E-08	8.62E-06
midbrain development	8.87E-08	1.40E-05
mesodermal cell migration	2.96E-07	3.79E-05
mesoderm formation	1.68E-06	1.79E-04
endoderm development	2.03E-06	2.12E-04
mesoderm morphogenesis	3.38E-06	3.34E-04
regulation of gene expression, epigenetic	7.97E-06	6.90E-04
ER-nucleus signaling pathway	6.14E-05	4.88E-03
neural tube patterning	1.22E-04	9.00E-03

Table 7.2.1, Top 10 enriched Gene ontology terms (biological process) associated with Brachyury binding sites calculated by GREAT

GO:Mouse phenotype	Binom raw P-value	Binom FDR q-value
abnormal rostral-caudal axis patterning	3.87E-14	1.80E-11
abnormal mesoderm development	4.97E-14	2.15E-11
abnormal axial mesoderm	1.52E-11	5.82E-09
abnormal neural fold formation	8.91E-10	2.52E-07
abnormal midbrain-hindbrain boundary development	1.93E-09	5.02E-07
abnormal optic vesicle formation	5.30E-09	1.28E-06
abnormal primitive streak morphology	6.90E-09	1.55E-06
abnormal allantois morphology	6.95E-09	1.51E-06
abnormal midbrain development	1.77E-08	3.48E-06
abnormal notochord morphology	1.89E-08	3.62E-06

Table 7.2.2, Top 10 enriched Gene ontology terms (Mouse phenotype) associated with Brachyury binding sites calculated by GREAT

GO:MSigDB	Binom raw P-value	Binom FDR q-value
Prostate cancer	7.56E-09	1.33E-06
Genes involved in AKT phosphorylates targets in the cytosol	2.99E-07	3.76E-05
Endometrial cancer	1.11E-05	9.76E-04
Genes involved in PI3K/AKT signalling	5.09E-05	3.20E-03
Genes involved in Apoptosis induced DNA fragmentation	9.38E-05	4.86E-03
Genes involved in Metabolism of mRNA	1.88E-04	7.86E-03
Genes involved in Apoptotic execution phase	2.74E-04	8.94E-03
Genes related to IL4 rceptor signaling in B lymphocytes	1.33E-03	2.99E-02

Table 7.2.3, Top 10 enriched Gene ontology terms (MSigDB) associated with Brachyury binding sites calculated by GREAT

7.3 Regions Bound by Brachyury Positively Influence Gene Expression

To further probe how Brachyury regulates its target genes, we next performed mRNA-seq to analyze gene expression changes after Brachyury depletion by two independent lentivirus-mediated shRNAs. Compared to non-target shRNA control, we detected 424 down-regulated genes and 580 up-regulated ones. Interestingly, although both down- and up- regulated genes were enriched for developmental genes, we only detected a biased association of Brachyury sites (by ChIP-seq) with genes that were down-regulated upon loss of Brachyury, while the opposite association was seen for the up-regulated genes (Figure 7.3.1). This genome-wide correlation together with our previous results from the gene ontology (GO) analysis strongly favors the notion that, as its induction occurs, Brachyury targets enhancer regions of pro-mesodermal genes and regulates their proper transcriptional activation. Consistent with this model, regions bound by Brachyury (8 out of the 12 tested) efficiently elevated the expression of a fire-fly luciferase reporter in synergy with Brachyury (a range of 2 to 25 fold increases compared to RFP controls) in human 293T cells (Figure 7.3.2).

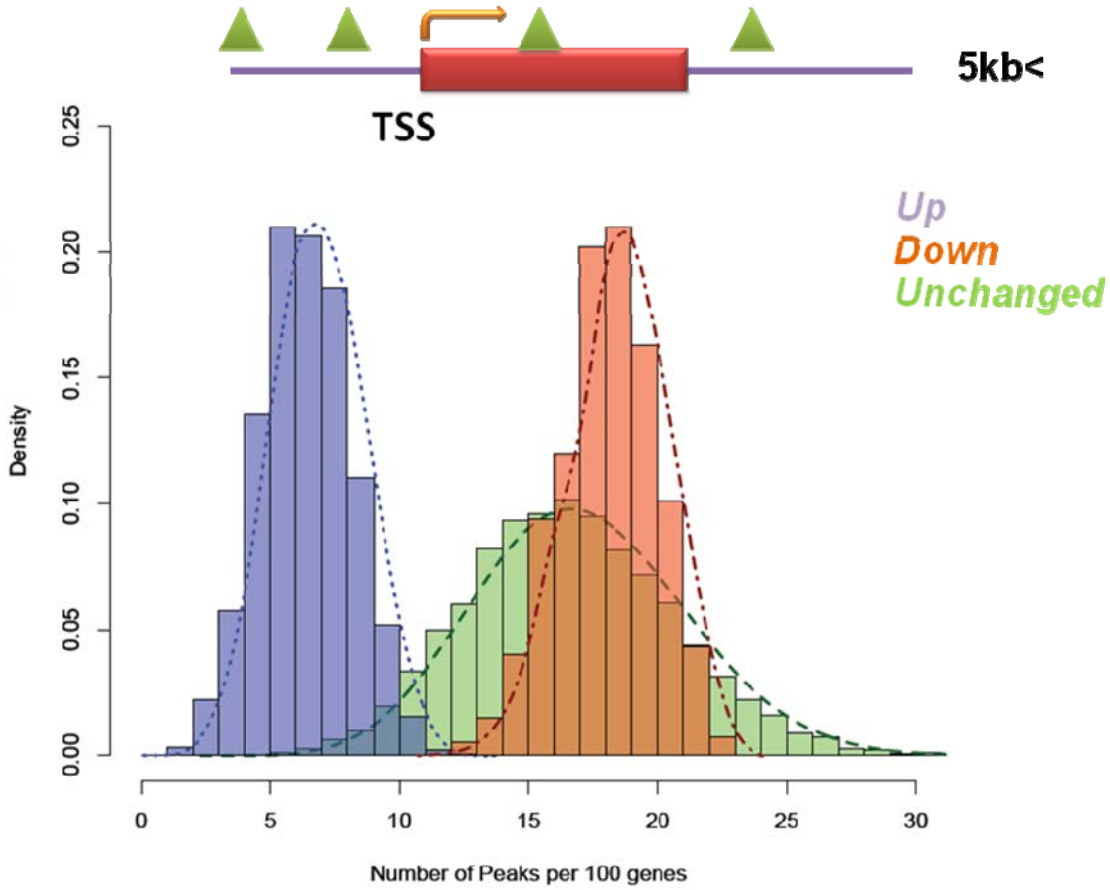


Figure 7.3.1, Number of Peaks associated with 100 random-sampled up-regulated (light blue), down-regulated (yellow) or unchanged genes (light green). (For Peak-to-Gene association: See *Chapter 8*)

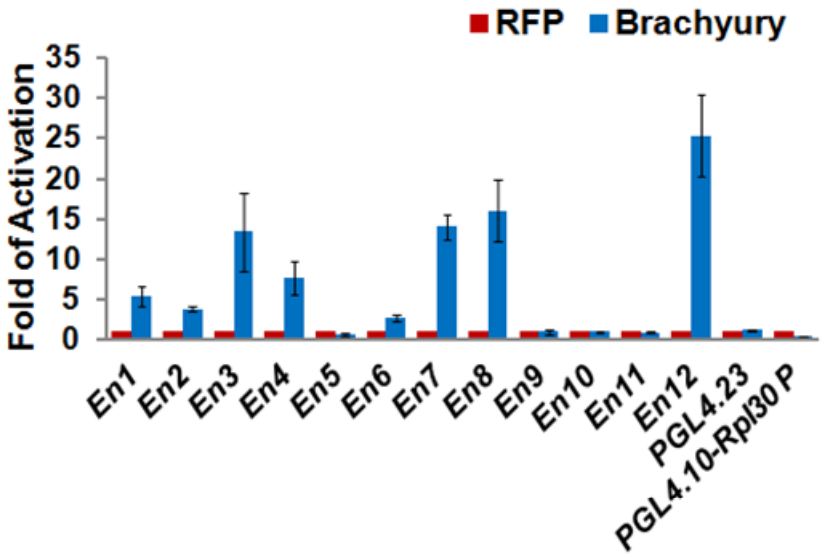


Figure 7.3.2, Luciferase reporter assays to assess the ability of 12 Brachyury-bound regions in directing transcriptional activation in synergy with Brachyury. RFP served as a transfection control.

