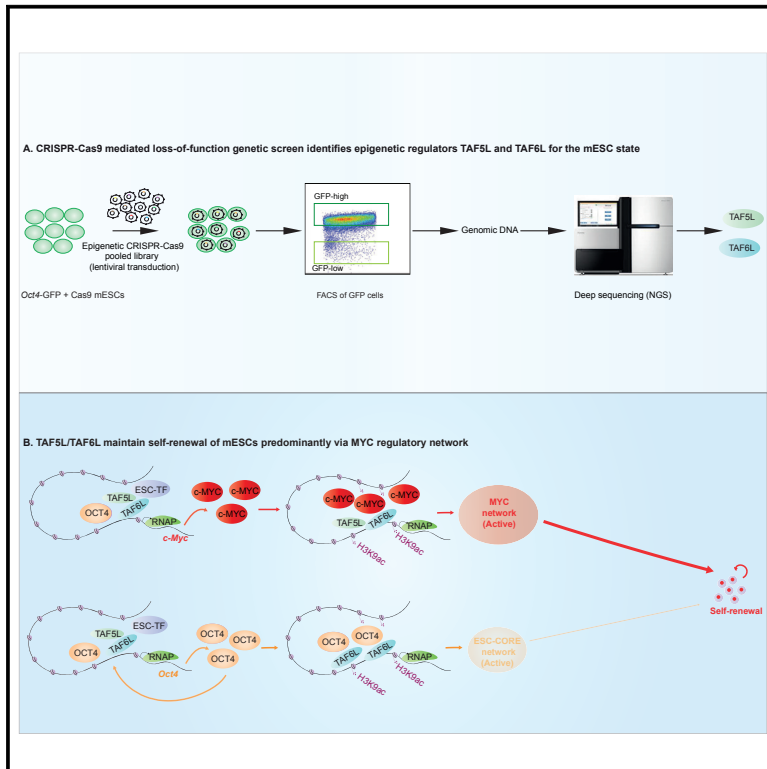# Molecular Cell

TAF5L and TAF6L Maintain Self-Renewal of Embryonic Stem Cells via the MYC Regulatory Network

Graphical Abstract

Authors

Davide Seruggia, Martin Oti, Pratibha Tripathi, ..., Jonghwan Kim, Stuart H. Orkin, Partha Pratim Das

Correspondence

orkin@bloodgroup.tch.harvard.edu (S.H.O.), partha.das@monash.edu (P.P.D.)

In Brief

Seruggia et al. identify TAF5L and TAF6L as epigenetic regulators for mESC state. Both transcriptionally activate c-Myc, Oct4, and their corresponding regulatory networks. Detailed molecular analysis uncovers that TAF5L/TAF6L predominantly regulate their target genes through H3K9ac and c-MYC recruitment that activates MYC regulatory network for self-renewal of mESCs.

Highlights

- CRISPR screen identifies epigenetic regulators TAF5L and TAF6L for mESC state

- TAF5L and TAF6L are essential for somatic reprogramming

- TAF5L/TAF6L transcriptionally activate c-Myc and Oct4 and their respective networks

- TAF5L/TAF6L predominantly regulate MYC network to control self-renewal of mESC state

CellPress

TAF5L and TAF6L Maintain Self-Renewal of Embryonic Stem Cells via the MYC Regulatory Network

Davide Seruggia,[1,10] Martin Oti,[2,3,9,10] Pratibha Tripathi,[2,3,10] Matthew C. Canver,[1] Lucy LeBlanc,[4]
Dafne C. Di Giammartino,[5] Michael J. Bullen,[2,3] Christian M. Nefzger,[2,3,6] Yu Bo Yang Sun,[2,3,6] Rick Farouni,[7]
Jose M. Polo,[2,3,6] Luca Pinello,[7] Effie Apostolou,[5] Jonghwan Kim,[4] Stuart H. Orkin,[1,8,11,12,*] and Partha Pratim Das[2,3,11,*]
[1]Division of Hematology/Oncology, Boston Children's Hospital and Department of Pediatric Oncology, Dana-Farber Cancer Institute (DFCI),
Harvard Stem Cell Institute, Harvard Medical School, Boston, MA 02115, USA
[2]Department of Anatomy and Developmental Biology, Monash University, Wellington Road, Clayton, VIC 3800, Australia
[3]Development and Stem Cells Program, Monash Biomedicine Discovery Institute, Wellington Road, Clayton, VIC 3800, Australia
[4]Department of Molecular Biosciences, Institute for Cellular and Molecular Biology, Center for Systems and Synthetic Biology,
The University of Texas at Austin, Austin, TX, USA
[5]Sanford I. Weill Department of Medicine, Sandra and Edward Meyer Cancer Center, Weill Cornell Medicine, New York, NY 10021, USA
[6]Australian Regenerative Medicine Institute, Monash University, Wellington Road, Clayton, VIC 3800, Australia
[7]Molecular Pathology & Cancer Center, Massachusetts General Hospital & Harvard Medical School, Boston, MA 02114, USA
[8]Howard Hughes Medical Institute, Boston, MA 02115, USA
[9]Present address: Institute of Molecular Biology GmbH (IMB), Ackermannweg 4, 55128 Mainz, Germany
[10]These authors contributed equally
[11]Senior author
[12]Lead Contact
*Correspondence: orkin@bloodgroup.tch.harvard.edu (S.H.O.), partha.das@monash.edu (P.P.D.)
https://doi.org/10.1016/j.molcel.2019.03.025

SUMMARY

Self-renewal and pluripotency of the embryonic stem cell (ESC) state are established and maintained by multiple regulatory networks that comprise transcription factors and epigenetic regulators. While much has been learned regarding transcription factors, the function of epigenetic regulators in these networks is less well defined. We conducted a CRISPR-Cas9-mediated loss-of-function genetic screen that identified two epigenetic regulators, TAF5L and TAF6L, components or co-activators of the GNAT-HAT complexes for the mouse ESC (mESC) state. Detailed molecular studies demonstrate that TAF5L/TAF6L transcriptionally activate c-Myc and Oct4 and their corresponding MYC and CORE regulatory networks. Besides, TAF5L/TAF6L predominantly regulate their target genes through H3K9ac deposition and c-MYC recruitment that eventually activate the MYC regulatory network for self-renewal of mESCs. Thus, our findings uncover a role of TAF5L/TAF6L in directing the MYC regulatory network that orchestrates gene expression programs to control self-renewal for the maintenance of mESC state.

INTRODUCTION

Embryonic stem cells (ESCs) isolated from inner cell mass of the blastocyst during embryonic development (Thomson et al.,

1998) have two unique properties: self-renewal, the ability to proliferate in the same state indefinitely; and pluripotency, the capacity to differentiate into all lineages of the organism (Zwaka and Thomson, 2005). Induced pluripotent stem cells (iPSCs) are molecularly and functionally equivalent to ESCs; however, iPSCs can be derived from diverse cell types through ectopic expression of specific factors (Takahashi and Yamanaka, 2006). These unique properties of ESCs and iPSCs make them attractive possible routes toward future cellular therapies (Robinton and Daley, 2012).

Self-renewal and pluripotency of the "ESC state" are established and maintained by the concerted action of signaling pathways that respond to external stimuli and intrinsically activate critical "transcription regulatory networks" (Young, 2011). Each of the transcription regulatory networks comprises transcription factors (TFs), co-factors, and chromatin/epigenetic regulators, and they form interconnected regulatory networks that control chromatin architecture and gene expression programs (Orkin and Hochedlinger, 2011; Young, 2011). Although much has been learned regarding specific transcription factors, the epigenetic components that establish and maintain the ESC state are less well defined.

Comprehensive studies using protein-DNA interaction data and expression of target genes have revealed three key functionally distinct regulatory modules or networks in mouse ESCs (mESCs) that maintain the ESC state: CORE (active), MYC (active), and Polycomb/PRC (repressive) (Chen et al., 2008; Kim et al., 2008, 2010). The CORE module is composed of "core" ESC TFs—OCT4, SOX2, and NANOG (OSN)—that cooperate with other ESC TFs, co-activators, and chromatin regulators to activate ESC-specific genes and repress lineage-specific genes (Kim et al., 2010; Young, 2011), whereas the PRC module is composed of both PRC1 and PRC2 components

and represses mainly lineage-specific genes to maintain the ESC state (Boyer et al., 2006; Laugesen and Helin, 2014). The MYC regulatory module or network is c-MYC centered and includes partner co-activators and chromatin modifiers. The MYC module primarily controls self-renewal of mESCs through cell cycle, metabolism, ribosome biogenesis, and protein-synthesis-associated genes but also regulates pluripotency-related genes to prevent differentiation to maintain the ESC state (Fagnocchi and Zippo, 2017; Fagnocchi et al., 2016; Scognamiglio et al., 2016; Smith et al., 2010).

c-MYC has been implicated in controlling chromatin architecture and epigenetic landscapes of ESCs in multiple ways. First, TIP60 (KAT5)-p400, a MYST (Moz/Morf, Ybf2, Sas3, Sas2, and Tip60) family member of the histone acetyltransferase (HAT) complex, interacts with c-MYC and functionally belongs to the MYC module (Kim et al., 2010). Notably, TIP60-p400 plays a crucial role in maintaining the ESC state through Nanog (Fazzio et al., 2008a). Likewise, GCN5/KAT2A (a catalytic component of SAGA and TFTC HAT complexes of the GNAT [GCN5 *N*-acetyltransferase] family) belongs to the MYC module in mESCs (Kim et al., 2010). These findings suggest that MYC acts with distinct HAT complexes through diverse mechanisms because different HAT complexes act on different substrates for gene regulation. For instance, TIP60-HAT acetylates H4K16, whereas GNAT-HAT preferentially acetylates H3K9 (Lee and Workman, 2007). c-MYC and the MYC module are both positively correlated with active H3ac, H4ac, and H3K4me3 marks and are negatively correlated with repressive H3K27me3 (Kim et al., 2008, 2010). Although the connection between c-MYC and HAT complexes has been established, the precise molecular mechanism by which c-MYC and HAT cooperatively regulate gene expression and maintain the ESC state is not well understood. Second, c-MYC targets several pluripotency factors, especially the core ESC transcription factor SOX2, and numerous chromatin-associated genes (Kim et al., 2010) for gene regulation to maintain the ESC state. Third, c-MYC also mediates gene repression of developmental genes through PRC2/H3K27me3 for mESC identity (Fagnocchi et al., 2016; Varlakhanova et al., 2011). Nonetheless, the repressive role of c-MYC is less well characterized compared to its pervasive role in gene activation.

Previously, short hairpin RNA (shRNA)-based knockdown screening approaches identified epigenetic factors necessary for maintenance of the mESC state (Bilodeau et al., 2009; Cooper and Brockdorff, 2013; Das et al., 2014; Fazzio et al., 2008a; Hu et al., 2009; Kagey et al., 2010). However, this approach remains technically challenging due to incomplete gene knockdown and the off-target activity of shRNAs, often making interpretation of the phenotypic changes difficult (Jackson and Linsley, 2010). Recently, clustered regularly interspaced short palindromic repeats (CRISPR)-Cas9-mediated gene-specific *knockouts* (*KOs*) have been used in multiple biological systems to circumvent partial knockdown and off-target effects of shRNAs (Hsu et al., 2014). Moreover, CRISPR-Cas9-mediated functional genetic screens successfully identified genes that are required for various biological processes (Shalem et al., 2014, 2015; Wang et al., 2014; Zhou et al., 2014). However, to date, CRISPR-Cas9-mediated functional genetic screens have

not described to identify epigenetic factors that control the mESC state.

Here, we performed a CRISPR-Cas9-mediated loss-of-function genetic screen, which identified two epigenetic regulators for the mESC state with high confidence: TAF5L and TAF6L, components of GNAT-HAT complexes. Through integrated molecular and genomics approaches, we have implicated the role of TAF5L/TAF6L in transcriptional activation of c-Myc and Oct4 and their respective MYC and CORE regulatory modules or networks. Furthermore, we revealed a detailed mechanism that shows TAF5L/TAF6L primarily regulate their target gene expression through H3K9ac deposition and c-MYC recruitment, which ultimately trigger the MYC regulatory network for self-renewal of mESCs. Thus, TAF5L/TAF6L establish a link between HAT complexes and c-MYC/MYC regulatory network to fine-tune gene expression programs for the maintenance of mESC state.

RESULTS

CRISPR-Cas9-Mediated Loss-of-Function Genetic Screen Identifies Potential Candidate Epigenetic Genes for the mESC State

The pooled sgRNAs used for the CRISPR-Cas9 screen targeted 323 genes of different classes of epigenetic regulators and ESC-specific TFs (Figure S1A; Table S1). We used six single guide RNAs (sgRNAs) per gene, targeting coding sequences from prior published work (Shalem et al., 2014) (Table S1). The "epigenetic CRISPR-Cas9 pooled library" contained 2,335 sgRNAs, including 1,938 sgRNAs targeting 323 epigenetic and ESC TF genes (Figure 1A; Table S1), 128 non-targeting sgRNAs as negative controls, 119 sgRNAs targeting GFP (of the *Oct4*-GFP reporter), and 150 sgRNAs targeting coding sequence of known mESC-specific TFs as positive controls (Figure 1A; Table S1). The pooled library was transduced into an *Oct4*-GFP reporter (OCT4 is the master regulator of ESCs) mESC line, which constitutively expressed Cas9 (Figures S1A–S1C). The *Oct4*-GFP reporter was used as a "readout" for the screen to measure the change in GFP levels upon perturbation of any of candidate epigenetic regulator genes by their corresponding sgRNAs. Lentiviral transduction of the pooled library was performed at low multiplicity (MOI) to ensure that selected cells contained roughly one sgRNA per cell (Figure S1A). After drug selection, the GFP-low and the GFP-high cells were sorted, genomic DNA was isolated, and next-generation sequencing (NGS) was employed to enumerate the sgRNAs in each cell populations (Figures S1A and S1D). The entire screen was performed in triplicate.

We calculated an "enrichment score" of each sgRNA by comparing its frequency in GFP-low versus GFP-high cells (Figure 1B). Enrichment scores were built based on the three best individual sgRNAs per gene (Table S2). The highest and lowest enrichment scores were obtained from GFP-targeting sgRNAs (positive controls) and non-targeting sgRNAs (negative controls), respectively, indicating the screen was technically successful (Figure 1B). Based on the enrichment scores, we identified several known and previously unreported distinct epigenetic regulator genes as candidates (24 out of 323 epigenetic genes in the pooled library) (Figures 1B and 1C), including epigenetic regulator genes (*Pcgf6*, *Kdm1a*, *Rybp, L3mbtl2*, *Smc1a*, *Nipbl*,
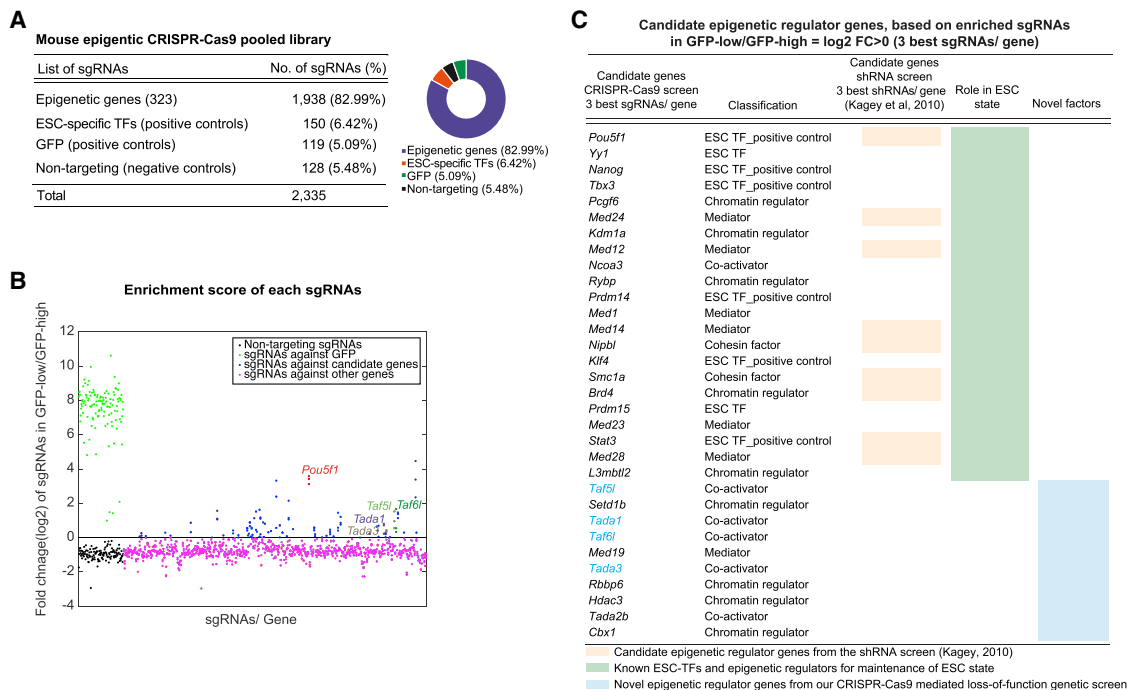
Figure 1. CRISPR-Cas9-Mediated Loss-of-Function Genetic Screen Identifies Potential Candidate Epigenetic Genes for the mESC State

(A) Mouse epigenetic CRISPR-Cas9 pooled library distribution.

