DOT1L interaction partner AF10 controls patterning of H3K79 methylation to maintain cell identity

Coral K. Wille[1], Edwin N. Neumann[1], Aniruddha J. Deshpande[2], Rupa Sridharan[1,3*]

1. Wisconsin Institute for Discovery, University of Wisconsin-Madison, Madison, WI 53715, USA.

2. Tumor Initiation and Maintenance Program, National Cancer Institute-Designated Cancer Center, Sanford Burnham Prebys Medical Discovery Institute, La Jolla, CA.

3. Department of Cell and Regenerative Biology, University of Wisconsin-Madison, Madison, WI 53715, USA.

*Corresponding author: Dr. Rupa Sridharan, Wisconsin Institute for Discovery, University of Wisconsin, 330 North Orchard Street, Room 2118, Madison, WI 53715, Email: rsridharan2@wisc.edu; phone: 608-316-4422

ABSTRACT

Histone-modifying enzymes function as part of protein complexes that alter the accessibility of chromatin to elicit differential gene expression patterns. DOT1L, the sole histone H3 lysine (K) 79 (H3K79) methyltransferase, is associated with proteins that recognize specific histone modifications and target DOT1L to particular chromatin contexts. Here we find that depletion of the DOT1L interacting protein AF10, which recognizes unmodified H3K27, mimics DOT1L catalytic inhibition to increase the efficiency of reprogramming somatic cells to induced pluripotent stem cells (iPSCs). AF10 deletion results in almost no steady state transcriptional changes yet is responsible for half of the DOT1L iPSC reprogramming phenotype. We find that the AF10-histone interaction domain is dispensable for increased iPSC generation. Instead, the AF10-DOT1L interaction domain that potentiates H3K79 di- and tri-methylation is a barrier to pluripotency acquisition. The loss of AF10 limits H3K79 methylation to a uniquely pluripotent pattern to facilitate iPSC formation. Taken together we reveal a mechanism for higher order histone methylation located at the gene body in reinforcing cell identity.

INTRODUCTION

Histone modifications shape the epigenome to be responsive to the transcriptional machinery that controls gene expression. The enzymes that write or erase histone modifications are found in complexes that can regulate their activity. While the function of histone modifications that determine chromatin compaction to allow transcription factor binding is intuitive, those that accompany transcriptional elongation have been obscure.

DOT1L is the exclusive methyltransferase of all degrees of histone 3 lysine (K) 79 methylation (H3K79me1, me2, and me3), a modification found on gene bodies (Vlaming and van Leeuwen, 2016). DOT1L is a distributive enzyme and longer residence time leads to greater degree of methylation (Frederiks et al., 2008). Thus the same loci tend to be enriched for H3K79me1, me2 and me3. While the individual functions of each modification remain unknown, all levels of H3K79me are enriched after transcription has begun (Steger et al., 2008). H3K79me2 in particular has been correlated with a rapid transcriptional rate (Duffy et al., 2018; Jonkers et al., 2014; Veloso et al., 2014).

DOT1L has a DNA-binding domain and an H2BK120 ubiquitin interaction domain that stabilizes DOT1L-histone binding and promotes methylation of H3K79 (Anderson et al., 2019; Worden et al., 2019). DOT1L also forms complexes with numerous associated proteins that have histone interaction domains and can boost higher degrees of H3K79 methylation, including: AF10 (*Mllt10*), AF17 (*Mllt6*), AF9 (*Mllt9*), and ENL (*Mllt1*) (Mohan et al., 2010a). AF10 and AF17 contain a PZP domain that interacts with unmodified H3K27 (Chen et al., 2015). AF10 localizes H3K79 methylation to gene bodies that lack H3K27me3 (Chen et al., 2015; Deshpande et al., 2014), and functions as a co-factor to increase DOT1L catalytic activity for both H3K79me2 and me3 (Deshpande et al., 2014). AF9 and ENL contain a YEATS domain that binds to H3K9 and H3K27 acylations (acetylation and crotonylation) to localize H3K79 methylation to transcriptional start sites (Li et al., 2014, 2016). Depletion of AF9 only decreases H3K79me3 without an effect on H3K79me2 (Mohan et al., 2010b). Thus DOT1L complex composition could greatly affect the recruitment of H3K79me to genomic loci depending on the surrounding chromatin environment. DOT1L targeting has also been found to be important in leukemias where AF10, AF9 and ENL are some of the most frequent fusion partners in rearrangements of the mixed-lineage leukemia (MLL) gene. In these MLL-rearranged (MLL-R) leukemias, DOT1L is recruited to MLL target loci with consequent aberrant H3K79me deposition (Bernt et al., 2011; Guenther et al., 2008; Krivtsov et al., 2008; Mohan et al., 2010a; Mueller et al., 2007; Okada et al., 2005).

DOT1L is essential for development as mouse knockouts are embryonic lethal around E11.5 (Feng et al., 2010; Jones et al., 2008; Liao and Szabó, 2020) and *in vitro* differentiation from embryonic stem cells (ESCs) is compromised upon DOT1L depletion (Barry et al., 2009; Cao et al., 2020). Inhibition or loss of DOT1L greatly enhances the reprogramming of induced pluripotent stem cells from somatic cells (Jackson et al., 2016; Onder et al., 2012; Wille and Sridharan, 2022a) and promotes formation of totipotent-like cells (Xu et al., 2022; Yang et al., 2022). Thus DOT1L activity is coincident with, and even promotes, a loss of potency (Wille and Sridharan, 2022b).

Here we mechanistically define the role of DOT1L-associated proteins and degrees of H3K79 methylation in the non-transformed and highly dynamic environment of somatic cell reprogramming to induced pluripotent stem cells (iPSCs). H3K79me2 is the most differentially

enriched histone modification in mouse embryonic fibroblasts (MEFs) compared to embryonic stem cells (ESCs), with a 7 fold greater abundance in MEFs (Sridharan et al., 2013). ESCs uniquely shuttle DOT1L to the cytoplasm (Michowski et al., 2020) suggesting that low H3K79me2 favors pluripotency. We find that DOT1L maintains cellular identity through AF10, but not AF9 or ENL. Deletion of AF10 increases reprogramming efficiency to almost half the levels seen by DOT1L catalytic inhibition. Structure-function analysis of AF10 reveals that the AF10-histone interaction is dispensable, but the AF10-DOT1L interaction is required for maintenance of somatic cell fate. AF10 depletion results in few transcriptional changes, the majority of which are not congruent with expression in ESCs. Further, despite AF10 deletion, highly expressed genes continue to retain H3K79methylation suggesting that DOT1L is associated with the transcriptional complex in non-transformed cells. Instead, deletion of AF10 results in the reduction of H3K79me2 at highly expressed housekeeping genes to resemble the pattern found in ESCs to facilitate RNA polymerase II (RNAPII) elongation and iPSC formation. Therefore, AF10 functions as a DOT1L enzymatic cofactor in the acquisition of pluripotency rather than a histone targeting module.

RESULTS

Expression of DOT1L associated factors is non-uniform during reprogramming

DOT1L is essential for development (Feng et al., 2010; Jones et al., 2008; Liao and Szabó, 2020) and a strong barrier to reprogramming to iPSCs (Jackson et al., 2016; Onder et al., 2012; Wille and Sridharan, 2022a). Therefore, pluripotency acquisition is an ideal platform to interrogate the mechanism of DOT1L and its associated proteins. To explore which of the DOT1L associated factors contribute to maintenance of somatic identity, we examined their expression (Wille and Sridharan, 2022a) in mouse somatic cells (MEFs), cells on day 2 and day 4 of reprogramming, and ESCs (Fig 1A). All assessed DOT1L associated factors were expressed throughout reprogramming, yet with different patterns (Fig 1A). Since reprogramming is a heterogenous process, single cell RNA-seq (scRNA-seq) (Tran et al., 2019) was analyzed to better understand the regulation of DOT1L associated factors during pluripotency acquisition. DOT1L was expressed at low levels in few cells regardless of pluripotency status (Fig 1B). AF9

was transiently expressed during mid-reprogramming. ENL was expressed across all cell types. AF10 was expressed in a higher percentage of pluripotent cells compared to MEFs and early reprogramming populations (Fig 1B). Co-expression analysis revealed that DOT1L was largely expressed with ENL in MEFs, yet was co-expressed predominately with AF9 during reprogramming (Fig S1A). DOT1L was co-expressed almost exclusively with AF10 in ESCs suggesting that reprogramming and pluripotency status may affect DOT1L complex composition (Fig S1A).

Given this variation throughout reprogramming, we tested the effect of DOT1L associated factors during the process. We initiated reprogramming in MEFs isolated from a reprogrammable mouse model in which the Yamanaka factors Oct4, Sox2, c-Myc and Klf4 are expressed from a single cassette under the control of a doxycycline inducible promoter. Each DOT1L associated factor was depleted using siRNA (Fig S1B) concurrent with the initiation of reprogramming. Successful iPSC generation was measured by the expression of the pluripotency protein NANOG. Interestingly, depletion of AF9 and ENL both decreased reprogramming efficiency (Fig 1C), perhaps due to their functions in transcription and unrelated to their DOT1L association (He et al., 2011). By contrast, like DOT1L, AF10 was a barrier for reprogramming (Fig 1C).

DOT1L maintains cellular identity through AF10

To better elucidate the role of AF10 in reprogramming, AF10 conditional mice (Deshpande et al., 2014) were bred with the inducible reprogrammable mouse line (Fig 1D). In somatic cells isolated from the resulting AF10 reprogrammable mouse line, both the transition to pluripotency and AF10 expression can be temporally controlled. We first compared the reprogramming efficiency of AF10 deletion (Δ) to the undeleted (fl) control. ΔAF10 increased NANOG+ colonies by greater than 3.5-fold in the presence of exogenous reprogramming factor expression (Fig 1E, Left). Deletion of AF10 also increased the formation of *bona fide* colonies that remained after the withdrawal of doxycycline-driven reprogramming factors, a more stringent test of pluripotency acquisition (Fig 1E, Right).

Since both AF10 and DOT1L can have pleiotropic functions, we next asked if AF10 functioned in the same pathway as DOT1L in pluripotency acquisition. We used a potent

DOT1L catalytic inhibitor SGC0946 (Yu et al., 2012), henceforth called DOT1Li, in combination with Cre-recombinase mediated deletion of AF10. Over 2-fold more colonies were obtained in DOT1Li as compared to ΔAF10 alone (Fig 1E, black bars). Importantly, AF10 deletion did not increase reprogramming beyond addition of inhibitor indicating that AF10 was in the same pathway as DOT1L (Fig 1E). To quantify the role of AF10 in DOT1L-mediated maintenance of cellular identity, a linear relationship was established between the percentage of AF10 deletion as measured by genotype PCR (Fig S1C) and percentage of total colonies obtained in ΔAF10, with DOT1Li treated cells set to 100% (Fig 1F). This estimate does not model the reduction in chromatin associated-DOT1L by chemical inhibitors (Nassa et al., 2019; Wu et al., 2021). Nonetheless, using this model, ΔAF10 is estimated to contribute about 40% of the DOT1Li phenotype in reprogramming (Fig 1F).

We further tested the relationship between DOT1Li and AF10 deletion using other commercially available inhibitors of DOT1L that have different efficacy: EPZ0044777, which is structurally similar to DOT1Li but 10-fold less potent, and EPZ-5676, a drug with similar potency to SGC0946 but with increased drug properties for use in clinic (Campbell et al., 2017; Daigle et al., 2013; Kaniskan et al., 2018; Yu et al., 2012). Irrespective of the drug concentration, deletion of AF10 always increased reprogramming efficiency as compared to the control (Fig S1D). The effect of AF10 deletion was most pronounced when DOT1L inhibition was the least potent. Upon AF10 deletion at the highest concentration of EPZ-5676, there was only a 1.5-fold increase of iPSC colonies as compared to the lower concentration of 2.5 μM which showed a 5-fold increase (Fig S1D). Thus, deletion of AF10 functions in the same pathway and is epistatic to DOT1L inhibition during reprogramming.

To identify if AF10 has a temporal effect during pluripotency acquisition, AF10 was deleted coincident with reprogramming factor induction (day 0), at the midpoint (day 2) and later (day 4) in a reprogramming timecourse, which assessed NANOG+ colony formation on day 6 (Fig S1E). Similar to DOT1L inhibition (Kim et al., 2021; Onder et al., 2012; Wille and Sridharan, 2022a), loss of AF10 most potently increased early and mid-reprogramming (Fig S1F). Taken together, AF10 collaborates with DOT1L to maintain somatic identity and its loss facilitates cell identity conversion.