7.4 Purification of Brachyury-associated Factors

Our previous results established that Brachyury is required for the efficient expression of its targets genes but did not provide a transcriptional activation mechanism. To address this, an ES cell line that stably expresses the Flag-GFP tagged Brachyury has been generated. After confirming that this fusion protein was properly nuclear-localized (Figure 7.4.1) and was competent in directing transcription activation in a luciferase reporter assay (Figure 7.4.2), we have now scaled up cell culture and obtained enough materials for purification of factors associated with Brachyury. Preliminary MudPit results were shown in Table 7.4.1 However, further experiments should be performed to study whether these protein factors were also associated with endogenous Brachyury protein during stem cell differentiation and by which mechanisms they contribute to the Brachyury-mediated molecular regulations.

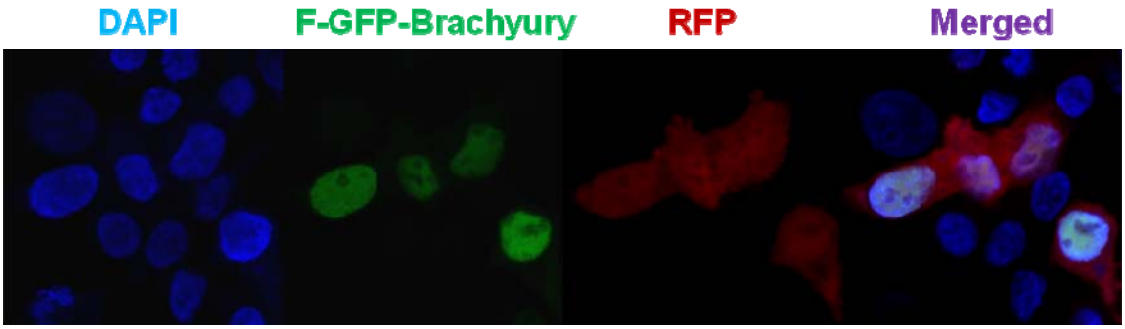


Figure 7.4.1, Nuclear-localization of Flag-GFP tagged Brachyury (Blue, DAPI; Green, GFP; Red, DsRed)

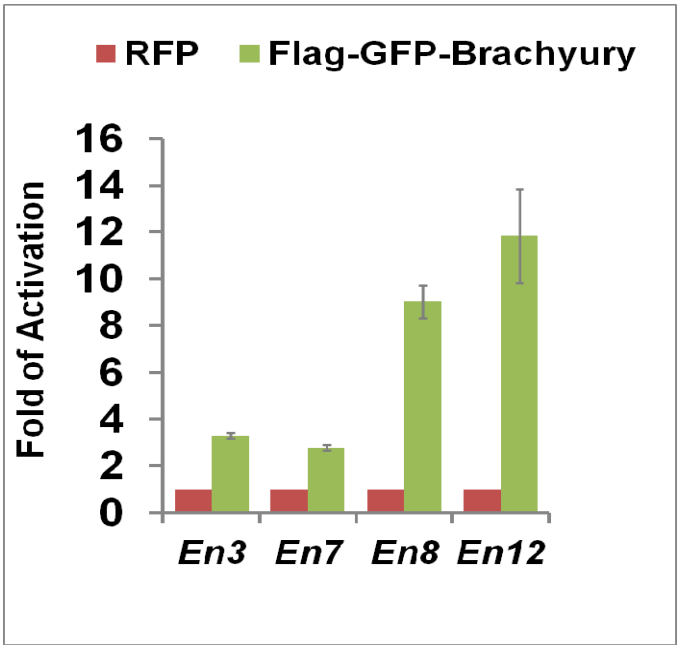


Figure 7.4.2, Flag-GFP tagged Brachyury was able to activate transcription with all four enhancers tested.

- GFP
- Brachyury
- *carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase*
- RuvB-like protein 2
- PTB-associated splicing factor
- DEAD (Asp-Glu-Ala-Asp) box polypeptide 5
- DEAD (Asp-Glu-Ala-Asp) box polypeptide 42
- Eukaryotic translation elongation factor 1
- Tripartite motif protein 28
- T-box 19
- Actin family
- Tubulin family

Table 7.4.1, Factor associated with Brachyury after the Tandem Affinity Purification

Charpter 8 Experimental Procedures

8.1 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).

8.2 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 8.2.1 and 8.2.2). 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 8.2.3).

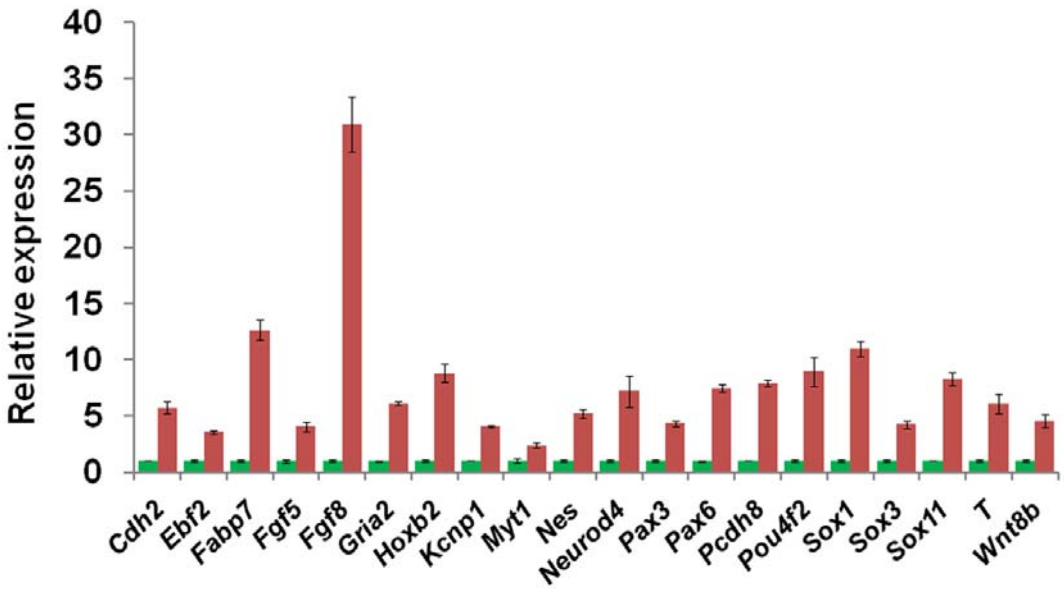


Figure 8.2.1, 20 up-regulated genes detected by mRNA-seq in shRNA A treated sample were confirmed by qRT-PCR using shRNA B treated sample.