(B) Dot plot analysis shows enrichment scores of sgRNAs by comparing their frequency in the GFP-low cells over the GFP-high cells. Enrichment scores of the three best individual sgRNAs per gene are represented. sgRNAs are ranked based on their corresponding target gene names, alphabetically from left to right on the x axis. sgRNAs targeting GFP (positive controls, in green), non-targeting sgRNAs (negative controls, in black), and sgRNAs targeting all candidate genes (in blue) are shown. The sgRNAs targeting candidate epigenetic genes (*Taf5l*, *Taf6l*, *Tada1*, and *Tada3*) are labeled. The sgRNAs targeting *Pou5f1/Oct4* (master regulator of mESCs, used as positive control; in red) provide high enrichment scores.

(C) List of candidate epigenetic regulator and TF genes ranked by enrichment scores. Candidate genes that have been previously identified through either shRNA screens (orange column) or individual functional studies (green column) related to mESC state are shown. Candidate epigenetic genes are also presented (blue column); among them *Taf5l*, *Taf6l*, *Tada1*, and *Tada3* were selected for further validation.

Related to Figure S1.

and *Brd4*), ESC TF genes (*Oct4*, *Nanog*, *Tbx3*, *Prdm14*, *Yy1*, *Prdm14*, and *Klf4*), and mediator genes (*Med24*, *Med12*, *Med1*, *Med14*, and *Med23*) (Figure 1C), which have been earlier linked to the mESC state (Whyte et al., 2012).

TAF5L and TAF6L Are Epigenetic Genes for the mESC State

As candidate epigenetic genes, we selected *Tada1*, *Tada3*, *Taf5l*, and *Taf6l* for further validation (Figures 1B and 1C). These factors belong to the GNAT family of HAT complexes (Lee and Workman, 2007). Each of the factors was highly expressed in undifferentiated mESCs except Tada1, which was moderately expressed. However, the expression levels of Tada3, Taf5l, and Taf6l were significantly reduced upon differentiation (24 and 48 h after differentiation) (Figure S1E), suggesting their possible roles in controlling the mESC state. Validation experiments were performed in two settings to target *Tada1*, *Tada3*, *Taf5l*, and *Taf6l* genes in wild-type mESCs. First, we used two "individual" sgRNAs (with highest enrichment scores from the screen) that target coding sequences of selected genes and create "indels" (insertions or deletions of bases) (Figure S2A). Second, we used "paired" sgRNAs that target exon(s) at the N-terminal end of each of the selected

candidate genes (Table S3) to create "homozygous deletion" or *KO* clones (Figures S2D–S2G). These *KO* clones were confirmed by genotyping PCR, Sanger sequencing, and quantitative RT-PCR (RT-qPCR) (Figures 2A and S2D–S2G).

The individual sgRNAs (that create indels) for all the target genes (*Tada1*, *Tada3*, *Taf5l*, and *Taf6l*) showed significant reduction of the endogenous levels of Oct4 with different degrees, except *Tada3* (Figures S2A–S2C). Also, individual sgRNAs targeting *Taf5l* and *Taf6l* showed substantial reduction of GFP levels of the *Oct4*-GFP reporter (Figure S2H), consistent with the primary outcome of the screen. Moreover, *Taf5l* and *Taf6l* *KO* clones consistently displayed a significant reduction of the endogenous Oct4 both at mRNA and protein levels (Figures 2B–2D and S2I). SSEA1, a mouse pluripotency marker, was also reduced in *Taf5l* and *Taf6l* *KO* mESCs (Figures S2J). Thus, TAF5L and TAF6L were chosen as epigenetic regulators for further mechanistic studies.

TAF5L and TAF6L Maintain Gene Expression Programs of the mESC State

Next, we asked the role of TAF5L and TAF6L in gene expression programs in mESCs. Transcriptomic profiling (RNA-seq)
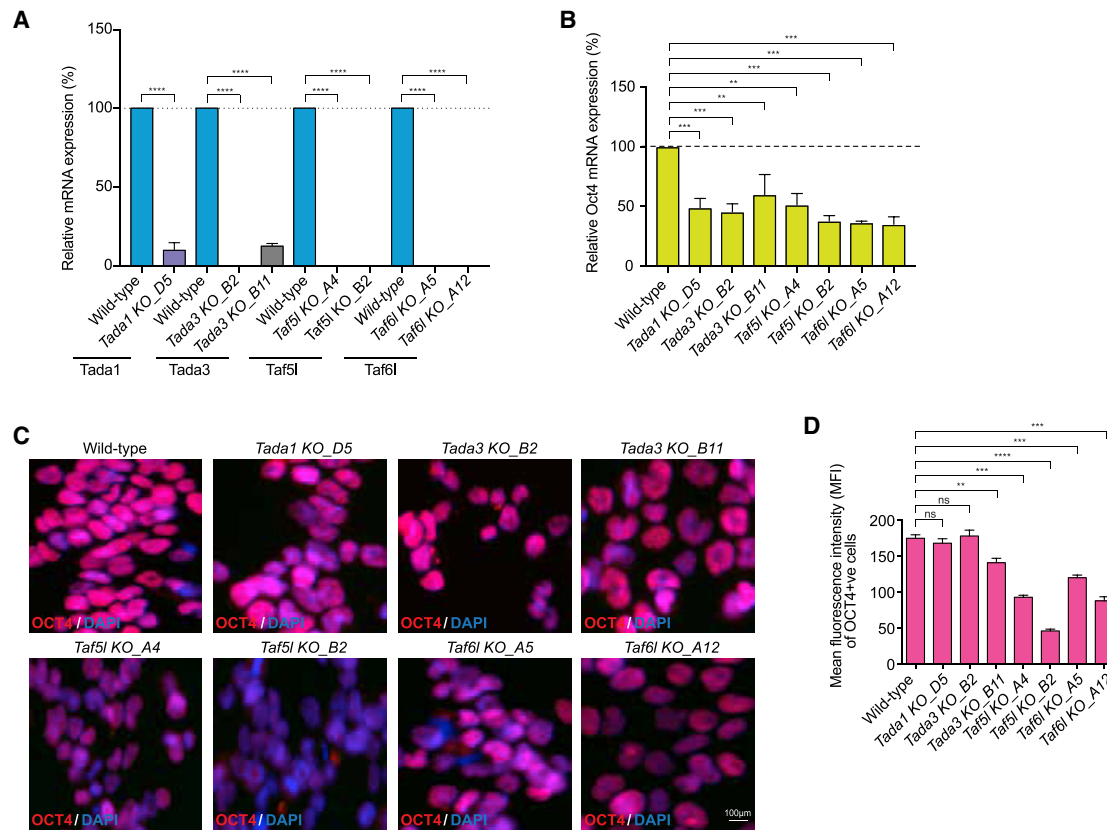
Figure 2. Validation Confirms TAF5L and TAF6L Are the Epigenetic Genes for the mESC State

(A) RT-qPCR data showing mRNA expression levels of Tada1, Tada3, Taf5l, and Taf6l in their corresponding *KOs* compared to wild type. mRNA levels were normalized to GAPDH. Paired sgRNAs were used to target exon(s) at the N terminus of *Tada1*, *Tada3*, *Taf5l*, and *Taf6l* candidate epigenetic genes to create homozygous or biallelic deletion or *KO* clones.

(B) Endogenous Oct4 mRNA expression levels in the *Tada1*, *Tada3*, *Taf5l*, and *Taf6l KOs*.

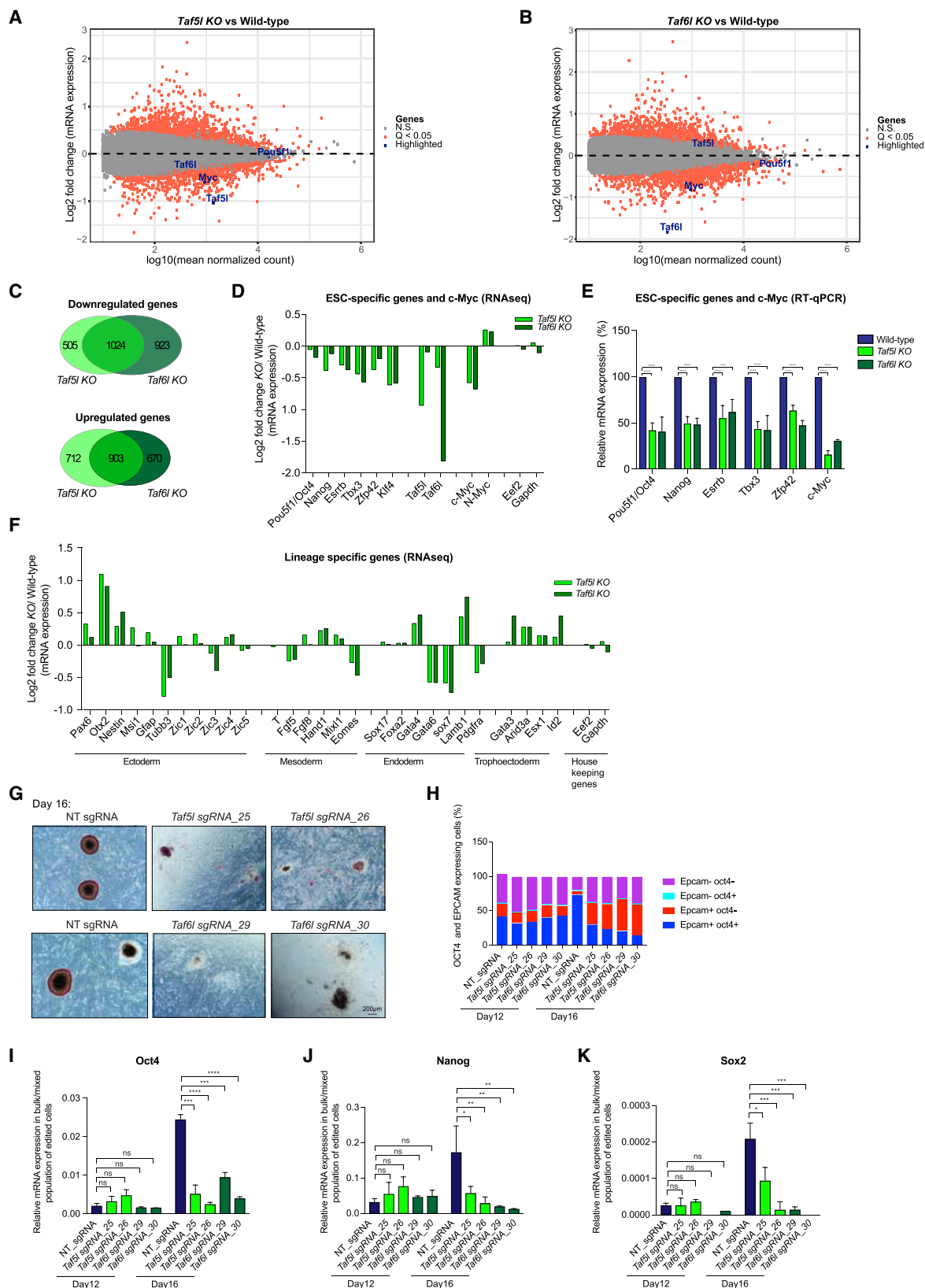(C) Immunofluorescence staining of OCT4 in the *Tada1*, *Tada3*, *Taf5l*, and *Taf6l KOs*. Scale bar: 100 μm.

(D) Quantification of mean fluorescence intensity (MFI) of OCT4 in the *Tada1*, *Tada3*, *Taf5l*, and *Taf6l KOs*.

Data are represented as mean ± SEM (n = 3); p values were calculated using ANOVA. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$; and ns (non-significant). Related to Figure S2.

performed in the absence of *Taf5l* and *Taf6l* revealed differentially expressed genes (p value [q value] < 0.05) in *Taf5l* and *Taf6l KOs* compared to wild-type mESCs (Figures 3A, 3B, and S3A) (Table S5) with a substantial overlap among upregulated as well as downregulated genes from *Taf5l KO* and *Taf6l KO* (Figure 3C). Expression of mESC-specific genes, including Oct4/Pou5f1, Nanog, Esrrb, Tbx3, Zfp42, and Klf4, as well as c-Myc, was downregulated (Figure 3D). Furthermore, quantitative RT-PCR (RT-qPCR) data confirmed a significant reduction of several mESC-specific genes, including Oct4, and dramatic reduction of c-Myc in *Taf5l* and *Taf6l KO* ESCs (Figures 3E, S3B, and S3C). The marked reduction of c-MYC was also monitored at the protein level in *Taf5l* and *Taf6l KOs* (Figure S3D). We also validated our findings in two independent ESC lines (J1 and CJ7) (Figures 3E, S3B, and S3C). Ectoderm-, mesoderm-, and endoderm-lineage-specific genes were both up- and downregulated without an obvious pattern; trophoectoderm (TE)-lineage-specific genes were consistently upregulated in *Taf5l* and *Taf6l KOs* (Figure 3F), implying that loss of *Taf5l* and *Taf6l* biases mESCs toward the TE

lineage. Importantly, mRNA expression of Taf5l and Taf6l itself was markedly downregulated in their corresponding *KOs*, as expected (Figures 3A, 3B, and 3D). We observed that exogenous expression of TAF6L fully rescued Oct4 and c-Myc expression in the *Taf6l KO*, whereas exogenous expression of TAF5L partially rescued Oct4 and c-Myc expression in the *Taf5l KO* (Figures S3E–S3H), suggesting that TAF5L and TAF6L regulate Oct4 and c-Myc gene expression to different extents.

To interrogate the possible functions of TAF5L and TAF6L in mESCs, gene ontology (GO) analysis was performed on differentially expressed genes from *Taf5l* and *Taf6l KOs*. *Taf5l KO*- and *Taf6l KO*-upregulated genes were considerably enriched for several biological processes, including cell-cell adhesion, cytoskeleton organization, and cell migration related to differentiation of mESCs; in contrast, *Taf5l KO*- and *Taf6l KO*-downregulated genes were significantly enriched with cell cycle, transcription, DNA replication, ribosome biogenesis/translation, nucleosome assembly, and chromatin-silencing biological functions related to self-renewal of mESCs (Figures S3I–S3L). Taken together,

A *Taf5l KO* vs Wild-type

B *Taf6l KO* vs Wild-type

C Downregulated genes / Upregulated genes

D ESC-specific genes and c-Myc (RNAseq)

E ESC-specific genes and c-Myc (RT-qPCR)

F Lineage specific genes (RNAseq)

G Day 16:

H OCT4 and EPCAM expressing cells (%)

I Oct4

J Nanog

K Sox2

(legend on next page)

these data demonstrate that TAF5L and TAF6L are required to maintain the gene expression programs of the mESC state.

