Cell

Regulation of Pluripotency and Self-Renewal of ESCs through Epigenetic-Threshold Modulation and mRNA Pruning

Junling Jia,[1,2] Xiaobin Zheng,[1] Gangqing Hu,[3] Kairong Cui,[3] Junqi Zhang,[1,6] Anying Zhang,[1,7] Hao Jiang,[1] Bingwen Lu,[5,8] John Yates III,[5] Chengyu Liu,[4] Keji Zhao,[3] and Yixian Zheng[1,2,*]

[1]Department of Embryology, Carnegie Institution for Science, 3520 San Martin Drive, Baltimore, MD 21218, USA
[2]Howard Hughes Medical Institute
[3]Systems Biology Center
[4]Transgenic Core Facility
National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD 20892, USA
[5]Department of Chemical Physiology, The Scripps Research Institute, La Jolla, CA 92037, USA
[6]Present address: Department of Medical Microbiology, Shanghai Medical College, Fudan University, Shanghai, China
[7]Present address: School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, China
[8]Present address: Pfizer, 401 North Middletown Road, Pearl River, NY 10965, USA
*Correspondence: zheng@ciwemb.edu
http://dx.doi.org/10.1016/j.cell.2012.09.023

SUMMARY

Embryonic stem cell (ESC) pluripotency requires bivalent epigenetic modifications of key developmental genes regulated by various transcription factors and chromatin-modifying enzymes. How these factors coordinate with one another to maintain the bivalent chromatin state so that ESCs can undergo rapid self-renewal while retaining pluripotency is poorly understood. We report that Utf1, a target of Oct4 and Sox2, is a bivalent chromatin component that buffers poised states of bivalent genes. By limiting PRC2 loading and histone 3 lysine-27 trimethylation, Utf1 sets proper activation thresholds for bivalent genes. It also promotes nuclear tagging of messenger RNAs (mRNAs) transcribed from insufficiently silenced bivalent genes for cytoplasmic degradation through mRNA decapping. These opposing functions of Utf1 promote coordinated differentiation. The mRNA degradation function also ensures rapid cell proliferation by blocking the Myc-Arf feedback control. Thus, Utf1 couples the core pluripotency factors with Myc and PRC2 networks to promote the pluripotency and proliferation of ESCs.

INTRODUCTION

Embryonic stem cells (ESCs) undergo rapid self-renewal and can differentiate into any cell type. These features depend on transcription factors including Oct4, Sox2, and Nanog (Boyer et al., 2005; Chen et al., 2008; Kim et al., 2008; Loh et al., 2006) that form the core pluripotency network (Orkin and Hochedlinger, 2011). This ESC-specific network interacts with both the Myc-based transcription network and a network of chromatin-modifying complexes including the polycomb-repressive complex 2 (PRC2). Together these three networks occupy and regulate a large number of target genes essential for the self-renewal and differentiation of ESCs (Boyer et al., 2006; Hu et al., 2009; Kim et al., 2010; Ku et al., 2008; Lee et al., 2006). Although a few factors have been shown to coordinate some aspects of these networks in ESCs (Orkin and Hochedlinger, 2011), how all three networks are functionally integrated is unknown.

The Myc-centered network promotes proliferation while regulating lineage-specific differentiation (Lin et al., 2009; Smith et al., 2010; Varlakhanova et al., 2010). However, Myc also activates the *Arf* tumor suppressor encoded by *Cdkn2a* (also called *Ink4a-Arf*), which in turn inhibits cell proliferation. In somatic cells, this feedback mechanism prevents uncontrolled cell proliferation, and its inactivation by mutations leads to tumorigenesis (Cleveland and Sherr, 2004; Eischen et al., 1999). Whereas cancer cells evade the Myc-Arf feedback loop through mutations, how ESCs block this feedback to allow rapid proliferation is unknown. Oct4, Sox2, and Nanog are required for ESCs to maintain a low level of *Arf* messenger RNA (mRNA) (Banito et al., 2009; Kawamura et al., 2009; Li et al., 2009; Utikal et al., 2009). As these transcription factors do not directly bind to *Cdkn2a*, it is unclear how the pluripotency core antagonizes the Myc-Arf feedback loop to ensure rapid ESC self-renewal.

Besides specific transcriptional regulation, two unique chromatin features contribute to the establishment and maintenance of ESC pluripotency. A less compacted chromatin structure compared to differentiated cells endows ESCs with highly dynamic chromatin organization (Meshorer and Misteli, 2006). Another chromatin feature of ESCs is the so-called poised bivalent state of developmentally regulated genes that harbor

both transcriptionally active (histone 3-lysine 4 trimethylation, H3K4me3) and transcriptionally repressive (histone 3-lysine 27 trimethylation, H3K27me3) epigenetic marks (Bernstein et al., 2006), which are catalyzed by SET/MLL complexes and PRC2, respectively (Ang et al., 2011; Boyer et al., 2006; Jiang et al., 2011; Landeira et al., 2010; Li et al., 2010; Peng et al., 2009; Shen et al., 2009; Pasini et al., 2010). These chromatin features lead to a global transcriptional activity so that even repressed developmentally regulated genes are expressed sporadically (Efroni et al., 2008). Given that these mRNAs neither accumulate to high levels nor translate into large amounts of proteins in ESCs, active mechanisms must exist to prevent their accumulation and translation. Understanding these mechanisms should shed light on these unique properties of ESCs.

The PRC2-catalyzed H3K27me3 is essential for maintaining both the silenced and poised states of bivalent genes. PRC2 recognizes both CpG islands and chromatin features within the GC-rich regions where most bivalent genes reside (Zhou et al., 2011). However, because both CpG-island densities and chromatin contexts of bivalent genes vary widely, additional regulatory mechanisms are likely required to ensure that appropriate amounts of PRC2 are loaded onto individual bivalent genes in ESCs so that they are neither overly nor insufficiently silenced. Indeed, recent studies of Jarid2, a component of PRC2, have revealed complicated mechanisms of promoting PRC2 loading and activities on bivalent genes. In addition to promoting PRC2 loading, ESCs may actively limit PRC2 binding via a presently unknown mechanism so that the bivalent genes are not overly silenced. The function of PRC2 on bivalent genes must also be coordinated with the core pluripotency factors. As no physical interaction between the core factors and PRC2 has been detected, the core factors may be coupled to PRC2 through their downstream effector(s) in ESCs.

The *u*ndifferentiated embryonic cell *t*ranscription *f*actor *1* (*Utf1*) is one of the direct downstream target genes of Oct4 and Sox2, and it is highly expressed in mouse and human ESCs (Nishimoto et al., 2005; Okuda et al., 1998; Tan et al., 2007). The highest amount of *Utf1* is found in the inner cell mass of mouse blastocysts. After implantation, *Utf1* expression is silenced in most cells with the exception of some embryonic tissues (Okuda et al., 1998). In ESCs, Utf1 is only found in the nucleus where it tightly associates with chromatin, and during ESC differentiation, Utf1 is rapidly downregulated. Although Utf1 has been implicated in regulating ESC proliferation and differentiation, the mechanism remains unclear (Nishimoto et al., 2005; van den Boom et al., 2007). We report here that Utf1 is a component of the bivalent chromatin that regulates gene expression in a context-dependent manner, which connects the pluripotency core to both Myc and PRC2 networks to ensure rapid proliferation and coordinated differentiation of ESCs.

RESULTS

Utf1 Binds to Bivalent Genes to Regulate Their Expression

We expressed biotin-Utf1 at less than 5% of the endogenous Utf1 level (Figure S1A available online), and we used biotin-mediated and crosslinked ChIP-sequencing (biotin-ChIP-seq) to map Utf1-binding sites in ESCs (Kim et al., 2009, 2010; Shen et al., 2009). We identified 75,029 chromatin regions with significant Utf1 enrichments (*f*alse *d*iscovery *r*ate [FDR] < 0.001). Utf1 binding was enriched in gene-rich regions with highest binding near the transcription start sites (TSS) of genes containing dense CpG islands (Figures 1A–1C and S1B). Gene ontology (GO) analyses revealed a striking enrichment of Utf1 on genes with functions in organ/system development and cell differentiation (Figure S1C and Table S1).

PRC2-ChIP-seq using an antibody to Suz12 (a subunit of PRC2) revealed that Utf1 binding strongly correlated with PRC2 binding ($r = 0.71$; Figures 1B–1D and S1B). Our Suz12-ChIP-seq data set was consistent with previously published data sets ($r = 0.82$) (Ku et al., 2008). Of the total 16,380 Utf1-bound genes, ~6,116 were bivalent genes (Table S2), and they exhibited significantly stronger Utf1 enrichment within 5 kb up- and downstream of the TSS than those nonbivalent genes (Figure 1C). Utf1 binding was highly correlated ($r = 0.6$) with H3K27me3 catalyzed by PRC2 but poorly correlated with H3K4me3 ($r = 0.21$) present on both bivalent promoters and promoters of strongly expressed genes (Figures 1B, S1B, and S1D). Utf1 binding was not correlated with H3K9me3 ($r = -0.17$) found on heterochromatin (Figure S1E). Using different motif-find methods (Bailey and Elkan, 1994; Zheng et al., 2011), we predicted two similar AG-rich motifs recognized by Utf1 within CpG islands (Figure 1E) closely resembling one of the two motifs previously predicted to bind to Jarid2 in PRC2 (Peng et al., 2009). Thus, Utf1 is preferentially enriched at the promoters of bivalent genes.

To study how Utf1 regulates gene expression, we generated *Utf1* null ($Utf1^{-/\Delta}$) ESCs by gene targeting (Figures S1F and S1G). Compared to control $Utf1^{+/f}$ ESCs, the $Utf1^{-/\Delta}$ ESCs expressed similar amounts of pluripotency proteins and Suz12 (Figure S1H) and maintained similar levels of euploidy under standard ESC culture conditions (Figure S1I). Using RNA-seq, we found that 792 genes exhibited changes of expression by at least 1.5-fold ($p < 10^{-5}$) in $Utf1^{-/\Delta}$ ESCs compared to $Utf1^{+/f}$ ESCs. Importantly, Utf1 directly bound to 86.6% of the down- and 90.3% of the upregulated genes (Figures 1F and 1G; Tables S3 and S4). The majority of these are bivalent genes. Thus Utf1 could regulate either repression or activation of bivalent genes in ESCs.

Utf1 Limits Bivalent Gene Silencing by Preventing Excessive PRC2 Binding and H3K27me3

Because Utf1 is strongly enriched on bivalent genes (Figures 1B and 1C), we focused our study on the bivalent genes. We first studied whether Utf1 could limit gene silencing by preventing excessive PRC2 binding and H3K27me3 on bivalent genes because Utf1 and PRC2 were predicted to bind to similar DNA sequences (Figure 1E). ChIP-seq showed that $Utf1^{-/\Delta}$ ESCs had up to 4-fold increase in PRC2 binding to bivalent genes compared to $Utf1^{+/f}$ ESCs, which corresponded to a significant increase in H3K27me3 on these genes (Figures 2A–2D and S2A). Further ChIP analyses of histone H3 and H3K4me3 demonstrated that increased PRC2 binding and H3K27me3 on bivalent genes were not due to some general change of chromatin upon Utf1 deletion (Figures 2E, 2F, S2B, and S2C).

We then used four means to show that Utf1 and PRC2 indeed competitively bound to the same bivalent genes. First, we used

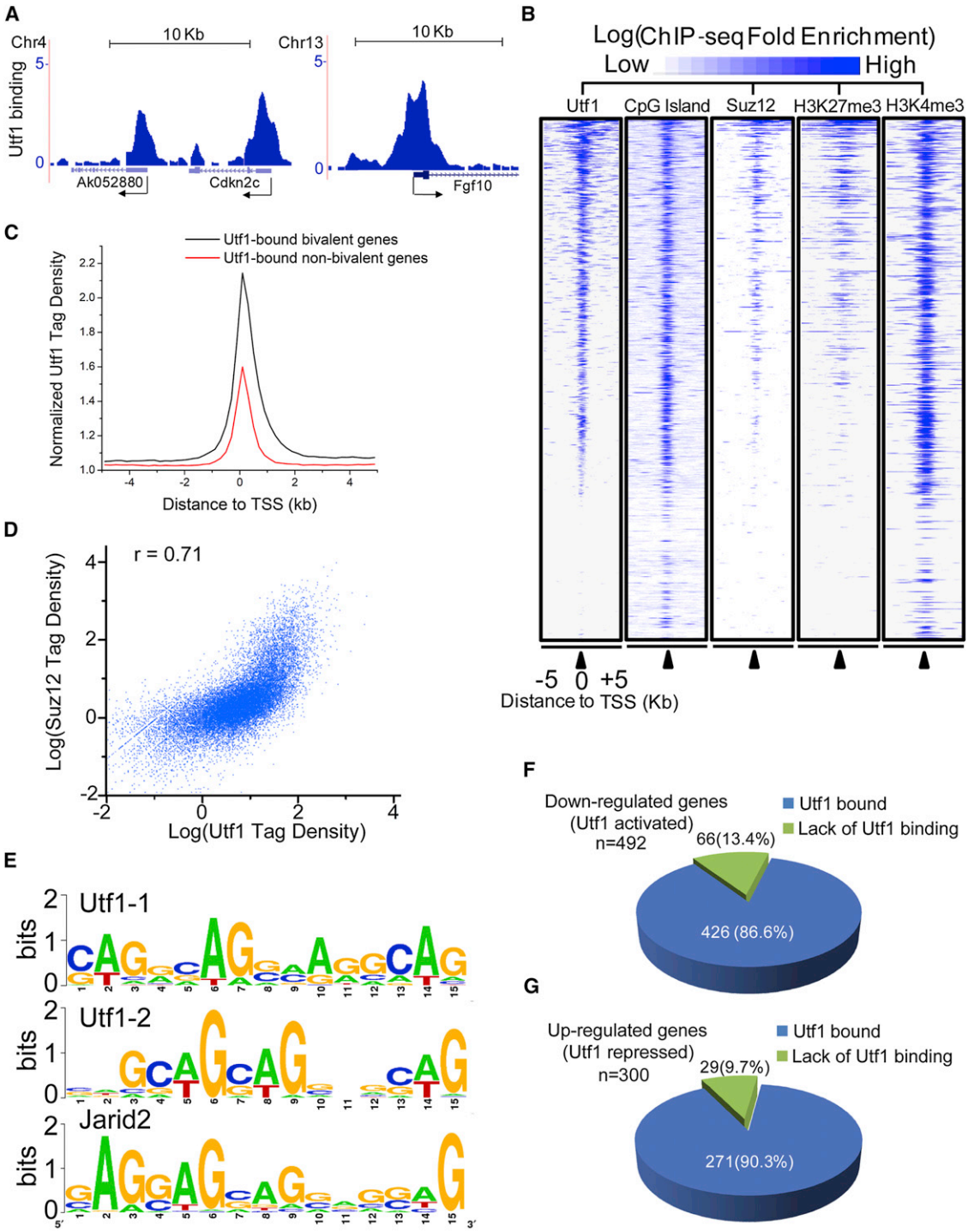

Figure 1. Utf1 Binds Bivalent Genes to Regulate Their Expression

(A) Utf1 binding at bivalent promoters in ESCs. Biotin-ChIP-seq of BirA-expressing ESCs showed no significant binding of background biotinylated proteins to chromatin (data not shown). Arrows, the start and direction of transcription. y axes, *r*eads *p*er *k*ilobases per *m*illion reads (RPKM).

(B) Heatmap of annotated mouse-gene promoters enriched for Utf1, CpG, Suz12, H3K27me3 (Mikkelsen et al., 2007), and H3K4me3 in ESCs. ChIP-seq enrichment (identified in this study) for Utf1, Suz12, and H3K4me3 was calculated as the ratio of normalized tag counts of the ChIP-seq and the input sequence in a 1 kb window that slides every 200 bp along 10 kb promoter regions. The heatmap is rank-ordered based on the enrichment of Utf1 (blue, enriched; white, not enriched).

(C) Utf1 tag density on bivalent genes is higher than on the nonbivalent genes. The normalized Utf1-tag densities, determined as the averaged ratio of normalized tag counts of Utf1-ChIP-seq and the input sample in the 200 bp window, are plotted within the 5 kb up- and downstream of TSS.

