# ***Dnmt3b*as regulates transcriptional induction and alternative splicing of *Dnmt3b***

Mohd Saleem Dar[1], Isaiah K Mensah[1], Sarah McGovern[1], Ming He[1], Mark C Hall[1,2], Hannah Christian Whitlock, Nina Elise Bippus, Madison Ceminsky, Humaira Gowher[1,2*]

[1]Department of Biochemistry; [2]Purdue University Center for Cancer Research; Purdue University; West Lafayette, Indiana 47907 USA

*Correspondence: hgowher@purdue.edu

**Running title:** Role of *Dnmt3b*as in *Dnmt3b* expression

## Summary

Our study identified the mechanism by which *Dnmt3bas* lncRNA and *Dnmt3b* cis-regulatory elements control inducible expression and alternative splicing of *Dnmt3b*. *Dnmt3bas* knockdown increases transcriptional induction and decreases H3K27me3 at Dnmt3b upstream elements, suggesting that PRC2 targeting by *Dnmt3bas* fine-tunes *Dnmt3b* induction. Moreover, we identified the catalytically inactive *Dnmt3b6* as a major isoform in naïve ESCs. Alternative splicing switches to include exons during transcriptional induction, thus generating the catalytically active isoform, *Dnmt3b1*. While *Dnmt3bas* overexpression attenuates *Dnmt3b* induction, we observed a relative increase in the *Dnmt3b1* isoform. Towards a mechanism, our data demonstrated that hnRNPL, a *Dnmt3bas* binding partner, promotes exon inclusion. Therefore, we propose that *Dnmt3bas* recruits hnRNPL to the Dnmt3b promoter, where its interaction with RNA Pol II coordinates alternative splicing with transcriptional induction. With this two-pronged approach, DNMT3B activity can be tightly controlled during the short developmental window, ensuring the fidelity and specificity of DNA methylation.

## Introduction

*Dnmt3b* catalyzes *de novo* DNA methylation during early embryogenesis, and its homozygous knockout in mice is embryonic lethal. *Dnmt3b*$^{-/-}$ ESC (murine embryonic stem cells) show defective differentiation potential and loss of DNA methylation at the regulatory elements of lineage-specific genes and minor satellite repeats, which are the preferred targets of the *Dnmt3b* enzyme (Gopalakrishnan et al., 2009; Norvil et al., 2020; Okano et al., 1999). In somatic cells, the expression of *Dnmt3b* is low and is controlled in a tissue-specific manner (Robertson et al., 2000; Robertson et al., 1999). The aberrant increase in *Dnmt3b* expression leads to DNA hypermethylation and loss of gene regulation in various human diseases (Esteller, 2005; Walton et al., 2014; Watanabe and Maekawa, 2010). Sequence polymorphisms in the *DNMT3B* promoter that increase promoter activity are associated with an increased risk of

several cancers (Montgomery et al., 2004; Shen et al., 2002; Singal et al., 2005). Loss of function mutations in human *DNMT3B* are highly prevalent in patients with ICF (immunodeficiency, centromeric instabilities, and facial abnormalities) syndrome and cause hypomethylation of repetitive elements and genomic instability in B cells (Gowher and Jeltsch, 2002; Okano et al., 1999; Xie et al., 2006; Xu et al., 1999). *Dnmt3b* is transcribed in more than 30 alternatively spliced isoforms, although only a few have been detected at the protein level (Ostler et al., 2007; Xie et al., 1999). Several of these isoforms are catalytically inactive due to the loss of key catalytic residues. The two major isoforms expressed in normal cells are *Dnmt3b1* and *Dnmt3b6*. The exclusion of exons 22 and 23 results in the loss of a significant part of the target recognition region in the Dnmt3b6 enzyme, thus rendering it inactive. However, Dnmt3b6 has been shown to interact and allosterically activate *Dnmt3b1* catalysis (Duymich et al., 2016; Ostler et al., 2007; Saito et al., 2002; Wang et al., 2007; Weisenberger et al., 2004; Zeng et al., 2020). Aberrant expression and splicing of *Dnmt3b* is linked to the loss of methylation at oncogenes and repetitive elements in diverse cancers, including colorectal, lung, and breast cancers (Guil and Esteller, 2009; Kulis and Esteller, 2010; Mensah et al., 2021; Szyf, 2005). Collectively, these observations indicate that the highly regulated spatiotemporal expression and alternative splicing of *Dnmt3b* are critical for cell differentiation, homeostasis, and survival (Huntriss et al., 2004; Okano et al., 1998).

The murine Dnmt3b gene locus comprises a CpG island promoter and an enhancer element ~8 kb upstream of the *Dnmt3b* promoter (Ishida et al., 2003; Jinawath et al., 2005). A divergently expressed antisense transcript (*Dnmt3bas*) initiates 1 kb downstream of the *Dnmt3b* promoter and is potentially a long non-coding RNA (lncRNA). LncRNAs have critical cellular functions, including transcription, splicing, translation, RNA export, chromatin looping, and DNA repair. Most lncRNAs are transcribed by RNA Pol II but at lower levels, and many undergo capping, splicing, and polyadenylation (Mele et al., 2017; Statello et al., 2021; Zuckerman and Ulitsky, 2019). LncRNAs are preferentially located in the nucleus, interact with chromatin proteins and

regulate the activation and silencing of genes (Derrien et al., 2012; Guo et al., 2020; Guttman et al., 2009; Guttman et al., 2011; Tian and Manley, 2017). They are often expressed from syntenic regions among animal species (Dinger et al., 2008; Ponjavic et al., 2007), and widespread antisense lncRNA transcription is evolutionarily conserved (Seila et al., 2008). Many lncRNAs such as HOTAIR (Rinn et al., 2007), Kcnq1ot1 (Thakur et al., 2004), Air (Sleutels et al., 2002), Evx1as(Luo et al., 2016), and Evf2as (Feng et al., 2006) are coordinately transcribed with a protein-coding gene such that the pair is expressed from one transcriptional locus (Engstrom et al., 2006). Several studies have used comparative expression analysis of lncRNA/mRNA pairs to support a functional relationship between the divergent lncRNA and associated protein-coding gene (Sigova et al., 2013). During ESC differentiation, several lncRNAs originate from the divergent transcription. These lncRNA/mRNA gene pairs are coordinately regulated and required for pluripotency and differentiation (Guttman et al., 2011; Luo et al., 2016).

*Dnmt3b* is dynamically expressed during development, with the highest expression in the early stages. During differentiation of naive pluripotent mouse embryonic stem cells (2i-ESCs), the temporal expression pattern of *Dnmt3b* is similar to that observed during embryonic development (Ficz et al., 2013; Leitch et al., 2013; Ying et al., 2008), making them an ideal model system. In undifferentiated 2i-ESCs, *Dnmt3b* is expressed at a low basal level. However, immediately upon the induction of differentiation, the expression of *Dnmt3b* is strongly upregulated. As differentiation proceeds, *Dnmt3b* expression is downregulated (Leitch et al., 2013). Despite the breadth of evidence supporting the critical role of *Dnmt3b* in differentiation and cell identity, little is known about the mechanisms that control its transcription.

Our study is the first to report the role and mechanism of *Dnmt3bas* in the regulation of *Dnmt3b* expression. Gene expression analysis of undifferentiated 2iESCs and cells post differentiation showed a coordinated yet contrasting expression pattern of *Dnmt3b* and *Dnmt3bas*, with the

*Dnmt3b* pattern mimicking the expression of *Dnmt3b* observed *in vivo*. The chromatin modification state of proximal and distal enhancers and enhancer-promoter looping complemented the *Dnmt3b* transcriptional state, thus establishing the role of the upstream cis-regulatory elements in regulating *Dnmt3b* expression. Whereas we observed no discernable effect on *Dnmt3b* basal expression, the induced expression of *Dnmt3b* was inversely affected by downregulation (KD) or overexpression (OE) of *Dnmt3bas*. Interestingly, we observed an increase in H3K27me3 at the cis-regulatory elements pre-induction in undifferentiated OE cells suggesting the role of Dnmt3bas in the regulation of PRC2 activity at these sites. Additionally, higher enrichment of H3K27Ac and an increased interaction frequency between enhancer and promoter were observed in the KD cells post-differentiation. Next, we asked if changes in *Dnmt3b* expression in *Dnmt3bas* manipulated cell lines affected the alternative splicing (AS) of *Dnmt3b*. This analysis revealed that in WT undifferentiated 2i-ESCs, catalytically inactive *Dnmt3b6* is expressed as the major isoform. Upon *Dnmt3b* induction, the AS switches to exon inclusion, promoting the expression of the catalytically active, *Dnmt3b1* as the major isoform. However, in Dnmt*3bas* OE cells, we observed increased exon inclusion. Our systematic experimental analysis determined the splicing factor, hnRNPL, as the binding partner of *Dnmt3bas* that facilitates exon inclusion in *Dnmt3b1* expression.

Overall, this comprehensive study elucidates how cis-regulatory factors control the transcriptional induction and alternative splicing of an essential developmental gene and establish a unique role of *Dnmt3bas* in transcriptional priming and alternative splicing of *Dnmt3b*.

**Keywords:** lncRNA, alternative splicing, DNA methylation, chromatin-modifications, promoters, enhancers, Dnmt3b, Dnmt3bas

## Results

### ***The upstream proximal and distal enhancers regulate Dnmt3b induction.***

We adapted serum cultured murine embryonic stem cells (s-ESCs) to 2i media (2i-ESCs). As reported earlier, during the adaptation process, naïve pluripotency markers were upregulated. Moreover, the expression of *Dnmt3a* and *Dnmt3b* were downregulated, with the concomitant loss of DNA methylation genome-wide (Figure S1A-1C). We differentiated 2i-ESCs and monitored the induction of *Dnmt3b* transcription. RT-qPCR analysis showed a 10-15 fold increase in *Dnmt3b* expression, which peaks at D3, followed by a substantial decrease that stays steady from D4 to D6 post-differentiation (Figure 1A). Since *Dnmt3b* is expressed at a low basal level in 2i-ESCs, the induction is significantly higher than observed during the differentiation of s-ESCs (Figure S1D).

The *Dnmt3b* gene locus contains a CpG island promoter and a putative proximal and distal enhancer elements located about 0.3 and 8 kilobases upstream of the transcription start site (TSS), respectively (Figure S1E) (ENCODE CRE) (Ficz et al.). To establish the engagement of the putative distal enhancer in the regulation of *Dnmt3b* expression, we examined the enhancer-promoter (E-P) looping interaction pre and post-differentiation in 2i-ESCs, using a chromatin conformation capture (3C) assay (Hagege et al., 2007; Naumova et al., 2012). We observed a potent and specific contact between the distal enhancer and promoter regions. Furthermore, a significant increase in the interaction frequency was observed on D3 post-differentiation. However, the E-P loop was prevalent in the undifferentiated 2i-ESC when *Dnmt3b* was expressed at the basal level. This observation suggests that pre-positioning of the enhancer next to the promoter enables a quick transcriptional response to the induction signal (Figure 1B).

Based on the transcriptional state of the associated gene, enhancer regions acquire different chromatin modifications. We measured the enrichment of H3K4me1, H3K27Ac, and H3K27me3 at *Dnmt3b* regulatory regions pre and post-differentiation. At D3 post-differentiation, we observed an expected increase in H3K4me1 and H3K27Ac concomitant with enhancer

activation and induction of *Dnmt3b* expression. Repression of *Dnmt3b* at D6 post-differentiation was associated with a decrease in H3K4me1 and H3K27Ac at both proximal and distal enhancer regions (Figures 1C and 1D). In contrast, H3K27me3 at the proximal and distal enhancer regions was higher in undifferentiated 2iESCs, followed by a decrease post-differentiation at D3 and D6. It is noteworthy that deacetylation of H3K27 at D6 post-differentiation is not followed by H3K27 methylation at these sites. H3 occupancy shows no significant difference in enrichment in various samples pre- and post-differentiation, further validating the significance of the observed chromatin modification changes (Figure S1G).

Previous studies have shown that during enhancer silencing, histone demethylation of H3K4me1 and deacetylation of H3K27Ac by Lsd1-Mi2NURD- complex poises the chromatin for DNMT3A activity (Petell et al., 2016). Therefore, we asked if DNA methylation at enhancer regions maintained the repression of *Dnmt3b* post-differentiation. Methylation-dependent qPCR (MD-qPCR) showed a significant gain of DNA methylation at distal and proximal enhancer regions on D6 post-differentiation (Figure 1F). To directly test the role of DNMT3A and DNMT3L, we compared DNA methylation in the WT, *Dnmt3a*$^{-/-,}$ and *Dnmt3l*$^{-/-}$ cells. We observed that compared to WT and *Dnmt3l*$^{-/-}$ cells, the *Dnmt3a*$^{-/-}$ cells showed a severe defect in gaining DNA methylation (Figure 1F). The *Dnmt3b* promoter also harbors a CpG island (CGI) around TSS, and we asked if DNA methylation spreads into the CGI. Therefore, we designed primers within the promoter and proximal enhancer region closer to the CGI. As previously reported for strong CGIs, the DNA methylation levels decreased with increasing proximity to the *Dnmt3b* promoter CGI (Figure 1G). However, unlike in the proximal and distal enhancers regions (Fig 1D), the CGI promoter region showed a recovery of H3K27me3 at D6 post-differentiation, ensuring the repressed state of the *Dnmt3b* promoter (Figure 1H). Overall, these data establish the activity of distal and proximal enhancers in regulating *Dnmt3b* expression during 2i-ESC differentiation.

**_Dnmt3bas_ transcript regulates the magnitude of _Dnmt3b_ induction post-differentiation.**

The divergent antisense (*as*) long non-coding RNA (lncRNA), *Dnmt3bas*, is a 2.8 kb transcript that initiates 800 bps downstream of TSS in the *Dnmt3b* promoter (Figure 2A). Sequence analysis of *Dnmt3bas* revealed a potential exon-intron structure. Structural modeling of the spliced *Dnmt3bas* using RNA Fold webserver (ViennaRNA) also showed a stem structure typical of lncRNAs (Figure S2A, S2B). Based on the predicted exon-intron boundaries, we designed primers to capture potential spliced and polyadenylated forms of *Dnmt3bas*. An expected unique band of nearly 716 bps, which corresponds to the size of spliced *Dnmt3bas*, confirmed splicing and polyadenylation of *Dnmt3bas* (Figure 2B). Subcellular localization assays showed that *Dnmt3bas* is predominantly present in the nucleus (Figure 2C).

We first asked if there is a correlation between the expression of *Dnmt3b* and *Dnmt3bas* during 2i medium adaptation and differentiation of ESCs. We used exon-specific primers to detect the expression of spliced *Dnmt3bas*, which also minimizes potential amplification from genomic DNA contamination in the RNA samples (Figure S2C). The RT-qPCR specificity was ensured by the absence of signal in the No-RT (no reverse transcriptase) control and by visualizing the amplified product as a single band on an agarose gel (Figure S2D). Expression analysis showed that *Dnmt3b* was downregulated at passage P6 and maintained a basal level in 2i-ESCs. Interestingly, *Dnmt3bas* expression constantly increased during the adaptation process (Figure S2E). However, during 2i-ESC differentiation, the expression of *Dnmt3bas* was downregulated concomitant with induction of the *Dnmt3b* (Figure 2D).