AF10 deletion has few transcriptional effects

To determine how the contribution of AF10 deletion matched or enhanced that of DOT1Li, we performed RNA-seq experiments. Based on the timecourse of maximum iPSC generation (Fig S1F), we performed RNA-seq on day 4 of reprogramming of cells that were treated with control or Cre recombinase at the initiation of reprogramming factors, and compared it to DOT1Li. The ΔAF10 expression profile is most correlated with DOT1Li treated cells, and not ESCs as expected for a sample from the midpoint of the process (Fig 2A-B). To observe the trend in functional categories that changed dynamically, we generated a heatmap of differentially expressed (DE) genes from all conditions. Of the two clusters (C1 and C2) that were altered by DOT1Li to resemble ESCs (Fig 2A), C2 contained genes that functioned in negative regulation of cellular proliferation and epithelium development (Fig S2A). However, most of the transcriptional changes are not representative of ESC gene expression (Fig 2A). In reprogramming unique clusters C3-C6, the magnitude of gene expression changes in ΔAF10 is lower than that in DOT1Li. Cluster 7 contains protein modification process genes that are modestly upregulated upon DOT1Li and in ESCs, but are downregulated in AF10 deleted cells (Fig S2B). None of the clusters showed a major decrease in MEF specific genes. Thus in contrast to human reprogramming (Uğurlu-Çimen et al., 2021), suppression of somatic expression by AF10 deletion is not the primary mode of facilitating iPSC generation.

Although deletion of AF10 positions cells closer to DOT1Li treatment (Fig 2B), few genes are significantly DE measured as 2-fold or more fold-change in expression, with a posterior probability of being DE greater than 0.95 (Methods). At the starting point in MEFs, fewer than 15 genes were DE (Fig 2C). On day 4 of reprogramming in control treated cells, ΔAF10 resulted in the upregulation of 3 genes relative to control, and there was a single upregulated gene in ΔAF10 in cells treated with DOT1Li (Fig 2C) further demonstrating that deletion of AF10 does not add to DOT1L inhibition. Treatment of cells with DOT1Li had the greatest effect with many more genes upregulated (153) than downregulated (8) relative to control treatment (Fig 2D). DOT1Li treatment of ΔAF10 cells produced 33 upregulated genes and 4 downregulated (Fig 2D). Twenty-two genes were upregulated by DOT1Li in both fl and ΔAF10 cells (Fig 2E), and only *Hoxd12* and *Osr1* were commonly downregulated. Of note, we have demonstrated that reduction of *Hoxd12* did not contribute to the DOT1Li reprogramming

phenotype (Wille and Sridharan, 2022a). Overall, the loss of AF10 yielded almost no significant transcriptional changes at the steady state mRNA level although it lessened the DOT1Li transcriptional effect (Fig 3D). Therefore, we examined epigenetic means by which ΔAF10 could contribute to pluripotency acquisition.

AF10 targeting of DOT1L to unmodified H3K27 is dispensable for cellular identity

Since the steady state mRNA levels remain largely unchanged upon AF10 deletion (Fig 2), we wondered if there was a change in the epigenetic circuitry that does not produce immediate transcriptional alterations. H3K27me3 and H3K79me have been suggested to reciprocally oppose one another on the genome (Deshpande et al., 2014), and AF10 cannot bind to any modification at H3K27 *in vitro* (Chen et al., 2015). Therefore, to uncover the role of epigenetic feedback, we defined the mechanistic function of the AF10 protein in reprogramming (Fig 3A). AF10 is comprised of an N-terminal PHD finger-Zn knuckle-PHD finger (PZP) module that interacts with unmodified histone H3K27 (Chen et al., 2015) and a C-terminal octapeptide motif–leucine-zipper (OMLZ) domain that interacts with DOT1L (Fig 3B).

We generated one mutant of AF10 that cannot interact with histones (NLS-CTD) and two mutants that omitted either the OMLZ domain alone (delta-dOMLZ) or in combination with the C-terminal domain (NTD-OMLZ) to prevent interaction with DOT1L (Fig 3B) in a vector with a fluorescent reporter. AF10 mutants were individually transduced into AF10 conditional reprogrammable MEFs followed by deletion of endogenous AF10 or control treatment (Fig 3A). MEFs containing exogenous AF10 mutants were isolated with flow cytometry (Fig S3A) and reprogramming was initiated (Fig 3A). Equivalent levels of each mutant protein were observed in the sorted reprogramming cells (Fig 3C). Therefore, each reprogramming condition expressed either exogenous AF10 mutants alone (Δ) (Fig 3D, blue bars) or mutant proteins in addition to the endogenous AF10 (fl) (Fig S3B, blue bars). While deletion of AF10 enhanced reprogramming (Fig 3D, S3B dotted line), expression of exogenous full length AF10 reduced the number of NANOG+ colonies during reprogramming (Fig 3D) confirming that this deletion-rescue strategy can be used to identify the domains of AF10 responsible for safeguarding cell identity. The AF10 mutant that cannot interact with histones phenocopied full-length AF10 (Fig

3B, D, S3B) similar to a study in human reprogramming (Uğurlu-Çimen et al., 2021). These data imply that histone targeting may not be the primary mode of AF10 in inhibiting reprogramming.

To further investigate the crosstalk between H3K27me3 and H3K79me2, we directly compromised the enzyme complex that mediates H3K27me3, Polycomb Repressive Complex 2 (PRC2). PRC2 targets chromatin for H3K27 methylation by the enzymatic complex component EZH1/2, and promotes its spread through the EED subunit (Laugesen et al., 2019). Using A395, an EED small molecule inhibitor (He et al., 2017), we reasoned that depletion of H3K27me3 will create more potential binding sites for AF10 which could further decrease pluripotency acquisition. Although A395 resulted in loss of H3K27me3 (Fig 3E), the number of pluripotent colonies was unchanged in the presence and absence of AF10 compared to control treatment (Fig 3F). EED inhibition did not affect H3K79me2 (Fig 3E) demonstrating that there is not widespread antagonism between H3K79me2 and H3K27me3 in pluripotency acquisition. Therefore, the histone modification specific DOT1L targeting function of AF10 is not the barrier to somatic cell reprogramming.

AF10 enforces cellular identity through DOT1L interaction

We next turned our attention to the two AF10 mutants that cannot interact with DOT1L (Fig 3B). There was an increase in the number of iPSCs similar to AF10 deletion when cells were rescued with AF10 mutants lacking the OMLZ domain (Fig 3D). Mechanistically, the OMLZ deficient mutants failed to maintain H3K79me2, unlike full-length AF10 that restored almost wild-type levels (Fig 4A). Therefore, the AF10 function in boosting DOT1L catalytic activity for higher order methylation on H3K79 demonstrated in leukemia (Deshpande et al., 2014), likely opposes cell fate conversion during reprogramming.

To determine the impact of AF10 deletion on the degrees of H3K79 methylation, we performed western blots throughout reprogramming (Fig 4B). As compared to control cells, AF10 deletion resulted in a global reduction of both H3K79me2 and H3K79me3 to almost the same levels observed upon addition of the DOT1Li (Fig 4B-C). By contrast, AF10 deletion reduced H3K79me1 to about half of unperturbed levels (Fig 4B), with almost 5.7-fold more H3K79me1 retained as compared to DOT1Li treatment on day 4 of reprogramming (Fig 4C).

Thus, the contribution of AF10 deletion to increased reprogramming efficiency is likely to occur through effects on higher order H3K79 methylations.

We next examined retention of H3K79me2 vs. H3K79me1 at specific locations using quantitative ChIP-Seq with a spike in control on day 4 of reprogramming. This timepoint captures the difference in H3K79me1/2 (Fig 4B-C), but precedes the formation of fully reprogrammed colonies (data not shown). We focused on the H3K79me1 and me2 enrichment per gene body because almost all the H3K79me1 and H3K79me2 peaks were found at genic locations (Fig S4A). We quantitated the reduction of H3K79me1 and H3K79me2 in ΔAF10 or DOT1Li compared to their respective controls on day 4 of reprogramming (Log2 fold change (L2FC) (AF10Δ/f) or (D/C)). Since reprogramming is the conversion from the somatic to the pluripotent state, we used the difference in H3K79me1/me2 ESCs relative to the starting MEF population (E/M) as the gold standard to find trends that normally occur in pluripotency acquisition. ESCs had lower enrichment of H3K79me2 but a similar median amount of H3K79me1 as compared to MEFs. While DOT1Li severely reduced both H3K79me2 and H3K79me1 levels, deletion of AF10 led to the retention of H3K79me1 compared to DOT1Li (Fig 4D). Therefore, we hypothesized that deletion of AF10 may promote an ESC-like H3K79me1 retention to enhance reprogramming efficiency.

Deletion of AF10 causes an ESC-like pattern of H3K79me1 at housekeeping genes

We next examined the specific sites of AF10 dependent H3K79me1/me2 enrichment during reprogramming. The deletion of AF10 did not lead to novel sites of either H3K79me1 or me2 enrichment since ΔAF10 did not substantially alter the number of genes with significant H3K79me1/2 peaks (Fig 5A, D). Genes with H3K79me2 peaks in the control cells (fl) are also enriched for H3K79me1 (>89%) on day 4 of reprogramming (Fig 5A-B), and vice versa (Fig 5D-E). The vast majority also contain H3K79me2 in both ESCs and MEFs (Fig 5C) and enriched for housekeeping functions in metabolism and gene expression. Thus, the effect of AF10 deletion on promoting pluripotency occurs on genes that are commonly H3K79 methylated across cell types and not at specific lineage genes.

To elucidate how the loss of AF10 regulates H3K79 methylation at specific genes, we performed k-means clustering based on extent of H3K79me1/me2 reduction (Fig S4B).

Interestingly, genes separated into groups based on total H3K79me levels (Fig 5F-H) and expression (Fig S4B-C). The top cluster (C1) had the greatest total abundance of H3K79me1 and me2 on day 4 of reprogramming (Fig 5F-H), and consisted of genes that were very highly expressed in both ESC and MEFs (Fig S4B-C). The subsequent 3 clusters (C2-4) had decreasing amounts of total H3K79me2/me1 (Fig 5G-H) and expression in ESC and MEFs (Fig S4B-C).

We next compared the pattern of enrichment of H3K79me1 and H3K79me2 in the four clusters upon AF10 deletion to MEFs and ESCs. The global depletion for H3K79me2 (Fig 4D) in ESCs as compared to MEFs was also reflected in each of the four clusters. The deletion of AF10 caused an ~50% reduction in the H3K79me2 levels in all four clusters mimicking the decrease observed between ESCs and MEFs (Fig 5G). A more interesting pattern emerged in the distribution of H3K79me1 across the gene body. In the unperturbed condition, H3K79me1 is depleted at the transcriptional start site (TSS) compared to the mid-gene body (Fig 5H, S4D) suggesting almost complete conversion of H3K79me1 to H3K79me2 at the TSS. By contrast, the pattern of residual H3K79me1 in ΔAF10 shows an apex of enrichment at the TSS (Fig 5H, S4D). This pattern resembles the H3K79me1 found in ESCs which are much more TSS-enriched for this modification compared to MEFs (Fig 5H). Residual H3K79me1 did not alter expression (Fig S4E). Thus deletion of AF10 is likely to enhance reprogramming by making the epigenome more ESC-like at predominantly housekeeping genes (Fig 6).

Furthermore, given that AF10 deleted cells retain H3K79 methylation (Fig 5G-H) combined with our data that the histone targeting domain of AF10 is not important for enhancing pluripotency (Fig 3D) implicate that DOT1L is targeted independently of AF10 to target loci. DOT1L has been shown to interact with the phosphorylated C-terminal tail of RNAPII (Kim et al., 2012) and methylates genes concurrent with active transcription (Steger et al., 2008). Therefore we next examined the quantitative enrichment of RNA Polymerase II (RNAPII) in ESCs and MEFs. In both cell types, RNAPII was most enriched in C1 followed by the C2, C3, and C4, mirroring gene expression (Fig 5I). Of note, ESCs have much less H3K79me2 at the TSS compared to MEFs at all clusters which correlates with their relative reduction of RNAPII at the TSS (Fig 5G, I).

