Dot1L interaction partner AF10 safeguards cell identity during the acquisition of pluripotency.

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. In contrast, reduced levels of Dot1L interactors AF9 and ENL that recognize H3 acetylation decrease iPSC generation. Despite the opposite effects in reprogramming of Dot1L interacting proteins with differing histone reader specificity, 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. Taken together we reveal a key role for higher order gene body histone methylation in safeguarding 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 the bodies of genes with a rapid elongation rate (Duffy et al., 2018; Veloso et al., 2014). Although the molecular function of H3K79me remains mysterious, Dot1L is essential for development. Dot1L mouse knockout models are embryonic lethal (Feng et al., 2010; Jones et al., 2008) around the stage of organogenesis (E8.5) (Mitiku and Baker, 2007). H3K79me2 is the most differentially enriched modification more abundant in mouse embryonic fibroblasts (MEFs) compared to embryonic

stem cells (ESCs) (Sridharan et al., 2013). This enrichment of H3K79me2 in somatic cells is a barrier for pluripotency acquisition as depletion of Dot1L greatly enhances the generation of induced pluripotent stem cells (Jackson et al., 2016; Onder et al., 2012).

How H3K79me is targeted to genes with ongoing transcriptional elongation, and to which degree of methylation, is still unknown. 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, including AF9 (*Mllt9*), AF10 (*Mllt10*), AF17 (*Mllt6*), and ENL (*Mllt1*) (Mohan et al., 2010a). AF10 and AF17 contain a PZP domain that interacts with unmodified H3K27 (Chen et al., 2015). AF9 and ENL contain a YEATS domain that bind to H3K9 and H3K27 acylations (acetylation and crotonylation) (Li et al., 2014, 2016). Dot1L recruitment via interacting proteins has been decisively demonstrated in leukemias driven by rearrangements of the mixed-lineage leukemia (MLL) gene (MLL-R leukemias). AF9, ENL, and AF10 are some of the most frequent MLL fusion partners that bring Dot1L, and subsequent H3K79me, to MLL target loci (Mohan et al., 2010a; Okada et al., 2005).

Each of these interaction partners has differing effects on the degree of H3K79 methylation. At the global level, ENL depletion leads to a small decrease of H3K79me2/3 (Mohan et al., 2010b; Mueller et al., 2007). In contrast, depletion of AF9 decreases H3K79me3 (Mohan et al., 2010b) while that of AF10 decreases both H3K79me3 and me2 (Deshpande et al., 2014; Lin et al., 2009; Mohan et al., 2010b). AF9 and AF10 bind different chromatin signatures and localize H3K79me near transcriptional start sites in the case of AF9 (Li et al., 2014), and to gene bodies that lack H3K27me3 in the case of AF10 (Chen et al., 2015; Deshpande et al., 2014). Additionally, AF10 functions as a co-factor to increase Dot1L catalytic activity of higher order methylations (Deshpande et al., 2014). Thus, Dot1L complex composition could greatly affect the recruitment of H3K79me to genomic loci depending on the surrounding chromatin environment.

AF10 and AF9 not only target Dot1L to unique chromosomal loci, but proteomic analyses suggest differential functions. While Dot1L interacts with a similar ratio of AF9, and AF10, AF9 enriches for AF4 family/ENL family/P-TEFb transcriptional complex 10-fold more

than for Dot1L (Park et al., 2010). AF9 complexes are also significantly enriched for signaling linked ubiquitin ligases (Mohan et al., 2010b). In contrast, AF10 interacts with Dot1L 7-fold more than with AF9 (Mohan et al., 2010b), suggesting AF10 and Dot1L form an independent complex (Deshpande et al., 2014). Thus, AF9 is a more promiscuous complex component likely to have functions in transcription beyond its Dot1L association.

Here we mechanistically define the role of Dot1L associated proteins in the non-transformed and highly dynamic environment of somatic cell reprogramming to induced pluripotent stem cells (iPSCs). We find that Dot1L maintains cellular identity through AF10, but not AF9 or ENL. Deletion of AF10 accounts for about half of the Dot1L chemical inhibition phenotype in reprogramming. Similar to Dot1L inhibition, AF10 depletion results in few transcriptional changes, the majority of which are not congruent with expression in ESCs. We find that AF10-Dot1L interaction is required for the maintenance of cellular identity, yet AF10-histone interaction is dispensable, suggesting that AF10 functions as a Dot1L enzymatic cofactor in the acquisition of pluripotency rather than a histone targeting module.

RESULTS

Dot1L associated factors are expressed throughout reprogramming

Dot1L is essential for development (Feng et al., 2010; Jones et al., 2008) and a strong barrier to reprogramming to iPSCs (Jackson et al., 2016; Onder et al., 2012). 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 (Fig 1A), we examined their expression (Wille and Sridharan, 2020) in somatic cells – MEFs, cells on day 2 of reprogramming, cells on day 4 of reprogramming, and ESCs (Fig 1B). All assessed Dot1L associated factors were expressed throughout reprogramming, yet with different patterns (Fig 1B). 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 1C). 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 1C). Co-expression analysis revealed that Dot1L was expressed with both AF9 and AF10 in a similar number of MEFs, yet was co-expressed predominately with AF9 during reprogramming (Fig S1A-B). Dot1L was co-expressed almost exclusively with AF10 in ESCs suggesting that reprogramming and pluripotency status may affect Dot1L complex composition (Fig S1A-B).

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 concurrent with the initiation of reprogramming. Successful iPSC generation was measured by the expression of the pluripotency protein NANOG (Fig 1D, S1C). Interestingly, depletion of AF9 and ENL both decreased reprogramming efficiency, which could be 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 1D). Thus, AF10 is likely to be in the same pathway as Dot1L in pluripotency acquisition.

