PTE, a novel module of mammalian Polycomb Response Elements

Sarina R. Cameron[1], Soumyadeep Nandi[1,2,¤], Tatyana G. Kahn[1,¤], Juan I. Barrasa[1,¤], Per Stenberg[1,2,3], Yuri B. Schwartz[1,#]

[1] Department of Molecular Biology, Umeå University, Umeå, Sweden

[2] Computational Life Science Cluster (CLiC), Umeå University, Umeå, Sweden

[3] Division of CBRN Security and Defence, FOI–Swedish Defence Research Agency, Umeå, Sweden

[#]correspondence should be addressed to: yuri.schwartz@umu.se

[¤]joint second authors

Running title: PRC1 targeting element

Key words: Polycomb/ Polycomb targeting/ PRC1/ epigenetics/ Cyclin D2

Abstract

Polycomb Group proteins are essential epigenetic repressors. They form multiple protein complexes of which two kinds, PRC1 and PRC2, are indispensable for repression. Although much is known about their biochemical properties, how PRC1 and PRC2 are targeted to specific genes is poorly understood. Here we establish the *Cyclin D2* (*CCND2*) oncogene as a simple model to address this question. We provide the evidence that coordinated recruitment of PRC1 and PRC2 complexes to *CCND2* involves a combination of a specialized PRC1 targeting element (PTE) and an adjacent CpG-island, which together act as a human Polycomb Response Element. Chromatin immunoprecipitation analysis of *CCND2* in different transcriptional states indicates that histone modifications produced by PRC1 and PRC2 are not sufficient to recruit either of the complexes. However, catalytic activity of PRC2 helps to anchor PRC1 at the PTE. Our analyses suggest that coordinated targeting of PRC1 and PRC2 complexes by juxtaposed AT-rich PTEs and CpG-islands is a general feature of Polycomb repression and advance our understanding of how the targeting has evolved.

Introduction

Polycomb Group (PcG) proteins comprise a family of epigenetic repressors that prevent unscheduled transcription of hundreds of developmental genes (Schuettengruber et al. 2007; Schwartz and Pirrotta 2007; Kohler and Hennig 2010; Geisler and Paro 2015). PcG proteins act in concert as multisubunit complexes. These are usually grouped into two evolutionary conserved classes called Polycomb Repressive Complex 1 (PRC1) and PRC2. Several auxiliary complexes, whose repertoire varies between species, further aid PRC1 and PRC2 to achieve robust repression (Schwartz and Pirrotta 2013).

PRC2 complexes contain a core with five subunits: EZH2 (or related protein EZH1), EED, SUZ12, RBBP4 (or closely related RBBP7) and AEBP2 (Ciferri et al. 2012; Kloet et al. 2016), which together act as a histone methyltransferase specific to Lysine 27 of histone H3 (Cao et al. 2002; Czermin et al. 2002; Kuzmichev et al. 2002; Muller et al. 2002). Tri-methylation of H3K27 (H3K27me3) at target genes is necessary for PcG repression (Pengelly et al. 2013). In addition to extensive tri-methylation of H3K27 at target genes, PRC2 complexes are also responsible for di-methylation of H3K27 within the entire transcriptionally inactive genome (Ferrari et al. 2013; Lee et al. 2015). This di-methylation happens via untargeted “hit-and-run” action and suppresses pervasive transcription (Lee et al. 2015).

The systematics of PRC1 group is more complicated as RING2 (or closely related RING1 protein), the central subunit of PRC1, are incorporated in large variety of protein complexes (Gao et al., 2012; Schwartz and Pirrotta, 2013). Here, for simplicity, we will use the PRC1 name for complexes that consist of one of the five variant chromodomain proteins (CBX2, CBX4, CBX6, CBX7 or CBX8), one of the three Polyhomeotic-like proteins (PHC1, PHC2 and PHC3), SCMH1 (or related SCML2) protein and a heterodimer between RING2 (or RING1) and one of the two closely related PCGF proteins: MEL18 (also known as PCGF2)

and BMI1 (also known as PCGF4). These complexes, sometimes called canonical PRC1, can recognize H3K27me3 produced by PRC2 via the chromodomain of their CBX subunits (Fischle et al. 2003; Bernstein et al. 2006; Kaustov et al. 2011). Mutations in genes encoding PRC1 subunits lead to embryonic lethality and miss-expression of *HOX* genes, indicating that PRC1 complexes are essential for PcG repression (van der Lugt et al. 1994; Bel et al. 1998; Akasaka et al. 2001; Isono et al. 2005).

Other RING2/RING1 complexes contain RYBP (or closely related YAF2 protein) and one of the four PCGF proteins (PCGF1, PCGF3, PCGF5 and PCGF6). Complexes with different PCGF proteins have distinct additional subunits (Gao et al. 2012; Schwartz and Pirrotta 2013). None of these RING-PCGF complexes include CBX and PHC proteins and, therefore, cannot recognize the H3K27me3 mark. While PRC1 complexes are integral for PcG repression, the role of other RING-PCGF complexes, sometimes called non-canonical PRC1, is less clear. One of the complexes, which incorporates the RING2-PCGF1 dimer and the characteristic KDM2B subunit, is involved in PcG repression in mouse embryonic stem cells (Farcas et al. 2012; Wu et al. 2013; Blackledge et al. 2014). However, some of the other RING-PCGF complexes have functions unrelated to PcG mechanisms (Gao et al. 2014).

PRC1 and other RING-PCGF complexes can act as E3 ligases *in vitro* to transfer a single ubiquitin group to Lysine 119 of histone H2A (H2AK119) (Gao et al. 2012; Tavares et al. 2012; Taherbhoy et al. 2015). How the *in vivo* ubiquitylase activity of PRC1 differs from that of RYBP-containing RING-PCGF complexes is a subject of debate and there is disagreement in whether PRC1 is less catalytically active than its RYBP-containing counterparts (Gao et al. 2012; Blackledge et al. 2014; Taherbhoy et al. 2015). PRC2 is able to recognize H2AK119ub via its core subunit AEBP2 (Kalb et al. 2014). However, the importance of H2AK119ub for PcG repression is not entirely clear. Thus mice, homozygous for the *Ring2* allele that encodes a catalytically inactive protein, show no missexpression of *Hox* genes and die much later than

mice lacking PRC1 and PRC2 complexes (Illingworth et al. 2015). Likewise, *Drosophila*, which have H2AK118 ubiquitylation (analog of mammalian H2AK119ub) abolished by replacement of the corresponding H2A lysine to arginine, express *Hox* genes in proper manner and survive until late larval stages (Pengelly et al. 2015).

In *Drosophila*, where PcG system was first discovered and is the most studied, the targeting of PRC1 and PRC2 depends on designated Polycomb Response Elements (PREs). These approximately 1kb-long DNA elements are necessary and sufficient to recruit both complexes and tend to repress reporter genes when integrated elsewhere in the genome (Kassis and Brown 2013). The ability of PREs to repress is not absolute. If a target gene gets activated at early embryonic stages, the de-repressed state is often inherited by the progenitor cells (Poux et al. 2002) and the binding of PcG complexes and H3K27 trimethylation is impaired (Schwartz et al. 2006; Petruk et al. 2008; Schwartz et al. 2010; Bowman et al. 2014).

PRC1 can specifically recognize the H3K27me3 modification catalyzed by PRC2 and PRC2 can recognize H2AK119ub produced by PRC1. However, neither histone modification is necessary or sufficient for targeting of PRC complexes to PREs (Lee et al. 2015; Kahn et al. 2016). Instead, multiple lines of evidence indicate that PREs contain combinations of recognition sequences for different DNA binding factors. These factors act cooperatively to recruit PRC1 and PRC2, which themselves cannot bind DNA in sequence specific fashion (reviewed in: (Kassis and Brown 2013)). Our recent work indicates that PRC1 can bind PREs in the absence of PRC2. However, at many PREs, PRC2 requires PRC1 for binding (Kahn et al. 2016). Consistently, in cells where PcG target genes are active they often lose H3K27me3 and PRC2 but retain significant amount of PRC1 bound at high-affinity PRE sites (Schwartz et al. 2006; Schwartz et al. 2010; Bowman et al. 2014).

PRC1 and PRC2 complexes are evolutionary conserved and target many of the same genes in *Drosophila* and mammals (Schwartz and Pirrotta 2007). It is, therefore, hard to imagine that mechanisms to bring PRC complexes to these genes have evolved completely independently. The question, therefore, is which parts of the targeting process are similar and which have diverged since flies and mammals split from their common ancestor several hundred million years ago. A number of mouse and human DNA fragments were reported to resemble PREs in their ability to recruit either PRC1 or PRC2 complexes when integrated elsewhere in the genome (Sing et al. 2009; Mendenhall et al. 2010; Woo et al. 2010; Cabianca et al. 2012; Lynch et al. 2012; Arnold et al. 2013; Schorderet et al. 2013; Jermann et al. 2014). Yet, the overall picture remains obscure. First, in contrast to paradigmatic *Drosophila* PREs, few of the reported mammalian PRE-like elements were mapped to high precision or their recruitment properties confirmed by independent investigations. Second, some of these elements were said to recruit both PRC1 and PRC2 while others appear to recruit only one of the complexes. Lastly, little is known about the mechanisms implicated in targeting.

Several independent investigations concur that DNA sequences with high density of unmethylated CpG di-nucleotides (so-called CpG islands) are sufficient to recruit PRC2 but not PRC1 in mouse embryonic stem cells (Mendenhall et al. 2010; Lynch et al. 2012; Jermann et al. 2014). Presence of binding sites for transcription factors REST and SNAIL within a CpG island may further help PRC2 recruitment (Dietrich et al. 2012; Arnold et al. 2013). Studies from Klose and Brockdorff groups put forward a model linking CpG-islands to PRC2. According to this model, mouse Ring2-Pcgf1-Kdm2B complex recognizes unmethylated CpG-islands via the CXXC domain of the Kdm2B subunit. Once recruited to a CpG island the complex mono-ubiquitylates surrounding nucleosomes at H2AK119 which, in turn, serve as a binding platform for PRC2 (Blackledge et al. 2014; Cooper et al. 2014). Clearly different from how PRC2 is recruited in *Drosophila* (Lee et al. 2015; Kahn et al.

2016), the model is supported by observations that tethering a minimal catalytically active H2AK119 ubiquitylase is sufficient for PRC2 recruitment (Blackledge et al. 2014). Yet, the central role of H2AK119ub implied by the model, is at odds with reports that H2AK119ub is dispensable for early mouse development (Illingworth et al. 2015). Furthermore, which DNA elements, if any, target PRC1 complexes and whether H3K27 methylation is involved in the process remains an open question.

Here we establish the *Cyclin D2* (*CCND2*) oncogene as a simple model system to address these questions. Using this system we find that coordinated recruitment of PRC1 and PRC2 complexes to *CCND2* involves a combination of a specialized PRC1 targeting element (PTE) and an adjacent CpG-island, which we view as a functional analogue of *Drosophila* PREs. Chromatin immunoprecipitation analysis of *CCND2* in different transcriptional states suggest that H2AK119ub and H3K27me3 do not serve as binding platforms for either complex. Yet, catalytic activity of PRC2 helps PRC1 to bind the PTE. Further analyses suggest that coordinated targeting of PRC1 and PRC2 complexes by juxtaposed PTEs and CpG-islands is not unique to the *CCND2* gene and that PRC1 targeting by compact DNA elements is an ancestral feature of Polycomb repression.

Results

PRC1 and PRC2 complexes usually act together to effect epigenetic repression. However, in experiments with cultured *Drosophila* cells we noted PcG target genes that retained PRC1 at PREs in the absence of PRC2 and H3K27me3 when they happened to be transcriptionally active (Schwartz et al. 2006; Schwartz et al. 2010). This fueled our interest to a region approximately 4.8kb upstream of the Transcription Start Site (TSS) of the human *Cyclin D2* (*CCND2*) gene which in the embryonic teratocarcinoma NT2-D1 cells is strongly immunoprecipitated with antibodies against BMI1 and MEL18 but very weakly with antibodies against EZH2 and H3K27me3 (Figure 1A-C). This is in stark contrast to all other sites on human chromosomes 8, 11 and 12 which show strong precipitation with anti-EZH2 and anti-H3K27me3 antibodies whenever they are strongly precipitated with antibodies against BMI1 or MEL18 (Kahn et al. 2014).

Quantitative RT-PCR measurements show that, like in the *Drosophila* cases, the *CCND2* gene of NT2-D1 cells is highly transcribed (Figure 1D). In contrast, in TIG-3 embryonic fibroblasts the *CCND2* gene is transcriptionally inactive (Figure 1D). In these cells, it is decorated with H3K27me3, binds CBX8 and SUZ12 and gets upregulated upon the knockdown of EZH2, EED, SUZ12 and BMI1 (Bracken et al. 2006). This indicates that *CCND2* is an ordinary PcG target gene which, when transcriptionally inactive, acquires the chromatin state characteristic of PcG repression but binds a lot of PRC1 and little PRC2 and H3K27me3 in cells where it is transcriptionally active. This finding is significant in three ways. First, it shows that PRC1 can bind target genes when very little H3K27me3 is present. Second, it suggests that, like in *Drosophila*, the binding of PRC1 in the vicinity of a gene is not by itself repressive. Third, it points to a location of a DNA element that may be responsible for PRC1 recruitment to *CCND2* (putative PRC1 Targeting Element). It also

suggests that *CCND2* is as a good model to investigate how the targeting of human Polycomb complexes is coordinated.

PRC1 and PRC2 binding to the *CCND2* gene in alternative chromatin states

With this in mind we performed quantitative ChIP analysis of PRC1, PRC2 and H3K27me3 binding at the *CCND2* gene in the NT2-D1 and TIG-3 cell lines and in parallel assayed expression of genes encoding PRC1 subunits to know which of the alternative subunits are available in each of the cell lines (Figure S1). *MEL18* is highly expressed in NT2-D1 cells and slightly weaker in TIG-3 cells (Figure S1). *BMI1* is expressed moderately and at the same level in both NT2-D1 and TIG-3 cells. *RING1* expression is low in both cell lines but *RING2* is strongly expressed in NT2-D1 cells and shows medium expression in TIG-3 cells (Figure S1). Out of five *CBX* genes implicated in PcG regulation (Schwartz and Pirrotta 2013), *CBX4*, *CBX6* and *CBX7* are weakly expressed in NT2-D1 and TIG-3 cells. *CBX8* is expressed exceedingly weakly in both cell lines and *CBX2* is expressed strongly in NT2-D1 but very weakly in TIG-3 cells.