DISCUSSION

Taken together our results favor a model in which DOT1L is targeted to loci by RNAPII (Fig 6). The presence of AF10 increases higher order H3K79 methylation which is unfavorable for pluripotency. Even without the complete loss of H3K79 methylation with rapid pharmacological inhibition of DOT1L (Fig 4B), merely reducing H3K79me2 and changing the pattern of H3K79me1 with AF10 deletion can cause a remarkable 4-fold gain in reprogrammed colonies (Fig 1E) which accounts to 40% of the DOT1L inhibited phenotype (Fig 1F). These changes occur especially at highly expressed housekeeping genes (Fig 5, S4). Compared to somatic cells, ESCs have both a lower TSS-enrichment of RNAPII (Fig 5I) and H3K79me2 (Fig 5G) to mitigate negative transcriptional feedback (Fig 6) (Wille et al., 2022). ESCs have a much greater need for biosynthetic output than somatic cells (Bulut-Karslioglu et al., 2018; Percharde et al., 2017) and are likely to regulate transcription of housekeeping gene in a distinct manner (Wille et al., 2022). Taken together, our data support a negative transcriptional feedback model for AF10 mediated enhancement of DOT1L catalytic activity in the maintenance of cell identity (Fig 6). Such a function for AF0 seems to be evolutionarily conserved since a greater enrichment and degree of H3K79me has been shown to negatively impact transcriptional rate in *C. elegans* (Cecere et al., 2013). Specifically, the presence of the AF10 homolog ZFP-1 inhibits the transition of RNAPII pausing to productive elongation (Cecere et al., 2013).

Although AF10 accounts for about half of the DOT1L phenotype (Fig 1), deletion of AF10 results in few transcriptional changes (Fig 2). This result mirrors our recent report that showed DOT1Li yields few significant transcriptional alterations despite global eviction of H3K79me, and furthermore, none of differentially expressed genes are causal for enhanced reprogramming efficiency (Wille and Sridharan, 2022a). Previous studies have shown that abrogation of AF9 (Li et al., 2014) and AF10 (Deshpande et al., 2014) dramatically decreases H3K79me at target genes, yet the transcriptional effect is modest. Similar to our findings, only a subset of genes is transcriptionally altered in leukemia despite the presence of H3K79me2 at thousands of genes (Deshpande et al., 2014). It is interesting to note that AF10 deletion tempers the already modest DOT1Li effect on steady state mRNA (Fig 2), indicating that they are in the same pathway. Therefore, the majority of steady state transcriptional alterations do not contribute

to pluripotency acquisition, and instead AF10 functions as an epigenome modifier to mediate higher order H3K79 methylation to maintain cell identity.

We do not find any evidence that H3K79me2 opposes spreading of H3K27me3 in reprogramming to iPSCs (Fig 3E). Inhibition of EED and loss of H3K27me3 also did not increase H3K79me2 or inhibit reprogramming (Fig 3E-F). A recent study in leukemia found reduction of H3K27me3 at genes where H3K79me2 is removed with lower, but not higher, concentrations of EPZ-5676 (Richter et al., 2021). Perhaps this reduced concentration of EPZ-5676 mimics DOT1L attenuation via ΔAF10 in our study. DOT1L knockout in lymphocytes downregulated EZH2 and de-repressed genes normally shut off by H3K27me3 (Aslam et al., 2021; Kwesi-Maliepaard et al., 2020). Inhibition of DOT1L and EZH2 have both shown promise in pre-clinical MLL-R studies suggesting that H3K79me and H3K27me3 are functionally connected in leukemia (Lenard et al., 2020). Furthermore, depletion of AF9 and ENL, other DOT1L-histone targeting proteins, indicates that they do not function in the H3K79me2-mediated maintenance of cell identity (Fig 1C). We did not observe an increase in cell death upon AF9 or ENL depletion (data not shown) but further studies can determine whether the DOT1L- or transcriptional complex-interaction functions of AF9 and ENL contribute to their phenotype in reprogramming.

Taken together our studies have uncovered a mechanism by which subtle changes in the epigenetic distribution of a gene body associated histone modification can cause a profound change in cell identity.

METHODS

Mice and breeding. Mice were maintained in agreement with our UW-Madison IACUC approved protocol. Reprogramming mice are homozygous for the Oct4-2A-Klf4-2A-IRES-Sox2-2A-c-Myc (OKSM) transgene at the *Col1a1* locus and either heterozygous or homozygous for the reverse tetracycline transactivator (rtTA) allele at the *Rosa26* locus (Sridharan et al., 2013). Conditional AF10 knockout mice with LoxP sites before exon 17 and after exon 18 (OM-LZ domain) were a kind gift from Dr. Aniruddha Deshpande (Sanford Burnham Prebys Medical Discovery Institute) and Cre-mediated deletion results in loss of AF10 protein (Deshpande et al.,

2014). These mice were bred to the reprogrammable mouse strain to generate strains homozygous for flox (fl)-AF10, homozygous for OKSM, and either heterozygous or homozygous for rtTA. DR4 mice are genetically resistant to geneticin (G418), puromycin, hygromycin, and 6-thioguanine and acquired from Jackson labs.

Cells and culture. MEFs were isolated from embryos at day E13.5 of time mated mice as previously described (Tran et al., 2015). MEFs were maintained in MEF media (DMEM, 10% FBS, 1x non-essential amino acids, 1x glutamax, 1x penicillin/streptomycin, and 2-Mercaptoethanol - 4 µl/500ml). DR4 feeder MEFs were expanded for 3 passages and irradiated with 9000 rad. V6.5 ESCs were grown in ESC media (knock-out DMEM, 15% FBS, 1x non-essential amino acids, 1x glutamax, 1x penicillin/streptomycin, 2-Mercaptoethanol - 4 µl/525ml, and leukemia inhibitory factor) on plates coated with 0.1% porcine gelatin and DR4 feeder MEFs. Platinum-E (PLAT-E) are 293T cells engineered to express retroviral packaging components Gag, Pol, and Env, and are maintained in DMEM, 10% FBS, 1x penicillin/streptomycin, 10 µg/mL puromycin, and 1 µg/mL blasticidin.

Genotyping. Genotyping of mouse litters was performed on ear clippings at 21 days. Genotyping of reprogramming experiments was performed 2 days post Cre recombinase treatment. Tissue or cells were digested in 100 mM Tris-HCl pH 8.5, 5 mM EDTA, 0.2% SDS, and 200 mM NaCl with 20 µg Proteinase K for 2 hours at 65°C, followed by DNA precipitation with an equal volume of isopropanol and 20 µg glycogen. AF10 genotyping was performed with: 5’-CACAGCCTACTTCAAAGAAC-3’, 5’-TAGTCATGGGAATGGAGATG-3’, and 5’-ATTAGAGTCCATCCCACTTC-3’ primers using the BioReady Taq system (Bulldog Bio, BSA12L050).

Reprogramming. Cells were plated in MEF media on glass coverslips (Warner Instruments, 64-0714) that had been coated with 0.1% porcine gelatin, at a density of 30,000-50,000 cells per 12-well or 10,000-20,000 per 24 well. Reprogramming (OSKM expression) was initiated with 2 µg/mL doxycycline hyclate (Sigma-Aldrich, D9891). Reprogramming cells were treated with vehicle (DMSO) control, SGC0946 (ApexBio, A4167), EPZ-5676 (MedChem Express, HY-15593), or EPZ004777 (MedChem Express, HY-15227) beginning at day 0 of reprogramming, at 5, 2.5, 1, 0.2 µM, as indicated. SGC0946 was used at 5 µM in all additional reprogramming experiments. Cells were treated with 1 uM EED inhibitor A-395 hydrochloride (Millipore Sigma,

SML1923) beginning on day 1 of reprogramming. Gamma-irradiated feeder DR4 MEFs were added at 0.5x confluency by day 2 of reprogramming. Media was switched to ESC media by day 2 of reprogramming. Media with fresh doxycycline and drugs was replenished every two days of reprogramming. In the case of AF10 deletion experiments, Cre recombinase was added 1-2 days before doxycycline, or at the indicated timepoints (Methods, Viral transduction).

In the case of siRNA depletion experiments, cells were plated 1 day post doxycycline treatment and transfected with 20 nmol siRNA overnight. Media was switched to ESC media containing knock-out serum replacement (KSR) and feeder MEFs were added on day 2. Cells were then transfected with 40 nmol siRNA on days 3, 5, and 7 to account for increased cell number.

Stable colony formation was measured by removing media containing drugs and doxycycline, washing cells with ESC media, and replacing with ESC media free of doxycycline and drugs. Coverslips were fixed 2-4 days post doxycycline removal and assessed for sustained NANOG expression. Reprogramming timing was carefully adjusted for individual MEFs such that transgenes were activated long enough to produce *bona fide* colonies, but not so long that counting NANOG positive colonies in the presence of doxycycline was impossible due to overcrowding.

Reprogramming statistical analysis. Reprogrammed colonies were quantitated using fluoroscopy and were comprised of a grouping of at least 4 NANOG+ cells. Each individual biological replicate consisted of at least 2 technical replicate coverslips that were counted and averaged. The control condition, or as specified, was set 1 to compare the fold reprogramming efficiency among at least 3 biological replicates for which biological replicate number (n) is included in all legends. Each biological replicate was performed on a separate day and/or with a separate MEF isolation. Reprogramming efficiency was statistically analyzed with GraphPad Prism unpaired two-tailed t tests or as specified in figure legends.

siRNA transfection. siRNA was diluted in up to 50 μL serum-free DMEM per 12 well. 1 μL of DharmaFECT was incubated with 49 μL of serum-free DMEM per 12 well, for 5 minutes. The transfection reagent was combined with the diluted siRNA and incubated at room temperature for 20 minutes. Cells were transfected dropwise. The following siRNAs were used in

reprogramming experiments and purchased from Dharmacon (horizon): siDot1l (J-057964-12), siMllt10 (J-042898-09), siMllt1 (D-063284-01), siMllt3 (D-050750-03), non-targeting control (D-001810-01).

Viral transduction: Cre recombinase (Ad5-CMV-Cre) and empty control (Ad5-CMV-Empty) purified adenoviruses were purchased from the Vector Development Lab (Baylor College of Medicine). $2.4x10^9$ pfu of virus was resuspended in 5 mL of serum-free DMEM and added to 1-2 million MEFs in a 10-cm dish. Virus containing media was removed after 1 hour and replaced with MEF media. Reprogramming was initiated 2 days post-transduction.

AF10 domain mutant retroviruses were packaged in PLAT-E cells. Cells were treated with 10 μg/mL puromycin and 1 μg/mL blasticidin for 4 days to select for cells expressing packing components. Media was changed to antibiotic free DMEM+10% FBS on 15-cm dishes at 60% confluency 2 hours before transfection. 15 μg DNA was incubated in 600 μL Opti-MEM and 60 μL of 1 mg/mL linear PEI (Polyscience, 23966-2) for 15 minutes at room temperature before adding to cells dropwise. Media was changed to MEF media plus 20 mM HEPES 4 hours post-transfection. Virus-containing media was collected 48- and 72-hours post-transfection, filtered through a 0.45 um PVDF filter, and added directly to MEFs or frozen at -80°C for long term storage. MEFs were infected three times with the following strategy: 4 hours of infection (50% viral media, 50% fresh MEF media, 10 μg/mL Hexadimethrine Bromide), 4 hours of rest (MEF media), overnight infection, 4 hours of rest, 4 hours of infection.

Immunofluorescence. Cells were fixed for 10 minutes at room temperature in 4% paraformaldehyde/1xPBS. Cells were permeabilized in 0.5% Trition-X/1xPBS and washed in wash buffer (0.2% Tween-20/1xPBS), each for 10 minutes. Coverslips were blocked for 30 min in blocking buffer (5% normal goat serum, 0.2% Tween-20, 0.2% fish skin gelatin, 1x PBS) and probed for 1 hour in blocking buffer containing α-NANOG (1:100; Cosmo Bio USA, REC-RCAB001P or 1:1000; Cell Signaling Technology, 8822S) on parafilm. Coverslips were washed twice in wash buffer and then probed with α-Rabbit IgG-Dylight 488 (1:1000; ThermoFisher, 35552) for 1 hour. Coverslips were washed once, incubated with 4′,6-Diamidino-2-phenylindole dihydrochloride (1:10,000; Millipore Sigma, D8417) in wash buffer, and then washed a second time before mounting with Aqua-Poly/Mount (Fisher Scientific, NC9439247).