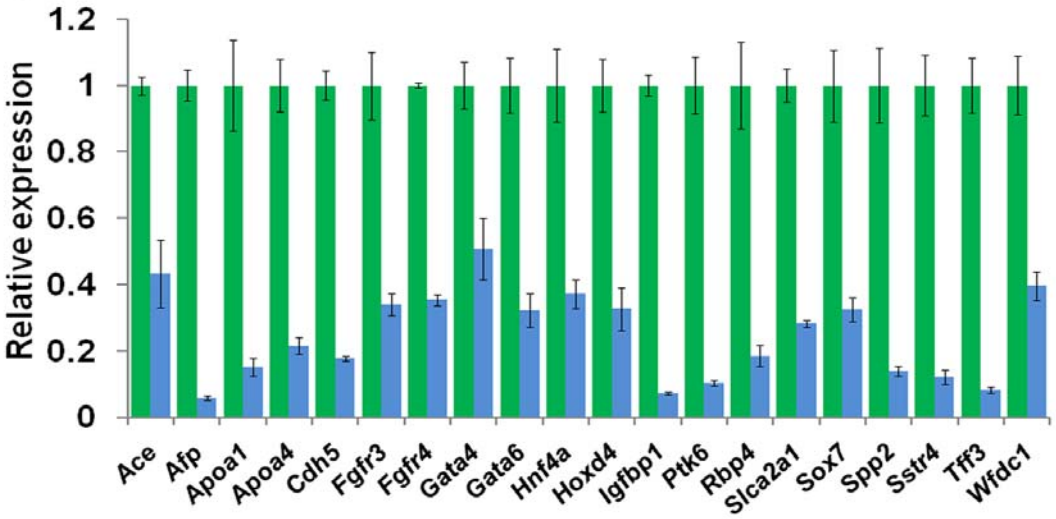


Figure 8.2.2, 20 up-regulated genes detected by mRNA-seq in shRNA A treated sample were confirmed by qRT-PCR using shRNA B treated sample.

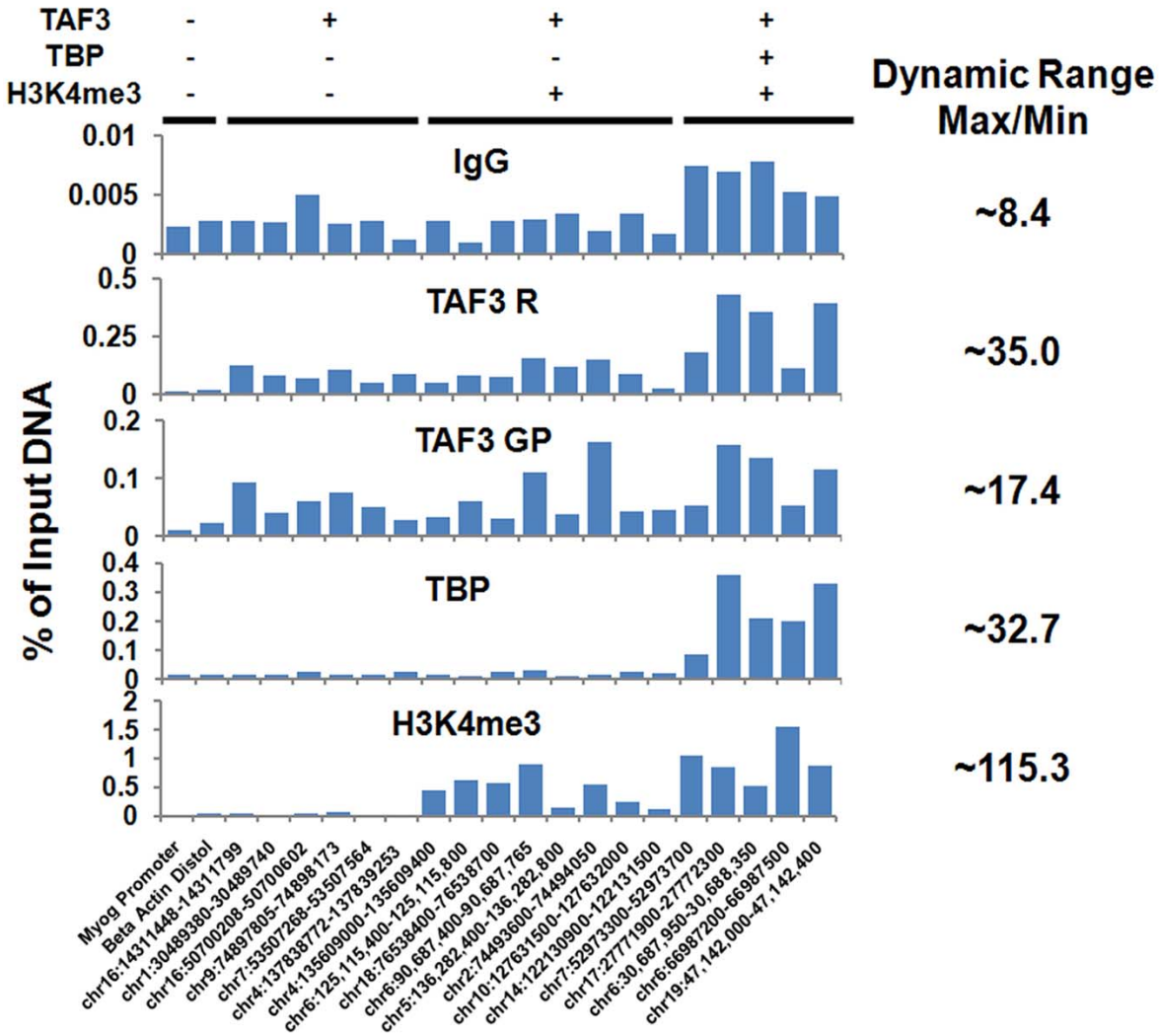


Figure 8.2.3, Enrichment of different combinations of factors as indicated at 21 genomic regions was confirmed by ChIP-qPCR. Two TAF3 antibodies were used: TAF3 R (rabbit polyclonal antibody) and TAF3 GP (guinea pig).

8.3 Antibody Generation and Purification

For antibody used in TAF3 western blot and staining, guinea pigs were immunized with 3 X flag - full length TAF3 fusion protein produced in the baculovirus expression system. For antibodies used in TAF1, TAF3 and Brachyury ChIP experiments, rabbits were immunized with residues 1823-1872 of TAF1, residues 102–175 of TAF3 or residues 249-436 GST fusion proteins. The antisera obtained were further affinity-purified using antigen (full length TAF3 or MBP tagged partial TAF1, TAF3 or Brachyury) immobilized on Affigel 10/15 resin (Bio-Rad). For Pol II ChIP, monoclonal anti-Pol II (8WG16) was concentrated from hybridoma supernatant with Protein G Sepharose beads (GE Healthcare).

8.4 Plasmid Contruction and Virus Production

Anti-TAF3 shRNA A (Sense: 5'-GCATGAACTAGAAGACTATAT-3'), anti-TAF3 shRNA B (Sense: 5'-GAGGAGGAGGAAGTCATCAAT-3'), anti-CTCF shRNA #1 (Sense: 5'-GCAGAGAAAGTAGTTGGTAAT-3') and anti-CTCF shRNA #2 (Sense: 5'-CAGCAGTGTACAGATGGTAAT-3') were cloned into pLKO.1 - TRC Cloning Vector (Moffat et al., 2006). For infecting 46C ES cells that carry *PuroR* at *Sox1* locus, the *PuroR* ORF in PLKO.1 vector was replaced with *NeoR* using BamHI and KpnI sites. Lentiviral particles were generated by transfecting TLA-HEK293T cells (Open Biosystems) with a PLKO.1 shRNA vector and packaging vectors, psPAX2 and pMD2.G (Addgene), according to the RNAi consurtium protocol (Moffat et al., 2006). Viruses were concentrated using Fast-Trap Lentivirus Purification and Concentration Kit (Millipore). The titer of virus stock was determined using QuickTiter™ Lentivirus Titer Kit (Cell Biolabs. Inc).