RNA interference (RNAi) to reduce the PRC2 subunit Jarid2 (Figure S2D) and performed biotin-Utf1-ChIP-qPCR on selected bivalent genes. This resulted in an increased binding of Utf1 to the bivalent genes (graph in Figure S2D). Second, we re-expressed Utf1 in *Utf1*$^{-/\Delta}$ ESCs (Figure S2E) and carried out Suz12-ChIP-qPCR on selected bivalent genes. After re-expressing Utf1, the increase of Suz12 binding to the bivalent genes in *Utf1*$^{-/\Delta}$ ESCs was reduced to levels similar to those seen in *Utf1*$^{+/f}$ ESCs (graph in Figure S2E). Third, we performed electrophoretic mobility shift assay (EMSA) to show that Utf1 and PRC2 competitively bound to the DNA oligo containing the predicted binding motifs for Utf1 and PRC2 (Figure S2F). Finally, sequential ChIP-qPCR analyses on selected bivalent genes showed that Utf1 and PRC2 co-occupied the same bivalent genes in wild-type ESCs (Figure S2G). Thus, Utf1 prevents oversilencing of bivalent genes by limiting PRC2 binding.

Utf1 Recruits the mRNA-Decapping Protein Dcp1a to Bivalent Genes

To understand how Utf1 could also repress gene expression, we used mass spectrometry (MS) to identify Utf1-interacting proteins. Among the proteins identified by two rounds of MS with a high peptide coverage (Table S5), we did not find either ATF2 or any subunit of TFIID that was shown to interact with Utf1 in somatic cells (Fukushima et al., 1998; Okuda et al., 1998). However, we identified components of the mRNA-decapping complex, Dcp1a, Ddx6, and Edc3, as candidate Utf1-interacting proteins in ESCs (Ling et al., 2011; Tritschler et al., 2009).

By pulling down biotin-Utf1 expressed in ESCs, we showed that whereas Dcp1a, Ddx6, and Edc3 coimmunoprecipitated with Utf1, the RNA polymerase II (Pol II) and Taf1, two abundant nuclear proteins, were not detectable in the immunoprecipitate (Figure 3A). Consistent with the absence of the catalytic subunit Dcp2 (cleaves the 5′-methyl cap of mRNAs) in the MS analyses (Table S5), Dcp2 was not detected in Utf1 immunoprecipitates (Figure 3A). We found that Dcp1a, Ddx6, and Edc3 were present in ESC nuclear extracts, whereas Dcp2 was only found in the whole-cell lysate (Figure 3B). Similarly, immunofluorescence staining revealed that whereas Dcp1a, Ddx6, and Edc3 were present in both the nucleus and the cytoplasm, Dcp2 was only found in the cytoplasm of ESCs (Figure 3C). Thus Utf1 binds to three noncatalytic subunits of the mRNA-decapping complex in the ESC nucleus.

A clear reduction of bright Dcp1a^{+} granules in both the nucleus and the cytoplasm was seen in ESCs upon Utf1 loss and was not due to the reduction of the decapping proteins (Figures 3D–3F). Instead, as formation of mRNP-processing granules requires efficient loading of proteins like Dcp1a onto mRNAs, the reduction of Dcp1a^{+} granules in *Utf1*$^{-/\Delta}$ ESCs suggests that Utf1 facilitates the recruitment of the noncatalytic subunits of the mRNA-decapping complex to bivalent promoters so they can be loaded onto newly transcribed mRNAs in the nucleus. Indeed, sequential ChIP-qPCR of randomly selected bivalent genes showed that Dcp1a and Utf1 co-occupied these genes (Figure 3G). Importantly, the binding of Dcp1a to bivalent genes decreased significantly upon Utf1 loss, which was rescued by re-introducing Utf1 into these ESCs (Figure 3H). Given that the reduction of Dcp1a binding in the absence of Utf1 is incomplete, additional factors may mediate Dcp1a recruiting.

Utf1 Promotes the Binding of Dcp1a to mRNA in the Nucleus for Cytoplasmic Degradation

RNA-seq showed that reduction of Dcp1a by two different short hairpin RNAs (shRNAs) in ESCs resulted in upregulation of many Utf1-bound bivalent genes, including *Arf* and *Hoxa1*, without affecting the expression of *Nanog*, *Oct4*, *Sox2*, *Utf1*, and *Suz12* (Figure 4A and Table S6). To show that Utf1 tags the mRNA from leaky bivalent promoters for cytoplasmic degradation by recruiting Dcp1a, we focused on the bivalent gene *Arf* because it exhibited a low level of expression in *Utf1*$^{+/f}$ ESCs, which was further upregulated in *Utf1*$^{-/\Delta}$ ESCs (Figures S3A–S3D). The low level of *Arf* expression in wild-type ESCs could be due to the binding of the transcriptional activator Myc (Figure S3A). Consistent with the idea that Utf1 represses *Arf* expression posttranscriptionally, both the nuclear run-on assay and the transcriptional elongation assay showed that Utf1 did not inhibit *Arf* transcription (Figures 4B and 4C). Treating ESCs with the Myc inhibitor 10058-F4 (Rahl et al., 2010) caused a similar reduction of *Arf* transcription in both *Utf1*$^{+/f}$ and *Utf1*$^{-/\Delta}$ ESCs (Figure 4B), which was consistent with a role of Myc in activating *Arf* transcription in ESCs. Finally, S1 nuclease protection assays showed that *Arf* mRNA was properly spliced in the nucleus of *Utf1*$^{+/f}$ ESCs (Figure S3E). Thus, Utf1 does not regulate *Arf* transcription or splicing in the nucleus.

We then performed RNA immunoprecipitation (RIP) with Dcp1a antibody and nuclear extracts made from *Utf1*$^{-/\Delta}$ and *Utf1*$^{+/f}$ ESCs. qPCR analyses showed that the binding of Dcp1a to *Arf* mRNA was significantly reduced upon Utf1 loss (Figure 4D). Next, we used flavopiridol to inhibit Pol II (Rahl et al., 2010) and northern blotting to probe for *Arf* mRNA levels. We found that *Utf1*$^{-/\Delta}$ ESCs were less efficient in degrading *Arf* mRNA compared to *Utf1*$^{+/f}$ ESCs (Figure 4E). Thus, Utf1 recruits Dcp1a to bivalent promoters, which facilitates the loading of Dcp1a to mRNAs transcribed from leaky bivalent genes for cytoplasmic degradation.

Utf1 Buffers Bivalent Gene Expression in a Context-Dependent Manner

The above findings reveal that whereas Utf1 limits PRC2 binding to prevent excessive H3K27me3 and bivalent gene silencing, it also recruits Dcp1a to bivalent genes to repress gene expression through mRNA pruning (Figure 5A). These dual functions could allow Utf1 to enforce the poised state of bivalent genes residing

(D) Scatterplot of the correlation between Utf1 and Suz12 binding on promoters. A window of 1 kb up- and downstream of the TSS of all genes was used to calculate the tag enrichment. Spearman's rank correlation coefficient, r = 0.71.

(E) Two AG-rich sequence motifs that are enriched in Utf1-bound loci are highly similar to the predicted Jarid2-bound AG-rich motif.

(F and G) Down- (F) or upregulated (G) genes in *Utf1*$^{-/\Delta}$ ESCs compared to *Utf1*$^{+/f}$ ESCs as shown by RNA-Seq.

See also Figure S1 and Tables S1, S2, S3, and S4.

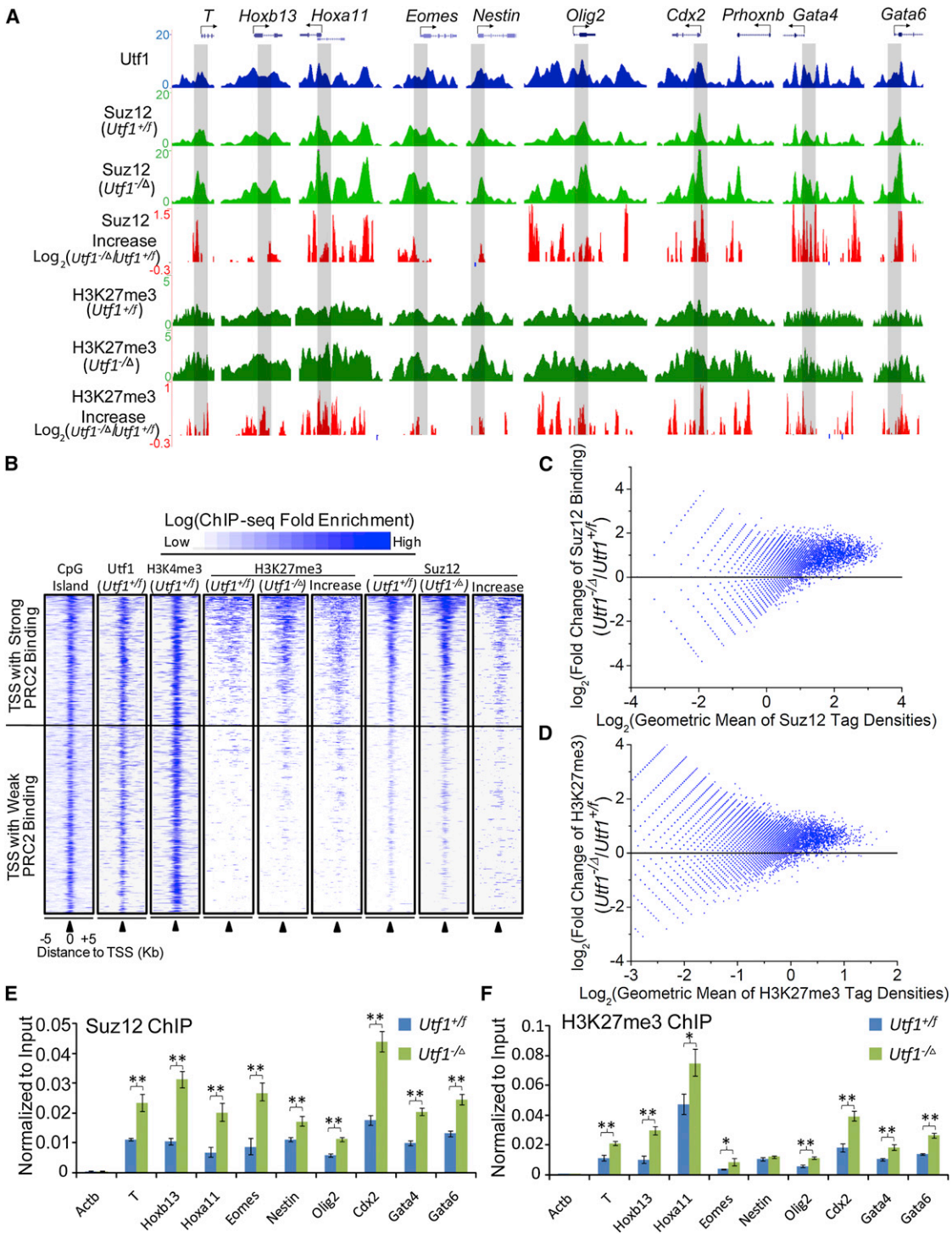

Figure 2. Utf1 Binds to Bivalent Genes to Limit PRC2 Loading and H3K27me3

(A) Plots of ChIP-seq peaks at ten genomic loci. y axes, RPKM of ChIP-seq of Utf1, Suz12, and H3K27me3. The increase (red peaks) or decrease (blue peaks below zero, hardly visible due to very little decrease) of Suz12 binding and H3K27me3 in response to *Utf1* depletion are plotted as $\log_2$ fold changes on y axes. Genomic regions that are further analyzed by ChIP-qPCR (see E and F below) are shaded gray.

(B) Heatmap of all bivalent genes identified in this study showing the enrichment of CpG, Utf1, H3K4me3, H3K27me3 (in $Utf1^{+/f}$ or $Utf1^{-/\Delta}$ ESCs), and Suz12 (in $Utf1^{+/f}$ or $Utf1^{-/\Delta}$ ESCs) and the relative increase of Suz12 binding or H3K27me3 upon *Utf1* depletion. The heatmap is rank-ordered based on the enrichment of Suz12 in control ESCs. The increase in Suz12 binding and H3K27me3 was calculated by the log-ratio of normalized Suz12 and H3K27me3 tag density between $Utf1^{-/\Delta}$ and $Utf1^{+/f}$ ESCs.

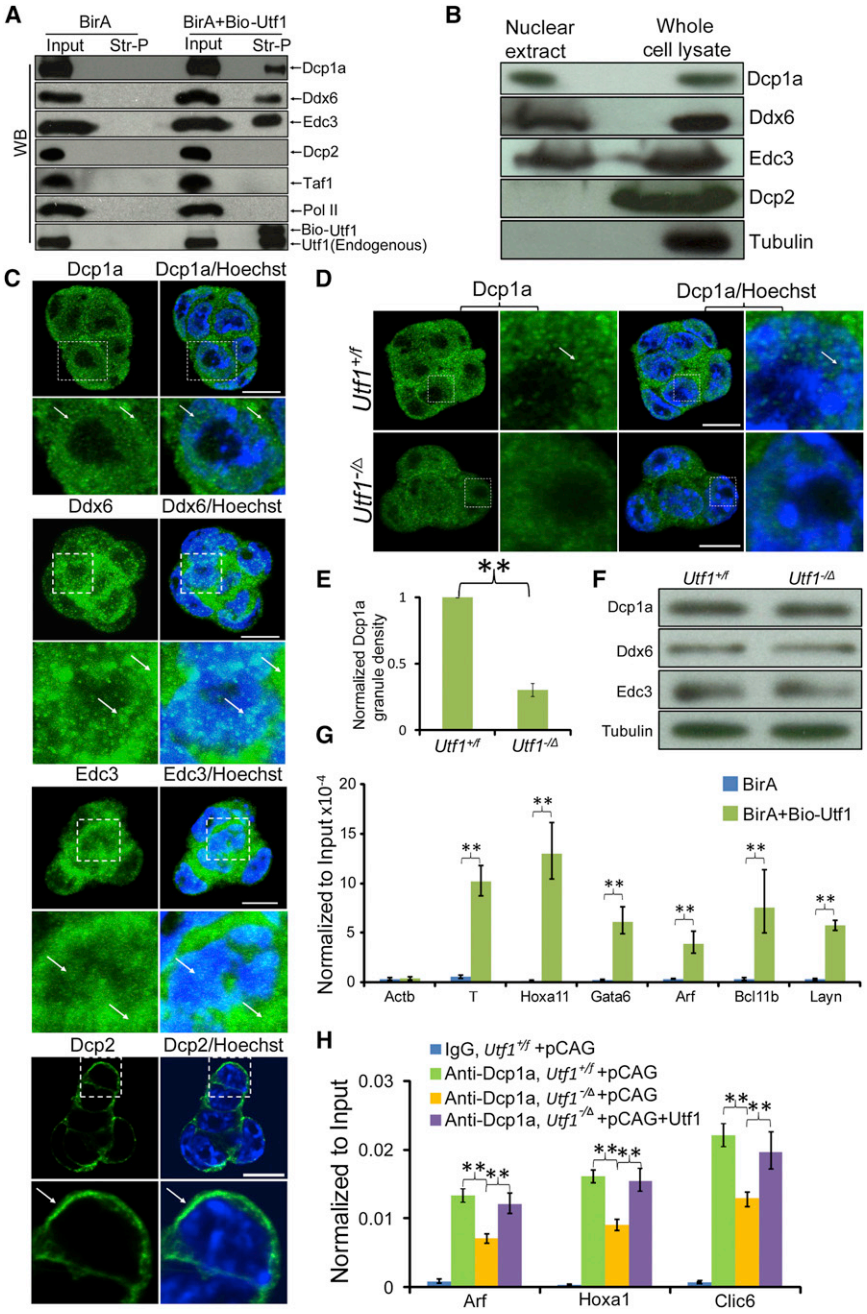

Figure 3. Utf1 Recruits the Noncatalytic Subunit of the mRNA-Decapping Complex to Bivalent Genes

(A) Biotin-Utf1 coprecipitates with Dcp1a, Ddx6, and Edc3, but not Dcp2, Taf1, and Pol II in wild-type ESCs. Streptavidin pulled-down (Str-P) of biotin-Utf1 was probed for the indicated antibodies by western blotting analyses.

(B) Western blotting analyses of nuclear extracts and whole-cell lysates of wild-type ESCs. Tubulin, cytoplasmic protein control.

(C) Immunolocalization of Dcp1a, Ddx6, Edc3, and Dcp2 in wild-type ESCs. Areas in white dashed squares are enlarged at the bottom of each image. Arrows, proteins in the region occupied by chromatin stained by Hoechst (blue) or in the cytoplasm. Scale bar, 10 μm.