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 2A). In somatic cells isolated from the resulting AF10 reprogrammable mouse line, both the transition to pluripotency and AF10 expression can be temporally controlled. AF10 deletion using Cre recombinase (Fig 2B) in reprogramming cells isolated from this new line resulted in global reduction of H3K79me2, and to a lesser extent H3K79me1 (Fig 2C). This corroborates results from MLL-R leukemia where deletion of AF10 reduces H3K79me2/3 but not H3K79me1 (Deshpande et al., 2014).

We first compared the reprogramming efficiency of AF10 deletion (Δ) to the undeleted (fl/fl) control. ΔAF10 increased NANOG+ colonies by greater than four-fold (Fig 2D) as well as

transgene independent iPSCs (Fig S2A). Because AF10 can participate in other complexes (Mohan et al., 2010b), and the expression of AF10 and Dot1L are not completely coincident (Fig 1, S1), we determined if AF10 deletion could further enhance Dot1L inhibition. We added SGC0946 (Dot1Li), a potent small molecule Dot1L inhibitor, in combination with Cre-recombinase mediated deletion. However, AF10 deletion did not increase reprogramming beyond addition of Dot1Li (Fig 2D, S2A). To quantify the role of AF10 in Dot1L-mediated maintenance of cellular identity, a linear relationship between the percentage of AF10 deletion as measured by genotype PCR and percentage of total colonies obtained in ΔAF10 (cells treated with Dot1Li set to 100%) was established (Fig 2E). Using this model, AF10 was found to account for 44% of the Dot1L phenotype in reprogramming.

We further tested this 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 (Kaniskan et al., 2018). Irrespective of the drug concentration, deletion of AF10 always increased reprogramming efficiency as compared to the fl/fl control (Fig 2F). The effect was of AF10 deletion was most pronounced when Dot1L inhibition was 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. Thus, deletion of AF10 functions in the same pathway and is likely to be 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) of a reprogramming timecourse and assessed for NANOG+ colony formation on day 6 (Fig 2G). Similar to Dot1L inhibition, loss of AF10 most potently increased early and mid-reprogramming (Fig 2H, S2B) (Onder et al., 2012; Wille and Sridharan, 2020). Taken together, AF10 and Dot1L collaboratively maintain cellular identity during somatic cell reprogramming.

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 change (Fig 2G-H), 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 (Fig 3B) and had not yet reached the ESC state as expected for a sample from the midpoint of the process. To observe the trend in functional categories that changed dynamically, we generated a heatmap of differentially expressed (DE) genes from all conditions. Two clusters (C1 and C2) were altered by Dot1L disruption to resemble ESCs (Fig 3A). In the larger of these clusters, C2, functional groups such as negative regulation of cellular proliferation and epithelium development pathways were enriched (Fig S3A). However, most of the transcriptional changes are not representative of ESC gene expression (Fig 3A). In clusters 3-6, the effect of Dot1Li generates an expression profile that does not resemble either MEF or ESCs. In all these clusters, the effect of AF10 deletion is milder than that with 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 S3B). Motif analysis revealed that cluster 7 is enriched for the MYC associated factor, MAX, binding sites (Fig S3B) but neither *Myc* nor *Max* are DE in any condition suggesting that expression of C7 is epigenetically regulated.

Although deletion of AF10 positions cells closer to Dot1Li treatment (Fig 3B), few genes are significantly DE. At the starting point in MEFs, fewer than 15 genes were DE at high confidence (Fig 3C, Methods). On day 4 of reprogramming in control treated cells, ΔAF10 only resulted in the upregulation of 3 genes relative to fl control (Fig 3C). There was a single upregulated gene in ΔAF10 relative to fl in cells treated with Dot1Li (Fig 3C) further demonstrating that deletion of AF10 does not add to Dot1L inhibition. Treatment of fl cells with Dot1Li had the greatest effect with many more genes upregulated (150) than downregulated (8) relative to control treatment (Fig 3D). Dot1Li treatment of ΔAF10 cells produced 37 upregulated genes and 3 downregulated (Fig 3D). Twenty-two genes were upregulated by Dot1Li in both fl and ΔAF10 cells (Fig 3E), and only *Hoxd12* and *Osr1* were commonly downregulated (Fig 3F). Overall, chemical inhibition of Dot1L resulted in few DE genes and the loss of AF10 yielded almost no significant transcriptional changes at the steady state mRNA level although it lessened the Dot1Li transcriptional effect.

As we previously found (Wille and Sridharan, 2020), most of the Dot1Li DE genes do not resemble their expression in ESCs (Fig 3A). This aberrant expression is observed with much lower magnitude in ΔAF10, which suggests that improper activation or failure to downregulate many of these genes does not impair reprogramming as Dot1Li is more efficient than deletion of AF10 in NANOG+ colony formation (Fig 2D). On the other hand, it is possible that deletion of AF10 may fail to achieve the full reprogramming efficiency of Dot1Li as genes in cluster 7 are aberrantly downregulated in ΔAF10 but not in Dot1Li treatment (Fig S3B).