RT-qPCR. RNA was isolated from cells with Isolate II RNA Mini Kit according to the manufacturer's instructions (Bioline, BIO-52702). 1 µg of RNA was converted to cDNA using qScript (Quanta, 95047). Relative expression was measured with 20 ng of cDNA (calculated based on starting RNA quantity) in 10 µL reactions with SYBR Green (Bio-Rad, 1725124). The following primers were used:

	FORWARD 5'-3'	REVERSE 5'-3'
AF9	ACGGCAGTGACAGTGAAAG	TCAAGGATCTGGTTGTTGTTAGT
AF10	GCAGATCCCCATCAACAATC	TGAGCTACAGGCCCAGTTTT
DOT1L	GCGGCTGTGTGACAAATACA	CCATACACCTCAGGGGAGAA
ENL	CCCGAAAAGATGCTGAAGAA	AAAGTCGAAGGTGGTGTTGG
GAPDH	TTCACCACCATGGAGAAGGC	CCCTTTTGGCTCCACCCT

Immunoblot. Approximately 1 million whole cells were lysed in SUMO buffer - a fresh mixture of ¼ part I (5% SDS, 0.15 M Tris-HCl pH 6.8, 30% glycerol), ¾ part II (25 mM Tris-HCl pH 8.3, 50 mM NaCl, 0.5% NP-40, 0.5% deoxycholate, 0.1% SDS), and 1x cOmplete protease inhibitors (Roche, 4693132001) - and sonicated with a microtip at 20% amplitude for 5 seconds on ice. Protein concentration was measured with DC Protein Assay kit II (BioRad, 5000112) against a BSA standard curve and measured with Synergy H4 microplate reader. 20-25 µg of protein was loaded on H3K79me1/2 gels, 10 µg on H3K27me3 and H3K79me3, and 15 µg on FLAG-AF10 gels. Blots were transferred to nitrocellulose membranes, and blocked and probed in 5% milk, PBS, 0.1% Tween-20. The following antibodies were used: FLAG-M2 (1:5000; Sigma-Aldrich, F3165), H3K79me1 (1:1000; Diagenode, C15410082), H3K79me2 (1:1000; Active Motif, 39143), H3K79me3 (1:1000; Diagenode, C15410068, or 1:1000; Abcam, ab2621), H3K27me3 (1:1000; Cell Signaling, 9733S), and α-TUBULIN (1:3000; Cell Signaling, 3873). Immunoblot images were captured with ECL on ImageQuant LAS 4000. Tiff files were quantitated with Image Studio Lite V5.2 using the Add Rectangle function to calculate signal and background intensities.

Plasmids. AF10 domain mutants in the MSCV_IRES_YFP retrovirus plasmid were provided by Dr. Aniruddha Deshpande (Sanford Burnham Prebys Medical Discovery Institute). The full length (FL) vector contains the human AF10 cDNA isoform X6 (XM_024448184.1), NTD-

OMLZ (NO) encodes aa 1-724, ΔOMLZ (ΔO) encodes aa 1-730, 798-1084 with the OMLZ domain replaced by a methionine and histidine, and NLS-CTD encodes aa 207-1084. 3xFLAG was inserted N-terminal of AF10 in all constructs between the EcoRI and BamHI sites to facilitate immunoblot detection.

Flow cytometry sorting. AF10 domain mutant (MSCV_AF10_IRES_YFP) containing cells were FACS sorted before reprogramming. Cells were trypsinized, washed in PBS, resuspended in MEF media, and filtered through cell strainer capped tubes (Corning, 352235) before collecting the 530^{+} population on a BD FACS Aria II (UW Carbone Cancer Center, Grant #: 1S10RR025483-01). Uninfected MEFs were used to set the gates (Fig S3A). Flow cytometry images were generated in FlowJo.

RNA-Seq. RNA was isolated from 2 independent biological replicate reprogramming experiments starting from independent MEF isolations, using 1 mL of TRIzol. After 5 minutes of incubation, 200 μL of chloroform was added and mixed thoroughly. The aqueous phase was isolated after spinning the samples at 21xg for 15 min at 4°C. 0.53 volumes of ethanol was added, mixed, and spun through Qiagen RNeasy column at 21xg for 30 seconds. DNA was digested on the column with 10 μL RNAse free DNase mixed with 70 μL RDD buffer (Qiagen, 79254) for 30 min at room temperature. The RNA was then washed with 500 μL of buffer RW1 and RPE, each. RNA was eluted with 30 μL of water. 2 μg of RNA was used as starting material for library construction with TruSeq RNA Sample Prep Kit V2 (Illumina, RS-122-2002) following the manufacturer's instructions. Libraries were assessed with Qubit dsDNA HS assay and bioanalyzer3.0 before sequencing on HiSeq 2500 (UW-Madison Biotechnology) single end, 100 bp. Samples had 20-30 million reads.

RNA-Seq analysis. Sequencing quality was assessed with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads were processed before alignment with the FASTX-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). First low-quality reads were removed with the following parameters: fastq_quality_trimmer -Q33 -t 28 -l 20. Adapters were removed with the following parameters: fastx_clipper -Q33 -a adapter_sequence. The 5' end of the read (length determined by FastQC per base sequence content) was removed with the following parameters: fastx_trimmer -Q33 -f 11. Reads were then aligned to mm9 with an allowance of up to 2 mismatches using RSEM (Li and Dewey, 2011) with the following

parameters: rsem-calculate-expression -bowtie-m 200 --bowtie-n 2 --forward-prob 0.5 --seed-length 28. DE gene analysis was initiated by generating matrices of the "expected counts" from RSEM .gene.results files of the two replicate samples for DE comparison on R. These matrices were analyzed using EBSeq (Leng et al., 2013) with the follow parameters: MedianNorm, 5 iterations, target FDR = 0.05, and posterior probability of being DE (PPDE) > 0.95 on R. Genes were considered DE if the posterior FC (PostFC) was greater than or equal to 2 (upregulated) or less than or equal to 0.5 (downregulated).

Expression changes were assessed with TPM values collected in a matrix across all samples from RSEM .gene.results files using R. TPM values of DE genes from all sample comparisons were assessed with a Pearson correlation and displayed using the R package "pheatmap". To cluster DE genes, TPM values from replicates were averaged, 1 was added, and then they were calculated as Log2 fold change relative to the MEF control sample (fl). The Log2 FC values were clustered using Gene Cluster 3.0 (de Hoon et al., 2004) with the following parameters: Organize genes, 100 runs, k-means, Euclidean distance. Various numbers of clusters were tested, of which 7 were chosen based on downstream analysis. Clustered gene Log2FC values were displayed using Java TreeView (Saldanha, 2004). Clusters were analyzed using DAVID gene ontology (Huang et al., 2009) with the following parameters: Gene Ontology (GOTERM_BP_4, GOTERM_MF_4), Functional annotation clustering. Categories with a p-value less than 5.0E-3, in cluster(s) with the highest enrichment score were displayed. Motif analysis was performed with HOMER (Heinz et al., 2010) with the following parameters: findMotifs.pl -start -1000 -end 100 -len 6,10.

ChIP. Reprogrammable AF10fl MEFs that had been treated with CRE-recombinase or control virus (day -2) were plated on day 0 of reprogramming at 2 million cells per 15 cm dish in ESC media with feeder MEFs (Methods, Reprogramming). Media was refreshed on day 2, and cells were fixed on day 4 of reprogramming. Cells were fixed for 10 minutes with 1% formaldehyde, in solution, rotating at room temperature. Fixing was quenched with 0.14 M glycine for 5 minutes, and the cells were spun down at 300xg for 3 minutes. Cells were washed 3 times with PBS and flash frozen. 25-30 million cells were resuspended in 1 mL lysis buffer (1% SDS, 50 mM Tris-HCl pH 8, 20 mM EDTA, 1x cOmplete (Roche, 4693132001) protease inhibitor), and sonicated on a Covaris S220 Focused-ultrasonicator with the following parameters: 21 cycles of

45 seconds ON (peak 170, duty factor 5, cycles/burst 200), 45 seconds OFF (rest) in 6-8°C degassed water. 25 µL samples were taken before (0 cycles), at the mid-point, and at the end to test sonication fragmentation. Sonication test samples were resuspended in 3 volumes of water with 10 µg of RNaseA and incubated at 37°C for 30 minutes. 20 µg of Proteinase K was added, and crosslinks were reversed overnight at 65°C. DNA was cleaned and isolated with phenol-chloroform extraction using phase-lock tubes (VWR, 10847-802). 5 µg was run on a 1.5% agarose gel to ensure fragments of 200-400 bp were generated. Sonicated DNA was quantified with Qubit dsDNA HS (ThermoFisher Scientific, Q32854), aliquoted, and stored at -80°C.

The immunoprecipitation was set up with 7.1 ug of reprogramming day 4 chromatin plus 1:53 (133 ng) of human 293T spike in chromatin (Orlando et al., 2014), generated as above. Chromatin was diluted 1:10 with dilution buffer (16.7 mM Tris-HCl pH 8, 0.01% SDS, 1.1% Trition-X, 1.2 mM EDTA, and 167 mM NaCl), and immunoprecipitated with 5 ug H3K79me1 (Abcam, ab2886, Lot GR274715-4) or H3K79me2 (Active Motif, 39143, Lot 16918002) antibody for 16 hours, rotating at 4°C. 25 µL of Protein A (ThermoFisher Scientific, 10002D) and 25 µL of Protein G (ThermoFisher Scientific, 10004D) Dynabeads that were washed once in PBS, 0.02% Tween-20 and once in dilution buffer, were added. Chromatin was incubated, rotating, with beads for 2 hours at 4°C. Beads were washed twice, rotating at 4°C for 5 minutes, with 1 mL of each of the following buffers: Low salt (50 mM HEPES pH 7.9, 0.1% SDS, 1% Triton X-100, 0.1% Deoxycholate, 1 mM EDTA pH 8.0, 140 mM NaCl), High salt (50 mM HEPES pH 7.9, 0.1% SDS, 1% Triton X-100, 0.1% Deoxycholate, 1 mM EDTA pH 8.0, 500 mM NaCl), LiCl (20 mM Tris-HCl pH 8, 0.5% NP-40, 0.5% Deoxycholate, 1 mM EDTA pH 8.0, 250 mM LiCl), and TE (10 mM Tris-HCl pH 8, 1 mM EDTA pH 8) using a magnetic rack. Beads were incubated with 250 µL elution buffer (50 mM Tris-HCl pH 8, 1 mM EDTA pH 8, 1% SDS) for 10 minutes at 65°C, shaking. 200 µL of TE with 0.67% SDS was added, and incubated shaking at 65°C for 10 minutes. 10 µg RNase A was added and incubated at 37°C for 30 minutes. 40 µg of Proteinase K was added, and incubated for 12-16 hours at 65°C to reverse crosslinks. DNA was cleaned with phenol-chloroform phase separation with phase lock tubes (VWR, 10847-802). DNA was precipitated with 1 volume isopropanol, 1/10th volume of 3M NaActate, and 20 µg glycogen, and washed with 70% ethanol. DNA was dissolved in 30 µL ultrapure water and concentrated to 5 µL using a SpeedVac for library preparation.

The entire ChIP sample was used as starting material for library construction with Ovation Ultralow System V2 using reagents at 0.5X and adapters diluted 1/5 in water, following the manufacturer's instructions (NuGEN, 0344). The libraries were separated on an 1.5% agarose gel, cut from 200-400 bp for an additional size selection step. The gel fragment was purified using the MinElute Gel Extraction Kit (Qiagen, 28604). Libraries were assessed with Qubit quantitation and Bioanalyzer3.0. Single-end 50 bp sequencing was performed on a HiSeq 4000 for a minimum of 20 million reads (NUSeq Core Northerwestern Medicine, Center for Genomic Medicine). See Table 1 for sequencing depth, antibody, and normalization details.

ChIP analysis. Reads were aligned to the mm9 mouse genome with Bowtie2 (Langmead and Salzberg, 2012). Samtools-1.2 (Li et al., 2009) was used to convert .sam files into .bam files using "view" and, and then sort .bam files with the "sort" command. Reads that did not align to the mouse genome were isolated from sorted bam files with Sam-1.2 "view -f4", converted to fastq files with "bam2fq", and then aligned to hg19 genome as above. The number of reads that uniquely aligned to the human genome were used for spike in normalization (see Table 1). Peaks were called with MACS2 (Feng et al., 2012; Zhang et al., 2008) with the following parameters: H3K29me2 --broad -p 0.0001 and H3K79me1 –broad -p 0.001. Peaks were annotated with HOMER (Heinz et al., 2010) to the mm9 genome using annotatePeaks.pl. High confidence genes were identified by isolating genes with H3K79me peaks within gene bodies (5'UTR, exon, intron, 3'UTR, and/or TTS) for each ChIP-Seq replicate, and then taking the consensus using Venny2.0 (https://bioinfogp.cnb.csic.es/tools/venny/).