TAF5L and TAF6L Are Required for Efficient Somatic Cell Reprogramming or iPSC Generation

Enforced expression of ESC-specific TFs OCT4, SOX2, KLF4, and c-MYC (OSKM) reprograms somatic cells to iPSCs that are molecularly and functionally equivalent to ESCs (Takahashi and Yamanaka, 2006). Epigenetic regulators serve as critical controllers of the reprogramming process and iPSC generation (Papp and Plath, 2013). Since we observed downregulation of mESC-specific genes (Oct4, Klf4) and c-Myc in mESCs lacking *Taf5l* and *Taf6l* (Figures 3D, 3E, S3C, and S3D), we sought to determine whether TAF5L and TAF6L influence somatic cell reprogramming or iPSC generation. We used doxycycline-inducible OKSM-driven reprogrammable mouse embryonic fibroblasts (MEFs) that carry an *Oct4*-GFP reporter (Stadtfeld et al., 2010). The reprogrammable MEFs were transduced with *Taf5l*- and *Taf6l*-targeting sgRNAs and non-targeting (control) sgRNAs; the efficiency of generation of stable transgene-independent iPSC colonies was measured. Indel frequencies of targeted *Taf5l* and *Taf6l* alleles were quantified by amplicon sequencing before and after the reprogramming. More than 80% of *Taf5l* and *Taf6l* alleles were edited (Figure S3S). The high indel frequencies of *Taf5l* and *Taf6l* alleles correlate with higher fractions of edited cells. Interestingly, *Taf5l*- and *Taf6l*-targeted reprogrammable MEFs generated similar percentages of OCT4+ EPCAM+ expressing cells as non-targeted (control) cells in the presence of doxycycline at day 12 (when reprogrammed cells are ready to acquire the full pluripotency but not yet established) (Figure 3H). However, shortly after doxycycline withdrawal at day 16 (when reprogrammed cells already acquire full pluripotency in a transgene-independent manner and become iPSCs), *Taf5l* and *Taf6l* bulk-edited, reprogrammed cells displayed reduced fractions of iPSCs and increased fractions of differentiated cells (Figures 3G, S3M, and S3N), along with decreased percentages of OCT4+ EPCAM+ and OCT4+ SSEA1+ expressing cells (Figures 3H and S3O). Furthermore, RNA expression analysis from *Taf5l* and *Taf6l* bulk-edited, reprogrammed cells demonstrated no significant changes of ESC-specific genes (endogenous Oct4, Nanog, Sox2) or TE-specific genes (Cdx2, Arid3a, Esx1) at

day 12 (Figures 3I–3K and S3P–S3R). Nonetheless, at day 16, *Taf5l* and *Taf6l* bulk-edited, reprogrammed cells exhibited reduced levels of ESC-specific genes (endogenous Oct4, Nanog, Sox2) and elevated levels of TE-specific genes (Cdx2, Arid3a, Esx1) (Figures 3I–3K and S3P–S3R), similar to gene expression changes in *Taf5l* and *Taf6l KO* mESCs (Figures 3D–3F). Taken together, these data suggest that both TAF5L and TAF6L are critically required for somatic cell reprogramming or iPSC generation induced by OSKM, precisely at the final stages of acquisition and/or maintenance of the iPSCs.

Genome-wide Distribution of TAF5L and TAF6L Binding Sites in mESCs

To understand the mechanisms by which TAF5L and TAF6L regulate gene expression and the mESC state, we determined their genome-wide occupancy using chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq). As we could not identify any antibodies that specifically recognize endogenous TAF5L and TAF6L, we generated flag-biotin (FB)-tagged TAF5L and TAF6L mESC lines (Figure S4A) and performed *in vivo* biotinylation-mediated ChIP-seq (Bio-ChIP-seq) of TAF5L and TAF6L (Das et al., 2014; Kim et al., 2009). The majority of TAF5L and TAF6L binding sites were distributed at the promoter and enhancer regions in mESCs (Figure 4A). Of note, mass spectroscopy analyses of streptavidin pull-down of biotinylated TAF5L-FB and TAF6L-FB identified their associated proteins (Figures S4B and S4C), including components of GNAT-HAT complexes (Lee and Workman, 2007). Hence, proteomics supports that FB-tagged versions of TAF5L and TAF6L are functional and suitable for Bio-ChIP-seq.

To determine the probable functions of TAF5L and TAF6L through their direct binding to the target genes, we performed GO analysis of the TAF5L- and TAF6L-binding target genes. TAF5L-binding target genes were enriched for positive regulation of transcription and stem cell maintenance biological functions, while TAF6L-binding target genes were highly enriched for positive regulation of transcription, cell cycle, cell division, and stem cell maintenance biological functions (Figures S4D and S4E). Furthermore, metagene analysis showed that both TAF5L- and TAF6L-binding target genes were highly and moderately expressed in mESCs (Figures S4F and S4G), implying that

---

Figure 3. TAF5L and TAF6L Are Required For Gene Expression Programs of mESC State and for Somatic Cell Reprogramming or iPSC Generation

(A and B) Scatterplots showing differentially expressed genes from *Taf5l KO* (A) and *Taf6l KO* (B), compared to wild-type mESCs. Orange dots are significantly up- and downregulated genes in *Taf5l and Taf6l KOs* with a q value < 0.05. Grey dots are unaltered genes in *KOs* compared to wild type. Genes of interest are labeled in the scatterplot.

(C) Venn diagrams represent overlapped up or downregulated genes between *Taf5l KO* and *Taf6l KO*.

(D) mRNA expression levels of mouse ESC-specific genes, c-Myc, and N-Myc in *Taf5l* and *Taf6l KOs* compared to wild type (from RNA-seq). Eef2 and Gapdh used as internal controls.

(E) mRNA expression levels of mouse ESC-specific genes and c-Myc in *Taf5l* and *Taf6l KOs* compared to wild type (from RT-qPCR). mRNA levels were normalized to GAPDH.
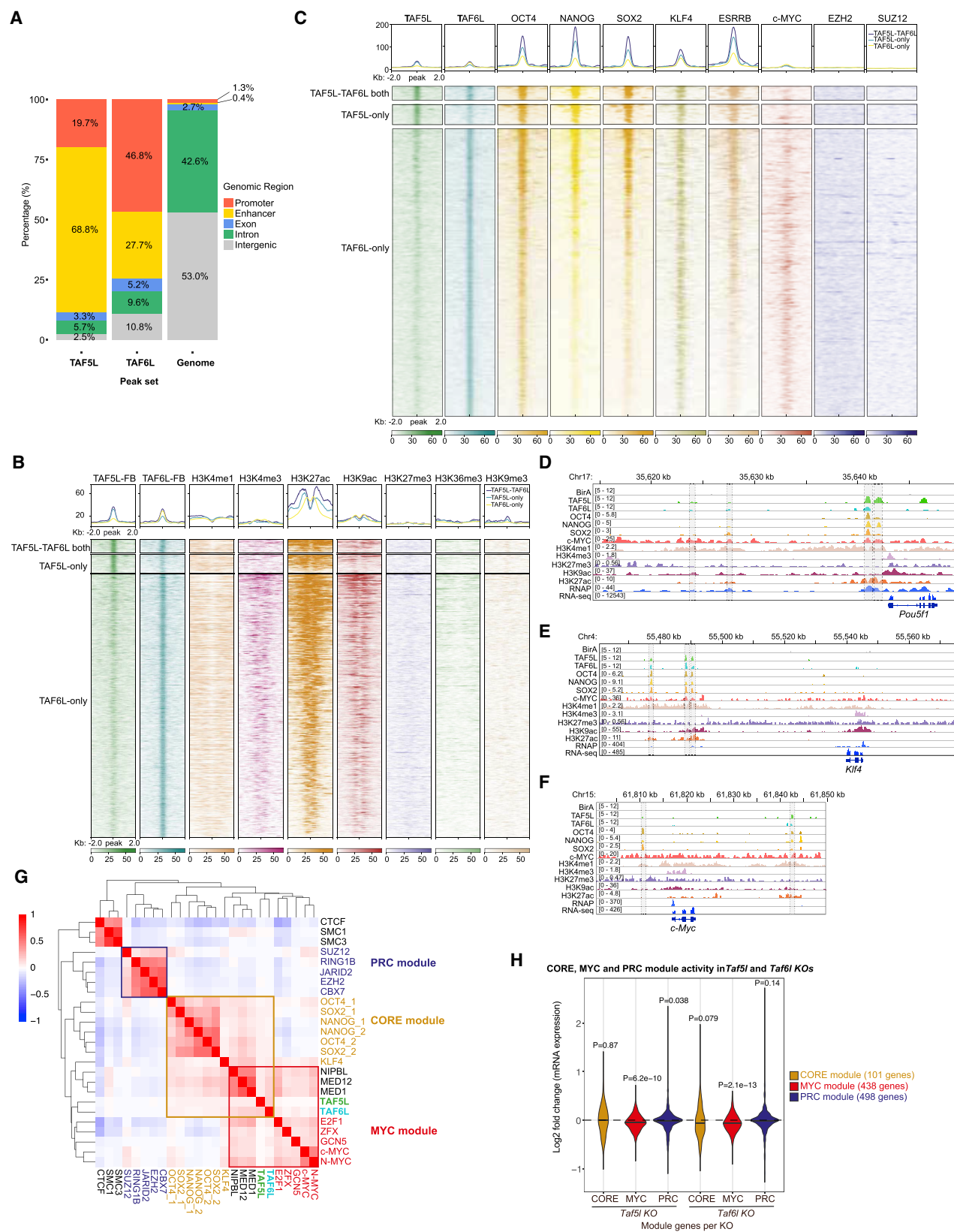
(F) mRNA expression levels of lineage-specific genes in *Taf5l* and *Taf6l KOs* compared to wild type (from RNA-seq).

(G) AP staining of transgene-independent iPSC colonies following reprogramming (at day 16). Scale bar: 200 μm.

(H) FACS analysis showing quantification of relative percentages of OCT4 and EPCAM expressing cells after reprogramming at day 12 and day 16.

(I–K) mRNA expression levels of ESC Core TFs—endogenous Oct4 (I), Nanog (J), and Sox2 (K) in bulk/mixed populations of edited cells after reprogramming at day 12 and day 16, upon perturbation of *Taf5l* and *Taf6l* using sgRNAs. Non-targeting sgRNA used as control.

Data are represented as mean ± SEM (n = 3); p values were calculated using ANOVA. *p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001; and ns (non-significant). Related to Figure S3.

A

B

C

D Chr17:

Pou5f1

E Chr4:

Klf4

F Chr15:

c-Myc

G

PRC module

CORE module

MYC module

H CORE, MYC and PRC module activity in *Taf5l* and *Taf6l* KOs

(legend on next page)

TAF5L and TAF6L most likely positively regulate their target genes for the maintenance of mESC state.

Both TAF5L and TAF6L Belong to the MYC and CORE Regulatory Modules but Predominantly Regulate the MYC Module Activity

To interrogate the significance of TAF5L and TAF6L binding, we correlated their occupancy with different active (H3K4me1, H3K4me3, H3K9ac, H3K27ac, and H3K36me3) and repressive (H3K27me3 and H3K9me3) histone marks (Zhou et al., 2011), as well as with ESC-TFs (OCT4, NANOG, SOX2, KLF4, and ESRRB), components of the PRC2 complex (EZH2 and SUZ12) and c-MYC binding sites. TAF5L and TAF6L binding regions overlapped with H3K4me1, H3K9ac, H3K27ac, and H3K4me3 active histone marks occupancy (Figure 4B). In addition, TAF5L and TAF6L binding sites overlapped with active TFs such as OCT4-, NANOG-, SOX2-, KLF4-, ESRRB-, and c-MYC-occupied sites, but not with repressive EZH2 and SUZ12 binding sites (Figure 4C). Similarly, co-occupancy of TAF5L, TAF6L, OCT4, NANOG, SOX2, c-MYC, H3K4me1, H3K9ac, H3K27ac, and H3K4me3 was observed at a few ESC-specific genes (Oct4/Pou5f1, Nanog, Klf4, and Esrrb) as well as at c-Myc, Taf5l, and Taf6l gene loci (Figures 4D–4F and S4H–S4K). Importantly, binding of TAF5L and TAF6L at these mESC-specific genes and c-Myc loci (Figures 4D–4F and S4H–S4L), and their reduced gene expression in Taf5l and Taf6l KOs (Figures 3D, 3E, and S3C) collectively indicate that TAF5L and TAF6L positively and directly regulate ESC-specific genes and c-Myc.

To obtain a global view of TAF5L and TAF6L binding in the mESC genome and sharing their occupancy with ESC-TFs and other epigenetic regulators, we combined previously published genome-wide binding datasets of several ESC factors (Das et al., 2014; Kim et al., 2010) together with the TAF5L and TAF6L binding dataset. Hierarchical clustering and a target correlation heatmap showed the degree of co-occupancy between factors and revealed three distinct ESC regulatory modules—CORE, MYC, and PRC, as defined previously (Kim et al., 2010). TAF5L and TAF6L binding sites were clustered mainly with the MYC module. However, binding sites of TAF5L and TAF6L were also correlated with the CORE module through NIPBL/MED1/MED12 binding (Figure 4G).

Because TAF5L and TAF6L appear to be members of the MYC and CORE modules, we monitored "module activity" by examining gene expression changes of the MYC and CORE module genes (i.e., MYC and CORE targeted or bound genes) in Taf5l

and Taf6l KOs. MYC module activity was significantly reduced in Taf5l and Taf6l KOs, whereas CORE module activity was modestly reduced only in the Taf6l KO (Figure 4H). Taken together, these data indicate that TAF5L and TAF6L belong to the MYC and CORE regulatory modules, but they predominantly regulate the MYC module activity.

TAF5L/TAF6L Regulate Target Gene Expression through H3K9ac Deposition and Predominant Recruitment of c-MYC over OCT4

Since both TAF5L and TAF6L belong to the MYC and CORE regulatory modules, we asked whether and how TAF5L and TAF6L cooperate with c-MYC (of MYC module) and OCT4 (of CORE module) for transcriptional gene regulation. To address this, genome-wide binding data of c-MYC and OCT4 were generated from wild-type Taf5l KO and Taf6l KO mESCs. The global occupancy of c-MYC was significantly reduced in Taf5l and Taf6l KOs compared to wild-type mESCs, whereas global occupancy of OCT4 was unchanged in these KOs (Figure S5A). The substantial loss of global binding of c-MYC was correlated with dramatic reduction of c-Myc levels in Taf5l and Taf6l KOs; however, global binding of OCT4 was not compromised by a modest reduction of Oct4 levels in Taf5l and Taf6l KOs (Figures 2B, 3D, 3E, S3C, S3D, and S5A). Besides, a significant decrease of c-MYC binding was observed at TAF5L bound genes in the Taf5l KO and at TAF6L bound genes in the Taf6l KO (Figures 5A and 5B). In contrast, OCT4 binding was considerably reduced only at TAF6L bound genes in the Taf6l KO compared to wild type (Figures 5C and 5D).

Next, we measured changes of H3K9ac and H3K4me3 marks in Taf5l and Taf6l KOs compared to wild type, as TAF5L and TAF6L are components of the GNAT-HAT complexes that are related to H3K9ac modification and gene activation (Lee and Workman, 2007), and c-MYC occupancy is correlated with the H3K4me3 signature (Kim et al., 2008, 2010). Our analysis exhibited a substantial reduction of H3K9ac globally (Figure S5B), as well as at TAF5L and TAF6L bound genes in the absence of Taf5l and Taf6l (Figures 5E and 5F). Nonetheless, binding of H3K4me3 was slightly increased globally, but its binding was unaltered at TAF5L and TAF6L bound genes in Taf5l and Taf6l KOs (Figures 5E, 5F, and S5B).