(D) A reduction of Dcp1a$^+$ granule numbers and intensity in *Utf1*$^{-/\Delta}$ ESCs compared to *Utf1*$^{+/f}$ ESCs. Areas in dashed squares are enlarged to the right of each image. Arrows, Dcp1a$^+$ granules in the Hoechst-stained chromatin region. Scale bar, 10 μm.

(E) Quantifications of Dcp1a$^+$ granules in *Utf1*$^{-/\Delta}$ and *Utf1*$^{+/f}$ ESCs.

(F) *Utf1*$^{-/\Delta}$ and *Utf1*$^{+/f}$ ESCs express similar amounts of Dcp1a, Ddx6, and Edc3. Loading controls, tubulin.

(G) Dcp1a-ChIP samples were re-ChIPed using streptavidin beads followed by qPCR of bivalent genes in *Utf1*$^{+/f}$ ESCs expressing BirA alone or BirA plus biotin-Utf1.

(H) ChIP-qPCR of Dcp1a binding at bivalent genes in *Utf1*$^{+/f}$ or *Utf1*$^{-/\Delta}$ ESCs transfected with control vector pCAG or pCAG-Utf1. Restoration of Utf1 expression (see Figure S2E for Utf1 western) rescued Dcp1a binding in *Utf1*$^{-/\Delta}$ ESCs. The IgG antibody was used as controls.

Error bars, SD of triplicates. Student's t test, **p < 0.01. See also Table S5.

in different DNA and chromatin contexts. We analyzed whether the downregulated bivalent genes upon Utf1 loss have different chromatin and DNA contents compared to the upregulated genes. We found that the upregulated genes have lower CpG island densities and weaker PRC2 binding than the downregulated genes (Figures 5B–5D). Importantly the upregulated genes exhibited less PRC2 binding increase upon Utf1 loss than the downregulated genes (Figures 5B, 5E, and 5F). These analyses suggest the following modes of context-dependent bivalent gene regulation by Utf1. For bivalent genes that have low CpG island densities and PRC2 binding or are bound by strong transcriptional activators (see *Arf* as an example), Utf1 would repress gene expression because the loss of mRNA pruning due to Utf1 loss could not be

(C and D) MA-plots of Suz12 binding (C) and H3K27me3 (D) on all promoters in *Utf1*$^{-/\Delta}$ and *Utf1*$^{+/f}$ ESCs. Two biological repeats give similar results. One data set is shown here. Log_2 of the normalized Suz12 or H3K27me3 tag densities within −1 kb to 1 kb up- and downstream of the TSS in *Utf1*$^{+/f}$ *and Utf1*$^{-/\Delta}$ ESCs were averaged and plotted on the x axes. The y axes are the log-ratio of Suz12 or H3K27me3 tag densities between *Utf1*$^{-/\Delta}$ and *Utf1*$^{+/f}$ ESCs.

(E and F) ChIP-qPCR of Suz12 binding (E) and H3K27me3 (F) at selected genes in *Utf1*$^{+/f}$ and *Utf1*$^{-/\Delta}$ ESCs. The *Actb* locus was used as controls. Error bars, standard deviations (SD) of triplicates. Student's t test, *p < 0.05, **p < 0.01.

See also Figure S2 and Table S2.

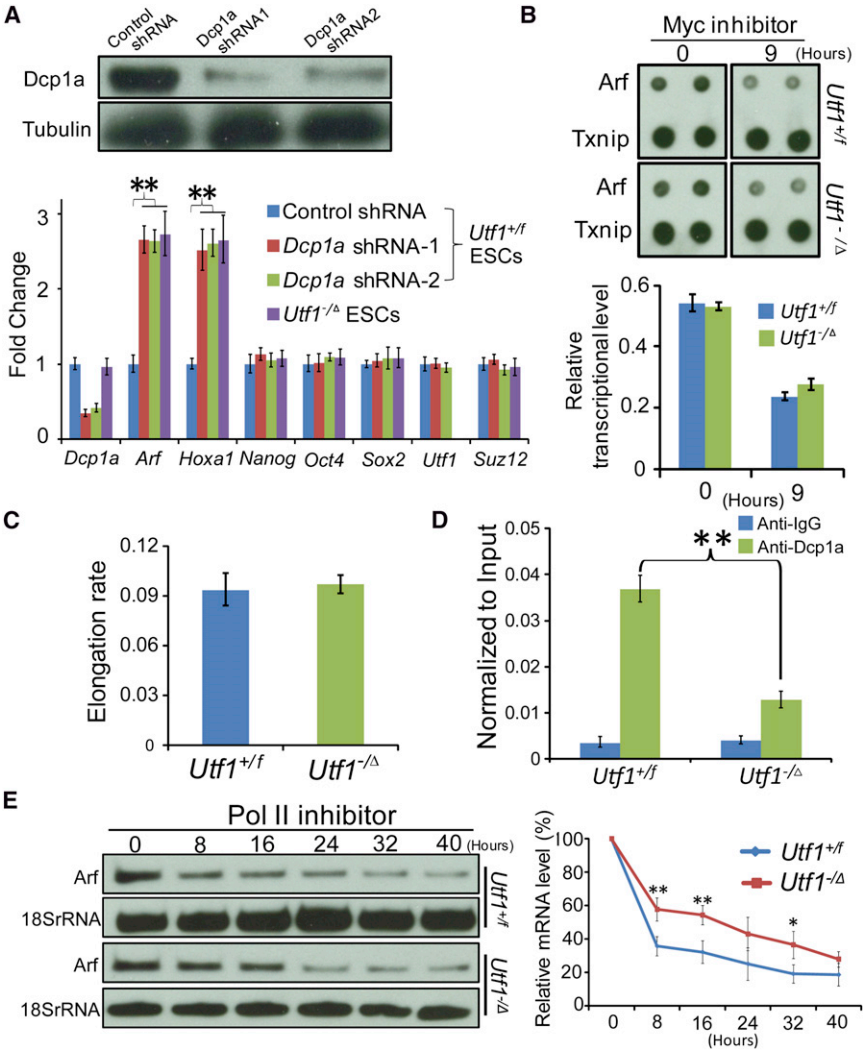

Figure 4. Utf1 Promotes the Binding of Dcp1a to mRNAs in the Nucleus for Cytoplasmic Degradation

(A) Western blotting analyses of *Utf1*$^{+/f}$ ESCs treated with control or Dcp1a shRNAs. Loading controls, tubulin. The graph shows RT-qPCR of mRNA levels of indicated genes in the control, *Dcp1a*-shRNA-treated *Utf1*$^{+/f}$ ESCs, or untreated *Utf1*$^{-/\Delta}$ ESCs.

(B) ^{32}P-labeled transcripts from Myc inhibitor-treated (9 hr) and untreated (0 hr) ESCs were hybridized to nitrocellulose filters spotted with *Arf* or *Txnip* cDNAs (two spots per gene). *Txnip* is a nonbivalent gene not regulated by Myc (Rahl et al., 2010). The graph shows quantifications of the nuclear run-on assays. The *Arf* hybridization intensity was normalized to the *Txnip* intensity and plotted.

(C) The elongation rate of Pol II at the *Arf* locus in *Utf1*$^{+/f}$ and *Utf1*$^{-/\Delta}$ ESCs measured by the ratio of ChIP-qPCR of Pol II at exon 2 and at the promoter region of *Arf*.

(D) RIP using Dcp1a antibody in *Utf1*$^{+/f}$ and *Utf1*$^{-/\Delta}$ ESCs followed by RT-qPCR analyses.

(E) *Arf* mRNA decay was determined by northern blotting in *Utf1*$^{+/f}$ and *Utf1*$^{-/\Delta}$ ESCs after transcriptional arrest by flavopiridol. Loading control, 18S rRNA. The *Arf* mRNA band intensity normalized to the 18S RNA band intensity at time 0 was defined as 100%. The normalized *Arf* band intensities at other time points were calculated relative to the time 0.

Error bars, SD. Student t test, *p < 0.05, **p < 0.01. See also Figure S3 and Table S6.

compensated for by the limited gaining of PRC2 (Figure 5G). By contrast, for bivalent genes with high CpG island densities and strong PRC2 binding, Utf1 would prevent gene silencing because upon Utf1 deletion, the loss of mRNA pruning would be overcompensated by the excessive gaining of PRC2 (Figure 5H).

We have shown that Utf1 uses Dcp1a to repress genes such as *Hoxa1* and *Arf* (Figure 4A; Tables S4 and S6), which is consistent with the model in Figure 5G. Because our studies suggest that Utf1 recruits Dcp1a to bivalent genes, we analyzed how the loss of Dcp1a could affect the expression of genes that use Utf1 to limit their repression (see Figure 5H). As these genes exhibit stronger PRC2 binding than the other bivalent genes, Dcp1a loss may or may not affect their expression depending on the strength of PRC2-mediated gene repression. Indeed, we found that 56% (164/295) of them did not show gene expression change upon Dcp1a loss, suggesting that PRC2 was sufficient for their repression. We did find that 41% (121 out of 295) of bivalent genes in this group underwent upregulation upon Dcp1a reduction, suggesting that they depended on both PRC2 and Dcp1a for repression. The above analyses predict that whereas Utf1 functions in ESCs to maintain rapid cell proliferation by repressing *Arf* expression through Dcp1a-mediated mRNA pruning, both the mRNA pruning and the epigenetic functions of Utf1 also ensure proper gene expression during ESC differentiation.

Utf1 Antagonizes Myc-Mediated Activation of *Arf* to Ensure Rapid Proliferation of ESCs

We found that Arf protein levels in *Utf1*$^{-/\Delta}$ ESCs were reduced either by reintroducing Utf1 or by inhibiting c-Myc (Figures S4A and S4B), demonstrating that Utf1 functions downstream of Oct4 and Sox2 to antagonize the Myc-Arf feedback loop. Although both *Utf1*$^{+/f}$ and *Utf1*$^{-/\Delta}$ ESCs expressed similar amounts of pluripotency proteins (Figures S1H and S4C), the *Utf1*$^{-/\Delta}$ ESCs had higher amounts of Arf in the nucleus, flatter colony morphologies, and slower proliferation rates compared to *Utf1*$^{+/f}$ ESCs (Figures 6A–6C). Consistently, teratomas formed by *Utf1*$^{-/\Delta}$ ESCs were much smaller than those of *Utf1*$^{+/f}$ ESCs (Figure S4D). Inhibiting *Arf* expression using shRNA (Figure 6D) resulted in a significant increase of cell proliferation rates in *Utf1*$^{-/\Delta}$ ESCs without affecting the rate of *Utf1*$^{+/f}$ ESCs (Figure 6C). Thus upregulation of *Arf* upon Utf1 deletion was

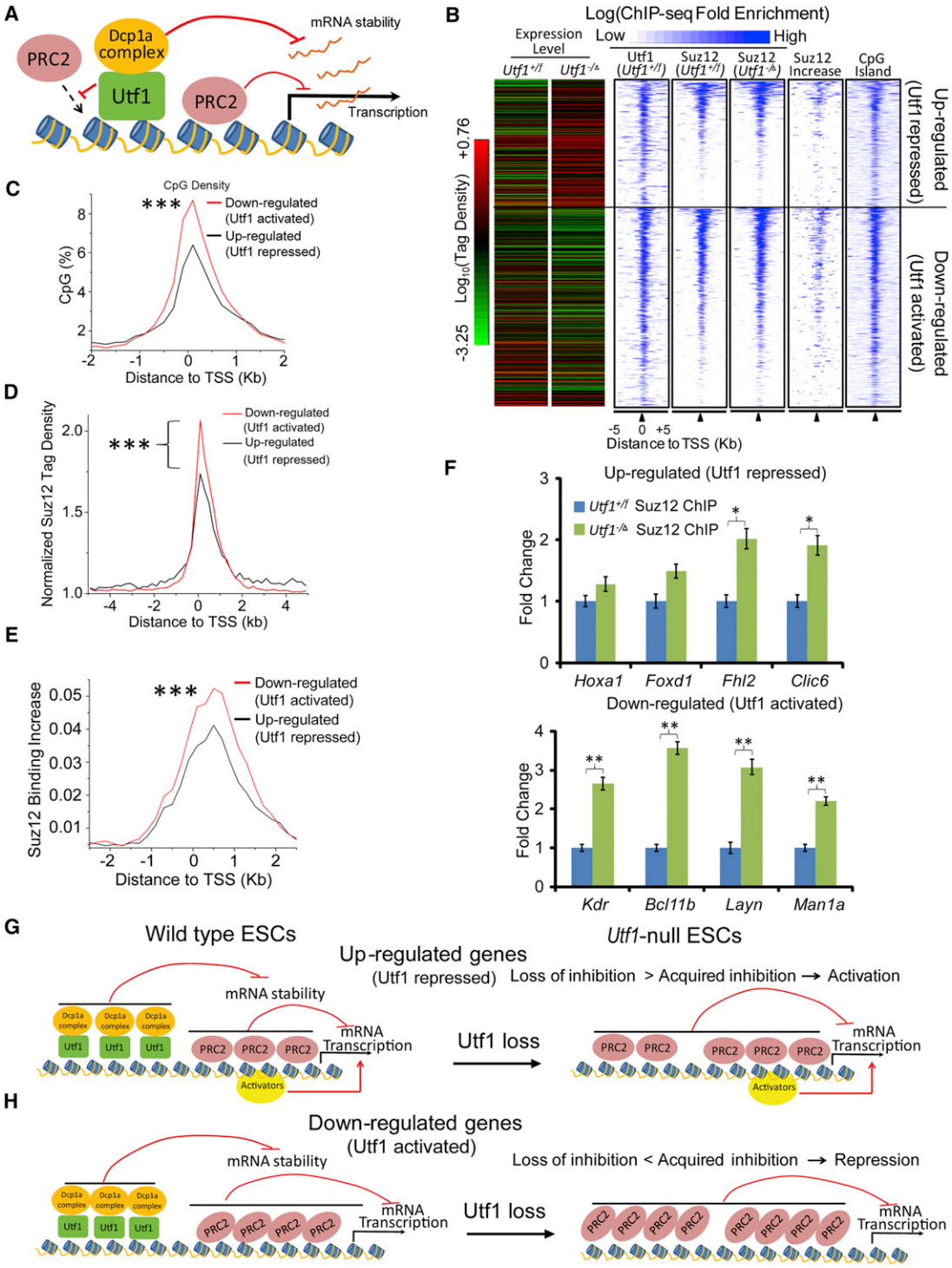

Figure 5. Utf1 Regulates Bivalent Genes in a Context-Dependent Manner

(A) Illustration of the dual functions of Utf1 on bivalent genes.

(B) The Heatmap of gene expression in *Utf1*$^{+/f}$ and *Utf1*$^{-/\Delta}$ ESCs (red and green, high and low expression, respectively). The profiles of Utf1 binding, Suz12 binding (in *Utf1*$^{+/f}$ or *Utf1*$^{-/\Delta}$ ESCs), Suz12-binding increase (in *Utf1*$^{-/\Delta}$ ESCs), and CpG densities of these genes are also shown as heatmaps that are rank-ordered based on Suz12-binding strength in *Utf1*$^{+/f}$ ESCs. The gene expression heatmap shows that Utf1 buffers gene expression.

(C) Quantifications of CpG densities at the TSS of Utf1-activated (downregulated upon Utf1 loss, red) and Utf1-repressed (upregulated upon Utf1 loss, black) genes. The CpG density is determined by the percentage of CpG dinucleotides in 200 bp windows and plotted within 2 kb up- and downstream of the TSS (Wilcoxon two-sample test, ***$p < 0.001$).

responsible for the slow ESC proliferation. Reduction of Arf in the *Utf1*$^{-/\Delta}$ ESCs also partially reversed the flattened colony morphology (Figure 6E). The incomplete reversal of the proliferation rate and colony morphology of Arf-RNAi-treated *Utf1*$^{-/\Delta}$ ESCs is consistent with our finding that *Arf* is only one of the targets of Utf1 in ESCs.

The p53-p21^{Cip1} pathway represents the best characterized mechanism by which Arf inhibits cell-cycle progression (Pomerantz et al., 1998; Zindy et al., 1998). The binding of Arf to Mdm2, an ubiquitin ligase for p53, leads to Mdm2 inhibition and p53 accumulation. The increase in p53 could lead to transcriptional upregulation of the cell-cycle inhibitor p21^{Cip1}, which would slow down cell-cycle progression. We found that although Arf upregulation led to a small increase of p53 levels in *Utf1*$^{-/\Delta}$ ESCs, p21^{Cip1} was not upregulated (Figures 6A and 6F). Similarly, although transient overexpression of Arf in *Utf1*$^{+/f}$ ESCs strongly increased p53 protein levels and inhibited cell-cycle progression, p21^{Cip1} expression was unchanged (Figures 6G, 6H, and S4E–S4G). Thus, Utf1 functions downstream of Oct4 and Sox2 to ensure rapid cell-cycle progression by blocking the Myc-Arf feedback loop independent of p53-p21^{Cip1} in ESCs.