Immunoprecipitations of the formaldehyde crosslinked NT2-D1 chromatin with antibodies against MEL18, BMI1, CBX2 and RING2 give essentially the same results (Figures 2, S2). The ChIP signals peak within a putative PRC1 Targeting Element (PTE) and recede steeply at both sides to reach a background level half way between the PTE and the *CCND2* TSS. In TIG-3 cells, where *CCND2* is transcriptionally inactive, the putative PTE remains the strongest precipitated site with ChIP signals similar to those detected in NT2-D1 cells. However, in these cells, in addition to the PTE the entire upstream region of the *CCND2* gene, including the TSS, is also immunoprecipitated, albeit at a much lower level. In contrast to PRC1, ChIPs with antibodies against SUZ12 and H3K27me3 give enrichment profiles that

differ dramatically between the two cell lines. In NT2-D1 cells, SUZ12 ChIP signals are very weak barely raising above the background at the PTE (Figure 2). These are paralleled by weak ChIP signals for H3K27me3. In contrast, in TIG-3 cells, ChIP signals for both antibodies are very strong. Strikingly, their profiles do not match those for PRC1. Instead, the SUZ12 and H3K27me3 profiles are broad and shifted away from the CpG-poor PTE into the adjacent CpG-rich region (Figures 2, 1B). The obvious difference between the PRC1 and SUZ12/H3K27me3 ChIP profiles is instructive. First, it indicates that a high level of H3K27me3 is not sufficient to recruit PRC1 to an extent seen at the putative PTE and that other mechanisms must contribute to PRC1 recruitment to this element. Second, it suggests that the putative PTE is not capable to recruit PRC2 as efficiently as activities linked to the adjacent CpG-rich region.

KDM2B, H2AK119 ubiquitylation and recruitment of PRC2

Mouse Ring2-Pcgf1-Kdm2B complex can bind unmethylated CpG islands and catalyze the monoubiquitylation of H2AK119 (Farcas et al. 2012; He et al. 2013; Wu et al. 2013), which was proposed to directly recruit PRC2 (Blackledge et al. 2014). According to data from the ENCODE project, in NT2-D1 cells the CpG-region upstream of the *CCND2* TSS is unmethylated (Consortium et al. 2012). We therefore asked to what extent the apparent strong binding of PRC2 over the CpG-rich region in TIG-3 cells is linked to the presence of KDM2B-RING2 complexes and H2AK119ub.

KDM2B is not crosslinked to chromatin with formaldehyde (Farcas et al. 2012), so we used a combination of formaldehyde and EGS (ethylene glycol bis(succinimidyl succinate)) to prepare chromatin for KDM2B mapping. Somewhat surprisingly, the KDM2B ChIP signals are higher in NT2-D1 cells, probably because in these cells the *KDM2B* gene is expressed

much stronger (Figure S1), but are qualitatively similar to those in TIG-3 cells (Figure 2). In contrast to PRC1, the KDM2B ChIP profile is shifted from the putative PTE into adjacent CpG-rich DNA. In NT2-D1 cells most of this region (amplicons 4 and 5) is not precipitated with the anti-RING2 antibodies regardless of whether we perform RING2 ChIP from chromatin crosslinked with formaldehyde and EGS or formaldehyde alone (Figure S3). It appears that when *CCND2* is transcriptionally active KDM2B binds the CpG-island independently of RING2. Consistently in NT2-D1 cells the H2AK119ub ChIP signal peaks at the putative PTE and recedes to the background level between the PTE and the *CCND2* TSS (Figure 2). In TIG-3 cells, the region precipitated with anti-H2AK119ub antibodies is broad and coextensive with that precipitated with anti-SUZ12 and anti-H3K27me3 antibodies. The strength of the H2AK119ub ChIP signal in TIG-3 cells is comparable to that over the putative PTE in NT2-D1 cells. This suggests that either H2A119ub is not by itself sufficient to recruit much PRC2 and additional processes must be in place for efficient recruitment, or that something prevents H2A119ub-mediated PRC2 recruitment when *CCND2* is transcriptionally active.

CCND2 PTE can autonomously recruit PRC1 and is evolutionary conserved

To investigate whether the DNA fragment underneath the *CCND2* PRC1 binding peak is indeed a PRC1 targeting element, we tested whether it can autonomously recruit PcG complexes when integrated elsewhere in the genome. To this effect we cloned the 2.4kb fragment covering the putative PTE into a lentiviral vector (Figure 3A) and integrated it back into the genome of NT2-D1 cells by viral transduction. To distinguish transgenic and endogenous copies of the putative PTE we identified a small stretch of nucleotides close to the summit of the MEL18/BMI1 peak that shows little conservation within mammalian species and substituted five of those nucleotides in the transgenic copy to create an annealing

site for a specific PCR primer (Figure S4). An analogous construct containing a 2.4kb fragment from a gene desert region on chromosome 12 that showed no binding of PcG proteins and H3K27me3 (Figure S4) and an empty vector were integrated in parallel as negative controls. The transduced cell lines were genotyped by PCR to validate their identity (Figure 3A, Figure S5).

As summarized in Figure 3B, the 2.4kb transgenic PTE fragment is immunoprecipitated with antibodies against BMI1 and MEL18. The precipitation is robust and as strong as that of the endogenous PTE upstream of *CCND2*. This is in contrast to the transgenic insertion of the negative control fragment from the gene desert (Figure 3C) or the empty vector (Figure S6) whose precipitation is very low and close to the ChIP background. The immunoprecipitation of the transgenic 2.4kb fragment with the antibodies against SUZ12 and H3K27me3 is weak and comparable to that of the endogenous site (Figure 3B). Taken together these observations suggest that the 2.4kb DNA fragment underneath the *CCND2* PRC1 binding peak indeed contains an element capable of autonomous recruitment of PRC1 complexes.

The *CCND2* PTE is evolutionary conserved suggesting that its function is important. Thus, mouse *Ccnd2* resides within large 32Mbp block of shared synteny between mouse chromosome 6 and human chromosome 12. Aside of the open reading frame, which has high DNA sequence conservation, the 10kb sequence upstream of the *Ccnd2* TSS contains multiple blocks predicted as evolutionary conserved DNA elements (Figure 4A). One of these blocks corresponds to ~300bp sequence that is directly underneath the PRC1 binding peak within the human *CCND2* PTE.

In mouse F9 testicular teratoma cells, where the *Ccnd2* gene is transcriptionally inactive, its upstream region is immunoprecipitated with antibodies against Cbx7 and Suz12 (Figure 4B). This suggests that the mouse *Ccnd2* is a PcG target gene. Similar to what is seen at inactive human *CCND2* in TIG-3 cells, the ChIP signal for Cbx7 is highest at the site corresponding to the evolutionary conserved block within the PTE but the highest Suz12 ChIP signal is shifted from the Cbx7 peak towards the TSS into the CpG-rich area. To extend this comparison further, we cloned the 3.6kb fragment covering the putative mouse PTE into the same lentiviral vector used to test the human PTE (Figure 4A) and integrated it into the genome of the human NT2-D1 cells by lentiviral transduction. ChIP-qPCR analysis indicates that the evolutionary conserved central part of the mouse 3.6kb fragment is robustly precipitated with antibodies against MEL18, CBX2 and RING2 and weakly with antibodies against SUZ12 (Figure 4C). This indicates that the mouse *Ccnd2* gene is also equipped with a PTE.

The *CCND2* PTE is modular and its activity depends on specific DNA sequences

Drosophila PREs, the closest analogues of the *CCND2* PTE, contain recognition sequences for multiple unrelated DNA binding proteins that cooperate to provide robust PcG repression. A typical PRE has a size of about 1kb. Often such PREs may be subdivided into fragments of a few hundred base pairs that can still recruit PcG proteins in transgenic assays but the recruitment is less robust (Hagstrom et al. 1997; Horard et al. 2000; Hodgson et al. 2001). We, therefore, wondered whether the *CCND2* PTE is smaller than the 2.4kb fragment tested in our initial transgenic assay and whether the PTE contains a single core recruiting element or multiple weaker elements that cooperate. To address these questions we examined PRC1 recruitment by sub-fragments of the 2.4kb *CCND2* PTE (Figure 5A). ChIP analysis of PRC1 binding in cells transduced with corresponding lentiviral constructs indicates that the 1kb

fragment (PTE 1.2) centered on the summit of the MEL18 and BMI1 binding peak and the larger overlapping fragments (PTE 1.1 and PTE 1.3) recruit MEL18, BMI1and CBX2 as efficiently as the full-length 2.4kb fragment or the endogenous *CCND2* PTE (Figure 5B-D). In contrast precipitation of the PTE 1.2 fragment with anti-SUZ12 antibodies is weak (Figure 5B) reinforcing our conclusion that the *CCND2* PTE is not an efficient PRC2 recruiter.

Further dissection indicates that smaller sub-fragments of PTE 1.2 cannot recruit PRC1 as efficiently as the full-length fragment (Figures 6, S7). When integrated elsewhere in the genome, the left (PTE 1-6) or the central (PTE 6-8) parts are immunoprecipitated with anti-MEL18 antibodies very weakly (Figure 6B-C) and their precipitation with anti-BMI1 antibodies is at the edge of detection (Figure S7). This is likely, because in NT2-D1 cells BMI1 is less abundant than MEL18 (Figure S1). The immunoprecipitation of larger fragments that combine the left and the central parts (PTE 1-8) or the central and the right parts (PTE 6-11) is more efficient (Figures 6D-E, S7). Consistent with the asymmetric shape of the MEL18/BMI1 ChIP-chip peak (Figure 1B), the PTE 6-11 fragment shows the strongest immunoprecipitation from all PTE 1.2 sub-fragments tested. This suggests that the central and the right parts of the 1kb PTE have greater contribution to PRC1 recruitment. Importantly, the synthetic fragment (PTE1.2Δ) that includes both the left and the right parts of PTE 1.2 but lacks the central part is still immunoprecipitated (Figure 6F, S7). Overall these results indicate that the *CCND2* PTE consists of at least three separable modules, all of which contribute to the recruitment with the central and the right modules being more important.

Both sequence specific DNA binding activity and *cis*-acting non-coding RNAs (ncRNAs) have been implicated in the recruitment of mammalian PcG proteins (Brockdorff 2013). Recent genome annotation indicates two long non-coding RNAs CCND2-AS1 and CCND2-

AS2, which originate within the second exon or the first intron of *CCND2* (Figure 1B). They are transcribed in the opposite direction to *CCND2* and traverse over the PTE to terminate some 27kb away from their TSS (Yates et al. 2016). These lncRNAs are unlikely to play a critical role in autonomous recruitment by *CCND2* PTE as their transcription start sites were not included in our lentiviral constructs. We therefore sought evidence of a specific DNA binding activity targeting the PTE 1.2 fragment. The analysis of the PTE 1.2 DNA sequence revealed no potential binding sites for YY1, RUNX1/CBFβ, REST and SNAIL (Figure S8A), the sequence specific DNA binding proteins previously implicated in the recruitment of mammalian PcG proteins (Woo et al. 2010; Dietrich et al. 2012; Yu et al. 2012; Arnold et al. 2013). Consistently, we and others detect no YY1 binding to the *CCND2* PTE (Consortium et al. 2012; Kahn et al. 2014). Mining the ENCODE data shows that the PTE 1.2 fragment does not bind any of the transcription factors mapped to date (Figure S8A) although multiple transcription factors bind the adjacent CpG-rich region.

Since we did not find any binding sites for known transcription factors, we analyzed the *CCND2* PTE for sequences that might bind proteins not yet implicated in PcG targeting. We hypothesized that some of the other PcG target genes, represented by high-confidence BMI1/MEL18 binding sites on the three human chromosomes surveyed by ChIP (Kahn et al. 2014), may have elements similar to the *CCND2* PTE. In this case, they may use some of the same DNA binding proteins for PcG recruitment and may be distinguished by the same sequence motifs. To test this conjecture, we applied multivariate modelling (Stenberg et al. 2005; Philip et al. 2012) and searched for all possible 6 nucleotide long sequence words predictive of high-confidence MEL18/BMI1 binding sites compared to 1000 random sequences not bound by MEL18 or BMI1 and matched for the distance to the closest TSS. This approach yielded two motifs that we dubbed “CGA” and “CGCG” (Figure 6G, S8B).

While the “CGCG”-motif is likely an outcome of the close proximity between PTEs and CpG-islands, the “CGA”-motif is interesting. First, it is present right at the summit of the *CCND2* MEL18/BMI1 peak and its position within the *CCND2* PTE is highly evolutionarily conserved (Figure S8C). Second, the “CGA”-motif is significantly enriched in high-confidence MEL18/BMI1 binding sites (Figure 6H). Overall, these observations suggest that the “CGA”-motif is a common feature of MEL18/BMI1 bound sites and may be involved in targeting PRC1 to *CCND2* PTE.

To test the role of the “CGA”-motif for recruitment by the *CCND2* PTE we generated a lentiviral construct that contained the 1kb PTE 1.2 fragment in which the CG dinucleotide within the “CGA”-motif was replaced with AA. ChIP-qPCR analysis of transduced NT2-D1 cells shows that antibodies against PRC1 components precipitate the mutated PTE1.2 fragment (PTE1.2mCGA) four-fold weaker than the wild-type variant (Figure 6I, S9). The reduction of ChIP signals is comparable to that seen after deleting the entire central part of the PTE (PTE1.2Δ construct). This suggests that the “CGA”-motif makes an important contribution to the recruitment by the evolutionary conserved central part of the PTE. The loss of ChIP signals upon mutation of the “CGA”-motif or deletion of the central part of the PTE is significant but not complete. This indicates that other sequences within the 1kb PTE contribute to recruitment. Consistently, Electrophoretic Mobility-Shift Assays (EMSA) with nuclear protein extracts from NT2-D1 cells and a set of partially overlapping 100-150bp fragments indicate that three different fragments (N2, N5 and N7) within the 1kb PTE 1.2 element bind nuclear proteins in sequence specific fashion (Figure S10). Overall it appears that the *CCND2* PTE is modular and depends on multiple sequence specific DNA binding proteins for PRC1 recruitment.