Based on these data, we hypothesized that TAF5L/TAF6L deposit H3K9ac and recruit c-MYC and/or OCT4 at target sites. Further analysis revealed a significant reduction of c-MYC occupancy at the common bound gene loci of TAF5L and H3K9ac (TAF5L+H3K9ac) and TAF6L and H3K9ac (TAF6L+H3K9ac) in

Figure 4. TAF5L and TAF6L Belong to the MYC and CORE Regulatory Modules but Mainly Regulate the MYC Module Activity
(A) A bar chart shows the genome-wide binding distribution of TAF5L-FB and TAF6L-FB.
(B) A heatmap displaying co-occupancy between different active and repressive histone marks with TAF5L-TAF6L both, TAF5L-only, and TAF6L-only binding regions.
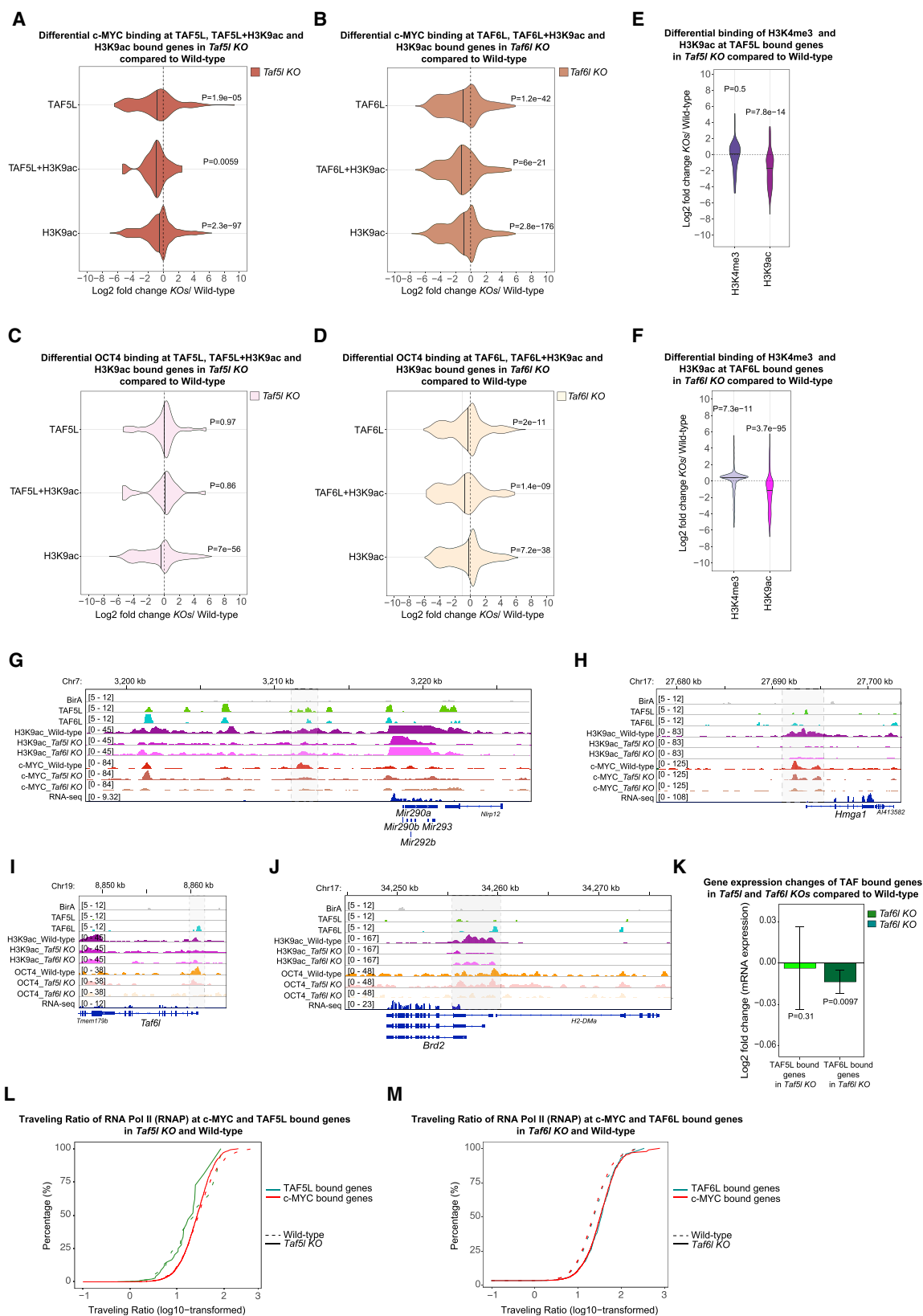(C) A heatmap representing co-occupied regions of ESC-TFs (OCT4, NANOG, SOX2, KLF4, and ESRRB); components of the PRC2 complex (EZH2 and SUZ12); and c-MYC binding sites with TAF5L-TAF6L both, TAF5L-only, and TAF6L-only binding regions.
(D–F) Genomic tracks of ChIP intensities of several factors, including TAF5L and TAF6L, and histone marks binding at Oct4/pou5f1 (D), Klf4 (E), and c-Myc (F) gene loci. RNA-seq tracks represent expression of these genes.
(G) A target correlation map of binding loci shows the degree of co-occupancy between selected TFs and chromatin or epigenetic regulators; three clusters or modules are presented: CORE, MYC, and PRC. The color scale depicts the Pearson correlation coefficient and the clustering tree is derived from hierarchical clustering. Red color indicates more frequent co-occupancy between factors.
(H) A violin plot representing changes in CORE, MYC, and PRC module gene expression in Taf5l and Taf6l KOs compared to wild type.
Related to Figure S4.

A Differential c-MYC binding at TAF5L, TAF5L+H3K9ac and H3K9ac bound genes in *Taf5l KO* compared to Wild-type

B Differential c-MYC binding at TAF6L, TAF6L+H3K9ac and H3K9ac bound genes in *Taf6l KO* compared to Wild-type

E Differential binding of H3K4me3 and H3K9ac at TAF5L bound genes in *Taf5l KO* compared to Wild-type

C Differential OCT4 binding at TAF5L, TAF5L+H3K9ac and H3K9ac bound genes in *Taf5l KO* compared to Wild-type

D Differential OCT4 binding at TAF6L, TAF6L+H3K9ac and H3K9ac bound genes in *Taf6l KO* compared to Wild-type

F Differential binding of H3K4me3 and H3K9ac at TAF6L bound genes in *Taf6l KO* compared to Wild-type

G

H

I

J

K Gene expression changes of TAF bound genes in *Taf5l* and *Taf6l KOs* compared to Wild-type

L Traveling Ratio of RNA Pol II (RNAP) at c-MYC and TAF5L bound genes in *Taf5l KO* and Wild-type

M Traveling Ratio of RNA Pol II (RNAP) at c-MYC and TAF6L bound genes in *Taf6l KO* and Wild-type

(legend on next page)

Taf5l KO and *Taf6l KO*, respectively (Figures 5A and 5B). Likewise, c-MYC occupancy was also reduced at H3K9ac bound genes both in *Taf5l* and *Taf6l KOs* (Figures 5A and 5B). Moreover, reduction of H3K9ac and c-MYC occupancy at the TAF5L and TAF6L binding sites of specific gene loci (ESC-specific *miR 290-295* cluster and *Hmga1*) in *Taf5l* and *Taf6l KOs* was also observed (Figures 5G and 5H). On the other hand, OCT4 occupancy was mainly reduced at common bound gene loci of TAF6L and H3K9ac (TAF6L+H3K9ac) in *Taf6l KO*; as well as at H3K9ac bound genes both in *Taf5l* and *Taf6l KOs* (Figures 5C and 5D). These events were also observed at specific gene loci, such as *Taf6l* and *Brd2* (Figures 5I and 5J).

Lastly, we checked gene expression changes of TAF5L and TAF6L bound genes in the *Taf5l KO* and *Taf6l KO*, respectively. TAF6L bound genes were significantly downregulated in *Taf6l KO*, whereas expression of TAF5L bound genes was unchanged in *Taf5l KO* (Figure 5K). Taken together, our data suggest that TAF5L/TAF6L regulate their target gene expression through H3K9ac deposition and predominate recruitment of c-MYC rather than OCT4, which might occur through indirect recruitment of c-MYC by TAF5L/TAF6L. Of note, we did not detect physical association of TAF5L/TAF6L and c-MYC (Figures S4B and S4C). However, it has been shown that TAF5L/TAF6L physically interact with c-MYC through other co-factors of the GNAT-HAT complexes in mESCs (Kim et al., 2010). Our data also indicate that TAF6L plays a more critical role compared to TAF5L in controlling gene expression. These inferences are consistent with the respective module activities, where both TAF5L and TAF6L impact the MYC module activity, yet only TAF6L regulates OCT4/CORE module activity (Figure 4H).

TAF5L/TAF6L Function with c-MYC to Activate Target Gene Expression through RNA Pol II Pause Release

TAF5L and TAF6L regulate target gene expression principally through c-MYC (Figure 5). c-MYC plays a crucial role in RNA Pol II pause release at target genes for active transcription in mESCs (Rahl et al., 2010). We speculated that TAF5L and TAF6L might function with c-MYC to activate target gene expression through RNA Pol II pause release. To gain insight into this mechanism, we generated binding data for RNA Pol II (RNAP) from *Taf5l KO*, *Taf6l KO*, and wild type. Next, we calculated the traveling ratio (TR) of RNAP that evaluates the ratio of

RNAP density or occupancy at promoter regions over transcribed or gene body regions (Rahl et al., 2010). We observed a substantial increase in the TR of RNAP at the TAF6L and c-MYC target genes in *Taf6l KO* compared to wild type, whereas no such difference in the TR of RNAP was observed at the TAF5L and c-MYC target genes in *Taf5l KO* (Figures 5L and 5M). Similarly, the TR of RNA Pol II-Ser 2p (transcription elongation) and RNA Pol II-Ser 5p (transcription initiation) was also significantly increased at the TAF6L and c-MYC bound genes in *Taf6l KO* (Figures S5D and S5F). A small but significant increase of the TR of RNA Pol II-Ser 2p (but not RNA Pol II-Ser 5p) was observed at the TAF5L and c-MYC bound genes in *Taf5l KO* (Figures S5C and S5E). Overall, these data suggest that primarily TAF6L functions with c-MYC for RNAP pause release at the promoters of TAF6L and c-MYC target genes for activation.

TAF5L/TAF6L Activate MYC Regulatory Network to Maintain Self-Renewal of mESCs
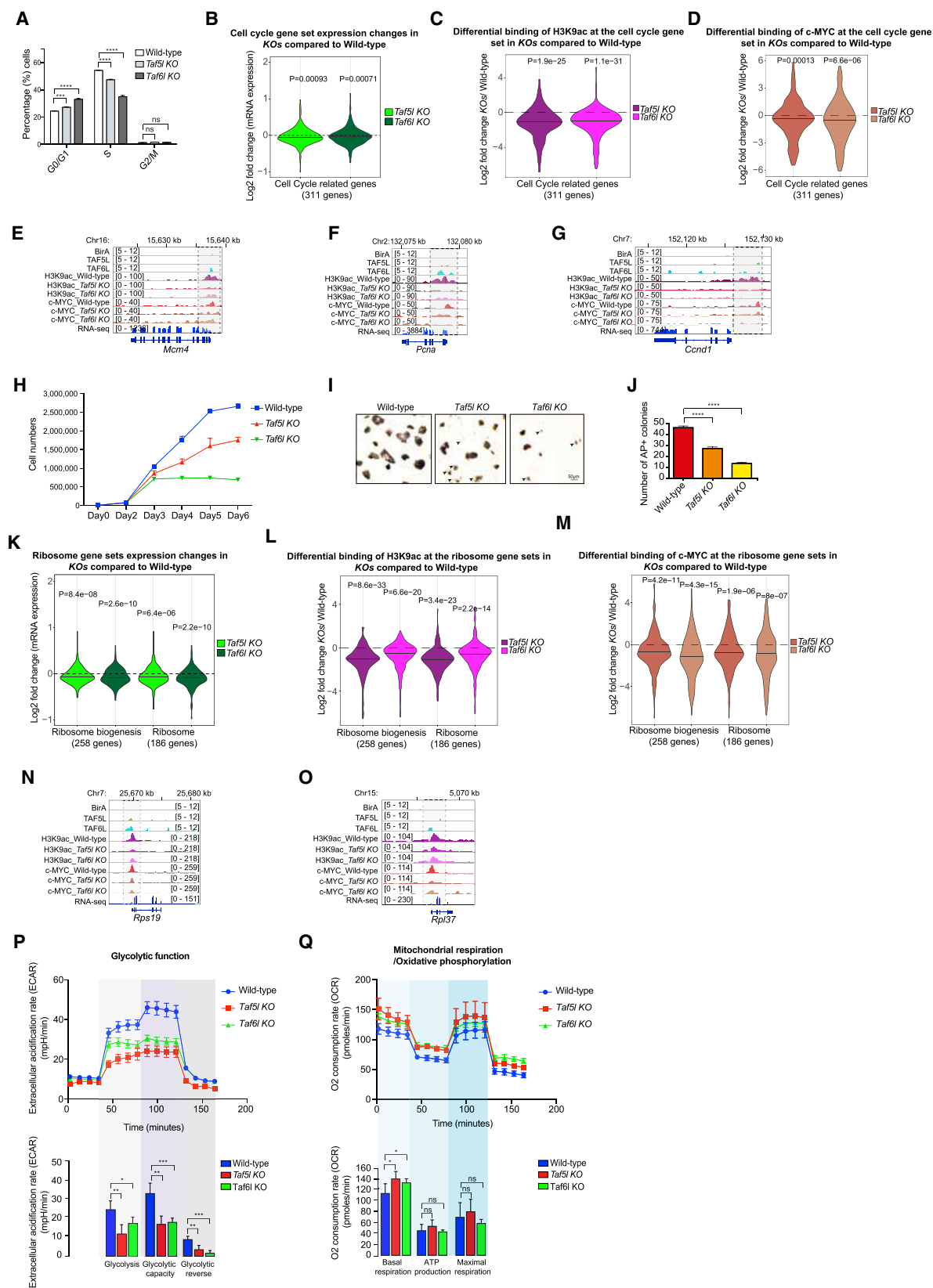
Several lines of evidence demonstrate that TAF5L and TAF6L not only transcriptionally regulate c-Myc expression (Figures 3E and 4F) but also collaborate with c-MYC to control expression of their target genes (Figure 5). As the connection between TAF5L/TAF6L and c-MYC is robust, we hypothesized that TAF5L/TAF6L might function with c-MYC in the same pathway to control the mESC state. c-MYC maintains the ESC state—preferentially self-renewal of ESCs—by regulating cell cycle, cell proliferation, ribosome biogenesis, and metabolism (Chappell and Dalton, 2013; Fagnocchi and Zippo, 2017). Therefore, we asked whether TAF5L and TAF6L are involved in controlling these biological processes.

First, cell cycle profiling showed lengthening of the G0/G1 phase, shortening of the S-phase (related to DNA synthesis) and an unaltered G2/M phase in *Taf5l* and *Taf6l KOs* (Figures 6A and S6A). Additionally, expression of the cell cycle gene set was downregulated in *Taf5l* and *Taf6l KOs* (Figure 6B), particularly gene sets and individual genes that are related to G1/S-phase, S-phase and DNA replication (Figures S6B–S6F). Moreover, the cell cycle gene set, including G1/S-phase, S-phase and DNA replication gene sets, revealed a reduction of H3K9ac and c-MYC binding at their gene loci in the absence of *Taf5l* and *Taf6l*, both globally and at individual gene loci

Figure 5. Predominantly TAF5L/TAF6L Modulate H3K9ac Deposition and c-MYC Recruitment at the TAF5L/TAF6L Target Genes to Activate Their Gene Expression through RNA Pol II Pause Release

(A) Differential binding of c-MYC at the TAF5L, TAF5L+H3K9ac, and H3K9ac bound genes in *Taf5l KO* compared to wild type.