8.5 Lentivirus Mediated RNA Interference

For infection, ES cells were incubated with lentiviral particles (See *Plasmid Contruction and Virus Production*) at a multiplicity of infection of 40 in ES cell medium containing 8μg/ml polybrene (Sigma) for 24 hours. Subsequently, cells were subjected to either 1μg/ml puromycin or 250μg/ml geneticin selection.

8.6 Western Blots

Whole cell extracts from ES Cells and EBs were isolated using RIPA buffer that contained Complete Protease Inhibitor Cocktail (Roche). Protein concentrations were measured using Bio-Rad Protein Assay against BSA standards. Protein from each sample was resolved by SDS-PAGE. Primary antibodies used: Oct4 (sc-5279X, Santa Cruz, 1:2000), Nanog (A300-398A, Bethyl, 1:5000), TBP (ab62126, Abcam, 1:5000), TAF1 (rabbit polyclonal, 2µg/ml), TAF3 (guinea pig polyclonal, 1µg/ml), TAF4 (612054, BD Biosciences, 1:250), Tubb3 (MAB1637, Millipore, 1:1000), Afp (AF5369, R&D Systems, 1 µg/ml), Myog (556358, BD Biosciences,1:250), beta-tubulin (E7), beta-actin (A2228, Sigma, 1:2000), CTCF (07-729, Millipore, 1:2000), Smc1a (A300-055A, Bethyl, 1:2000), Smc3 (ab9263, abcam, 1:2000), Med12 (A300-774A, Bethyl, 1:2000) and HA-Tag (ab13834, abcam, 1:500). HRP conjugated secondary antibodies (Pierce) were used at a dilution of 1:5000. Western Lightning Plus–ECL (Perkin) was used for chemiluminescent detection.

8.7 Quantitative-RT-PCR

Total RNA was extracted with Trizol LS Reagent (Invitrogen), and reverse-transcribed by SuperScript III First-Strand Synthesis System with oligo-dT primer (Invitrogen). cDNA corresponding to 10ng of total RNA was used in each iQ Sybr Green Supermix (Bio-Rad) reaction. Reactions were performed in triplicates on a DNA Engine Opticon 2 Real Time cycler. Cycle threshold values ($\log_2$ scale) were obtained in MJ opticon analysis software (Bio-Rad). Housekeeping genes, *beta-actin* and *Gapdh*, served as internal controls. Relative gene expression was computed for each sample by comparing to control ES cells.

8.8 In vitro Differentiation of ES cells

Embryoid bodies were formed by plating cells (150,000cells/ml) in ES cell medium without LIF on 6 well Ultra Low Cluster Plate (Costar) for the indicated number of days. Medium was renewed every other day. Serum free *in vitro* differentiation of EBs was performed as previously described (Gadue et al., 2006) with minor modifications. Specifically, cells were plated at a concentration of 75,000cells/ml. After 2 days, 100ng/ml Activin A (R&D) was added into medium without dissociation/re-aggregation. At EB day4, definitive endoderm and mesoderm cells were analyzed by flow cytometry. Or, EBs were harvested for Brachyury ChIP-seq experiments.

8.9 Flow Cytometric Analysis

The percentages of G1, S, and G2/M cells in a population were determined by Click-iT EdU Alexa Fluor 488 Flow Cytometry Assay Kit (Invitrogen). Specifically, cells were pulsed with 10μM 5-ethynyl-2'-deoxyuridine for 30mins. The percentage of SSEA1+/Oct4+ cells in a population was measured using Human and Mouse Pluripotent Stem Cell Analysis Kit (BD Biosciences). For analyzing definitive endoderm ($CXCR4^+c\text{-}kit^+$; (Gouon-Evans et al., 2006)) and mesoderm ($CXCR4^+Flk\text{-}1^+$; (Nelson et al., 2008)) cells, EBs were dissociated with Trypsin/EDTA and stained using the following antibodies: PE rat anti-mouse CD117(c-kit), PE-Cy7 rat anti-mouse Flk-1and APC rat anti-mouse CD184 (CXCR4) (BD Pharmingen). Samples were analyzed on an LSR II (BD Biosciences). Data analysis was performed with FlowJo Software (Tree Star Inc.).

8.10 Histochemistry

For alkaline phosphatase staining, cells (R1) were plated onto feeder cells (MEF). After 3 days, ES cell colonies were stained with Alkaline Phosphatase Detection Kit (Millipore) and counted. For immunostaining, ES cells or EBs were cultured on chamber glass slides pre-coated with Matrigel (BD Biosciences) or Laminin (Sigma). Samples were fixed with 4% paraformaldehyde, permeabilized with PBST (PBS plus 0.2% Triton X-100) and blocked with 10% FCS and 1% BSA in PBST. Samples were stained with primary antibody in blocking solution. Alternatively, whole mount staining was performed for EBs in suspension. For haematoxylin and eosin staining, EBs were fixed, cyro-embedded (Tissue-Tek O.C.T. compound) and sectioned (20 µm). Primary antibodies used: Oct4 (sc-5279X, Santa Cruz, 1:1000), Nanog (560278, BD Biosciences, 1:5, Alexa Fluor 488 conjugated), Tubb3 (MAB353, Chemicon, 1:200), TAF3 (guinea pig polyclonal, 5µg/ml), Afp (AF5369, R&D Systems, 10 µg/ml), Gata4 (sc1237X, Santa Cruz, 1:500), and Tubb3 (MAB1637, Millipore, 1:50), Secondary antibodies: DyLight 488/549 conjugated secondary antibodies (anti-mouse, anti-goat, or anti- guinea pig, 1:400, Jackson ImmunoResearch). Nuclei were counterstained with DAPI.

8.11 Teratoma Formation Assays

2×10^6 clonally selected control (non-target) or TAF3 K/D (shRNA A clone 2 or shRNA B clone 1) cells were injected subcutaneously in the flanks of 8 week old SCID mice (NOD.CB17-*Prkdc*^{scid}/J, the Jackson Laboratory). Tumour tissue samples developed after 6 weeks were fixed overnight in 4% paraformaldehyde and then embedded into paraffin. Sections were stained with haematoxylin and eosin using a standard protocol (Sigma, Procedure No.GHS). Total RNA and protein were isolated from fine tissue powder made by grinding liquid nitrogen snap-frozen teratomas with mortar and pestle.

8.12 Gel Filtration of ES Cell Nuclear Extract

ES cell (D3) nuclear extract was prepared as described (Dignam et al., 1983) with modified Buffer C (20mM Hepes, pH 7.9, 25% glycerol, 0.6M KCl, 15mM MgCl2, 0.2mM EDTA, 1mM DTT, 0.5mM PMSF). 1 ml of nuclear extract after dialysis against Running Buffer (20mM Hepes, pH 7.9, 10% glycerol, 0.2M NaCl, 0.1% NP-40, 0.2mM EDTA, 1mM DTT, 0.5mM PMSF) was resolved by a calibrated pre-packed Superose 6 column (GE Heathcare). Fractions were blotted with anti-TAF3 and anti-TBP antibodies.