PRC2 activity promotes PRC1 binding to the *CCND2* PTE

The PTE binds little PRC2 and H3K27me3 when *CCND2* is active. Considering that in TIG-3 cells, high levels of H3K27me3 and PRC2 are not sufficient to recruit much PRC1 to the region between the PTE and the *CCND2* TSS, we expected that at the low levels seen in NT2-D1 cells PRC2 and H3K27 methylation are dispensable for recruitment of PRC1 to the PTE. To our surprise, the CRISPR/Cas9-mediated knock-down of SUZ12 in NT2-D1 cells (Figure 7A, C) leads to a strong reduction of the PTE immunoprecipitation by the antibodies against PRC1 subunits (Figure 7B). Western-blot analysis argues that reduced immunoprecipitation of the *CCND2* PTE is not due to lower PRC1 abundance in the SUZ12 depleted cells (Figure 7A). In fact, NT2-D1 cells seem to require at least a low level of PRC2 for proliferation as after multiple attempts we failed to recover any cell lines completely deficient for SUZ12. In the clonal isolate used here (S1H6), the level of SUZ12 dropped to a few percent of that in the original NT2-D1 cells and the overall H3K27me3 is reduced about 10-fold (Figure 7A, C). Under these conditions SUZ12 and H3K27me3 are no longer detected at the *CCND2* PTE. At the same time, a very low SUZ12 ChIP signal is still detectable at the *ALX4* gene (used as a positive control example of a PcG-repressed gene in NT2-D1 cells). With this level of SUZ12 binding, *ALX4* retains significant immunoprecipitation with anti-H3K27me3 antibodies and ChIP signals for CBX2 and MEL18 are nearly unchanged (Figure 7B). Taken together, these results suggest that the physical presence of PRC2 or its enzymatic activity in some way contribute to robust recruitment of PRC1 to the *CCND2* PTE.

To distinguish between the two possibilities, we inhibited PRC2 methyltransferase activity with a small molecule inhibitor UNC1999 shown to be highly specific for EZH2 and EZH1 and competitive with the cofactor SAM but not with the H3 substrate (Konze et al. 2013).

Notably, UNC1999 does not lead to degradation of PRC2 and does not prevent its binding to chromatin (Xu et al. 2015). Consistent with observations in other cultured cell lines (Konze et al. 2013; Xu et al. 2015), twelve day treatment of NT2-D1 cells with recommended (2μM) concentration of UNC1999 leads to partial inhibition of the PRC2 activity without reducing overall levels of PRC2 and PRC1 (Figure 7D). Under these conditions, ChIP signals for H3K27me3 at the *CCND2* PTE drop to a background level but reduce less than two-fold at the control *ALX4* gene. At both sites the SUZ12 ChIP signals remain unchanged (Figure 7E). Importantly, following the UNC1999 treatment the MEL18 and CBX2 ChIP signals decrease 4-fold at the *CCND2* PTE but remain unchanged at *ALX4* (Figure 7E). To exclude the possibility that UNC1999 treatment causes indiscriminate loss of proteins bound next to active genes, we tested the binding of transcription factor CTCF upstream of the *CLBP*, *VDAC2*, *STK17A* genes. According to published expression array profile of the NT2-D1 cells (Fong et al. 2011), these genes are expressed within the same range (2 to 16 percent of *GAPDH*) as *CCND2*. As illustrated by Figure 7F, the CTCF ChIP signals are not affected by UNC1999 treatment. Taken together, our experiments suggest that enzymatic activity of PRC2 but not its physical presence are critical for robust binding of PRC1 to the PTE.

PRC1 targeting elements may be common

The upstream region of the *CCND2* gene is organized in a particular way. It consists of the A/T-rich PTE, which is adjacent to the CpG-island that "links" the PTE to the *CCND2* TSS. How common is such "PTE – CpG-island – TSS" arrangement? The question is not trivial to address. First, we cannot unambiguously pinpoint other putative PTEs from genome-wide binding profiles since at transcriptionally inactive target genes PRC1 binding is rather broad. Second, it is possible that some of the PcG target genes use alternative ways to recruit PRC1. Taking this into account we reasoned that, if PTEs are sufficiently common and, like in the

case of *CCND2*, correspond to sites of the strongest PRC1 binding, we should be able to detect the trace of the "PTE – CpG-island – TSS" arrangement by plotting aggregate distributions of di-nucleotide combinations around the highest BMI1/MEL18 binding peaks within PcG repressed genes.

Indeed, plotted such that all DNA sequences around the highest BMI1/MEL18 peaks on the human chromosomes 8, 11 and 12 are oriented to have the BMI1/MEL18 binding peak at the centre and the closest TSS to the right of the peak, the di-nucleotide distributions show three striking features (Figure 8A). First, the DNA sequences underneath the BMI1/MEL18 peaks tend to be AA-rich but CG-poor. Second, the DNA around BMI1/MEL18 peaks is CG-rich with clear bias towards TSSs. Third, the DNA encompassing BMI1/MEL18 peaks is depleted of TG/CA nucleotides. This is in contrast to average di-nucleotide distributions in upstream regions of human genes (Figure S11). These control distributions show no AA-rich hot-spots and matching CG-poor gap. They also display no TG/CA depletion. Regions upstream of TSS appear generally CG-rich, however their CG enrichment is lower than that downstream of the BMI1/MEL18 peaks. The picture of di-nucleotide distributions around BMI1/MEL18 peaks parallels the arrangement seen at the *CCND2* locus (compare Figure 8A to Figure 1C). This suggests that many of the DNA sequences underneath MEL18/BMI1 peaks contain AA-rich elements similar to the *CCND2* PTE and that the arrangement "AA-rich MEL18/BMI1 peak – CpG-island – TSS" is common for many PcG target genes.

PRC2 binding was linked to unmethylated CpG islands (Mendenhall et al. 2010; Lynch et al. 2012; Jermann et al. 2014; Riising et al. 2014) and in TIG-3 cells, where the *CCND2* gene is repressed, the ChIP signal for SUZ12 is offset from the PTE to the adjacent CpG-rich region. Strikingly, the average SUZ12 distribution around MEL18/BMI1 peaks in NT2-D1 cells

(Consortium et al. 2012) parallels that of the unmethylated CpG di-nucleotides (Figure 8B). We see that the SUZ12 binding is generally offset to the sides of the MEL18/BMI1 peaks (Figure 8C). This observation is not unique for NT2-D1 cells. Comparing the ChIP-seq profiles of CBX7 and RING1 in primary fibroblasts (Pemberton et al. 2014) to the ChIP-seq profile of SUZ12 in human embryonic stem cells (Consortium et al. 2012) we see essentially the same picture (Figure 8D). Overall, these results suggest that a significant fraction of human PcG target genes recruit PRC complexes via a combination of AA-rich PTEs and CpG-islands.

To further evaluate the hypothesis that *CCND2* is not the only gene equipped with a PTE, we searched for more examples. As demonstrated by the *CCND2* case, a PTE can be pinpointed as localized PRC1 binding site when its target gene is transcriptionally active and has little PRC2 and H3K27me3. We expect such cases to be rare. First, a large fraction of Polycomb target genes encodes tissue specific transcription factors (Boyer et al. 2006; Bracken et al. 2006; Lee et al. 2006; Schwartz et al. 2006) and, therefore, most of these genes will be inactive in any given cell type. Second, there is likely more than one way to target PRC1 and, therefore, not all Polycomb target genes are equipped with PTEs. To screen for additional PTE-equipped genes, we mapped MEL18 and SUZ12 in NT2-D1 cells genome-wide with ChIP-seq. We identified MEL18-bound regions that had an annotated TSS within 20kb distance and displayed MEL18 ChIP-seq signal comparable to that at *CCND2*. These regions were further ranked by MEL18/SUZ12 ChIP-seq signal ratios. Consistent with our original ChIP-chip mapping, *CCND2* locus was ranked second and had the highest rank from all loci on chromosomes 8, 11 and 12 that were tiled by our Affymetrix array (Table S1). To be conservative, we focused on the site that scored better than *CCND2* (i.e. was ranked first and

had higher MEL18 ChIP-seq signal). It corresponded to a narrow MEL18 bound region approximately 10kb downstream of the *ZIC2* gene (Figure S12A). Like the *CCND2* PTE, this putative PTE region is CpG-poor but surrounded by CpG-islands. RT-qPCR analysis confirmed our expectation that *ZIC2*, the likely target of PcG regulation, is highly expressed in NT2-D1 cells. Fortunately, *ZIC2* is inactive in TIG-3 cells providing essential comparison (Figure S12B). The putative *ZIC2* PTE is traversed by a long non-coding RNA *LINC00554* of unknown function, which is not detectable in either cell line (Figure S12B).

Detailed ChIP-qPCR analysis shows that, just like *CCND2*, the *ZIC2* locus binds PRC1 at the putative PTE in both NT2-D1 and TIG-3 cells regardless of transcriptional status of the *ZIC2* gene (Figure S12C). However, the antibodies against SUZ12 and H3K27me3 efficiently immunoprecipitate the locus only from the chromatin of TIG-3 cells, where *ZIC2* is transcriptionally inactive (Figure S12C). In line with what is seen at the *CCND2* gene (Figure 2B) and at an average PcG target locus (Figures 8C-D), the SUZ12 and H3K27me3 ChIP signals in TIG-3 cells are offset from the putative PTE into neighboring CpG-rich regions. Taken together, the *ZIC2* example and the bioinformatic analyses above strongly suggest that *CCND2* is not the only human gene equipped with a PRC1 targeting element.

Discussion

Epigenetic repression by Polycomb Group mechanisms is critical for development of all multicellular animals and is frequently disrupted in cancers (Schwartz and Pirrotta 2007; Scelfo et al. 2015; Comet et al. 2016). Yet, how it is targeted to specific genes is not well understood. More is known about the process in *Drosophila*. There we know that the targeting is mediated by compact DNA elements (PREs), that PREs are sufficient to recruit all Polycomb complexes and that the recruitment requires combinatorial action of sequence

specific DNA binding adaptor proteins (Muller and Kassis 2006; Schwartz and Pirrotta 2008; Kassis and Brown 2013). We also know that PREs can bind PRC1 independently of PRC2 and that, at many PREs, PRC1 has to be present for robust PRC2 binding (Kahn et al. 2016). Targeting PcG repression in mammalian cells is likely somewhat different. First, mammalian genomes encode greater variety of RING complexes that contribute to PcG repression. Second, as a consequence of CpG methylation, mammalian genomes are AT-rich and their genes and corresponding regulatory elements are often embedded within CpG-rich islands protected from DNA methylation. This feature, absent from *Drosophila*, seems to help targeting of mammalian PRC2 complexes (Mendenhall et al. 2010; Lynch et al. 2012; Jermann et al. 2014; Riising et al. 2014). On the other hand, many of the same developmental genes are regulated by Polycomb mechanisms in flies and mammals. This suggests that some of the common target genes have been regulated by PcG complexes in the last common ancestor of vertebrates and insects and that at least part of the original targeting mechanism is still in place in both animal branches.

Here we used the peculiar binding of PRC1 to the transcriptionally active *CCND2* gene as an entry point to look for similarities and differences in the targeting of PcG complexes in *Drosophila* and humans. We show that, similar to flies, PRC1 complexes are autonomously targeted to *CCND2*, and possibly many other human PcG target genes, by a compact DNA element (PTE). Like fly PREs, the *CCND2* PTE consists of several partially redundant modules recognized by distinct sequence specific DNA binding activities. However, the *CCND2* PTE also differs from *Drosophila* PREs. PREs can robustly recruit both PRC1 and PRC2 and therefore are self-sufficient to install PcG repression. In contrast, the *CCND2* PTE seems to recruit PRC1 very well but PRC2 only poorly. Comparison of PRC1 and PRC2 binding to *CCND2* in alternative transcriptional states suggests that robust targeting of both complexes involves multiple DNA elements: A/T-rich PTE, responsible for PRC1

recruitment, and adjacent CpG-island that acts as focus of PRC2 targeting. We propose that the combination of these two elements corresponds to a functional analogue of a *Drosophila* PRE (Figure 8E). In this view, the targeting of PRC1 by a compact DNA element is a common ancestral feature while the targeting of PRC2 has diverged. Flies, which lost DNA methylation, must have evolved different means to target PRC2 in a way that is more hardwired to the targeting of PRC1 (Kang et al. 2015; Kahn et al. 2016).

Consistent with this view, unmethylated CpG-islands are recognized by the mammalian RING2-PCGF1-KDM2B complex via the CXXC domain of KDM2B (Farcas et al. 2012; He et al. 2013; Wu et al. 2013). This complex mono-ubiquitylates H2AK119 and ubiquitylation is directly recognized by PRC2 and may help to recruit it (Blackledge et al. 2014; Cooper et al. 2014; Kalb et al. 2014). In contrast, the fly genome encodes no PCGF1 protein and the corresponding KDM2B complex does not exist. Furthermore, the fly H2AK118ub is not sufficient and, at least at some genes, not necessary for stable PRC2 binding or Polycomb repression (Pengelly et al. 2013; Lee et al. 2015; Kahn et al. 2016).

In TIG-3 cells immunoprecipitation of the *CCND2* CpG-island with H2AK119ub antibodies correlates with strong ChIP signals for PRC2 and H3K27me3. However, comparable H2AK119ub signal over the PTE in NT2-D1 cells does not. This suggests that H2AK119ub is not sufficient to recruit much PRC2 or that it is prevented from doing so by a process linked to the *CCND2* expression. Interestingly, KDM2B seems to bind the *CCND2* CpG-island regardless of the transcriptional status of the gene. However, it is accompanied by RING2 only when *CCND2* is repressed. It is tempting to speculate that, when *CCND2* is transcriptionally inactive, the CpG-anchored KDM2B “receives” its RING2 from PRC1 anchored at the nearby PTE to form RING2-PCGF1-KDM2B ubiquitylase.