(B) Differential binding of c-MYC at the TAF6L, TAF6L+H3K9ac, and H3K9ac bound genes in *Taf6l KO* compared to wild type.

(C) Differential binding of OCT4 at the TAF5L, TAF5L+H3K9ac, and H3K9ac bound genes in *Taf5l KO* compared to wild type.

(D) Differential binding of OCT4 at the TAF6L, TAF6L+H3K9ac, and H3K9ac bound genes in *Taf6l KO* compared to wild type.

(E) Differential binding of H3K9ac and H3K4me3 at the TAF5L bound genes in *Taf5l KO* compared to wild type.

(F) Differential binding of H3K9ac and H3K4me3 at the TAF6L bound genes in *Taf6l KO* compared to wild type.

(G and H) Genomic tracks of H3K9ac and c-MYC binding at *miR 290-295* cluster (G) and *Hmga1* (H) gene loci from wild type, *Taf5l*, and *Taf6l KOs*. Genomic tracks of BirA (control), TAF5L, and TAF6L binding are presented at the same gene loci. Highlighted regions show changes of H3K9ac and c-MYC binding.

(I and J) Genomic tracks of H3K9ac and OCT4 binding at *Taf6l* (I) and *Brd2* (J) gene loci from wild-type, *Taf5l*, and *Taf6l KOs*. Genomic tracks of BirA (control), TAF5L, and TAF6L binding are presented at the same gene loci. Highlighted regions show changes of H3K9ac and OCT4 binding.

(K) Gene expression changes of TAF5L and TAF6L bound genes in *Taf5l* and *Taf6l KOs*, respectively.

(L) The traveling ratio (TR) of RNA Pol II (RNAP) at c-MYC and TAF5L bound genes in *Taf5l KO* and wild type.

(M) The traveling ratio (TR) of RNA Pol II (RNAP) at c-MYC and TAF6L bound genes in *Taf6l KO* and wild type.

Related to Figure S5.

A

B Cell cycle gene set expression changes in *KOs* compared to Wild-type

C Differential binding of H3K9ac at the cell cycle gene set in *KOs* compared to Wild-type

D Differential binding of c-MYC at the cell cycle gene set in *KOs* compared to Wild-type

E Chr16: *Mcm4*

F Chr2: *Pcna*

G Chr7: *Ccnd1*

H

I Wild-type / *Taf5l KO* / *Taf6l KO*

J

K Ribosome gene sets expression changes in *KOs* compared to Wild-type

L Differential binding of H3K9ac at the ribosome gene sets in *KOs* compared to Wild-type

M Differential binding of c-MYC at the ribosome gene sets in *KOs* compared to Wild-type

N Chr7: *Rps19*

O Chr15: *Rpl37*

P Glycolytic function

Q Mitochondrial respiration /Oxidative phosphorylation

(legend on next page)

(Figures 6C–6G, S6G, and S6H). These data suggest that TAF5L and TAF6L regulate the cell cycle, specifically G1/S-phase, S-phase, and DNA replication—similar to the function of c-MYC in controlling cell cycle (Cartwright et al., 2005; Scognamiglio et al., 2016; Singh and Dalton, 2009).

Second, we observed a reduction in cell proliferation in *Taf5l* and *Taf6l KOs* (Figure 6H), consistent with a cell-cycle defect. Alkaline phosphatase (AP) staining revealed smaller and fewer dome-shaped AP+ colonies, with a fraction of differentiated cells in *Taf5l* and *Taf6l KO* mESCs (Figures 6I and 6J). Moreover, we observed relatively fewer AP+ colonies and more differentiated cells in *Taf6l KO* compared to *Taf5l KO* (Figures 6I and 6J). These findings suggest that TAF5L and TAF6L regulate cell proliferation of mESCs in a manner comparable to c-MYC (Eilers and Eisenman, 2008; Singh and Dalton, 2009).

Third, we observed reduced gene expression of ribosome gene sets (Figure 6K) and MYC module (MYC targeted) ribosome genes (Figure S6I) in *Taf5l*- and *Taf6l*-deleted mESCs. Ribosome gene sets showed a reduction of H3K9ac and c-MYC binding at their gene loci in the absence of *Taf5l* and *Taf6l*, both globally and at individual gene loci (Figures 6L–6O). These data suggest that TAF5L and TAF6L regulate ribosome biogenesis through c-MYC, similar to the function of c-MYC in ribosome biogenesis (Kim et al., 2008; van Riggelen et al., 2010).

Fourth, we assessed two major metabolic pathways, glycolysis and oxidative phosphorylation (OxPhos). Glycolytic function was significantly reduced in the absence of *Taf5l* and *Taf6l* (Figure 6P) and correlated with reduced expression of glucose transporter genes (important for glucose uptake) in *Taf5l* and *Taf6l KOs* (Figure S6J). A high rate of glucose uptake and glycolysis is required for rapid proliferation of mESCs, similar to cancer cells (Vander Heiden et al., 2009). In contrast, mitochondrial respiration/OxPhos function was unaffected, except for an upregulation of basal respiration level in *Taf5l* and *Taf6l KOs* (Figure 6Q), which may be linked to the low level of ongoing differentiation (Figure 6I) (Shyh-Chang et al., 2013). These functional data suggest that TAF5L and TAF6L play a pivotal role in glycolysis metabolism. Recent evidence also underlines the importance of c-MYC in regulating the glycolytic

metabolism of both human and mouse ESCs (Cao et al., 2015; Gu et al., 2016).

Lastly, we examined whether loss of *c-Myc* function phenocopies loss of *Taf5l* and *Taf6l* function in mESCs. Gene expression analysis demonstrated a partial but significant overlap between differentially expressed genes in *c-Myc*, *Taf5l*, and *Taf6l KOs* (Figures S7A–S7E), which implies that TAF5L/TAF6L work together with c-MYC and regulate gene expression. However, a significant fraction of non-overlapping differentially expressed genes was observed in *c-Myc*, *Taf5l*, and *Taf6l KOs* (Figure S7E), which indicates that TAF5L/TAF6L and c-MYC work independently as well.

Taken together, our data demonstrate that TAF5L/TAF6L activate the MYC regulatory network through multidimensional control of cell cycle, cell proliferation, ribosome biogenesis, and metabolism to maintain self-renewal of mESCs.

DISCUSSION

TAF5L- and TAF6L-Mediated Gene Regulation

Here, we identified TAF5L and TAF6L as epigenetic regulators for the mESC state (Figures 1 and 2). TAF5L and TAF6L are components or co-activators of different STAGA- (or SAGA), PCAF-, and TFTC-HAT complexes that belong to the GNAT family of HAT complexes (Carrozza et al., 2003; Lee and Workman, 2007). GCN5/KAT2A is the catalytic component of STAGA- and TFTC-HAT complexes, whereas PCAF/KAT2B is the catalytic component of the PCAF-HAT complex (Lee and Workman, 2007). HAT complexes exhibit different functions depending on their precise composition. Specific subunits of HAT complexes may select particular histone residues for acetylation and/or guide complexes to the genome to recruit other TFs, co-activators, and general transcription machinery for gene activation (Lee and Workman, 2007). We demonstrate that both TAF5L and TAF6L are part of the STAGA-, PCAF- and TFTC-HAT complexes (Figures S4B and S4C) and belong to the MYC and CORE regulatory modules, but predominantly regulate the MYC module activity (Figures 4G and 4H). Previous studies reveal a functional relationship between GCN5/KAT2A of HAT complexes and

---

Figure 6. TAF5L/TAF6L Maintain Self-Renewal of mESCs through MYC Regulatory Module or Network

(A) Quantitative analysis of different cell cycle phases from wild-type, *Taf5l*, and *Taf6l KOs*.

(B) Cell cycle gene set expression changes in *Taf5l* and *Taf6l KOs* compared to wild type.

(C and D) Differential binding of H3K9ac (C) c-MYC (D) at the cell cycle gene set in *Taf5l* and *Taf6l KOs* compared to wild type.

(E–G) Genomic tracks of H3K9ac and c-MYC binding at *Mcm4* (E), *Pcna* (F), and *Ccnd1* (G) gene loci from wild-type, *Taf5l*, and *Taf6l KOs*. Genomic tracks of BirA, TAF5L, and TAF6L binding are presented at the same gene loci. RNA-seq tracks show the expression of these genes in mESCs. Highlighted regions display changes of H3K9ac and c-MYC occupancy.

(H) Cell proliferation assay from wild-type, *Taf5l*, and *Taf6l KOs* (day 0 to day 6).

(I) Alkaline phosphatase (AP)+ colonies from wild-type, *Taf5l*, and *Taf6l KO* mESCs. Arrowheads indicate differentiating cells. Scale bar: 50 μm.

(J) Quantification of numbers of AP+ colonies (12-well scale).

(K) Expression changes of two ribosome gene sets in *Taf5l* and *Taf6l KOs* compared to wild type.

(L and M) Differential binding of H3K9ac (L) and c-MYC (M) at the ribosome gene sets in *Taf5l* and *Taf6l KOs* compared to wild type.

(N and O) Genomic tracks of H3K9ac and c-MYC binding at ribosome (*Rps19*, N, and *Rpl37*, O) gene loci from wild-type, *Taf5l*, and *Taf6l KOs*. Genomic tracks of BirA, TAF5L, and TAF6L binding are presented at the same loci. RNA-seq tracks show the expression of these genes in mESCs. Highlighted regions show changes of H3K9ac and c-MYC occupancy.

(P) Glycolytic function measurement using extracellular acidification rate (ECAR) from wild-type, *Taf5l*, and *Taf6l KOs*.

(Q) Oxidative phosphorylation activity presented based on oxygen consumption rate from wild-type, *Taf5l*, and *Taf6l KOs*.

Data are represented as mean ± SEM (n = 3); p values were calculated using ANOVA. *p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001; and ns (non-significant). Related to Figures S6 and S7.
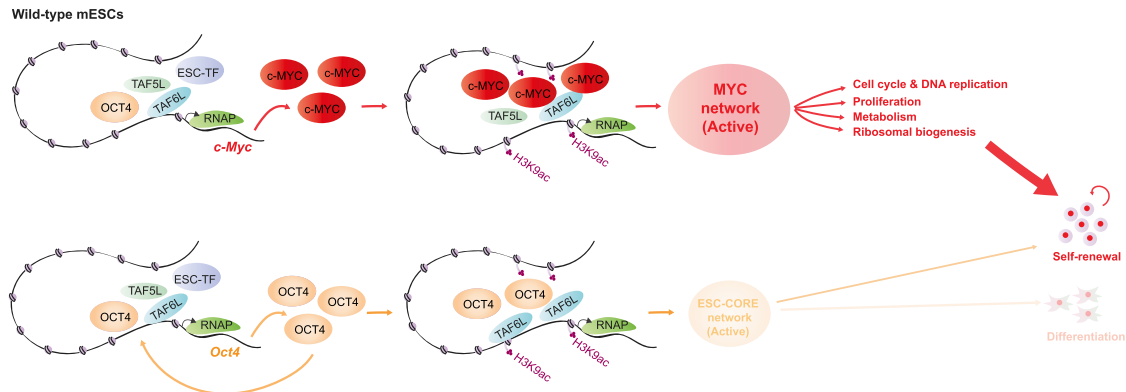
Figure 7. A Model Represents the Detailed Functions of TAF5L and TAF6L in mESCs

The proposed model describes how TAF5L and TAF6L transcriptionally activate c-Myc gene expression. Also, TAF5L/TAF6L regulate their target genes through H3K9ac deposition and c-MYC recruitment that eventually activate the MYC regulatory network for self-renewal of mESCs. Similarly, TAF5L and TAF6L transcriptionally activate Oct4 gene expression. Particularly, TAF6L modulate H3K9ac deposition and Oct4 recruitment at their target sites to activate ESC/CORE regulatory network that controls self-renewal and a low-level of pluripotency/differentiation of mESCs. Our findings suggest that TAF5L/TAF6L predominantly activate c-Myc and the MYC regulatory network over Oct4 and the CORE regulatory network to mainly control self-renewal for the maintenance of mESC state. TAF5L-/TAF6L-mediated gene activation of c-Myc and Oct4 and their corresponding activated MYC and CORE networks are shown. MYC network display in larger bold font (bright red), compared to CORE network (light orange)—represents predominant MYC network activity over CORE network. The thick arrowhead (red) depicts the activated MYC network that primarily directs the self-renewal of mESCs. Simultaneously, the thin arrowhead (light orange) illustrates an activated CORE network that controls a low level of self-renewal and differentiation of mESCs.

c-MYC (Hirsch et al., 2015; Kim et al., 2010). However, molecular mechanisms underlying this relationship are ill defined. Our findings close this gap in understanding. We show that c-Myc is a direct downstream target of TAF5L/TAF6L and that TAF5L/TAF6L transcriptionally activate c-Myc gene expression (Figures 3 and 4). In addition, mechanistic studies illustrate how TAF5L and TAF6L modulate H3K9ac deposition and recruit c-MYC at the TAF5L and TAF6L target genes to activate the MYC gene regulatory network or module (Figures 5 and 6). Likewise, TAF5L/TAF6L transcriptionally activate Oct4 (Figures 3 and 4); particularly, TAF6L controls H3K9ac deposition and OCT4 recruitment at the TAF6L bound genes to trigger the CORE regulatory network (Figures 5 and 6). These events are consistent with the observation that both TAF5L and TAF6L influence MYC module activity, whereas only TAF6L affects OCT4/CORE module activity (Figure 4H). Overall, our data indicate that TAF5L/TAF6L regulate both c-Myc and the MYC module, as well as Oct4 and the CORE module, but predominately regulate c-Myc and the MYC network or module for self-renewal of mESCs (Figures 3, 4, 5, and 6).

Our systematic analyses show that TAF6L plays a more critical role than TAF5L in gene regulation based on the following: (1) TAF6L controls the MYC module activity to a greater degree than TAF5L (Figure 4H); (2) H3K9ac deposition and c-MYC binding/recruitment are more dependent on TAF6L than TAF5L at their bound genes for activation (Figures 5A, 5B, 5E, 5F, S5A, and S5B); (3) TAF6L primarily functions with c-MYC for RNAP pause release at the promoters of TAF6L and c-MYC target genes for activation (Figures 5L, 5M, and S5C–S5F); (4) only TAF6L controls OCT4 (of CORE module) recruitment at the TAF6L bound genes, and regulates the CORE module activity (Figures 4H, 5C, 5D, 5I, and 5J); (5) the correlation between gene expression and binding data of TAF5L and TAF6L demon-

strates that only TAF6L bound genes are significantly downregulated in the absence of Taf6l (Figure 5K). However, we cannot exclude a positive role for TAF5L in controlling some critical ESC-specific genes, such as Oct4, Nanog, Klf4, Esrrb, and c-Myc (Figures 3, 4, and 5). Indeed, metagene and GO analyses show that TAF5L bound genes are highly or moderately expressed in the mESCs, and they are related to positive regulation of transcription and stem cells maintenance functions, same as TAF6L (Figures S4D–S4G). Moreover, TAF5L also regulates the cell cycle, DNA replication, and ribosomal genes, similar to TAF6L (Figures 6 and S6); suggesting that TAF5L works with TAF6L to regulate important gene sets and individual genes that control the mESC state.