The clustered heatmap (Fig 5D) was created by calculating H3K79me enrichment per gene body with BEDTools2.0 (Quinlan and Hall, 2010) multibamCov. Reads were normalized per scaling factor (see Table 1), and then per kilobase of gene length. The two ChIP-seq replicate normalized quantitations were averaged per gene and the Log2 fold change (L2FC) was calculated per indicated conditions. All box and whiskers plots were generated in Graphpad Prism as: the box limits - 25th to 75th percentiles, the center line - median, and the whiskers - minimum and maximum values. The heatmap (Fig 4D) included genes with an H3K79me1 peak in fl control on day 4 of reprogramming since this captured more genes than the H3K79me2 peak list. Clustering was performed with H3K79me1 and me2 L2FC in ΔAF10/fl and DOT1Li/Control using Gene Cluster 3.0 (de Hoon et al., 2004) with the following parameters:

Organize genes, 100 runs, k-means, Euclidean distance. Various numbers of clusters were tested, of which 4 were chosen based on downstream analysis. The expression measured as transcripts per million (displayed on the heatmap as Log2 (TPM divided by 25)) and L2FC H3K79me1 and me2 in ESCs relative to MEFs were matched back to the clustered list. Clusters were displayed with Java TreeView (Saldanha, 2004). Clusters were analyzed using DAVID gene ontology (Huang et al., 2009) with the following parameters: Gene Ontology (GOTERM_BP_4, GOTERM_MF_4), Functional annotation clustering. Categories with a p-value less than 5.0E-3, in cluster(s) with the highest enrichment score were considered.

Deeptools (Ramírez et al., 2016) was used to generate metaplots. First, a bigwig file normalized per human spike-in, that was the average of the two sorted .bam ChIP-Seq replicates was created with the following parameters: bamCompare --scaleFactors (see Table 1) --operation mean --binSize 10 --ignoreForNormalization chrX. A matrix of reads per cluster was generated with the Deeptools computeMatrix command. Profiles were created with: scale-regions -b 3000 -a 3000 --regionBodyLength 5000, or as listed. Deeptools was used to compare ChIP-Seq replicates using a Pearson correlation of normalized bigwig files. A matrix of enrichment per 1 kb bins over the genome was made with: multiBigwigSummary -bs 1000, which was displayed with: plotCorrelation --corMethod pearson --skipZeros --removeOutliers. IGV was used to visualize ChIP-Seq tracks of normalized bigwig files.

Published datasets. Expression data and H3K79me1/2 ChIP-seq data in ΔAF10 and fl cells can be obtained at NCBI GEO SuperSeries GSE214143. ESC RNA-Seq data and bulk expression of DOT1L and associated factors was assessed from data previously generated (Wille and Sridharan, 2022a) and can be obtained at NCBI GEO GSE160580. Single cell expression and co-expression of DOT1L and associated factors were analyzed as described (Tran et al., 2019) and can be obtained at NCBI GEO GSE108222. MEF, ESC, and day 4 of reprogramming in Control or DOT1Li H3K79me1/2 and RNAPII ChIP-seq datasets were processed as described in Methods, ChIP-seq analysis, and can be obtained at NCBI GEO GSE190391.

TABLES

Table 1. ChIP-Seq sample information

Antibody	Sample	Total Reads	Mouse Reads	Human Reads	Scaling Factor
H3K79 me1	fl/fl-1	24762326	23842691	349996	1
	fl/fl-2	32365629	24367241	654398	1
	ΔAF10-1	28590630	27108358	621731	0.56
	ΔAF10-2	36902973	29245857	999328	0.65
H3K79 me2	fl/fl-1	21430788	16177048	104951	1
	fl/fl-2	22934694	19042337	152464	0.69
	ΔAF10-1	32619715	29228855	652995	0.16
	ΔAF10-2	26589996	23180361	404187	0.26

ACKNOWLEDGEMENTS

This work was supported by a UW-Madison Stem Cell and Regenerative Medicine Center postdoctoral award to CKW, and NIH-NIGMS R01GM113033 to the RS lab. We thank Stefan Pietrzak for scRNA-seq analysis and the R.S. lab members for critical reading of the manuscript.

AUTHOR CONTRIBUTIONS

C.K.W. and E.N.N. performed experiments, C.K.W. completed bioinformatic analysis, C.K.W. and R.S. wrote the manuscript, R.S. acquired funding, A.D. provided reagents and mice, R.S. and C.K.W. conceived and directed the project, and all authors edited and reviewed the manuscript.

COMPETING INTERESTS

The authors declare that they have no competing interests.

REFERENCES

Anderson, C.J., Baird, M.R., Hsu, A., Barbour, E.H., Koyama, Y., Borgnia, M.J., and McGinty, R.K. (2019). Structural Basis for Recognition of Ubiquitylated Nucleosome by Dot1L Methyltransferase. Cell Rep *26*, 1681-1690.e5. https://doi.org/10.1016/j.celrep.2019.01.058.

Aslam, M.A., Alemdehy, M.F., Kwesi-Maliepaard, E.M., Muhaimin, F.I., Caganova, M., Pardieck, I.N., van den Brand, T., van Welsem, T., de Rink, I., Song, J.-Y., et al. (2021). Histone methyltransferase DOT1L controls state-specific identity during B cell differentiation. EMBO Rep *22*, e51184. https://doi.org/10.15252/embr.202051184.

Barry, E.R., Krueger, W., Jakuba, C.M., Veilleux, E., Ambrosi, D.J., Nelson, C.E., and Rasmussen, T.P. (2009). ES cell cycle progression and differentiation require the action of the histone methyltransferase Dot1L. Stem Cells *27*, 1538–1547. https://doi.org/10.1002/stem.86.

Bernt, K.M., Zhu, N., Sinha, A.U., Vempati, S., Faber, J., Krivtsov, A.V., Feng, Z., Punt, N., Daigle, A., Bullinger, L., et al. (2011). MLL-rearranged leukemia is dependent on aberrant H3K79 methylation by DOT1L. Cancer Cell *20*, 66–78. https://doi.org/10.1016/j.ccr.2011.06.010.

Bulut-Karslioglu, A., Macrae, T.A., Oses-Prieto, J.A., Covarrubias, S., Percharde, M., Ku, G., Diaz, A., McManus, M.T., Burlingame, A.L., and Ramalho-Santos, M. (2018). The Transcriptionally Permissive Chromatin State of Embryonic Stem Cells Is Acutely Tuned to Translational Output. Cell Stem Cell *22*, 369-383.e8. https://doi.org/10.1016/j.stem.2018.02.004.

Campbell, C.T., Haladyna, J.N., Drubin, D.A., Thomson, T.M., Maria, M.J., Yamauchi, T., Waters, N.J., Olhava, E.J., Pollock, R.M., Smith, J.J., et al. (2017). Mechanisms of Pinometostat (EPZ-5676) Treatment-Emergent Resistance in MLL-Rearranged Leukemia. Mol Cancer Ther *16*, 1669–1679. https://doi.org/10.1158/1535-7163.MCT-16-0693.

Cao, K., Ugarenko, M., Ozark, P.A., Wang, J., Marshall, S.A., Rendleman, E.J., Liang, K., Wang, L., Zou, L., Smith, E.R., et al. (2020). DOT1L-controlled cell-fate determination and transcription elongation are independent of H3K79 methylation. Proc Natl Acad Sci U S A *117*, 27365–27373. https://doi.org/10.1073/pnas.2001075117.

Cecere, G., Hoersch, S., Jensen, M.B., Dixit, S., and Grishok, A. (2013). The ZFP-1(AF10)/DOT-1 complex opposes H2B ubiquitination to reduce Pol II transcription. Mol Cell *50*, 894–907. https://doi.org/10.1016/j.molcel.2013.06.002.

Chen, S., Yang, Z., Wilkinson, A.W., Deshpande, A.J., Sidoli, S., Krajewski, K., Strahl, B.D., Garcia, B.A., Armstrong, S.A., Patel, D.J., et al. (2015). The PZP Domain of AF10 Senses Unmodified H3K27 to Regulate DOT1L-Mediated Methylation of H3K79. Mol. Cell *60*, 319–327. https://doi.org/10.1016/j.molcel.2015.08.019.

Daigle, S.R., Olhava, E.J., Therkelsen, C.A., Basavapathruni, A., Jin, L., Boriack-Sjodin, P.A., Allain, C.J., Klaus, C.R., Raimondi, A., Scott, M.P., et al. (2013). Potent inhibition of DOT1L as treatment of MLL-fusion leukemia. Blood *122*, 1017–1025. https://doi.org/10.1182/blood-2013-04-497644.

Deshpande, A.J., Deshpande, A., Sinha, A.U., Chen, L., Chang, J., Cihan, A., Fazio, M., Chen, C., Zhu, N., Koche, R., et al. (2014). AF10 Regulates Progressive H3K79 Methylation and HOX Gene Expression in Diverse AML Subtypes. Cancer Cell *26*, 896–908. https://doi.org/10.1016/j.ccell.2014.10.009.

Duffy, E.E., Canzio, D., Maniatis, T., and Simon, M.D. (2018). Solid phase chemistry to covalently and reversibly capture thiolated RNA. Nucleic Acids Res *46*, 6996–7005. https://doi.org/10.1093/nar/gky556.

Feng, J., Liu, T., Qin, B., Zhang, Y., and Liu, X.S. (2012). Identifying ChIP-seq enrichment using MACS. Nat Protoc *7*, 1728–1740. https://doi.org/10.1038/nprot.2012.101.

Feng, Y., Yang, Y., Ortega, M.M., Copeland, J.N., Zhang, M., Jacob, J.B., Fields, T.A., Vivian, J.L., and Fields, P.E. (2010). Early mammalian erythropoiesis requires the Dot1L methyltransferase. Blood *116*, 4483–4491. https://doi.org/10.1182/blood-2010-03-276501.

Frederiks, F., Tzouros, M., Oudgenoeg, G., van Welsem, T., Fornerod, M., Krijgsveld, J., and van Leeuwen, F. (2008). Nonprocessive methylation by Dot1 leads to functional redundancy of histone H3K79 methylation states. Nat Struct Mol Biol *15*, 550–557. https://doi.org/10.1038/nsmb.1432.

Guenther, M.G., Lawton, L.N., Rozovskaia, T., Frampton, G.M., Levine, S.S., Volkert, T.L., Croce, C.M., Nakamura, T., Canaani, E., and Young, R.A. (2008). Aberrant chromatin at genes encoding stem cell regulators in human mixed-lineage leukemia. Genes Dev *22*, 3403–3408. https://doi.org/10.1101/gad.1741408.

He, N., Chan, C.K., Sobhian, B., Chou, S., Xue, Y., Liu, M., Alber, T., Benkirane, M., and Zhou, Q. (2011). Human Polymerase-Associated Factor complex (PAFc) connects the Super Elongation Complex (SEC) to RNA polymerase II on chromatin. PNAS *108*, E636–E645. https://doi.org/10.1073/pnas.1107107108.

He, Y., Selvaraju, S., Curtin, M.L., Jakob, C.G., Zhu, H., Comess, K.M., Shaw, B., The, J., Lima-Fernandes, E., Szewczyk, M.M., et al. (2017). The EED protein-protein interaction inhibitor A-395 inactivates the PRC2 complex. Nat Chem Biol *13*, 389–395. https://doi.org/10.1038/nchembio.2306.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589. https://doi.org/10.1016/j.molcel.2010.05.004.

de Hoon, M.J.L., Imoto, S., Nolan, J., and Miyano, S. (2004). Open source clustering software. Bioinformatics *20*, 1453–1454. https://doi.org/10.1093/bioinformatics/bth078.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc *4*, 44–57. https://doi.org/10.1038/nprot.2008.211.

Jackson, S.A., Olufs, Z.P.G., Tran, K.A., Zaidan, N.Z., and Sridharan, R. (2016). Alternative Routes to Induced Pluripotent Stem Cells Revealed by Reprogramming of the Neural Lineage. Stem Cell Reports *6*, 302–311. https://doi.org/10.1016/j.stemcr.2016.01.009.

Jones, B., Su, H., Bhat, A., Lei, H., Bajko, J., Hevi, S., Baltus, G.A., Kadam, S., Zhai, H., Valdez, R., et al. (2008). The histone H3K79 methyltransferase Dot1L is essential for mammalian development and heterochromatin structure. PLoS Genet. *4*, e1000190. https://doi.org/10.1371/journal.pgen.1000190.

Jonkers, I., Kwak, H., and Lis, J.T. (2014). Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. Elife *3*, e02407. https://doi.org/10.7554/eLife.02407.

Kaniskan, H.Ü., Martini, M.L., and Jin, J. (2018). Inhibitors of Protein Methyltransferases and Demethylases. Chem Rev *118*, 989–1068. https://doi.org/10.1021/acs.chemrev.6b00801.

Kim, K.-P., Choi, J., Yoon, J., Bruder, J.M., Shin, B., Kim, J., Arauzo-Bravo, M.J., Han, D., Wu, G., Han, D.W., et al. (2021). Permissive epigenomes endow reprogramming competence to transcriptional regulators. Nat Chem Biol *17*, 47–56. https://doi.org/10.1038/s41589-020-0618-6.

