Intergenic Polycomb target sites are dynamically marked by non-coding transcription during lineage commitment

Betül Hekimoglu-Balkan,[1,3] András Aszodi,[2] Robert Heinen,[1] Markus Jaritz[2] and Leonie Ringrose[1,*]

[1]IMBA, Institute of Molecular Biotechnology GmbH; Vienna, Austria; [2]IMP, Institute of Molecular Pathology; Vienna, Austria; [3]MRC Human Genetics Unit, Chromosomes & Gene Expression Section, Western General Hospital; Edinburgh, UK

Non-coding (nc) RNAs are involved both in recruitment of vertebrate Polycomb (PcG) proteins to chromatin, and in activation of PcG target genes. Here we investigate dynamic changes in the relationship between ncRNA transcription and recruitment of PcG proteins to chromatin during differentiation. Profiling of purified cell populations from different stages of a defined murine in vitro neural differentiation system shows that over 50% of regulated intergenic non-coding transcripts precisely correspond to PcG target sites. We designate these PcG recruiting elements as Transcribed Intergenic Polycomb (TIP) sites. The relationship between TIP transcription and PcG recruitment switches dynamically during differentiation between different states, in which transcription and PcG recruitment exclude each other, or in which both are present. Reporter assays show that transcribed TIP sites can repress a flanking gene. Knockdown experiments demonstrate that TIP ncRNAs are themselves required for repression of target genes both in *cis* and in *trans*. We propose that TIP transcription may ensure coordinated regulation of gene networks via dynamic switching and recruitment of PcG proteins both in *cis* and in *trans* during lineage commitment.

Introduction

Polycomb (PcG) proteins are essential for self-renewal and differentiation processes.[1,2] These highly conserved proteins silence several hundred developmentally important genes, and change their targets dynamically upon differentiation.[3-5] In both flies and vertebrates, many PcG binding sites give rise to developmentally regulated non-coding (nc) RNAs,[6-9] reviewed in.[10] In flies, PcG proteins are recruited to Polycomb Response Elements (PREs) by DNA binding proteins,[11] and the PRE ncRNAs are thought to modulate PcG function at PREs.[12] Nc transcription from fly PREs is highly dynamic and transient at different stages of development, and has been reported to be involved in both activation[13] and silencing[14] of gene activity (reviewed in 12).

In vertebrates, two PRE elements have recently been identified.[15,16] Sequence mining and functional studies have uncovered DNA features that are enriched in PcG binding sites[17-19] and are required for PcG recruitment.[16,20] However in contrast to flies, DNA sequence requirements for mammalian PRE definition have proved elusive, and there is much current interest in the potential role of ncRNAs in recruiting PcG proteins to chromatin. Many ncRNAs and short promoter RNAs have been reported to associate specifically with mammalian PcG proteins.[8,21-26] Several of these ncRNAs have been shown to be required for PcG recruitment to specific chromatin sites, and in one recent report, via a DNA-RNA bridging protein.[27]

In addition to the PcG associated ncRNAs, there is a vast traffic of additional regulated ncRNA transcription in vertebrates, some of which has been shown to play a role in activating, rather than silencing, of associated genes.[7,8,28-30] These studies have led to models in which specific nc or short promoter RNAs are required to recruit mammalian PcG proteins to particular sites, while other classes of nc RNAs play an activating role at distinct sites. However, the involvement of ncRNAs in both silencing and activation mechanisms has high potential for dynamic regulation of PcG recruitment and function at the same sites during differentiation. Studies to date have mostly been performed in single cell types or in mixed populations of cells, and the relationship between ncRNA transcription and PcG recruitment during a single defined differentiation pathway has not been addressed.

Here we compare mammalian intergenic Polycomb target sites with sites of non-coding transcription during commitment to a defined lineage. To detect dynamic changes in the relationship between intergenic PcG binding and nc transcription, we performed profiling in purified cell populations from different stages of a defined in vitro neural differentiation system. This analysis identifies a novel class of regulatory elements, which we designate as Transcribed Intergenic Polycomb (TIP) target sites. Data from profiling, reporter assays, and RNA knockdown experiments are consistent with a model in which PcG interaction with TIP sites undergoes regulated transitions between DNA-mediated recruitment, RNA-mediated recruitment, and transcriptional antagon-

ism of both these states. This state transition model offers a conceptual framework for understanding the changing role of nc RNAs in PcG recruitment during differentiation.

Results

Analysis of transcription during in vitro neural differentiation. To investigate changes in transcription and Polycomb binding during mammalian cell differentiation, we used an in vitro differentiation system, in which neuroectodermal precursors and neurons are generated from mouse embryonic stem cells (ESCs) in adherent monoculture[31,32] (**Fig. 1A**). To obtain homogenous cell populations, we used ESC lines derived from Sox1[GFP] and Tau[GFP] knock-in mice,[33,34] enabling FACS sorting of neural progenitors and neurons (**Fig. 1B**). The purity of isolated populations was confirmed by qPCR analysis (**Fig. 1C**). To analyze changes in both genic and nearby intergenic transcription from Polycomb target regions, we designed oligonucleotide tiling arrays, covering 5kb up-and downstream of the transcription start site (TSS) of Polycomb target genes,[35,36] and additional genes of interest. Analysis of cDNA by hybridization to the array (**Fig. S1**) and by

qPCR (**Fig. 1C**) showed that the array results were fully consistent with those of the qPCR analysis. Thus, array-based analysis of transcription faithfully reflects the changes in transcription status of known genes at different stages of neural differentiation, and is therefore a reliable tool for the identification of non-annotated intergenic transcripts.

Combined analysis of transcription and Polycomb binding during neural differentiation defines Transcribed Intergenic Polycomb target sites. To systematically compare changes in transcription and PcG binding, we performed transcription and ChIP analyses at different stages of neural differentiation (**Fig. 2, S2, S3**). We mapped the PRC2 component SUZ12, the histone modification H3K27me3 as an indicator of PRC2 recruitment.[3] The antibody to H3K27me3 was more robust in ChIP on differentiated cells than the SUZ12 antibody, which led to difficulties in obtaining good ChIP signals with the SUZ12 antibody in neurons. However the H3K27me3 and SUZ12 signals in ESCs and progenitors were highly correlated (**Fig. S3F, G**) demonstrating that H3K27me3 is a good indicator of SUZ12 presence. We also mapped H3K4me3 to detect active chromatin,[37] in addition to cDNA hybridizations to detect transcription as described
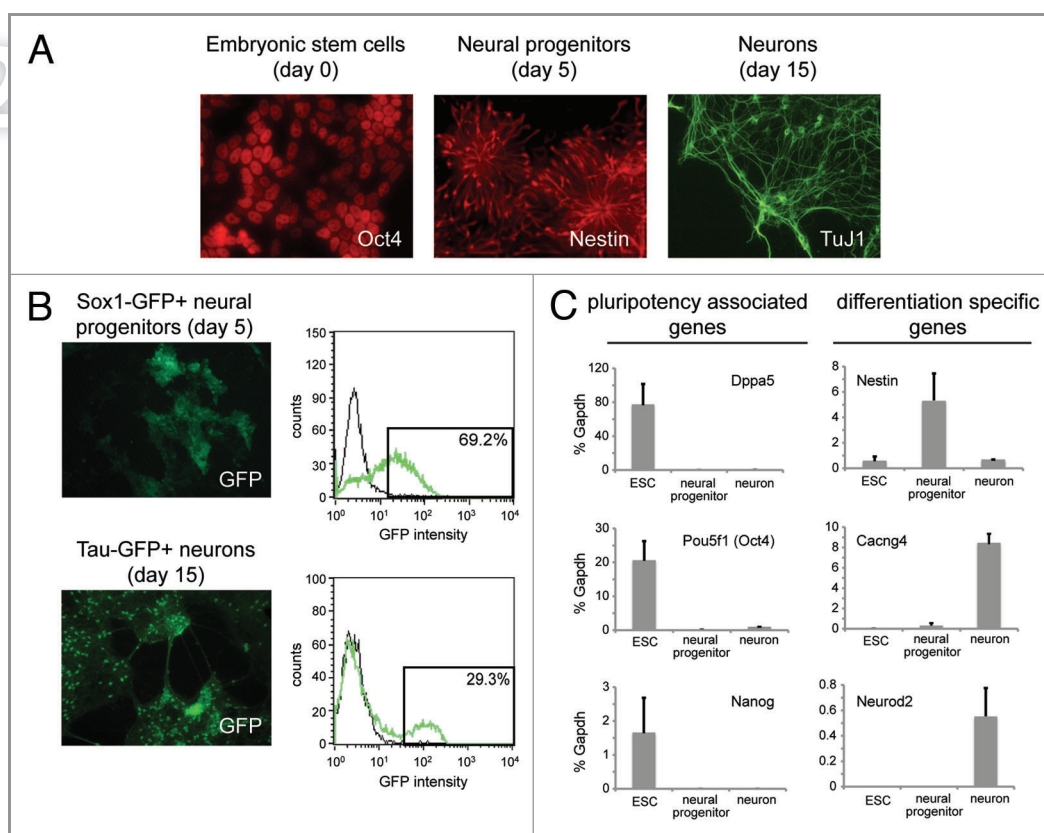