AF10 enforces cellular identity through Dot1L interaction but not histone interaction

Since the steady state mRNA levels remain largely unchanged upon AF10 deletion, 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 modular functions of the AF10 protein in reprogramming. 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 4A). We generated one mutant of AF10 that cannot interact with histones (NLS-CTD) and two mutants that omitted either the OMLZ domain alone (ΔOMLZ) or in combination with the C-terminal domain (NTD-OMLZ) to prevent interaction with Dot1L (Fig 4A) in a vector with a fluorescent reporter. AF10 mutants were individually transduced into AF10 conditional reprogrammable MEFs and sorted by expression of the fluorescent marker with flow cytometry (Fig S4A) to isolate only the cells that included the mutant version of AF10. Equivalent levels of each mutant protein were observed in the sorted AF10 conditional reprogrammable MEFs (Fig 4B). Endogenous AF10 was deleted and reprogramming was induced (Fig 4C). Therefore, each reprogramming condition expressed either exogenous AF10 mutants alone (Δ) or mutant proteins in addition to the endogenous AF10 (fl) (Fig 4C-D). The abundance of H3K79me2 was reduced by AF10 deletion and rescued with the full-length and histone interaction deficient (NLS-CTD) AF10 mutant (Fig S4B). Thus, the NLS-CTD of AF10 preserves its function as a cofactor for boosting H3K79me2 levels. As previously demonstrated (Fig 2D, S2A), deletion of AF10 enhanced reprogramming (Fig 4D,

S4C). Expression of exogenous full length AF10 reduced the number of NANOG+ colonies during reprogramming (Fig 4D, S4C) confirming that this deletion rescue strategy could be used to identify the domains of AF10 responsible for safeguarding cell identity. The two AF10 mutants that cannot interact with Dot1L phenocopied AF10 deletion (Fig 4A, D, S4C), and resulted in an increase of iPSCs. In contrast, the AF10 mutant that cannot interact with histones phenocopied endogenous (fl) AF10 (Fig 4A, D, S4C). Therefore, AF10 functions in maintaining cell identity by increasing higher order methylation on H3K79 by boosting Dot1L catalytic activity

These data implicated 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 (Laugesen et al., 2019). PRC2 also contains EED as one of the core subunits that binds H3K27me3 and increases enzymatic activity to facilitate H3K27me3 spread (Laugesen et al., 2019). Cells were treated with A395, an EED small molecule inhibitor (He et al., 2017). If AF10-histone interaction is causal, depletion of H3K27me3 will create more potential binding sites for AF10 which could further decrease pluripotency acquisition. Although A395 resulted in loss of H3K27me3 detection (Fig 4E), the number of pluripotent colonies was unchanged in the presence and absence of AF10 compared to control treatment (Fig 4F). Notably, neither Dot1L nor EED inhibition affected the opposing mark globally indicating there is not widespread antagonism between H3K79me2 and H3K27me3 in pluripotency acquisition (Fig 4E). Therefore AF10-Dot1L interaction, but not histone targeting, is required as a barrier of somatic cell reprogramming.

The enrichment of H3K79me2/3 was suggested to preclude H3K27me3 at specific loci in leukemia (Deshpande et al., 2014). Therefore, we investigated whether the ability of AF10 to block reprogramming was due to its function in repelling local gain of H3K27me3. We performed a ChIP for H3K27me3 in MEFs and AF10-deleted cells of day 4 of reprogramming. Genes marked by H3K79me2 in MEFs, ESCs, or shared in both had very low enrichment for H3K27me3 that did not increase upon AF10 deletion (Fig S4D). Additionally, we assessed the role of AF10 in generation of bivalent domains. Bivalency is an epigenetic domain important in

pluripotent stem cells where both activating (H3K4me3) and repressive (H3K27me3) modifications are enriched on lineage specifying genes so that cells are poised for facile differentiation. ΔAF10 did not allow for H3K27me3 to spread to the correct genomic positions at bivalent genes in ESC (Fig S4E). Thus, AF10 safeguards cellular identity through protection of higher-order H3K79 methylations, but not in connection with H3K27me3.

DISCUSSION

We and others have demonstrated that Dot1L is a strong barrier for pluripotency acquisition (Jackson et al., 2016; Onder et al., 2012; Wille and Sridharan, 2020), yet surprisingly, H3K79me2 is enriched on many of the same genes in ESCs and MEFs (Wille and Sridharan, 2020). However, the global level of H3K79me2 is much higher in somatic cells such as MEFs and keratinocytes compared to ESCs (Sridharan et al., 2013; Wille and Sridharan, 2020). Here we find that pluripotency acquisition is directly correlated to Dot1L catalytic activity. Deletion of the Dot1L cofactor AF10 reduces H3K79me2 (Fig 2C) and accounts for about half of the inhibited Dot1L phenotype in reprogramming (Fig 2E). ΔAF10 cells likely only reach half of the Dot1Li reprogramming efficiency due to retained H3K79me1 (Fig 2C). Furthermore, we find that the AF10-Dot1L interaction domain that mediates higher order H3K79 methylation (Fig S4B) (Deshpande et al., 2014) is necessary to safeguard cellular identity relative to wild-type AF10 (Fig 4). Thus, maximum pluripotency acquisition requires low enrichment of all forms H3K79me.

The sustained levels of H3K79me1 demonstrate that Dot1L does not require AF10 for targeting to genomic loci during reprogramming. Accordingly, we show that the AF10-histone interaction domain is dispensable for maintaining cellular identity (Fig 4) and H3K79me2 (Fig S4B). This result is surprising as AF10 histone interaction is required for global H3K79me2 levels and local enrichment at Dot1L target genes in cancer cells, as well as leukemia cell growth (Chen et al., 2015). Thus, Dot1L may have a specific targeting complex in non-transformed cells or in pluripotency. On the other hand, AF10 contains a AT-hook motif that can directly bind to DNA (Linder et al., 2000) and can be further explored for its function in Dot1L localization.

We do not find any evidence that H3K79me2 and H3K27me3 reciprocally regulate one another in reprogramming to iPSCs (Fig 4). Inhibition of EED and loss of H3K27me3 did not increase H3K79me2 or inhibit reprogramming (Fig 4E-F), and deletion of AF10 did not allow for spread of H3K27me3 into genes that contain H3K79me2 or are bivalent in ESCs (Fig S4D-E). 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). Interestingly, in leukemia cell lines, Dot1L inhibition reduced protein translation by downregulation of ribosomal genes, which was partially rescued by an EZH2 inhibitor (Lenard et al., 2020). However, other differentially expressed genes were both synergistically and antagonistically affected by combined Dot1L/EZH2 inhibition (Lenard et al., 2020). Ribosomal genes were not downregulated in reprogramming and thus Dot1L may function differentially in non-transformed cells where it is not tethered to MLL-R fusions.