The coordinated yet contrasting expression pattern of the *Dnmt3b*/*Dnmt3bas* pair suggests a potential role for *Dnmt3bas* in regulating *Dnmt3b* transcription. Therefore, we generated stable ESCs lines expressing anti-*Dnmt3bas* shRNA (KD) or overexpressing (OE) spliced and unspliced forms of *Dnmt3bas* (Figure S2F). Post adaptation to 2i medium, the *Dnmt3bas* KD cells showed a significant reduction of *Dnmt3bas* by nearly 2.5 fold compared to the vector

control (VC) cells. Conversely, *Dnmt3bas* OE cells showed nearly 8-9-fold higher expression of *Dnmt3bas* compared to VC cells (Figure 2E). Next, we differentiated the transgenic cell lines and analyzed the expression of *Dnmt3b* (Figure 2F). Similar to our observation in Figure 1A, the VC cells showed a 15-fold induction of *Dnmt3b* at D3 before reaching the basal level at D6 post-differentiation. Interestingly, in the KD cells, the induction of *Dnmt3b* at D3 increased to nearly 30-fold and returned to the basal level at D6 post-differentiation. In OE cells, an opposite effect on *Dnmt3b* induction was observed, with only a 7-fold increase in *Dnmt3b* expression on D3 post-differentiation, supporting a predicted trans effect of *Dnmt3bas* transcript on *Dnmt3b* induction. The repressive activity of *Dnmt3bas* was specific for only the spliced isoform since cells overexpressing the unspliced *Dnmt3bas* behaved similar to VC cells. Notably, modulation of *Dnmt3bas* expression specifically affected the induced expression of *Dnmt3b* without changing its basal expression pre- and post-differentiation (Figure 2F). Moreover, this effect on *Dnmt3b* induction was rescued by overexpression of *Dnmt3bas* in KD cells (KD-OE), confirming the trans effect of *Dnmt3bas* transcript on *Dnmt3b* induction post-differentiation (Figure 2G).

These data suggest that *Dnmt3bas* affects the kinetics and the magnitude of *Dnmt3b* induction. To test this, we performed a time-course experiment by collecting samples each day post-differentiation and measured the temporal increase in transcript levels of *Dnmt3b*. The data were fit to a non-linear regression using GraphPad Prism. Whereas an apparent change in the magnitude of induction was recorded, it peaked at the same time point in all three cell lines, suggesting the role of *Dnmt3bas* in controlling the level of *Dnmt3b* expression (Figure 2H). We also ruled out the potential confounding impact of impaired pluripotency in the *Dnmt3bas* manipulated cell lines (Figure S2F).

Collectively, these data confirm the role of the spliced *Dnmt3bas* transcript in *Dnmt3b* transcriptional induction with no effect on the basal expression of the gene. These observations suggest a mechanistic model in which the repressive effect of *Dnmt3bas* maintains *Dnmt3b*

promoter/proximal enhancer in the primed state in undifferentiated cells to fine-tune the magnitude of *Dnmt3b* induction in response to differentiation signals.

***Dnmt3bas modulates the chromatin modification state at the Dnmt3b regulatory regions.***

To understand the mechanism of *Dnmt3bas* in the regulation of *Dnmt3b* induction, we measured chromatin modifications, i.e., H3K4me1, H3K27me3, and H3K27Ac at *Dnmt3b* regulatory regions during differentiation and compared changes in KD and OE cells to VC cells. In all three cell lines, an increase in H3K4me1 was observed at both proximal and distal enhancer regions at D3 post-differentiation compared to the undifferentiated (UD) cells (Figure 3A). Similarly, the loss of H3K27me3 was accompanied by the gain of H3K27Ac at the proximal and distal enhancer region in D3 cells (Figure 3B). Interestingly, we observed a lower enrichment of H3K27Ac on D3 in OE cells than in KD and VC cells (Figure 3C), corresponding to the alleviated induction of *Dnmt3b* gene expression (Figure 2F).

We next measured H3K27me3 and H3K27Ac close to CGI at the Dnmt3b promoter. Our data showed that H3K27me3 demethylation is accompanied by H3K27 acetylation during the differentiation of all three cell lines (Figure 3D, 3E). However, compared to the VC cells, we observed a lower and a higher enrichment of H3K27me3 in undifferentiated KD and OE cells, respectively (Figure 3D). Expectedly, an opposite trend in H3K27Ac was observed in these cells (Figure 3E), conforming to the difference in Dnmt3b induction. We also tested the potential effect of *Dnmt3bas* on DNA methylation at the promoter and proximal enhancer (PE) region. Similar to the VC cells, a gain of DNA methylation was observed in promoter regions farther from the CGI, except that a small but significant increase was prominent in OE cells at D6 post-differentiation (Figure 3F).

Overall, these data demonstrate that manipulating *Dnmt3bas* levels significantly affects the deposition of H3K27me3 at the promoter/PE region, affecting the gain of H3K27Ac and induction of *Dnmt3b* post-differentiation. We speculate that *Dnmt3bas* recruits the PRC2

complex to *Dnmt3b* cis-regulatory regions, similar to several other lncRNAs known to escort the PRC2 complex to their target sites (listed in (Schertzer et al., 2019; Statello et al., 2021)). The absence of predicted changes in H3K27me3 at the distal enhancer in *Dnmt3bas* KD and OE cells is possibly due to the low enrichment of this modification, limiting its detection.

**_Dnmt3bas_ is localized at the _Dnmt3b_ promoter and enhancer and affects E-P looping.**

Given that *Dnmt3bas* regulates the chromatin modification state of Dnmt3b regulatory regions, we asked if that affects the distal enhancer-promoter interaction. Therefore, we performed a 3C assay in VC, KD, and OE cell lines. The data show little or no effect on the interaction frequency of *Dnmt3b* distal enhancer and promoter in the undifferentiated 2i-ESCs (Figure 4A). However, we observed a higher interaction frequency between enhancer and promoter in KD cells than in VC cells (Figure 4B). These data suggest that an increase in H3K27Ac at *Dnmt3b* regulatory regions reinforces the E-P loop formation in KD cells, potentially affecting the induction level of the *Dnmt3b* gene.

Based on the observation that *Dnmt3bas* affects H3K27Ac at the distal enhancer and its interaction with the promoter, we asked if *Dnmt3bas* RNA is associated with the chromatin near the distal enhancer. We performed the Chromatin Isolation by RNA purification (ChIRP) assay in 2i-ESCs pre and post-differentiation. Using probes specific to exon or intron regions of *Dnmt3bas*, RNA recovery analysis showed a 10-fold higher enrichment of spliced than unspliced *Dnmt3bas* (Figure S4A). A high and specific recovery of the spliced *Dnmt3bas* was measured from the RNA fraction of undifferentiated and D3 differentiated cells (Figure 4C). From the DNA fraction, approximately 10-fold higher enrichment of the *Dnmt3b* promoter region was measured compared to the distal enhancer. We also observed an increased recovery of both *Dnmt3bas* and *Dnmt3b* promoter at the D3 in the RNA and DNA fractions, respectively.

These data confirm that spliced *Dnmt3bas* is the predominant transcript recruited to the *Dnmt3b* promoter and enhancer. An increase in *Dnmt3bas* enrichment at D3 post-differentiation suggests an additional role of the lncRNA during transcriptional induction of *Dnmt3b*.

### ***hnRNPL binds Dnmt3bas* at a CA repeat region in exon 2**

To determine the mechanism of *Dnmt3bas* in *Dnmt3b* promoter priming and transcriptional induction, we performed an RNA pull-down assay to identify *Dnmt3bas* interacting proteins. *In vitro* transcribed biotinylated *Dnmt3bas* and control antisense-*Dnmt3bas* RNAs were immobilized on streptavidin beads and incubated with nuclear extract from 2i-ESCs (Figure 5A). The bound proteins were visualized on a Coomassie-stained SDS-PAGE. A distinct protein band at around 60 KDa was explicitly observed in the *Dnmt3bas* elution and was absent in the control sample (Figure S5A). We performed mass spectrometry (MS) analysis of proteins from control and experimental lanes within the excised fragment to identify proteins in the specific band. The data from MS showed a strong enrichment of hnRNPL in the experimental sample. The identity of this band as hnRNPL was confirmed by the western blot of the RNA pull-down protein samples (Figure 5B). To examine the *in vivo* binding of hnRNPL to *Dnmt3bas*, we performed western blot analysis of the protein fraction from the *Dnmt3bas* ChIRP assay. The data showed a strong signal of hnRNPL in the experimental sample that was absent in the RNase treated control sample (Figure 5C). ChIP assay using anti-hnRNPL antibody showed significant and specific enrichment of hnRNPL at the *Dnmt3b* promoter, confirming its *in vivo* binding at these sites (Figure 5D).

hnRNPL binds CA repeats with high affinity (Smith et al., 2013), and *Dnmt3bas* contains two clusters of CA-repeats within exon 2. To locate the hnRNPL binding sites, we performed RNA pull-down assays using *Dnmt3bas* fragments (Figure 5E). Compared to a weak binding in the exon 1 fragment, strong hnRNPL binding was observed in the exon 2 and the intermediate RNA fragment containing both CA clusters (Figure 5F). Next, we systematically mutated the predicted

hnRNPL binding sites in the full-length *Dnmt3bas* and used the variants for the RNA pull-down assay. Whereas hnRNPL binding of variant 1 is similar to that of WT *Dnmt3bas*, a very weak binding was detected for variants 2 and 3 (Figure 5G). Variant 1 has mutations in cluster 2, and variants 2 and 3 have mutations in both clusters, indicating that hnRNPL predominantly binds to the cluster 1 site in exon 2 of *Dnmt3bas*. Thus, we conclude that hnRNPL specifically binds *Dnmt3bas* and is recruited to the *Dnmt3b* promoter region.

**_Dnmt3bas_ modulates alternative splicing of _Dnmt3b_ through hnRNPL activity.**

*Dnmt3b* transcripts comprise two major alternatively spliced isoforms, *Dnmt3b1*, a full-length transcript, and *Dnmt3b6*, in which exons 22 and 23 are excluded (Figure 6A) (Ref). Whereas, *Dnmt3b1* codes for an active DNA methyltransferase enzyme, *Dnmt3b6* protein is inactive due to the absence of motifs critical for catalytic activity. Both isoforms are expressed in ESCs grown in FBS media conditions. Previous studies have shown that the RNA-binding protein, hnRNPL regulates alternative splicing (33-35). Therefore, we asked if *Dnmt3bas* had a role in the AS of *Dnmt3b.* A primer pair in exons were used in an RT-PCR to amplify both isoforms from the VC, KD, and OE samples (Figure 6A). Interestingly, we noticed that in undifferentiated 2i-ESC, the basal expression of *Dnmt3b* mainly comprised of the shorter isoform *Dnmt3b6*, indicating that exon exclusion is preferred. However, upon *Dnmt3b* induction, the alternative splicing favors exon inclusion producing *Dnmt3b1* as a major isoform (Figure 6B, S6A). Although this trend was prevalent in all the three cell lines, we observed a peculiar increase in the ratio of long to short isoform (*Dnmt3b1*: *Dnmt3b6*) in OE cells at D3 post-differentiation. An opposite effect on the isoform ratio was observed in the KD cells, albeit with lower significance. These results suggested the role of hnRNPL in exon inclusion of *Dnmt3b* transcript. To directly test the role of hnRNPL in the AS of *Dnmt3b*, we transfected ESCs with *Dnmt3bas* variants 2 and 3, with mutations in hnRNPL binding sites. Transgenic cell lines were differentiated, and alternative splicing of *Dnmt3b* was analyzed at D3 post-differentiation. Compared to cells

overexpressing WT *Dnmt3bas*, the ratio of *Dnmt3b1*:*Dnmt3b6* expression in V2 and V3 variants expressing cells was similar to that in the VC cells (Figure 6C, 6D). This data confirm the direct role of the hnRNPL/*Dnmt3bas* complex in regulating Dnmt3b alternative splicing.

Based on our data, we speculate that post-induction, the *Dnmt3bas*-hnRNPL complex facilitates the interaction of hnRNPL with elongating RNA Pol II, which ferries hnRNPL to alternatively spliced sites. In *Dnmt3bas* OE cells, more hnRNPL is recruited at the promoter and binds to RNA Poll, increasing the number of exon inclusion events. Previous studies have shown that the interaction of SETD2 with hnRNPL and elongating RNA Pol II couples transcriptional elongation and alternative splicing (Bhattacharya et al., 2021a; Bhattacharya et al., 2021b; Yuan et al., 2009). We observed no significant change in the deposition of H3K36me3 at the 3' end of the *Dnmt3b* gene in *Dnmt3bas* OE cells compared to the VC cells (Figure S6B). However, this observation does not exclude the role of SETD2 in mediating the interaction of hnRNPL with elongating Pol II. However, it confirms that the activities of SETD2 and hnRNPL are not dependent on their interaction with each other (Yuan et al., 2009).

***A mechanistic model of Dnmt3bas***

Our study comprehensively examined the regulatory mechanism that allows for the spatiotemporal expression of *Dnmt3b* during early development. The data revealed the unique role of *Dnmt3bas* in both transcriptional priming and optimal activation of the *Dnmt3b* gene and regulation of alternative splicing of the Dnmt*3b* transcript. In the undifferentiated 2i-ESCs, *Dnmt3bas* maintains the *Dnmt3b* promoter in a primed state by regulating PRC2 activity. When differentiation is triggered, *Dnmt3b* expression is induced by enhancement of E-P interaction and deposition of H3K27Ac. Higher expression of *Dnmt3b* is also accompanied by increased exon inclusion and switching of alternative splicing in favor of *Dnmt3b1* transcription. The alternative splicing switch is facilitated by hnRNPL, a *Dnmt3bas* binding protein. At the transcriptionally active *Dnmt3b* promoter, *Dnmt3bas* bridges the interaction of hnRNPL with

elongating RNA Pol II, potentially via SETD2, ensuring its delivery to alternative splice sites. Following induction, the post-differentiation repression of *Dnmt3b* is mediated by DNA methylation at proximal and distal enhancers. Notably, the CGI promoter regains the H3K27me3 mark suggesting the role of a lower but persistent expression of *Dnmt3bas*, which is essential for stabilizing the activity of the PRC2 complex at the CGI promoter.

## Discussion

Gene expression is controlled by the concerted activity of the transcriptional coactivator complex and a permissive chromatin environment at its cis-regulatory elements. Additionally, divergent long non-coding RNAs (lncRNA) regulate the expression of the nearby protein-coding genes, including those crucial for normal development. LncRNAs act as modulators of gene expression through multiple mechanisms, including the recruitment of chromatin remodelers (Akhade et al.; Guttman et al.). Transcriptional regulation by lncRNAs can be accomplished by either the transcript itself or by the act of transcription, which creates a chromatin environment that influences the expression of the neighboring gene (Gil and Ulitsky, 2020). This study elucidated how *Dnmt3b* expression is regulated by the combined activity of the lncRNA, *Dnmt3bas*, and enhancer elements. Our study characterized *Dnmt3bas* as a promoter-associated divergent lncRNAs, expressed coordinately with *Dnmt3b* mRNA. Mechanistically, *Dnmt3bas* acts in trans and regulates the induction and alternative splicing of *Dnmt3b*.