Figure 1. In vitro neural differentiation of mouse ESCs. (A) Mouse ESCs were differentiated into neural progenitors, and further, into neurons. Cell populations were visualized by immunofluorescence staining for OCT4 (ESCs), NESTIN (progenitors) and class III β-TUBULIN (TuJ1) (neurons) on days 0, 5 and 15 of differentiation. (B) Left: GFP expression from Sox1-GFP and Tau-GFP cell lines on days 5 and 15 of differentiation was observed by microscopy. Right: GFP+ cells were gated for sorting by flow cytometry. Green line: GFP positive cells. Black line: undifferentiated cells. Black box: percentages of GFP+ neural progenitors or neurons within the live cell population are shown for one representative experiment. (C) Expression of cell type specific genes in ESCs and sorted progenitors and neurons was analyzed by qPCR. Expression levels were normalized to *Gapdh*. Error bars indicate standard deviation of averages from 2 biological replicates.
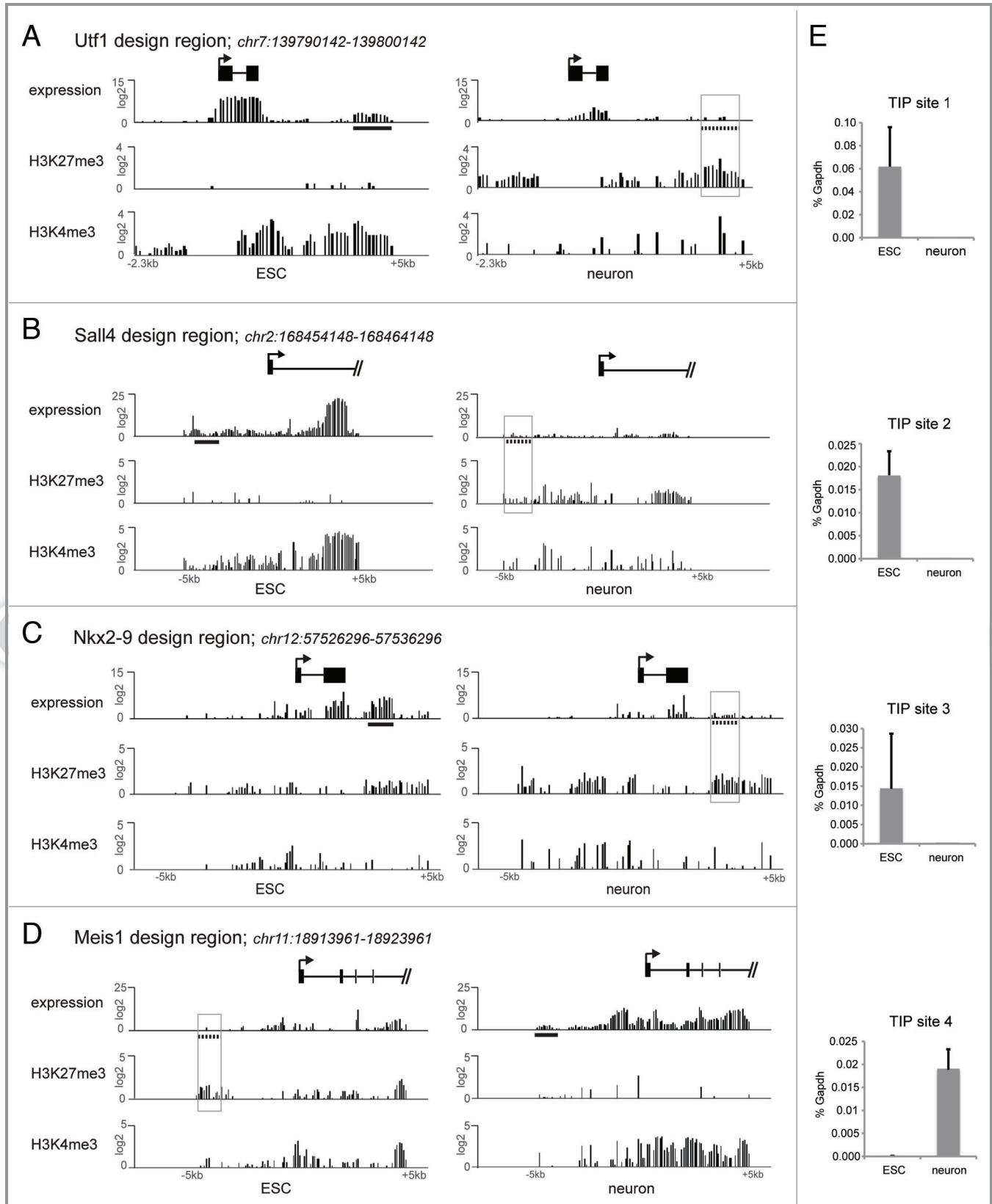
Figure 2. Identification of Transcribed Intergenic Polycomb (TIP) target sites. Expression, H3K27me3 ChIP and H3K4me3 ChIP tracks are shown for design regions (A) Utf1, (B) Sall4, (C) Nkx2–9, and (D) Meis1 Top; exon-intron structure of gene or part of the gene that is within the design region is indicated. Intergenic transcripts are indicated below expression tracks with solid black bars when expressed, and dashed black bars when not expressed in the corresponding cell type. Grey boxes show the overlap of the sites of intergenic transcripts with H3K27me3 enrichment. Each track is an average of at least 2 biological replicates. (E) qPCR analysis of transcripts from TIP sites in ESCs and neurons. Values were normalized to *Gapdh*. Error bars indicate standard deviation of averages from 2 biological replicates.

above. Interestingly, in addition to transcription from exons, we observed regulated transcription from many intergenic and intronic sites at different stages of differentiation (**Fig. 2**, **S5**). For further analysis, we focused on the intergenic transcripts, as these are readily distinguishable from transcripts arising from coding regions.

To validate the existence and the non-coding nature of these intergenic transcripts we performed several filtering steps. Transcripts were first screened for signals in both ES cell lines (see **Supplementary Information**). For all detected transcripts, comparisons to the Uniref. 90 and EST databases were performed to exclude coding transcripts (see **Supplementary Information**). Expression of intergenic transcripts from selected design regions (solid black lines in **Fig. 2A–D**) was validated by qPCR (**Fig. 2E**), confirming genuine transcription of these intergenic sites. In addition, analysis of 9 intergenic transcripts by RT-PCR confirmed that 7 were distinct from the mRNA of the adjacent gene (**Fig. S4**), thus arguing against the majority being mis-annotated exons. The above analysis indentified 238 nc transcripts arising from intergenic regions. 5' RACE analysis (**Fig. S6**) and comparison with FANTOM3 CAGE tag data fantom31p.gsc.riken. jp/cage_analysis/export/mm5 confirmed strand specificity for (**Fig. S7A–C**). The presence of CAGE tags and overlap with Pol II ChIP enrichments in ESCs[38] indicates that the majority of these ncRNAs are Pol II transcripts (**Fig. S7D**).