In NT2-D1 cells the endogenous and transgenic *CCND2* PTE binds little PRC2 and displays little H3K27 tri-methylation. Paradoxically, the binding of PRC1 to the PTE is strongly reduced when PRC2 enzymatic activity is impaired. This dependence is hard to explain by postulating that H3K27me3 serves as a platform for PRC1 binding and therefore the binding is abolished when PRC2 activity and H3K27me3 are lost. First, the PRC1 ChIP signals at the PTE are very similar between NT2-D1 and TIG-3 cells although the extent of immunoprecipitation with anti-H3K27me3 antibodies is very different. More importantly, in TIG-3 cells a much higher level of H3K27me3 over the CpG-island does not lead to PRC1 binding comparable to that at the PTE suggesting that H3K27me3 is not sufficient for robust recruitment. Conceivably, H3K27me3 serves a "catalytic" function by binding chromodomain of a CBX subunit and introducing allosteric changes to the PRC1 complexes that promote interaction with DNA binding adapter proteins. Alternatively, the effect may be indirect and stem not from the PRC2 complex bound at the PTE but from the loss of untargeted PRC2 activity. This activity was shown to globally suppress chromatin accessibility and could be, in some way, beneficial for PRC1 binding (Ferrari et al. 2013; Lee et al. 2015). The puzzling PRC1 dependence on PRC2 for binding the *CCND2* PTE is not without a precedent. A DNA element upstream of the mouse *MafB* gene, which resembles the *CCND2* PTE in its ability to recruit PRC1 but not PRC2, was also reported to inexplicably lose PRC1 binding after the SUZ12 knock-down by RNAi (Sing et al. 2009).

Serendipitous discovery of the PTE turned the *CCND2* locus into a fruitful model that gave us mechanistic insights in principles of PcG protein targeting and shed light on the parallels between the processes in *Drosophila* and mammals. We have little doubt that *CCND2* has much more to offer. Of obvious interest is the nature of the DNA binding proteins implicated at the *CCND2* PTE as those are likely to be involved in the recruitment of PRC1 complexes to other human genes. Other questions that could be addressed using *CCND2* as a model

include the roles of H2AK119 ubiquitylation and the RING2-PCGF1-KDM2B complexes in targeting of PRC2 and the identity of the processes that prevent PRC2 binding when a target gene is active. It will also be interesting to explore whether the *CCND2* PTE, or PTEs in general, have preferences for PRC1 variants with different CBX subunits. For example, CBX2 has the chromodomain coupled to an AT-hook and in early mouse embryos the two domains cooperate to target CBX2 to paternal AT-rich satellite DNA (Tardat et al. 2015). Conceivably, the same feature may help the CBX2-containing PRC1 complexes to bind the AT-rich PTEs. Finally, *CCND2* is an oncogene (Burns et al. 2003; Mermelshtein et al. 2005; Cole et al. 2010) and multiple lines of evidence link PcG mechanisms to cancer progression (Scelfo et al. 2015; Koppens and van Lohuizen 2016). It is still puzzling why some tumour types depend on the overexpression of PcG proteins while others require the loss of PcG function. The oncogenic effect of the overexpression has to some extent been explained by erroneous silencing of the *INK4A/ARF* locus (Sparmann and van Lohuizen 2006) but the link between malignant transformation and the loss of PcG proteins remains elusive. We speculate that a failure to repress *CCND2* may, at least in part, explain this link. From this we predict that the *CCND2* PTE may be disrupted in some cancers that overexpress *CCND2* but otherwise have the PcG function intact.

Materials and Methods

Transgenic constructs

Lentiviral constructs were produced by *in vitro* recombination of fragments of interest into the Eco47III site of the pLenti-CMVTRE3G-eGFP-ICR-Puro or pLenti-ICR-Puro vectors. *In vitro* recombination was done using In-Fusion HD system (Clontech). All fragments were

amplified using high-fidelity Pfu DNA polymerase (Thermo Scientific) and human or mouse genomic DNA or, in case of *CCND2* PTE sub-fragments, DNA of the PTE 2.4kb construct as a template. PCR primers and their sequences are indicated in Tables S2, S3.

pLenti-CMVTRE3G-eGFP-ICR-Puro was generated based on the pLenti-CMVTRE3G-eGFP-Puro backbone (Campeau et al. 2009). As first step, pLenti-CMVTRE3G-eGFP-Puro was cut with HpaI and EcoRI and recombined with two DNA fragments produced by PCR with following pairs of primers: 5'-TCGACGGTATCGGTTAACTT-3'; 5'-AGCGCTAGTCTCGTGATCGATAAA-3' and 5'-CACGAGACTAGCGCTGAGAGTTGG-3'; 5'-CTACCCGGTAGAATTCCACGTGGGGAG-3' and DNA of pLenti-CMVTRE3G-eGFP-Puro as a template. This step introduced unique PmlI site between *eGFP* and *Pac* (puromycin resistance gene) and unique Eco47III site upstream of the *eGFP* gene. As second step, the resulting construct was digested with PmlI and recombined with the mouse *Igf2/H19* ICR insulator sequence PCR amplified from mouse genomic DNA using the forward 5'-CCGGTAGAATTCCACTGTCACAGCGGACCCCAACCTATG-3' and the reverse: 5'-GGGCCGCCTCCCCACTCGTGGACTCGGACTCCCAAATCA-3' primers. The native *Igf2/H19* ICR sequence contained one Eco47III site. As final step, this site was removed by cutting the above construct with Eco47III and recombining it with the DNA fragment produced by PCR with the following primer pair: 5'-GATCACGAGACTAGCGCTGAGAG-3' and reverse 5'-TTTTCACACAATGGCGCTGATGGCC-3', using DNA of the above construct as a template.

We originally planned to use the pLenti-CMVTRE3G-eGFP-ICR-Puro based constructs and integrate them into the NT2-D1 cells that had been modified to express TetR protein from constitutive CMV promoter. In theory this should have allowed us to induce eGFP by adding the doxycycline to the media. Unfortunately, we have soon discovered that the CMV promoter is not active in NT2-D1 cells. Therefore to simplify and reduce the size of the

transgenes we have removed the TRE3G-eGFP part from pLenti-CMVTRE3G-eGFP-ICR-Puro to yield the pLenti-ICR-Puro vector (Figure S6A). This was done by digesting pLenti-CMVTRE3G-eGFP-ICR-Puro with EcoRV and Eco47III and recombining it with short dsDNA fragment produced by the annealing of the 5'-GATCACGAGACTAGCGCTGAGAGTTGGCTTCACGTGCTAGACCCAGCTTTC-3' and 5'-GAAAGCTGGGTCTAGCACGTGAAGCCAACTCTCAGCGCTAGTCTCGTGATC-3' oligonucleotides. Side-by-side comparison of the PRC1 binding to transgenic PTE 1.2 in either vector showed that the presence of the CMVTRE3G-eGFP part had no effect on the recruitment of PRC1. The analyses on Figure 3B-C and Figure S6B were done with transgenic constructs based on pLenti-CMVTRE3G-eGFP-ICR-Puro. All other analyses used transgenes based on pLenti-ICR-Puro.

Cell culture and lentiviral transduction

NTERA-2 (NT2-D1 ATCC® CRL-1973™), TIG-3 (gift from Dr. K. Helin, BRIC), 293T (ATCC® CRL-3216™) and F9 (ATCC® CRL-1720™) cells were cultured in high-glucose DMEM (Gibco) supplemented with 10% Fetal Bovine Serum (Sigma), Penicillin/Streptomycin (Gibco) and 1.5g/L Sodium Bicarbonate (Sigma) at 37^0C in atmosphere of 5% CO_2. The F9 cells were plated in T-flasks pre-treated with 0.1% gelatin in H_2O for 1 hour prior to plating.

To produce viral particles, $5x10^6$ 293T cells were plated in a $75cm^2$ flask 24 hours prior to transfection. The packaging plasmids pCMV-dR8.2dvpr (6μg) and pCMV-VSV-G (3μg) (gift from of M. Roth, UMDNJ) were combined with a transfer construct (9μg) and co-transfected using X-tremeGene HP (Roche) at 1:3 ratio of DNA to transfection reagent. After 24 hours of incubation, the medium was changed. Lentiviral supernatant was collected after another 24 hours, filtered (0.45 μm filters) and used for infection directly or stored at -80^0C.

For lentiviral infection, cells were plated at a confluence of 40 to 60% 24 hours in advance. Viral supernatant was added in serial dilution to cells in combination with 8µg/ml Polybrene (Millipore). After overnight incubation, the medium was changed to remove Polybrene. Transduced cells were selected for 14 days by growth on culture medium supplemented with 4µg/ml puromycin (Invitrogen).

To generate the SUZ12 knock-out NT2 cell line the SUZ12g1.1 (caccgGGTGGCGGCGGCGACGGCTT) and SUZ12g1.2 (aaacAAGCCGTCGCCGCCGCCACCc) DNA oligonucleotides were annealed and cloned into lentiCRISPRv2 plasmid (Addgene #52961) linearized by digestion with *BsmB*I. The construct was introduced in NT2-D1 cells by lentiviral infection and transduced cells selected by growth on the culture medium supplemented with 4µg/ml puromycin (Invitrogen). Cells were cloned and knock-down assayed by Western blot and immunofluorescence with the antibodies listed in Table S4.

To inhibit EZH2/EZH1, the NT2-D1 cells were grown in complete DMEM supplemented with 2µM of UNC1999 (Cayman Chemical Company #14621) or 0.2% DMSO as a negative control. The media were replaced every second day for 12 days and cells crosslinked for ChIP analyses. Fractions of the same cell cultures were taken before crosslinking and their total nuclear protein analyzed by western blot.

ChIP and RT-qPCR analyses

ChIP reactions were performed as described in (Schwartz et al. 2006; Kahn et al. 2014). For some experiments the procedure was modified to include additional crosslinking with EGS (ethylene glycol bis(succinimidyl succinate)). Cells were scraped, resuspended in 10ml of 1xPBS and EGS solution in DMSO added to final concentration of 2mM. The crosslinking reaction was allowed to proceed for 1 hour at room temperature after which formaldehyde

was added to final concentration 1% and crosslinking continued for another 15 minutes at room temperature. The reaction was stopped by adding an excess of glycine. The antibodies used for ChIP are listed in the Table S4. Total RNA from cultured cells was isolated using TRI Reagent (Sigma) according to manufactures instructions. cDNA was prepared with the First Strand cDNA synthesis kit (Thermo Scientific) using 2µg of RNA and random hexamer primers and purified as described (Schwartz et al. 2010). qPCR analysis of cDNA and ChIP products was performed essentially as described in (Kahn et al. 2006; Schwartz et al. 2010) except that iQ5 Real-Time PCR Detection system (Bio-Rad) and KAPPA SYBR FAST qPCR Kit (Kappa Biosystems) were used for all analyses. The primers used for qPCR analyses are described in Tables S5, S6, S7. Sequencing of ChIP products was done as described in (Kahn et al. 2016) except that NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB #E7645S) was used for sequencing library preparation.

Electrophoretic Mobility-Shift Assay

To prepare the nuclear extract, the NT2-D1 cells from four 182.5cm^2 T-flasks were disrupted by resuspension in 400µl of hypotonic Cell Lysis Buffer (5% Sucrose, 5mM Tris-HCl pH8, 5mM NaCl, 1.5mM $MgCl_2$, 0.1% Triton X-100, 1mM PMSF, 2mM dithiolthreitol, 1x Protease Inhibitor Cocktail). After discarding the cytosolic fraction, nuclei were extracted for 1 hour at 4°C with 110µl of Nuclear Extract Buffer (20mM HEPES, 20% Glycerol, 500mM NaCl, 1.5mM $MgCl_2$, 0.2mM EDTA, 0.5mM dithiolthreitol, 0.5mM PMSF, 1x Protease Inhibitor Cocktail). The nuclear extract was cleared by centrifugation for 20 minutes at 16000g, 4°C and used immediately or stored at -80 ^{o}C.

dsDNA fragments were labeled with αP^{32} dATP (Perkin Elmer) using PCR and the DNA of the PTE 2.4kb construct as a template and purified by passing through Illustra MicroSpin

S300 HR Columns (GE Healthcare). The corresponding PCR primers are indicated in the Table S8.

Binding reactions (final volume of 20µl) were assembled by combining 5µl of 4x binding buffer (80mM HEPES pH7.4-7.9, 20mM $MgCl_2$, 20% Glycerol, 400mM NaCl, 4mM dithiolthreitol, 4mM EDTA), 5µg to 10µg of nuclear protein, 1µg to 5µg of Salmon Sperm DNA (Life Technologies), 40fmol of labeled probe and 150 fold excess of specific or non-specific competitor. The binding was allowed to proceed for 20 minutes at room temperature after which samples were run on 8% acrylamide gel (29:1 acrylamide to bis-acrylamide solution) on 1xTBE for 4 to 6 hours at 160V. The resulted gels were dried for 2 hours and exposed to x-ray film (AGFA Healthcare).

Computational analyses

Definition of bound regions

The MEL18 and BMI1 bound regions were defined as coordinates of clusters of microarray features that satisfied the following three criteria. i) smoothed ChIP/Input hybridization intensity ratios of the features were above the 99.8 percentile cut-off; ii) the maximum distance to the neighboring feature above the intensity cut-off was equal or greater than 200bp; and iii) the length of the cluster was equal or greater than 200bp. The peak center of a bound region was set at the center of the 5 consecutive microarray features with the highest hybridization intensity values. Regions of 1kb, centered on overlapping binding peaks of MEL18 and BMI1 were considered as high-confidence MEL18/BMI1 binding sites. To avoid overlapping regions in the analysis for figures 1C, 8A and 8B we increased the ii-criterion above to 20kb. CpG-islands were defined using default parameters in the EMBOSS package (Rice et al. 2000).

Motif analysis

The 317 most predictive words from the 1kb MEL18/BMI1 high-confidence regions derived by multivariate modelling as described in (Stenberg et al. 2005; Philip et al. 2012) were pair-wise aligned and each alignment was assigned a score reflecting the maximum number of identical nucleotides in the alignment. Based on these scores we generated a hierarchical tree (Euclidian distance and complete linkage) using Cluster 3.0 (de Hoon et al. 2004). The tree was divided into eight groups and the words from each group were realigned using Muscle (Edgar 2004). The aligned words were used to build the position weight matrix (PWM) following the method proposed by Staden (Touzet and Varre 2007). As final step, the resulted PWMs were optimized with the Bound/Surveyed Sequence Discrimination Algorithm (Liu et al. 2012). The prediction of potential binding sites for the YY1, RUNX1/CBFβ, REST and SNAIL proteins was done as described in (Nandi et al. 2013).