Utf1 Coordinates Gene Expression during ESC Differentiation

We next studied whether Utf1 is required for proper bivalent gene expression during ESC differentiation by investigating their expression in either ectoderm (such as *Olig2* and *Nestin*) or mesoendoderm (such as *T* and *Hoxa1*). *Olig2*, *Nestin*, and *T* exhibited a significant increase in PRC2 binding and H3K27me3 upon Utf1 loss (see Figures 2A, 2E, and 2F). *Hoxa1* was upregulated either upon Utf1 loss or upon *Dcp1* RNAi (see Figure 4A; Tables S4 and S6), suggesting that the mRNA-pruning mechanism is important for controlling *Hoxa1* expression. We used embryoid body (EB)-based differentiation and qPCR assays and found that whereas *Olig2*, *Nestin*, and *T* were insufficiently upregulated, *Hoxa1* underwent excessive upregulation during EB differentiation of *Utf1*$^{-/\Delta}$ ESCs compared to *Utf1*$^{+/f}$ ESCs (Figure 7A). Additional qPCR analyses of other bivalent genes expressed in mesoendoderm (*Gata6* and *Bmp4*) or during epithelium-to-mesenchyme transition (EMT, *Cd44*) revealed a similar misregulation (Figure 7A).

To study how Utf1 could affect gene expression and lineage-specific differentiation, we used ZHBTc4 ESCs, in which the endogenous *Oct4* was replaced by a tetracyclin (Tc)-regulated *Oct4* transgene (Niwa et al., 2000). ZHBTc4 ESCs undergo efficient trophectoderm (TE) differentiation in the presence of Tc. The upregulation of the transcription factor Cdx2 during TE development in vivo is required for establishing progenitor cells for TE (Niwa et al., 2005). In vitro, Cdx2 is not essential for terminal TE differentiation from ESCs, but it is required for the establishment and maintenance of trophoblast stem (TS) cells that undergo further terminal differentiation into TE lineages (Niwa et al., 2005). Upon Utf1 loss, the *Cdx2* gene locus exhibited a significant increase of PRC2 binding and H3K27me3 in ESCs (see Figures 2E and 2F), predicting that Cdx2 would not be sufficiently upregulated when Utf1-knockdown ESCs were induced to differentiate toward TE. Consistently, we found that addition of Tc to ZHBTc4 ESCs treated with Utf1 shRNA led to insufficient *Cdx2* upregulation (Figure 7B) and rapid terminal differentiation as judged by the formation of large and flat terminally differentiated TE cells, whereas the control shRNA-treated ZHBTc4 ESCs formed many colonies made of small TS-like cells upon Tc and FGF4 treatment (Figures 7C and 7D). Thus although Utf1 deletion did not block the differentiation of ESCs to TE, it disrupted proper coupling between proliferation and differentiation.

DISCUSSION

Using a combination of genomic, cell biology, proteomic, and gene-targeting approaches, we uncover a mechanism that integrates the pluripotency core factors with the PRC2-based epigenetic network and Myc regulation. This mechanism has three features. First, the core factors use their direct downstream target Utf1, which itself is a chromatin-associated protein enriched at bivalent genes, to control both self-renewal and proper differentiation. Second, the core factors use Utf1 to block the Myc-Arf feedback loop to ensure rapid cell proliferation. An important function of the core factors in reprogramming is to block the expression of *Arf* and *Ink4a*. However, because none of the core factors directly binds to *Cdkn2a*, the mechanism of repression has remained unknown. Our findings provide the critical missing molecular link that allows ESCs to evade the Myc-Arf feedback control. Third, Utf1 integrates all three networks using both epigenetic regulation and a previously unappreciated mRNA pruning mechanism (Figure 7E).

The differential PRC2 loading on bivalent genes invites the question of how the genes with high or low PRC2 binding are neither oversilenced nor derepressed, respectively. We show that Utf1 and PRC2 bind to bivalent genes competitively because they recognize similar AG-rich motifs, which allows Utf1 to prevent excessive loading of PRC2 in ESCs. How Utf1 or PRC2 binds to chromatin is unclear. Interestingly, H2AZ is one of the Utf1-interacting proteins identified in our MS analyses. Because H2AZ binds to PRC2-bound genes (Creyghton et al., 2008), studying the functional relationship between H2AZ and Utf1 may help to understand how Utf1 is loaded onto bivalent genes. We note that our biotin-Utf1-ChIP-seq data are not

(D) Quantifications of Suz12 binding at the TSS of Utf1-activated (downregulated upon Utf1 loss) and Utf1-repressed (upregulated upon Utf1 loss) bivalent genes within 5 kb up- and downstream of the TSS (Wilcoxon two-sample test, ***$p < 0.001$).

(E) Quantifications of Suz12-binding increase upon Utf1 loss at the TSS of Utf1-activated (downregulated upon Utf1 loss) and Utf1-repressed (upregulated upon Utf1 loss) genes within 5 kb up- and downstream of the TSS (Wilcoxon two-sample test, ***$p < 0.001$).

(F) Suz12-ChIP-qPCR of Utf1-repressed (upregulated upon Utf1 loss, top plot) and Utf1-activated (downregulated upon Utf1 loss, bottom plot) genes found by RNA-Seq. Error bars, SD of triplicates. Student's t test, *$p < 0.05$, **$p < 0.01$.

(G and H) Illustrations for the context-dependent buffering of bivalent genes by Utf1. The dual functions of Utf1 in repressing (by recruiting Dcp1a) and activating (by limiting PRC2 binding) genes are indicated. Upon Utf1 loss, bivalent genes may undergo upregulation (G) or downregulation (H), depending on the degree of PRC2-binding increase, Dcp1a loss, and other transcription activators (such as Myc in the *Arf* locus) on these genes.

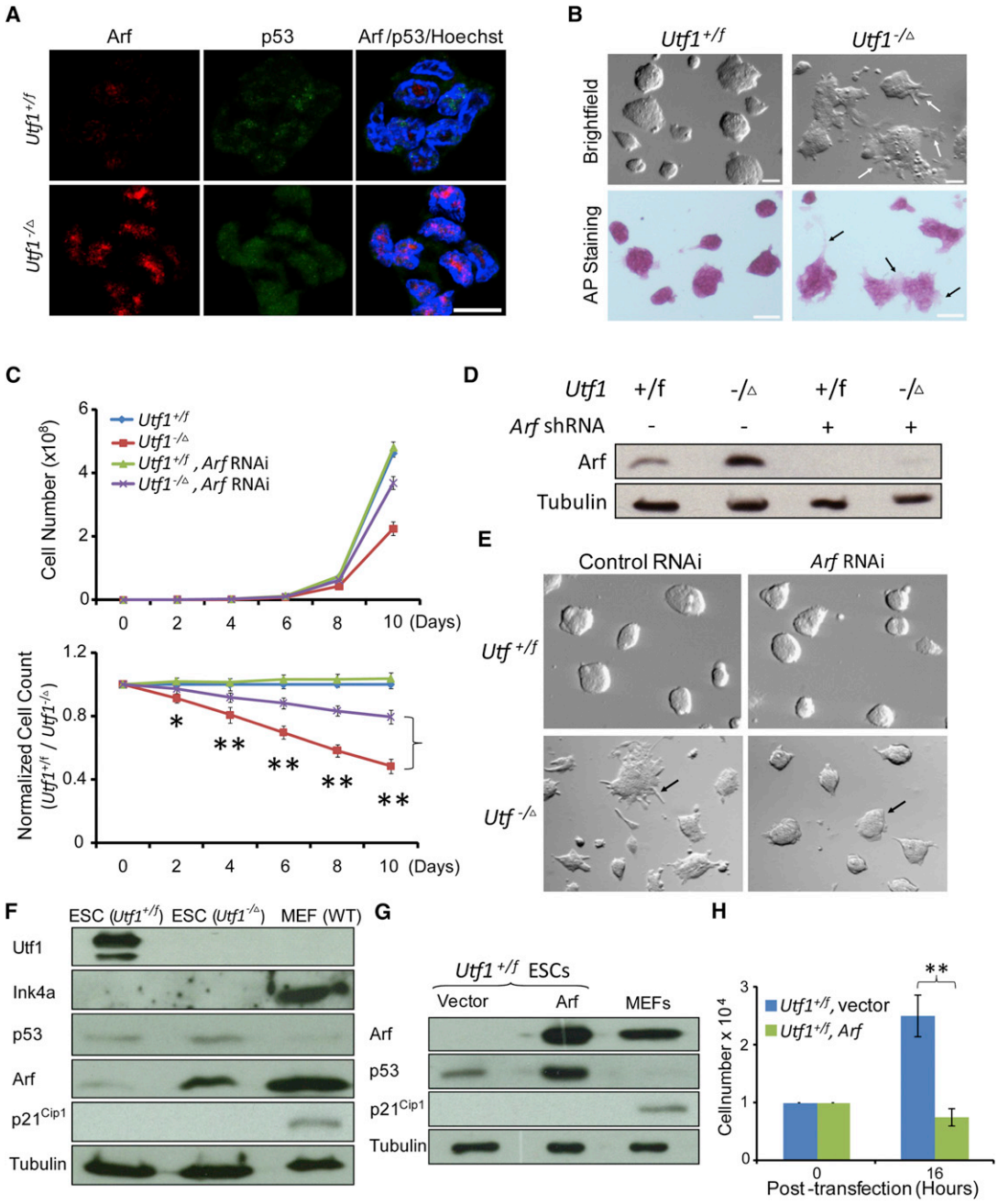

Figure 6. Utf1 Promotes Proliferation of ESCs through Inhibiting *Arf* Expression

(A) Representative immunofluorescence images of Arf and p53 proteins in *Utf1*$^{+/f}$ and *Utf1*$^{-/\Delta}$ ESCs. Scale bar, 10 μm.

(B) *Utf1*$^{-/\Delta}$ ESCs appear flatter than *Utf1*$^{+/f}$ ESCs. Both types of ESCs are positive for the pluripotency marker alkaline phosphatase (AP). Arrows indicate spontaneous differentiation at the edges of *Utf1*$^{-/\Delta}$ ESC colonies.

(C) *Utf1*$^{-/\Delta}$ ESCs proliferated slower than *Utf1*$^{+/f}$ ESCs. Reduction of *Arf* by shRNA significantly enhanced *Utf1*$^{-/\Delta}$ ESC proliferation without affecting *Utf1*$^{+/f}$ ESCs. The top graph plots cell numbers. The bottom graph plots the percentages of *Utf1*$^{-/\Delta}$ ESCs normalized against *Utf1*$^{+/f}$ ESCs. Error bars, SD of triplicates.

(D) shRNA reduction of Arf expression in both *Utf1*$^{+/f}$ and *Utf1*$^{-/\Delta}$ ESCs. Loading controls, tubulin.

(E) Inhibiting Arf expression reverted the flattened ESC colonies back to the morphology similar to the *Utf1*$^{+/f}$ ESCs.

(F) Western blotting analyses of Ink4a, p53, Arf, and p21^{Cip1} in *Utf1*$^{+/f}$ and *Utf1*$^{-/\Delta}$ ESCs. Controls, wild-type MEFs. Loading controls, tubulin.

(G) Western blotting analyses of Arf overexpression 16 hr post-transfection of ESCs, which had increased p53 but not p21^{Cip1}. Controls, wild-type MEFs. Loading controls, tubulin.

(H) Overexpression of Arf in *Utf1*$^{+/f}$ ESCs resulted in a significant reduction of cell proliferation 16 hr post-transfection.

Error bars, SD of triplicates. Student's t test in (C) and (H), *p < 0.05, **p < 0.01. See also Figure S4.

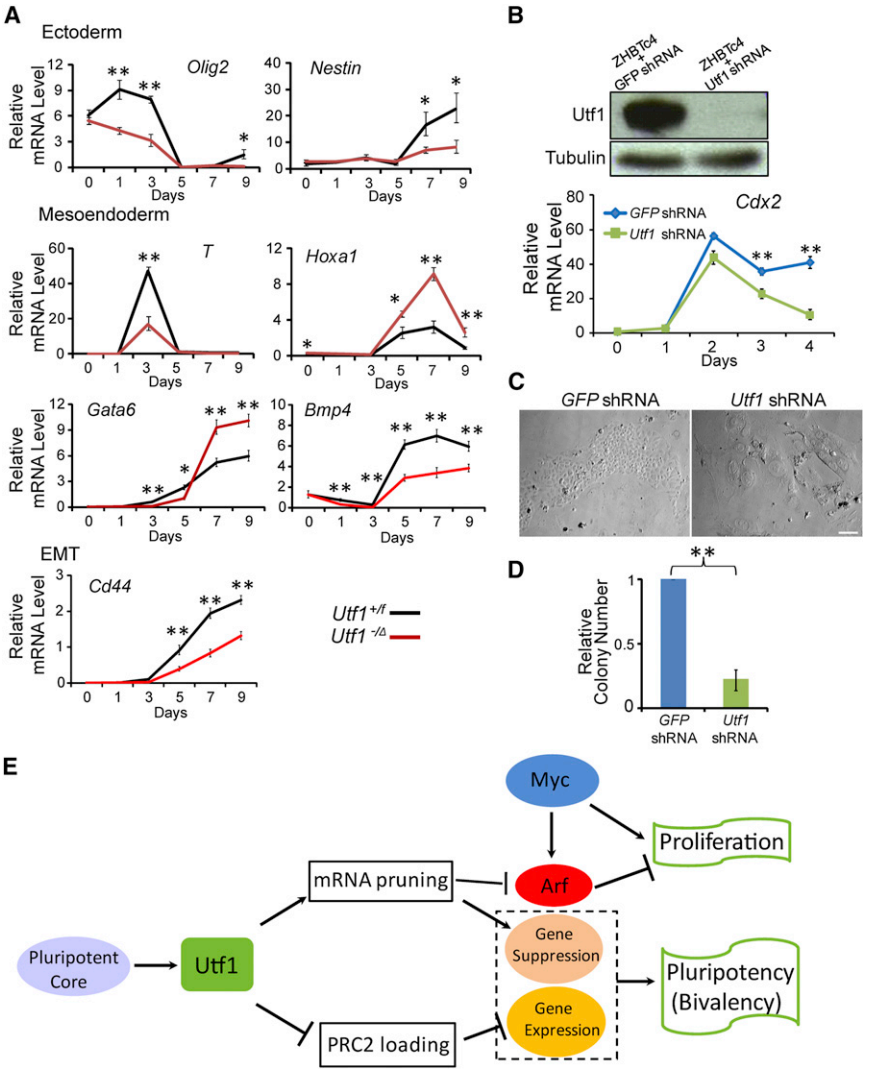

Figure 7. Utf1 Regulates Proper Differentiation by Coordinating Gene Expression in ESCs

(A) Time-course RT-qPCR analyses of developmental regulators normalized to *GAPDH* during EB differentiation in *Utf1*$^{+/f}$ and *Utf1*$^{-/\Delta}$ ESCs. Error bars, SD of triplicates.

(B) Western blotting analyses showed that Utf1 shRNA reduced Utf1 protein levels in ZHBTc4 ESCs compared to control GFP shRNA. Time-course RT-qPCR analyses of Cdx2 mRNA during TE differentiation was normalized to GAPDH and plotted. Error bars, SD of triplicates.

(C) Images of TS cells or terminally differentiated TE cells derived from GFP-shRNA- or Utf1-shRNA-treated ZHBTc4 ESCs, respectively.

(D) Quantifications of TS colonies in GFP-shRNA- and Utf1-shRNA-treated cells 6 days after differentiation. Error bars, SD of triplicates. Student's t test in (A), (B), and (D), *p < 0.05, **p < 0.01.

(E) A model. Utf1 couples three regulatory networks to ensure proper proliferation and differentiation of ESCs.

consistent with the native-ChIP-on-chip studies of Utf1 (Kooistra et al., 2010). The native-ChIP procedure involved micrococcal nuclease treatments and additional manipulations for extended time without cross-linking, which could fail to capture the true Utf1-chromatin interactions.