We next asked whether these intergenic transcripts originated from PcG target sites. Strikingly, for DNA sites of several intergenic transcripts, we detected H3K27me3 enrichment either at a different stage of differentiation than the intergenic transcription itself was detected (gray boxes in **Fig. 2A–D**) or in the same cell type as intergenic transcription (gray box in **Fig. 2C**, left). The relationship between H3K27me3 enrichment and intergenic transcription showed complex and dynamic changes during differentiation. This dynamic behavior is documented in **Fig. 3** and is described in detail below. In total, of the 238 intergenic transcripts detected at different stages of differentiation, 138 transcripts overlapped with intergenic PcG binding sites at one or other differentiation stage. Thus over 50% of regulated intergenic transcription was associated with PcG binding. We designate these sites as Transcribed Intergenic Polycomb (TIP) target sites. A full list of TIP sites is provided in **Table S1**.

We found that conservation of TIP sites was lower than that of exons but higher than intronic sequences, implying evolutionary constraint (**Fig. S8A**). It has been proposed that CpG islands in H3K27me3 enriched promoters are required for PcG recruitment.[3,17,18,20] Surprisingly, we found that only a small minority of TIP sites contained a predicted CpG island (**Fig. S8B and C**), indicating that CpG islands are not required for recruitment of PcG to the majority of TIP sites. In summary this analysis identifies a novel class of intergenic, dynamically regulated PcG recruiting sites that are conserved, that do not contain CpG islands, and that give rise to non-coding transcripts during differentiation.

Sites of TIP transcription correspond precisely to sites of PcG recruitment. To analyze the localization of PcG on TIP sites, we examined average H3K27me3 enrichment profiles. The length of TIP transcripts detected on the array ranged between 400 bp and 3 kb (**Fig. S8D**). Each TIP transcription site was scaled to a unit length, and the average H3K27me3 enrichment across this unit length and the flanking regions was plotted (**Fig. 3A–C**). Similar data were obtained for SUZ12 in ESCs and progenitors (**Fig. S3D, E**). Remarkably, both H3K27me3 and SUZ12 showed a specific average enrichment profile, precisely co-localizing with the region in which transcription was detected, and with sharp depletion on the flanking regions. This indicates a precise correspondence between sites of TIP transcription and sites of PcG recruitment.

TIP transcription correlates both negatively and positively with PcG recruitment. We observed that TIP transcription levels changed dynamically during differentiation (**Fig. S5**). To examine the relationship between PcG recruitment and TIP transcription, we compared H3K27me3 enrichments to TIP transcript levels (**Fig. 3D–F**). This analysis revealed three main categories present in each cell type: TIPs that were enriched for H3K27me3 in the absence of detectable transcription (colored black in **Fig. 3D–F**); TIPs for which both H3K27me3 and transcription were present (green in **Fig. 3D–F**), and TIPs showing transcription but no detectable H3K27me3 enrichment (red in **Fig. 3D–F**). The four TIPs shown in **Fig. 2** are plotted on **Fig. 3** as examples. Furthermore, we observed that each cell type favored particular TIP categories (**Fig. 3G**). For example, the proportion of TIPs with H3K27me3 enrichment but no detectable transcription (black on **Fig. 3G**) decreased from ESCs to progenitors to neurons, while the proportion of TIPs with both H3K27me3 and transcription (green on **Fig. 3G**) increased during differentiation.

To examine the relationship between TIP status and nearby gene activity we analyzed the transcription of coding exons within the design region containing each TIP (**Fig. 3H**). This analysis revealed a clear correlation between TIP status and gene activity in ESCs, suggesting co-regulation. Interestingly, gene expression status and TIP status were no longer correlated in progenitors and neurons, suggesting that TIP regulation and gene regulation become uncoupled at later stages. Taken together these results indicate that TIPs are dynamically regulated during differentiation, and demonstrate that PcG recruitment to TIPs can be both negatively and positively correlated with TIP transcription status.

The relationship between TIP transcription and PcG recruitment changes dynamically upon differentiation. The above observations prompted us to ask whether the relationship between TIP transcription and PcG recruitment is TIP-specific, or whether TIPs can change this status upon differentiation. To address this, we examined the behavior of each class of TIPs (black, green or red) upon differentiation. In **Fig. 3I–N**, TIPs are separated according to their class in ESCs (**Fig. 3D, I, K, M**) or in progenitors (**Fig. 3E, J, L, N**). Changes in transcription and H3K27me3 enrichments upon the ESC-progenitor transition (**Fig. 3I, K, M**) and progenitor-neuron transition (**Fig. 3J, L, N**) are plotted for each class. Grey boxes depict transitions in which the relationship between TIP transcription and H3K27me3 recruitment is preserved upon differentiation. For example, many TIPs of the "black" class in ESCs, showing H3K27me3
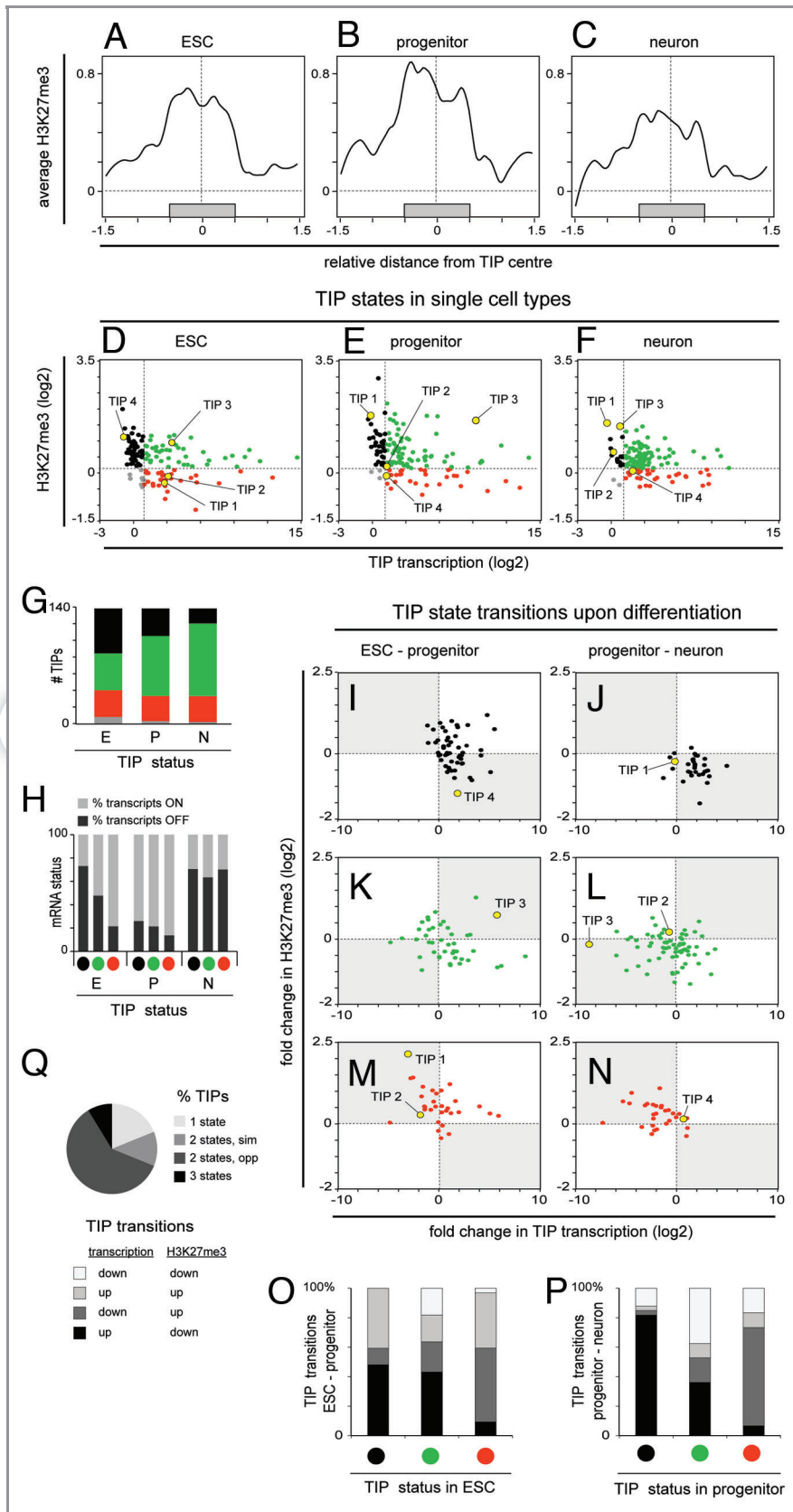
Figure 3. TIP state transitions upon differentiation. (A-C) TIP sites are shown scaled to unit length. Average H3K27me3 enrichment profiles across TIPs and flanking region are shown for each cell type, for those TIPs for which the H3K27me3 enrichment score averaged over the whole region was higher than 0.1. (D–F) Scatter plots showing log2 H3K27me3 enrichment (vertical scale) and log2 transcription (horizontal scale; see **Supplementary Information** for calculation of transcription levels) for TIPs in each cell type. Enrichments were calculated as average values across each TIP site from microarray data. Data points are colored according to category. Dotted line is set at 0.1 for H3K27me3 enrichment and at 1.0 for transcription, based on qPCR validation of array data. Yellow data points show TIPs 1–4, illustrated in **Figures 2 and 5**. (G) Summary of (D,E,F) according to color code described above. E, ESC; P, progenitor; N, neuron. (H) Expression of coding genes in each cell type (E,P,N) separated according to TIP status (black, green or red). % of total ON and OFF coding exons within each TIP category is shown for each cell type (see **Supplementary Information**). (I-N) TIP state transitions upon differentiation are shown as log2 fold change in H3K27me3 enrichment (vertical scale) and log2 fold change in TIP transcription (horizontal scale). TIPs are separated and color coded according to their status in ESCs (D;I,K,M) or in progenitors (E;J,L,N). Grey boxes indicate transitions that are consistent with the relationship between H3K27me3 and transcription in the starting cell type (see main text for details). (O,P) Summary of data in J-N. (Q) Pie chart showing number of TIPs that occupy one, two or three states (black, red or green as defined in D-E), during differentiation. "2 states, sim": TIPs occupy red and black states, indicating similar relationship between H3K27me3 and transcription in the two states (i.e., both states show an antagonistic relationship). "2 states, opp": TIPs occupy either black and green or red and green states, indicating opposite relationships between H3K27me3 and transcription in the two states (ie one state shows an antagonistic relationship, the other a permissive relationship).