### ***Dnmt3b expression in ESC model system***

Serum-cultured ESCs are a mix of two interchangeable cell populations; naïve pluripotent cells and epiblast stem cells. Epiblast stem cells are primed for differentiation, and their gene expression profile shows the expression of early differentiation genes. In contrast, naïve ESCs have high pluripotency genes expression (Tosolini and Jouneau, 2016; Ying et al., 2008). Interestingly, in contrast to serum-ESCs, the expression of *Dnmt3b* is strongly downregulated in 2i-ESCs. Upon induction of differentiation, the transcript levels of *Dnmt3b* increase substantially

within a small window of 2-3 days, after which the *Dnmt3b* gene is completely repressed (Ficz et al., 2013; Leitch et al., 2013). This is akin to the dynamic expression of *Dnmt3b* during early embryogenesis that is restricted to the inner mass cells from E4.5-E8.5, after which *Dnmt3b* expression is strongly downregulated in somatic lineages (Hirasawa and Sasaki, 2009; Okano et al., 1999; Watanabe et al., 2002). We used 2i-adapted ESC differentiation to study the dynamic regulation of *Dnmt3b* expression by the combined activity of its cis-regulatory elements and *Dnmt3bas*. During 2i medium adaptation and ESC differentiation, we observed a coordinated yet contrasting pattern of *Dnmt3b* and *Dnmt3bas* expression with the *Dnmt3b* pattern mimicking the expression of *Dnmt3b* observed *in vivo*. The observation that *Dnmt3bas* expression is lowest when the *Dnmt3b* gene is expressed at high levels suggests a potential transcriptional interference mechanism wherein a high rate of sense transcription at the *Dnmt3b* promoter will inhibit the antisense transcription of *Dnmt3bas* (Shuman, 2020). Nevertheless, post-differentiation repression of *Dnmt3b* is not accompanied by an increase in *Dnmt3bas* expression, suggesting a mechanism different from transcriptional interference and confirming the regulatory role of *Dnmt3bas* in maintaining the primed state of the *Dnmt3b* promoter during the pluripotent state.

***Role of Dnmt3bas in Dnmt3b transcriptional induction***

Given that *Dnmt3bas* is expressed within the promoter of the *Dnmt3b* gene, genetic manipulations using CRISPR/Cas9 could not be employed since it could have a confounding effect on *Dnmt3b* expression. However, we observed a significant effect of *Dnmt3bas* on *Dnmt3b* expression by using shRNA-mediated knockdown and overexpression of *Dnmt3bas* RNA in the 2i-ESCs. Whereas the overexpression of *Dnmt3bas* suppresses the magnitude of *Dnmt3b* induction, knockdown of *Dnmt3bas* has an opposite effect, complementing the observed inverse correlation between their expression pre-induction and during *Dnmt3b* induction. The chromatin modification state of the *Dnmt3b* cis-regulatory elements shows the

impact of *Dnmt3bas* expression on H3K27me3 at CGI promoter and enhancer elements. Our data suggest that *Dnmt3bas* recruits the PRC2 complex to maintain the primed state of the *Dnmt3b* promoter and enhancer elements in undifferentiated 2i-ESCs. In response to the differentiation signal, a concomitant loss of H3K27me3 precedes the gain of H3K27Ac for activation of *Dnmt3b* enhancers, which engage strongly with the promoter to facilitate transcriptional induction.

Some lncRNAs can help stabilize the E-P loop by interacting with components of the Mediator complex like Med1 and Med12 (Luo et al., 2016) or cooperate with CTCF-mediated chromatin interactions to affect the transcription of genes (Saldana-Meyer et al., 2019; Xiang et al., 2014; Yang et al., 2015). However, we observed an increase in the enhancer-promoter (E-P) interaction and H3K27Ac at the distal enhancer in *Dnmt3bas* KD cells, suggesting the role of *Dnmt3bas* in stabilizing the repressive PRC2 complex binding. Furthermore, the proximity of the distal enhancer to the Dnmt3b promoter in undifferentiated cells implies a dual role of the E-P loop, i) in facilitating the recruitment of Dnmt3bas and PRC2 complex and ii) mediating a quick transcriptional response to the differentiation signal.

Interestingly neither the expression of *Dnmt3bas* nor enrichment of H3K27me3 at distal regulatory elements increased during the post-differentiation repression of *Dnmt3b*. Instead, we observed a loss of H3K27Ac and a gain of DNA methylation in regions farther upstream and flanking the 5’ end of the CGI promoter. We also showed that DNA methylation was deposited by the methyltransferase Dnmt3a and its allosteric activator protein, Dnmt3l. However, at the CGI promoter, which is maintained in an unmethylated state post differentiation, we observed an increase in H3K27me3, suggesting the role of the PRC2 complex in maintaining the *Dnmt3b* promoter in a repressed state post-differentiation. This observation also justifies the low but significant expression of *Dnmt3bas* post-differentiation that could be involved in maintenance of H3K27me3 at the CGI. Interestingly in the OE cells, DNA methylation at the proximal enhancer

region also increasedpost-differentiation, suggesting positive reinforcement of the repressive pathways. These observations are comparable with recent studies showing that the lncRNA potency in spreading the PRC2 complex is tightly linked to lncRNA abundance, genome architecture, and CpG island DNA (Schertzer et al., 2019).

***Role of Dnmt3bas in Dnmt3b alternative splicing***

In ESCs, alternative splicing of the *Dnmt3b* transcript results in two main isoforms: *Dnmt3b6*, exons 22 and 23 are excluded, and *Dnmt3b1*, the full-length transcript (Gowher et al., 2008). Our data here show for the first time that *Dnmt3b* transcripts undergo a splicing switch in 2i-ESCs post-differentiation. Previous studies reporting *Dnmt3b1* as a major isoform used ESCs cultured in FBS. However, the 2i-ESCs, which are a homogenous population of cells in the ground state of pluripotency, express catalytically inactive *Dnmt3b6* as the major isoform, similar to the somatic expression of *Dnmt3b*3. These data suggest that besides downregulated expression, alternative splicing of Dnmt3b contributes to the genome-wide loss of DNA methylation in 2i-ESCs. Concurrently, basal transcription could also retain the Dnmt3b promoter in a primed state for prompt activation. However, this would also necessitate the promoter activation process to append a mechanism that facilitates exon inclusion to generate the full-length *Dnmt3b1*.

Interestingly, among several unique proteins bound to *Dnmt3bas*, RNA pull-down assays identified hnRNPL with the highest specific enrichment. HnRNPL functions as a regulator of alternative splicing (Cole et al., 2015; Hung et al., 2008; Preussner et al., 2012), RNA stability (Hui et al., 2003), and transcriptional activation or repression in partnership with multiple lncRNAs (Atianand et al., 2016; Li et al., 2014; Ruan et al., 2016). Previous studies have shown that the binding of hnRNPs, including hnRNPK and hnRNPU and U1 snRNP binding, facilitate chromatin tethering and nuclear retention of lncRNAs (Hacisuleyman et al., 2016; Lubelsky and Ulitsky, 2018; Shukla et al., 2018; Yin et al., 2020). The interaction of hnRNPL with *Dnmt3bas* at

the *Dnmt3b* promoter was also confirmed by ChIP and ChIRP assays. Med23 was also shown to recruit hnRNPL to the promoter of genes, a role similar to *Dnmt3bas* (Huang et al., 2012). These observations suggest that the nucleoprotein constitution of promoters can determine both transcription and splicing outcomes (Cramer et al., 1997). Functional characterization of *Dnmt3bas*/hnRNPL interaction showed increased inclusion of exons 22 and 23 in *Dnmt3b* transcript in the *Dnmt3bas* OE cells was reversed in cells overexpressing *Dnmt3bas* with a mutation in the hnRNPL binding site. This observation suggests that hnRNPL is the regulator of inducible exon inclusion, a process critical for expressing catalytically active, *Dnmt3b1* during differentiation.

***Implications of Dnmt3bas-hnRNPL interaction***

Recent studies have shown the interaction of SETD2 with hnRNPL and proposed that SETD2 hitchhikes to the elongating Pol II and transports hnRNPL to the splice sites as they emerge during chain elongation(Yuan et al., 2009) (Bhattacharya et al., 2021a; Bhattacharya et al., 2021b). Our study suggests that recruitment of hnRNPL by *Dnmt3bas* to the promoter of *Dnmt3b* facilitates hnRNPL- SETD2 transfer. Given that the RNA binding domain of hnRNPL, RRM2, also interacts with SETD2, the interaction of hnRNPL with SETD2 and *Dnmt3bas* could be mutually exclusive. This prediction is supported by the absence of SETD2 in the RIP/Mass spectrometry analysis of *Dnmt3bas* binding proteins. Therefore, interaction choice could be dictated by proximity and binding strength of the RRM2 with RNA versus SETD2.

In contrast to hnRNPL, notably, the knockdown of hnRNPI (PTBP1) increases exon inclusion in the *Dnmt3b* mRNA (Gowher et al., 2012), suggesting that the antagonistic activity of hnRNPI/hnRNPL potentially determines the splice site choice in *Dnmt3b* pre-mRNA. Both proteins bind to similar polypyrimidine tracts (CA repeats) on RNA through their RRM domains and interact with each other (Hahm et al., 1998). Interestingly PTBP1 is recruited to chromatin by H3K36me3 (Luco et al., 2010), a modification deposited by SETD2. We observed an

increase in H3K36me3 post-differentiation (Fig S6), which could contribute to hnRNPI/hnRNPL recruitment at splice sites. However, the mechanism by which their antagonistic activity at alternative splice sites is regulated is unknown. It would be interesting to determine the mechanism of potential cross-talk between antagonist splice factors and chromatin modifiers that fine-tunes the relative levels of the alternatively spliced isoforms in response to varying physiological conditions. Also, given that *Dnmt3bas* acts in trans, future studies will determine the potential global effect of its activity in regulating expression and alternative splicing of other genes.

**Acknowledgements**

We are thankful to Gowher lab members for discussions. This work was supported by NSF award 1716678. The authors gratefully acknowledge the Walter Cancer Foundation, DNA Sequencing Facility, and support from the Purdue University Center for Cancer Research, P30CA023168.

**Authors Contributions**

M.S.D., I.K.M., S.M., M.He, M.H. H. C. W., N.E.B., and M.C.,performed the experiments. M.H., H.G.. analyzed the mass spectrometry data. M.S.D, I.K.M. and H.G. wrote the manuscript

**Declaration of Interests**

None declared.

**Figure Legends**

**Figure 1. Proximal and distal enhancers regulate Dnmt3b induction.**

UD: undifferentiated; D1- D6: Days post-induction of differentiation; ESCs: embryonic stem cells; RA: Retinoic Acid; LIF: Leukocyte Inhibitory factor; 2i: Signaling pathway Inhibitors, CHIR99021 and PD184352; PE: proximal enhancer; Pro: promoter; DE: Distal enhancer, CGI: CpG island; WT: wild type; 3a: Dnmt3a; 3l: Dnmt3L

(A) RT-qPCR of Dnmt3b during 2iESC differentiation as illustrated. The threshold cycle (Ct) values were normalized to Gapdh and expression is shown relative to that in UD cells. The data show a steady increase in Dnmt3b expression up to D3 which decreases at D4 post differentiation.

(B) Chromatin conformation assay (3C) shows an interaction of the distal enhancer with the promoter of Dnmt3b. The location of the distal enhancer and promoter in the Dnmt3b locus, the 3C PCR primers, and the corresponding HaeIII sites (tick marks) are shown. The 3C PCR primers are numbered. qPCR analysis was done using the primer at the DE as an anchor (Chr2 position: 153466046) and interactions with downstream sites were measured in 2iESC at undifferentiated and D3 post differentiation state. Plotted are the relative interactions in arbitrary units of the Dnmt3b DE with six HaeIII sites. The X-axis represents the chromosome 2 location (153465120- 153477120) of these HaeIII sites. A significant increase in the interaction frequency was observed between DE and Pro of Dnmt3b gene at D3 post differentiation.

(C, D, E, H) Chromatin immunoprecipitation (ChIP)-qPCR shows percent enrichment over input. Histone modifications at Dnmt3b regulatory elements (C) H3K4me1 (D) H3K27Ac and (E, H) H3K27me3 in 2iESCs pre- and post-differentiation. Concomitant with Dnmt3b induction, a systematic increase and decrease in the H3K4me1 and H3K27Ac signal was observed post-differentiation. Whereas H3K37me3 demethylation is maintained post differentiation at the Pro/PE and DE, it is regained at the CGI region of the Pro, close to the TSS of the Dnmt3b gene. % Enrichment = fold enrichment over input x 100. p-values were derived from Student's t test: $^{**}p < 0.01$.

(F, G) DNA methylation was measured by MD-qPCR. The DNA was restricted using the methylation-dependent enzyme MspJI which cuts at methylated cytosines. The specified regions in the regulatory elements were amplified using the cleaved DNA as the template by qPCR. The Cq values for differentiated samples were normalized to that of the undifferentiated

sample. An increase in Cq value indicates a gain of DNA methylation. (F) Genomic DNA was collected from the WT, Dnmt3a$^{-/-}$ and Dnmt3l$^{-/-}$ ESC cell lines pre-and post- differentiation. DNA methylation analysis shows a reduced gain in both Dnmt3a$^{-/-}$ and Dnmt3l$^{-/-}$ cells. (G) DNA methylation analysis in 2iESCs pre and post-differentiation at Dnmt3b Pro and CGI regions. The tested CpG sites are represented on X-axis as base pairs upstream of the Dnmt3b TSS. The data show a steady gain of DNA methylation corresponding to the distance away from the CGI, which is maintained in an unmethylated state. Results are presented as normalized mean values ± SEM for n = 3. See also Figure S1.

**Figure 2. Spatiotemporal expression of Dnmt3b in dnmt3bas manipulated ESCs**

UD: undifferentiated; D1- D6: Days post-induction of differentiation; ESCs: embryonic stem cells; 3bas: dnmt3bas lncRNA; VC: Vector control; KD: shRNA mediated knockdown of dnmt3bas; OE: overexpression of dnmt3bas lncRNA;

(A) Illustration of Dnmt3b promoter flanking regions shows the relative position, exon/intron structure, and size of unspliced and spliced transcripts of the lncRNA dnmt3bas, which is transcribed in antisense direction within the Dnmt3b promoter.

(B) Ethidium bromide-stained agarose gel. RT-PCR of dnmt3bas using exonic (lane 1) and Oligo dT (lane 2) primers show a single band corresponding to the spliced transcript.

(C, D, E, G,) RT-qPCR analysis of dnmt3bas expression. (C), RT-qPCR of dnmt3bas from nuclear and cytosolic fractions of RNA showing about 2-fold higher enrichment in the nucleus compared to the cytosol. In contrast, Dnmt3b and Gapdh mRNA transcripts showed a significantly higher enrichment in cytosol compared to the nucleus as expected. (D) RT-qPCR shows dnmt3bas expression in 2iESCs pre- and post-differentiation, with the highest expression in the UD state. (E) RT-qPCR of dnmt3bas from transgenic 2i-ESC lines expressing dnmt3bas shRNA (KD), spliced and unspliced transcripts of dnmt3bas (OE), and spliced dnmt3bas in KD cells (-KD-OE). The Ct values were normalized to Gapdh and expression is shown relative to

that in vector control (VC) cells. (F, G) RT-qPCR of Dnmt3b from the VC, KD, and OE and KD-OE cells, pre-and post-differentiation. (F) The left Y-axis shows Dnmt3b expression data for UD and D6 cells whereas the right Y-axis shows the expression for the D3 cells. (G) In KD-OE cells, Dnmt3b induction is rescued to the level in VC cells on D3 post differentiation. (H) Dnmt3b expression in VC, OE, and KD cell lines was measured every 24 hours post differentiation. The data were fit to a non-linear regression in PRISM to estimate the time of the Dnmt3b induction peak. % Enrichment = fold enrichment over input x 100. p values were derived from Student's t test: *$p < 0.05$; **$p < 0.01$. Results are presented as normalized mean values ± SEM. See also Figure S2.