Our data suggest that by binding to Dcp1a, Ddx6, and Edc3, Utf1 ensures that mRNAs transcribed from leaky bivalent genes are tagged for efficient degradation by the mRNA-decapping pathway. The mRNA-decapping complex, which has only been studied in fungi and in animal somatic cells, promotes the formation of mRNP-processing granules that either prevent mRNA translation or facilitate mRNA degradation (Ling et al., 2011; Tritschler et al., 2009). Our findings have revealed a previously unappreciated function of the mRNA-decapping complex in silencing bivalent genes that are insufficiently repressed in ESCs. As a component of the mRNA-decapping complex, Dcp1a could promote both the binding of mRNA to Dcp2 and the decapping activity of Dcp2 to facilitate mRNA degradation in the ESC cytoplasm. Dcp1a could also promote mRNA turnover through microRNAs in ESCs (Ling et al., 2011; Tritschler et al., 2009).

The promoter-mediated mRNA-pruning function defined here may represent an important mechanism to ensure ESC maintenance. Interestingly, two recent studies in yeast have shown that promoters can dictate mRNA stability by recruiting proteins that target mRNAs for cytoplasmic degradation (Bregman et al., 2011; Trcek et al., 2011). Although the proteins involved in the promoter-based marking of mRNAs are different in the yeast studies and in our study, our findings strongly suggest that the promoter-mediated control of mRNA stability is evolutionarily conserved.

Dcp1a, Edc3, and Dcp2 have been found in the same complex with TTF2 (a transcriptional terminator), and they have been implicated in facilitating mRNA decapping during transcription, which inhibits transcriptional elongation in both HeLa and HEK293 cells (Brannan et al., 2012). Dcp1a has also been shown to bind to *Smad4* to activate target genes upon TGF-β signaling in different somatic tissue culture cells (Bai et al., 2002). However, unlike these previous reports, our studies show that the Utf1-mediated recruiting of Dcp1a to bivalent promoters in ESCs does not affect transcription, but it tags mRNA for post-transcriptional degradation in the cytoplasm.

The buffer function of Utf1 that we uncovered suggests that although removing Utf1 may not block lineage specification and differentiation, it would lead to inappropriate coupling of developmental processes and cell proliferation. Our preliminary analyses of *Utf1* null mice are consistent with this prediction (data not shown). *Utf1* is localized on the human chromosome

10q26.3, near the telomere. Humans bearing a heterozygous deletion of 10q26.3 exhibit a number of developmental abnormalities and retarded growth (Kehrer-Sawatzki et al., 2005). Our findings suggest that Utf1 heterozygosity could be responsible for some of the abnormalities in the 10q26.3 patients.

EXPERIMENTAL PROCEDURES

Karyotype Analyses of ESCs and Differentiation Assays

These procedures were performed as described previously (Creyghton et al., 2008; Kim et al., 2011; Niwa et al., 2000; Vong et al., 2010) and in the Extended Experimental Procedures.

Utf1 Pull-Down, Western Blotting, and MS

V6.5 ESCs (1–2 × 10^9) were used to make nuclear extracts. After dialyzing the extract into a low-salt buffer, a Utf1 polyclonal antibody was used to immunoprecipitate Utf1 and its associated proteins. ESCs (1 × 10^7) expressing either BirA alone or both BirA and biotin-Utf1 were used for streptavidin pull-down of biotin-Utf1. See the Extended Experimental Procedures for more details.

Cell Proliferation Assay

To perform long-term proliferation assays, 10^5 *Utf1*$^{+/f}$ or *Utf1*$^{-/\Delta}$ ESCs were seeded into a single well of a 6-well plate. The cells in each well were counted and replated every 2 days. The *Utf1*$^{+/f}$ ESCs were diluted so that 10^5 cells were plated in a fresh well. The same dilution factor was used to dilute the *Utf1*$^{-/\Delta}$ ESCs into a new well. As *Utf1*$^{-/\Delta}$ ESCs grew slower, fewer *Utf1*$^{-/\Delta}$ ESCs than *Utf1*$^{+/f}$ ESCs were plated each time. The cumulative total cell number was calculated at each time point and used to plot the growth curve. The cell number ratios of control-treated *Utf1*$^{+/f}$ ESCs to other experimental groups of ESCs at each time point were used to plot the normalized growth curve.

RNAi-Mediated Gene Silencing in ESCs

shRNAs were purchased from Open Biosystems to knock down the mRNA of *Arf* (TRCN0000077816), *Utf1* (TRCN0000081708), *Jarid2* (TRCN0000096642), and *Dcp1a* (shRNA-1:TRCN0000096664, shRNA-2:TRCN0000096665). See the Extended Experimental Procedures for more details.

ChIP Assays

The ChIP-qPCR and ChIP-seq assays were performed according to the conditions suggested by the manufacturers of ChIP-grade antibodies to H3K27me3 (Abcam, ab6002), H3K4me3 (Cell Signaling, 9751), Suz12 (Cell Signaling, 3737), and Dcp1a (WH0055802) with modifications (Jia et al., 2007). Biotin-ChIP was performed as previously described (Kim et al., 2009). See the Extended Experimental Procedures for detailed procedures, data analyses, and primer information.

Sequential ChIP Assay

Sample treatment and immunoprecipitations were performed using standard ChIP assay procedures. After the first immunoprecipitation, chromatin was eluted in a solution of 30 mM DTT, 500 mM NaCl, and 0.1% SDS at 37°C as described (Bernstein et al., 2006). Eluted chromatin was diluted 50-fold and subjected to biotin-Utf1-ChIP as described in the Extended Experimental Procedures.

Whole-Transcriptome Shotgun Sequencing (RNA-Seq)

Total RNA was isolated from 10^7 ESCs with the RNeasy Plus kit (QIAGEN). The poly(A)-containing mRNAs were purified, and libraries were built following Illumina TruSeq RNA protocols. Libraries were sequenced using an Illumina HiSeq 2000. 100 bp-long reads from both ends were obtained. See the Extended Experimental Procedures for more details.

Cell-Cycle Analyses

ESCs were trypsinized and resuspended in 0.5 ml PBS followed by the addition of 0.5 ml of 100% ice-cold ethanol in a drop-wise manner while vortexing. After incubation for 20 min on ice, cells were harvested and washed by PBS containing 1% FBS. Next, cells were incubated in PBS with 25 μg/ml RNase A at 37°C for 30 min to digest RNA. Finally, cells were stained by 50 μg/ml propidium iodide for 10 min at room temperature and analyzed by flow-cytometry using BD-FACS Calibur.

Determining mRNA Degradation Profile

ESCs were cultured as described above. Cells were treated with 1 μM flavopiridol (Sigma cat #F3055) and harvested at different time points. mRNA extraction and northern blotting analyses are described in the Extended Experimental Procedures.

Nuclear Run-on and S1 Nuclease Protection Assays

ESCs were cultured as described above. Cells were treated with 50 μM c-Myc inhibitor 10058-F4 (Sigma cat #F3680) and harvested after 9 hr. Whole ESC and nuclear extract were harvested as described above. S1 nuclease protection assay was performed as previously described (Mendrysa and Perry, 2000). For details, see the Extended Experimental Procedures.

EMSA

A 74 bp mouse genomic sequence containing the predicted Utf1- and Jarid2-binding motifs was labeled by γ-^{32}P-ATP and incubated with ESC nuclear extract. The unlabeled competitor probe was used at 100-fold excess to the labeled probe. Super-shift was performed by incubating the reactions with antibodies against Utf1 or Suz12. For details, see the Extended Experimental Procedures.

ACCESSION NUMBERS

All ChIP-seq and RNA-seq data sets have been deposited into the GEO repository (GSE39513).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, four figures, and seven tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2012.09.023.

ACKNOWLEDGMENTS

We thank Allison Pinder for deep-seq; Nick Ingolia for advice on data analyses; and Chen-ming Fan, Nick Ingolia, Xin Chen, Haiqing Zhao, Dong Yang, Ma Wan, Max Guo, and members of the Zheng lab for comments. This work was supported by NIH (GM56312 to Y.Z.; MH067880 and P41 RR011823 to J.R.Y.) and by NHLBI intramural research (G.H., K.C., C.L., and K.Z.). This work was also supported by R01 GM056312 (Y.Z., J.Z., A.Z.), HHMI (J.J., Y.Z.), P41 RR011823 and R01 HL079442-08 (J.Y.), and Cystic Fibrosis Foundation Therapeutics, Inc. BALCH05X5 (B.L.).

Received: December 1, 2011
Revised: May 24, 2012
Accepted: August 17, 2012
Published: October 25, 2012

REFERENCES

Ang, Y.S., Tsai, S.Y., Lee, D.F., Monk, J., Su, J., Ratnakumar, K., Ding, J., Ge, Y., Darr, H., Chang, B., et al. (2011). Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. Cell *145*, 183–197.

Bai, R.Y., Koester, C., Ouyang, T., Hahn, S.A., Hammerschmidt, M., Peschel, C., and Duyster, J. (2002). SMIF, a Smad4-interacting protein that functions as a co-activator in TGFbeta signalling. Nat. Cell Biol. *4*, 181–190.

Bailey, T.L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proceedings/International Conference on Intelligent Systems for Molecular Biology; ISMB *2*, 28–36.

Banito, A., Rashid, S.T., Acosta, J.C., Li, S., Pereira, C.F., Geti, I., Pinho, S., Silva, J.C., Azuara, V., Walsh, M., et al. (2009). Senescence impairs successful reprogramming to pluripotent stem cells. Genes Dev. *23*, 2134–2139.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315–326.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., et al. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell *122*, 947–956.

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.

Brannan, K., Kim, H., Erickson, B., Glover-Cutter, K., Kim, S., Fong, N., Kiemele, L., Hansen, K., Davis, R., Lykke-Andersen, J., and Bentley, D.L. (2012). mRNA decapping factors and the exonuclease Xrn2 function in widespread premature termination of RNA polymerase II transcription. Mol. Cell *46*, 311–324.

Bregman, A., Avraham-Kelbert, M., Barkai, O., Duek, L., Guterman, A., and Choder, M. (2011). Promoter elements regulate cytoplasmic mRNA decay. Cell *147*, 1473–1483.

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.

Cleveland, J.L., and Sherr, C.J. (2004). Antagonism of Myc functions by Arf. Cancer Cell *6*, 309–311.

Creyghton, M.P., Markoulaki, S., Levine, S.S., Hanna, J., Lodato, M.A., Sha, K., Young, R.A., Jaenisch, R., and Boyer, L.A. (2008). H2AZ is enriched at polycomb complex target genes in ES cells and is necessary for lineage commitment. Cell *135*, 649–661.

Efroni, S., Duttagupta, R., Cheng, J., Dehghani, H., Hoeppner, D.J., Dash, C., Bazett-Jones, D.P., Le Grice, S., McKay, R.D., Buetow, K.H., et al. (2008). Global transcription in pluripotent embryonic stem cells. Cell Stem Cell *2*, 437–447.

Eischen, C.M., Weber, J.D., Roussel, M.F., Sherr, C.J., and Cleveland, J.L. (1999). Disruption of the ARF-Mdm2-p53 tumor suppressor pathway in Myc-induced lymphomagenesis. Genes Dev. *13*, 2658–2669.

Fukushima, A., Okuda, A., Nishimoto, M., Seki, N., Hori, T.A., and Muramatsu, M. (1998). Characterization of functional domains of an embryonic stem cell coactivator UTF1 which are conserved and essential for potentiation of ATF-2 activity. J. Biol. Chem. *273*, 25840–25849.

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.

Jia, J., Lin, M., Zhang, L., York, J.P., and Zhang, P. (2007). The Notch signaling pathway controls the size of the ocular lens by directly suppressing p57Kip2 expression. Mol. Cell. Biol. *27*, 7236–7247.

Jiang, H., Shukla, A., Wang, X., Chen, W.Y., Bernstein, B.E., and Roeder, R.G. (2011). Role for Dpy-30 in ES cell-fate specification by regulation of H3K4 methylation within bivalent domains. Cell *144*, 513–525.

Kawamura, T., Suzuki, J., Wang, Y.V., Menendez, S., Morera, L.B., Raya, A., Wahl, G.M., and Izpisúa Belmonte, J.C. (2009). Linking the p53 tumour suppressor pathway to somatic cell reprogramming. Nature *460*, 1140–1144.

Kehrer-Sawatzki, H., Daumiller, E., Müller-Navia, J., Kendziorra, H., Rossier, E., du Bois, G., and Barbi, G. (2005). Interstitial deletion del(10)(q25.2q25.3 approximately 26.11)—case report and review of the literature. Prenat. Diagn. *25*, 954–959.

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.

Kim, Y., Sharov, A.A., McDole, K., Cheng, M., Hao, H., Fan, C.M., Gaiano, N., Ko, M.S., and Zheng, Y. (2011). Mouse B-type lamins are required for proper organogenesis but not by embryonic stem cells. Science *334*, 1706–1710.

Kooistra, S.M., van den Boom, V., Thummer, R.P., Johannes, F., Wardenaar, R., Tesson, B.M., Veenhoff, L.M., Fusetti, F., O'Neill, L.P., Turner, B.M., et al. (2010). Undifferentiated embryonic cell transcription factor 1 regulates ESC chromatin organization and gene expression. Stem Cells *28*, 1703–1714.

Ku, M., Koche, R.P., Rheinbay, E., Mendenhall, E.M., Endoh, M., Mikkelsen, T.S., Presser, A., Nusbaum, C., Xie, X., Chi, A.S., et al. (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. PLoS Genet. *4*, e1000242.

Landeira, D., Sauer, S., Poot, R., Dvorkina, M., Mazzarella, L., Jørgensen, H.F., Pereira, C.F., Leleu, M., Piccolo, F.M., Spivakov, M., et al. (2010). Jarid2 is a PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA Polymerase II to developmental regulators. Nat. Cell Biol. *12*, 618–624.

Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K., et al. (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. Cell *125*, 301–313.

Li, G., Margueron, R., Ku, M., Chambon, P., Bernstein, B.E., and Reinberg, D. (2010). Jarid2 and PRC2, partners in regulating gene expression. Genes Dev. *24*, 368–380.

Li, H., Collado, M., Villasante, A., Strati, K., Ortega, S., Cañamero, M., Blasco, M.A., and Serrano, M. (2009). The Ink4/Arf locus is a barrier for iPS cell reprogramming. Nature *460*, 1136–1139.

Lin, C.H., Jackson, A.L., Guo, J., Linsley, P.S., and Eisenman, R.N. (2009). Myc-regulated microRNAs attenuate embryonic stem cell differentiation. EMBO J. *28*, 3157–3170.

Ling, S.H., Qamra, R., and Song, H. (2011). Structural and functional insights into eukaryotic mRNA decapping. Wiley Interdisciplinary Reviews *2*, 193–208.

Loh, Y.H., Wu, Q., Chew, J.L., Vega, V.B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J., et al. (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat. Genet. *38*, 431–440.

Mendrysa, S.M., and Perry, M.E. (2000). The p53 tumor suppressor protein does not regulate expression of its own inhibitor, MDM2, except under conditions of stress. Mol. Cell. Biol. *20*, 2023–2030.

Meshorer, E., and Misteli, T. (2006). Chromatin in pluripotent embryonic stem cells and differentiation. Nat. Rev. Mol. Cell Biol. *7*, 540–546.

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

Nishimoto, M., Miyagi, S., Yamagishi, T., Sakaguchi, T., Niwa, H., Muramatsu, M., and Okuda, A. (2005). Oct-3/4 maintains the proliferative embryonic stem cell state via specific binding to a variant octamer sequence in the regulatory region of the UTF1 locus. Mol. Cell. Biol. *25*, 5084–5094.

Niwa, H., Miyazaki, J., and Smith, A.G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat. Genet. *24*, 372–376.

Niwa, H., Toyooka, Y., Shimosato, D., Strumpf, D., Takahashi, K., Yagi, R., and Rossant, J. (2005). Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. Cell *123*, 917–929.

Okuda, A., Fukushima, A., Nishimoto, M., Orimo, A., Yamagishi, T., Nabeshima, Y., Kuro-o, M., Nabeshima, Y., Boon, K., Keaveney, M., et al. (1998).

UTF1, a novel transcriptional coactivator expressed in pluripotent embryonic stem cells and extra-embryonic cells. EMBO J. *17*, 2019–2032.

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

Pasini, D., Cloos, P.A., Walfridsson, J., Olsson, L., Bukowski, J.P., Johansen, J.V., Bak, M., Tommerup, N., Rappsilber, J., and Helin, K. (2010). JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. Nature *464*, 306–310.