enrichment but no detectable transcription (**Fig. 3D**), showed a gain of transcription and loss of H3K27me3 upon the ESC to progenitor transition (gray box, TIP4, **Fig. 3I**), consistent with an antagonistic relationship between PcG recruitment and transcription in both cell types. However, other TIPs of this class showed a gain of transcription and gain of H3K27me3 upon differentiation (white box, upper right in **Fig. 3I**), consistent with a switch from an antagonistic relationship toward a permissive or cooperative relationship between transcription and PcG recruitment. This analysis showed that with the exception of "black" and "red" TIPs in the progenitor-neuron transition (gray boxes in **Fig. 3J and N**), all classes of TIPs contain many that switch the relationship between H3K27me3 and transcription from one cell type to the next (white boxes in **Fig. 3I–N**, summarized in **Fig. 3O and P**).

To gain an overview of TIP behaviors for all three cell types, we examined the number of states ("black," "red" or "green"), that were occupied by each TIP during differentiation. Interestingly, 80% of TIPs switched between at least two states (**Fig. 3Q**). Furthermore, 85% of these transitions resulted in a switch between negatively and positively correlated relationships of TIP transcription to H3K27me3 recruitment. Taken together these results demonstrate that for most TIPs, the relationship between PcG recruitment and TIP transcription is not TIP-specific but changes dynamically upon changes in cell identity.

TIP sites repress transcription of a reporter gene. PcG target sites have been shown to repress a flanking reporter gene.[16] In contrast, transcribed intergenic sites have been shown to activate reporters.[29] To address the effect of TIP sites on reporter activity, we used a luciferase assay. TIP site 1 and TIP site 3 (Utf1 and Nkx2–9 design regions, respectively; **Fig. 2A and C**), were cloned upstream of a luciferase reporter in either forward (fwd) or reverse (rev) orientation with respect to the direction of transcription
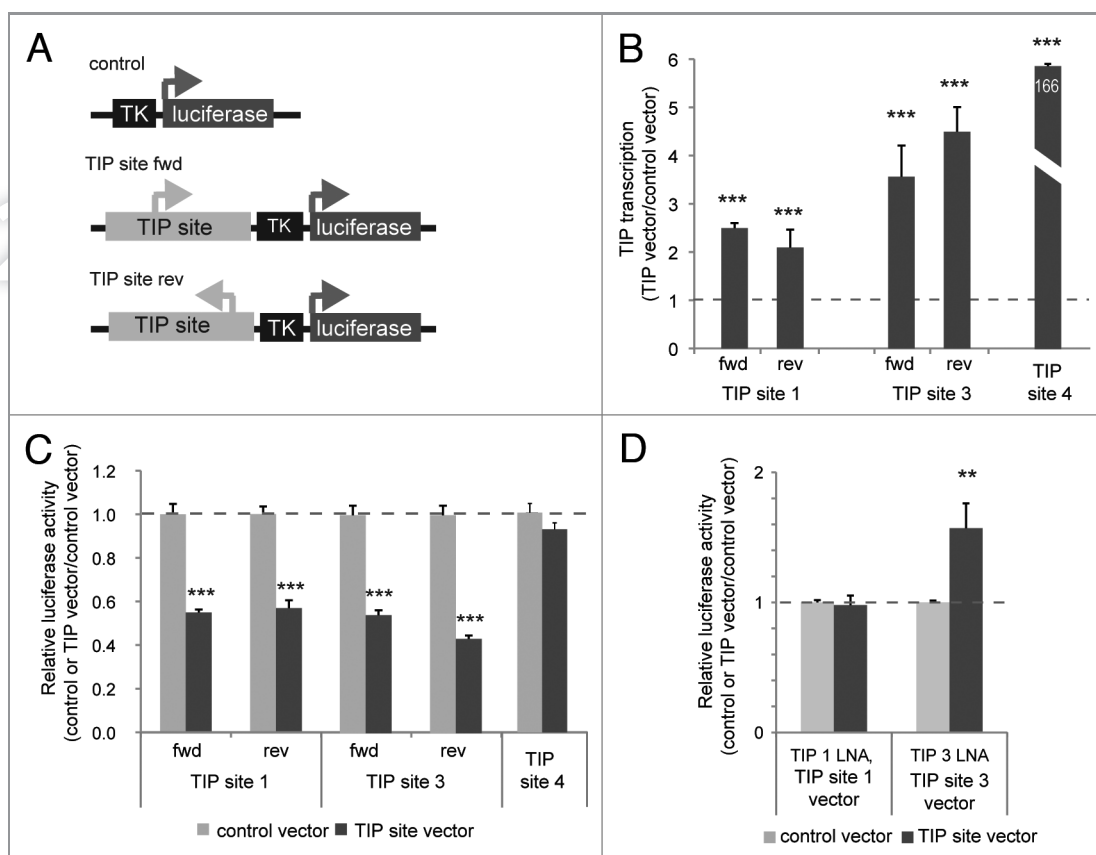

Figure 4. TIP sites repress transcription from a reporter construct. (A) Luciferase reporter constructs with Thymidine Kinase promoter (TK) are shown schematically for the control vector (control) and for the TIP site vectors that contain the TIP site either in forward (TIP site fwd) or in reverse (TIP site rev) orientation with respect to the direction of transcription from the TIP site. TIP site 4 showed transcription from both strands, and was cloned in one orientation only. (B) qPCR analysis of transcription from TIP site 1, TIP site 3 and TIP site 4 reporter constructs are shown as ratio of transcript detected upon transfection of the TIP site vector to endogenous levels detected upon transfection of the control vector. Values above 1 (dotted line) indicate transcription from the TIP site vector above endogenous levels. Error bars indicate +/−SEM of averages from 3 biological replicates. ***p < 0.001. (C) The change in luciferase expression upon transfection of TIP site 1 fwd/rev, TIP site 3 fwd/rev or TIP site 4 constructs into ESCs is shown as fold change relative to the control vector, taking molar concentrations into account. At least 2 plasmid preparations were used for each data point. Error bars indicate +/−SEM of averages from at least 6 biological replicates. ***p-value < 10^{-8} by two-tailed Student's t test. (D) The effect of knock-down of TIP 1 and 3 transcripts on luciferase gene expression is shown as the ratio between luciferase activity of the vector of interest (light gray bars: control vector; dark gray bars: TIP site vector) and that of the control vector upon LNA knockdown, normalized against control LNA knockdown. Control LNA or specific TIP site LNA oligos were co-transfected with control or TIP site rev vectors. Error bars indicate +/−SEM of averages from at least 4 biological replicates. **p-value < 0.005 by two-tailed Student's t test.