**Figure 3. Chromatin modification at Dnmt3b regulatory elements in dnmt3bas manipulated ESCs**

UD: undifferentiated; D3, D6: Days post-induction of differentiation; PE: proximal enhancer; Pro: promoter; DE: Distal enhancer; CGI: CpG island; VC: Vector control; KD: shRNA mediated knockdown of dnmt3bas; OE: overexpression of dnmt3bas lncRNA

(A, B, C, D, E) ChIP-qPCR assays show percent enrichment over input. Histone modifications at Dnmt3b regulatory elements (A) H3K4me1 (B, D) H3K27me3 and (C, E) H3K27Ac pre- and D3 post-differentiation in VC, KD, and OE cells. (A) On D3 post differentiation, an increase in the H3K4me1 signal was observed in KD and OE cells similar to the VC cells in Pro/PE and DE regions. (B, D) All cell lines show H3K37me3 demethylation on D3 post differentiation. However, a higher enrichment of H3K27me3 is observed in the UD state of OE cells at the Dnmt3b promoter including PE and CGI regions. (C, E) H3K27Ac shows an increase at D3 post differentiation, albeit less in OE cells at all the cis-regulatory elements. A significantly higher enrichment in KD cells compared to VC cells is also notable at CGI region on D3 post differentiation.

(F) DNA methylation at the Dnmt3b promoter was measured by MD-qPCR. After digestion of genomic DNA from VC, KD and OE cells pre and post-differentiation using FspE1 enzyme, the specified regions in the regulatory elements were amplified using qPCR. The Cq values for differentiated samples were normalized to that of the VC-UD sample. An increase in Cq value indicates a gain of DNA methylation. The tested CpG sites are represented on X-axis as base pairs upstream of the Dnmt3b TSS. The data show a steady gain of DNA methylation corresponding to the distance away from the CGI, which is maintained in an unmethylated state.

% Enrichment = fold enrichment over input x 100. p values were derived from Student's t test: **$p < 0.01$; ***$p < 0.005$. Results are presented as normalized mean values ± SEM. See also Figure S3.

**Figure 4. dnmt3bas associates with Dnmt3b enhancers and influences E-P looping**

UD: undifferentiated; D3: Days post-induction of differentiation; PE: proximal enhancer; Pro: promoter; DE: Distal enhancer; VC: Vector control; KD: shRNA mediated knockdown of dnmt3bas; OE: overexpression of dnmt3bas lncRNA

(A, B) Chromatin interaction assay, 3C, shows an interaction of the distal enhancer with the promoter of Dnmt3b. 3C assays were performed using undifferentiated and D3 differentiated VC, KD and OE cells. Primers as described in Figure 1 legend were used and DE primer was used as a bait. The data in (A) show no change in E-P interaction in the undifferentiated cells, whereas in (B) a stronger interaction was observed in KD cells at D3 post differentiation. (C, D) ChIRP (chromatin isolation by RNA purification) was performed using dnmt3bas specific biotinylated probes. The RNA and DNA fractions from the eluate were separated and used to probe for (C) dnmt3bas transcript and (D) Dnmt3b Pro/PE and DE regions respectively. (C) We observed prominent recovery of dnmt3bas compared to Gapdh mRNA. (D) From the DNA fraction, the Pro/PE region of Dnmt3b enriched 100-180-fold and DE showed 10-15-fold

recovery compared to a non-specific region in the mouse genome. Results are presented as normalized mean values ± SEM. See also Figure S4.

**Figure 5. hnRNPL binds to a CA repeat region in exon 2 of Dnmt3bas**

3bas: dnmt3bas transcript, A-3bas: Antisense of dnmt3bas transcript, bt-UTP: biotinylated UTP, IN: Input, NC: negative control, E1: Exon1, M: Regions from Exon 1 and Exon 2, E2: Exon2, Cluster 1 and Cluster 2: Predicted hnRNPL binding sites on Dnmt3bas exonic region, V1, V2: Variant of dnmt3bas transcript with substitutions in (V1) cluster 2, and (V2) cluster 1 of hnRNPL binding.

(A, E) Schematic showing the experimental design for in vitro transcription of (A) dnmt3bas, antisense control transcript and (E) various dnmt3bas fragments used for RNA pull down assays. The in vitro transcription was performed using biotinylated UTP and nuclear extract from 2i-ESCs. D) ChIP-qPCR assays show percent enrichment over input of hnRNPL. (B, C F, G) Western Blot analysis of hnRNPL; (B) RNA pull down assay showing specific binding of hnRNPL to dnmt3bas and not to antisense control transcript; (C) ChIRP (Chromatin Isolation by RNA Purification) showing hnRNPL in the protein fraction from the eluate that was absent in RNAse treated control. (F) RNA pull-down assays using fragments of dnmt3bas transcript shows strong and specific binding of hnRNPL to a region in E2. (G) Variants of Dnmt3bas were used for RNA Pull down assays.% Enrichment = fold enrichment over input x 100. Results are presented as normalized mean values ± SEM. See also Figure S5.

**Figure 6. hnRNPL regulates alternative splicing of Dnmt3b**

UD: undifferentiated; D3: Days post-induction of differentiation; VC: Vector control; KD: shRNA mediated knockdown of dnmt3bas; OE: overexpression of dnmt3bas lncRNA; V2: Variant 2; V3: variant 3; 3b1: Dnmt3b long isoform which includes exons 22 and 23; 3b6: Dnmt3b short isoform which excludes exons 22 and 23

A) Illustration of alternatively spliced exons, 22 and 23. Red arrows represent the position of primers used to amplify around the alternatively spliced exons of the Dnmt3b gene. B) RT-PCR analysis of *Dnmt3b1* and *Dnmt3b6* expression amplified in the same reaction by using primers flanking the AS exons. The bar graph shows ratio of the quantified band intensity of the long and the short isoforms of Dnmt3b. A switch in splice isoform from short to long is observed as cells are induced to differentiate, and is enhanced in OE cells. C) Ratio of *Dnmt3b1*/*Dnmt3b6* isoforms at D3 from cells expressing wildtype Dnmt3bas, Varient 2 (V2) and variant 3 (V3) mutant *Dnmt3bas*, and vector control. The bar graph shows ratio of the quantified band intensity of the long and the short isoforms of Dnmt3b. Results are presented as normalized mean values ± SEM. See also Figure S6.

**Figure 7. Model of epigenetic changes at pluripotency gene enhancers during stem cell differentiation.** In the undifferentiated state (UD), *Dnmt3bas* maintains the *Dntm3b* promoter in a primed state, where the proximal enhancer and promoter regions are characterized by H3K27Me/3 chromatin modifications, facilitating basal *Dnmt3b* expression. In response to differentiation signals (D3), *Dnmt3b* expression is induced via E-P interactions and the deposition of H3K27Ac marks. In concert with induced *Dnmt3b* expression, a switch in alternative splicing increases exon inclusion, promoting *Dnmt3b1* transcription. *Dnmt3bas* directly bind and recruit hnRNPL, an obvious splicing factor, to the transcriptionally active *Dnmt3b* promoter. Once at the promoter, hnRNPL hitchhikes RNA Pol II, potentially through an adaptor protein, ensuring its delivery to *Dnmt3b* splice sites. Post-differentiation (D6), DNA methylation by DNMT3A at the promoter and enhancers reduces the expression of *Dnmt3b* back to basal levels. In particular, the CGI promoter regains H3K27me3 suggesting the role of a lower but persistent expression of *Dnmt3bas*, which is required for stabilizing the activity of the PRC2 complex at the CGI promoter.

## STAR Methods

### Cell culture, adaptation, and differentiation:

As previously described by Alabdi et al., mouse embryonic stem cells (mESCs, E14Tg2A) were cultured and maintained in gelatin-coated tissue culture plates. Differentiation of mESCs was induced by plating $10X10^6$ cells in low attachment 10 cm petri dishes with a concurrent withdrawal of LIF. After Day 3, embryoid bodies were treated with 1µM Retinoic acid (RA), the medium was replenished every day and cells were harvested each day, and samples were collected for protein, RNA, DNA, and chromatin until Day 9.

### 2i Adaptation

Serum-mESCs were seeded in a T-25 flask and passaged once before adapting them into 2i media. Briefly, the serum-mESCs were washed with PBS and detached with trypsin. FBS was added to inactivate the trypsin, DMEM/F12 media was added, and cells were pelleted. The cells were again washed to remove any traces of FBS, and then finally plated and grown on gelatin-coated tissue culture plates in 2i media (3µM CHIR99021 and 1µM PD0325901). The cells were adapted until passage 10, and at every passage, the cells were harvested and checked for gene expression and DNA methylation. The differentiation was induced by removing LIF from the media, and at D3, 1µM RA was added as explained earlier. The samples were harvested at different time points for the analysis.

### Generation of transgenic cell lines

The different transgenic lines used in this study are 2i-cells stably expressing GFP as a vector control, shRNA against dnmt3bas, overexpressing dnmt3bas spliced, overexpressing dnmt3bas unspliced. These stable cell lines were generated by the treatment of antibiotic selection method over the course of time. The cells were selected using the antibiotic selection method and were

resistant to the antibiotics used. CRISPR KO dnmt3a and dnmt3l cell lines were procured from Dr. Taiping Chen, MD Anderson.

**Gene expression and alternative splicing analysis by qPCR and RT-PCR**

Briefly, total RNA was isolated using the TRIzol reagent (Invitrogen, 15596026) according to the manufacturer's protocol. RNA samples were digested with DNAse (Roche, 04716728001) at 37°C, for two hours, and purified using a Quick-RNA$^{TM}$ MiniPrep Plus Kit (ZymoReseach, R1057). Reverse-transcription quantitative PCR (RT-qPCR) was performed either using Verso One-Step RT-qPCR kits (Thermo Scientific, AB-4104A) or cDNA synthesized using the Tetro cDNA Synthesis Kit (Bioline, BIO-65043). For Verso, we used 5-50ng of RNA per reaction, and to synthesize cDNA, we used 1.2-2.5µg of RNA. cDNA was synthesized either through the gene-specific primers (GSP) or random hexamers or with oligo (dT) as per the manufacturer's instructions. Gene expression was calculated as $\Delta C_t$ which is $C_t$(Gene)- $C_t$(*Gapdh or Beta-actin*). Change in gene expression is reported as fold change relative to that in control undifferentiated cells, which was set to 1. Supplementary Table S1 details the list of primers used in this study. To analyze alternative splicing, a 1 step RT-PCR kit (Invitrogen, ) was used to reverse transcribe Trizol purified and DNAse (Roche, 04716728001) treated RNA from ESCs. We designed primers at exon-exon boundaries (View primer list) to range in product sizes of 200 – 500 bps.

**Chromosome Conformation Capture Assay (3C).**

3C assays followed the protocol from (Hagege et al., 2007; Naumova et al., 2012) with a few modifications. Murine 2i cells were crosslinked with 1% formaldehyde for 10 minutes at room temperature. Chromatin was suspended in NEB buffer 2.1 and digested with 1000 units HaeIII at 37°C overnight. Crosslinked fragments were ligated with 45 units T4 DNA ligase at 16 °C for 2 hours. A random ligation control template was generated using a BAC clone RP23-474F18

covering the Dnmt3b locus and digested with HaeIII and ligated with T4 ligase. Gapdh locus was used as an endogenous control. Compared to a known concentration of E14 genomic DNA by qPCR with primer pair GaploadF and GaploadR to determine 3C samples concentration. According to the result, adjust each 3C sample to the same template amount for the following quantitative PCR. Interaction frequency between anchor fragment (F1) and distant fragment detection was done in triplicate by quantitative PCR using 3C products and BAC products as a template and 3C primers. Primers position and sequence are indicated in (Fig 1) and Table S1. The relative interaction frequency of each fragment to anchor was calculated as $2^{(Ct\ BAC\ -Ct\ 3C)}$.

**RNA-Pulldown and hnRNPL binding assays**

The WT and mutant *Dnmt3bas* templates for *in vitro* transcription assays were obtained by PCR using primers containing the T7 promoter sequence (View primer list). Biotin-labeled *Dnmt3bas* transcripts, *Dnmt3bas* antisense,  and *Dntm3bas* variants 1, 2, and 3 were transcribed *in vitro* using the T7 RNA polymerase (ThermoFisher Scientific, AM1333) with the following modifications. Unlabeled ATP, GTP, and CTP were used at a concentration of 7.5 mM, except for unlabeled UTP, which was used at 5.63 mM. Notably, the reaction was supplemented with 1.8 mM biotin-UTP (Roche, 11093070910). Next, pre-cleared nuclear protein extracts were incubated with phenol-chloroform purified transcripts and streptavidin-coated magnetic beads for 45 mins at 4 $^{0}$C. Following washes, beads were eluted using SDS-loading buffer by boiling for 5 mins as previously described (Panda et al., 2016). Eluates were analyzed using Coomassie staining and Western blot.

**Subcellular fractionation**

20 million mESCs were harvested and used for the subcellular fractionation, according to (Pandya-Jones and Black, 2009; Wuarin and Schibler, 1994) protocol. Briefly, the cells were trypsinized and centrifuged at low speed at RT for 5min. Cells were lysed by adding the NP-40 lysis buffer for 5 mins. The samples were centrifuged and the supernatant was saved as the

cytosolic fraction. The pellet was rinsed with the lysed buffer, to remove any cytoplasmic leftover. The pellet was then resuspended in glycerol buffer by gentle flicking. This was followed by the addition of nuclei lysis buffer. After 2 minutes of incubation, the sample was centrifuged and the supernatant taken as the soluble nuclear fraction/nucleoplasm. Part of the fractions was used to perform the western blot, and the remaining fraction was used to isolate the RNA for RT-qPCR analysis.

**Chromatin Isolation by RNA Purification (ChIRP)**

Chromatin Isolation by RNA Purification (ChIRP) was performed by following (Au - Chu et al., 2012). Briefly >20million cells were grown and cross-linked 1% glutaraldehyde at room temperature, followed by quenching for 5 minutes. The cell pellets were flash freezed in liquid nitrogen and stored at -80 °C indefinitely. The cells were thawed at room temperature and dislodged. For 100mg of cell pellet, we used 1ml of lysis buffer containing Protease Inhibitor cocktail, PMSF, and SUPERase•In™ RNase Inhibitor (Invitrogen).The cells were sonicated using a Bioruptor in 15 ml tubes for 2-3 hours. The sonicated sample was centrifuged and the supernatant was hybridized with the biotinylated DNA probes specific to lncRNAs, Dnmt3bas for 4 hours. In the meantime, C-1 magnetic beads were washed and added to the probe-sample mix. The complex was incubated at 37°C for 30 minutes and then washed 5 times with the 1ml wash buffer. The beads were resuspended in the final wash, and 100ul of the sample was kept for RNA purification. The complex was processed accordingly for the downstream analysis. For RNA purification, the bead complex was resuspended in TriZol, and for DNA isolation, the complex was reverse cross-linked, and RNase treated. The purified ChIRP-DNA is confirmed for the genomic region by a qPCR. For isolation of proteins, the bead complex was used to elute the protein using a biotin elution buffer as described by the above protocol. The eluted protein samples were used for western blot to identify binding partners. Supplementary Table S2 lists sequences of probes used in this study.