To dissect the function of Dot1L in normal cells, we examined Dot1L interacting proteins with histone binding domains. We uncovered that AF10, but not AF9 or ENL, functions in the same pathway as Dot1L in reprogramming (Fig 1). AF9 and ENL are mutually exclusive interactors with Dot1L and the AF4 family/ENL family/P-TEFb (AEP) complex (He et al., 2011). AEP directly promotes transcription by CDK9 phosphorylation of RNA polymerase II, DSIF, and NELF. The AF9 and ENL YEATS domain bridges this transcriptional complex to chromatin via interaction with the polymerase-associated factor complex (PAFc) (He et al., 2011). 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 interaction function of AF9 and ENL contribute to their phenotype in reprogramming.

While the expression of Dot1L does not change throughout reprogramming, AF10 is most highly expressed in ESCs (Fig 1B-C). Additionally, AF10 is uniquely co-expressed with Dot1L in the highest proportion of ESCs compared to MEFs or reprogramming cells (Fig S1A-B). Thus, deletion of AF10 may enhance pluripotency acquisition by preventing this late interaction. At the same time, we find that deletion of AF10 increased reprogramming most efficiently at the beginning or mid-point of reprogramming (Fig 2G-H). This suggests that the AF10 protein may be long lived, or AF10-Dot1L complexes irreversibly maintain the somatic cell epigenome early in reprogramming.

Although AF10 accounts for about half of the Dot1L phenotype, deletion of AF10 results in few transcriptional changes (Fig 3). 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, 2020). 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 3), indicating that they are in the same pathway. AF10 deletion similarly modulates genes that are differentially expressed to resemble ESCs as Dot1Li treatment (Fig 3E, clusters C1-2). However, AF10 deletion has a much smaller burden of genes that are upregulated in opposition to ESCs relative to MEFs (Fig 3E, C3-6) even though it has a lower reprogramming efficiency compared to Dot1Li (Fig 2). Therefore, the majority of steady state transcriptional alterations do not contribute to pluripotency acquisition, and instead AF10 functions as an Dot1L cofactor to mediate higher order H3K79 methylation to maintain 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 60°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 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), EPZ5676 (MedChem Express, HY-15593), or EPZ004777 (MedChem Express, HY-15227) beginning at day 0 of reprogramming, at the indicated concentrations. SGC0946 was used at 5 µM in all additional reprogramming experiments. 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. NANOG positive colonies were quantitated using fluoroscopy and statistically analyzed with GraphPad Prism unpaired two-tailed t tests 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, mixed, and allowed to incubate for 20 minutes at room temperature. 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). 2.4×10^9 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. Cells were allowed to rest for 1-2 days before reprogramming initiation.

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 beginning at 1 day post thaw. 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) 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 pH8.3, 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 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; Abcam, ab2886), H3K79me2 (1:1000; Active Motif, 39143), 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 S4A). Flow cytometry images were generated in FlowJo.

RNA-Seq. RNA was isolated from 2 independent biological replicate reprogramming experiments 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 HiSeq2500 (UW-Madison Biotechnology) single end, 100 bp. Samples had a 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 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 AF10ff 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: 16 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), mid (8 cycles) and at the end (16 cycles) to test sonication fragmentation. Sonication test samples were resuspended in 3 volumes of water with 10 µg of RNAse and incubated at 37°C for 30 minutes. 20 µg of Proteinase K was added, and samples were allowed to reverse crosslink overnight at 60°C. DNA was cleaned and isolated with phenol-chloroform extraction using phase-lock tubes. 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.

15 µg of 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). Chromatin was immunoprecipitated with 5 µg with anti-H3K27me3 (Active Motif, 39155) 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 60°C to reverse crosslinks. DNA was cleaned with phenol-chloroform phase separation with phase lock tubes. 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). Libraries were assessed with Qubit quantitation and Bioanalyzer3.0. Single-end 50 bp sequencing was performed on a HiSeq2500 (UW-Madison Biotechnology Center) yielding about 8 million reads per sample.

ChIP analysis. Reads were aligned to the mm9 mouse genome with bwa-0.6.2 (Li and Durbin, 2009) using the "aln" command. Sam files were generated from .bai files with bwa-0.6.2 with "samse". Samtools-1.2 was used to convert .sam files into .bam files using the "view" and then sort .bam files with the "sort" command. Peaks were called with MACS2 (Feng et al., 2012; Zhang et al., 2008) with --broad -p 0.0001 parameters for H3K79me2 ChIPs (Chronis et al., 2017). Peaks were annotated in EaSeq (Lerdrup et al., 2016) from the center of the peak to the center of the nearest gene. Genes with H3K79me2 in MEFs, ESCs, or shared in both were determined by overlap of gene lists with a called intergenic peak using Venny2.0. Bivalent genes

were aggregated from previous datasets (Maherali et al., 2007; Mikkelsen et al., 2008) and had an H3K4me3 and H3K27me3 peak. Heatmaps and metaplots were created using deepTools2 (Ramírez et al., 2016). Briefly, bigwig files were generated with the following parameters: bamcoverage --bin size 10 --normalizeUsing CPM --ignoreForNormalization chrX. Matrices of ChIP signal for specific gene sets were made with the following parameters: computeMatrix scale-regions -b 3000 -a 3000 --regionBodyLength 5000. Using the ChIP signal matrices, heatmaps were created with the "plotHeatmap" command and metaplots with the "plotProfile" command.

Published datasets. Bulk expression of Dot1L and associated factors was assessed from data previously generated (Wille and Sridharan, 2020). 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. H3K79me2 ChIP-seq datasets were acquired from NCBI GEO GSE90893 (Chronis et al., 2017) and processed as described in Methods, ChIP-seq analysis.