Kim, S.-K., Jung, I., Lee, H., Kang, K., Kim, M., Jeong, K., Kwon, C.S., Han, Y.-M., Kim, Y.S., Kim, D., et al. (2012). Human histone H3K79 methyltransferase DOT1L protein binds actively transcribing RNA polymerase II to regulate gene expression. J Biol Chem *287*, 39698–39709. https://doi.org/10.1074/jbc.M112.384057.

Krivtsov, A.V., Feng, Z., Lemieux, M.E., Faber, J., Vempati, S., Sinha, A.U., Xia, X., Jesneck, J., Bracken, A.P., Silverman, L.B., et al. (2008). H3K79 methylation profiles define murine and human MLL-AF4 leukemias. Cancer Cell *14*, 355–368. https://doi.org/10.1016/j.ccr.2008.10.001.

Kwesi-Maliepaard, E.M., Aslam, M.A., Alemdehy, M.F., van den Brand, T., McLean, C., Vlaming, H., van Welsem, T., Korthout, T., Lancini, C., Hendriks, S., et al. (2020). The histone methyltransferase DOT1L prevents antigen-independent differentiation and safeguards epigenetic identity of CD8+ T cells. Proc Natl Acad Sci U S A *117*, 20706–20716. https://doi.org/10.1073/pnas.1920372117.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357–359. https://doi.org/10.1038/nmeth.1923.

Laugesen, A., Højfeldt, J.W., and Helin, K. (2019). Molecular Mechanisms Directing PRC2 Recruitment and H3K27 Methylation. Mol Cell *74*, 8–18. https://doi.org/10.1016/j.molcel.2019.03.011.

Lenard, A., Xie, H.M., Pastuer, T., Shank, T., Libbrecht, C., Kingsley, M., Riedel, S.S., Yuan, Z.-F., Zhu, N., Neff, T., et al. (2020). Epigenetic regulation of protein translation in KMT2A-rearranged AML. Exp Hematol *85*, 57–69. https://doi.org/10.1016/j.exphem.2020.04.007.

Leng, N., Dawson, J.A., Thomson, J.A., Ruotti, V., Rissman, A.I., Smits, B.M.G., Haag, J.D., Gould, M.N., Stewart, R.M., and Kendziorski, C. (2013). EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. Bioinformatics *29*, 1035–1043. https://doi.org/10.1093/bioinformatics/btt087.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12*, 323. https://doi.org/10.1186/1471-2105-12-323.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

Li, Y., Wen, H., Xi, Y., Tanaka, K., Wang, H., Peng, D., Ren, Y., Jin, Q., Dent, S.Y.R., Li, W., et al. (2014). AF9 YEATS domain links histone acetylation to DOT1L-mediated H3K79 methylation. Cell *159*, 558–571. https://doi.org/10.1016/j.cell.2014.09.049.

Li, Y., Sabari, B.R., Panchenko, T., Wen, H., Zhao, D., Guan, H., Wan, L., Huang, H., Tang, Z., Zhao, Y., et al. (2016). Molecular Coupling of Histone Crotonylation and Active Transcription by AF9 YEATS Domain. Mol. Cell *62*, 181–193. https://doi.org/10.1016/j.molcel.2016.03.028.

Liao, J., and Szabó, P.E. (2020). Maternal DOT1L is dispensable for mouse development. Sci Rep *10*, 20636. https://doi.org/10.1038/s41598-020-77545-6.

Michowski, W., Chick, J.M., Chu, C., Kolodziejczyk, A., Wang, Y., Suski, J.M., Abraham, B., Anders, L., Day, D., Dunkl, L.M., et al. (2020). Cdk1 Controls Global Epigenetic Landscape in Embryonic Stem Cells. Mol Cell *78*, 459-476.e13. https://doi.org/10.1016/j.molcel.2020.03.010.

Mohan, M., Lin, C., Guest, E., and Shilatifard, A. (2010a). Licensed to elongate: a molecular mechanism for MLL-based leukaemogenesis. Nat. Rev. Cancer *10*, 721–728. https://doi.org/10.1038/nrc2915.

Mohan, M., Herz, H.-M., Takahashi, Y.-H., Lin, C., Lai, K.C., Zhang, Y., Washburn, M.P., Florens, L., and Shilatifard, A. (2010b). Linking H3K79 trimethylation to Wnt signaling through a novel Dot1-containing complex (DotCom). Genes Dev. *24*, 574–589. https://doi.org/10.1101/gad.1898410.

Mueller, D., Bach, C., Zeisig, D., Garcia-Cuellar, M.-P., Monroe, S., Sreekumar, A., Zhou, R., Nesvizhskii, A., Chinnaiyan, A., Hess, J.L., et al. (2007). A role for the MLL fusion partner ENL in transcriptional elongation and chromatin modification. Blood *110*, 4445–4454. https://doi.org/10.1182/blood-2007-05-090514.

Nassa, G., Salvati, A., Tarallo, R., Gigantino, V., Alexandrova, E., Memoli, D., Sellitto, A., Rizzo, F., Malanga, D., Mirante, T., et al. (2019). Inhibition of histone methyltransferase DOT1L silences ERα gene and blocks proliferation of antiestrogen-resistant breast cancer cells. Sci Adv *5*, eaav5590. https://doi.org/10.1126/sciadv.aav5590.

Okada, Y., Feng, Q., Lin, Y., Jiang, Q., Li, Y., Coffield, V.M., Su, L., Xu, G., and Zhang, Y. (2005). hDOT1L Links Histone Methylation to Leukemogenesis. Cell *121*, 167–178. https://doi.org/10.1016/j.cell.2005.02.020.

Onder, T.T., Kara, N., Cherry, A., Sinha, A.U., Zhu, N., Bernt, K.M., Cahan, P., Marcarci, B.O., Unternaehrer, J., Gupta, P.B., et al. (2012). Chromatin-modifying enzymes as modulators of reprogramming. Nature *483*, 598–602. https://doi.org/10.1038/nature10953.

Orlando, D.A., Chen, M.W., Brown, V.E., Solanki, S., Choi, Y.J., Olson, E.R., Fritz, C.C., Bradner, J.E., and Guenther, M.G. (2014). Quantitative ChIP-Seq normalization reveals global modulation of the epigenome. Cell Rep *9*, 1163–1170. https://doi.org/10.1016/j.celrep.2014.10.018.

Percharde, M., Bulut-Karslioglu, A., and Ramalho-Santos, M. (2017). Hypertranscription in Development, Stem Cells, and Regeneration. Dev Cell *40*, 9–21. https://doi.org/10.1016/j.devcel.2016.11.010.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res *44*, W160-165. https://doi.org/10.1093/nar/gkw257.

Richter, W.F., Shah, R.N., and Ruthenburg, A.J. (2021). Non-canonical H3K79me2-dependent pathways promote the survival of MLL-rearranged leukemia. Elife *10*, e64960. https://doi.org/10.7554/eLife.64960.

Saldanha, A.J. (2004). Java Treeview--extensible visualization of microarray data. Bioinformatics *20*, 3246–3248. https://doi.org/10.1093/bioinformatics/bth349.

Sridharan, R., Gonzales-Cope, M., Chronis, C., Bonora, G., McKee, R., Huang, C., Patel, S., Lopez, D., Mishra, N., Pellegrini, M., et al. (2013). Proteomic and genomic approaches reveal critical functions of H3K9 methylation and heterochromatin protein-1γ in reprogramming to pluripotency. Nat. Cell Biol. *15*, 872–882. https://doi.org/10.1038/ncb2768.

Steger, D.J., Lefterova, M.I., Ying, L., Stonestrom, A.J., Schupp, M., Zhuo, D., Vakoc, A.L., Kim, J.-E., Chen, J., Lazar, M.A., et al. (2008). DOT1L/KMT4 recruitment and H3K79 methylation are ubiquitously coupled with gene transcription in mammalian cells. Mol Cell Biol *28*, 2825–2839. https://doi.org/10.1128/MCB.02076-07.

Tran, K.A., Jackson, S.A., Olufs, Z.P.G., Zaidan, N.Z., Leng, N., Kendziorski, C., Roy, S., and Sridharan, R. (2015). Collaborative rewiring of the pluripotency network by chromatin and signalling modulating pathways. Nat Commun *6*, 6188. https://doi.org/10.1038/ncomms7188.

Tran, K.A., Pietrzak, S.J., Zaidan, N.Z., Siahpirani, A.F., McCalla, S.G., Zhou, A.S., Iyer, G., Roy, S., and Sridharan, R. (2019). Defining Reprogramming Checkpoints from Single-Cell

Analyses of Induced Pluripotency. Cell Rep *27*, 1726-1741.e5. https://doi.org/10.1016/j.celrep.2019.04.056.

Uğurlu-Çimen, D., Odluyurt, D., Sevinç, K., Özkan-Küçük, N.E., Özçimen, B., Demirtaş, D., Enüstün, E., Aztekin, C., Philpott, M., Oppermann, U., et al. (2021). AF10 (MLLT10) prevents somatic cell reprogramming through regulation of DOT1L-mediated H3K79 methylation. Epigenetics Chromatin *14*, 32. https://doi.org/10.1186/s13072-021-00406-7.

Veloso, A., Kirkconnell, K.S., Magnuson, B., Biewen, B., Paulsen, M.T., Wilson, T.E., and Ljungman, M. (2014). Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications. Genome Res. *24*, 896–905. https://doi.org/10.1101/gr.171405.113.

Vlaming, H., and van Leeuwen, F. (2016). The upstreams and downstreams of H3K79 methylation by DOT1L. Chromosoma *125*, 593–605. https://doi.org/10.1007/s00412-015-0570-5.

Wille, C.K., and Sridharan, R. (2022a). DOT1L inhibition enhances pluripotency beyond acquisition of epithelial identity and without immediate suppression of the somatic transcriptome. Stem Cell Reports *17*, 384–396. https://doi.org/10.1016/j.stemcr.2021.12.004.

Wille, C.K., and Sridharan, R. (2022b). Connecting the DOTs on Cell Identity. Frontiers in Cell and Developmental Biology *10*. .

Wille, C.K., Zhang, X., Haws, S., Denu, J., and Sridharan, R. (2022). Antagonistic H3K79me-H3K9ac crosstalk determines elongation at housekeeping genes to promote pluripotency. 2022.09.26.509534. https://doi.org/10.1101/2022.09.26.509534.

Worden, E.J., Hoffmann, N.A., Hicks, C.W., and Wolberger, C. (2019). Mechanism of Cross-talk between H2B Ubiquitination and H3 Methylation by Dot1L. Cell *176*, 1490-1501.e12. https://doi.org/10.1016/j.cell.2019.02.002.

Wu, A., Zhi, J., Tian, T., Cihan, A., Cevher, M.A., Liu, Z., David, Y., Muir, T.W., Roeder, R.G., and Yu, M. (2021). DOT1L complex regulates transcriptional initiation in human erythroleukemic cells. Proc Natl Acad Sci U S A *118*, e2106148118. https://doi.org/10.1073/pnas.2106148118.

Xu, Y., Zhao, J., Ren, Y., Wang, X., Lyu, Y., Xie, B., Sun, Y., Yuan, X., Liu, H., Yang, W., et al. (2022). Derivation of totipotent-like stem cells with blastocyst-like structure forming potential. Cell Res https://doi.org/10.1038/s41422-022-00668-0.

Yang, M., Yu, H., Yu, X., Liang, S., Hu, Y., Luo, Y., Izsvák, Z., Sun, C., and Wang, J. (2022). Chemical-induced chromatin remodeling reprograms mouse ESCs to totipotent-like stem cells. Cell Stem Cell S1934-5909(22)00010-8. https://doi.org/10.1016/j.stem.2022.01.010.

Yu, W., Chory, E.J., Wernimont, A.K., Tempel, W., Scopton, A., Federation, A., Marineau, J.J., Qi, J., Barsyte-Lovejoy, D., Yi, J., et al. (2012). Catalytic site remodelling of the DOT1L

methyltransferase by selective inhibitors. Nat Commun *3*, 1288. https://doi.org/10.1038/ncomms2304.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137. https://doi.org/10.1186/gb-2008-9-9-r137.