from the TIP sites (**Fig. 4A**). TIP site 4 (Meis 1 design region) showed transcription from both strands (**Fig. S6**), and was cloned in one orientation only. Upon transfection of each TIP construct into ESCs, expression of the corresponding TIP transcripts were detected above endogenous levels, indicating that the cloned TIP site regions contain active promoters (**Fig. 4B**). Interestingly, TIP sites 1 and 3, which were transcribed at moderate levels (between 2 and 5 fold above background) gave substantial repression of the reporter compared with the control vector lacking the TIP site (**Fig. 4C**). Furthermore, this repression was independent of the direction of transcription from the TIP sites as both fwd and rev constructs showed comparable levels of repression (**Fig. 4C**), indicating that these TIP sites do not repress by transcriptional interference with the luciferase promoter. In contrast, TIP site 4, which was highly transcribed in ESCs (166-fold above background) did not repress the reporter. This result again indicates that transcription from a TIP site does not interfere with transcription from the luciferase promoter. However, the high level of transcription from TIP site 4 also did not lead to activation of the reporter, suggesting that TIP and luciferase transcription are not simply co-dependent. We note that the high transcription observed from TIP site 4 in the reporter context is in contrast to the endogenous situation in ESCs, in which transcription of TIP site 4 was not detected (**Fig. 3D**). This result suggests that the endogenous TIP site 4 locus is subject to repression via additional sequences that are not present in the cloned reporter construct. Transfection of selected constructs in N2A cells did not show repression of luciferase (**Fig. S9A**), demonstrating that the repression by TIP sites in ESCs is cell type specific.

We next asked whether repression of the reporter by TIP sites 1 and 3 is dependent on the TIP transcripts. To investigate the effect of the transcripts themselves rather than an act of transcription running toward the TK promoter, we used the TIP site rev constructs (**Fig. 4A**). Custom design Locked Nucleic Acid (LNA) oligos were used to knock down TIP 1 or TIP 3 transcripts. The level of luciferase expression after knockdown of the TIP site transcript was compared with its level detected after transfection with the control scrambled LNA oligo. Depletion of the TIP 1 transcript (**Fig. S9B**) had no effect on luciferase expression (**Fig. 4D**). In contrast, depletion of the TIP 3 transcript lead to an increase of reporter activity of approximately 1.5 fold (**Fig. 4D**). Together, these results indicate that when placed adjacent to a reporter gene, TIP sites 1 and 3 repress transcription, and that the TIP 3 transcript is required for full repression.

Endogenous TIP RNAs repress target genes in *trans*. To address the role of non-coding transcripts from endogenous TIP sites, we performed knock-down experiments in ESCs using locked nucleic acid (LNA) oligos against TIP 1 and TIP 3 transcripts, and analyzed changes in gene expression. To evaluate whether the transcripts are required for regulation of their nearby genes, we performed qPCR on the genes flanking TIP1 and TIP3. However, no change in expression level of these genes was detected upon knockdown of the corresponding transcript. This result was surprising in view of the requirement for the TIP 3 transcript for *cis*-repression in the reporter assay. Thus the

endogenous TIP 3 locus may be subject to additional layers of regulation that are not recapitulated in the reporter assay.

In order to address whether the TIP 1 and TIP 3 transcripts have regulatory roles elsewhere in the genome, we performed expression microarray analysis, and validated results by qPCR (**Figs. 5A–D**). This analysis identified several genes that were upregulated in both TIP 1 and 3 knockdowns. Successful knockdown of TIP 1 RNA (**Fig. 5A**) resulted in upregulation of a lincRNA transcript and the *Syn1* gene (**Fig. 5B**). Similarly, knockdown of TIP 3 RNA (**Fig. 5C**) led to upregulation of the genes *Lzic* and *Stox2* (**Fig. 5D**). Thus these targets are sensitive to the levels of a single ncRNA, and may be poised for activation upon TIP downregulation. To ask how these genes behave upon downregulation of the endogenous TIP transcript during differentiation, we analyzed expression levels in ESCs and neurons. Both TIP 1 and TIP 3 transcripts were downregulated in neurons. LincRNA expression did not change significantly during differentiation from ESCs to neurons (**Fig. 5E**). However, expression of the genes *Syn1*, *Lzic* and *Stox2* were dramatically upregulated in neurons (**Figs. 5E and F**). These results demonstrate that TIP transcripts are required for repression of specific targets in ESCs in *trans*, and are consistent with a model in which this repression is released during differentiation into neurons as the TIP transcript itself is downregulated.

Discussion

Using a combination of transcription and ChIP analysis in a well-defined in vitro neural differentiation system, we identify a novel class of vertebrate PcG target sites, which we designate as TIP (Transcribed Intergenic Polycomb) target sites. Unlike transcribed intergenic sites in the mouse and human Hox complexes[7,8] TIP sites recruit PcG proteins precisely to their site of transcription (**Fig. 3A–C; S3D and E**). They are conserved, and are distinct from promoter-proximal PcG target sites in terms of CpG island content (**Fig. S8**).[17,18,20]

Our results highlight parallels between fly and vertebrate PcG function, consistent with a recent study in Drosophila.[9] These authors showed by ChIP seq profiling of PcG proteins in Drosophila S2 cells, that a large proportion (approximately 50%) of PcG binding sites map to annotated coding transcription start sites, similar to the situation in vertebrate ES cells: 73% of total SUZ12 peaks map to annotated promoters.[19] In the Drosophila study,[9] profiling of embryos and S2 cells revealed that a further 10% of total PcG binding sites mapped to intergenic and intronic sites of transcription initiation. The majority of fly PRE elements that have so far been functionally characterized are found in intronic or intergenic positions, and many are transcribed into non coding RNA in a developmentally regulated manner (reviewed in 10,12). This raises the intriguing possibility that TIP sites may represent a class of vertebrate PRE elements that are analogous to the intergenic fly PREs. Stable integration of TIP site reporters into chromatin will be required to address the question of whether they share properties of mammalian PREs,[15,16] and whether their function is dependent on RNA.
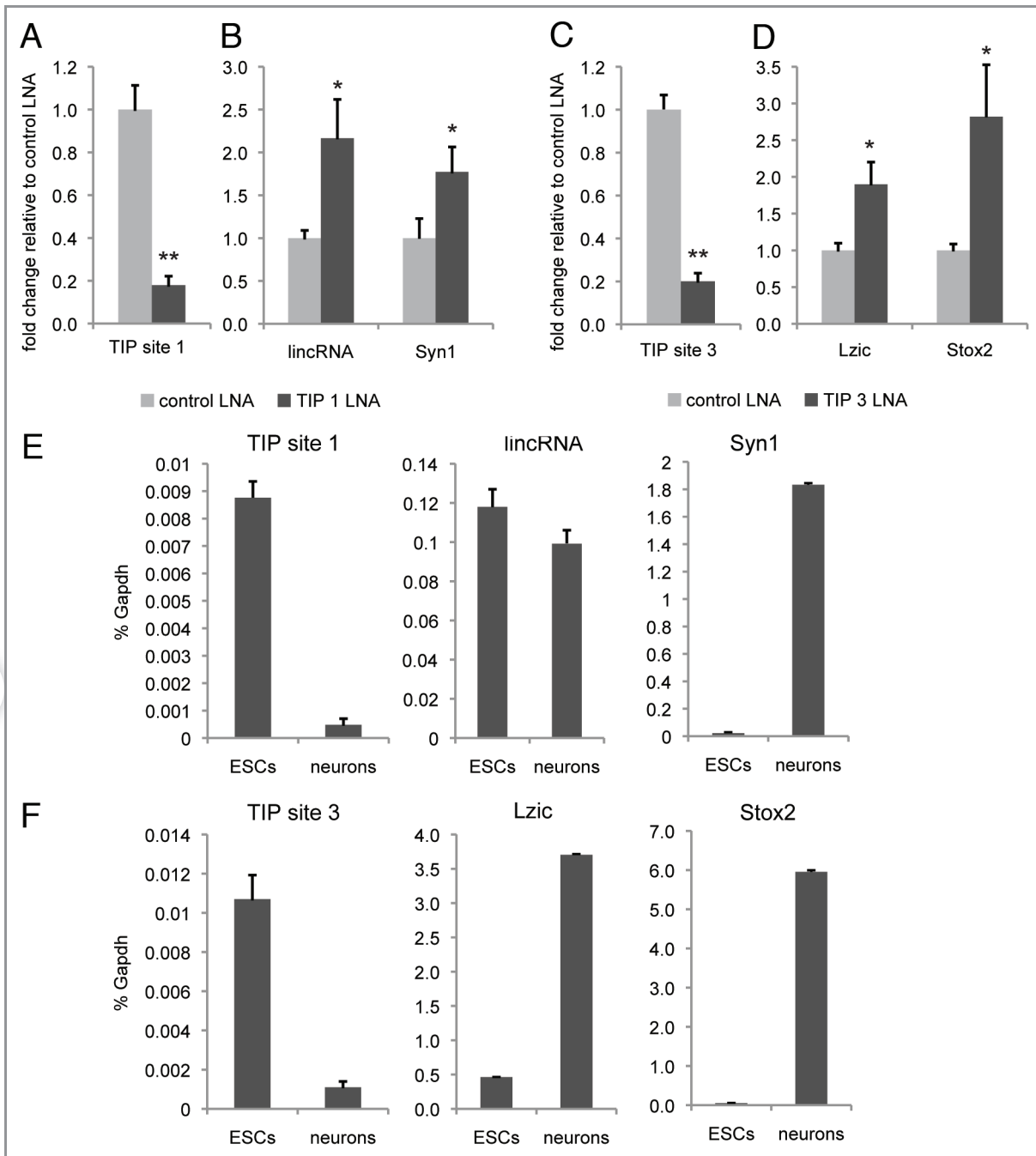
Figure 5. Transcripts from endogenous TIP sites repress their target transcripts in *trans*. qPCR analysis of LNA knock-down experiments on (A) TIP 1 transcript (from the Utf1 design region) and (C) TIP 3 transcript (from the Nkx2–9 design region) are shown as fold change relative to control LNA treatment of cells. (B) Validation of the effect of LNA knock-down of TIP 1 transcript on the targets lincRNA (mm9 coordinates; chr5:137392195–137426421) and *Syn1* (mm9 coordinates; chrX:20437637–20498022). (D) Validation of the effect of LNA knock-down of TIP 3 transcript on the targets *Lzic* (mm9 coordinates; chr4:148859338–148870777) and *Stox2* (mm9 coordinates; chr8:48265402–48437702). Error bars indicate standard deviation of averages from at least 2 biological replicates. *p $<$ 0.05, **p $<$ 0.0005 by one-tailed Student's t test in A and C, and by two-tailed Student's t test in B and D. (E) Expression levels of TIP 1, lincRNA and *Syn1* gene in ESCs and in neurons. (F) Expression levels of TIP 3, *Lzic* and *Stox2* genes in ESCs and in neurons. Error bars indicate standard deviation of averages from two cDNA preparations.