## DNA Methylation-Dependent qPCR Assay (MD-qPCR)

Genomic DNA (gDNA) was purified using the standard phenol-chloroform method. The extracted gDNA was treated with RNAse (Roche) overnight at 37 C. Following another round of Phenol-Chloroform DNA extraction, gDNA was subjected to FspE1 (NEB, R0662S) digestion overnight at 37 C. The digested gDNA was purified by the Phenol-Chloroform method and quantified using the NanoDrop 3300 fluorospectrometer through PicoGreen dye according to the manufacturer's protocol (Life Technologies, P11495). Quantitative PCR was performed using an equal amount of DNA for each sample. The change in DNA methylation is represented by relative fold change in the Cq value as follows: 2ˆ(Cq(U)-Cq(I)), where Cq(U) is the Cq for the undifferentiated ESC sample, and Cq(I) represents day 3 or day 7 differentiated ESCs. The primers used for DNA methylation-dependent qPCR analysis have been previously described [30]. Standard deviations represent three technical and two biological replicates.

## Chromatin Immunoprecipitation

ChIP was performed as described (Petell et al., 2016). Briefly nuclei were isolated from the x-linked cells and were sonicated with the biorupture (instrument), according to the manufacturer's protocol. A total of 8µg of sheared crosslinked chromatin was incubated with 8µg of antibody pre-loaded on a 1:1 ratio of protein A and protein G magnetic beads (Life Technologies, 10002D and 10004D, respectively). After washing the beads, the samples were eluted in 1% SDS, 10 mM EDTA, 50 mM Tris-HCl, pH 8.0. Crosslinking was reversed by incubation at 65°C for 30 min with shaking. Samples were treated with RNase (Roche, 11119915001) for 2 h at 37°C, and subsequently treated with Proteinase K (Worthington, LS004222) for 2 h at 55°C. DNA was purified by phenol: chloroform extraction followed by ethanol precipitation and quantified using PicoGreen (Life Technologies, P11495) and NanoDrop 3300 fluorospectrometer. qPCR and data analysis was performed as previously described (Petell et al., 2016). Percent enrichment = Fold enrichment X 100. Significance of change was determined via *p*-value, which was

calculated by GraphPad Prism using Student's *t*-test. Supplementary Table S4 lists sequences of primers used.

**Western blot**

Standard western blot analysis was conducted on nuclear protein extracts and RNA-pulldown eluates using anti-hnRNP L, 1:1000 (Abcam, ab32680), anti-hnRNP K, 1:1000 (Abclonal, A1701), and anti-rabbit, 1:40,000 (Jackson Immunoresearch, 111-035-003). Chemiluminescence was performed according to the manufacturer's protocol (Thermo-Fisher Scientific 34580). For Dnmt3b expression, Protein samples were purified using standard RIPA protein purification (Peach et al., 2015), and blotted with anti-Dnmt3b antibodies (sc-20704), and β-actin (sc-47778) as a loading control. Images were taken using the ChemiDoc MP imaging system (Biorad, 170001402).

**Supplementary figures legends**

**Figure S1. Proximal and distal enhancers regulate Dnmt3b induction**

s-ESCs: serum cultured embryonic stem cells; 2i-ESCs: 2i cultured embryonic stem cells 2i: Signaling pathway Inhibitors, CHIR99021, and PD184352; UD: undifferentiated; D3, D6: Days post-induction of differentiation: U: Uncut; H: Hpa II cut; M: Msp I cut

A, D) Relative expression analysis by RT-qPCR of A) various early developmental genes in 2iESCs compared to expression in s-ESCs D) Dnmt3b expression in s-ESC pre and post differentiation. The Ct values were normalized to Gapdh and expression is shown relative to that in UD cells. The data are an average and standard deviation of three biological replicates. B) Alkaline phosphatase staining and SSEA-1 immunofluorescence of s-ESCs and 2i-ESCs. A positive signal indicates pluripotency, which is strong in 2i-ESCs. The scale bar is 100 µm. C) Genomic DNA from 2iESC and s-ESCs was digested with restriction enzymes Hpa II, which cuts unmethylated DNA at the 5’-CCGG-3’ site, and MspI which cuts DNA at the same site

irrespective of the DNA methylation. E) Schematic of the Dnmt3b gene showing the position of the promoter-associated lncRNA Dnmt3bas and upstream enhancer. F) ChIP-qPCR assays show percent enrichment over input of H3.

% Enrichment = fold enrichment over input x 100.

**Figure S2. Spatiotemporal expression of Dnmt3b in Dnmt3bas manipulated ESCs**

A) Sequence of dnmt3bas gene showing exons (red) and one intron (green). B) Predicted minimum free energy structure of spliced Dnmt3bas transcript, determined using ViennaRNA RNA Fold webserver. The structure is colored by base-pairing probabilities. For unpaired regions, the color denotes the probability of being unpaired, from blue to red going from the highest to the lowest. C) Relative gene expression of lncRNAs, Dnmt3bas and Evx1as and mRNAs Dnmt3a, Dnmt3b, and Prdm14 in 2i-ESCs compared to s-ESCs. D) Ethidium bromide-stained 6% acrylamide gel showing a specific amplification in the qRT-PCR reactions with no signal in the reactions without Reverse Transcriptase (NRT). E) Relative gene expression of Dnmt3b and Dnmt3bas during adaptation (P5-P10) of s-ESCs to 2i medium conditions. F) Relative expression of master pluripotency genes pre and post differentiation of the transgenic cell lines, VC, KD, OE. G) Illustration of the transgenic cell line generation and adaptation to 2i media conditions. The bar graph shows an increase in the expression of prdm14 indicating successful adaptation to 2i media conditions.

**Figure S3. Chromatin modification at Dnmt3b regulatory elements in dnmt3bas manipulated ESCs**

ChIP-qPCR assay showing percent enrichment over input of H3K27me3 histone modifications at Dnmt3b regulatory elements. Enrichment = fold enrichment over input x 100.

**Figure S4. Dnmt3bas associates with Dnmt3b enhancers and influences E-P looping**

ChIRP was performed using Dnmt3bas biotinylated probes specific for spliced and unspliced transcript. The RNA fraction from the eluate were separated and used to probe for spliced and unspliced Dnmt3bas transcript. The data show about 10 fold higher enrichment of spliced transcript compared to the unspliced.

**Figure S5. hnRNPL binds to a CA repeat region in exon 2 of Dnmt3bas**

Coomassie stained SDS-PAGE showing proteins eluted after RNA pull down using Dnmt3bas or anti-Dnmt3bas control transcript. The boxes represent the gel slices that were used for extraction of proteins and detection by mass spectrometry. A prominent band is visible in Dnmt3bas lane compared to the control transcript.

**Figure S6. Dnmt3bas associates with Dnmt3b enhancers and influences E-P looping**

(A) Western blot showing the expression of Dnmt3b in undifferentiated (UD) ESCs, Day 3 and Day 6 post-differentiation (D3 and D6 respectively) cells. β-actin was used as loading control.
(B) ChIP-qPCR assay showing percent enrichment over input of H3K36me3 histone modifications at the Dnmt3b gene locus. Enrichment = fold enrichment over input x 100.

**Table S1. Primers used in this study.**

A list of all PCR primers used in this study (5’ to 3’), separated by technique

**Table S2. ChIRP probes used in this study.**

A list of all ChIRP probes used in this study (5’ to 3’) with 3' biotin modification

817

818 Akhade, V.S., Pal, D., and Kanduri, C. (2017). Long Noncoding RNA: Genome Organization and
819 Mechanism of Action. Adv Exp Med Biol *1008*, 47-74.

820 Atianand, M.K., Hu, W., Satpathy, A.T., Shen, Y., Ricci, E.P., Alvarez-Dominguez, J.R., Bhatta, A.,
821 Schattgen, S.A., McGowan, J.D., Blin, J.*, et al.* (2016). A Long Noncoding RNA lincRNA-EPS Acts as a
822 Transcriptional Brake to Restrain Inflammation. Cell *165*, 1672-1685.

823 Au - Chu, C., Au - Quinn, J., and Au - Chang, H.Y. (2012). Chromatin Isolation by RNA Purification (ChIRP).
824 JoVE, e3912.

825 Bhattacharya, S., Levy, M.J., Zhang, N., Li, H., Florens, L., Washburn, M.P., and Workman, J.L. (2021a).
826 The methyltransferase SETD2 couples transcription and splicing by engaging mRNA processing factors
827 through its SHI domain. Nat Commun *12*, 1443.

828 Bhattacharya, S., Wang, S., Reddy, D., Shen, S., Zhang, Y., Zhang, N., Li, H., Washburn, M.P., Florens, L.,
829 Shi, Y.*, et al.* (2021b). Structural basis of the interaction between SETD2 methyltransferase and hnRNP L
830 paralogs for governing co-transcriptional splicing. Nat Commun *12*, 6452.

831 Cole, B.S., Tapescu, I., Allon, S.J., Mallory, M.J., Qiu, J., Lake, R.J., Fan, H.Y., Fu, X.D., and Lynch, K.W.
832 (2015). Global analysis of physical and functional RNA targets of hnRNP L reveals distinct sequence and
833 epigenetic features of repressed and enhanced exons. RNA *21*, 2053-2066.

834 Cramer, P., Pesce, C.G., Baralle, F.E., and Kornblihtt, A.R. (1997). Functional association between
835 promoter structure and transcript alternative splicing. Proc Natl Acad Sci U S A *94*, 11456-11460.

836 Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel,
837 A., Knowles, D.G.*, et al.* (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of
838 their gene structure, evolution, and expression. Genome Res *22*, 1775-1789.

839 Dinger, M.E., Pang, K.C., Mercer, T.R., and Mattick, J.S. (2008). Differentiating protein-coding and
840 noncoding RNA: challenges and ambiguities. PLoS computational biology *4*, e1000176.

841 Duymich, C.E., Charlet, J., Yang, X., Jones, P.A., and Liang, G. (2016). DNMT3B isoforms without catalytic
842 activity stimulate gene body methylation as accessory proteins in somatic cells. Nat Commun *7*, 11453.

843 Engstrom, P.G., Suzuki, H., Ninomiya, N., Akalin, A., Sessa, L., Lavorgna, G., Brozzi, A., Luzi, L., Tan, S.L.,
844 Yang, L.*, et al.* (2006). Complex Loci in human and mouse genomes. PLoS genetics *2*, e47.

845 Esteller, M. (2005). Aberrant DNA methylation as a cancer-inducing mechanism. Annual review of
846 pharmacology and toxicology *45*, 629-656.

847 Feng, J., Bi, C., Clark, B.S., Mady, R., Shah, P., and Kohtz, J.D. (2006). The Evf-2 noncoding RNA is
848 transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator.
849 Genes Dev *20*, 1470-1484.

850 Ficz, G., Hore, T.A., Santos, F., Lee, H.J., Dean, W., Arand, J., Krueger, F., Oxley, D., Paul, Y.L., Walter, J.*, et*
851 *al.* (2013). FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic
852 ground state of pluripotency. Cell Stem Cell *13*, 351-359.

853 Gil, N., and Ulitsky, I. (2020). Regulation of gene expression by cis-acting long non-coding RNAs. Nat Rev
854 Genet *21*, 102-117.

855 Gopalakrishnan, S., Sullivan, B.A., Trazzi, S., Della Valle, G., and Robertson, K.D. (2009). DNMT3B
856 interacts with constitutive centromere protein CENP-C to modulate DNA methylation and the histone
857 code at centromeric regions. Human molecular genetics *18*, 3178-3193.

858 Gowher, H., Brick, K., Camerini-Otero, R.D., and Felsenfeld, G. (2012). Vezf1 protein binding sites
859 genome-wide are associated with pausing of elongating RNA polymerase II. Proc Natl Acad Sci U S A *109*,
860 2370-2375.

861 Gowher, H., and Jeltsch, A. (2002). Molecular enzymology of the catalytic domains of the Dnmt3a and
862 Dnmt3b DNA methyltransferases. The Journal of biological chemistry *277*, 20409-20414.

863 Gowher, H., Stuhlmann, H., and Felsenfeld, G. (2008). Vezf1 regulates genomic DNA methylation
864 through its effects on expression of DNA methyltransferase Dnmt3b. Genes Dev *22*, 2075-2084.

865 Guil, S., and Esteller, M. (2009). DNA methylomes, histone codes and miRNAs: tying it all together. The
866 international journal of biochemistry & cell biology *41*, 87-95.

867 Guo, C.J., Ma, X.K., Xing, Y.H., Zheng, C.C., Xu, Y.F., Shan, L., Zhang, J., Wang, S., Wang, Y., Carmichael,
868 G.G.*, et al.* (2020). Distinct Processing of lncRNAs Contributes to Non-conserved Functions in Stem Cells.
869 Cell *181*, 621-636 e622.

870 Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W.,
871 Cassady, J.P.*, et al.* (2009). Chromatin signature reveals over a thousand highly conserved large non-
872 coding RNAs in mammals. Nature *458*, 223-227.

873 Guttman, M., Donaghey, J., Carey, B.W., Garber, M., Grenier, J.K., Munson, G., Young, G., Lucas, A.B.,
874 Ach, R., Bruhn, L.*, et al.* (2011). lincRNAs act in the circuitry controlling pluripotency and differentiation.
875 Nature *477*, 295-300.

876 Hacisuleyman, E., Shukla, C.J., Weiner, C.L., and Rinn, J.L. (2016). Function and evolution of local repeats
877 in the Firre locus. Nat Commun *7*, 11021.

878 Hagege, H., Klous, P., Braem, C., Splinter, E., Dekker, J., Cathala, G., de Laat, W., and Forne, T. (2007).
879 Quantitative analysis of chromosome conformation capture assays (3C-qPCR). Nat Protoc *2*, 1722-1733.

880 Hahm, B., Cho, O.H., Kim, J.E., Kim, Y.K., Kim, J.H., Oh, Y.L., and Jang, S.K. (1998). Polypyrimidine tract-
881 binding protein interacts with HnRNP L. FEBS Lett *425*, 401-406.

882 Hirasawa, R., and Sasaki, H. (2009). Dynamic transition of Dnmt3b expression in mouse pre- and early
883 post-implantation embryos. Gene Expr Patterns *9*, 27-30.

884 Huang, Y., Li, W., Yao, X., Lin, Q.J., Yin, J.W., Liang, Y., Heiner, M., Tian, B., Hui, J., and Wang, G. (2012).
885 Mediator complex regulates alternative mRNA processing via the MED23 subunit. Mol Cell *45*, 459-469.

886 Hui, J., Reither, G., and Bindereif, A. (2003). Novel functional role of CA repeats and hnRNP L in RNA
887 stability. RNA *9*, 931-936.

888 Hung, L.H., Heiner, M., Hui, J., Schreiner, S., Benes, V., and Bindereif, A. (2008). Diverse roles of hnRNP L
889 in mammalian mRNA processing: a combined microarray and RNAi analysis. RNA *14*, 284-296.

890 Huntriss, J., Hinkins, M., Oliver, B., Harris, S.E., Beazley, J.C., Rutherford, A.J., Gosden, R.G., Lanzendorf,
891 S.E., and Picton, H.M. (2004). Expression of mRNAs for DNA methyltransferases and methyl-CpG-binding
892 proteins in the human female germ line, preimplantation embryos, and embryonic stem cells. Molecular
893 reproduction and development *67*, 323-336.

894 Ishida, C., Ura, K., Hirao, A., Sasaki, H., Toyoda, A., Sakaki, Y., Niwa, H., Li, E., and Kaneda, Y. (2003).
895 Genomic organization and promoter analysis of the Dnmt3b gene. Gene *310*, 151-159.