Figure 1

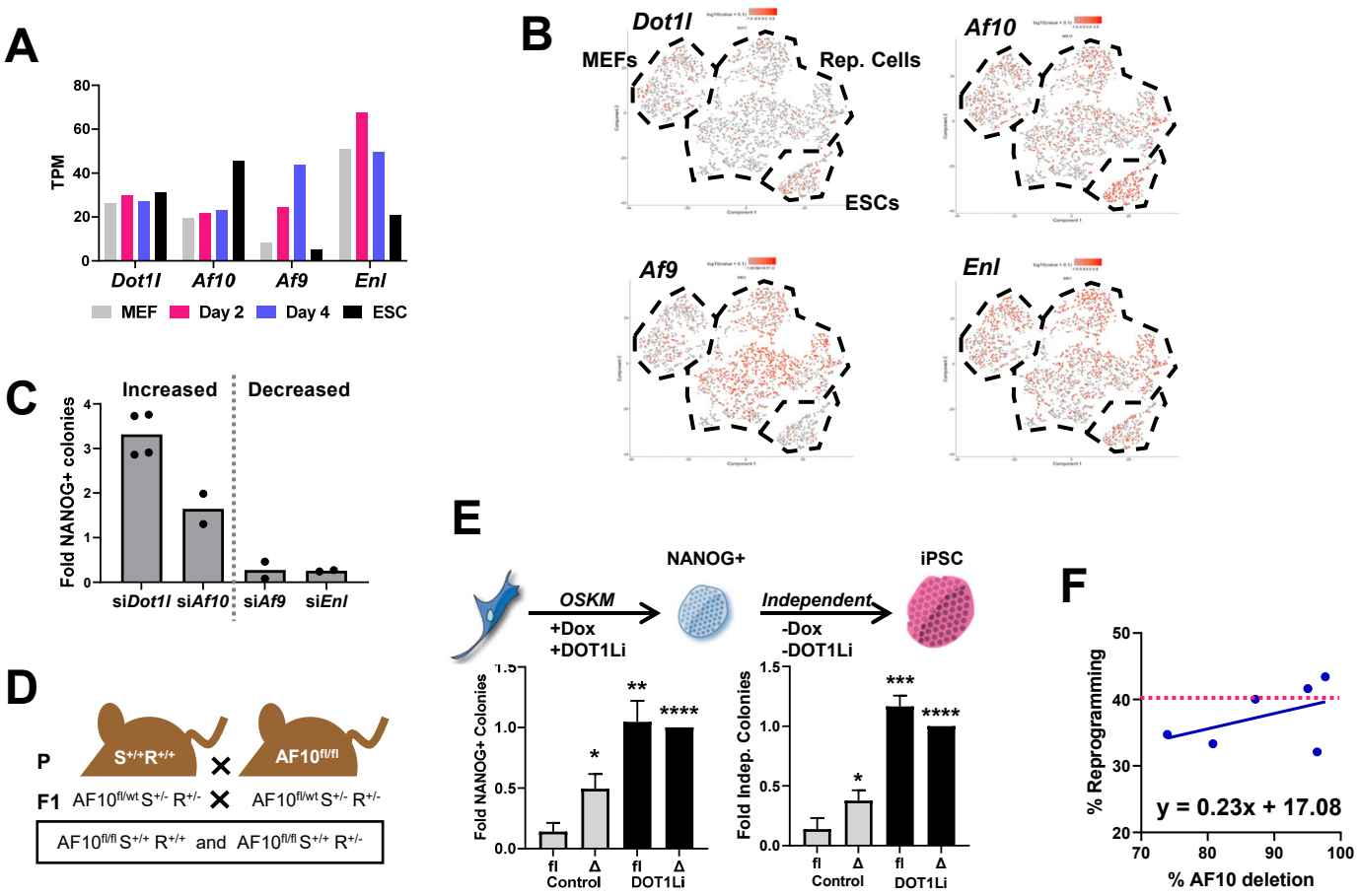

Figure 1. DOT1L maintains cellular identity through AF10.

A. Expression of DOT1L associated factors determined by bulk RNA-Seq (Wille and Sridharan, 2022) in MEFs, day 2 of reprogramming, day 4 of reprogramming, and ESCs measured as transcripts per million (TPM).

B. scRNA-seq (Tran et al., 2019) t-SNE of MEFs, Reprogramming (Rep.) cells on days 3, 6, 9, and 12, and ESCs. Every cell is represented by a dot and cells expressing the indicated factor are indicated in red.

C. Reprogramming upon siRNA depletion of indicated genes. Colonies obtained in non-targeting control set to 1. Data are the mean of at least 2 biological replicates, depicted as dots.

D. Breeding scheme to generate conditional AF10 mice with inducible reprogramming factors. AF10ff.SS.RR and AF10ff.SS.Rw were used for all studies.

E. Top: Reprogramming scheme. Bottom: OSKM transgene dependent (Left) and independent (Right) NANOG+ reprogrammed colonies of cells infected with empty adenovirus (fl) or adenovirus-Cre (ΔAF10), with and without DOT1Li treatment. Colonies obtained in ΔAF10 treated with DOT1Li set to 1. Data are the mean + S.D. (n = 3). ****P<0.0001, ***P<0.001, **P<0.01, and *P<0.05 by unpaired two-tailed t-test.

F. Linear relationship of the percentage of AF10 genomic deletion measured by genotyping PCR, to the percentage of reprogrammed colonies in AF10 deleted conditions (colonies formed in DOT1Li set to 100%).

Figure 2

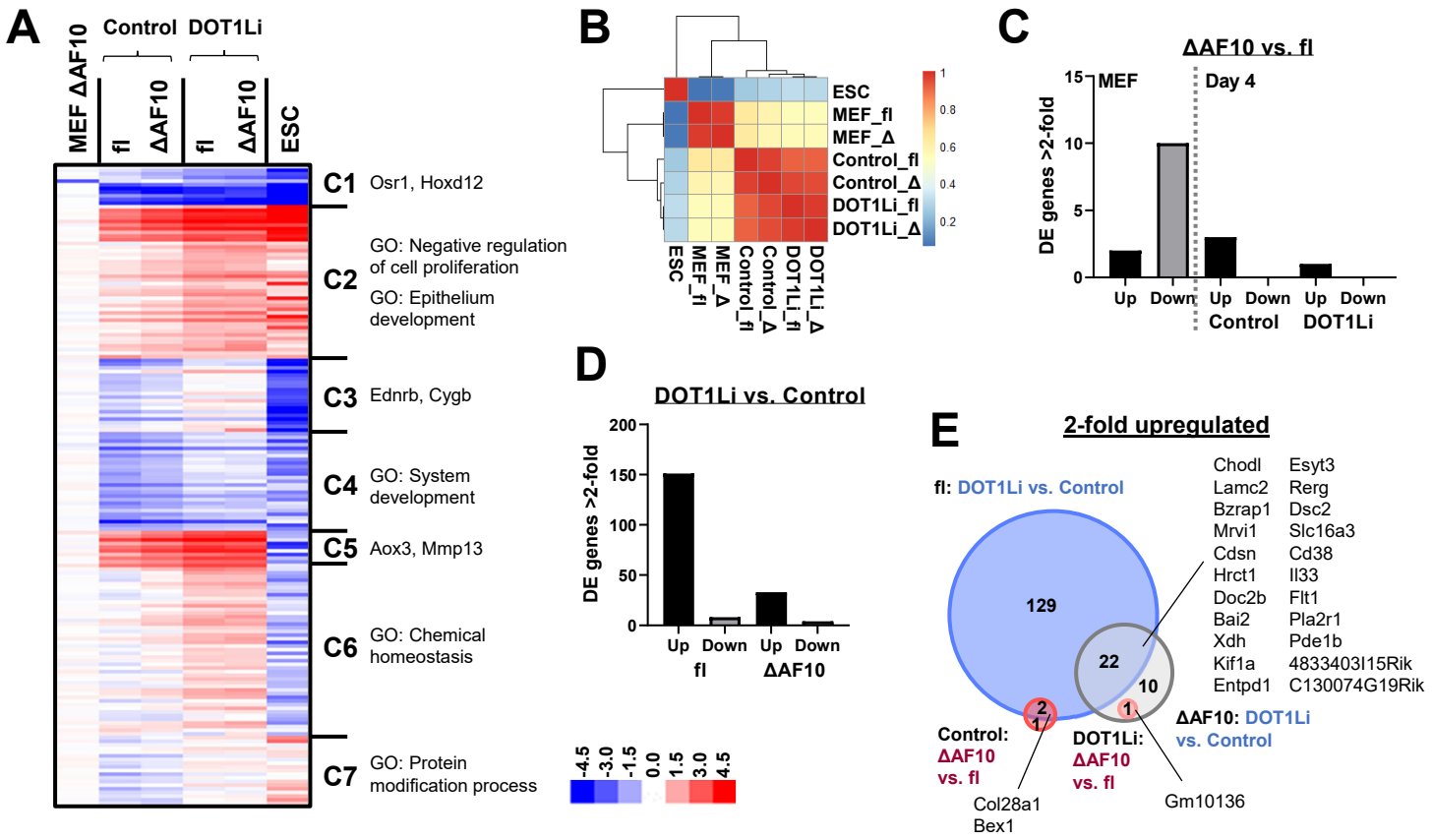

Figure 2. AF10 deletion has few transcriptional effects.

A. k-means clustered heatmap of Log2 fold change in transcripts per million (TPM) relative to fl MEFs of DE genes from all conditions.

B. Heatmap of Pearson correlations with hierarchical clustering of TPM values of DE genes from all conditions.

C-D. Number of DE genes (2-fold or more with a posterior probability of differential expression 0.95 or greater) upregulated (black) or downregulated (gray) in: C. ΔAF10 vs. fl cells: in MEFs (left) or day 4 of reprogramming (right, as indicated); D. DOT1Li vs. Control on day 4 of reprogramming of: fl cells (left) and ΔAF10 cells (right).

E. Overlap of upregulated genes on day 4 of reprogramming in: DOT1Li vs. Control in fl (blue), DOT1Li vs. Control in ΔAF10 (gray), ΔAF10 vs. fl in control (red), and ΔAF10 vs. fl in DOT1Li (pink) conditions.

Figure 3

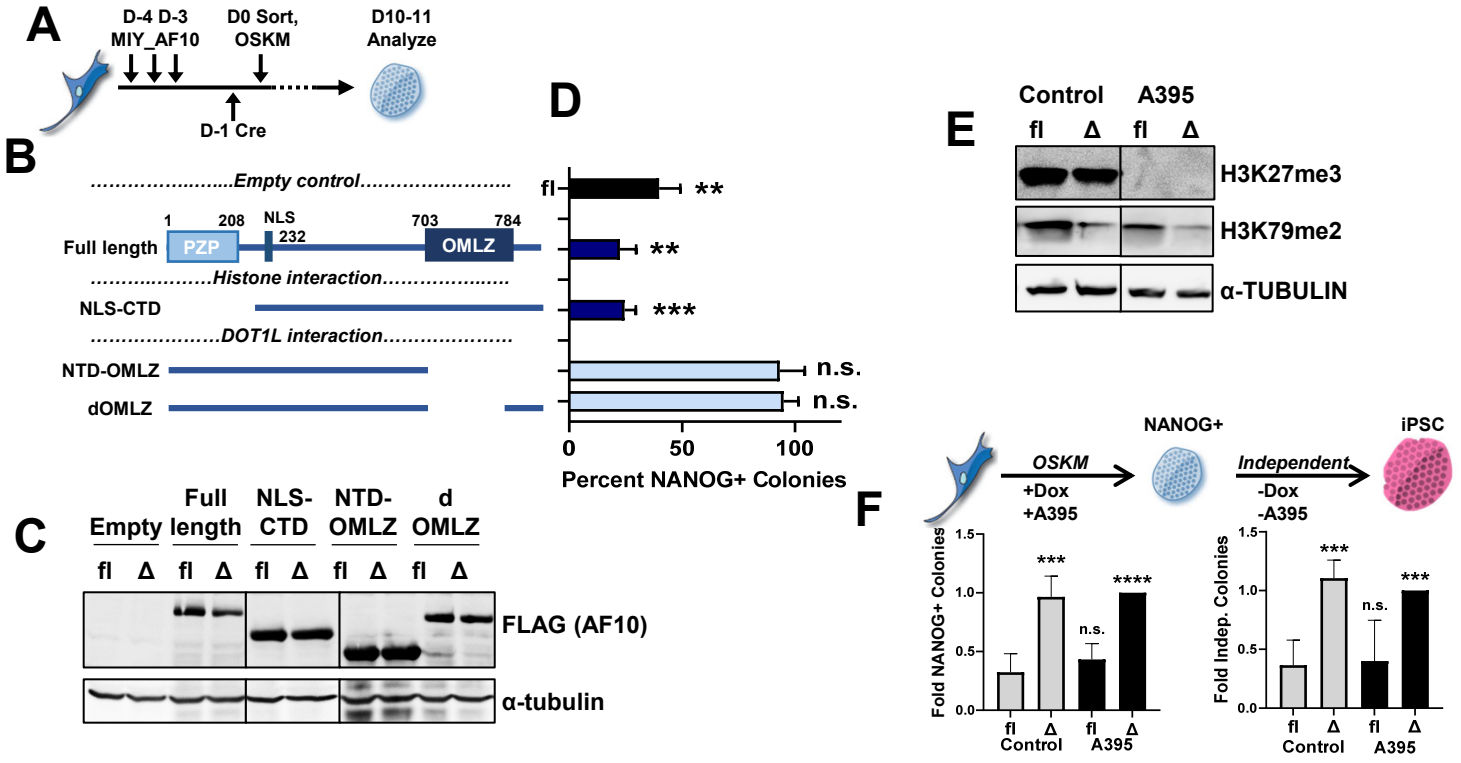

Figure 3. AF10 targeting of DOT1L to unmodified H3K27 is dispensable for cellular identity.