Peng, J.C., Valouev, A., Swigut, T., Zhang, J., Zhao, Y., Sidow, A., and Wysocka, J. (2009). Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. Cell *139*, 1290–1302.

Pomerantz, J., Schreiber-Agus, N., Liégeois, N.J., Silverman, A., Alland, L., Chin, L., Potes, J., Chen, K., Orlow, I., Lee, H.W., et al. (1998). The Ink4a tumor suppressor gene product, p19Arf, interacts with MDM2 and neutralizes MDM2's inhibition of p53. Cell *92*, 713–723.

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.

Shen, X., Kim, W., Fujiwara, Y., Simon, M.D., Liu, Y., Mysliwiec, M.R., Yuan, G.C., Lee, Y., and Orkin, S.H. (2009). Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells. Cell *139*, 1303–1314.

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.

Tan, S.M., Wang, S.T., Hentze, H., and Dröge, P. (2007). A UTF1-based selection system for stable homogeneously pluripotent human embryonic stem cell cultures. Nucleic Acids Res. *35*, e118.

Trcek, T., Larson, D.R., Moldón, A., Query, C.C., and Singer, R.H. (2011). Single-molecule mRNA decay measurements reveal promoter- regulated mRNA stability in yeast. Cell *147*, 1484–1497.

Tritschler, F., Braun, J.E., Motz, C., Igreja, C., Haas, G., Truffault, V., Izaurralde, E., and Weichenrieder, O. (2009). DCP1 forms asymmetric trimers to assemble into active mRNA decapping complexes in metazoa. Proc. Natl. Acad. Sci. USA *106*, 21591–21596.

Utikal, J., Polo, J.M., Stadtfeld, M., Maherali, N., Kulalert, W., Walsh, R.M., Khalil, A., Rheinwald, J.G., and Hochedlinger, K. (2009). Immortalization eliminates a roadblock during cellular reprogramming into iPS cells. Nature *460*, 1145–1148.

van den Boom, V., Kooistra, S.M., Boesjes, M., Geverts, B., Houtsmuller, A.B., Monzen, K., Komuro, I., Essers, J., Drenth-Diephuis, L.J., and Eggen, B.J. (2007). UTF1 is a chromatin-associated protein involved in ES cell differentiation. J. Cell Biol. *178*, 913–924.

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.

Vong, Q.P., Liu, Z., Yoo, J.G., Chen, R., Xie, W., Sharov, A.A., Fan, C.M., Liu, C., Ko, M.S., and Zheng, Y. (2010). A role for borg5 during trophectoderm differentiation. Stem Cells *28*, 1030–1038.

Zheng, X., Hu, G.Q., She, Z.S., and Zhu, H. (2011). Leaderless genes in bacteria: clue to the evolution of translation initiation mechanisms in prokaryotes. BMC Genomics *12*, 361.

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.

Zindy, F., Eischen, C.M., Randle, D.H., Kamijo, T., Cleveland, J.L., Sherr, C.J., and Roussel, M.F. (1998). Myc signaling via the ARF tumor suppressor regulates p53-dependent apoptosis and immortalization. Genes Dev. *12*, 2424–2433.

Cell

EXTENDED EXPERIMENTAL PROCEDURES

Mouse ESCs and Culturing

V6.5 ESCs (derived from the F1 hybrid of 129SvJae/C57BL/6) were maintained in DMEM (Dulbecco's modified Eagle's medium) supplemented with 15% fetal calf serum, 0.1 mM β-mercaptoethanol, 2 mM L-glutamine, 0.1 mM nonessential amino acid, 1000 U/ml recombinant leukemia inhibitory factor (LIF; Millipore), and 30 U/ml penicillin/streptomycin (the ESC medium) either with or without a feeder layer. For feeder-free culture, V6.5 ESCs and their derivatives were grown on plates coated with gelatin (Millipore).

Generation of *Utf1* null ESC Lines

The mouse genomic DNA targeting sequences containing *Utf1* gene were isolated from BAC clones (BMQ-251M19 from Sanger Institute for the conditional allele, RP24-292M18 from Children's Hospital Oakland Research Institute for the conventional knockout allele) by gap repair (http://recombineering.ncifcrf.gov). Briefly, primer pairs a1 and a2, b1, and b2 (Table S7) were used to PCR amplify two small fragments corresponding to the 5′ and 3′ noncoding regions of *Utf1* gene, respectively. The two fragments were subcloned into pDTA4B by digesting with ApaI and SacII, BglII and Xho I, respectively. This vector (pDTA4B-*Utf1*) was then used to recombine with the BAC clones above to acquire the whole genomic region of *Utf1* locus. For conventional targeting, primers e1 and e2 were used to PCR amplify PGK-Neo cassette and then recombined with the pDTA4B-*Utf1* to delete the whole *Utf1* coding sequence. This gives rise to the conventional targeting vector. For *Utf1* conditional knockout vector, the primer pair d1 and d2 was used to PCR amplify the fragment containing Loxp-PGK-Neo-Loxp in pL452. The fragment was then recombined with pDTA4B-*Utf1* and then the Cre-expressing bacteria was used to remove PGK-Neo, which introduced a Loxp site at the 3′ end of *Utf1* to produce pDTA4B-*Utf1*-Loxp. The primer pair c1 and c2 was used to PCR amplify the fragment containing Loxp-FRT-PGK-neo-FRT in pL451. This fragment was recombined into pDTA4B-*Utf1*-Loxp to produce the final conditional targeting vector. Primers used are listed in Table S7. The *Utf1* conditional knockout vector was linearized by Pme1 digestion and electroporated into V6.5 ESCs cultured on a feeder layer to generate a pre-floxed allele next to the PGK-Neo cassette. G418-resistant heterozygous *Utf1* pre-floxed ESCs were transiently transfected with a FLP-recombinase-encoding plasmid (pCAG-Flpe, Addgene plasmid 13787), which converts the pre-floxed allele into a floxed allele with one loxP site on either side of *Utf1*. The $Utf1^{+/f}$ ESCs were then electroporated with the conventional *Utf1* knockout vector linearized by Pme1 and selected in G418 to generate $Utf1^{-/f}$ cells. In order to generate $Utf1^{-/\Delta}$ cells, $Utf1^{-/f}$ cells were infected with a low-titer Cre-recombinase-encoding lentivirus (Addgene plasmid 17408) followed by selection using puromycin. Genotyping primers and Southern probe sequences are listed in Table S7 and illustrated in Figures S1F and S1G.

Utf1 Immunoprecipitation and MS Analyses

ESCs ($1–2 \times 10^9$) were lysed with 10 volumes of hypotonic buffer (0.001 mM $MgCl_2$, 0.003 mM KCl, 0.1 mM Tris pH7.3). Nuclei were pelleted at 5000 g for 15 min. The nuclear pellet was incubated with 1 volume of high-salt buffer (0.5 M KCl, 0.0001 mM EDTA, 0.0025 mM $MgCl_2$, 0.4 mM Tris, pH7.3, 6% glycerol) for 30 min. After sedimentation at 25,000 g for 20 min at 4°C, the supernatant was dialyzed against 2-l low-salt buffer (0.15 mM KCl, 0.0001 mM EDTA, 0.0025 mM $MgCl_2$, 0.4 mM Tris, pH7.3, 6% glycerol) for 30 min. After dialysis, the extract was incubated with 20 μg anti-Utf1 (Abcam, ab24273) antibody at 4°C for 2 hr with rocking and then with 100 μl protein A beads (Santa Cruz) for 1 hr. The protein-A beads were washed with 2 ml NETN (20 mM Tris at pH 8.0, 100 mM NaCl, 1 mM EDTA and 0.5% NP-40) four times.

For MS, the beads were incubated with 150 μl of 8M urea in 100 mM Tris HCl (pH 8.5) to denature proteins on-bead. The proteins together with the beads were then reduced with TCEP at room temperature for 20 min and alkylated using iodoacetamide (IAM) for 15 min at room temperature in the dark. Subsequently 600 μl of 100 mM Tris HCl (pH 8.5) was added to the reaction tube to dilute the urea concentration. Trypsin was then added at the enzyme:substrate ratio of 1:100 to digest the proteins on-bead overnight. The digestion process was stopped by adding formic acid to a final concentration of 1%. Beads were removed by centrifugation at 14,000 rpm for 10 min.

The protein digest was pressure-loaded onto a fused silica capillary desalting column containing 3 cm of 5 μm Aqua C18 material (Phenomenex, Ventura, CA, USA) and 3 cm 5 μm Partisphere strong cation exchanger (Whatman, Clifton, NJ, USA) packed into a 250 μm i.d. capillary with a 2 μm filtered union (UpChurch Scientific, Oak Harbor, WA, USA). The desalting column was washed with buffer containing 95% water, 5% acetonitrile, and 0.1% formic acid. After desalting, a 100 μm i.d. capillary with a 5-μm pulled tip packed with 10 cm 5 μm Aqua C18 material (Phenomenex) was attached to the filter union and the entire split-column (desalting column–filter union–analytical column) was placed inline with an Agilent 1100 quaternary HPLC (Palo Alto, CA, USA) and analyzed using a modified 7-step separation described previously (Washburn et al., 2001). The buffer solutions used were 5% acetonitrile/0.1% formic acid (buffer A), 80% acetonitrile/0.1% formic acid (buffer B), and 500 mM ammonium acetate/5% acetonitrile/0.1% formic acid (buffer C). Step 1 consisted of a 75 min gradient from 0%–100% buffer B. Steps 2–7 had the following profile: 5 min of 100% buffer A, 5 min of X% buffer C, a 95 min gradient from 0%–100% buffer B, and 15 min of 100% buffer A. The 5 min buffer C percentages (X%) were 10, 20, 30, 50, 70, 100%, respectively, for the step 2 to step 7 analyses.

As peptides eluted from the microcapillary column, they were electrosprayed directly into a LTQ-Orbitrap mass spectrometer (ThermoFinnigan, Palo Alto, CA, USA) with the application of a distal 2.5 kV spray voltage. A cycle of one full-scan mass spectrum

followed by 6 data-dependent MS/MS spectra at 35% normalized collision energy was repeated continuously throughout each step of the multidimensional separation. Applications of mass spectrometer scan functions and HPLC solvent gradients were controlled by the Xcalibur datasystem.

Interpretation of MS Data

Tandem mass spectra were extracted from raw files. The tandem mass spectra were searched using the SEQUEST algorithm (Eng et al., 1994) against the EBI IPI mouse protein database (version 3.52, downloaded on Nov 21, 2008) with its reversed decoy attached (Peng et al., 2003). SEQUEST searches were done with peptide mass tolerance set to 50 ppm. No differential modifications were considered. The mass of the amino acid cysteine was statically modified by +57.02146 Da, due to carboxyamidomethylation of the sample. No enzymatic cleavage conditions were imposed on the database search. The validity of peptide/spectrum matches was assessed in DTASelect (Tabb et al., 2002) using SEQUEST-defined parameters, the cross-correlation score (XCorr), and normalized difference in cross-correlation scores (DeltaCN). The distribution of XCorr and DeltaCN values for (a) direct and (b) decoy database hits was obtained, and the two subsets were separated by quadratic discriminative analysis. The discriminative score was set such that a false-positive rate of 1% at the protein level was determined based on the number of accepted decoy proteins.

Two sets of either Utf1 or control IgG immunoprecipitations were sequenced. Candidate proteins present in two rounds of Utf1 MS analyses were kept if they were not present in the IgG control MS, or have 15-, 35-, or 5-fold higher SequenceCount, spectrumeCount, or peptide coverage, respectively, than those in the IgG MS. Among these candidate proteins, only those that have more than 5 SequenceCount, 10 spectrume Count, and 20% coverage in at least one MS analysis were kept (see Table S5).

Precipitation of Biotin-Utf1 by Streptavidin-Coupled Beads and Western Blotting Analyses

BirA only ESCs or biotin-Utf1 ESCs (1×10^7 each) were treated with hypotonic buffer to prepare nuclei (see above). The nuclear pellets were lysed by 2.5 volume M-NETN (1M Tris, pH 8.0, 5 mM $CaCl_2$, 100 mM NaCl, 0.5% NP-40, 4000 gel unit/ml Micrococcal nuclease). After further clarification of the nuclear extract, 100 μl streptavidin-coupled bead slurry was used to precipitate biotin-Utf1 for 2 hr. The beads were washed four times with 2 ml each of NETN buffer, boiled in sample buffer, and then analyzed by western blotting.

EB, Teratoma, and TE Differentiation Assays

For EB differentiation, ESCs were diluted to 10,000 cells/ml in ESC medium lacking LIF, and 20 μl drops were arranged on the surface of a 15cm Petri dish lid. The lid was gently inverted onto the plate (with water in the plate to keep the drops from evaporating) and incubated at 37°C with 5% CO_2 for 3 days to allow aggregation. After 3 days EBs were transferred to low-attachment 96-well plates (Corning) at 1 EB/well in the same culture medium. EBs were cultured up to 10–12 days, and samples were analyzed at indicated time points. For teratoma generation, the *Utf1*$^{+/f}$ and *Utf1*$^{-/\Delta}$ ESCs (1×10^6) were injected into nude mice subcutaneously. Tumors were allowed to grow for 30 days before they were dissected out and analyzed. To induce TE or TS differentiation, ZHBTc4 ESCs were used as described previously (Niwa et al., 2000).

Karyotype Analyses of ESCs

One day before the karyotyping, a 70% confluent ESC plate was split by 1:2. ESCs (10^5) were trypsinized and resuspended in 5 ml 0.56% KCl in water. After incubation at 37°C for 10 min, 1 ml 25% freshly made fixative solution (methanol:glacial acetic acid = 3:1 by volume) in water was added. Cells were pelleted and resuspended in 1 ml 100% fixative solution three times. To make cell spreads, one drop of the cell suspension was dropped (from a height of ~0.5 m) onto a glass slide and allowed to air dry before DAPI staining. The mitotic chromosome number was counted under the fluorescence microscope.

Quantitative Real-Time PCR and Northern Blotting

RNA was isolated from 1×10^7 ESCs or 100 EBs using the RNeasy Plus kit (QIAGEN). First-strand cDNA synthesis was done with the cDNA synthesis kit (Bio-Rad). The relative transcript level of each gene in *Utf1*$^{+/f}$ and *Utf1*$^{-/\Delta}$ samples was determined by normalizing to *GAPDH* mRNA level. qRT-PCR was performed with SYBR Green Master Mix (Bio-Rad) on a PCT-200/CFD-3200 Opticon system (MJ Research). Primers for tested genes are listed in Table S7.

RNA was isolated from 1×10^7 ESCs using the RNeasy Plus kit (QIAGEN), and northern blotting analyses were performed according to standard procedures. Briefly, RNA samples were separated on a formaldehyde–agarose gel and transferred to nitrocellulose filters. The filters were hybridized with radiolabeled *Arf* or 18S rRNA cDNA probes.

Nuclear Run-on Assay

Nuclear run-on assay was performed as described (Smale, 2009). Briefly, ESCs were first treated by a Myc inhibitor 10058-F4 (Sigma cat #F3680) and then harvested after 16 hr treatment. In order to pause the RNA polymerase, cells were chilled on ice, and then permeabilized by NP-40 lysis buffer. The nuclei were then incubated for 30 min at 37°C in the presence of nucleoside triphosphates (NTPs) and α-^{32}P-UTP to restart the transcription. To determine the relative transcripts level of *Arf* and *Txnip* in each sample, the radiolabeled RNAs were purified and hybridized to a membrane containing immobilized cDNAs of the corresponding gene. The densities of dots on membranes were quantified by PhosphorImager.

RIP-qPCR

RIP was performed following manufacture's procedure and conditions as described (Millipore). Nuclear extract from 10^8 ESCs was used for each RIP-qPCR. Nuclear extract was lysed by RIP lysis buffer (100 mM KCl, 5 mM MgCl2, 10 mM HEPES pH 7.0, 0.5% NP40, 1 mM DTT, DNase, RNase inhibitor [Roche], protease inhibitor cocktail). The lysate was diluted in the RIP dilution buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM $MgCl_2$, 0.05% NP40) and incubated with 10 μg of Dcp1a antibody at 4°C overnight. Then, the RNA/antibody complex was precipitated by protein-A beads. After several washes, the RNA was extract by phenol:chloroform:isoamyl alcohol and precipitated. Next, first-strand cDNA synthesis was done with the cDNA synthesis kit (Bio-Rad). The *Arf* RNA level was determined by qRT-PCR performed with SYBR Green Master Mix (Bio-Rad) on a PCT-200/CFD-3200 Opticon system (MJ Research).