ACKNOWLEDGEMENTS

This work was supported by a UW-Madison Stem Cell and Regenerative Medicine Center postdoctoral award to C.K.W. and Shaw Scientist award to the R.S. 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.

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.

Chronis, C., Fiziev, P., Papp, B., Butz, S., Bonora, G., Sabri, S., Ernst, J., and Plath, K. (2017). Cooperative Binding of Transcription Factors Orchestrates Reprogramming. Cell *168*, 442-459.e20.

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.

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.

Feng, J., Liu, T., Qin, B., Zhang, Y., and Liu, X.S. (2012). Identifying ChIP-seq enrichment using MACS. Nat Protoc *7*, 1728–1740.

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.

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.

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.

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.

de Hoon, M.J.L., Imoto, S., Nolan, J., and Miyano, S. (2004). Open source clustering software. Bioinformatics *20*, 1453–1454.

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.

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.

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.

Kaniskan, H.Ü., Martini, M.L., and Jin, J. (2018). Inhibitors of Protein Methyltransferases and Demethylases. Chem Rev *118*, 989–1068.

Laugesen, A., Højfeldt, J.W., and Helin, K. (2019). Molecular Mechanisms Directing PRC2 Recruitment and H3K27 Methylation. Mol Cell *74*, 8–18.

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.

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.

Lerdrup, M., Johansen, J.V., Agrawal-Singh, S., and Hansen, K. (2016). An interactive environment for agile analysis and visualization of ChIP-sequencing data. Nat. Struct. Mol. Biol. *23*, 349–357.

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.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

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.

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.

Lin, Y.-H., Kakadia, P.M., Chen, Y., Li, Y.-Q., Deshpande, A.J., Buske, C., Zhang, K.-L., Zhang, Y., Xu, G.-L., and Bohlander, S.K. (2009). Global reduction of the epigenetic H3K79

methylation mark and increased chromosomal instability in CALM-AF10-positive leukemias. Blood *114*, 651–658.

Linder, B., Newman, R., Jones, L.K., Debernardi, S., Young, B.D., Freemont, P., Verrijzer, C.P., and Saha, V. (2000). Biochemical analyses of the AF10 protein: the extended LAP/PHD-finger mediates oligomerisation. J Mol Biol *299*, 369–378.

Maherali, N., Sridharan, R., Xie, W., Utikal, J., Eminli, S., Arnold, K., Stadtfeld, M., Yachechko, R., Tchieu, J., Jaenisch, R., et al. (2007). Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. Cell Stem Cell *1*, 55–70.

Mikkelsen, T.S., Hanna, J., Zhang, X., Ku, M., Wernig, M., Schorderet, P., Bernstein, B.E., Jaenisch, R., Lander, E.S., and Meissner, A. (2008). Dissecting direct reprogramming through integrative genomic analysis. Nature *454*, 49–55.

Mitiku, N., and Baker, J.C. (2007). Genomic analysis of gastrulation and organogenesis in the mouse. Dev Cell *13*, 897–907.

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.

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.

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.

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.

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.

Park, G., Gong, Z., Chen, J., and Kim, J.-E. (2010). Characterization of the DOT1L network: implications of diverse roles for DOT1L. Protein J. *29*, 213–223.

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.

Saldanha, A.J. (2004). Java Treeview--extensible visualization of microarray data. Bioinformatics *20*, 3246–3248.

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.

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.

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.

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.

Wille, C.K., and Sridharan, R. (2020). Dot1L methyltransferase activity is a barrier to acquisition of pluripotency but not transdifferentiation. BioRxiv 835413.

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.

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.