TAF5L and TAF6L Predominantly Regulate Self-Renewal of mESCs

Our data provide evidence that TAF5L/TAF6L work together with c-MYC to activate the MYC regulatory network that regulates cell cycle, DNA replication, ribosome biogenesis, and metabolism (particularly glycolysis) to maintain mainly cellular proliferation and self-renewal of mESCs (Figures 4, 5, 6) (Fagnocchi and Zippo, 2017; Shyh-Chang et al., 2013; Singh and Dalton, 2009; van Riggelen et al., 2010). However, low-level differentiation of mESCs was also detected in the absence of Taf5l and Taf6l (Figure 6I)—caused by upregulation of TE-specific genes and/or downregulation of ESC-specific genes (e.g., Pou5f1/Oct4, Nanog, Klf4, and Tbx3) in Taf5l and Taf6l KOs (Figures 3D–3F), possibly through ESC-specific genes, as they are the direct targets of TAF5L/TAF6L and part of the CORE regulatory network (Figures 4 and 5). Overall, we demonstrate a predominant self-renewal defect and a low-level of differentiation in the absence of Taf5l and Taf6l (Figures 6H–6J), which is connected to TAF5L-/TAF6L-directed predominant regulation of the

c-MYC/MYC network over the OCT4/CORE network in the mouse ESCs (Figures 3, 4, 5, and 6).

Several studies demonstrate that MYC regulates cell cycle, DNA replication, ribosome biogenesis, and metabolism for stem cell proliferation and self-renewal of ESCs (Chen et al., 2008; Eilers and Eisenman, 2008; Kim et al., 2008). Furthermore, MYC controls proliferation and self-renewal, but not pluripotency or differentiation in the ground-state of mESCs using 2i+LIF condition (Scognamiglio et al., 2016). ESCs and cancer cells and/or cancer stem cells are linked to self-renewal and proliferation, based on overlapping ESC/CORE and MYC module activities (Kim et al., 2010; Wong et al., 2008). While the majority of the studies emphasize the central role of MYC in proliferation and self-renewal, others demonstrate the role of MYC in differentiation or pluripotency of ESCs as well (Smith et al., 2010; Varlakhanova et al., 2010, 2011).

The self-renewal defect observed in *c-Myc* and *N-Myc double knockout* (*DKO*) ESCs, but not in *single KO* ESCs maintained in serum+LIF or 2i+LIF conditions (Scognamiglio et al., 2016; Varlakhanova et al., 2010), suggests overlapping functions of c-MYC and N-MYC. We demonstrate a self-renewal defect in *Taf5l* and *Ta6l single KO* mESCs (Figures 6H–6J), which is similar to the self-renewal defect in *c-Myc* and *N-Myc DKO* mESCs, but not that seen in *c-Myc single KO* mESCs (Scognamiglio et al., 2016; Varlakhanova et al., 2010). These observations imply that TAF5l/TAF6L regulate self-renewal through both c-MYC and N-MYC (of the MYC regulatory network) and that they have a compensatory function in mESCs.

In conclusion, we demonstrate that TAF5L/TAF6L transcriptionally activate c-Myc and Oct4 and their respective MYC and CORE regulatory modules. Furthermore, we reveal a mechanism by which TAF5L/TAF6L predominantly activate the MYC regulatory module or network that fine-tunes gene expression programs to control self-renewal for the maintenance of mESC state (Figure 7).

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - ○ Mouse embryonic stem cells (mESCs)
  - ○ Reprogrammable mouse embryonic fibroblasts
  - ○ Induced pluripotent stem cells (iPSCs)
  - ○ Human embryonic kidney cells (HEK293T cells)
- METHOD DETAILS
  - ○ Mouse epigenetic CRISPR-Cas9 pooled library design
  - ○ Mouse epigenetic CRISPR-Cas9 pooled library construction
  - ○ Lentiviral production
  - ○ CRISPR-Cas9 mediated screen in mESCs
  - ○ Generation of mESC *KO* deletion clones
  - ○ Somatic cell reprogramming or iPSC generation
  - ○ Flow cytometry
  - ○ Intracellular FACS

- ○ Immunofluorescence microscopy
- ○ Alkaline Phosphatase (AP) staining
- ○ Cell Cycle analysis
- ○ Cell proliferation assay
- ○ Measurement of metabolic functions in mESCs
- ○ Generation of Flag-Biotin (FB) tagged mESC lines
- ○ Pull down of Flag-Biotin tagged TAF5L and TAF6L, and identification of their interacting partners through mass spectrometry
- ○ Western Blot
- ○ RNA isolation and RT-qPCR
- ○ RNA-seq
- ○ ChIP-seq
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - ○ CRISPR-Cas9 screen data analysis
  - ○ RNA-seq analysis
  - ○ GO (Gene Ontology)
  - ○ ChIP-seq analysis
  - ○ Heat-map
  - ○ Hierarchical clustering
  - ○ Module activity
  - ○ Violin Plots/ differential binding
  - ○ Calculation of traveling ratio of RNAP, RNAP-Ser 5p and RNAP-Ser 2p
  - ○ Statistical analysis
- DATA AND SOFTWARE AVAILABILITY

AUTHOR CONTRIBUTIONS

D.S., P.T., M.C.C., L.L., D.C.D.G., M.J.B., C.M.N., Y.B.Y.S., and P.P.D. performed the experiments and analyzed the data. D.S., M.O., P.T., E.A., J.M.P., J.K., S.H.O., and P.P.D. interpreted the data. M.O., R.F., and L.P. performed bioinformatics analysis. D.S., M.O., S.H.O., and P.P.D. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

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.; The Gene Ontology Consortium (2000). Gene ontology: tool for the unification of biology. Nat. Genet. 25, 25–29.

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

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.

Canver, M.C., Lessard, S., Pinello, L., Wu, Y., Ilboudo, Y., Stern, E.N., Needleman, A.J., Galactéros, F., Brugnara, C., Kutlar, A., et al. (2017). Variant-aware saturating mutagenesis using multiple Cas9 nucleases identifies regulatory elements at trait-associated loci. Nat. Genet. 49, 625–634.

Cao, Y., Guo, W.T., Tian, S., He, X., Wang, X.W., Liu, X., Gu, K.L., Ma, X., Huang, D., Hu, L., et al. (2015). miR-290/371-Mbd2-Myc circuit regulates glycolytic metabolism to promote pluripotency. EMBO J. 34, 609–623.

Carrozza, M.J., Utley, R.T., Workman, J.L., and Côté, J. (2003). The diverse functions of histone acetyltransferase complexes. Trends Genet. 19, 321–329.

Cartwright, P., McLean, C., Sheppard, A., Rivett, D., Jones, K., and Dalton, S. (2005). LIF/STAT3 controls ES cell self-renewal and pluripotency by a Myc-dependent mechanism. Development 132, 885–896.

Chappell, J., and Dalton, S. (2013). Roles for MYC in the establishment and maintenance of pluripotency. Cold Spring Harb. Perspect. Med. 3, a014381.

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.

Cooper, S., and Brockdorff, N. (2013). Genome-wide shRNA screening to identify factors mediating Gata6 repression in mouse embryonic stem cells. Development 140, 4110–4115.

Das, P.P., Shao, Z., Beyaz, S., Apostolou, E., Pinello, L., De Los Angeles, A., O'Brien, K., Atsma, J.M., Fujiwara, Y., Nguyen, M., et al. (2014). Distinct and combinatorial functions of Jmjd2b/Kdm4b and Jmjd2c/Kdm4c in mouse embryonic stem cell identity. Mol. Cell 53, 32–48.

Das, P.P., Hendrix, D.A., Apostolou, E., Buchner, A.H., Canver, M.C., Beyaz, S., Ljuboja, D., Kuintzle, R., Kim, W., Karnik, R., et al. (2015). PRC2 is required to maintain expression of the maternal Gtl2-Rian-Mirg locus by preventing de novo DNA methylation in mouse embryonic stem cells. Cell Rep. 12, 1456–1470.

Dennis, G., Jr., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., and Lempicki, R.A. (2003). DAVID: database for annotation, visualization, and integrated discovery. Genome Biol. 4, 3.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21.

Eilers, M., and Eisenman, R.N. (2008). Myc's broad reach. Genes Dev. 22, 2755–2766.

Fagnocchi, L., and Zippo, A. (2017). Multiple roles of MYC in integrating regulatory networks of pluripotent stem cells. Front. Cell Dev. Biol. 5, 7.

Fagnocchi, L., Cherubini, A., Hatsuda, H., Fasciani, A., Mazzoleni, S., Poli, V., Berno, V., Rossi, R.L., Reinbold, R., Endele, M., et al. (2016). A Myc-driven self-reinforcing regulatory network maintains mouse embryonic stem cell identity. Nat. Commun. 7, 11903.

Fazzio, T.G., Huff, J.T., and Panning, B. (2008a). An RNAi screen of chromatin proteins identifies Tip60-p400 as a regulator of embryonic stem cell identity. Cell 134, 162–174.

Gu, W., Gaeta, X., Sahakyan, A., Chan, A.B., Hong, C.S., Kim, R., Braas, D., Plath, K., Lowry, W.E., and Christofk, H.R. (2016). Glycolytic metabolism plays a functional role in regulating human pluripotent stem cell state. Cell Stem Cell 19, 476–490.

Hirsch, C.L., Coban Akdemir, Z., Wang, L., Jayakumaran, G., Trcka, D., Weiss, A., Hernandez, J.J., Pan, Q., Han, H., Xu, X., et al. (2015). Myc and SAGA rewire an alternative splicing network during early somatic cell reprogramming. Genes Dev. 29, 803–816.

Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. Cell 157, 1262–1278.

Hu, G., Kim, J., Xu, Q., Leng, Y., Orkin, S.H., and Elledge, S.J. (2009). A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. Genes Dev. 23, 837–848.

Huber, W., Carey, V.J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B.S., Bravo, H.C., Davis, S., Gatto, L., Girke, T., et al. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. Nat. Methods 12, 115–121.

Jackson, A.L., and Linsley, P.S. (2010). Recognizing and avoiding siRNA off-target effects for target identification and therapeutic application. Nat. Rev. Drug. Discov. 9, 57–67.

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.

Kent, W.J., Zweig, A.S., Barber, G., Hinrichs, A.S., and Karolchik, D. (2010). BigWig and BigBed: enabling browsing of large distributed datasets. Bioinformatics 26, 2204–2207.

Kim, J., Chu, J., Shen, X., Wang, J., and Orkin, S.H. (2008). An extended transcriptional network for pluripotency of embryonic stem cells. Cell 132, 1049–1061.

Kim, J., Cantor, A.B., Orkin, S.H., and Wang, J. (2009). Use of in vivo biotinylation to study protein-protein and protein-DNA interactions in mouse embryonic stem cells. Nat. Protoc. 4, 506–517.

Kim, J., Woo, A.J., Chu, J., Snow, J.W., Fujiwara, Y., Kim, C.G., Cantor, A.B., and Orkin, S.H. (2010). A Myc network accounts for similarities between embryonic stem and cancer cell transcription programs. Cell 143, 313–324.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359.

Laugesen, A., and Helin, K. (2014). Chromatin repressive complexes in stem cells, development, and cancer. Cell Stem Cell 14, 735–751.

Lee, K.K., and Workman, J.L. (2007). Histone acetyltransferase complexes: one size doesn't fit all. Nat. Rev. Mol. Cell Biol. 8, 284–295.

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

Mudge, J.M., and Harrow, J. (2015). Creating reference gene annotation for the mouse C57BL6/J genome assembly. Mamm. Genome 26, 366–378.

Orkin, S.H., and Hochedlinger, K. (2011). Chromatin connections to pluripotency and cellular reprogramming. Cell 145, 835–850.

Papp, B., and Plath, K. (2013). Epigenetics of reprogramming to induced pluripotency. Cell 152, 1324–1343.

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods 14, 417–419.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842.

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

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation

web server for deep-sequencing data analysis. Nucleic Acids Res. *44* (W1), W160–W165.

Robinton, D.A., and Daley, G.Q. (2012). The promise of induced pluripotent stem cells in research and therapy. Nature *481*, 295–305.

Ross-Innes, C.S., Stark, R., Teschendorff, A.E., Holmes, K.A., Ali, H.R., Dunning, M.J., Brown, G.D., Gojis, O., Ellis, I.O., Green, A.R., et al. (2012). Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature *481*, 389–393.

Sanjana, N.E., Shalem, O., and Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. Nat. Methods *11*, 783–784.

Scognamiglio, R., Cabezas-Wallscheid, N., Thier, M.C., Altamura, S., Reyes, A., Prendergast, Á.M., Baumgärtner, D., Carnevalli, L.S., Atzberger, A., Haas, S., et al. (2016). Myc Depletion Induces a Pluripotent Dormant State Mimicking Diapause. Cell *164*, 668–680.

Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelson, T., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G., and Zhang, F. (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. Science *343*, 84–87.

Shalem, O., Sanjana, N.E., and Zhang, F. (2015). High-throughput functional genomics using CRISPR-Cas9. Nat. Rev. Genet. *16*, 299–311.

Shyh-Chang, N., Daley, G.Q., and Cantley, L.C. (2013). Stem cell metabolism in tissue development and aging. Development *140*, 2535–2547.

Singh, A.M., and Dalton, S. (2009). The cell cycle and Myc intersect with mechanisms that regulate pluripotency and reprogramming. Cell Stem Cell *5*, 141–149.

Smith, K.N., Singh, A.M., and Dalton, S. (2010). Myc represses primitive endoderm differentiation in pluripotent stem cells. Cell Stem Cell *7*, 343–354.

Stadtfeld, M., Maherali, N., Borkent, M., and Hochedlinger, K. (2010). A reprogrammable mouse strain from gene-targeted embryonic stem cells. Nat. Methods *7*, 53–55. Published online December 13, 2009.

Stamatoyannopoulos, J.A., Snyder, M., Hardison, R., Ren, B., Gingeras, T., Gilbert, D.M., Groudine, M., Bender, M., Kaul, R., Canfield, T., et al.; Mouse ENCODE Consortium (2012). An encyclopedia of mouse DNA elements (Mouse ENCODE). Genome Biol. *13*, 418.

Sun, Y.B.Y., Qu, X., Howard, V., Dai, L., Jiang, X., Ren, Y., Fu, P., Puelles, V.G., Nikolic-Paterson, D.J., Caruana, G., et al. (2015). Smad3 deficiency protects mice from obesity-induced podocyte injury that precedes insulin resistance. Kidney Int. *88*, 286–298.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663–676.

Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., and Jones, J.M. (1998). Embryonic stem cell lines derived from human blastocysts. Science *282*, 1145–1147.

van Riggelen, J., Yetil, A., and Felsher, D.W. (2010). MYC as a regulator of ribosome biogenesis and protein synthesis. Nat. Rev. Cancer *10*, 301–309.

Vander Heiden, M.G., Cantley, L.C., and Thompson, C.B. (2009). Understanding the Warburg effect: the metabolic requirements of cell proliferation. Science *324*, 1029–1033.

Varlakhanova, N.V., Cotterman, R.F., deVries, W.N., Morgan, J., Donahue, L.R., Murray, S., Knowles, B.B., and Knoepfler, P.S. (2010). myc maintains embryonic stem cell pluripotency and self-renewal. Differentiation *80*, 9–19.

Varlakhanova, N., Cotterman, R., Bradnam, K., Korf, I., and Knoepfler, P.S. (2011). Myc and Miz-1 have coordinate genomic functions including targeting Hox genes in human embryonic stem cells. Epigenetics Chromatin *4*, 20.

Wang, T., Wei, J.J., Sabatini, D.M., and Lander, E.S. (2014). Genetic screens in human cells using the CRISPR-Cas9 system. Science *343*, 80–84.

Whyte, W.A., Bilodeau, S., Orlando, D.A., Hoke, H.A., Frampton, G.M., Foster, C.T., Cowley, S.M., and Young, R.A. (2012). Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. Nature *482*, 221–225.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307–319.

Wong, D.J., Liu, H., Ridky, T.W., Cassarino, D., Segal, E., and Chang, H.Y. (2008). Module map of stem cell genes guides creation of epithelial cancer stem cells. Cell Stem Cell *2*, 333–344.

Young, R.A. (2011). Control of the embryonic stem cell state. Cell *144*, 940–954.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (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. Published online November 30, 2010.

Zhou, Y., Zhu, S., Cai, C., Yuan, P., Li, C., Huang, Y., and Wei, W. (2014). High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. Nature *509*, 487–491.