In addition, we document transitions in the transcription state and PcG occupancy of TIP sites upon differentiation (**Fig. 6**). Due to our experimental design, TIPs that were not included in the array, or that are not transcribed or do not recruit PcG in the neural differentiation pathway we have studied would not have been detected. Thus we propose that many more intergenic genomic sites[19] may act as TIPs in different lineages. Further mapping of intergenic transcription and PcG binding
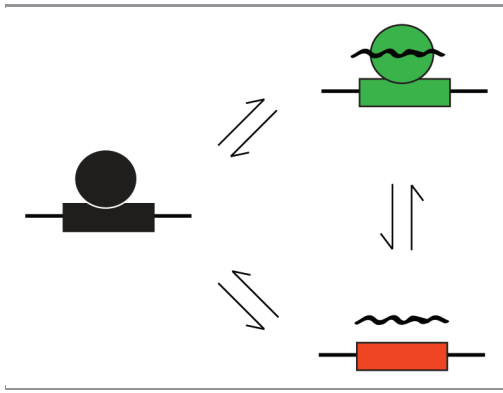
Figure 6. Proposed model for TIP state transitions. TIP sites (rectangles) in black, green or red states are shown. Black: RNA-independent PcG recruitment. Green: RNA mediated PcG recruitment. Red: the act of transcription prevents PcG binding. Arrows show transitions between states. See main text for details.

in distinct differentiation systems would enable this model to be tested.

Black TIPs recruit PcG in the absence of transcription. In all three cell types, a proportion of TIP sites showed PcG recruitment in the absence of detectable transcription (black in **Fig. 3D–F**). qPCR validation of transcript levels for selected TIPs showed that the array gave a faithful readout of transcript levels (**Fig. 2E**), thus we conclude that these TIPs recruit PcG proteins in the absence of transcription of the TIP itself (**Fig. 6**). This observation suggests either that PcG proteins repress TIP transcription, or alternatively that the absence of transcription is permissive for PcG binding. However, siRNA mediated knockdown of the PRC2 proteins Suz12 and Ezh2 in ESCs did not activate four of the TIP transcripts that are normally inactive (*Meis1, Phox2b, Khx5 and Cxxc4;* data not shown). This result suggests that the repression of these TIP sites may be regulated by other factors than PcG proteins. An alternative reason may be that additional factors, which are normally upregulated only upon differentiation, are required to antagonize PcG repression and activate the TIP sites. In this case, downregulation of PcG proteins in ESCs would not be sufficient to activate the TIP sites. It is not possible to distinguish between these two hypotheses on the basis of the present data. Future work will aim to elucidate how TIP sites are regulated at the transcriptional level.

Recruitment of PcG to this class of TIPs may occur through other RNAs in *trans*,[8,24,26] or through DNA-protein interactions in the absence of an RNA cofactor.[16,20] If TIPs recruit PcG directly via DNA binding proteins, this may represent a default state, that can be switched to a regulatable state upon TIP transcription. The TIP sites represent a class of DNA elements that recruit PcG proteins but are distinct from coding gene promoters, and thus may offer an interesting data set for sequence mining. Future work will aim to analyze the motif content of TIP sites to identify sites for DNA binding factors, which may be involved in recruitment of PcG proteins to these sites. The observation that TIP sites contain evolutionarily conserved nucleotides within

rodent and primate species will provide a starting point for this analysis.

Red TIPs are transcribed and lack PcG. In each cell type, a proportion of TIPs showed transcription but lacked detectable H3K27me3 (red in **Figure 3D–F**). Thus TIP transcription at these sites may remove PcG proteins, either by recruitment of activators,[28-30,39] or by removal of PcG proteins upon the act of transcription.[13] A similar relationship between intergenic transcription and removal of PcG proteins has been observed in the mouse and human HOX complexes.[7,8] Importantly in,[7] transcription was shown to occur before PcG removal, supporting a model in which the act of transcription counteracts PcG occupancy (**Fig. 6**). TIP site 1 is transcribed and lacks PcG in ESCs (**Fig. 3D**), and shows loss of transcription and gain of PcG recruitment upon differentiation to progenitors (**Fig. 3E**), consistent with an antagonistic relationship between TIP transcription and PcG binding. However, in the reporter assay, TIP site 1 was transcribed in ESCs and gave repression of the reporter, suggesting that when taken out of context, this TIP site may adopt an intermediate state in ESCs, in which transcription and PcG binding are in balance. The results from the knockdown experiments, in which removal of the TIP transcript did not result in loss of repression, argue against a role for the transcript itself in recruiting repressors.

Green TIPs are transcribed and recruit PcG. In each cell type we observed TIPs that were transcribed, and enriched for H3K27me3 (green in **Fig. 3D–F**). Interestingly, the proportion of TIPs in this class increased during differentiation (**Fig. 3G**). Several lines of evidence suggest that these TIP transcripts themselves may be involved in recruiting PcG to these sites. First, we show in reporter assays that the repressive activity of TIP site 3 is dependent on the non-coding transcript itself (**Fig. 4D**), thus the transcript may be involved in recruiting PcG proteins to the reporter. However, these results were obtained by transient transfection of plasmids, which may not be correctly chromatinised, and thus may not recapitulate important aspects of PcG regulation. A key goal for future studies will be to integrate TIP reporters into chromatin, to examine PcG recruitment and requirement for TIP RNA in the correct chromatin context.