Figure 1

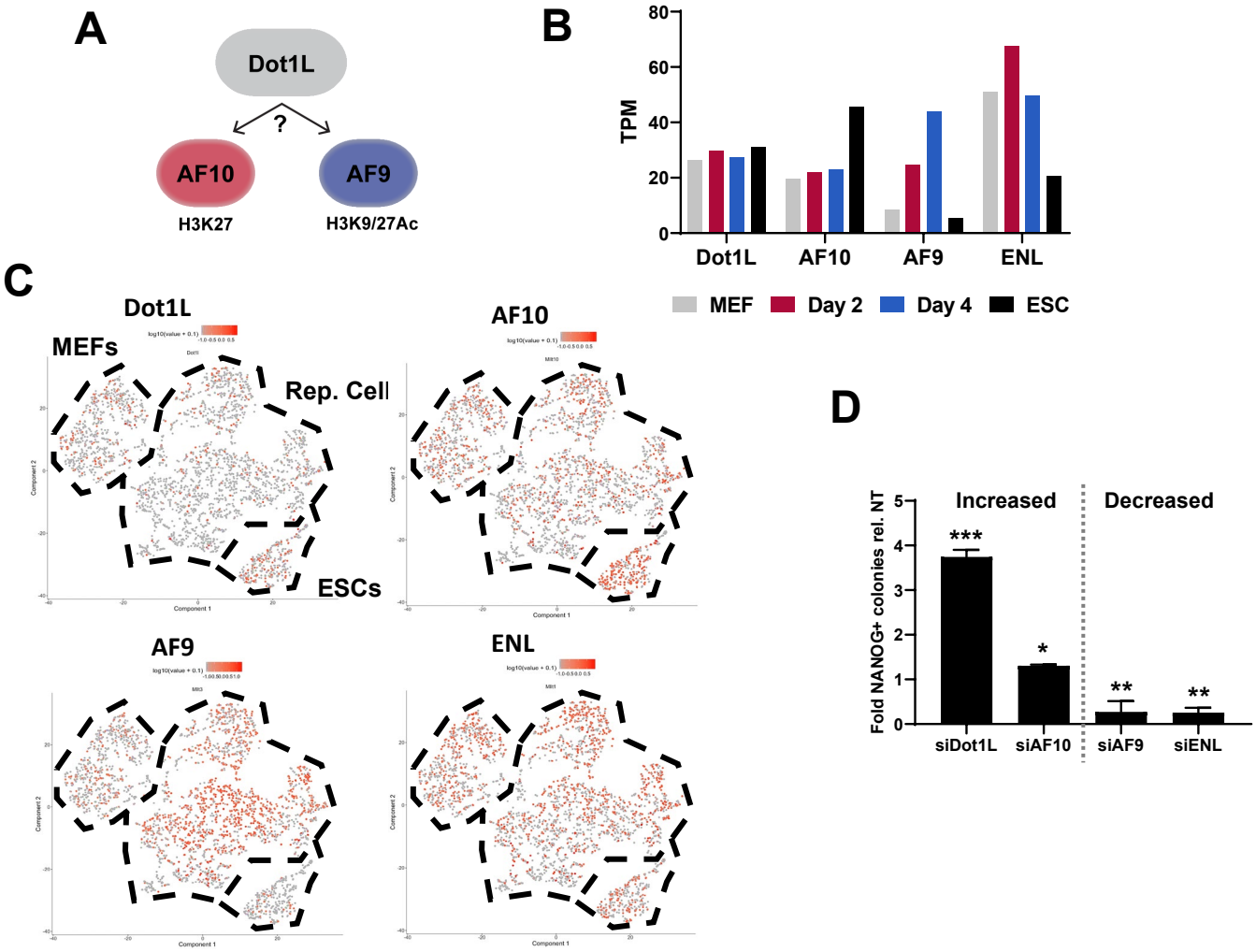

Figure 1. Dot1L associated factors are expressed throughout reprogramming.

A. Schematic of Dot1L associated factors. AF10 has a PZP domain that binds to unmodified H3K27. AF9 YEATS domain binds to H3K9ac and H3K27ac. AF9 also interacts with transcriptional complexes super elongation complex (SEC) and the polymerase associated factor complex (PAFc).

B. Expression of Dot1L associated factors determined by bulk RNA-Seq (Wille and Sridharan, 2020) in MEFs, day 2 of reprogramming, day 4 of reprogramming, and ESCs.

C. 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.

D. Fold NANOG+ colonies upon siRNA depletion of indicated genes relative to non-targeting control. Error bars represent the standard deviation of two biological replicates each composed of two technical replicates. ***P<0.001, **P<0.01, and *P<0.05 by unpaired t-test.

Figure 2

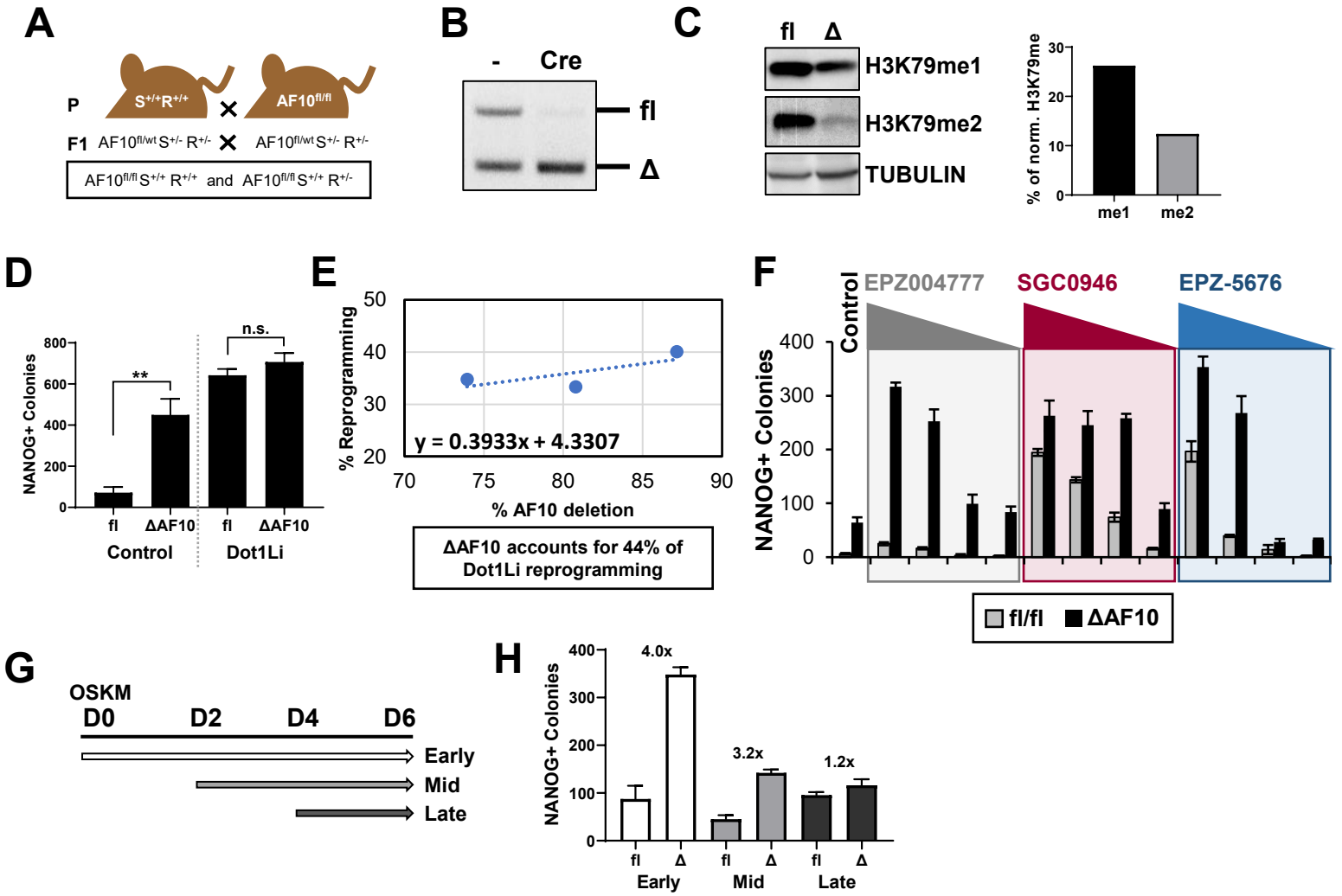

Figure 2. Dot1L maintains cellular identity through AF10.