Zwaka, T.P., and Thomson, J.A. (2005). Differentiation of human embryonic stem cells occurs through symmetric cell division. Stem Cells *23*, 146–149.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Streptavidin-HRP	GE Healthcare	Cat#RPN1231;
Oct-3/4 (N-19) antibody	Santa Cruz	Cat#sc-8628; RRID: AB_653551
c-Myc antibody	Cell Signaling Technology	Cat#9402; RRID: AB_2151827
H3K4me3 antibody	Abcam	Cat#ab1012; RRID: AB_442796
H3K9ac antibody	Abcam	Cat#ab4441; RRID: AB_2118292
RNA Pol II CTD, phos. Serine 5 [3E8] antibody	Chromotek	Cat#3e8-5; RRID: AB_2631404
RNA Pol II CTD, phos. Serine 2 [3E10] antibody	Chromotek	Cat#3e10-5; RRID: AB_2631403
RNA Pol II CTD, unphosphorylated CTD [1C7] antibody	Chromotek	Cat#1c7-5; RRID: AB_2631402
Bacterial and Virus Strains		
E.cloni 10G ELITE electrocompetent cells	Lucigen	Cat#60052-4
Critical Commercial Assays		
QIAGEN MinElute PCR purification kit	QIAGEN	Cat#28006
APC BrdU Flow Kit	BD	Cat#552598
Vector Red Substrate Kit	Vector Lab	Cat#SK-5100
Agilent Seahorse XF Cell Mito Stress Test Kit	Agilent	Cat#103015-100
iQ SYBR Green Supermix	Bio-Rad	Cat#1708882
iScript cDNA Synthesis Kit	Bio-Rad	Cat#1708891
RNeasy Mini Kit	QIAGEN	Cat#74106
rRNA Depletion Kit	NEB	Cat#E6310L
NEBNext Ultra Directional RNA Library Prep Kit	NEB	Cat#E7420L
NEBNext ChIP-seq Library Prep Master Mix Set for Illumina	NEB	Cat#6240L
NEBNext Ultra DNA Library Prep Kit for Illumina	NEB	Cat#E7370L
Dynabeads MyOne Streptavidin T1 beads	Thermo Fisher Scientific	Cat#65601
Dynabeads Protein A	Thermo Fisher Scientific	Cat#10002D
Dynabeads Protein G	Thermo Fisher Scientific	Cat#10004D
NE-PER reagents	Thermo Fisher Scientific	Cat#78833
Deposited Data		
RNA-seq and ChIP-seq data	This paper	GEO: GSE113335
ChIP-seq data	Das et al., 2014	GEO: GSE43231
Mendeley dataset	This paper	https://doi.org/10.17632/gkp97zr5zy.1
Experimental Models: Cell Lines		
J1 mouse embryonic stem cells	ATCC	Cat#SCRC-1010, RRID:CVCL_6412
Taf5l, *Taf6l*, *c-Myc KO* mESCs	This paper	N/A
HEK293T	ATCC	Cat#CRL-3216; RRID: CVCL_0063
Experimental Models: Organisms/Strains		
Mouse embryonic fibroblast isolated from ROSA26-rtTA, Collagen-OKSM, Oct4-GFP mouse embryos	Stadtfeld et al., 2010	JAX: 011001; 008214; 006965
Recombinant DNA		
Lentiguide-Puro	Sanjana et al., 2014	Addgene Plasmid # 52963
LentiCRISPR-mCherry V6	Unpublished	Addgene Plasmid #99154
LentiCas9-Blast	Sanjana et al., 2014	Addgene Plasmid #52962

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and Algorithms		
BEDTools (2.26.0)	Quinlan and Hall, 2010	http://bedtools.readthedocs.io/en/latest/
Bowtie 2 (2.3.2)	Langmead and Salzberg, 2012	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
MACS2 (2.1.1.20160309)	Zhang et al., 2008	https://github.com/taoliu/MACS
DeepTools (2.5.3)	Ramírez et al., 2016	https://deeptools.readthedocs.io/en/develop/
STAR (2.5.3a)	Dobin et al., 2013	https://github.com/alexdobin/STAR
Salmon (0.8.2)	Patro et al., 2017	https://combine-lab.github.io/salmon/
wigToBigWig, Kent tools (v4)	Kent et al., 2010	http://hgdownload.cse.ucsc.edu/admin/exe/
DESeq2 R package (1.14.1; Bioconductor 3.4)	Love et al., 2014	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
DiffBind R package (2.2.12; Bioconductor 3.4)	Ross-Innes et al., 2012	https://bioconductor.org/packages/release/bioc/html/DiffBind.html
R statistical program (3.3.2)	N/A	https://www.R-project.org/
ggplot2 R package (2.2.1)	N/A	http://ggplot2.tidyverse.org
pheatmap R package (1.0.8)	N/A	https://cran.r-project.org/web/packages/pheatmap/index.html
NIH DAVID website (6.8)	Dennis et al., 2003	https://david.ncifcrf.gov/
Mendeley dataset	This paper	https://doi.org/10.17632/gkp97zr5zy.1

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact: Stuart Orkin (orkin@bloodgroup.tch.harvard.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mouse embryonic stem cells (mESCs)

Mouse ESCs (mESCs) were cultured in mouse ESC media that contains DMEM (Dulbecco's modified Eagle's medium) (Thermo Fisher Scientific) supplemented with 15% fetal calf serum (FCS) (Omega Scientific), 0.1mM b-mercaptoethanol (Sigma-Aldrich), 2mM L-glutamine (Thermo Fisher Scientific), 0.1mM nonessential amino acid (Thermo Fisher Scientific), 1% of nucleoside mix (Merck Millipore), 1000 U/mL recombinant leukemia inhibitory factor (LIF/ESGRO) (Merck Millipore), and 50U/mL Penicillin/Streptomycin (Thermo Fisher Scientific). mESCs were cultured at 37°C, 5% CO_2.

Reprogrammable mouse embryonic fibroblasts

Reprogrammable mouse embryonic fibroblasts (MEFs) cultures were prepared from rtTA3-OKSM embryos (embryonic day E13.5) that were heterozygous for the *Oct4*-GFP reporter, OKSM cassette and rtTA3 (Stadtfeld et al., 2010). Mice were mainained and handled under the approved IACUC protocol 2014-0044 (PI: Apostolou) at the Weill Cornell Medicine. Doxycycline-inducible-OKSM driven reprogrammable MEFs harboring *Oct4*-GFP reporter were cultured with DMEM supplemented with 10% fetal bovine serum (FBS) (Omega Scientific) and 2% penicillin-streptomycin at 37°C, 5% CO_2. For reprogramming into iPSCs, MEFs were cultured at 37°C, 5% CO_2 in mESC media supplemented with 2μg/mL doxycycline (dox) (Sigma-Aldrich) and 50ug/mL ascorbic acid (Sigma-Aldrich).

Induced pluripotent stem cells (iPSCs)

iPSCs were cultured on irradiated MEFs (Thermo Fisher Scientific) at 37°C, 5% CO_2 in mESC media.

Human embryonic kidney cells (HEK293T cells)

HEK293T cells were cultured with DMEM supplemented with 10% fetal bovine serum (FBS) (Omega Scientific) and 2% penicillin-streptomycin (Thermo Fisher Scientific). These cells were cultured at 37°C, 5% CO_2.

METHOD DETAILS

Mouse epigenetic CRISPR-Cas9 pooled library design

A list of epigenetic genes was generated through literature mining, which cover different classes of epigenetic regulators and ESC-specific TFs (Table S1). Positive controls were selected as genes with known roles in embryonic stem cell state (25 genes). All possible sgRNAs targeting GFP (of *Oct4*-GFP reporter) were also included as positive controls. Previously published non-targeting sgRNAs were included as negative controls (Canver et al., 2017). sgRNAs (6 sgRNAs/ gene) targeting coding sequences of all the selected genes were retrieved from the mouse GeCKOv2.0 library (Sanjana et al., 2014). In total, the epigenetic CRISPR-Cas9 pooled library contained of 2,335 sgRNAs, including 1,938 sgRNAs targeting the 323 identified epigenetic and ESC TF genes, 150 sgRNAs targeting the 25 genes with known roles in embryonic stem cell state, 119 sgRNAs targeting GFP and 128 non-targeting sgRNAs.

Mouse epigenetic CRISPR-Cas9 pooled library construction

All the sgRNA oligonucleotides of the library were synthesized as previously described (Shalem et al., 2014) using a B3 synthesizer (CustomArray, Inc.), pooled together, PCR amplified and cloned into Esp3I-digested plentiGuide-Puro (Addgene plasmid ID: 52963) lentiviral vector, using a Gibson assembly master mix (New England Biolabs). Gibson assembly products were transformed into electrocompetent cells (*E. cloni*, Lucigen) and plated on 245mm x 245mm square LB-agar plates to obtain sufficient number of bacterial colonies at a ~50 × library coverage. Bacterial colonies were collected from the plates, genomic DNA was isolated and plasmid libraries were prepared for high-throughput sequencing to confirm the representation of each sgRNAs in the pooled library.

Lentiviral production

HEK293T cells were seeded onto 15cm dishes ~24hrs prior to transfection. Cells were transfected at 80% confluence in 16ml of media with 8.75 μg of VSVG, 16.25 μg of psPAX2, and 25 μg of the CRISPR-Cas9 pooled lentiviral plasmids, using 150 μg of linear polyethylenimine (PEI) (Sigma-Aldrich). Media was changed with fresh media 16–24hrs of post-transfection. Lentiviral supernatant was collected at 48 and 72hrs post-transfection and subsequently concentrated by ultracentrifugation (24000 rpm, 4°C, 2hrs) (Beckman Coulter SW32).

CRISPR-Cas9 mediated screen in mESCs

Oct4-GFP reporter mESCs with stably expressed Cas9 were transduced at low multiplicity of infection (MOI) to avoid more than one lentiviral integration per cell. Test transductions were performed to estimate the amount of virus required to ensure 30% transduction rate. 10 μg/mL blasticidin (InvivoGen) and 1 μg/mL puromycin (Sigma-Aldrich) were added 24hrs after transduction to select for lentiviral library integrants (puromycin resistant) in cells with Cas9 (blasticidin resistant). Cells were selected for next 5days, before FACS sorting. The GFP-low and GFP-high cells were sorted, genomic DNA was isolated, and libraries were prepared for deep sequencing to enumerate the existence of sgRNAs in these cell populations. The screening was performed in biological triplicates.

Generation of mESC *KO* deletion clones

Paired sgRNAs were designed to delete critical coding exons of the selected candidate genes. These sgRNAs were cloned into lentiguide-Puro (Addgene ID: 52963) plasmid, using Golden Gate Cloning approach as mentioned previously. Wild-type (J1) mESCs were transduced with Cas9-blast virus (generated from pLentiCas9-Blast, Addgene ID: 52962) and selected with 10 μg/mL blasticidin (InvivoGen) to generate the J1 mESC cell line with stable expression of Cas9 (mESC+Cas9). 50,000 of mESC+Cas9 cells were transfected with 500ng of each sgRNAs (5 and 3′ sgRNAs) using Lipofectamine 2000 (Thermo Fisher Scientific), and cells were selected with 1 μg/mL puromycin (Sigma-Aldrich) for 3-4 days. Next, puromycin-resistant cells were re-plated on a 15cm plate as single cells to grow individual clones. The individual clones were picked, expanded and genotyped PCR was performed to detect the homozygous or biallelic deletion or *knockout (KO)* mESC clones.

Somatic cell reprogramming or iPSC generation

Reprogramming was performed by seeding Doxycycline-inducible-OKSM driven reprogrammable MEFs harboring *Oct4*-GFP reporter, onto gelatin-coated 6-well plates (Stadtfeld et al., 2010). These cells were transduced with lentiviral particles of sgRNAs carrying Cas9 and mCherry. 24hrs post-transduction, mCherry+ MEFs were sorted, ~50,000 mCherry+ MEFs were plated on irradiated MEFs at 6-well format, and cultured in mESC media with 2μg/mL doxycycline (Sigma-Aldrich) and 50ug/mL ascorbic acid (Sigma-Aldrich). At day12, doxycycline was withdrawn, and cells were cultured for an additional 4 days with mESC media to obtain transgene-independent iPSCs. However, we obtained different percentages of mixed populations of iPSCs and differentiated cells after the reprogramming upon perturbation of *Taf5l* and *Taf6l* using their targeting sgRNAs. Non-targeting sgRNA was used as a control. The bulk/mixed populations of iPSCs and differentiated cells after reprogramming were used for RT-PCR analysis, FACS and Alkaline Phosphatase (AP) staining. Indel frequencies of *Taf5l* and *Taf6l* genes were assessed by targeted amplicon sequencing before (from mCherry+ reprogrammable MEFs transduced with viral particles of *Taf5l*- and *Taf6l*-targeting sgRNAs that carry mCherry and Cas9) and after (mixed populations of iPSCs and differentiated cells) the reprogramming. Indel frequencies of *Taf5l* and *Taf6l* alleles were quantified using the web-based tool– CRISPRESSO2 (http://crispresso.pinellolab.partners.org/).

Flow cytometry

Cells were dissociated using trypsin, washed with 1XPBS, followed by sorting. i) During the CRISPR-Cas9 screening, mESCs were sorted based on GFP-low and GFP-high; ii) percentages (%) of GFP (high and low) cells were quantified from *Oct4*-GFP reporter mESCs upon depletion of *Taf5l* and *Taf6l* using their targeting sgRNAs; iii) SSEA1+ve cells were measured from wild-type, *Taf5l* and *Taf6l KOs;* iv) During reprogramming, reprogrammable MEFs were also sorted based on mCherry.

Intracellular FACS

Cells were dissociated using trypsin; then fixed and permeabilized using eBioscience Fixation & Permeabilization Buffer Set (88-8824, eBioscience). Permeabilized Cells were incubated with PE-Conjugated-OCT3/4 Monoclonal Antibody (1:100; 12-5841-80: eBioscience) and its Isotype control (1:100; 12-4321-80:eBioscience) overnight at 4°C. Next, cells were washed 3 times and resuspended in 200 μL of permeabilization wash buffer, and analyzed through FACS Calibur.

Immunofluorescence microscopy

mESCs were fixed with 4% paraformaldehyde. The primary antibody against OCT4 (N-19: sc-8628, Santa Cruz) was used, and the secondary antibody was obtained from Jackson ImmunoResearch (705-585-003). Cells were imaged using a Nikon Eclipse fluorescence microscope. Mean fluorescence intensity (MFI) of OCT4+ cells were quantified using ImageJ.

Alkaline Phosphatase (AP) staining

AP staining of ESCs and iPSCs was performed according to the manufacture's protocol (Vector Red Substrate Kit, SK-5100, Vector Lab).

Cell Cycle analysis

Cell cycle analysis was performed by BrdU staining using the APC BrdU Flow Kit (BD). Cells were incubated with BrdU for 15 min and analyzed using a BD Accuri C6 Plus flow cytometer.

Cell proliferation assay

∼10,000 mESCs were plated on each well of the 6-well plates and cultured in mouse ESC (mESC) media up to 6 days. Cells were dissociated using trypsin, washed with 1XPBS, stained with trypan blue, and the live cells were counted using Countess II Automated Cell Counter (thermos Fisher Scientific) at Day 2, 3, 4, 5 and 6.