A. MEFs were infected 3 times with retroviruses containing AF10 domain mutants over 3-4 days prior to reprogramming. Endogenous AF10 was deleted with Cre recombinase or cells treated with control adenovirus on day -1. The next day, cells were sorted by flow cytometry, plated, and reprogramming (OSKM) was initiated.

B. AF10 contains two major functional domains: 1) PHD finger-Zn knuckle-PHD finger (PZP) module and 2) octapeptide motif-leucine zipper (OMLZ). Domain mutants that ablate histone interaction (NLS-CTD) or DOT1L interaction (NTD-OMLZ and dOMLZ) are depicted.

C. Immunoblot of AF10 domain mutant expression on day 5 of reprogramming in control (fl) and AF10 deleted (Δ) cells.

D. Reprogramming efficiency of AF10-Cre deleted cells (ΔAF10), with the indicated exogenous AF10 domain mutants (Fig 4B). Colonies in ΔAF10 infected with empty vector control set to 100%. Reprogramming of wild type cells (fl) infected with empty vector shown as a control (black bar). Data are mean + S.E.M. (n = 4). ***P<0.001, **P<0.01, and not significant (n.s.) P>0.05 by paired t-test against ΔAF10 with empty vector control.

E. Immunoblot of H3K79me2, H3K27me3, and α-TUBULIN in cells treated with control or A395 (EED inhibitor) on day 5 of reprogramming.

F. Transgene dependent (Left) and independent (Right) reprogramming of fl or ΔAF10 cells, treated with A395 or control. Data are the mean + S.D. (n = 5). ****P<0.0001, ***P<0.001, and not significant (n.s.) P>0.05 by unpaired two-tailed t-test.

Figure 4

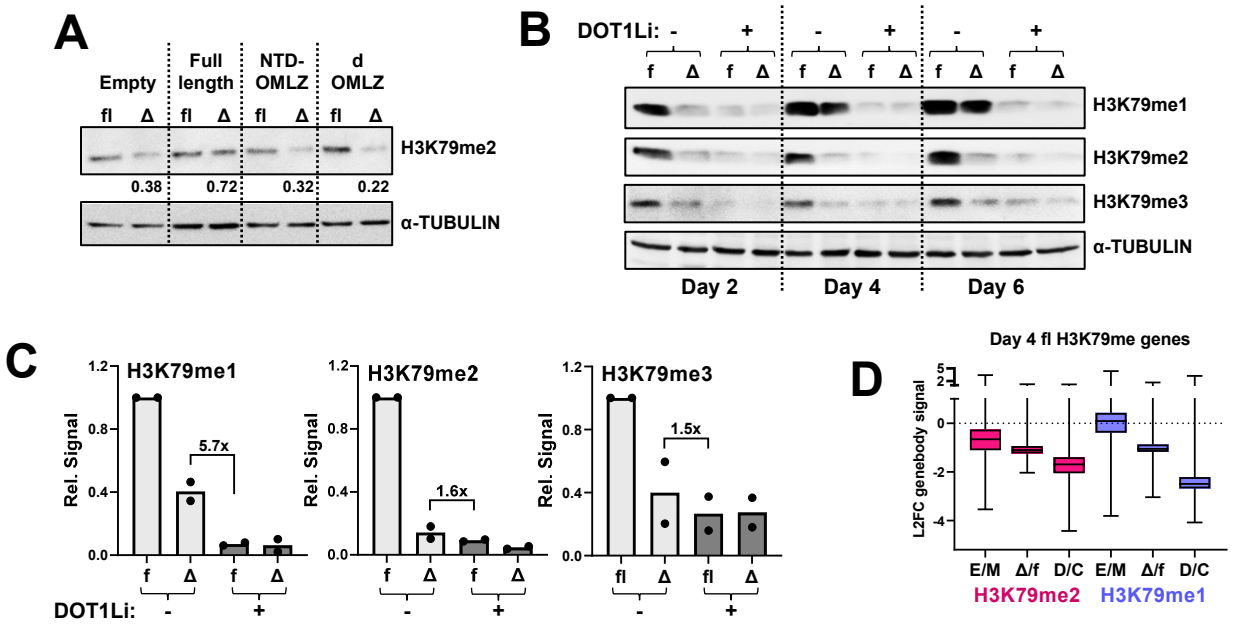

Figure 4. AF10 enforces cellular identity through DOT1L interaction.

A. Immunoblot of H3K79me2 in cells with endogenous (fl) or deleted (Δ) AF10, exogenously expressing AF10 domain mutants, on day 5 of reprogramming. H3K79me2 signal normalized to α-TUBULIN was calculated in ΔAF10 relative to matched fl conditions (Bottom).

B. Immunoblot of H3K79me1, me2, and me3 in AF10 wt (fl) or Cre deleted (ΔAF10) conditions during reprogramming.

C. Relative H3K79me1/2/3 immunoblot signal on day 4 of reprogramming (n = 2). Control treated fl condition set to 1.

D. Log2 fold change (L2FC) of normalized reads per gene body that have H3K79me2 (pink) or H3K79me1 (blue) peaks in day 4 AF10 wt (fl) in: ESCs relative to MEFs (E/M), deleted AF10 relative to fl control (Δ/f), and DOT1Li relative to control treatment (D/C). Normalized reads are the mean (n = 2).

Figure 5

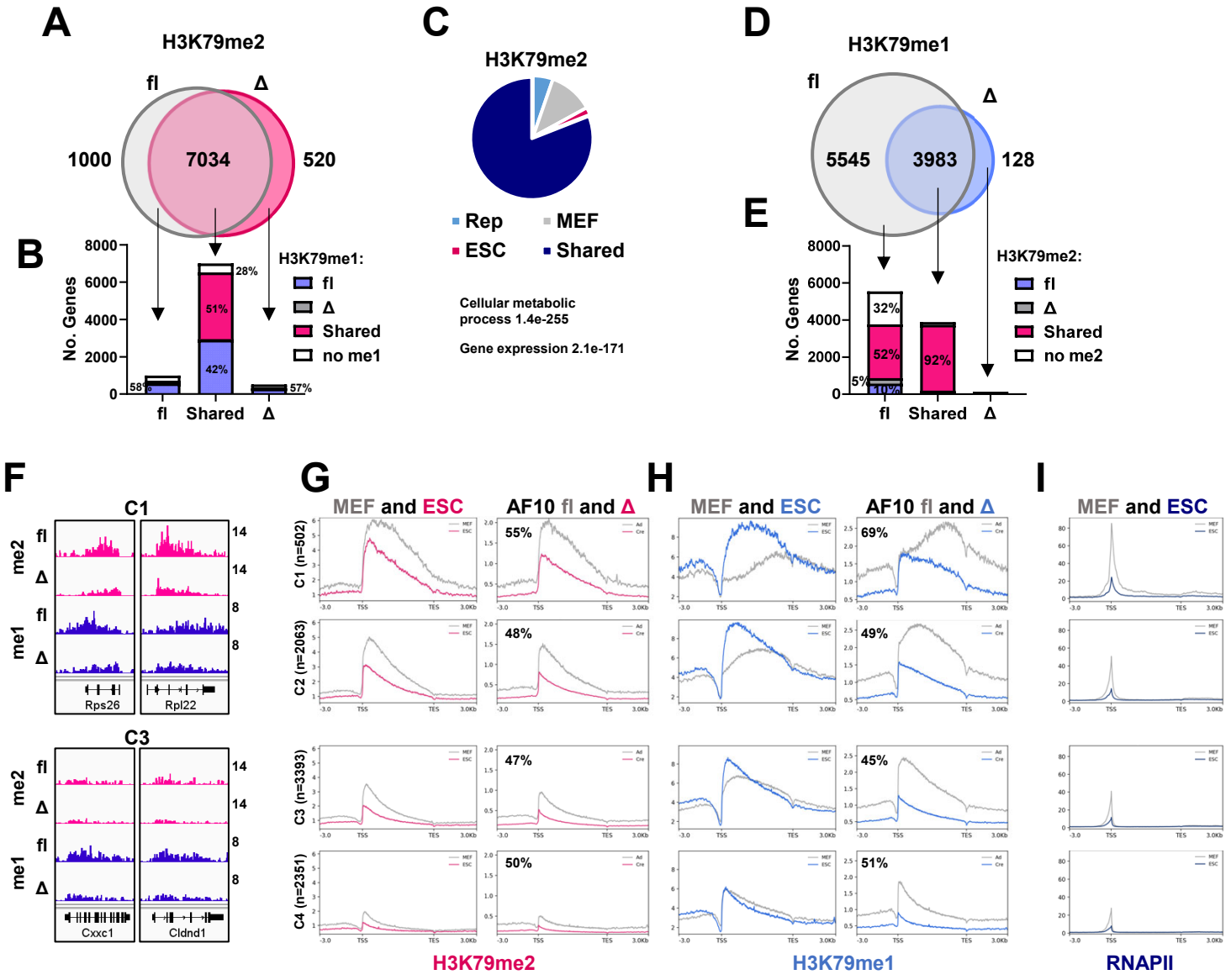

Figure 5. Deletion of AF10 causes an ESC-like pattern of H3K79me1 at housekeeping genes.

A. Overlap of genes with a significant H3K79me2 peak in AF10 wt (fl) or Cre deleted (Δ) on day 4 of reprogramming.

B. H3K79me1 peak status of genes with H3K79me2 peaks on day 4 of reprogramming.

C. H3K79me2 MEF and ESC peak status of genes with H3K79me2 peaks in AF10 wt (fl) (Fig 5A, n=8034). Below: Gene Ontology (GO) of genes with an H3K79me2 peak shared in ESCs and MEFs.

D. Overlap of gene with a significant H3K79me1 peak in AF10 wt (fl) or Cre deleted (Δ) on day 4 of reprogramming.

E. H3K79me2 peak status of genes with H3K79me1 peaks on day 4 of reprogramming.

F-I. Clusters from the heatmap of genes with an H3K79me1 peak in AF10 wt (fl) on day 4 of reprogramming (Fig S4B) were analyzed for:

F. Example H3K79me1/2 IGV tracks.

G. H3K79me2 metaplots in MEFs/ESCs and AF10 wt (fl)/Cre deleted (Δ).

H. H3K79me1 metaplots in MEFs/ESCs and AF10 wt (fl)/Cre deleted (Δ).

I. RNAPII metaplots in MEFs and ESCs.

Figure 6

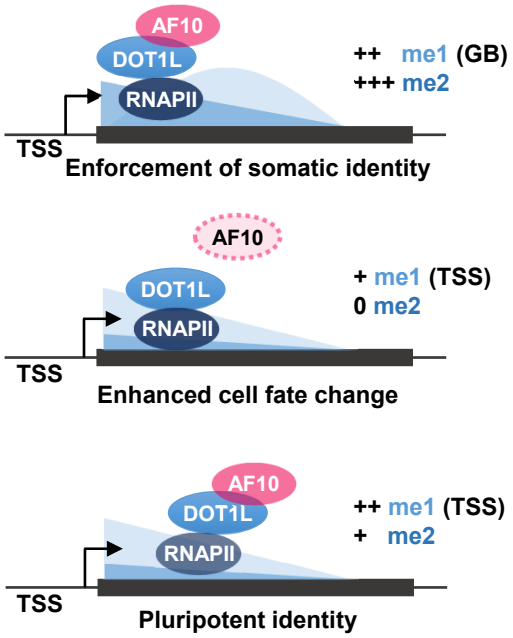

Figure 6. AF10 maintains cell identity by enforcing a somatic H3K79-epigenome

In somatic cells, the presence of AF10 and DOT1L is recruited to highly expressed housekeeping genes by RNAPII yielding the maximum amount of H3K79me1 and me2. At the TSS, H3K79me1 is fully converted to H3K79me2/3 leading to a peak in the mid-gene body (GB). In ESCs, the relatively reduced TSS-associated RNAPII leads to lower H3K79me2 and a TSS apex of H3K79me1. ESCs have increased transcription of biosynthetic genes suggesting H3K79me2 functions in a negative feedback loop. Deleting AF10 reduces DOT1L catalytic function, and although it is still recruited by RNAPII, the epigenome resembles ESCs which enhances acquisition of the pluripotent fate.