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

B. Addition of Cre-recombinase deletes AF10 from the genome.

C. Left: Immunoblot of H3K79me1 and me2 in AF10 wt (fl) or Cre deleted (ΔAF10) conditions on day 6 of reprogramming. Right: Percent of retained H3K79me signal (normalized to TUBULIN) in ΔAF10, fl set to 100%.

D. NANOG+ colonies on day 6 of reprogramming in cells treated with empty adenovirus (fl) or adenovirus-Cre (ΔAF10), with and without Dot1Li treatment. Error bars represent the standard deviation of three technical replicates of a representative biological replicate. **P<0.01 and not significant (n.s.) P>0.05 by unpaired t-test.

E. Linear relationship of the percentage of AF10 genomic deletion measured by genotyping PCR, to the percentage of iPSC formation in AF10 deleted conditions (colonies formed in Dot1Li set to 100%).

F. NANOG+ colonies on day 5 of reprogramming in control fl (gray bars) and AF10 deleted (black bars) MEFs treated with decreasing amounts (5, 2.5, 1, 0.2 μM) of Dot1L inhibitors: EPZ004777 (gray), SGC0946 (red), and EPZ-5676 (blue). Error bars represent the standard deviation of two technical replicates of a representative biological replicate.

G. AF10 timecourse reprogramming scheme. Cells were treated with control or Cre recombinase adenovirus Early on day 0 (D0), Mid on day 2 (D2), or Late on day 4 (D4) of reprogramming. NANOG+ colonies were analyzed by immunofluorescence on day 6.

H. AF10 timecourse fold reprogramming efficiency is calculated as the number of NANOG+ colonies in ΔAF10 relative to the time-matched fl control. Error bars represent the standard deviation of two technical replicates.

Figure 3

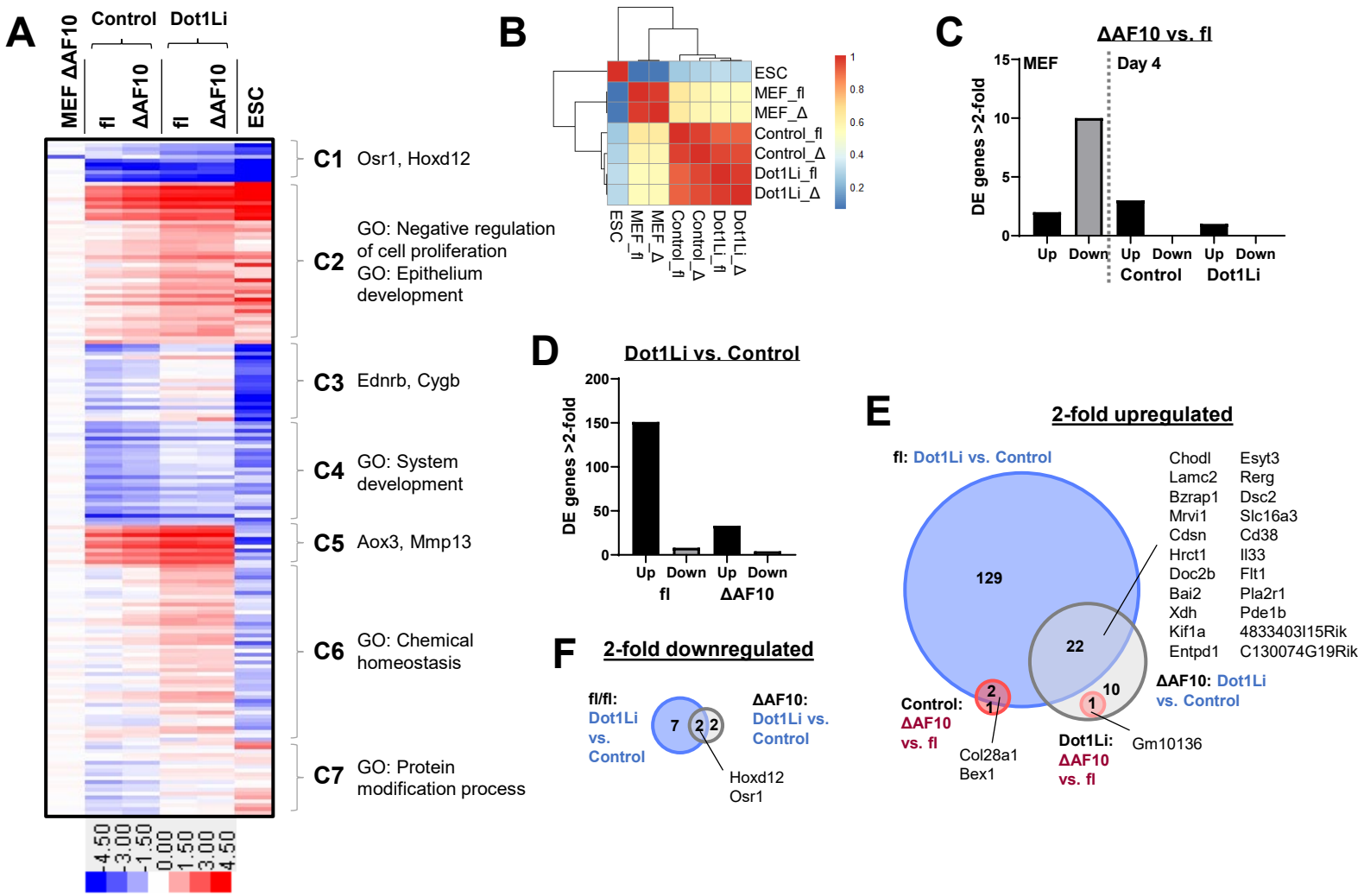

Figure 3. AF10 deletion has few transcriptional effects.