ChIP-sequencing data analysis

Sequencing reads were aligned to the Hg19 human reference genome with bowtie2 (Langmead and Salzberg 2012) using default parameters. Resulted SAM files were filtered using samtools (Li et al. 2009) to remove reads with q-scores lower than 10. The reads were extended to 180bp (the average insert size in sequencing libraries) and read density profiles generated using Pyicos (Althammer et al. 2011). To compare MEL18 and SUZ12 binding patterns, the read density profiles were filtered to remove all data points below 99th percentile and bound regions defined as clusters of five or more genomic positions separated by less than 501bp that had read densities greater than 0. For each detected region, the “binding” value was set to an average of the five consecutive highest read density values and the position of the corresponding binding peak to their middle coordinate. To calculate

MEL18/SUZ12 binding ratios, the binding value of each MEL18 peak was divided by that of the "strongest" SUZ12 binding peak in the 10kb neighborhood.

External data

The ChIP-chip profiles of MEL18, BMI1, EZH2 and H3K27me3 on Chromosomes 8, 11 and 12 of NT2-D1 cells were from GSE41854 (Kahn et al. 2014). The NT2-D1 DNA methylation profiles were from GSM683929 and GSM683804. The SUZ12 ChIP-seq profiles in NT2-D1 and human embryonic stem cells were from GSM935423 and GSM1003573, respectively. The CBX7 and RING1 ChIP-seq profiles in primary human fibroblasts were from GSE40740.

Data access

The ChIP-seq data are available from GEO at the accession number GSE101538

Acknowledgements

We are grateful to Dr. Kristian Helin for the gift of TIG-3 cells and Dr. Monica Roth for the gift of lentiviral packaging plasmids. We thank Dr. Haruhiko Koseki and Dr. Robert Klose for generous gifts of anti-RING2 and anti-KDM2B antibodies. We are grateful to Dr. Jana Šmigová for evaluation of anti-CTCF antibodies. This work was supported in part by grants from Åke Wiberg and Magnus Bergvall foundations to PS, grants from Cancerfonden, Umeå Universitet Insamlingsstiftelsen and Strategic Research grant from Medical Faculty, Umeå University to YBS and grants from Knut and Alice Wallenberg Foundation and Kempestiftelserna to EpiCoN (YBS and PS co-PIs).

Author contributions

S.C. generated all transgenic cell lines and performed ChIP analyses of human transgenic lines and EMSA. S.C. and J.I.B. performed ChIP analysis of mouse *Ccnd2* PTE. S.C. and T.G.K. analyzed PcG protein binding to *CCND2* locus in NT2-D1 and TIG-3 cells. T.G.K. performed PRC2 knock-down experiments, genome wide mapping of MEL18 and SUZ12 and characterized *ZIC2* PTE. J.I.B. helped with characterization of the PRC2 knock-down cell lines. S.N. performed most of the computational analyses. P.S. supervised S.N. and performed computational analyses. S.C., S.N., T.G.K., J.I.B., P.S. and Y.B.S. interpreted the data. Y.B.S. conceived and supervised the project and wrote the manuscript with input from all co-authors.

Competing interests

The authors declare that no competing interests exist

References

Akasaka T, van Lohuizen M, van der Lugt N, Mizutani-Koseki Y, Kanno M, Taniguchi M, Vidal M, Alkema M, Berns A, Koseki H. 2001. Mice doubly deficient for the Polycomb Group genes Mel18 and Bmi1 reveal synergy and requirement for maintenance but not initiation of Hox gene expression. *Development* **128**: 1587-1597.

Althammer S, Gonzalez-Vallinas J, Ballare C, Beato M, Eyras E. 2011. Pyicos: a versatile toolkit for the analysis of high-throughput sequencing data. *Bioinformatics* **27**: 3333-3340.

Arnold P, Scholer A, Pachkov M, Balwierz PJ, Jorgensen H, Stadler MB, van Nimwegen E, Schubeler D. 2013. Modeling of epigenome dynamics identifies transcription factors that mediate Polycomb targeting. *Genome research* **23**: 60-73.

Bel S, Core N, Djabali M, Kieboom K, Van der Lugt N, Alkema MJ, Van Lohuizen M. 1998. Genetic interactions and dosage effects of Polycomb group genes in mice. *Development* **125**: 3543-3551.

Bernstein E, Duncan EM, Masui O, Gil J, Heard E, Allis CD. 2006. Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. *Molecular and cellular biology* **26**: 2560-2569.

Blackledge NP, Farcas AM, Kondo T, King HW, McGouran JF, Hanssen LL, Ito S, Cooper S, Kondo K, Koseki Y et al. 2014. Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. *Cell* **157**: 1445-1459.

Bowman SK, Deaton AM, Domingues H, Wang PI, Sadreyev RI, Kingston RE, Bender W. 2014. H3K27 modifications define segmental regulatory domains in the Drosophila bithorax complex. *eLife* **3**: e02833.

Boyer LA, Plath K, Zeitlinger J, Brambrink T, Medeiros LA, Lee TI, Levine SS, Wernig M, Tajonar A, Ray MK et al. 2006. Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**: 349-353.

Bracken AP, Dietrich N, Pasini D, Hansen KH, Helin K. 2006. Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes & development* **20**: 1123-1136.

Brockdorff N. 2013. Noncoding RNA and Polycomb recruitment. *Rna* **19**: 429-442.

Burns KH, Agno JE, Sicinski P, Matzuk MM. 2003. Cyclin D2 and p27 are tissue-specific regulators of tumorigenesis in inhibin alpha knockout mice. *Mol Endocrinol* **17**: 2053-2069.

Cabianca DS, Casa V, Bodega B, Xynos A, Ginelli E, Tanaka Y, Gabellini D. 2012. A long ncRNA links copy number variation to a polycomb/trithorax epigenetic switch in FSHD muscular dystrophy. *Cell* **149**: 819-831.

Campeau E, Ruhl VE, Rodier F, Smith CL, Rahmberg BL, Fuss JO, Campisi J, Yaswen P, Cooper PK, Kaufman PD. 2009. A versatile viral system for expression and depletion of proteins in mammalian cells. *PloS one* **4**: e6529.

Cao R, Wang L, Wang H, Xia L, Erdjument-Bromage H, Tempst P, Jones RS, Zhang Y. 2002. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**: 1039-1043.

Ciferri C, Lander GC, Maiolica A, Herzog F, Aebersold R, Nogales E. 2012. Molecular architecture of human polycomb repressive complex 2. *eLife* **1**: e00005.

Cole AM, Myant K, Reed KR, Ridgway RA, Athineos D, Van den Brink GR, Muncan V, Clevers H, Clarke AR, Sicinski P et al. 2010. Cyclin D2-cyclin-dependent kinase 4/6

is required for efficient proliferation and tumorigenesis following Apc loss. *Cancer Res* **70**: 8149-8158.

Comet I, Riising EM, Leblanc B, Helin K. 2016. Maintaining cell identity: PRC2-mediated regulation of transcription and cancer. *Nat Rev Cancer* **16**: 803-810.

Consortium EP, Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**: 57-74.

Cooper S, Dienstbier M, Hassan R, Schermelleh L, Sharif J, Blackledge NP, De Marco V, Elderkin S, Koseki H, Klose R et al. 2014. Targeting polycomb to pericentric heterochromatin in embryonic stem cells reveals a role for H2AK119u1 in PRC2 recruitment. *Cell Rep* **7**: 1456-1470.

Czermin B, Melfi R, McCabe D, Seitz V, Imhof A, Pirrotta V. 2002. Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell* **111**: 185-196.

de Hoon MJ, Imoto S, Nolan J, Miyano S. 2004. Open source clustering software. *Bioinformatics* **20**: 1453-1454.

Dietrich N, Lerdrup M, Landt E, Agrawal-Singh S, Bak M, Tommerup N, Rappsilber J, Sodersten E, Hansen K. 2012. REST-mediated recruitment of polycomb repressor complexes in mammalian cells. *PLoS genetics* **8**: e1002494.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **32**: 1792-1797.

Farcas AM, Blackledge NP, Sudbery I, Long HK, McGouran JF, Rose NR, Lee S, Sims D, Cerase A, Sheahan TW et al. 2012. KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. *eLife* **1**: e00205.

Ferrari KJ, Scelfo A, Jammula S, Cuomo A, Barozzi I, Stutzer A, Fischle W, Bonaldi T, Pasini D. 2013. Polycomb-Dependent H3K27me1 and H3K27me2 Regulate Active Transcription and Enhancer Fidelity. *Molecular cell*.

Fischle W, Wang Y, Jacobs SA, Kim Y, Allis CD, Khorasanizadeh S. 2003. Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes & development* **17**: 1870-1881.

Fong CY, Chak LL, Biswas A, Tan JH, Gauthaman K, Chan WK, Bongso A. 2011. Human Wharton's jelly stem cells have unique transcriptome profiles compared to human embryonic stem cells and other mesenchymal stem cells. *Stem Cell Rev* **7**: 1-16.

Gao Z, Lee P, Stafford JM, von Schimmelmann M, Schaefer A, Reinberg D. 2014. An AUTS2-Polycomb complex activates gene expression in the CNS. *Nature* **516**: 349-354.

Gao Z, Zhang J, Bonasio R, Strino F, Sawai A, Parisi F, Kluger Y, Reinberg D. 2012. PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. *Molecular cell* **45**: 344-356.

Geisler SJ, Paro R. 2015. Trithorax and Polycomb group-dependent regulation: a tale of opposing activities. *Development* **142**: 2876-2887.

Hagstrom K, Muller M, Schedl P. 1997. A Polycomb and GAGA dependent silencer adjoins the Fab-7 boundary in the Drosophila bithorax complex. *Genetics* **146**: 1365-1380.

He J, Shen L, Wan M, Taranova O, Wu H, Zhang Y. 2013. Kdm2b maintains murine embryonic stem cell status by recruiting PRC1 complex to CpG islands of developmental genes. *Nature cell biology* **15**: 373-384.

Hodgson JW, Argiropoulos B, Brock HW. 2001. Site-specific recognition of a 70-base-pair element containing d(GA)(n) repeats mediates bithoraxoid polycomb group response element-dependent silencing. *Molecular and cellular biology* **21**: 4528-4543.

Horard B, Tatout C, Poux S, Pirrotta V. 2000. Structure of a polycomb response element and in vitro binding of polycomb group complexes containing GAGA factor. *Molecular and cellular biology* **20**: 3187-3197.

Illingworth RS, Moffat M, Mann AR, Read D, Hunter CJ, Pradeepa MM, Adams IR, Bickmore WA. 2015. The E3 ubiquitin ligase activity of RING1B is not essential for early mouse development. *Genes & development* **29**: 1897-1902.

Isono K, Fujimura Y, Shinga J, Yamaki M, J OW, Takihara Y, Murahashi Y, Takada Y, Mizutani-Koseki Y, Koseki H. 2005. Mammalian polyhomeotic homologues Phc2 and Phc1 act in synergy to mediate polycomb repression of Hox genes. *Molecular and cellular biology* **25**: 6694-6706.

Jermann P, Hoerner L, Burger L, Schubeler D. 2014. Short sequences can efficiently recruit histone H3 lysine 27 trimethylation in the absence of enhancer activity and DNA methylation. *Proceedings of the National Academy of Sciences of the United States of America* **111**: E3415-3421.

Kahn TG, Dorafshan E, Schultheis D, Zare A, Stenberg P, Reim I, Pirrotta V, Schwartz YB. 2016. Interdependence of PRC1 and PRC2 for recruitment to Polycomb Response Elements. *Nucleic acids research* **44**: 10132-10149.

Kahn TG, Schwartz YB, Dellino GI, Pirrotta V. 2006. Polycomb complexes and the propagation of the methylation mark at the Drosophila ubx gene. *The Journal of biological chemistry* **281**: 29064-29075.

Kahn TG, Stenberg P, Pirrotta V, Schwartz YB. 2014. Combinatorial interactions are required for the efficient recruitment of pho repressive complex (PhoRC) to polycomb response elements. *PLoS genetics* **10**: e1004495.

Kalb R, Latwiel S, Baymaz HI, Jansen PW, Muller CW, Vermeulen M, Muller J. 2014. Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. *Nature structural & molecular biology* **21**: 569-571.

Kang H, McElroy KA, Jung YL, Alekseyenko AA, Zee BM, Park PJ, Kuroda MI. 2015. Sex comb on midleg (Scm) is a functional link between PcG-repressive complexes in Drosophila. *Genes & development* **29**: 1136-1150.

Kassis JA, Brown JL. 2013. Polycomb group response elements in Drosophila and vertebrates. *Advances in genetics* **81**: 83-118.

Kaustov L, Ouyang H, Amaya M, Lemak A, Nady N, Duan S, Wasney GA, Li Z, Vedadi M, Schapira M et al. 2011. Recognition and specificity determinants of the human cbx chromodomains. *The Journal of biological chemistry* **286**: 521-529.

Kloet SL, Makowski MM, Baymaz HI, van Voorthuijsen L, Karemaker ID, Santanach A, Jansen PW, Di Croce L, Vermeulen M. 2016. The dynamic interactome and genomic targets of Polycomb complexes during stem-cell differentiation. *Nature structural & molecular biology* **23**: 682-690.

Kohler C, Hennig L. 2010. Regulation of cell identity by plant Polycomb and trithorax group proteins. *Current opinion in genetics & development* **20**: 541-547.

Konze KD, Ma A, Li F, Barsyte-Lovejoy D, Parton T, Macnevin CJ, Liu F, Gao C, Huang XP, Kuznetsova E et al. 2013. An orally bioavailable chemical probe of the Lysine Methyltransferases EZH2 and EZH1. *ACS Chem Biol* **8**: 1324-1334.

Koppens M, van Lohuizen M. 2016. Context-dependent actions of Polycomb repressors in cancer. *Oncogene* **35**: 1341-1352.

Kuzmichev A, Nishioka K, Erdjument-Bromage H, Tempst P, Reinberg D. 2002. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes & development* **16**: 2893-2905.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**: 357-359.

Lee HG, Kahn TG, Simcox A, Schwartz YB, Pirrotta V. 2015. Genome-wide activities of Polycomb complexes control pervasive transcription. *Genome research* **25**: 1170-1181.