8.13 Sample Preparation for mRNA-seq

ES cells (D3) were infected with lentivirus carrying either non-target shRNA or anti-TAF3 shRNA A. At 72 hours post infection (90% of TAF3 was depleted as judged by western blot), control and TAF3 K/D cells were then induced to form EBs. Total RNAs were isolated at three time points (ES cell, EB day 3 and day 6).

8.14 Chromatin Immunoprecipitation

Chromatin Immunoprecipitation (ChIP) was performed according to (Boyer et al., 2006) with minor modifications. Briefly, cross-linked ES cell (D3) chromatin was sheared using Covaris S2 system to a size range of 100bp~ 400bp. Immunoprecipitation was conducted with either a specific antibody or mock IgG conjugated Protein A (rabbit and guinea pig) or G (mouse) Sepharose beads (GE Healthcare). The reverse cross-linking was performed at 70^{o}C overnight. After RNAse A and Proteinase K treatment, sample was deproteinized with UltraPure Phenol: Chloroform: Isoamyl Alcohol (Invitrogen) and further purified with Qiaquick PCR purification kit (Qiagen). DNA concentration was measured with Quant-iT™ PicoGreen dsDNA Reagent (invitrogen). Monoclonal anti-TBP (ab62126, Abcam) and rabbit polyclonal anti-H3K4me3 (ab8580, Abcam) were used for ChIP.

8.15 Deep Sequencing Library Preparation

8μg of the total RNA from each sample (See *Sample Preparation for mRNA-seq*) was convert to mRNA-seq library using mRNA-Seq Sample Prep Kit (Illumina). mRNA-seq libraries for the biological replicate were generated independently following the same procedures. About 10ng of ChIP enriched DNA (See *Chromatin Immunoprecipitation*) was converted to ChIP-seq library using ChIP-Seq DNA Sample Prep Kit (Illumina).

8.16 Expression Level Estimation, Differential Expression Testing and Gene Ontology Analysis

Two biological replicates each of shRNA-treated and control libraries were sequenced in 36bp single-end format (SE reads; 1 lane Illumina GAII per sample) at each of three timepoints (See *Sample Preparation for mRNA-seq Library*). Reads with a unique best match to the mouse genome (mm9) according to Bowtie (--best --strata -m 1; (Langmead et al., 2009)) were retained for differential expression testing. We also sequenced two sets of ES cell samples in 76bp paired-end format (PE reads; 1 lane Illumina GAII per sample). Transcript isoforms were reconstructed from PE reads using Tophat (Trapnell et al., 2009) and abundances estimated using Cufflinks (Trapnell et al., 2010). In addition, we trimmed 10bp from the start and end of the first PE read, the remainder of which was then mapped and retained for differential expression testing as described for SE reads. Read counts were tallied for each Ensembl annotated protein-coding gene (Ensembl 56; (Flicek et al., 2010); http://sep2000.archive.ensembl.org) incremented by 1 and differential expression tested using EdgeR (Robinson et al., 2010) using all qualified samples at each timepoint. Genes significant at the 5% level after Benjamini-Hochberg correction were considered significant. Gene Ontology (Ashburner et al., 2000) analysis was performed separately on up-regulated and down-regulated genes at each timepoint using GOSeq (Young et al., 2010) with Wallenius gene-length correction. Categories significant at the 5% level after Benjamini-Hochberg correction were considered significant.

8.17 Tissue Bias Metric

Tags from tissue- and stage-specific mouse SAGE libraries (Khattra et al., 2007) mapped perfectly and uniquely to the mouse genome (mm9) by Bowtie were grouped by the gene they mapped to (Ensembl 56). Genes that contained tags from only one of the two libraries in a given comparison were considered library-specific markers. The number of library-specific markers was forced to be equal between libraries. To estimate TAF3 K/D associated tissue bias, we ordered genes by the fold change in expression and counted the number of library-specific markers in overlapping 200 gene windows along the expression change axis. Counts were Z-score transformed and smoothed using a lowess smoothing fraction of 0.1 prior to plotting.

8.18 Comparison between TAF3 K/D and Sox17 O/E Differentially Expressed Genes

GSE10809 was downloaded from the NCBI GEO database (Seguin et al., 2008) in series matrix format, quantile normalized using the 'preprocessCore' Bioconductor library (Gentleman et al., 2004) and analyzed using the SAM implementation 'Siggenes' (Schwender and Ickstadt, 2008). Probes with a delta value ≥1 were considered significantly differentially regulated. Probes were mapped from HG-U133_Plus_2 microarray platform to mouse Ensembl 56 using the Bioconductor library 'annotationTools' (Kuhn et al., 2008). The total number of mappable orthologous genes was 14,400 and this value was used to compute the hypergeometric P-value of the observed overlaps between significantly up-/down-regulated genes in the Sox17 O/E and TAF3 K/D experiments.

8.19 ChIP-seq Peak Calling, Bound Region Definition and Enrichment Calculations

We sequenced chromatin-IP and control libraries (See *Chromatin Immunoprecipitation*) in 36bp single-end format for each of TAF3, TAF1, TBP and Pol II (2 lanes Illumina GAII per sample). Reads were mapped to the mouse genome (mm9) using Eland (eland_extended; Illumina, Inc.) and peaks called using MACS (1.4.0beta; (Zhang et al., 2008)) at an approximate 1% FDR ("bound regions"). We merged overlapping bound regions (minimum overlap 1bp; union of all bases retained) to produce a set of 43,910 regions statistically enriched for one or more of the factors we assayed ("44K dataset"). We computed the enrichment of all four factors in all 45,799 regions using a modified version of MACS (1.4.0beta; request from D.S.) that operates on arbitrary genomic regions. We further merged the 44K dataset with overlapping bound regions for a variety of other factors (Chen et al., 2008; Kagey et al., 2010; Marson et al., 2008; Visel et al., 2009) to obtain a dataset of 124,558 regions statistically enriched for one or more factors ("125K dataset"). Within each of these regions we computed the enrichment of all factors and various histone modifications (Mikkelsen et al., 2007) by supplying the authors original read mappings to our modified version of MACS. mm8 coordinates were converted to mm9 using liftOver (Hinrichs et al., 2006).

8.20 Principal Components Analysis and Bound Region Clustering

We unit normalized the enrichments for each factor across all bound regions in the 125K dataset (See *ChIP-seq Peak Calling, Bound Region Definition and Enrichment Calculations*) and performed principal components analysis on the resulting values using the princomp method in *R*. We clustered all bound regions that overlapped a region determined to be statistically enriched for TAF3 binding by MACS on the first three principal components using the kmeans clustering method in *R* (centers=4, iter.max=20, nstart=20). The number of clusters was chosen using the Calinski-Harabasz index as implemented in the pamk method in the *R* library 'fpc'.

8.21 Testing Overlap Between TAF3 Bound Regions and Other Factors

We tested for a concentration of TAF3 bound regions in regions bound by other factors (Chen et al., 2008; Kagey et al., 2010; Marson et al., 2008; Mikkelsen et al., 2007; Visel et al., 2009) by comparing the observed number of overlapping regions to a null distribution obtained by randomizing the location of TAF3 peaks. We controlled for distance biases with respect to transcription start sites and mappabilty (Rozowsky et al., 2009) by first associating each bound region with a panel of other regions that are a similar distance from the nearest TSS. We then sampled from this panel based on their mappability computed for 30bp matches on a 1Kb scale. The resulting distributions closely approximate the real data.