896 Jinawath, A., Miyake, S., Yanagisawa, Y., Akiyama, Y., and Yuasa, Y. (2005). Transcriptional regulation of
897 the human DNA methyltransferase 3A and 3B genes by Sp3 and Sp1 zinc finger proteins. The
898 Biochemical journal *385*, 557-564.

899 Kulis, M., and Esteller, M. (2010). DNA methylation and cancer. Advances in genetics *70*, 27-56.

900 Leitch, H.G., McEwen, K.R., Turp, A., Encheva, V., Carroll, T., Grabole, N., Mansfield, W., Nashun, B.,
901 Knezovich, J.G., Smith, A.*, et al.* (2013). Naive pluripotency is associated with global DNA
902 hypomethylation. Nature structural & molecular biology *20*, 311-316.

903 Li, Z., Chao, T.C., Chang, K.Y., Lin, N., Patil, V.S., Shimizu, C., Head, S.R., Burns, J.C., and Rana, T.M. (2014).
904 The long noncoding RNA THRIL regulates TNFalpha expression through its interaction with hnRNPL. Proc
905 Natl Acad Sci U S A *111*, 1002-1007.

906 Lubelsky, Y., and Ulitsky, I. (2018). Sequences enriched in Alu repeats drive nuclear localization of long
907 RNAs in human cells. Nature *555*, 107-111.

908 Luco, R.F., Pan, Q., Tominaga, K., Blencowe, B.J., Pereira-Smith, O.M., and Misteli, T. (2010). Regulation
909 of alternative splicing by histone modifications. Science *327*, 996-1000.

910 Luo, S., Lu, J.Y., Liu, L., Yin, Y., Chen, C., Han, X., Wu, B., Xu, R., Liu, W., Yan, P.*, et al.* (2016). Divergent
911 lncRNAs Regulate Gene Expression and Lineage Differentiation in Pluripotent Cells. Cell Stem Cell *18*,
912 637-652.

913 Mele, M., Mattioli, K., Mallard, W., Shechner, D.M., Gerhardinger, C., and Rinn, J.L. (2017). Chromatin
914 environment, transcriptional regulation, and splicing distinguish lincRNAs and mRNAs. Genome Res *27*,
915 27-37.

916 Mensah, I.K., Norvil, A.B., AlAbdi, L., McGovern, S., Petell, C.J., He, M., and Gowher, H. (2021).
917 Misregulation of the expression and activity of DNA methyltransferases in cancer. NAR Cancer *3*,
918 zcab045.

919 Montgomery, K.G., Liu, M.C., Eccles, D.M., and Campbell, I.G. (2004). The DNMT3B C-->T promoter
920 polymorphism and risk of breast cancer in a British population: a case-control study. Breast cancer
921 research : BCR *6*, R390-394.

922 Naumova, N., Smith, E.M., Zhan, Y., and Dekker, J. (2012). Analysis of long-range chromatin interactions
923 using Chromosome Conformation Capture. Methods *58*, 192-203.

924 Norvil, A.B., AlAbdi, L., Liu, B., Tu, Y.H., Forstoffer, N.E., Michie, A.R., Chen, T., and Gowher, H. (2020).
925 The acute myeloid leukemia variant DNMT3A Arg882His is a DNMT3B-like enzyme. Nucleic Acids Res *48*,
926 3761-3775.

927 Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are
928 essential for de novo methylation and mammalian development. Cell *99*, 247-257.

929 Okano, M., Xie, S., and Li, E. (1998). Cloning and characterization of a family of novel mammalian DNA
930 (cytosine-5) methyltransferases. Nature genetics *19*, 219-220.

931 Ostler, K.R., Davis, E.M., Payne, S.L., Gosalia, B.B., Exposito-Cespedes, J., Le Beau, M.M., and Godley, L.A.
932 (2007). Cancer cells express aberrant DNMT3B transcripts encoding truncated proteins. Oncogene *26*,
933 5553-5563.

934 Panda, A.C., Martindale, J.L., and Gorospe, M. (2016). Affinity Pulldown of Biotinylated RNA for
935 Detection of Protein-RNA Complexes. Bio-protocol *6*, e2062.

936 Pandya-Jones, A., and Black, D.L. (2009). Co-transcriptional splicing of constitutive and alternative exons.
937 Rna *15*, 1896-1908.

938 Peach, M., Marsh, N., Miskiewicz, E.I., and MacPhee, D.J. (2015). Solubilization of proteins: the
939 importance of lysis buffer choice. Methods Mol Biol *1312*, 49-60.

940 Petell, C.J., Alabdi, L., He, M., San Miguel, P., Rose, R., and Gowher, H. (2016). An epigenetic switch
941 regulates de novo DNA methylation at a subset of pluripotency gene enhancers during embryonic stem
942 cell differentiation. Nucleic Acids Res *44*, 7605-7617.

943 Ponjavic, J., Ponting, C.P., and Lunter, G. (2007). Functionality or transcriptional noise? Evidence for
944 selection within long noncoding RNAs. Genome Res *17*, 556-565.

945 Preussner, M., Schreiner, S., Hung, L.H., Porstner, M., Jack, H.M., Benes, V., Ratsch, G., and Bindereif, A.
946 (2012). HnRNP L and L-like cooperate in multiple-exon regulation of CD45 alternative splicing. Nucleic
947 acids research *40*, 5666-5678.

948 Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A.,
949 Farnham, P.J., Segal, E.*, et al.* (2007). Functional demarcation of active and silent chromatin domains in
950 human HOX loci by noncoding RNAs. Cell *129*, 1311-1323.

951 Robertson, K.D., Keyomarsi, K., Gonzales, F.A., Velicescu, M., and Jones, P.A. (2000). Differential mRNA
952 expression of the human DNA methyltransferases (DNMTs) 1, 3a and 3b during the G(0)/G(1) to S phase
953 transition in normal and tumor cells. Nucleic acids research *28*, 2108-2113.

954 Robertson, K.D., Uzvolgyi, E., Liang, G., Talmadge, C., Sumegi, J., Gonzales, F.A., and Jones, P.A. (1999).
955 The human DNA methyltransferases (DNMTs) 1, 3a and 3b: coordinate mRNA expression in normal
956 tissues and overexpression in tumors. Nucleic acids research *27*, 2291-2298.

957 Ruan, X., Li, P., Cangelosi, A., Yang, L., and Cao, H. (2016). A Long Non-coding RNA, lncLGR, Regulates
958 Hepatic Glucokinase Expression and Glycogen Storage during Fasting. Cell Rep *14*, 1867-1875.

959 Saito, Y., Kanai, Y., Sakamoto, M., Saito, H., Ishii, H., and Hirohashi, S. (2002). Overexpression of a splice
960 variant of DNA methyltransferase 3b, DNMT3b4, associated with DNA hypomethylation on
961 pericentromeric satellite regions during human hepatocarcinogenesis. Proc Natl Acad Sci U S A *99*,
962 10060-10065.

963 Saldana-Meyer, R., Rodriguez-Hernaez, J., Escobar, T., Nishana, M., Jacome-Lopez, K., Nora, E.P.,
964 Bruneau, B.G., Tsirigos, A., Furlan-Magaril, M., Skok, J.*, et al.* (2019). RNA Interactions Are Essential for
965 CTCF-Mediated Genome Organization. Mol Cell *76*, 412-422 e415.

966 Schertzer, M.D., Braceros, K.C.A., Starmer, J., Cherney, R.E., Lee, D.M., Salazar, G., Justice, M., Bischoff,
967 S.R., Cowley, D.O., Ariel, P.*, et al.* (2019). lncRNA-Induced Spread of Polycomb Controlled by Genome
968 Architecture, RNA Abundance, and CpG Island DNA. Mol Cell *75*, 523-537 e510.

969 Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A., and Sharp, P.A.
970 (2008). Divergent transcription from active promoters. Science *322*, 1849-1851.

971 Shen, H., Wang, L., Spitz, M.R., Hong, W.K., Mao, L., and Wei, Q. (2002). A novel polymorphism in human
972 cytosine DNA-methyltransferase-3B promoter is associated with an increased risk of lung cancer. Cancer
973 research *62*, 4992-4995.

974 Shukla, C.J., McCorkindale, A.L., Gerhardinger, C., Korthauer, K.D., Cabili, M.N., Shechner, D.M., Irizarry,
975 R.A., Maass, P.G., and Rinn, J.L. (2018). High-throughput identification of RNA nuclear enrichment
976 sequences. EMBO J *37*.

977 Shuman, S. (2020). Transcriptional interference at tandem lncRNA and protein-coding genes: an
978 emerging theme in regulation of cellular nutrient homeostasis. Nucleic acids research *48*, 8243-8254.

979 Sigova, A.A., Mullen, A.C., Molinie, B., Gupta, S., Orlando, D.A., Guenther, M.G., Almada, A.E., Lin, C.,
980 Sharp, P.A., Giallourakis, C.C.*, et al.* (2013). Divergent transcription of long noncoding RNA/mRNA gene
981 pairs in embryonic stem cells. Proc Natl Acad Sci U S A *110*, 2876-2881.

982 Singal, R., Das, P.M., Manoharan, M., Reis, I.M., and Schlesselman, J.J. (2005). Polymorphisms in the DNA
983 methyltransferase 3b gene and prostate cancer risk. Oncology reports *14*, 569-573.

984 Sleutels, F., Zwart, R., and Barlow, D.P. (2002). The non-coding Air RNA is required for silencing
985 autosomal imprinted genes. Nature *415*, 810-813.

986 Smith, S.A., Ray, D., Cook, K.B., Mallory, M.J., Hughes, T.R., and Lynch, K.W. (2013). Paralogs hnRNP L
987 and hnRNP LL exhibit overlapping but distinct RNA binding constraints. PLoS One *8*, e80701.

988 Statello, L., Guo, C.J., Chen, L.L., and Huarte, M. (2021). Gene regulation by long non-coding RNAs and its
989 biological functions. Nat Rev Mol Cell Biol *22*, 96-118.

990 Szyf, M. (2005). Therapeutic implications of DNA methylation. Future Oncol *1*, 125-135.

991 Thakur, N., Tiwari, V.K., Thomassin, H., Pandey, R.R., Kanduri, M., Gondor, A., Grange, T., Ohlsson, R.,
992 and Kanduri, C. (2004). An antisense RNA regulates the bidirectional silencing property of the Kcnq1
993 imprinting control region. Molecular and cellular biology *24*, 7855-7862.

994 Tian, B., and Manley, J.L. (2017). Alternative polyadenylation of mRNA precursors. Nat Rev Mol Cell Biol
995 *18*, 18-30.

996 Tosolini, M., and Jouneau, A. (2016). Acquiring Ground State Pluripotency: Switching Mouse Embryonic
997 Stem Cells from Serum/LIF Medium to 2i/LIF Medium. Methods Mol Biol *1341*, 41-48.

998 Walton, E.L., Francastel, C., and Velasco, G. (2014). Dnmt3b Prefers Germ Line Genes and Centromeric
999 Regions: Lessons from the ICF Syndrome and Cancer and Implications for Diseases. Biology *3*, 578-605.

1000 Wang, J., Bhutani, M., Pathak, A.K., Lang, W., Ren, H., Jelinek, J., He, R., Shen, L., Issa, J.P., and Mao, L.
1001 (2007). Delta DNMT3B variants regulate DNA methylation in a promoter-specific manner. Cancer
1002 research *67*, 10647-10652.

1003 Watanabe, D., Suetake, I., Tada, T., and Tajima, S. (2002). Stage- and cell-specific expression of Dnmt3a
1004 and Dnmt3b during embryogenesis. Mech Dev *118*, 187-190.

1005 Watanabe, Y., and Maekawa, M. (2010). Methylation of DNA in cancer. Advances in clinical chemistry
1006 *52*, 145-167.

1007 Weisenberger, D.J., Velicescu, M., Cheng, J.C., Gonzales, F.A., Liang, G., and Jones, P.A. (2004). Role of
1008 the DNA methyltransferase variant DNMT3b3 in DNA methylation. Mol Cancer Res *2*, 62-72.

1009 Wuarin, J., and Schibler, U. (1994). Physical isolation of nascent RNA chains transcribed by RNA
1010 polymerase II: evidence for cotranscriptional splicing. Mol Cell Biol *14*, 7219-7225.

1011 Xiang, J.F., Yin, Q.F., Chen, T., Zhang, Y., Zhang, X.O., Wu, Z., Zhang, S., Wang, H.B., Ge, J., Lu, X.*, et al.*
1012 (2014). Human colorectal cancer-specific CCAT1-L lncRNA regulates long-range chromatin interactions at
1013 the MYC locus. Cell Res *24*, 513-531.

1014 Xie, S., Wang, Z., Okano, M., Nogami, M., Li, Y., He, W.W., Okumura, K., and Li, E. (1999). Cloning,
1015 expression and chromosome locations of the human DNMT3 gene family. Gene *236*, 87-95.

1016 Xie, Z.H., Huang, Y.N., Chen, Z.X., Riggs, A.D., Ding, J.P., Gowher, H., Jeltsch, A., Sasaki, H., Hata, K., and
1017 Xu, G.L. (2006). Mutations in DNA methyltransferase DNMT3B in ICF syndrome affect its regulation by
1018 DNMT3L. Hum Mol Genet *15*, 1375-1385.

1019 Xu, G.L., Bestor, T.H., Bourc'his, D., Hsieh, C.L., Tommerup, N., Bugge, M., Hulten, M., Qu, X., Russo, J.J.,
1020 and Viegas-Pequignot, E. (1999). Chromosome instability and immunodeficiency syndrome caused by
1021 mutations in a DNA methyltransferase gene. Nature *402*, 187-191.

1022 Yang, F., Deng, X., Ma, W., Berletch, J.B., Rabaia, N., Wei, G., Moore, J.M., Filippova, G.N., Xu, J., Liu, Y.*,*
1023 *et al.* (2015). The lncRNA Firre anchors the inactive X chromosome to the nucleolus by binding CTCF and
1024 maintains H3K27me3 methylation. Genome biology *16*, 52.

1025 Yin, Y., Lu, J.Y., Zhang, X., Shao, W., Xu, Y., Li, P., Hong, Y., Cui, L., Shan, G., Tian, B.*, et al.* (2020). U1
1026 snRNP regulates chromatin retention of noncoding RNAs. Nature *580*, 147-150.