Lee TI, Jenner RG, Boyer LA, Guenther MG, Levine SS, Kumar RM, Chevalier B, Johnstone SE, Cole MF, Isono K et al. 2006. Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125**: 301-313.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078-2079.

Liu Y, Nandi S, Martel A, Antoun A, Ioshikhes I, Blais A. 2012. Discovery, optimization and validation of an optimal DNA-binding sequence for the Six1 homeodomain transcription factor. *Nucleic acids research* **40**: 8227-8239.

Lynch MD, Smith AJ, De Gobbi M, Flenley M, Hughes JR, Vernimmen D, Ayyub H, Sharpe JA, Sloane-Stanley JA, Sutherland L et al. 2012. An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment. *The EMBO journal* **31**: 317-329.

Mendenhall EM, Koche RP, Truong T, Zhou VW, Issac B, Chi AS, Ku M, Bernstein BE. 2010. GC-rich sequence elements recruit PRC2 in mammalian ES cells. *PLoS genetics* **6**: e1001244.

Mermelshtein A, Gerson A, Walfisch S, Delgado B, Shechter-Maor G, Delgado J, Fich A, Gheber L. 2005. Expression of D-type cyclins in colon cancer and in cell lines from colon carcinomas. *Br J Cancer* **93**: 338-345.

Muller J, Hart CM, Francis NJ, Vargas ML, Sengupta A, Wild B, Miller EL, O'Connor MB, Kingston RE, Simon JA. 2002. Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. *Cell* **111**: 197-208.

Muller J, Kassis JA. 2006. Polycomb response elements and targeting of Polycomb group proteins in Drosophila. *Current opinion in genetics & development* **16**: 476-484.

Nandi S, Blais A, Ioshikhes I. 2013. Identification of cis-regulatory modules in promoters of human genes exploiting mutual positioning of transcription factors. *Nucleic acids research* **41**: 8822-8841.

Pemberton H, Anderton E, Patel H, Brookes S, Chandler H, Palermo R, Stock J, Rodriguez-Niedenfuhr M, Racek T, de Breed L et al. 2014. Genome-wide co-localization of Polycomb orthologs and their effects on gene expression in human fibroblasts. *Genome biology* **15**: R23.

Pengelly AR, Copur O, Jackle H, Herzig A, Muller J. 2013. A histone mutant reproduces the phenotype caused by loss of histone-modifying factor Polycomb. *Science* **339**: 698-699.

Pengelly AR, Kalb R, Finkl K, Muller J. 2015. Transcriptional repression by PRC1 in the absence of H2A monoubiquitylation. *Genes & development* **29**: 1487-1492.

Petruk S, Smith ST, Sedkov Y, Mazo A. 2008. Association of trxG and PcG proteins with the bxd maintenance element depends on transcriptional activity. *Development* **135**: 2383-2390.

Philip P, Pettersson F, Stenberg P. 2012. Sequence signatures involved in targeting the Male-Specific Lethal complex to X-chromosomal genes in Drosophila melanogaster. *BMC genomics* **13**: 97.

Poux S, Horard B, Sigrist CJ, Pirrotta V. 2002. The Drosophila trithorax protein is a coactivator required to prevent re-establishment of polycomb silencing. *Development* **129**: 2483-2493.

Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends in genetics : TIG* **16**: 276-277.

Riising EM, Comet I, Leblanc B, Wu X, Johansen JV, Helin K. 2014. Gene silencing triggers polycomb repressive complex 2 recruitment to CpG islands genome wide. *Molecular cell* **55**: 347-360.

Scelfo A, Piunti A, Pasini D. 2015. The controversial role of the Polycomb group proteins in transcription and cancer: how much do we not understand Polycomb proteins? *FEBS J* **282**: 1703-1722.

Schorderet P, Lonfat N, Darbellay F, Tschopp P, Gitto S, Soshnikova N, Duboule D. 2013. A genetic approach to the recruitment of PRC2 at the HoxD locus. *PLoS genetics* **9**: e1003951.

Schuettengruber B, Chourrout D, Vervoort M, Leblanc B, Cavalli G. 2007. Genome regulation by polycomb and trithorax proteins. *Cell* **128**: 735-745.

Schwartz YB, Kahn TG, Nix DA, Li XY, Bourgon R, Biggin M, Pirrotta V. 2006. Genome-wide analysis of Polycomb targets in Drosophila melanogaster. *Nature genetics* **38**: 700-705.

Schwartz YB, Kahn TG, Stenberg P, Ohno K, Bourgon R, Pirrotta V. 2010. Alternative epigenetic chromatin states of polycomb target genes. *PLoS genetics* **6**: e1000805.

Schwartz YB, Pirrotta V. 2007. Polycomb silencing mechanisms and the management of genomic programmes. *Nature reviews Genetics* **8**: 9-22.

-. 2008. Polycomb complexes and epigenetic states. *Current opinion in cell biology* **20**: 266-273.

-. 2013. A new world of Polycombs: unexpected partnerships and emerging functions. *Nature reviews Genetics* **14**: 853-864.

Sing A, Pannell D, Karaiskakis A, Sturgeon K, Djabali M, Ellis J, Lipshitz HD, Cordes SP. 2009. A vertebrate Polycomb response element governs segmentation of the posterior hindbrain. *Cell* **138**: 885-897.

Sparmann A, van Lohuizen M. 2006. Polycomb silencers control cell fate, development and cancer. *Nat Rev Cancer* **6**: 846-856.

Stenberg P, Pettersson F, Saura AO, Berglund A, Larsson J. 2005. Sequence signature analysis of chromosome identity in three Drosophila species. *BMC bioinformatics* **6**: 158.

Taherbhoy AM, Huang OW, Cochran AG. 2015. BMI1-RING1B is an autoinhibited RING E3 ubiquitin ligase. *Nat Commun* **6**: 7621.

Tardat M, Albert M, Kunzmann R, Liu Z, Kaustov L, Thierry R, Duan S, Brykczynska U, Arrowsmith CH, Peters AH. 2015. Cbx2 targets PRC1 to constitutive heterochromatin in mouse zygotes in a parent-of-origin-dependent manner. *Molecular cell* **58**: 157-171.

Tavares L, Dimitrova E, Oxley D, Webster J, Poot R, Demmers J, Bezstarosti K, Taylor S, Ura H, Koide H et al. 2012. RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. *Cell* **148**: 664-678.

Touzet H, Varre JS. 2007. Efficient and accurate P-value computation for Position Weight Matrices. *Algorithms for molecular biology : AMB* **2**: 15.

van der Lugt NM, Domen J, Linders K, van Roon M, Robanus-Maandag E, te Riele H, van der Valk M, Deschamps J, Sofroniew M, van Lohuizen M et al. 1994. Posterior transformation, neurological abnormalities, and severe hematopoietic defects in mice with a targeted deletion of the bmi-1 proto-oncogene. *Genes & development* **8**: 757-769.

Woo CJ, Kharchenko PV, Daheron L, Park PJ, Kingston RE. 2010. A region of the human HOXD cluster that confers polycomb-group responsiveness. *Cell* **140**: 99-110.

Wu X, Johansen JV, Helin K. 2013. Fbxl10/Kdm2b recruits polycomb repressive complex 1 to CpG islands and regulates H2A ubiquitylation. *Molecular cell* **49**: 1134-1146.

Xu B, On DM, Ma A, Parton T, Konze KD, Pattenden SG, Allison DF, Cai L, Rockowitz S, Liu S et al. 2015. Selective inhibition of EZH2 and EZH1 enzymatic activity by a small molecule suppresses MLL-rearranged leukemia. *Blood* **125**: 346-357.

Yates A, Akanni W, Amode MR, Barrell D, Billis K, Carvalho-Silva D, Cummins C, Clapham P, Fitzgerald S, Gil L et al. 2016. Ensembl 2016. *Nucleic acids research* **44**: D710-716.

Yu M, Mazor T, Huang H, Huang HT, Kathrein KL, Woo AJ, Chouinard CR, Labadorf A, Akie TE, Moran TB et al. 2012. Direct recruitment of polycomb repressive complex 1 to chromatin by core binding transcription factors. *Molecular cell* **45**: 330-343.

Figure legends

Figure 1. Disparate binding of PRC1 and PRC2/H3K27me3 reveals putative PRC1 targeting element upstream of the *CCND2* gene. A. ChIP profiles of MEL18, BMI1, H3K27me3 and EZH2 along the 55Mb of the left half of the human Chromosome 12. The gap corresponds to repetitive sequences around the centromere. Asterisk indicates the position of the *HOXC* cluster, the classic PcG target. Arrows mark strong anti-MEL18 and anti-BMI1 ChIP signals at the *CCND2* locus. **B**. Close up view of MEL18 ChIP profile at the *CCND2* locus. The heat-map underneath MEL18 binding profile shows the fraction of CpG nucleotides within 100bp sliding window (ranging from dark blue - lowest to bright yellow - highest). The DNA sequence underneath MEL18 peak (white rectangle, PTE) is CpG poor but immediately adjacent is the CpG-rich region that extends towards the *CCND2* TSS. The *CCND2* gene is transcribed from left to right. The lncRNAs *CCND2-AS1* and *CCND2-AS2* are transcribed from right to left. The scale indicates genomic positions in Hg19 coordinates. **C**. With an exception of *CCND2*, genomic regions that show strong ChIP-chip signals for MEL18 and BMI1 are also strongly precipitated with anti-EZH2 and H3K27me3 antibodies. The heat-map plots show ChIP-chip signals over 10kb regions centered on BMI1 ChIP signal peaks. Each row represents one region and these are sorted by the BMI1 ChIP signal strength. The enrichment values for each protein are scaled such that the strongest genomic binding corresponds to 1. **D**. RT-qPCR measurements indicate that *CCND2* is highly expressed in NT2-D1 (NT2) but very little in TIG-3 cells. Histograms show the average and the scatter (whiskers) between two independent experiments. Values are normalized to the expression of the housekeeping *GAPDH* gene.

Figure 2. Polycomb proteins and associated histone modifications at the active and repressed *CCND2* gene. A. The schematic of the *CCND2* locus. Positions of PCR amplicons analyzed in **B** are shown below the coordinate scale. Note that amplicon N2 corresponds to the summit of MEL18 peak and amplicon N5 corresponds to the *CCND2* TSS (see: Tables S4, S5 for sequence information). **B.** ChIP profiles of MEL18, RING2, SUZ12, H3K27me3 KDM2B and H2AK119ub in NT2-D1 (left column) and TIG-3 (right column) cells. The immunoprecipitation of ALX4 gene repressed by PcG in NT2-D1 and TIG-3 cells was used as positive control and the immunoprecipitation of a gene desert region (Chr.12: 127711096-127713493; Hg19) was used as a negative control. Histograms for both are shown to the right of the each graph and are to the same y-axes scale. ALX4* positive control region for H3K27me3 ChIPs is offset from the ALX4 region used for the profiling of other proteins. This is done to account for relatively low H3K27me3 at the site of the highest PcG occupancy overlapped by ALX4 amplicon. All histograms and graphs show the average and the scatter (whiskers) between two independent experiments.

Figure 3. The *CCND2* PTE can autonomously recruit PRC1. A. The schematic of experimental setup. Tested fragments, represented by 2.4kb PTE fragment, were cloned in the lentiviral vector. The vector contained *Pac* puromycin resistance gene shielded from potential repression by the chromatin insulator from mouse *Igf2/H19* Imprinting Control Region. The position of an engineered annealing site for transgene-specific PCR primer is indicated by white circle. Positions of additional amplicons are indicated as short lines. Following viral transduction and puromycin selection, cells were genotyped and collected for ChIP-qPCR assays. The results of a representative analysis of PCR products shows that transgene-specific primers amplify the product of expected size from genomic DNA of cells transduced with putative 2.4kb PTE but not with a control (gene desert) construct (see Figure S4 for

additional details). Comparison of immunoprecipitations of the 2.4kb PTE fragment (**B**) and the control 2.4kb fragment from Chromosome 12 gene desert (**C**) indicates that the PTE 2.4kb robustly recruits PRC1 and little PRC2 or H3K27me3. The precipitation of transgenic (tran) and endogenous (endo) PTE fragments was assayed using corresponding "Core" amplicons. The immunoprecipitations of the ALX4 and Chromosome 12 gene desert regions were assayed in parallel as positive and negative controls. The PCR amplicon used to assay the immunoprecipitation of the gene desert fragment in (**C**) does not discriminate between transgenic and endogenous copies. Since the endogenous site does not bind PcG any elevated ChIP signal would come from the transgenic copy. All histograms show the average and the scatter (whiskers) between two independent experiments.

Figure 4. Mouse *Ccnd2* locus contains a PTE. A. The schematic of the mouse *Ccnd2* locus. PhastCons sequence conservation scores based on the alignment of 60 vertebrate genomes illustrate that putative mouse PTE is an evolutionary conserved element. The heat-map underneath the conservation score graph shows the number of CpG nucleotides within 100bp sliding window (ranging from dark blue=0 to bright yellow=10). Similar to the human *CCND2* locus, the CpG-poor PTE is adjacent to a CpG-island that extends towards the *Ccnd2* TSS. The *Ccnd2* gene is transcribed from left to right. Also shown is lncRNA (*9330179D12Rik*) of unknown function. Positions of PCR amplicons analyzed in **B** and **C** are shown below the scale in mm10 genomic coordinates. Dashed rectangle indicates the 3.6kb fragment tested for autonomous recruitment of PRC1 in human cells. Black rectangle marks position of evolutionary conserved DNA sequence that corresponds to ~300bp fragment directly underneath the PRC1 binding peak within the human *CCND2* PTE. **B.** Cbx7 and Suz12 ChIP profiles across upstream region of *Ccnd2* in the mouse F9 cells. Note that Suz12 profile is offset from the PTE and the Cbx7 peak into the CpG-rich region. The

immunoprecipitation of *β-actin* gene (shown as bars to the right of the each graph) was used as negative control. Here and in **C** bar plots and graphs are to the same scale and show the average and the scatter (whiskers) between two independent experiments. **C.** Chip analysis of 3.6kb mouse PTE fragment integrated in the genome of human NT2-D1 cells shows that it robustly recruits PRC1 but little PRC2. Amplicons 4 and 5 shown in **A** were used to assay the transgenic mouse fragment. Immunoprecipitation of the endogenous human *CCND2* PTE and the *ALX4* locus were assayed in parallel as positive controls and precipitation of the Chromosome 12 gene desert was assayed as negative control.