Measurement of metabolic functions in mESCs

A Seahorse Bioscience XFe24 Extracellular Flux Analyzer (Agilent) was used to measure– the oxygen consumption rate (OCR), related to OxPhos; and extracellular acidification rate (ECAR), related to glycolytic function from mESCs in XFe24 FluxPak Mini cell culture microplates as described previously (Sun et al., 2015). mESCs were seeded onto gelatin-coated XFe24 microplates at a density of 1×10^5 per well in 500μl of mESC media, and then incubated at 37°C with 5% CO2. After cell attachment (∼5 hours later) in mESC media, the media was replaced with 500ul of Seahorse assay media. Measurements of oxygen consumption rate (OCR) and extracellular acidification rate (ECAR) were performed after equilibration of cells at 37°C in an anaerobic incubator for 1 hour. OCR was measured in cells that were treated with oligomycin (1 μM), FCCP (1 μM) and rotenone (1 μM) (Agilent Seahorse XF Cell Mito Stress Test Kit (103015-100, Agilent); ECAR was measured in cells that were treated with glucose (10mM), oligomycin (1 μM) and 2-DG (50mM) (Agilent Seahorse XF Glycolysis Stress Test Kit (103020-100, Agilent) according to the manufacturer's instructions. OCR is reported in the unit of picomoles per minute and ECAR is reported in milli-pH units (mpH) per minute. OCR and ECAR were normalized to protein content (determined via Bradford assay) in each well. Basal respiration, ATP production, maximum respiratory capacity (related to OxPhos); and glycolysis, glycolytic capacity and glycolytic reserve (related to glycolytic function) were calculated as per the manufacturer's instructions.

Generation of Flag-Biotin (FB) tagged mESC lines

Mouse *Taf5l* and *Taf6l* ORFs were synthesized (IDT), and cloned into BamHI-digested pEF1a-Flagbio(FB)-puromycin vector, using Gibson Assembly Master Mix (NEB). Positive clones were analyzed by Sanger sequencing. 10μg of *Taf5l*-FB and *Taf6l*-FB constructs were electroporated into 5×10^6 J1 wild-type mESCs, which constitutively express BirA ligase (neomycin). The electroporated cells were plated on a 15cm dish, with mESC media. After ∼24hrs, media was replaced with fresh mESC media containing 1 μg/mL of puromycin (Sigma-Aldrich) and 1 μg/mL of neomycin (Sigma-Aldrich), and cells were selected for 4-5days. Individual ESC colonies were picked, expanded, and tested by western blot using streptavidin-HRP (GE healthcare) (dilution 1:2000 in 5% BSA) to detect the clones that expressed TAF5L-FB and TAF6L-FB separately.

Pull down of Flag-Biotin tagged TAF5L and TAF6L, and identification of their interacting partners through mass spectrometry

One-step affinity purification with streptavidin-beads was performed as previously reported (Kim et al., 2009). In brief, nuclear extracts were obtained from 300millions cells using NE-PER reagents (Thermo Fisher Scientific) and quantified using DC Protein Assay

(Biorad). 5mg of nuclear extract from mESCs either expressing only BirA (control) or BirA plus biotin-tagged protein of interest (for example, TAF5L-FB/TAF6L-FB) were diluted in 10 mL of IP150 or IP350 buffer (20mM Tris-HCl (pH 7.5), 0.3% (vol/vol) NP-40, 1mM EDTA, 10% (vol/vol) glycerol (vol/vol), 1mM DTT, 0.2mM PMSF and protease inhibitor cocktail (Sigma 8340, 1: 1000), and 150 mM or 350 mM NaCl). 100ul Dybabeads MyOne Streptavidin T1 beads (Thermo Fisher Scientific) were added and incubated with rotation overnight at 4°C. Beads were washed 4 times with 10ml of IP buffer and eluted in 250ul NP40 Lysis buffer (50mM Tris-HCl, 150mM NaCl, 1% NP-40, and 5mM EDTA) supplemented with 1% SDS. Samples were concentrated using TCA, boiled in Laemmli buffer for 5 min, and resolved on a 10% SDS-polyacrylamide gel. Whole lane LC-MS/MS sequencing and peptide identification were performed at the Taplin Biological Mass Spectrometry Facility at Harvard Medical School.

Western Blot
Protein extract was mixed with Laemmli buffer, boiled for 5 min, and resolved on a 4%–12% gradient Bis-Tris gel (Bio-Rad). Proteins are on the gel transferred to the PVDF membrane, and specific antibodies were used to detect proteins of interest.

RNA isolation and RT-qPCR
RNA was isolated from cells using RNeasy Mini Kit (QIAGEN), and cDNA was prepared using iScript cDNA Synthesis Kit (Bio-Rad). RT-qPCR was performed using iQ SYBR Green Supermix (Bio-Rad) on Bio-Rad iCycler RT-PCR detection system.

RNA-seq
Total RNA was isolated, and ribosomal RNAs were depleted using ribosomal RNA depletion kit (rRNA Depletion Kit, NEB). Ribosomal depleted RNA was used to make the RNA-seq libraries using NEBNext Ultra Directional RNA Library Prep Kit for Illumina (NEB). All the libraries were checked through a Bio-analyzer for quality control purpose. Single End (SE) 100bp reads were generated using a HiSeq-2500 sequencer (Illumina). RNA-seq were performed in triplicates.

ChIP-seq
Chromatin Immunoprecipitation (ChIP) was performed as described previously (Das et al., 2014; Kim et al., 2009). For bioChIP reactions, streptavidin beads (Dynabeads MyOne Streptavidin T1- Invitrogen) were used for the precipitation of chromatin, and 2% SDS was applied for first washing step. All other steps were same as conventional ChIP protocol. BirA expressing J1 mESCs were used as a control. Conventional ChIP reactions were performed as described previously (Das et al., 2014, 2015). Input genomic DNA was used for the reference sample. Briefly, cells were tryipsinized from 15cm dishes, washed twice with 1XPBS, and crosslinked with 37% formaldehyde solution (Calbiochem) to a final concentration of 1% for 8 min at room temperature with gentle shaking. The reaction was quenched by adding 2.5M glycine to a final concentration of 0.125M. Cells were washed twice with 1XPBS, and cell pellet was resuspended in SDS-ChIP buffer (20 mM Tris-HCl pH 8.0, 150mM NaCl, 2 mM EDTA, 0.1% SDS, 1% Triton X-100 and protease inhibitor), and chromatin was sonicated to around 200-500 bp. Sonicated chromatin was incubated with 5∼10μg of antibody overnight at 4°C. After overnight incubation, protein A/G Dynabeads magnetic beads (Thermo Fisher Scientific), were added to the ChIP reactions and incubated for 2-3 hours at 4°C to immunoprecipitate chromatin. Subsequently, beads were washed twice with 1 mL of low salt wash buffer (50mM HEPES pH 7.5, 150mM NaCl, 1mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate), once with 1ml of high salt wash buffer (50mM HEPES pH 7.5, 500mM NaCl, 1mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate), once with 1ml of LiCl wash buffer (10mM Tris-HCl pH 8.0, 1mM EDTA, 0.5% sodium deoxycholate, 0.5% NP-40, 250 mM LiCl), and twice with 1ml of TE buffer (10mM Tris-HCl pH 8.0, 1mM EDTA, pH 8.0). The chromatin was eluted twice in SDS elution buffer (1% SDS, 10mM EDTA, 50mM Tris-HCl, pH 8.0) at 72°C, each time with 150μl of SDS elution buffer. Eluted chromatin was reverse-crosslinked at 65°C overnight. The next day, an equal volume of TE was added (300μl). ChIP DNA was treated with 1μl of RNaseA (10mg/mL) for 1hr, and with 3μl of proteinase K (20mg/mL) for 3hrs at 37°C, and purified using phenol-chloroform extraction, followed by QIAquick PCR purification spin columns (QIAGEN). Finally, ChIP-DNA was eluted from the column with 40μl of water. For several factors, we used multiple ChIPs. At the end, all eluted ChIP-DNA samples were pooled and precipitated to enrich the ChIP-DNA material to make the libraries for high-throughput sequencing. Input ChIP samples were reserved before adding the antibodies, and these input samples were processed from reverse cross-linking step until the end, same as other ChIP samples.

2-10 ng of purified ChIP DNA was used to prepare sequencing libraries, using NEBNext ChIP-seq Library Prep Master Mix Set for Illumina (NEB) and NEBNext Ultra DNA Library Prep Kit for Illumina (NEB) according to the manufacturer's instructions. All libraries were checked through a Bio-analyzer for quality control purpose. ChIP sequencing (50bp SE reads) was performed using Hiseq-2500 (Illumina).

QUANTIFICATION AND STATISTICAL ANALYSIS

CRISPR-Cas9 screen data analysis
sgRNA sequences were enumerated from the GFP-low and GFP-high populations. An enrichment score for each sgRNA was calculated as the log2 transformation of the median number of occurrences of a given sgRNA in the GFP-low population divided by the median number of occurrences of the same sgRNA in the GFP-high population. The enrichment score for each gene was calculated as the median of the 3 most enriched sgRNA targeting each gene across three biological replicates.

RNA-seq analysis

Reads were mapped to the mm9 mouse genome assembly and the GENCODE version M1 mouse transcriptome (Mudge and Harrow, 2015) using the STAR RNA-seq read alignment program (v2.5.3a) (Dobin et al., 2013) with default parameters. Bigwig tracks were generated using STAR to generate RPM-normalized wiggle files. Subsequently the wigToBigWig tool (v4) from the UCSC genome browser Kent tools (Kent et al., 2010) was used to convert the wiggle files to BigWig format. Gene expression quantification, both raw counts and TPM (Transcripts Per Million) values, were obtained using Salmon (v0.8.2) (Patro et al., 2017). Differential gene expression analysis was done using the DESeq2 (v1.14.1) (Love et al., 2014) R statistical programming language (v3.3.2) software package from the Bioconductor project (v3.4) (Huber et al., 2015) using the GENCODE version M1 mouse transcriptome. DESeq2 was also used to calculate the normalized counts and the FPKM (Fragments Per Kilobase per Million mapped reads) expression values. MA plots were created employing the ggplot2 R package (v2.2.1) that used log2 fold changes and mean normalized counts calculated by the DESeq2 R package.

GO (Gene Ontology)

Gene Ontology (Ashburner et al., 2000) analysis was performed using the NIH DAVID website (v6.8) (Dennis et al., 2003); default settings and the whole genome were used as background. Genes were identified using their Ensembl Gene IDs for the DAVID analysis, and their associated gene names were obtained from the GENCODE vM1 mouse transcriptome annotation and added to the resulting GO term enrichment tables.

ChIP-seq analysis

The 50bp single end (SE) reads were mapped to the mm9 mouse genome assembly using Bowtie 2 (v2.3.2) (Langmead and Salzberg, 2012). BigWig tracks were generated using the 'bamCoverage' program from the DeepTools software suite (v2.5.3) (Ramírez et al., 2016), with reads extended to 300bp and tracks normalized using the RPKM (Reads Per Kilobase per Million mapped reads) algorithm. Peak calling was done with the MACS2 program (v2.1.1.20160309) (Zhang et al., 2008). For most ChIP-seq datasets default MACS2 settings were used. For the TAF5L-FB and TAF6L-FB datasets, MACS2 was run with the "nomodel" option. Artifact peaks were filtered out based on the mouse ENCODE project (Stamatoyannopoulos et al., 2012), They were removed using "bedtools intersect" from the BEDTools software suite (v2.26.0) (Quinlan and Hall, 2010). Mapping of peaks to the nearest gene was done using "bedtools closest." Genomic distribution of peaks was performed using "bedtools intersect" using annotated genomic regions taken from the UCSC mm9 known gene transcriptome as implemented in the Bioconductor "TxDb.Mmusculus.UCSC.mm9.knownGene" package, as well as mouse enhancers taken from the literature (Whyte et al., 2013). As single peak can overlap multiple genomic regions, the following annotation hierarchy was used to prevent counting the same peak multiple times: promoters, enhancers, exons, introns, intergenic regions. Each peak was mapped only to the first annotation type it overlapped.

Heat-map

ChIP-seq datasets were generated using the "computeMatrix" and "plotHeatmap" programs of the DeepTools software suite (v.2.5.3) (Ramírez et al., 2016) for ChIP-seq heatmaps and profile plots of histone marks and transcription factors binding. Regions from −2kb to +2kb around the center of the peaks were used, split into 50bp bins (default for "computeMatrix"), with each bin receiving the mean of the BigWig signal scores across it.

Hierarchical clustering

Hierarchical clustering of genome-wide binding profiles for different transcription factors was accomplished as follows: First, 500bp tiling windows were created along the entire mm9 genome using "bedtools makewindows" from the BEDTools software suite (v2.26.0) (Quinlan and Hall, 2010). These genomic windows were overlapped with the peak BED files for all the transcription factors using "bedtools annotate," resulting in a peak overlap matrix. This matrix was loaded into the R statistical program (v3.3.2) and converted into a binary matrix, thus counting each window-peak overlap only once even if the window overlaps multiple peaks from the same transcription factor ChIP-seq dataset. Empty rows, corresponding to universally unbound genomic windows, were discarded. This binary matrix was subsequently used to calculate a Pearson correlation matrix between transcription factors. Finally, the correlation matrix was hierarchically clustered and plotted using the pheatmap (v1.0.8) R package, using default settings for the hierarchical clustering (Euclidean distance metric and complete linkage clustering).

Module activity

Differential expression of CORE, MYC and PRC module genes were plotted based on the log2 fold changes (log2FC) between *Taf5l/Taf6l KO* and wild-type normalized count-based gene expression, as calculated by the DESeq2 R package (v1.14.1) (Love et al., 2014). Log2FC for module genes were plotted as violin plots using the ggplot2 (v2.2.1) R package, and the statistical significance of differential expression was calculated using the Wilcoxon rank sum test with continuity correction, as implemented in the R statistical program (v3.3.2).

Violin Plots/ differential binding

Differential binding analyses were performed using the DiffBind (v2.2.12) (Ross-Innes et al., 2012) R package from the Bioconductor project (v3.4) (Huber et al., 2015). For each *KO* versus wild-type (WT) differential binding comparison, peak sets were merged between *KO* and *WT* and the differential binding signal was calculated from the BAM file reads of the merged peak sets. Violin plots of the *KO* versus *WT* log2 fold change of normalized read densities, as calculated by DiffBind, were generated using the ggplot2 (v2.2.1) R package. Statistical significance of overall differential binding was calculated using the Wilcoxon rank sum test with continuity correction as implemented in the R statistical program (v3.3.2).

Calculation of traveling ratio of RNAP, RNAP-Ser 5p and RNAP-Ser 2p

Traveling Ratio calculations were performed according to the method used by Rahl et al., 2010 (Rahl et al., 2010). The GENCODE version M1 mouse transcriptome was used (Mudge and Harrow, 2015), and the transcripts were filtered to remove very short transcripts (< 1kb genomic length). In accordance with Rahl et al., 2010, promoters were defined as the region from 30bp upstream of the Transcription Start Site (TSS-30bp) to 300bp downstream of it (TSS+300bp). Gene bodies were defined as TSS+300bp to TES (Transcription End Site). Promoter and gene body RPKMs were calculated using "bedtools intersect" (v2.26.0) to count overlapping RNAP/RNAP-Ser 5p/RNAP-Ser 2p reads with transcript promoters & gene bodies, followed by processing with Linux "awk" command (GNU Awk 4.0.2) to calculate the RPKMs. Traveling Ratios (TRs) were calculated per transcript from promoter & gene body RPKMs using the Linux "awk" command. Transcript-level TRs were subsequently converted into gene-level TRs by selecting the highest-expressed transcript per gene. This was done using R (v3.3.2), by selecting the promoter with highest RPKM per gene, followed by the gene body with highest RPKM for this promoter, to obtain one canonical transcript per gene. For comparability across the WT and both *Taf5l* & *Taf6l KOs*, the canonical transcripts were determined in the WT samples, with the same transcript set being used for both WT and *KOs*.

Metagene profile plots of RNAP2 binding along the gene body were created using the "computeMatrix" and "plotProfile" tools of the DeepTools software suite (v2.5.3) (Ramírez et al., 2016), whereby the gene body length was scaled to 10kb with 2kb flanking regions, and mean BigWig scores were calculated per bin.

Statistical analysis

All statistical analyses for the RNA-seq and ChIP-seq data were performed using the R statistical program (v3.3.2).

DATA AND SOFTWARE AVAILABILITY

The accession number for all the NGS data is GSE113335.