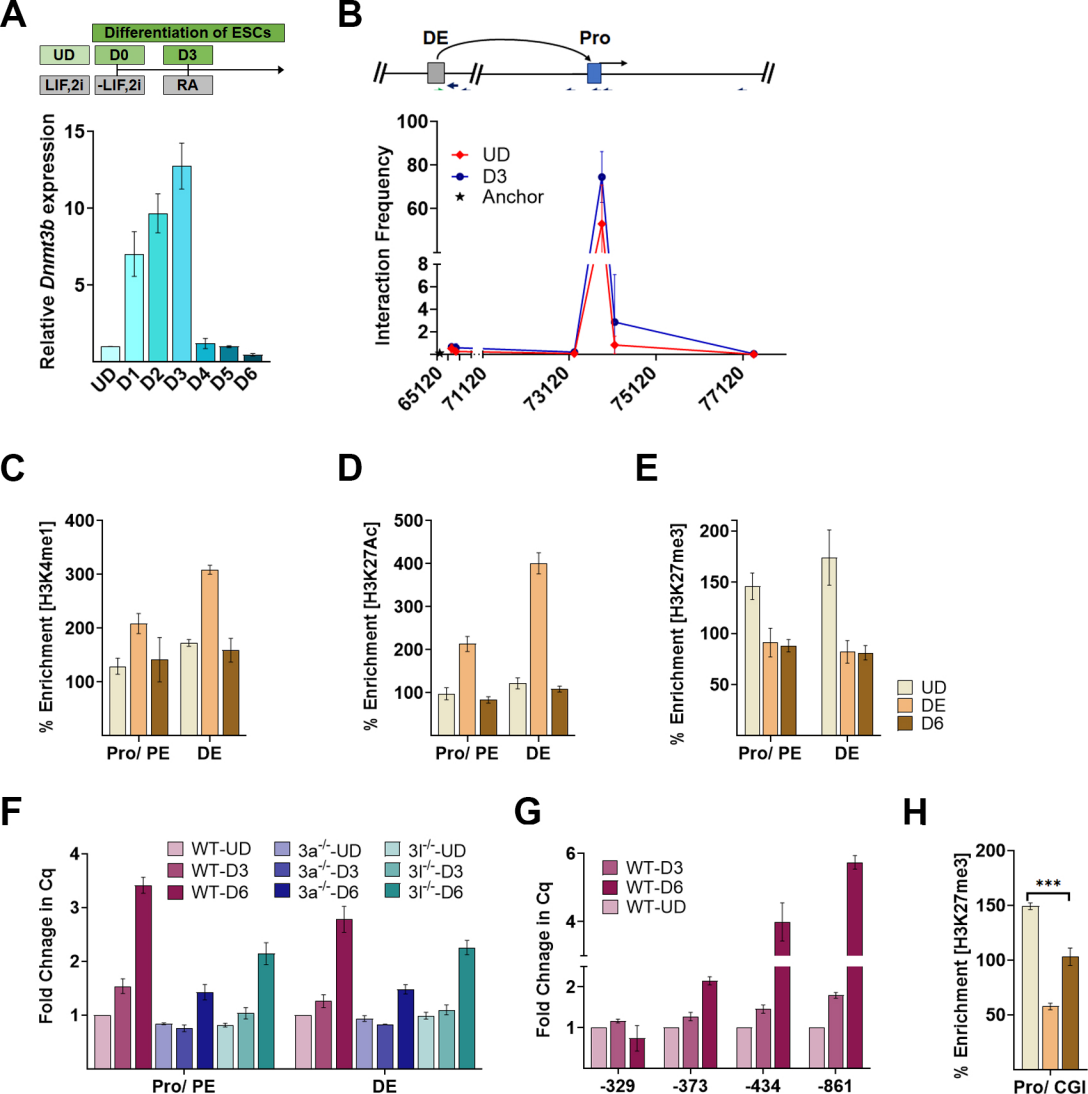
1027 Ying, Q.L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008).
1028 The ground state of embryonic stem cell self-renewal. Nature *453*, 519-523.

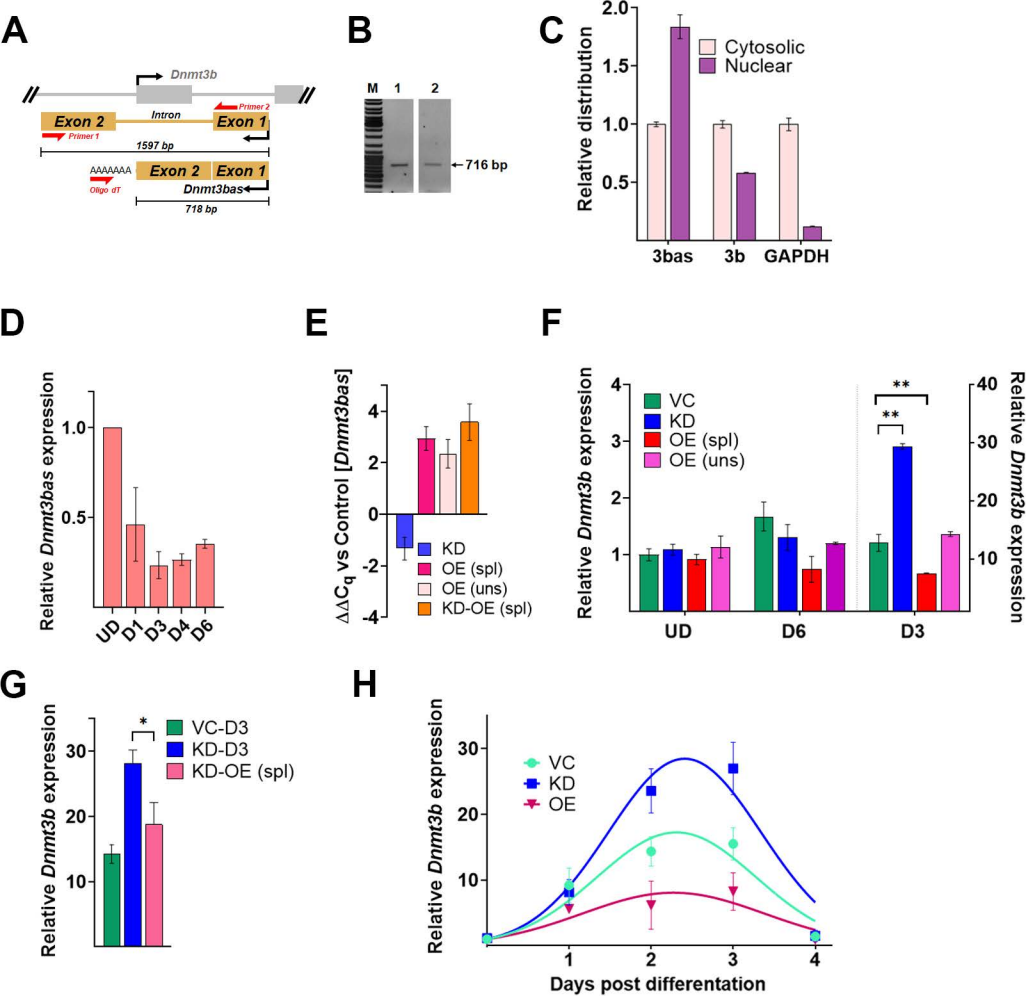
1029 Yuan, W., Xie, J., Long, C., Erdjument-Bromage, H., Ding, X., Zheng, Y., Tempst, P., Chen, S., Zhu, B., and
1030 Reinberg, D. (2009). Heterogeneous nuclear ribonucleoprotein L Is a subunit of human KMT3a/Set2
1031 complex required for H3 Lys-36 trimethylation activity in vivo. The Journal of biological chemistry *284*,
1032 15701-15707.

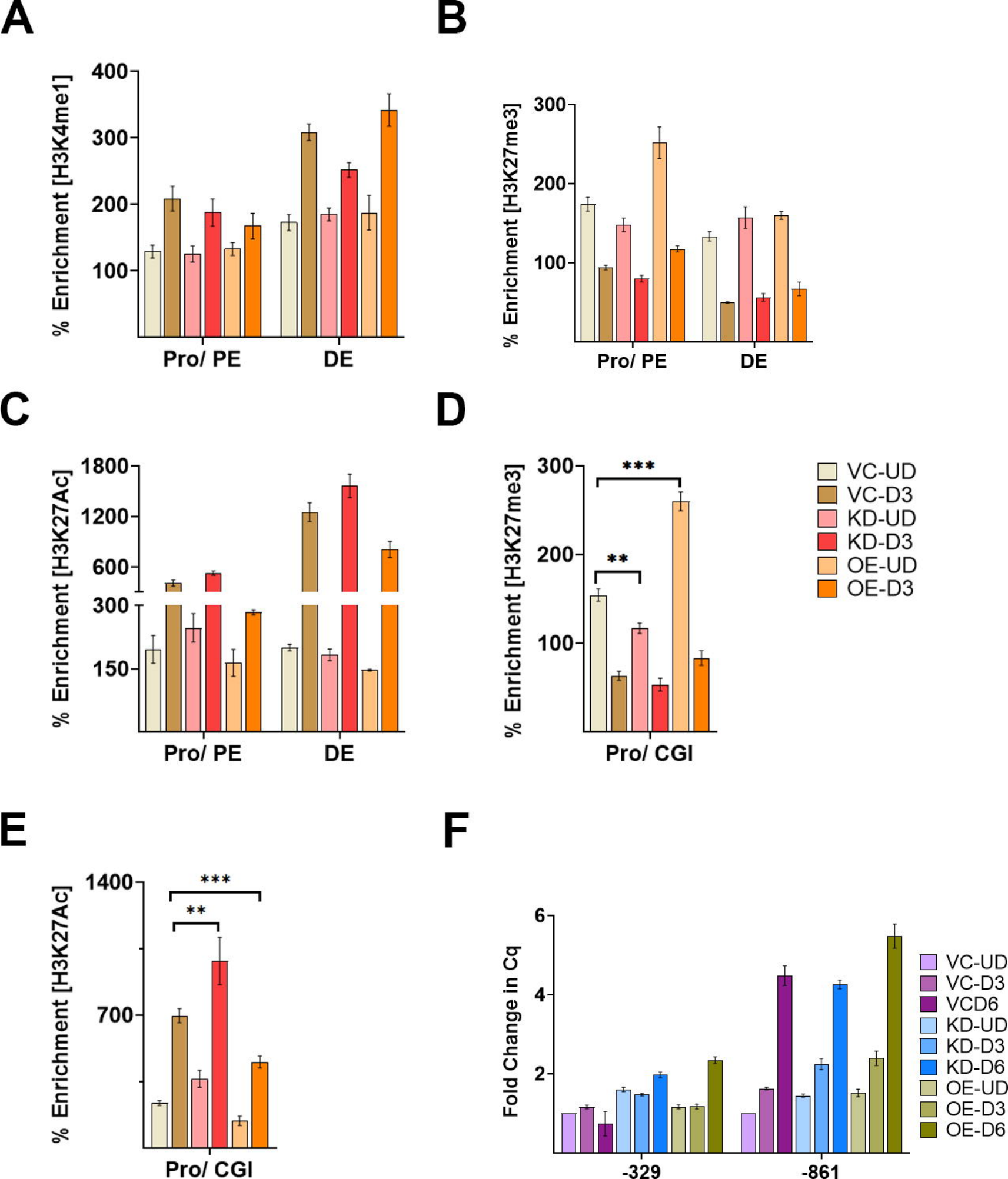
1033 Zeng, Y., Ren, R., Kaur, G., Hardikar, S., Ying, Z., Babcock, L., Gupta, E., Zhang, X., Chen, T., and Cheng, X.
1034 (2020). The inactive Dnmt3b3 isoform preferentially enhances Dnmt3b-mediated DNA methylation.
1035 Genes Dev *34*, 1546-1558.

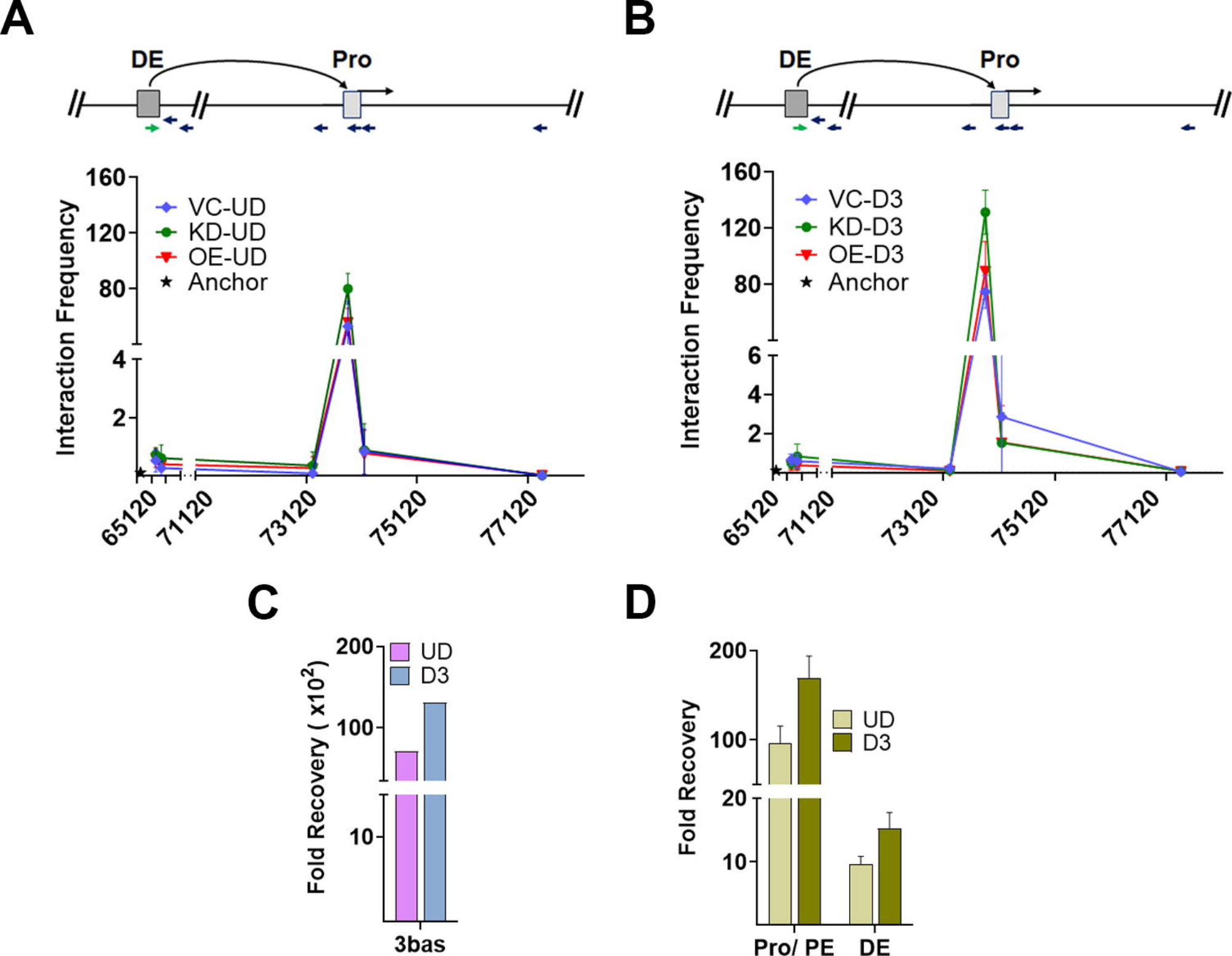
1036 Zuckerman, B., and Ulitsky, I. (2019). Predictive models of subcellular localization of long RNAs. RNA *25*,
1037 557-572.