Figure 5. Transgenic dissection of the *CCND2* PTE. A. Schematic representation of PRC1 recruitment capacity of different fragments from the *CCND2* PTE region. "++++" indicates recruitment comparable to endogenous locus, "+++" indicates robust recruitment but weaker than at endogenous locus, "++" indicates intermediate recruitment, "+" indicates weak recruitment. The fragment positions are shown in relation to *CCND2* TSS and Hg19 genomic coordinates. The minimal 1kb PTE 1.2 fragment capable of robust PRC1 recruitment is indicated in bold. The PTE 1.2mCGA construct corresponds to PTE 1.2 fragments with mutated "CGA" motif. **B.** ChIP-qPCR analysis of the PTE 1.2 transgene. The precipitation of transgenic DNA (tran) is compared to that of the endogenous *CCND2* PTE region (endo), unrelated PcG repressed gene ALX4 and the gene desert region (desert). Here and in **C**, **D** histograms show the average and the scatter (whiskers) between the two independent experiments. ChIP-qPCR analyses of transgenes containing PTE 1.1 (**C.**) and PTE 1.3 (**D.**) indicate that both fragments bind PRC1 to an extent seen at the endogenous *CCND2* PTE.

Figure 6. Multiple modules and the "CGA" motif contribute to PRC1 recruitment by the *CCND2* PTE. A.-F. ChIP analysis of MEL18 binding to a set of transgenic constructs

spanning different parts of PTE 1.2 indicates that it consists of at least three separable modules, all of which contribute to the recruitment but the central and the right modules contributing more. Positions of various amplicons along the transgenic constructs (grey box) are illustrated in Figure 3A. Note that some transgenic constructs lack some of the amplicons. Precipitation of transgenic fragments is compared to internal positive and negative controls (white box). The histograms show average ChIP yields normalized to that of the endogenous *CCND2* PTE (amplicon 2) with whiskers indicating the scatter between two independent experiments. **G.** Logo representation of the "CGA"- motif. **H.** The occurrence of the "CGA"-motif within high confidence genomic BMI1/MEL18 bound sites (blue) is much higher than that within control regions (red). The whiskers mark 95% confidence interval of the mean based on hundred permutations of the control group. **I.** ChIP analysis of MEL18 binding to transgenic 1kb PTE 1.2 fragment with the mutated "CGA"-motif. Four-fold reduction of ChIP signals suggests that the "CGA"-motif is important for PRC1 recruitment by the PTE.

Figure 7. Catalytic activity of PRC2 promotes PRC1 binding to the PTE. A. Western-blot analysis of two-fold serial dilutions of nuclear protein from control and SUZ12 depleted cells shows that drastic reduction of SUZ12 and H3K27me3 does not affect the overall levels of PRC1 components. Here and in **D** detection of histone H3 is used as a loading control. **B.** ChIP analysis of SUZ12, H3K27me3, MEL18 and CBX2 binding at the *CCND2* PTE and control loci in SUZ12 depleted (blue bars) and control (red bars) cells. Positions of PTE amplicons are as indicated in Figure 2A. Here and in **E** and **F** histograms show the average and the scatter (whiskers) between two independent experiments. **C**. Immunofluorescent analysis of H3K27me3 reduction in SUZ12 depleted cells. Staining for total histone H3 is used as positive control and DAPI staining as nuclear marker. **D.** Western-blot analysis of two-fold serial dilutions of nuclear protein from UNC1999-treated and control cells shows

that H3K27me3 is significantly reduced while the levels of SUZ12 and CBX2 remain unchanged. **E.** ChIP analysis of SUZ12, H3K27me3, MEL18 and CBX2 in control (red bars) and UNC1999-treated (blue bars) cells. **F.** ChIP analysis of CTCF occupancy at *CLPB*, *VDAC2*, *STK17A* genes and negative control locus (desert) in UNC1999-treated (blue bars) and control (red bars) cells.

Figure 8. Combined presence of PTE and CpG-islands paralleled by segregation of PRC1 and PRC2 binding are common features of PcG target genes. A. The heat-map of di-nucleotide distributions around high confidence BMI1 peaks. Each square represents the difference between the mean di-nucleotide count within a 100 bp window (observed) and the average count throughout the genome (expected). **B.** Heat-maps showing the enrichment of BMI1, MEL18, the methylation status of the CpG-islands and the enrichment of SUZ12 around BMI1 peaks. Here and in **C** and **D** the values are adjusted by subtracting the genomic background and linear scaling between 0 and 1. **C.** The average ChIP signals for MEL18 (red), BMI1 (blue) and SUZ12 (black) in NT2-D1 cells are plotted in 100bp increments around high confidence BMI1 binding sites. Note the clear difference between the BMI1/MEL18 and the SUZ12 profiles. **D.** The relative shift between PRC1 components and SUZ12 is also seen when the average ChIP signals for CBX7 (green), RING1 (orange) in primary fibroblasts and the SUZ12 signals (black) from embryonic stem cells are plotted around the same BMI1 binding positions. **E.** A speculative model of *CCND2* regulation by the PcG proteins. Our results suggest that *CCND2* is equipped with two DNA elements: the PTE and the CpG-island, which serves as focus of KDM2B and PRC2 binding. When *CCND2* is transcriptionally inactive, CpG-anchored KDM2B may acquire RING2-PCGF1 dimer with the help of PTE-anchored PRC1. We propose that the PTE and the CpG-island cooperate to act as the *CCND2* PRE.

Supplementary figure legends

Figure S1. The expression of genes encoding components of various RING1/RING2 complexes. Histograms show the average and the scatter (whiskers) between two independent experiments. Values are normalized to the expression of the housekeeping *GAPDH* gene.

Figure S2. BMI1 and CBX2 at the active and repressed *CCND2* gene. A. The schematic of the *CCND2* locus. Positions of PCR amplicons analyzed in **B** are shown below the coordinate scale. **B.** ChIP profiles of BMI1 and CBX2 in NT2-D1 (left column) and TIG-3 (right column) cells. The immunoprecipitation of the *ALX4* gene and a gene desert region were used as positive and negative controls. All histograms and graphs show the average and the scatter (whiskers) between two independent experiments. The CBX2 ChIP signals anywhere in the *CCND2* locus or in the positive control locus (*ALX4*) in TIG-3 cells are very low likely because the corresponding gene is poorly expressed in these cells.

Figure S3. Immunoprecipitations of RING2 from chromatins crosslinked with formaldehyde or with a combination of EGS and formaldehyde give the same result. A. The schematic of the *CCND2* locus. RING2 ChIP profiles from variously crosslinked NT2-D1 (**B.**) and TIG-3 (**C.**) cells. The immunoprecipitation of the *ALX4* gene and a gene desert region were used as positive and negative controls. All histograms and graphs show the average and the scatter (whiskers) between two independent experiments.

Figure S4. Evolutionary conservation of the DNA sequence underneath the *CCND2* MEL18/BMI1 peak. **A.** The comparison of the MEL18 binding profile with the PhastCons sequence conservation scores based on the alignment of 46 mammalian genomes highlights a

small stretch of nucleotides close to the summit of the MEL18 peak that shows little conservation (rectangular box). The lower panel shows the location of five nucleotides (shown in red under the sequence alignment) substituted to generate an annealing site for a transgene-specific PCR primer. The dashed lines mark the boundaries of the 2.4kb DNA fragment used in transgenic assays. The genomic coordinates are in release Hg 19. **B.** The gene desert region from Chromosome 12 selected as negative control. The dashed lines mark the boundaries of the DNA fragment used in transgenic assays. The ChIP-chip signals for EZH2, H3K27me3 and Mel18 indicate the lack of binding paralleled by poor evolutionary conservation.

Figure S5. PCR genotyping of transgenic cell lines. All transgenic lines were PCR genotyped using ChIP input material as a DNA template. **A.** The results of a representative analysis of PCR products shows that transgene-specific primers (PTE tran) amplify the product of expected size from genomic DNA of the cells transduced with PTE 2.4kb construct but not with the control (Gene desert) construct. This verifies both the identity of the cell lines and the specificity of the "PTE tran" primers. **B.** The analyses similar to those in **A.** were performed for cell lines transduced with PTE sub-fragments. Sequences of corresponding primers are indicated in the Table S3. **C.** Anticipated results of PCR genotyping are summarized in the table that indicates expected sizes of PRC products. The dash indicates that no product should be produced. Note that the results in **B.** match the expected outcomes.

Figure S6. The empty lentiviral vector does not recruit PcG complexes. A. The schematic of lentiviral vector backbone. The vector contains LTR sequences required for genomic

integration and *Pac* puromycin resistance gene flanked by the chromatin insulator from mouse *Igf2/H19* Imprinting Control Region (ICR). Positions of amplicons used for ChIP-qPCR analyses in **B.** are indicated as short lines. Note that LF1/RF amplicon overlaps the restriction site (arrowhead) used for cloning of PTE and control fragments. **B.** The ChIP-qPCR analyses with the indicated antibodies show the background binding that is much weaker than that seen at the endogenous *CCND2* PTE or *ALX4* gene. The histograms show the average and the scatter (whiskers) between the two independent experiments.

Figure S7. ChIP analysis of the BMI1 recruitment by PTE 1.2 sub-fragments. Positions of various amplicons along the transgenic constructs (grey box) are as in Figure 3A. Note that some transgenic constructs lack some of the amplicons. Precipitation of transgenic fragments is compared to internal positive and negative controls (white box). The histograms show the average and the scatter (whiskers) between the two independent experiments.

Figure S8. Distribution of sequence motifs in the vicinity of the *CCND2* PTE. A. The top tracks show the distribution of predicted recognition sequences for YY1, REST and SNAIL proteins. A composite distribution of all transcription factors (TFBS) mapped by ENCODE and positions of predicted CpG-islands (green rectangles) are shown below. **B.** Logo representation of the “CGCG”-motif. **C.** The alignment of DNA sequences from 21 vertebrate species and PhastCons score (lower panel) illustrate that the sequence of “CGA”-motif (red boxes) within *CCND2* PTE is highly evolutionary conserved. The majority of the species have one “CGA”-motif in the same position as in the human sequence. Some species have the second instance of the “CGA”-motif. In sheep (*Ovis aries*) the first instance of the motif is absent but the second one is present.

Figure S9. ChIP analysis of the CBX2 binding to the PTE 1.2mCGA transgene. Precipitation of transgenic fragment (grey box) is compared to internal positive and negative controls (white box). Positions of amplicons within the construct are indicated in Figure 3A.The histograms show the average and the scatter (whiskers) between the two independent experiments.

Figure S10. Sequence specific binding of nuclear proteins to the 1kb PTE 1.2. A. 100-150bp DNA fragments tiling 1kb PTE 1.2 are shown as rectangles. Fragments that show *in vitro* binding are marked in black. The vertical dashed line indicates the position of the *CCND2* MEL18/BMI1 binding peak summit. **B**. Representative EMSA experiments demonstrate that the fragments 2, 5 and 7 but not the fragment 6 shift the electrophoretic mobility (black circles) after incubation with the nuclear extract (NE). This shift is inhibited by the addition of an excess of specific unlabeled competitor (SC) while the addition of non-specific competitor (NSC) has no effect. In this example fragment 5 was used as a non-specific competitor for the EMSA with fragment 2 and vice versa. This illustrates that the fragments 2 and 5 are targeted by distinct sequence specific DNA binding activities. Similar tests were done for the fragment 7 (not shown). Note that the image of the EMSA with fragment 7 was acquired after much shorter exposure than those for fragments 2, 5 and 6.

Figure S11. The control heat-map of di-nucleotide distributions within 5 prime regions of human genes. Each square represents the difference between the mean di-nucleotide count within a 100 bp window (observed) and the average count throughout the genome (expected). The midpoint (dashed line) corresponds to position +2355bp upstream of all annotated human TSS, which matches the median distance from a high-confidence BMI1 peak to the nearest TSS.

Figure S12. *ZIC2* gene contains a putative PTE. A. The schematic of human *ZIC2* locus. ChIP-seq with chromatin from NT2-D1 cells indicates strong binding of MEL18 but not SUZ12 to a narrow region (a putative PTE) downstream of *ZIC2* gene. The heat-map underneath ChIP-seq profiles shows the number of CpG nucleotides within 100bp sliding window (ranging from dark blue=0 to bright yellow=10). The *ZIC2* gene is transcribed from left to right and the lncRNA *LINC00554* of unknown function is transcribed from right to left. Positions of PCR amplicons analyzed in **B** and **C** are shown below the scale in Hg19 genomic coordinates. **B.** RT-qPCR measurements indicate that *ZIC2* is highly expressed in NT2-D1 (NT2) but very little in TIG-3 cells. *LINC00554* is not expressed in either of the cell lines. Histograms show the average and the scatter (whiskers) between two independent experiments. Values are normalized to the expression of the housekeeping *GAPDH* gene. **C.** ChIP profiles of MEL18, SUZ12 and H3K27me3 in NT2-D1 (left column) and TIG-3 (right column) cells. The immunoprecipitation of ALX4 gene repressed by PcG in NT2-D1 and TIG-3 cells was used as positive control and Chromosome 12 gene desert was assayed as negative control. Histograms for both are shown to the right of the each graph and are to the same y-axes scale. ALX4* positive control region for H3K27me3 ChIPs is offset from the ALX4 region used for the profiling of other proteins. All histograms and graphs show the average and the scatter (whiskers) between two independent experiments. Note that in TIG-3 cells H3K27me3 and Suz12 profiles are offset from the MEL18 peak into neighboring CpG-rich regions.