8.22 Association between Differentially Expressed Genes and TAF3 Bound Regions

We used GREAT ((McLean et al., 2010); basalPlusExtension - maxExtension=100000) to define regulatory regions around protein-coding genes (Ensembl 56). We then computed the fraction of the total regulatory space that fell within the regulatory domains of the 200 most up- or down-regulated genes following TAF3 K/D. Similarly, we computed the fraction of TAF3 bound regions in each of the four classes (See *Principal Components Analysis and Bound Region Clustering*) that fell within the regulatory domains of the 200 most up- or down-regulated genes. We used the ratio of these numbers ("bound region density") as our test statistic. We generated a null distribution by recomputing our test statistic for 100 sets of sampled genes that had similar level of expression in wild-type ES cells and a comparable amount of surrounding non-coding DNA as the original genes (*i.e.* 200 up-/down-regulated genes). For each controlled parameters we sorted genes by the relevant statistic and divided the data into twelve equally-sized bins, from which all random samples were subsequently drawn. We expressed the relationship between the real data and the null distribution as a Z-score.

8.23 TAF3 Immunoprecipitation

Preimmune IgG and TAF3 antibody (guinea pig) were first conjugated to protein A beads. IP were performed using 0.5ml of nuclear extract after dialysis against IP Buffer (20mM Hepes, pH 7.9, 10% glycerol, 0.2M NaCl, 0.5% NP-40, 0.2mM EDTA, 1mM DTT, 0.5mM PMSF) at 4 ^{o}C overnight. 4 washes with 1ml IP Buffer were then performed. For B+E treatment, one benzonase wash (250u/ml) in Digestion Buffer (50 mM Tris-HCl pH 8.0, 100mM NaCl, 1 mM $MgCl_2$) and one ethidium bromide wash (0.1mg/ml) in IP buffer were performed after the initial 4 washes. Final IP products were resolved by SDS-PAGE and were then blotted against indicated antibodies.

8.24 Mapping the CTCF-interacting Domain of TAF3 protein

293T cells were co-transfected with PCAG plasmids over-expressing Flag-HA tagged full length/truncated TAF3/RFP and CTCF proteins. After 2 days, cells were harvested and lysed with Co-IP Buffer (20mM Hepes, pH 7.9, 10% glycerol, 0.3M NaCl, 0.5% NP-40, 0.2mM EDTA, 1mM DTT, 1 X Roche Complete Protease Inhibitor) by passing through 25G syringe needle for 10 times. Flag IPs were performed overnight at 4 ^{o}C followed by 5 washes with 1ml Co-IP Buffer. The relative enrichment of TAF3 or CTCF in IPs were analyzed by anti-HA or anti-CTCF western blot. Coomassie staining was further used to assess the association between Flag-HA-TAF3 (501-730aa) and CTCF.

8.25 Testing the Interdependence of Recruitment between TAF3 and CTCF

We computationally identified 3 distinct types of peak, 1) 6 promoter regions highly enriched for core promoter factors (TAF3, TBP and Pol II), 2) 11 promoter distal regions for TAF3 and CTCF but with low levels of other core promoter factors (TBP, TAF1 and Pol II) and 3) 6 regions for CTCF but with low levels of TAF3 (Table S3). ChIP-qPCR analysis was performed to compare the enrichment of TAF3 at type 1 and 2 regions in control and CTCF K/D (shRNA #1) cells. Similarly, the relative enrichment of CTCF was measured at type 2 and 3 regions in control and TAF3 K/D (shRNA A2) cells. We normalized the amount of input chromatin and antibody used for each ChIP reaction. The ChIP signals obtained were first normalized to % of input DNA and were then compared between samples.

8.26 Chromatin Conformation Capture (3C) Experiments

3C experiments were performed as previously described (Hagege et al., 2007; Miele and Dekker, 2009). Briefly, cells (non-target and shRNA A2) in single-cell suspension were cross-linked with 1.6% formaldehyde for 10mins at room temperature. Reactions were quenched by 250mM Glycine. Chromatin per 1 X 10^7 cells was digested with DpnII (1000U) or HaeIII (1000U) at 37 ^oC overnight. Digestion efficiency was assessed by qPCR. Above 90% digestion efficiencies were achieved. After inactivation of the restriction enzyme at 65^oC for 30mins, chromatin fragments were ligated with T4 DNA ligase (100 Weiss Units, M1794, Promega) at 4^oC for 4 hours. The reverse crosslinking was performed at 65^oC overnight in the presence of 300ug/ml Proteinase K. After RNAse A treatment, sample was then deproteinized with UltraPure Phenol: Chloroform: Isoamyl Alcohol and DNA was harvested by ethanol precipitation. 200ng of DNA was used for each qPCR reaction in the triplicate. To prepare 3C control template, BAC DNA (*Mapk3*, RP23-245M8 or *Psmd1*, RP24-151H21) was digested with HaeIII or Sau3AI (isoschizomer of DpnII) at 37^oC overnight and was then ligated with T4 DNA ligase at room temperature for 2 hours. 200ng of ligated BAC DNA was used for each qPCR reaction as the BAC control. Each data point was first corrected for bias of PCR amplification by dividing the average of three PCR signals by the average signal in the BAC control template (percentile scale). To compare between samples, data from control and TAF3 K/D cells were normalized to each other using the interaction frequencies between fragments in control regions (Gene_desert 1, 2 & 3 for *Mapk3*; Gene_desert 4, 5 & 6 for *Psmd1*). Values shown were averages from two biological replicates.

8.27 Brachyury-bound Regions and Differential Gene Expression Association

Brachyury binding sites that are within 5kb from the TSS were assigned to *Ensembl 61* annotated genes. Then, 100 genes were randomly sampled from genes that were down-regulated, up-regulated or unchanged upon Brachyury depletion. The total number of Brachyury sites associated with the 100 genes was calculated. The histograms of 1000 such simulations for each differential gene expression group was plotted in Figure 7.3.1

8.28 Tandem Affinity Purification of Brachyury-associated Factors

Total 80X p15 plates of an ES cell line expressing Flag-GFP tagged Brachyury were harvested. Whole cell lysate was made by douncing cells 40 times in the Extraction Buffer that contains 20mM Hepes (pH 7.6), 300mM NaCl, 1mM MgCl2, 0.2mM EDTA and 0.1% NP-40 alternative. After ultra-centrifugation, supernatant was further dialysed against the IP Buffer that contains 150mM NaCl (other compositions were the same to the Extraction Buffer). A 4-hour Anti-Flag IP was performed and followed by 4 washes with the IP buffer and then a Flag-peptide (200μg/ml) elution. The procedures of the anti-GFP IP were identical to the steps of the anti-Flag IP. After the anti-GFP pull-down, the IP product was eluted with 80μl of the Elution Buffer (50mM Glycine, pH 2.5, 150mM NaCl) and was neuralized to pH 7.8 with 100mM Tris-Cl.

References

Arnold, S.J., and Robertson, E.J. (2009). Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. Nat Rev Mol Cell Biol *10*, 91-103.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., *et al.* (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet *25*, 25-29.

Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K., *et al.* (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature *441*, 349-353.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., *et al.* (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell *133*, 1106-1117.

Cheng, Y., Buffone, M.G., Kouadio, M., Goodheart, M., Page, D.C., Gerton, G.L., Davidson, I., and Wang, P.J. (2007). Abnormal sperm in mice lacking the Taf7l gene. Mol Cell Biol *27*, 2582-2589.

Chopra, V.S., Srinivasan, A., Kumar, R.P., Mishra, K., Basquin, D., Docquier, M., Seum, C., Pauli, D., and Mishra, R.K. (2008). Transcriptional activation by GAGA factor is through its direct interaction with dmTAF3. Dev Biol *317*, 660-670.

Clapier, C.R., and Cairns, B.R. (2009). The biology of chromatin remodeling complexes. Annu Rev Biochem *78*, 273-304.