At the endogenous TIP site 3, we observed an increase of TIP transcription and H3K27me3 recruitment in the ESC to progenitor transition (**Fig. 3K**), again consistent with the idea that the TIP RNA may recruit PcG to the site. In addition, upon comparison with a recent analysis of transcripts that were found to be physically associated with PRC2 in mouse ESCs,[24] we found 32 TIP transcripts (**Table S1**). Many of these TIP sites were also enriched for H3K27me3 in our ESC data set, and thus may recruit PRC2 to their own site of transcription (**Fig. 6**). In summary, these data are consistent with RNA mediated recruitment of PcG proteins to TIP sites, but do not prove conclusively that this is the case. Future work will aim to elucidate the molecular mechanisms by which endogenous TIP sites recruit PcG proteins, and the extent to which TIP RNAs are involved in this process.

TIP state transitions. One of the most surprising findings from this study is the demonstration that many TIP sites undergo

transitions between opposing states upon differentiation, in which transcription of the same TIP site can accompany PcG recruitment or removal depending on cellular context (**Fig. 6**). Approximately 2/3 of all TIPs showed this behavior (**Fig. 3Q**). We show in reporter assays in ESCs, that TIP sites 1 and 3 were both transcribed, and repressed the luciferase gene (**Fig. 4B and C**). For TIP site 3 we show that the TIP transcript itself is required for repression. However we also show that the repressive function of both TIP sites tested is cell-type dependent (**Fig. 4C, S9A**), thus we propose that additional cell type specific factors determine the repressive activity of TIP sites. It will be extremely interesting in future to examine the behavior of TIP reporters integrated into the genome, upon differentiation of ESCs to progenitors and neurons, or to other pathways, to further examine the role of cell identity in TIP activity. Other sites of non-coding transcription have recently been shown to have enhancer-like functions, with direct contribution of ncRNAs to this activity.[29] Thus whether TIP elements act as transcriptional enhancers in other cell types will be an important question for future studies.

It will also be of great interest to identify cell type specific factors that determine the effect of TIP transcription on PcG recruitment or removal. These may include for example, the degradation rate of the RNA itself, which should influence the extent of RNA mediated recruitment, but not the extent of transcription-mediated PcG removal. Similarly, recruitment may be regulated by the RNA binding properties of the PcG, which have been shown to depend on PcG phosphorylation status and to be cell-cycle dependent.[26] Finally the presence of cell-type specific bridging factors, that may link PcG-bound ncRNAs to DNA sites[27] will be a key question for future studies.

How does TIP regulation relate to gene regulation? By profiling, we observe a correlation between PcG recruitment to transcriptionally silent TIPs and silencing of flanking genes in ESCs (**Fig. 3H**). These findings suggest a model in which transcriptionally silent TIP sites contribute to silencing of their neighboring genes in ESCs by recruiting PcG proteins to the locus via direct DNA mediated PcG recruitment. At later developmental stages, at the global level, TIP transcription increases, and gene transcription decreases, thus TIP and gene transcription states become uncoupled (**Fig. 3H**). We propose that this uncoupling may be an essential consequence of TIP state transitions: if a TIP switches to the green state, in which PcG proteins are recruited by TIP transcription, then nearby genes may be silenced by PcG action, while the TIP itself is transcriptionally active. This idea is consistent with the observation that the proportion of green TIPs increases upon differentiation, concomitant with an increase in silencing of nearby genes (**Fig. 3G, H**).

However, it is important to note that these hypotheses are based on correlations in the whole data set, and will require experimental testing based on integrated reporters for specific TIPs, and evaluation of their behavior in terms of TIP transcription, reporter transcription, and PcG recruitment during differentiation. By identifying the novel class of TIP elements, and demonstrating dynamic transitions in the behavior of the

endogenous sites, our work provides a starting point for the design of reporter assays that will enable the dissection of the molecular mechanisms underlying these transitions.

If correct, the model we propose, in which TIP transcription becomes uncoupled from coding gene transcription during differentiation, has two important implications for gene regulation. First, it enables RNA mediated PcG recruitment mechanisms to operate at transcribed TIP sites, independently of adjacent coding gene transcription. Second, we have shown that TIP transcripts repress a highly specific set of target genes in *trans* (**Fig. 5**). Thus TIP transcription may ensure coordinated regulation of gene networks via recruitment of PcG proteins both in *cis* and in *trans*. Indeed, the *Stox2* target gene of TIP site 3 is a PcG target in ESCs[40] suggesting that the TIP site 3 transcript may be a key component of PcG recruitment at this locus.

Materials and Methods

Cell culture and in vitro neural differentiation 46c (Sox1-GFP) and Tk23 (Tau-GFP) cell lines were cultivated feeder-free in 10% FCS medium and neural differentiation was performed essentially as described previously.[31] Details of differentiation and immunofluorescence staining are provided in Supplementary Information.

FACS purification. 46c (Sox1-GFP) and Tk23 (Tau-GFP) cells were trypsinized (0.1%) on days 5 and 15 of differentiation, respectively. Viable cells were gated by their forward and side scatter characteristics, and gates were set to purify either Sox1-GFP+ or Tau- GFP+ cells.

Chromatin Immunoprecipitation (ChIP) and antibodies. The ChIP protocol was adapted from Martens et al.[41] as described in Supplementary Information. Antibodies used were αH3K27me3 (kind gift from Thomas Jenuwein), αH3K4me3 (Diagenode), αSUZ12 (Abcam) and control IgG (Abcam).

Array design and bioinformatics. Details of array design, steps of normalization and detection of intergenic transcripts are provided in Supplementary Information.

Sample preparation for microarray experiments. For transcription experiments, total RNA sample of each cell type was amplified using the MessageAmp II kit (Ambion) according to the manufacturer's instructions and cDNA was prepared using the Retroscript kit (Ambion) with an equal mixture of random and oligo-dT primers. Genomic DNA was used as background control. For ChIP-on-chip experiments, ChIP DNA was amplified by whole genome amplification (GenomePlex Complete WGA2, Sigma). These samples were also used in site-directed qPCR for validation of microarray enrichments (**Fig. S2**). Sequences of validation primers are available on request.

Supplemental Materials

Supplemental materials can be found at:
www.landesbioscience.com/journals/rnabiology/article/19102

References

1. Fisher CL, Fisher AG. Chromatin states in pluripotent, differentiated, and reprogrammed cells. Curr Opin Genet Dev 2011; 21:140-6; PMID:21316216; http://dx.doi.org/10.1016/j.gde.2011.01.015

2. Margueron R, Reinberg D. The Polycomb complex PRC2 and its mark in life. Nature 2011; 469:343-9; PMID:21248841; http://dx.doi.org/10.1038/nature09784

3. Mohn F, Weber M, Rebhan M, Roloff TC, Richter J, Stadler MB, et al. Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. Mol Cell 2008; 30:755-66; PMID:18514006; http://dx.doi.org/10.1016/j.molcel.2008.05.007

4. Schwartz YB, Pirrotta V. Polycomb complexes and epigenetic states. Curr Opin Cell Biol 2008; 20:266-73; PMID:18439810; http://dx.doi.org/10.1016/j.ceb.2008.03.002

5. Schwartz YB, Kahn TG, Stenberg P, Ohno K, Bourgon R, Pirrotta V. Alternative epigenetic chromatin states of polycomb target genes. PLoS Genet 2010; 6:e1000805; PMID:20062800; http://dx.doi.org/10.1371/journal.pgen.1000805

6. Bae E, Calhoun VC, Levine M, Lewis EB, Drewell RA. Characterization of the intergenic RNA profile at abdominal-A and Abdominal-B in the Drosophila bithorax complex. Proc Natl Acad Sci U S A 2002; 99:16847-52; PMID:12481037; http://dx.doi.org/10.1073/pnas.222671299

7. Sessa L, Breiling A, Lavorgna G, Silvestri L, Casari G, Orlando V. Noncoding RNA synthesis and loss of Polycomb group repression accompanies the colinear activation of the human HOXA cluster. RNA 2007; 13:223-39; PMID:17185360; http://dx.doi.org/10.1261/rna.266707

8. Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell 2007; 129:1311-23; PMID:17604720; http://dx.doi.org/10.1016/j.cell.2007.05.022

9. Enderle D, Beisel C, Stadler MB, Gerstung M, Athri P, Paro R. Polycomb preferentially targets stalled promoters of coding and noncoding transcripts. Genome Res 2011; 21:216-26; PMID:21177970; http://dx.doi.org/10.1101/gr.114348.110