RNAi-Mediated Gene Silencing in ESCs

shRNAs were purchased from Open Biosystems to knockdown mRNAs of *Arf* (TRCN0000077816), *Utf1* (TRCN0000081708), and *Dcp1a* (shRNA-1:TRCN0000096664, shRNA-2:TRCN0000096665). To package the shRNA-expressing lentivirus, HEK293T cells were plated in 6-well plates at 6 × 10^5 cells/well. The shRNA plasmids and lentiviral components were cotransfected into HEK293T cells. Viral supernatants were collected 48 hr after cotransfection and the ESCs were directly infected with the viral supernatant with 2 mM polybrene using a low -speed spin (300 g) for 45 min at room temperature. The infected cells were selected 24 hr post-infection with ESC medium containing 1.5 μg/ml puromycin for 2 days before further analyses.

Antibodies for Western Blotting and Immunostaining Analyses

Cells were harvested and lysed in SDS-PAGE sample buffer. Equal amounts of total protein were loaded in each lane. Proteins were resolved by SDS-PAGE, transferred to nitrocellulose membranes, and probed with the indicated antibodies. The antibodies used for western blotting and immunostaining were as follows: anti-Utf1 (Abcam ab24273), anti-Suz12 (Cell Signaling 3737), anti-α-tubulin (Sigma DM1α), anti-Oct4 (Abcam ab27985), anti-Sox2 (Abcam ab15830), anti-Nanog (Abcam ab14959), anti-p53 (EMB pab421), anti-$p21^{Cip}$ (Santa Cruz F-5), anti-Arf (Abcam ab80), anti-Ink4a (Santa Cruz M-156), anti-Dcp1a (Sigma WH0055802-M6).

ChIP Assays

The ChIP assay was performed according to the condition suggested by the manufacturers of ChIP-grade antibodies to H3K27me3 (Abcam, ab6002), H3K4me3 (Cell Signaling, 9751), Suz12 (Cell Signaling, 3737), and Dcp1a (Sigma WH0055802-M6). Briefly, formaldehyde was added to cultured ESCs to cross-link proteins to DNA and the cells (10^7) were lysed in 200 μl lysis buffer (50 mM HEPES, pH 7.5, 140 mM NaCl; 1 mM EDTA; 1% Triton X-100; 0.1% sodium deoxycholate; 0.1% sodium dodecyl sulfate) (Jia et al., 2007). Cell lysates were sonicated using a Bioruptor ultrasonic cell disruptor (Diagenode) to shear genomic DNA to an average fragment size of 200 to 500 bp and then immunoprecipitated with antibodies. The precipitated DNA was purified using a QIAquick PCR purification kit (QIAGEN), followed by qRT-PCR performed with iQ SYBR Green Supermix (Bio-Rad) on a PCT-200/CFD-3200 Opticon system (MJ research). The primers are listed in Table S7. At least three biological replicates were performed.

Biotin-ChIP Assay

Biotin-ChIP was performed as described (Kim et al., 2009) with modifications. Briefly, the BirA-expressing plasmid (pEF1α*BirA*V5-neo) was stably transfected into V6.5 ESCs. *Utf1* CDS was PCR amplified using primer pair f1 and f2 (Table S7) and subcloned into the biotin-tag plasmid (pEF1αFlagbio-puro) with the biotin-tag at the N terminus. After transfecting the biotin-Utf1 expressing plasmid into V6.5-BirA ESCs, several puromycin-resistant clones were picked. Clones that expressed biotin-Utf1 at less than 5% of the endogenous Utf1 as judged by western blotting analyses (probing with HRP-streptavidin and Utf1 antibody) were used for biotin-ChIP. 100 μl streptavidin-coupled beads (Dynabeads MyOne streptavidin T1) were incubated with sheared DNA made from 10^7 cross-linked biotin-Utf1-expressing ESCs. After precipitation, the beads were washed twice with 2 ml 2% SDS followed by the washing steps described in the protocol from the Millipore ChIP assay kit (catalog#17-295). Biotin-ChIP reactions from BirA-expressing V6.5 ESCs were used as controls. For ChIP-qPCR 10^7 ESCs were used to make genomic DNA. The primer sequences are included in Table S7. At least three biological replicates were performed.

ChIP-Seq and Data Analyses

For ChIP-seq, 5 × 10^7 cells were used and genomic DNA was sheared to an average fragment size of 150–250 bp. After streptavidin-coupled bead precipitation or immunoprecipitation, ChIP-seq libraries were prepared following Illumina protocols. For biotin-Utf1-ChIP-seq and input carried out in unmodified wild-type ESCs, the libraries were sequenced using an Illumina Genome Analyzer GAII. Reads that were 25 base pairs were obtained. For ChIP-seq of H3K4me3, Suz12, H3K27me3, and H3 in $Utf1^{+/f}$ or $Utf1^{-/\Delta}$ ESCs, the libraries were sequenced using Illumina HiSeq 2000 and 50 bp long reads were obtained. All sequences were mapped to the mouse genome (mm9) by Bowtie allowing 2 mismatches per read (Langmead et al., 2009). Only uniquely mapped reads were retained. For redundant reads that fall on exactly the same position on the genome, only one was retained. Since our analysis incorporated direct comparisons for ChIP-seq result of Suz12, H3K27me3, and H3 between $Utf1^{+/f}$ ESCs and $Utf1^{-/\Delta}$ ESCs, to avoid bias introduced by the differential sequencing depth, the data were trimmed to the same size in each pair of samples.

The densities of short sequence tags of Utf1, Suz12, H3K27me3, and H3 were normalized to the library size and the number of sites, namely *r*eads *p*er *k*ilo-bases per *m*illion reads (RPKM). The 5′ and 3′ tags were shifted +100 bp and −100 bp, respectively, and merged to obtain the density. The tag density profiles in Figures 1A, 2A, S2A, and S3A were smoothened using a 500 bp window sliding along the chromosomes with a step length of 100 bp.

ChIP-seq fold enrichment is defined as the ratio of ChIP-seq tag density to that of input controls. For statistical robustness, the confidence interval of this ratio is calculated (exact Poisson test, confidence level 0.95). If the lower confidence limit is larger than 1, the ratio is set to this lower limit; or if the upper limit is smaller than 1, the ratio is set to this upper limit; otherwise the ratio is set to 1. The fold enrichment profiles in Figures 1B–1D, S1D, S1E, 2B, 5B, and 5D were calculated with this method by comparing corresponding ChIP-Seq data to the input control in 1 kb windows sliding every 200 bp through the promoter region. The profiles of increased Suz12 binding and H3K27me3 upon loss of Utf1 in Figures 2A, 2B, S2A, 5B, and 5E were calculated by the same method comparing Suz12-ChIP-Seq and H3K27me3-ChIP-seq of $Utf1^{-/\Delta}$ to $Utf1^{+/f}$ ESCs. Wilcoxon test on Suz12 increase (Figure 5E) and CpG density (Figure 5C) were performed on Suz12 increase and CpG densities within −1 kb to 1 kb of TSS of corresponding gene groups.

ChIP-seq peaks for Utf1, H3K4me3, Suz12, and H3K27me3 were identified by SICER (Zang et al., 2009) using input as the control and FDR = 0.001. Utf1-binding genes are defined when the gene promoter regions overlap with the Utf1 peaks. Bivalent genes are defined by the overlapping of both H3K27me3 islands and Suz12 islands ($Utf1^{+/f}$) with their promoters. The 2501 strong PRC2-binding bivalent genes are bivalent genes that have the highest normalized Suz12 ($Utf1^{+/f}$) tag density in −1 kb to 1 kb of TSS.

Functional enrichment analyses of the top 3,000 strong Utf1-binding genes were calculated by DAVID (Huang et al., 2009a, 2009b). For motif finding, Utf1 ChIP-Seq peaks were identified by MACS (Zhang et al., 2008) with p value < 10^{-15}. Further filtering with criteria normalized tag density > 5 and fold enrichment > 10 resulted in 3,790 peaks. 50 bp up- and downstream of the peak summits were supplied to the motif-finding software MEME using the ZOOPS model (Bailey and Elkan, 1994).

Unless specified, all analyses were done by custom C++ programs (available upon request). Heatmaps were drawn by the R statistical package.

Whole-Transcriptome Shotgun Sequencing (RNA-Seq) and Data Analyses

The total RNA was isolated from 10^7 ESCs with the RNeasy Plus kit (QIAGEN). The poly(A)-containing mRNA molecules were purified, and libraries were built following Illumina TruSeq RNA protocols. Libraries were sequenced using an Illumina HiSeq 2000. 100 bp-long reads from both ends were obtained. Low-quality ends of reads (quality score below 20) were trimmed and reads shorter than 36 bp after trimming were filtered out. All remaining sequences were mapped to the mouse genome (mm9) using TopHat with default settings. RefSeq annotation downloaded from the UCSC table browser in June 2011 was supplied to TopHat as reference for exon junctions (Trapnell et al., 2009). Differentially expressed genes were called by edgeR with p < 10^{-5} and fold change > 1.5 (Robinson et al., 2010). Only uniquely mapped reads were counted and TMM scaling factor normalization was used in running edgeR. The heatmap of gene expression levels in Figure 5B was plotted by defining gene expression levels as the number of RNA-Seq tags normalized by mRNA length. Table S6 lists the genes that are upregulated by both of the Dcp1a-targeting shRNAs in ESCs.

S1 Nuclease Protection Assay

Total RNA from whole cells and nuclear extract was hybridized overnight at 55°C to ^{32}P- labeled 80 nt long antisense probes with 40 nt corresponding to *Arf* exons 1, 2, and 3 sequences at the 5′ end and 40 nt correspond to the adjacent intron sequences. A fully spliced *Arf* mRNA would protect the 40 nt exon sequences in the probes. As a positive control, the probes were hybridized to their complementary 80 nt sequences. Hybridization products were digested with 100 U of S1 nuclease (Invitrogen) and separated on a 15% polyacrylamide gel. Probe sequences can be found in Table S7.

EMSA

Utf1 wild-type and knockout V6.5 ESCs were cultured as described above. To overexpress biotin-Utf1 and biotin-Utf1(P1 mutant), the corresponding expression plasmids were transiently transfected into Utf1 null ESCs and the overexpression was verified by western blotting. Nuclear extracts from $Utf1^{-/\Delta}$ ESCs, $Utf1^{+/f}$ ESCs, and $Utf1^{-/\Delta}$ ESCs overexpressing biotin-Utf1 or biotin-Utf1-P1 mutant that does not bind to DNA were prepared as described above.

A 74 bp DNA sequence from the promoter of *Lypd1* containing Utf1- and Jarid2-binding motifs was used as the positive probe (Probe1). A 60 bp mouse genomic DNA sequence without the predicted Utf1 and Jarid2 motif was used as the negative control probe (Probe2). Probes were labeled by γ-^{32}P-ATP (Perkin Elmer). Unincorporated γ-^{32}P-ATP was removed by G-25 spin columns (Roche).

EMSAs were performed as follows: DNA-binding reactions (20 μl) containing 10 mM Tris (pH 7.5), 50 mM KCl, 5 mM MgCl2, 0.2 mM EDTA, 2.5% glycerol, 50 ng/μl poly(dI-dC), 0.05% NP-40, and 10 μl nuclear extract were preincubated with the corresponding probe and antibodies for 20 min. Each reaction mixture was then loaded onto a native 10% TBE polyacrylamide gel (Bio-Rad) and electrophoresed in 0.5 × TBE. After electrophoresis, the gels were fixed in the gel fixing solution (10% acetic acid, 20% methanol, 70% water), dried at 80°C for 1 hr and exposed to X-ray films.

SUPPLEMENTAL REFERENCES

Bailey, T.L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proceedings/International Conference on Intelligent Systems for Molecular Biology; ISMB 2, 28–36.

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.

Eng, J., McCormack, A., and Yates, J. (1994). An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J. Am. Soc. Mass Spectrom. 5, 976–989.

Huang, W., Sherman, B.T., and Lempicki, R.A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 37, 1–13.

Huang, W., Sherman, B.T., and Lempicki, R.A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4, 44–57.

Jia, J., Lin, M., Zhang, L., York, J.P., and Zhang, P. (2007). The Notch signaling pathway controls the size of the ocular lens by directly suppressing p57Kip2 expression. Mol. Cell. Biol. 27, 7236–7247.

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.

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

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

Niwa, H., Miyazaki, J., and Smith, A.G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat. Genet. 24, 372–376.

Peng, J., Elias, J.E., Thoreen, C.C., Licklider, L.J., and Gygi, S.P. (2003). Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. J. Proteome Res. 2, 43–50.

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.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140.

Smale, S.T. (2009). Nuclear run-on assay. Cold Spring Harb Protoc. 2009, t5329.

Tabb, D.L., McDonald, W.H., and Yates, J.R., 3rd (2002). DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. J. Proteome Res. 1, 21–26.

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

Washburn, M.P., Wolters, D., and Yates, J.R., 3rd (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat. Biotechnol. 19, 242–247.

Zang, C., Schones, D.E., Zeng, C., Cui, K., Zhao, K., and Peng, W. (2009). A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. Bioinformatics 25, 1952–1958.

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.

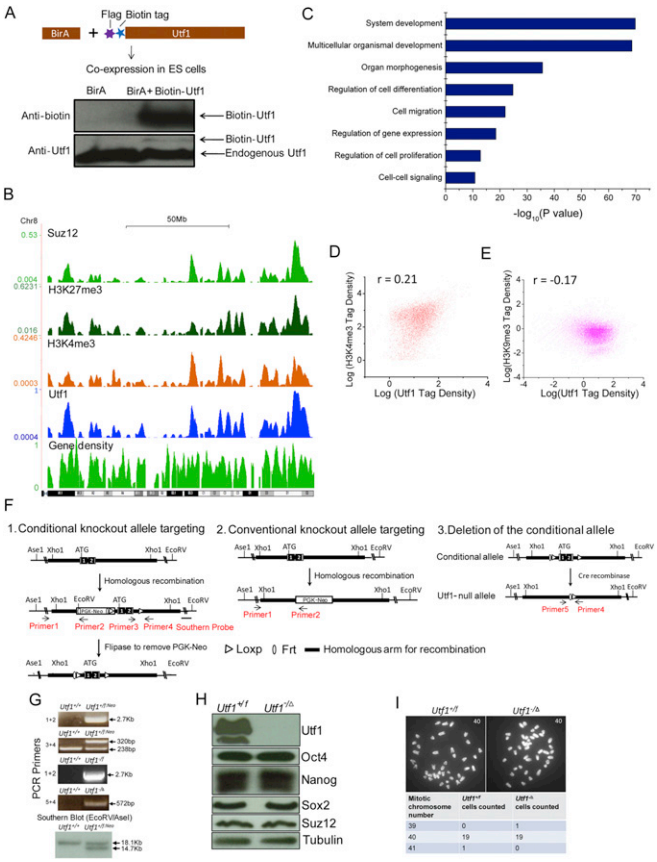

Figure S1. Utf1 Occupies Bivalent Genes to Regulate Their Expression, Related to Figure 1

(A) Establishment of ESCs expressing the subendogenous amount of biotin-Utf1. Utf1 was fused to a Flag epitope tag and a BirA recognition peptide sequence (the biotin tag) at its N terminus (Kim et al., 2009). This tagged Utf1 was coexpressed with the bacteria biotin ligase (BirA) in ESCs. Biotin-Utf1, detected by both HRP-streptavidin and anti-Utf1 antibody, was estimated to be expressed at less than 5% of endogenous Utf1 levels.

(B) The island densities of Suz12 (this study), H3K27me3 (Mikkelsen et al., 2007), H3K4me3 (this study), and Utf1 (this study) along chromosome 8. Utf1-binding islands are determined by SICER with the input as controls (FDR < 0.001). Utf1-binding island densities were determined within each 10 kb window along chromosomes and plotted (the values on the y axis are calculated by dividing the length of Utf1 binding islands by 10 kb). The gene density (the bottom plot) was also determined in the same 10 kb window by calculating the percentage of the total transcribed sequences (plotted on the y axis). Densities plotted on y axes are calculated as in Figure 1A. The input was used as controls because the control biotin-ChIP-seq of ESCs expressing only BirA showed no enrichment throughout the genome (data not shown).

(C) Biological processes of Utf1-occupied genes as determined by Gene Ontology (GO) analyses. y axis shows gene ontology terms and x axis represents the p values for the Utf1-bound genes. See Table S1 for a complete list.

(D and E) Scatterplots of Utf1 versus H3K4me3, r = 0.21 (Spearman's rank correlation coefficient of the region with Utf1 > 0 and K3K4me3 > 0]) (D) and Utf1 versus H3K9me3, r = −0.17 (E).