Cunliffe, V., and Smith, J.C. (1992). Ectopic mesoderm formation in Xenopus embryos caused by widespread expression of a Brachyury homologue. Nature *358*, 427-430.

D'Alessio, J.A., Wright, K.J., and Tjian, R. (2009). Shifting players and paradigms in cell-specific transcription. Mol Cell *36*, 924-931.

Deato, M.D., Marr, M.T., Sottero, T., Inouye, C., Hu, P., and Tjian, R. (2008). MyoD targets TAF3/TRF3 to activate myogenin transcription. Mol Cell *32*, 96-105.

Deato, M.D., and Tjian, R. (2007). Switching of the core transcription machinery during myogenesis. Genes Dev *21*, 2137-2149.

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.

Fernando, R.I., Castillo, M.D., Litzinger, M., Hamilton, D.H., and Palena, C. (2011). IL-8 signaling plays a critical role in the epithelial-mesenchymal transition of human carcinoma cells. Cancer Res *71*, 5296-5306.

Fernando, R.I., Litzinger, M., Trono, P., Hamilton, D.H., Schlom, J., and Palena, C. (2010). The T-box transcription factor Brachyury promotes epithelial-mesenchymal transition in human tumor cells. J Clin Invest *120*, 533-544.

Flicek, P., Aken, B.L., Ballester, B., Beal, K., Bragin, E., Brent, S., Chen, Y., Clapham, P., Coates, G., Fairley, S., *et al.* (2010). Ensembl's 10th year. Nucleic Acids Res *38*, D557-562.

Freiman, R.N., Albright, S.R., Zheng, S., Sha, W.C., Hammer, R.E., and Tjian, R. (2001). Requirement of tissue-selective TBP-associated factor TAFII105 in ovarian development. Science *293*, 2084-2087.

Gadue, P., Huber, T.L., Paddison, P.J., and Keller, G.M. (2006). 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 *103*, 16806-16811.
Gangloff, Y.G., Pointud, J.C., Thuault, S., Carre, L., Romier, C., Muratoglu, S., Brand, M., Tora, L., Couderc, J.L., and Davidson, I. (2001). 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 *21*, 5109-5121.
Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J.*, et al.* (2004). Bioconductor: open software development for computational biology and bioinformatics. Genome Biol *5*, R80.
Goodrich, J.A., and Tjian, R. (2010). Unexpected roles for core promoter recognition factors in cell-type-specific transcription and gene regulation. Nat Rev Genet *11*, 549-558.
Gouon-Evans, V., Boussemart, L., Gadue, P., Nierhoff, D., Koehler, C.I., Kubo, A., Shafritz, D.A., and Keller, G. (2006). BMP-4 is required for hepatic specification of mouse embryonic stem cell-derived definitive endoderm. Nat Biotechnol *24*, 1402-1411.
Grapin-Botton, A., and Constam, D. (2007). Evolution of the mechanisms and molecular control of endoderm formation. Mech Dev *124*, 253-278.
Grontved, L., Madsen, M.S., Boergesen, M., Roeder, R.G., and Mandrup, S. (2010). MED14 tethers mediator to the N-terminal domain of peroxisome proliferator-activated receptor gamma and is required for full transcriptional activity and adipogenesis. Mol Cell Biol *30*, 2155-2169.
Hagege, H., Klous, P., Braem, C., Splinter, E., Dekker, J., Cathala, G., de Laat, W., and Forne, T. (2007). Quantitative analysis of chromosome conformation capture assays (3C-qPCR). Nat Protoc *2*, 1722-1733.
Hart, D.O., Santra, M.K., Raha, T., and Green, M.R. (2009). Selective Interaction Between Trf3 and Taf3 Required for Early Development and Hematopoiesis. Dev Dynam *238*, 2540-2549.
Herrmann, B.G., Labeit, S., Poustka, A., King, T.R., and Lehrach, H. (1990). Cloning of the T gene required in mesoderm formation in the mouse. Nature *343*, 617-622.
Hinrichs, A.S., Karolchik, D., Baertsch, R., Barber, G.P., Bejerano, G., Clawson, H., Diekhans, M., Furey, T.S., Harte, R.A., Hsu, F.*, et al.* (2006). The UCSC Genome Browser Database: update 2006. Nucleic Acids Res *34*, D590-598.
Hochheimer, A., and Tjian, R. (2003). Diversified transcription initiation complexes expand promoter selectivity and tissue-specific gene expression. Genes Dev *17*, 1309-1320.
Jaenisch, R., and Young, R. (2008). Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. Cell *132*, 567-582.
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.
Khattra, J., Delaney, A.D., Zhao, Y., Siddiqui, A., Asano, J., McDonald, H., Pandoh, P., Dhalla, N., Prabhu, A.L., Ma, K.*, et al.* (2007). Large-scale production of SAGE libraries from microdissected tissues, flow-sorted cells, and cell lines. Genome Res *17*, 108-116.
Kuhn, A., Luthi-Carter, R., and Delorenzi, M. (2008). Cross-species and cross-platform gene expression studies with the Bioconductor-compliant R package 'annotationTools'. BMC Bioinformatics *9*, 26.

Lago, C., Clerici, E., Mizzi, L., Colombo, L., and Kater, M.M. (2004). TBP-associated factors in Arabidopsis. Gene *342*, 231-241.
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.
Levine, M., and Tjian, R. (2003). Transcription regulation and animal diversity. Nature *424*, 147-151.
Li, L., Sun, L., Gao, F., Jiang, J., Yang, Y., Li, C., Gu, J., Wei, Z., Yang, A., Lu, R.*, et al.* (2010). Stk40 links the pluripotency factor Oct4 to the Erk/MAPK pathway and controls extraembryonic endoderm differentiation. Proc Natl Acad Sci U S A *107*, 1402-1407.
Li, X.Y., MacArthur, S., Bourgon, R., Nix, D., Pollard, D.A., Iyer, V.N., Hechmer, A., Simirenko, L., Stapleton, M., Luengo Hendriks, C.L.*, et al.* (2008). Transcription factors bind thousands of active and inactive regions in the Drosophila blastoderm. PLoS Biol *6*, e27.
Liu, W.L., Coleman, R.A., Ma, E., Grob, P., Yang, J.L., Zhang, Y., Dailey, G., Nogales, E., and Tjian, R. (2009). Structures of three distinct activator-TFIID complexes. Genes Dev *23*, 1510-1521.
MacArthur, S., Li, X.Y., Li, J., Brown, J.B., Chu, H.C., Zeng, L., Grondona, B.P., Hechmer, A., Simirenko, L., Keranen, S.V.*, et al.* (2009). 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 *10*, R80.
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.
McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol *28*, 495-501.
Meshorer, E., and Misteli, T. (2006). Chromatin in pluripotent embryonic stem cells and differentiation. Nat Rev Mol Cell Biol *7*, 540-546.
Miele, A., and Dekker, J. (2009). Mapping cis- and trans- chromatin interaction networks using chromosome conformation capture (3C). Methods Mol Biol *464*, 105-121.
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.
Moffat, J., Grueneberg, D.A., Yang, X., Kim, S.Y., Kloepfer, A.M., Hinkle, G., Piqani, B., Eisenhaure, T.M., Luo, B., Grenier, J.K.*, et al.* (2006). A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. Cell *124*, 1283-1298.
Morey, L., and Helin, K. (2010). Polycomb group protein-mediated repression of transcription. Trends Biochem Sci *35*, 323-332.
Naar, A.M., Lemon, B.D., and Tjian, R. (2001). Transcriptional coactivator complexes. Annu Rev Biochem *70*, 475-501.
Naiche, L.A., Harrelson, Z., Kelly, R.G., and Papaioannou, V.E. (2005). T-box genes in vertebrate development. Annu Rev Genet *39*, 219-239.