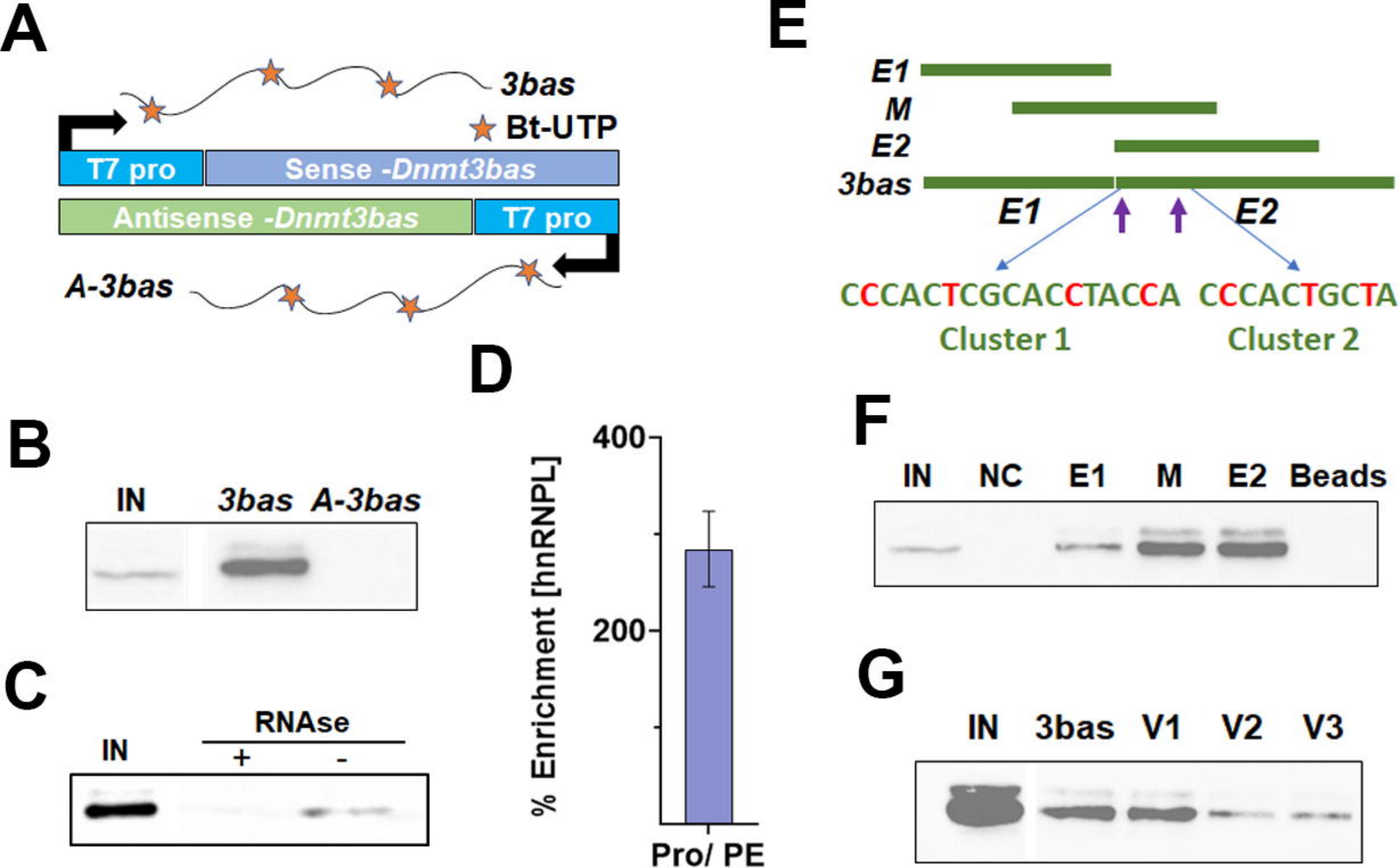
1038

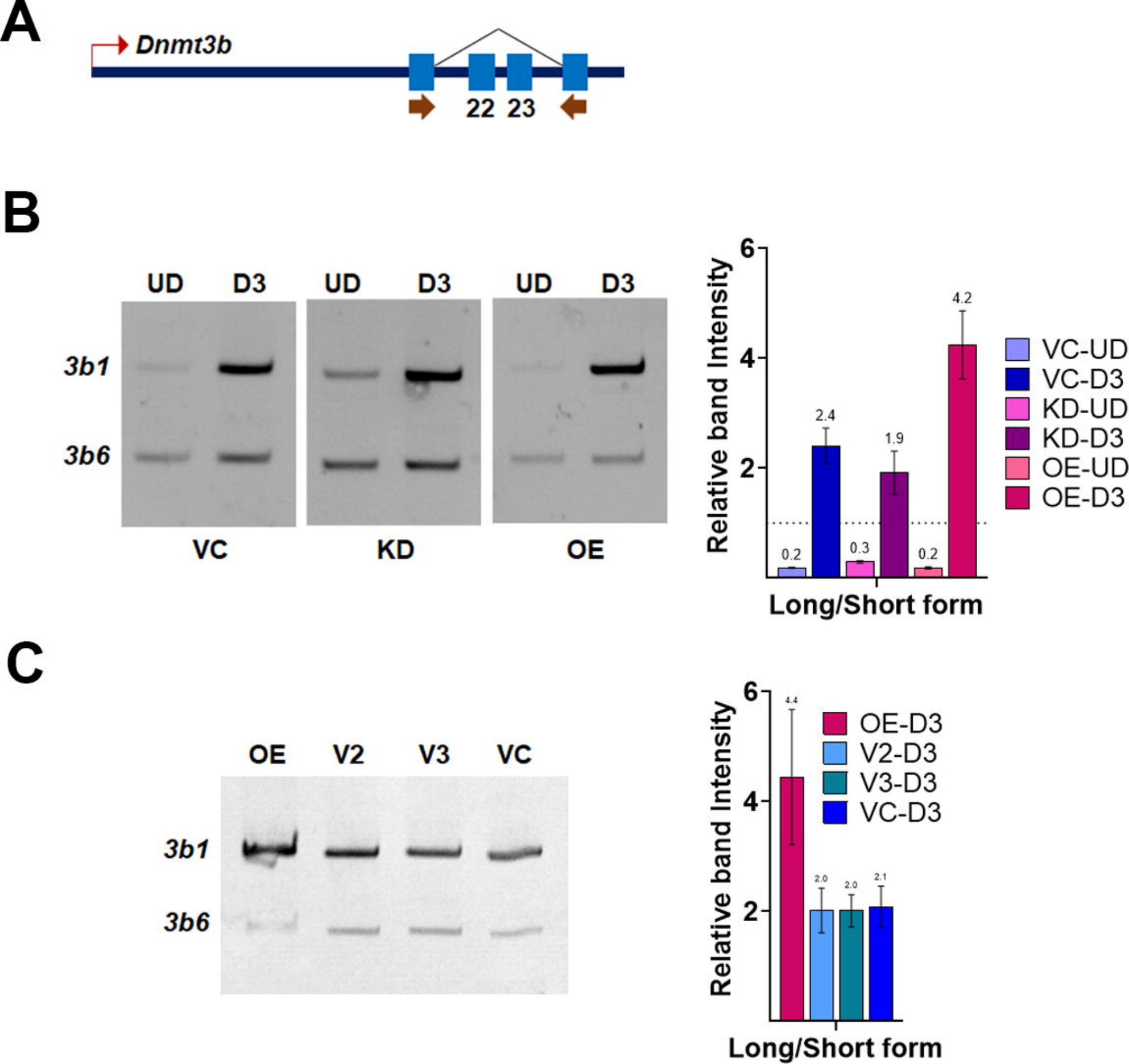

A
Differentiation of ESCs
UD
D0
D3
LIF,2i
-LIF,2i
RA
Relative Dnmt3b expression
15
10
5
UD
D1
D2
D3
D4
D5
D6
B
DE
Pro
Interaction Frequency
100
80
60
8
6
4
2
UD
D3
Anchor
65120
71120
73120
75120
77120
C
% Enrichment [H3K4me1]
400
300
200
100
Pro/ PE
DE
D
% Enrichment [H3K27Ac]
500
400
300
200
100
Pro/ PE
DE
E
% Enrichment [H3K27me3]
200
150
100
50
Pro/ PE
DE
UD
DE
D6
F
Fold Chnage in Cq
4
3
2
1
WT-UD
WT-D3
WT-D6
3a-/--UD
3a-/--D3
3a-/--D6
3l-/--UD
3l-/--D3
3l-/--D6
Pro/ PE
DE
G
Fold Chnage in Cq
6
4
2
1
WT-D3
WT-D6
WT-UD
-329
-373
-434
-861
H
% Enrichment [H3K27me3]
200
150
100
50
***
Pro/ CGI

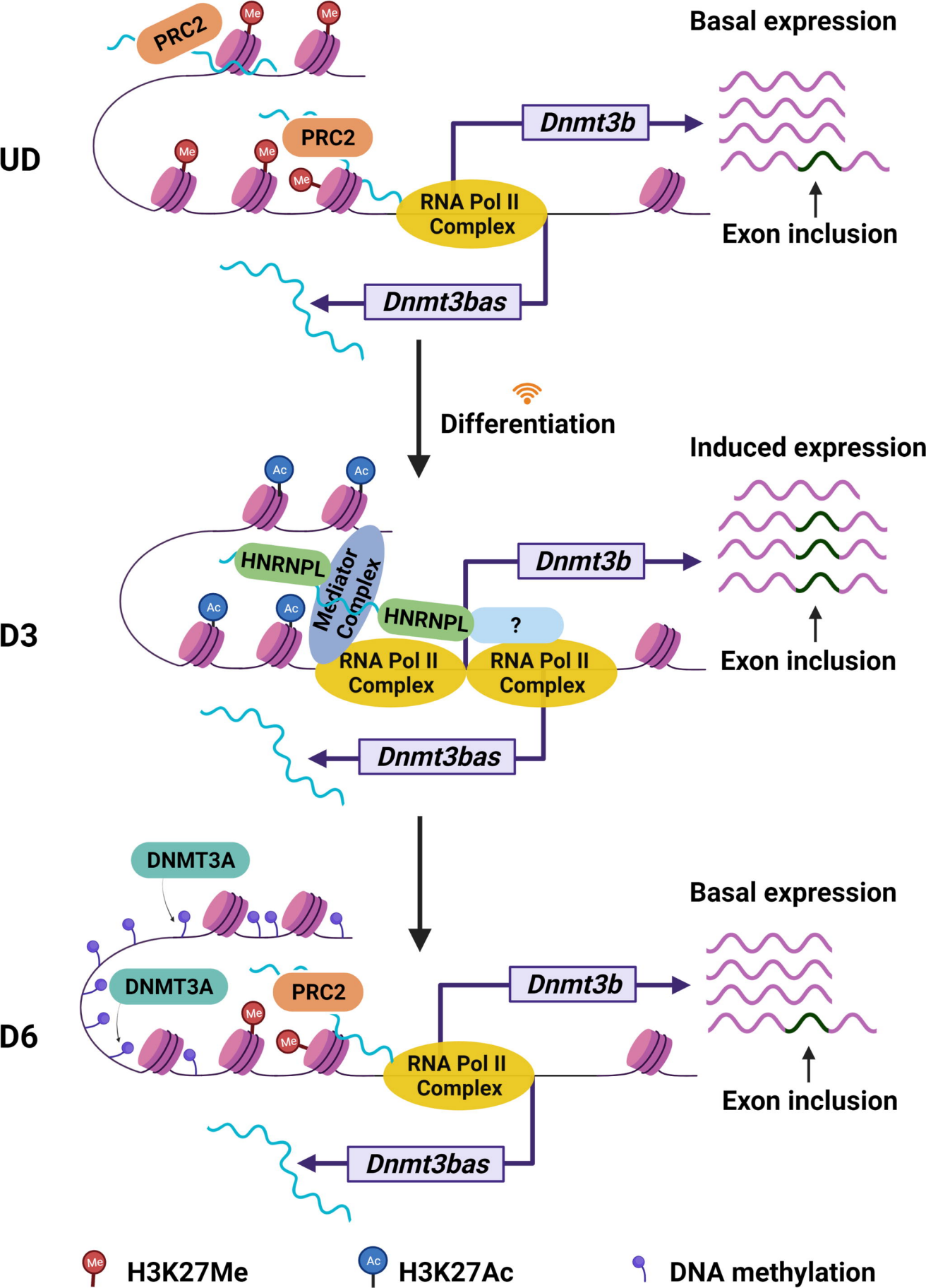

A
Dnmt3b
Intron
Exon 2
Exon 1
Primer 2
Primer 1
1597 bp
AAAAAAA
Oligo dT
Exon 2
Exon 1
Dnmt3bas
718 bp
B
M
1
2
716 bp
C
Relative distribution
Cytosolic
Nuclear
3bas
3b
GAPDH
D
Relative Dnmt3bas expression
UD
D1
D3
D4
D6
E
ΔΔCq vs Control [Dnmt3bas]
KD
OE (spl)
OE (uns)
KD-OE (spl)
F
Relative Dnmt3b expression
VC
KD
OE (spl)
OE (uns)
UD
D6
D3
G
Relative Dnmt3b expression
VC-D3
KD-D3
KD-OE (spl)
H
Relative Dnmt3b expression
VC
KD
OE
Days post differentation


A
% Enrichment [H3K4me1]
400
300
200
100
Pro/ PE
DE
B
% Enrichment [H3K27me3]
300
200
100
Pro/ PE
DE
C
% Enrichment [H3K27Ac]
1800
1200
600
300
150
Pro/ PE
DE
D
% Enrichment [H3K27me3]
300
200
100
***
**
Pro/ CGI
VC-UD
VC-D3
KD-UD
KD-D3
OE-UD
OE-D3
E
% Enrichment [H3K27Ac]
1400
700
***
**
Pro/ CGI
F
Fold Change in Cq
6
4
2
-329
-861
VC-UD
VC-D3
VCD6
KD-UD
KD-D3
KD-D6
OE-UD
OE-D3
OE-D6


A
DE
Pro
Interaction Frequency
160
120
80
4
2
VC-UD
KD-UD
OE-UD
Anchor
65120
71120
73120
75120
77120
B
DE
Pro
Interaction Frequency
160
120
80
6
4
2
VC-D3
KD-D3
OE-D3
Anchor
65120
71120
73120
75120
77120
C
Fold Recovery ( x10$^2$)
200
100
10
UD
D3
3bas
D
Fold Recovery
200
100
20
10
UD
D3
Pro/ PE
DE


A
3bas
Bt-UTP
T7 pro
Sense -Dnmt3bas
Antisense -Dnmt3bas
T7 pro
A-3bas
E
E1
M
E2
3bas
E1
E2
CCCACTCGCACCTACCA
CCCACTGCTA
Cluster 1
Cluster 2
D
B
IN
3bas
A-3bas
400
200
% Enrichment [hnRNPL]
Pro/ PE
F
IN
NC
E1
M
E2
Beads
C
RNAse
IN
+
-
G
IN
3bas
V1
V2
V3


A
Dnmt3b
22 23
B
UD D3 UD D3 UD D3
3b1
3b6
VC KD OE
Relative band Intensity
6
4
2
0.2
2.4
0.3
1.9
0.2
4.2
Long/Short form
VC-UD
VC-D3
KD-UD
KD-D3
OE-UD
OE-D3
C
OE V2 V3 VC
3b1
3b6
Relative band Intensity
6
4
2
4.4
2.0
2.0
2.1
Long/Short form
OE-D3
V2-D3
V3-D3
VC-D3


UD
PRC2
Me
Me
PRC2
Me
Me
Me
Basal expression
Dnmt3b
RNA Pol II Complex
Exon inclusion
Dnmt3bas
Differentiation
D3
Ac
Ac
HNRNPL
Mediator Complex
Ac
Ac
HNRNPL
?
Induced expression
Dnmt3b
RNA Pol II Complex
RNA Pol II Complex
Exon inclusion
Dnmt3bas
D6
DNMT3A
DNMT3A
PRC2
Me
Me
Basal expression
Dnmt3b
RNA Pol II Complex
Exon inclusion
Dnmt3bas
Me H3K27Me
Ac H3K27Ac
DNA methylation