10. Hekimoglu B, Ringrose L. Non-coding RNAs in polycomb/trithorax regulation. RNA Biol 2009; 6:129-37; PMID:19270511; http://dx.doi.org/10.4161/rna.6.2.8178

11. Müller J, Kassis JA. Polycomb response elements and targeting of Polycomb group proteins in Drosophila. Curr Opin Genet Dev 2006; 16:476-84; PMID:16914306; http://dx.doi.org/10.1016/j.gde.2006.08.005

12. Lempradl A, Ringrose L. How does noncoding transcription regulate Hox genes? Bioessays 2008; 30:110-21; PMID:18200528; http://dx.doi.org/10.1002/bies.20704

13. Schmitt S, Prestel M, Paro R. Intergenic transcription through a polycomb group response element counteracts silencing. Genes Dev 2005; 19:697-708; PMID:15741315; http://dx.doi.org/10.1101/gad.326205

14. Petruk S, Sedkov Y, Riley KM, Hodgson J, Schweisguth F, Hirose S, et al. Transcription of bxd noncoding RNAs promoted by trithorax represses Ubx in cis by transcriptional interference. Cell 2006; 127:1209-21; PMID:17174895; http://dx.doi.org/10.1016/j.cell.2006.10.039

15. Sing A, Pannell D, Karaiskakis A, Sturgeon K, Djabali M, Ellis J, et al. A vertebrate Polycomb response element governs segmentation of the posterior hindbrain. Cell 2009; 138:885-97; PMID:19737517; http://dx.doi.org/10.1016/j.cell.2009.08.020

16. Woo CJ, Kharchenko PV, Daheron L, Park PJ, Kingston RE. A region of the human HOXD cluster that confers polycomb-group responsiveness. Cell 2010; 140:99-110; PMID:20085705; http://dx.doi.org/10.1016/j.cell.2009.12.022

17. Tanay A, O'Donnell AH, Damelin M, Bestor TH. Hyperconserved CpG domains underlie Polycomb-binding sites. Proc Natl Acad Sci U S A 2007; 104:5521-6; PMID:17376869; http://dx.doi.org/10.1073/pnas.0609746104

18. Ku M, Koche RP, Rheinbay E, Mendenhall EM, Endoh M, Mikkelsen TS, et al. Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. PLoS Genet 2008; 4:e1000242; PMID:18974828; http://dx.doi.org/10.1371/journal.pgen.1000242

19. Peng JC, Valouev A, Swigut T, Zhang J, Zhao Y, Sidow A, et al. Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. Cell 2009; 139:1290-302; PMID:20064375; http://dx.doi.org/10.1016/j.cell.2009.12.002

20. Mendenhall EM, Koche RP, Truong T, Zhou VW, Issac B, Chi AS, et al. GC-rich sequence elements recruit PRC2 in mammalian ES cells. PLoS Genet 2010; 6:e1001244; PMID:21170310; http://dx.doi.org/10.1371/journal.pgen.1001244

21. Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science 2008; 322:750-6; PMID:18974356; http://dx.doi.org/10.1126/science.1163045

22. Pandey RR, Mondal T, Mohammad F, Enroth S, Redrup L, Komorowski J, et al. Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol Cell 2008; 32:232-46; PMID:18951091; http://dx.doi.org/10.1016/j.molcel.2008.08.022

23. Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, et al. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc Natl Acad Sci U S A 2009; 106:11667-72; PMID:19571010; http://dx.doi.org/10.1073/pnas.0904715106

24. Zhao J, Ohsumi TK, Kung JT, Ogawa Y, Grau DJ, Sarma K, et al. Genome-wide identification of polycomb-associated RNAs by RIP-seq. Mol Cell 2010; 40:939-53; PMID:21172659; http://dx.doi.org/10.1016/j.molcel.2010.12.011

25. Kanhere A, Viiri K, Araújo CC, Rasaiyaah J, Bouwman RD, Whyte WA, et al. Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2. Mol Cell 2010; 38:675-88; PMID:20542000; http://dx.doi.org/10.1016/j.molcel.2010.03.019

26. Kaneko S, Li G, Son J, Xu CF, Margueron R, Neubert TA, et al. Phosphorylation of the PRC2 component Ezh2 is cell cycle-regulated and up-regulates its binding to ncRNA. Genes Dev 2010; 24:2615-20; PMID:21123648; http://dx.doi.org/10.1101/gad.1983810

27. Jeon Y, Lee JT. YY1 tethers Xist RNA to the inactive X nucleation center. Cell 2011; 146:119-33; PMID:21729784; http://dx.doi.org/10.1016/j.cell.2011.06.026

28. Dinger ME, Amaral PP, Mercer TR, Pang KC, Bruce SJ, Gardiner BB, et al. Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. Genome Res 2008; 18:1433-45; PMID:18562676; http://dx.doi.org/10.1101/gr.078378.108

29. Ørom UA, Derrien T, Beringer M, Gumireddy K, Gardini A, Bussotti G, et al. Long noncoding RNAs with enhancer-like function in human cells. Cell 2010; 143:46-58; PMID:20887892; http://dx.doi.org/10.1016/j.cell.2010.09.001

30. Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, et al. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature 2011; 472:120-4; PMID:21423168; http://dx.doi.org/10.1038/nature09819

31. Pollard SM, Benchoua A, Lowell S. Neural stem cells, neurons, and glia. Methods Enzymol 2006; 418:151-69; PMID:17141035; http://dx.doi.org/10.1016/S0076-6879(06)18010-6

32. Ying QL, Smith AG. Defined conditions for neural commitment and differentiation. Methods Enzymol 2003; 365:327-41; PMID:14696356; http://dx.doi.org/10.1016/S0076-6879(03)65023-8

33. Aubert J, Stavridis MP, Tweedie S, O'Reilly M, Vierlinger K, Li M, et al. Screening for mammalian neural genes via fluorescence-activated cell sorter purification of neural precursors from Sox1-gfp knock-in mice. Proc Natl Acad Sci U S A 2003; 100(Suppl 1):11836-41; PMID:12923295; http://dx.doi.org/10.1073/pnas.1734197100

34. Tucker KL, Meyer M, Barde YA. Neurotrophins are required for nerve growth during development. Nat Neurosci 2001; 4:29-37; PMID:11135642; http://dx.doi.org/10.1038/82868

35. Boyer LA, Plath K, Zeitlinger J, Brambrink T, Medeiros LA, Lee TI, et al. Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature 2006; 441:349-53; PMID:16625203; http://dx.doi.org/10.1038/nature04733

36. Lee TI, Jenner RG, Boyer LA, Guenther MG, Levine SS, Kumar RM, et al. Control of developmental regulators by Polycomb in human embryonic stem cells. Cell 2006; 125:301-13; PMID:16630818; http://dx.doi.org/10.1016/j.cell.2006.02.043

37. Shilatifard A. Molecular implementation and physiological roles for histone H3 lysine 4 (H3K4) methylation. Curr Opin Cell Biol 2008; 20:341-8; PMID:18508253; http://dx.doi.org/10.1016/j.ceb.2008.03.019

38. Barrera LO, Li Z, Smith AD, Arden KC, Cavenee WK, Zhang MQ, et al. Genome-wide mapping and analysis of active promoters in mouse embryonic stem cells and adult organs. Genome Res 2008; 18:46-59; PMID:18042645; http://dx.doi.org/10.1101/gr.6654808

39. Sanchez-Elsner T, Gou D, Kremmer E, Sauer F. Noncoding RNAs of trithorax response elements recruit Drosophila Ash1 to Ultrabithorax. Science 2006; 311:1118-23; PMID:16497925; http://dx.doi.org/10.1126/science.1117705

40. Shen X, Kim W, Fujiwara Y, Simon MD, Liu Y, Mysliwiec MR, et al. Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells. Cell 2009; 139:1303-14; PMID:20064376; http://dx.doi.org/10.1016/j.cell.2009.12.003

41. Martens JHA, O'Sullivan RJ, Braunschweig U, Opravil S, Radolf M, Steinlein P, et al. The profile of repeat-associated histone lysine methylation states in the mouse epigenome. EMBO J 2005; 24:800-12; PMID:15678104; http://dx.doi.org/10.1038/sj.emboj.7600545