(F) Illustrations of strategies used to generate *Utf1* null ESCs. (1) For creating the conditional knockout allele, a targeting vector containing the Neomycin selection cassette (PGK-Neo), Frt sites, and loxP sites (flanking the genomic region corresponding to the *Utf1* gene) was recombined with the wild-type allele in ESCs to create a conditional knockout allele of *Utf1*. The Neo cassette was removed by Flipase to generate ESCs containing a conditional allele of $Utf1^{+/f}$. (2) To create a conventional knockout allele, a PGK-Neo cassette was recombined with the other wild-type allele in the $Utf1^{+/f}$ ESCs to delete the second allele of *Utf1*. (3) Cre recombinase was used to delete the floxed *Utf1* allele.

(G) PCR using the indicated primer combinations (primers 1, 2, 3, 4, and 5) to determine the correct gene targeting. Southern blotting shows correct targeting of the conditional allele using the indicated Southern-blot probe. Sequences for the PCR primers and the Southern-blot probe can be found in Table S7.

(H) The expression of Utf1 and pluripotency proteins. The $Utf1^{-/\Delta}$ ESCs lack the expression of Utf1 but express similar amounts of Oct4, Nanog, Sox2, and Suz12 as the $Utf1^{+/f}$ ESCs. Loading controls, tubulin.

(I) Karyotype analyses of $Utf1^{+/f}$ and $Utf1^{-/\Delta}$ ESCs. 95% of both types of ESCs maintained their proper karyotype after 40 successive passages in vitro.

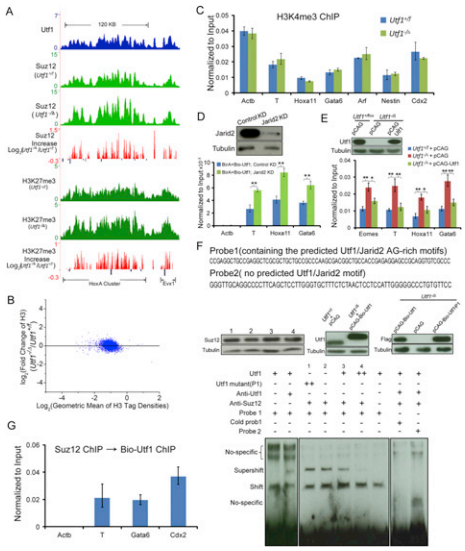

Figure S2. Utf1 Binds to Bivalent Genes to Limit PRC2 Loading and H3K27me3, Related to Figure 2

(A) Genome browser representations showing similar spatial patterns of Utf1, Suz12, and H3K27me3 enrichment at a genomic domain (~200 kb) that encompasses the *HoxA* gene cluster. The y axes plot the number of *r*eads *p*er *k*ilobases per *m*illion reads (RPKM). The increase (red peaks) or decrease (blue peaks, below zero) of Suz12 and H3K27me3 (Log_2 fold change) in the absence of Utf1 in this region is plotted.

(B) An MA-plot of one of the two histone H3-ChIP-seq biological repeats on promoters exhibiting Suz12 increase upon Utf1 deletion. The two sets of histone H3-ChIP-seq data are similar. Log_2 of the normalized H3 tag densities within −1 kb to 1 kb of the TSS in *Utf1*$^{+/f}$ *and Utf1*$^{-/\Delta}$ ESCs were averaged and plotted on the x axis. The y axis plots the log-ratio of H3 tag densities between *Utf1*$^{-/\Delta}$ and *Utf1*$^{+/f}$ ESCs.

(C) H3K4me3-ChIP-qPCR of selected bivalent genes that exhibit increase of PRC2 binding and H3K27me3 upon Utf1 loss. No increase of H3K4me3 was observed upon Utf1 deletion on these genes. Two biological repeats of H3K4me3-ChIP-seq of *Utf1*$^{+/f}$ *and Utf1*$^{-/\Delta}$ ESCs also showed no increase of H3K4me3 upon Utf1 deletion on bivalent genes (data not shown). Error bars, SD of triplicates.

(D) shRNA knockdown (KD) of *Jarid2* expression in *Utf1*$^{+/f}$ ESCs. Loading controls, tubulin. The graph plots biotin-ChIP-qPCR analyses of Utf1 binding at selected target genes in *Utf1*$^{+/f}$ ESCs after reduction of *Jarid2* expression. The *Actb* locus was used as controls. Error bars, SD of triplicates. Student's t test, $^{**}p < 0.01$.

(E) Restoration of *Utf1* expression in *Utf1*$^{-/\Delta}$ ESCs by a pCAG-Utf1 plasmid. The control plasmid is the parent pCAG plasmid alone. Loading controls, tubulin. The graph plots ChIP-qPCR analyses of Suz12 binding at selected bivalent genes in *Utf1*$^{+/f}$ or *Utf1*$^{-/\Delta}$ ESCs transfected with pCAG or pGAG-Utf1 as indicated. Re-expression of Utf1 restored the Suz12-binding levels on the selected promoters in *Utf1*$^{-/\Delta}$ ESCs to those seen in *Utf1*$^{+/f}$ ESCs. The *Actb* locus was used as ChIP controls. Error bars, SD of triplicates. Student's t test, $^{*}p < 0.05$, $^{**}p < 0.01$.

(F) Utf1 and PRC2 compete with one another to bind to DNA as revealed by the EMSA. Probes for EMSA: probe1 containing two predicted AG-rich motifs bound by Utf1/Jarid2 are derived from the promoter of the bivalent gene *Lypd1*. Probe2 containing no Utf1/Jarid2-binding motif is derived from a noncoding region of the genome. Western blotting analyses: First panel, the lysates used for EMSA were probed by tubulin and Suz12 antibodies to show equivalent loading (lanes 1–4 on the western blot correspond to the EMSA gel lanes 1–4 below). Second panel, *Utf1*$^{+/f}$ ESCs transiently transfected with control pCAG plasmid and *Utf1*$^{-/\Delta}$ ESCs transiently transfected with pCAG-biotin-Utf1 were analyzed using Utf1 antibodies, which show that biotin-Utf1 was overexpressed in *Utf1*$^{-/\Delta}$ ESCs compared to *Utf1*$^{+/f}$ ESCs (probed using the Utf1 antibody). Third panel, overexpression of mutant biotin-Utf1-P1 (failed to bind to DNA due to the change of four leucines at amino acids 291, 294, 298, and 301 to arginines, data not shown) in *Utf1*$^{-/\Delta}$ ESCs is compared to *Utf1*$^{-/\Delta}$ ESCs overexpressing biotin-Utf1, which shows that biotin-Utf1-P1 was overexpressed to a higher level than biotin-Utf1 (Flag antibody was used to probe for Utf1). Loading controls, tubulin. EMSA assay: Probe1 was incubated with nuclear extracts from different ESC lysates as indicated. Utf1− or Utf1 + indicates lysates made from Utf1 null or *Utf1*$^{+/f}$ ESCs, respectively. Utf1 ++ or Utf1 mutant (P1) ++ indicates lysates made from *Utf1*$^{-/\Delta}$ ESCs overexpressing wild-type or DNA-binding deficient Utf1 mutant, respectively. The nonradiolabelled Probe1 was used as a competitor. Probe2 was used as a negative control. Antibodies against Utf1 and Suz12 were used to induce super-shift.

(G) Sequential ChIP using Suz12 antibody followed by biotin-ChIP-Utf1. Q-PCR analyses show that Suz12 and Utf1 bind to the same bivalent promoters. Error bars, SD of triplicates.

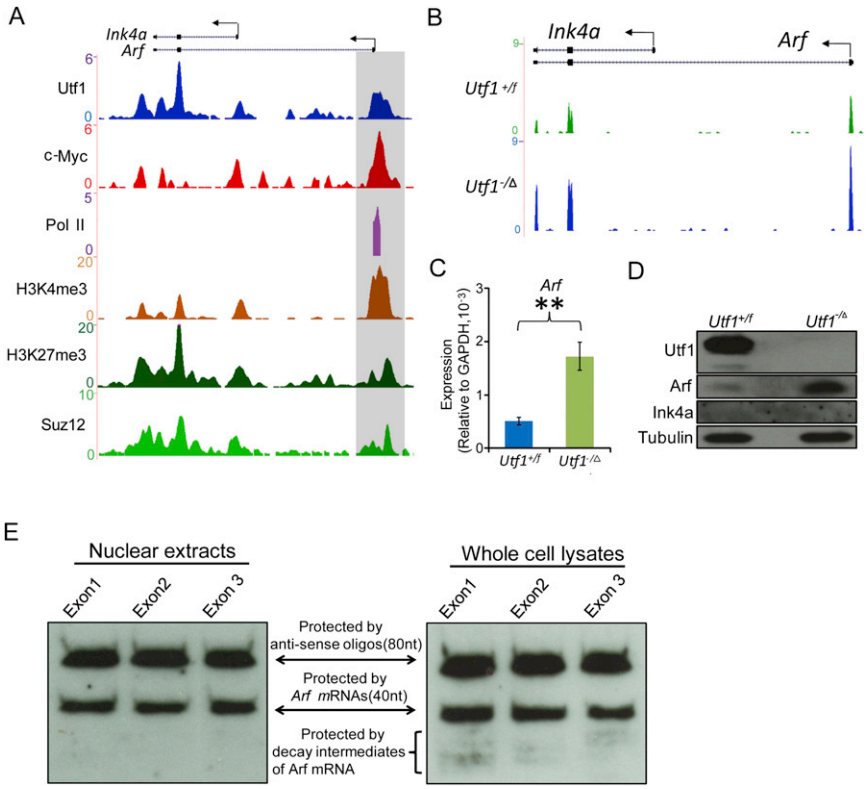

Figure S3. Utf1 Promotes the Binding of Dcp1a to mRNAs in the Nucleus for Cytoplasmic Degradation, Related to Figure 4

(A) Genome browser view of ChIP-seq peaks of Utf1 (this study), c-Myc (Chen et al., 2008), Pol II (Rahl et al., 2010), H3K4me3 (this study), H3K27me3 (Mikkelsen et al., 2007), and Suz12 (this study) at the *Cdkn2a* (*Ink4a-Arf*) locus. Arrows indicate the TSS and directions. Utf1, Suz12, Pol II, and c-Myc are all enriched at the *Arf* promoter (gray shading).

(B) Genome browser view of RNA-seq peaks of the *Cdkn2a* locus in $Utf1^{+/f}$ and $Utf1^{-/\Delta}$ ESCs. y axes plot the RPKM values.

(C) RT-qPCR analyses of *Arf* expression in $Utf1^{+/f}$ and $Utf1^{-/\Delta}$ ESCs. Error bars, SD of triplicates. Student's t test, $^{**}p < 0.01$.

(D) Western blotting analyses of Arf and Ink4a in $Utf1^{+/f}$ and $Utf1^{-/\Delta}$ ESCs. Loading controls, tubulin.

(E) S1 nuclease protection assays. The presence of 40 nt bands indicates that the *Arf* mRNA is properly spliced, which would only protect the 40 nt of the total 80 nt probe corresponding to exons 1, 2, or 3. The presence of 80 nt bands indicates that the 80 nt complementary sequences to the probes fully protected the probes. The presence of weak bands smaller than 40 nt in the whole-cell lysates, but not in the nuclear extracts, suggests that *Arf* mRNA is degraded in the cytoplasm. The probe sequences are listed in Table S7.

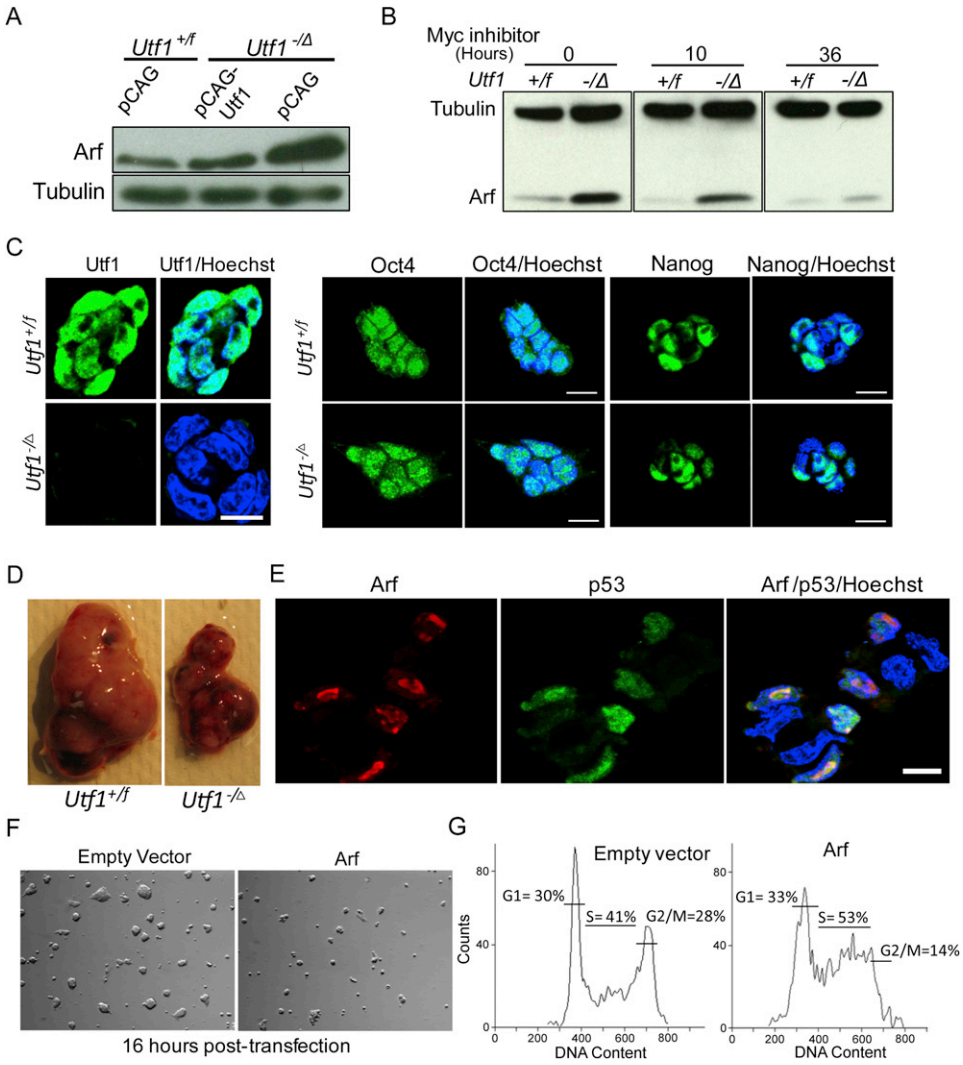

Figure S4. Utf1 Promotes Proliferation of ESCs through Inhibiting *Arf* Expression, Related to Figure 6

(A) Western blotting analyses of the expression level of Arf protein in *Utf1*$^{+/f}$ + pCAG, *Utf1*$^{-/\Delta}$ + pCAG-Utf1, and *Utf1*$^{-/\Delta}$ + pCAG ESCs. Restoring the expression of Utf1 in *Utf1*$^{-/\Delta}$ ESC (see the Utf1 western blotting in Figure S2E) decreased the expression of Arf protein. Loading controls, tubulin.

(B) *Utf1*$^{+/f}$ and *Utf1*$^{-/\Delta}$ ESCs were harvested at the indicated time points following Myc inhibitor (10058-F4) treatment. The Arf protein levels were determined by western blotting probing with the Utf1 antibody. Loading controls, tubulin.

(C) Immunofluorescence images revealing that *Utf1*$^{-/\Delta}$ ESCs do not express *Utf1* but maintain similar amounts of Oct4 and Nanog in their nuclei compared to *Utf1*$^{+/f}$ ESCs. Scale bar, 10 μm.

(D) *Utf1*$^{-/\Delta}$ ESCs produced smaller teratomas than those of *Utf1*$^{+/f}$ ESCs.

(E) Overexpression of Arf by transient transfection of the Arf-expressing plasmid in ESCs. The expression of Arf and p53 were detected by immunofluorescence staining. Scale bar, 10 μm

(F) Images of ESC colonies 16 hr post-transfection of either the empty vector or the Arf-expressing vector. The ESCs overexpressing Arf have smaller colonies compared to the controls.

(G) FACS analyses of the transfected ESCs. Overexpression of Arf caused a reduction of G2/M phase cells and an increase of S phase cells 16 hr post-transfection.