Nelson, T.J., Faustino, R.S., Chiriac, A., Crespo-Diaz, R., Behfar, A., and Terzic, A. (2008). CXCR4+/FLK-1+ biomarkers select a cardiopoietic lineage from embryonic stem cells. Stem Cells *26*, 1464-1473.
Palena, C., Polev, D.E., Tsang, K.Y., Fernando, R.I., Litzinger, M., Krukovskaya, L.L., Baranova, A.V., Kozlov, A.P., and Schlom, J. (2007). The human T-box mesodermal transcription factor Brachyury is a candidate target for T-cell-mediated cancer immunotherapy. Clin Cancer Res *13*, 2471-2478.
Phillips, J.E., and Corces, V.G. (2009). CTCF: master weaver of the genome. Cell *137*, 1194-1211.
Pijnappel, W.P., Kolkman, A., Baltissen, M.P., Heck, A., Jr., and Timmers, H.M. (2009). Quantitative mass spectrometry of TATA binding protein-containing complexes and subunit phosphorylations during the cell cycle. Proteome Sci *7*, 46.
Pointud, J.C., Larsson, J., Dastugue, B., and Couderc, J.L. (2001). The BTB/POZ domain of the regulatory proteins Bric a brac 1 (BAB1) and Bric a brac 2 (BAB2) interacts with the novel Drosophila TAF(II) factor BIP2/dTAF(II)155. Dev Biol *237*, 368-380.
Prince, F., Katsuyama, T., Oshima, Y., Plaza, S., Resendez-Perez, D., Berry, M., Kurata, S., and Gehring, W.J. (2008). The YPWM motif links Antennapedia to the basal transcriptional machinery. Development *135*, 1669-1679.
Robertson, K.D. (2005). DNA methylation and human disease. Nat Rev Genet *6*, 597-610.
Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139-140.
Rozowsky, J., Euskirchen, G., Auerbach, R.K., Zhang, Z.D., Gibson, T., Bjornson, R., Carriero, N., Snyder, M., and Gerstein, M.B. (2009). PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls. Nat Biotechnol *27*, 66-75.
Sarkar, D., Shields, B., Davies, M.L., Muller, J., and J, A.W. (2011). BRACHYURY confers cancer stem cell characteristics on colorectal cancer cells. Int J Cancer.
Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B., and Cavalli, G. (2007). Genome regulation by polycomb and trithorax proteins. Cell *128*, 735-745.
Schwender, H., and Ickstadt, K. (2008). Empirical Bayes analysis of single nucleotide polymorphisms. BMC Bioinformatics *9*, 144.
Seguin, C.A., Draper, J.S., Nagy, A., and Rossant, J. (2008). Establishment of endoderm progenitors by SOX transcription factor expression in human embryonic stem cells. Cell Stem Cell *3*, 182-195.
Shen, W.C., Bhaumik, S.R., Causton, H.C., Simon, I., Zhu, X., Jennings, E.G., Wang, T.H., Young, R.A., and Green, M.R. (2003). Systematic analysis of essential yeast TAFs in genome-wide transcription and preinitiation complex assembly. EMBO J *22*, 3395-3402.
Smale, S.T., and Kadonaga, J.T. (2003). The RNA polymerase II core promoter. Annu Rev Biochem *72*, 449-479.
Stott, D., Kispert, A., and Herrmann, B.G. (1993). Rescue of the tail defect of Brachyury mice. Genes Dev *7*, 197-203.
Stumpf, M., Waskow, C., Krotschel, M., van Essen, D., Rodriguez, P., Zhang, X., Guyot, B., Roeder, R.G., and Borggrefe, T. (2006). The mediator complex functions as a coactivator for GATA-1 in erythropoiesis via subunit Med1/TRAP220. Proc Natl Acad Sci U S A *103*, 18504-18509.

Taatjes, D.J., Marr, M.T., and Tjian, R. (2004). Regulatory diversity among metazoan co-activator complexes. Nat Rev Mol Cell Biol *5*, 403-410.
Tada, S., Era, T., Furusawa, C., Sakurai, H., Nishikawa, S., Kinoshita, M., Nakao, K., and Chiba, T. (2005). Characterization of mesendoderm: a diverging point of the definitive endoderm and mesoderm in embryonic stem cell differentiation culture. Development *132*, 4363-4374.
Takahashi, H., Parmely, T.J., Sato, S., Tomomori-Sato, C., Banks, C.A., Kong, S.E., Szutorisz, H., Swanson, S.K., Martin-Brown, S., Washburn, M.P.*, et al.* (2011). Human mediator subunit MED26 functions as a docking site for transcription elongation factors. Cell *146*, 92-104.
Tam, P.P., and Loebel, D.A. (2007). Gene function in mouse embryogenesis: get set for gastrulation. Nat Rev Genet *8*, 368-381.
Tatarakis, A., Margaritis, T., Martinez-Jimenez, C.P., Kouskouti, A., Mohan, W.S., 2nd, Haroniti, A., Kafetzopoulos, D., Tora, L., and Talianidis, I. (2008). Dominant and redundant functions of TFIID involved in the regulation of hepatic genes. Mol Cell *31*, 531-543.
Tjian, R., and Maniatis, T. (1994). Transcriptional activation: a complex puzzle with few easy pieces. Cell *77*, 5-8.
Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105-1111.
Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol *28*, 511-515.
Tutter, A.V., Kowalski, M.P., Baltus, G.A., Iourgenko, V., Labow, M., Li, E., and Kadam, S. (2009). Role for Med12 in regulation of Nanog and Nanog target genes. J Biol Chem *284*, 3709-3718.
Verheijen, M.H., Wolthuis, R.M., Bos, J.L., and Defize, L.H. (1999). The Ras/Erk pathway induces primitive endoderm but prevents parietal endoderm differentiation of F9 embryonal carcinoma cells. J Biol Chem *274*, 1487-1494.
Vermeulen, M., Mulder, K.W., Denissov, S., Pijnappel, W.W., van Schaik, F.M., Varier, R.A., Baltissen, M.P., Stunnenberg, H.G., Mann, M., and Timmers, H.T. (2007). Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. Cell *131*, 58-69.
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.
Wilkinson, D.G., Bhatt, S., and Herrmann, B.G. (1990). Expression pattern of the mouse T gene and its role in mesoderm formation. Nature *343*, 657-659.
Wright, K.J., Marr, M.T., 2nd, and Tjian, R. (2006). TAF4 nucleates a core subcomplex of TFIID and mediates activated transcription from a TATA-less promoter. Proc Natl Acad Sci U S A *103*, 12347-12352.
Yao, J., Fetter, R.D., Hu, P., Betzig, E., and Tjian, R. (2011). Subnuclear segregation of genes and core promoter factors in myogenesis. Genes Dev *25*, 569-580.

Ying, Q.L., Stavridis, M., Griffiths, D., Li, M., and Smith, A. (2003). Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. Nat Biotechnol *21*, 183-186.
Young, M.D., Wakefield, M.J., Smyth, G.K., and Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol *11*, R14.
Zhang, D., Penttila, T.L., Morris, P.L., Teichmann, M., and Roeder, R.G. (2001). Spermiogenesis deficiency in mice lacking the Trf2 gene. Science *292*, 1153-1155.
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.
Zhou, V.W., Goren, A., and Bernstein, B.E. (2011). Charting histone modifications and the functional organization of mammalian genomes. Nat Rev Genet *12*, 7-18.