A. k-means clustered heatmap of Log2FC TPM of DE genes from all conditions relative to fl MEFs.

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-F. Overlap of upregulated (E) or downregulated (F) 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 4

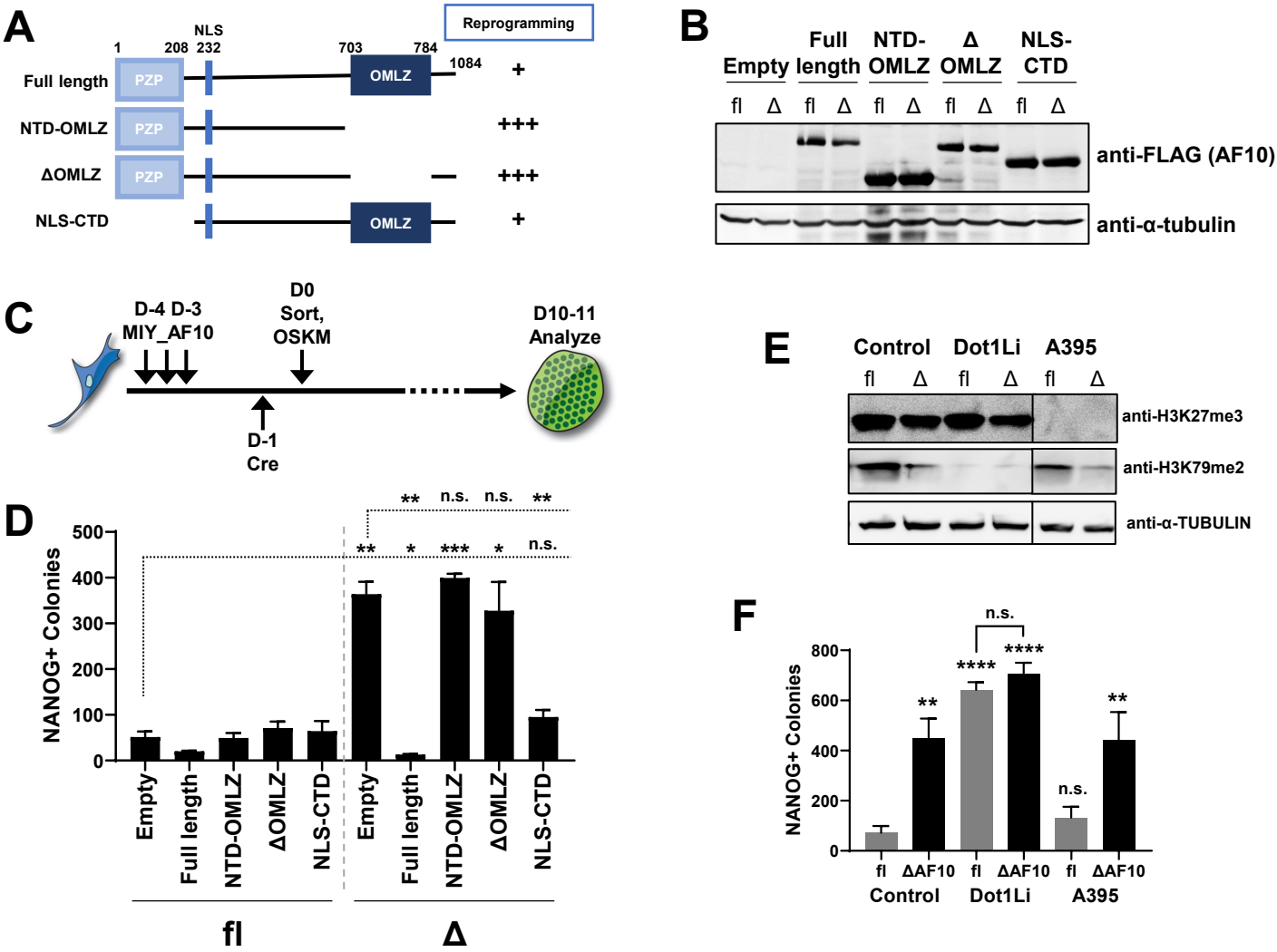

Figure 4. AF10 enforces cellular identity through Dot1L interaction but not histone interaction.

A. 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 Dot1L interaction (NTD-OMLZ and ΔOMLZ) or histone interaction (NLS-CTD) are shown on left, and reprogramming efficiency in the absence of endogenous AF10 shown on right.

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

C. 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.

D. Reprogramming efficiency of AF10 domain mutant containing cells was assessed by NANOG immunofluorescence on day 10. Error bars represent the standard deviation of three technical replicates of a representative biological replicate. ***P<0.001, **P<0.01, *P<0.05, and not significant (n.s.) P>0.05 by unpaired t-test.

E. Immunoblot of H3K79me2, H3K27me3, and α-TUBULIN in cells treated with control, Dot1Li, or A395 (EED inhibitor) on day 5 of reprogramming.

F. NANOG+ colonies on day 6 of reprogramming of fl or ΔAF10 cells treated with Dot1Li, A395, or control. Error bars indicate the standard deviation of 3 technical replicates of a representative biological replicate. ****P<0.0001, **P<0.01, and not significant (n.s.) P>0.05 by unpaired t-test.