Figure 1

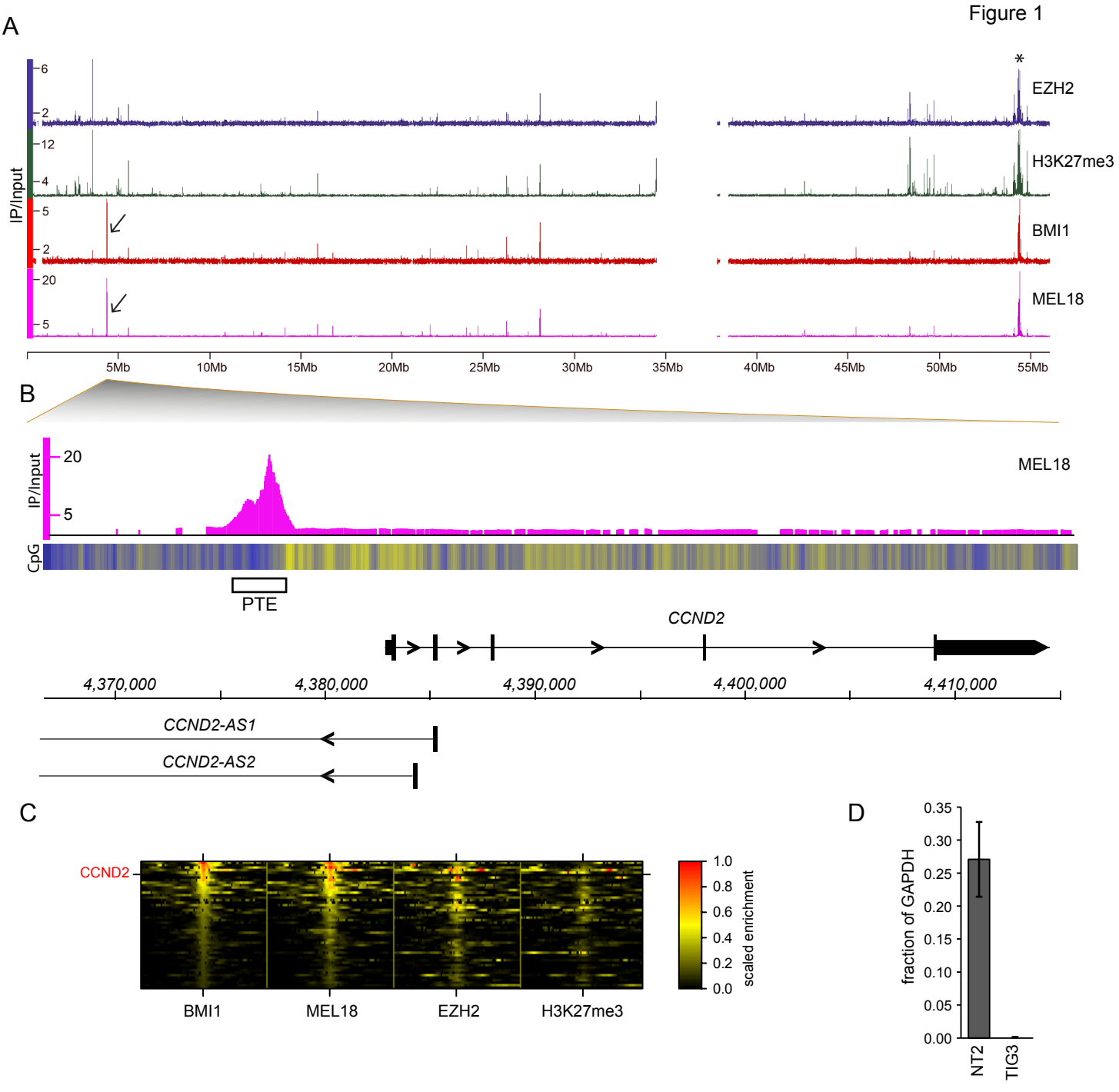
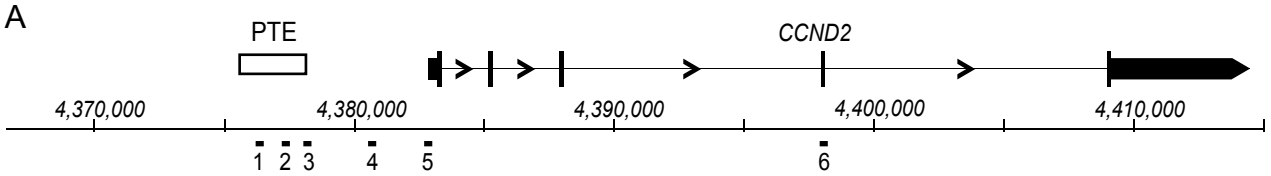

A
IP/Input
6
2
12
4
5
2
20
5
EZH2
H3K27me3
BMI1
MEL18
5Mb
10Mb
15Mb
20Mb
25Mb
30Mb
35Mb
40Mb
45Mb
50Mb
55Mb
B
IP/Input
20
5
MEL18
CpG
PTE
CCND2
4,370,000
4,380,000
4,390,000
4,400,000
4,410,000
CCND2-AS1
CCND2-AS2
C
CCND2
BMI1
MEL18
EZH2
H3K27me3
1.0
0.8
0.6
0.4
0.2
0.0
scaled enrichment
D
fraction of GAPDH
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
NT2
TIG3

Figure 2


A
PTE
CCND2
4,370,000
4,380,000
4,390,000
4,400,000
4,410,000
1 2 3 4 5 6

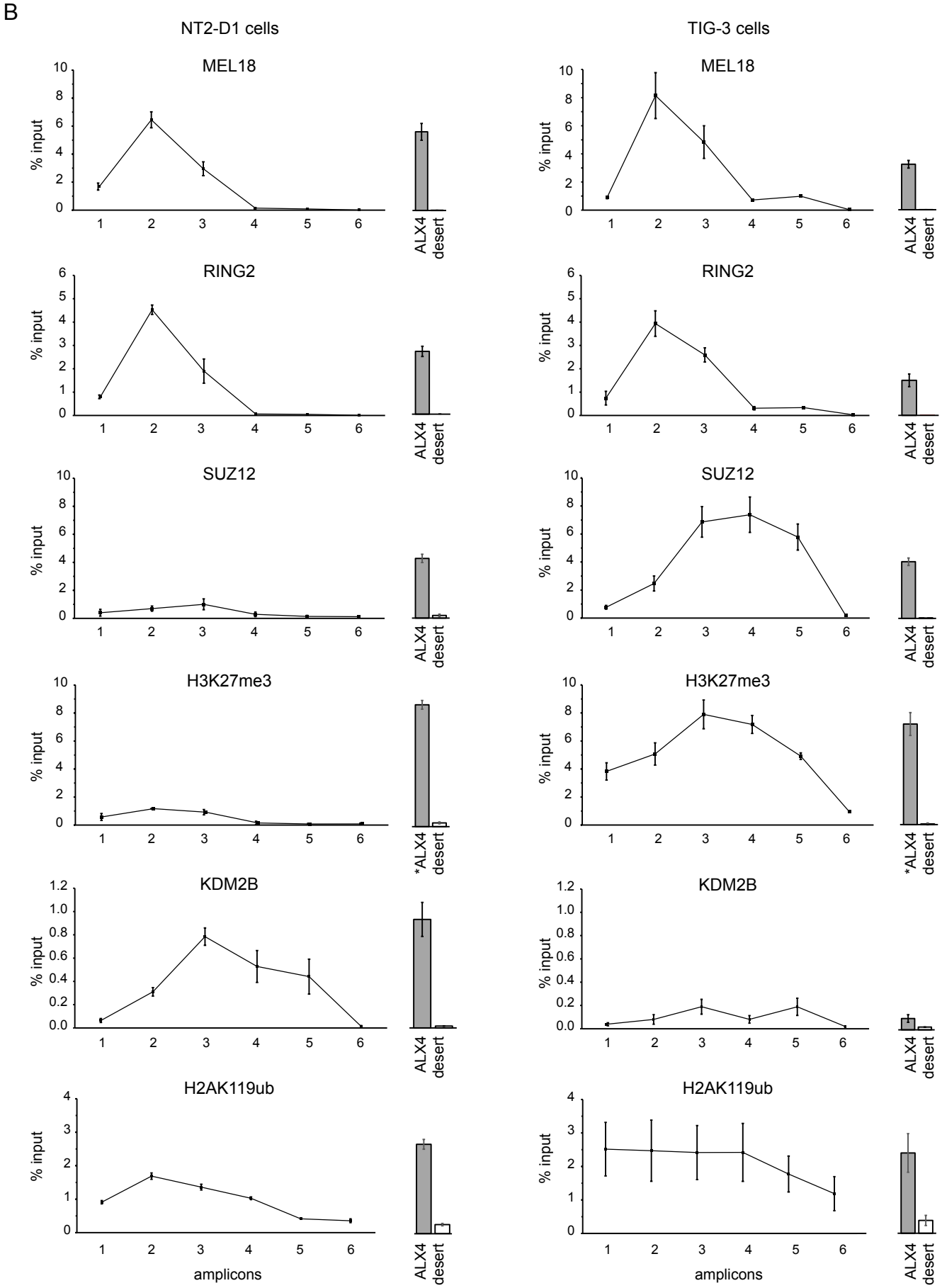
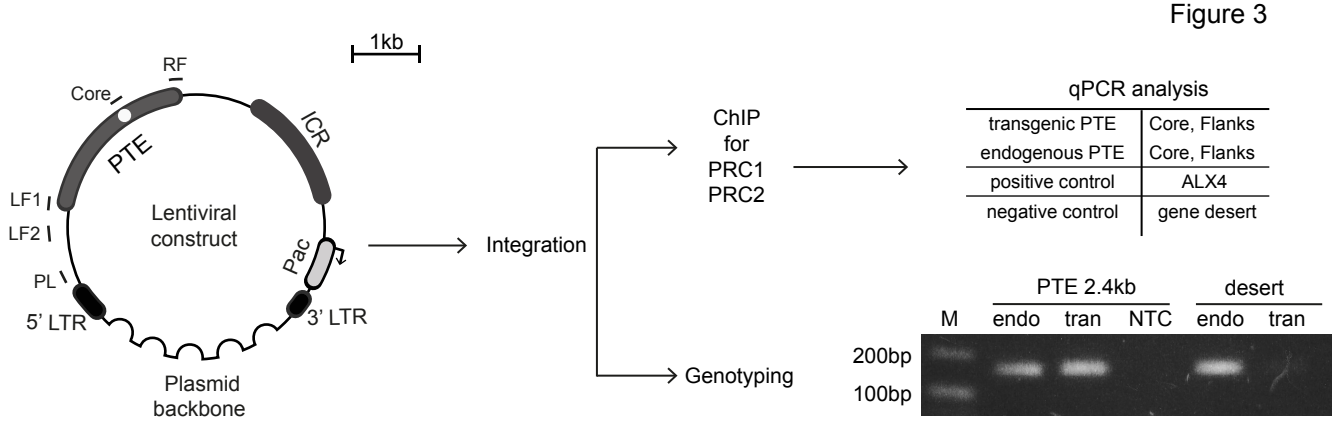
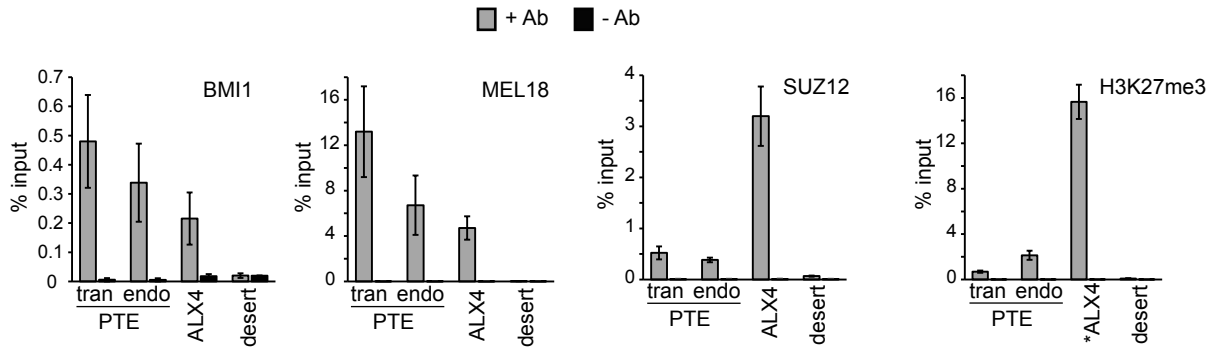

B
NT2-D1 cells
TIG-3 cells
MEL18
RING2
SUZ12
H3K27me3
KDM2B
H2AK119ub
% input
amplicons
ALX4
desert
*ALX4

Figure 3

A


1kb
RF
Core
PTE
ICR
LF1
LF2
Lentiviral construct
Pac
PL
5' LTR
3' LTR
Plasmid backbone
Integration
ChIP for PRC1 PRC2
Genotyping
qPCR analysis
transgenic PTE
Core, Flanks
endogenous PTE
Core, Flanks
positive control
ALX4
negative control
gene desert
PTE 2.4kb
desert
M
endo
tran
NTC
endo
tran
200bp
100bp

B


+ Ab
- Ab
BMI1
MEL18
SUZ12
H3K27me3
% input
tran
endo
PTE
ALX4
desert
*ALX4

C

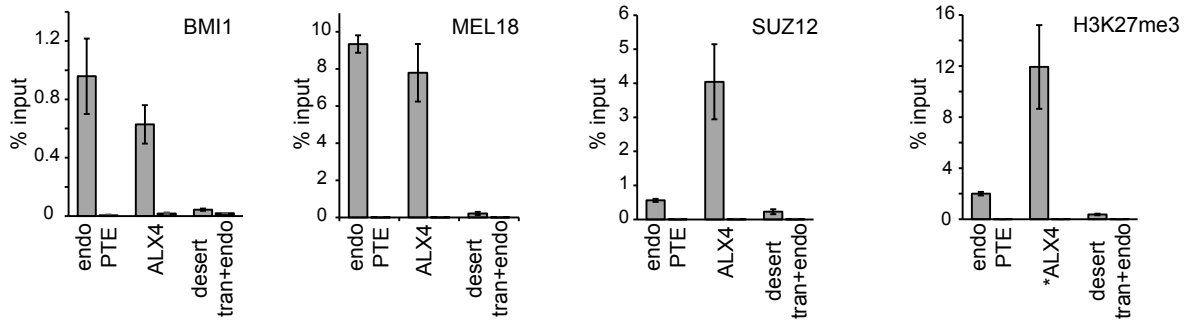

BMI1
MEL18
SUZ12
H3K27me3
% input
endo
PTE
ALX4
desert
tran+endo
*ALX4

Figure 4

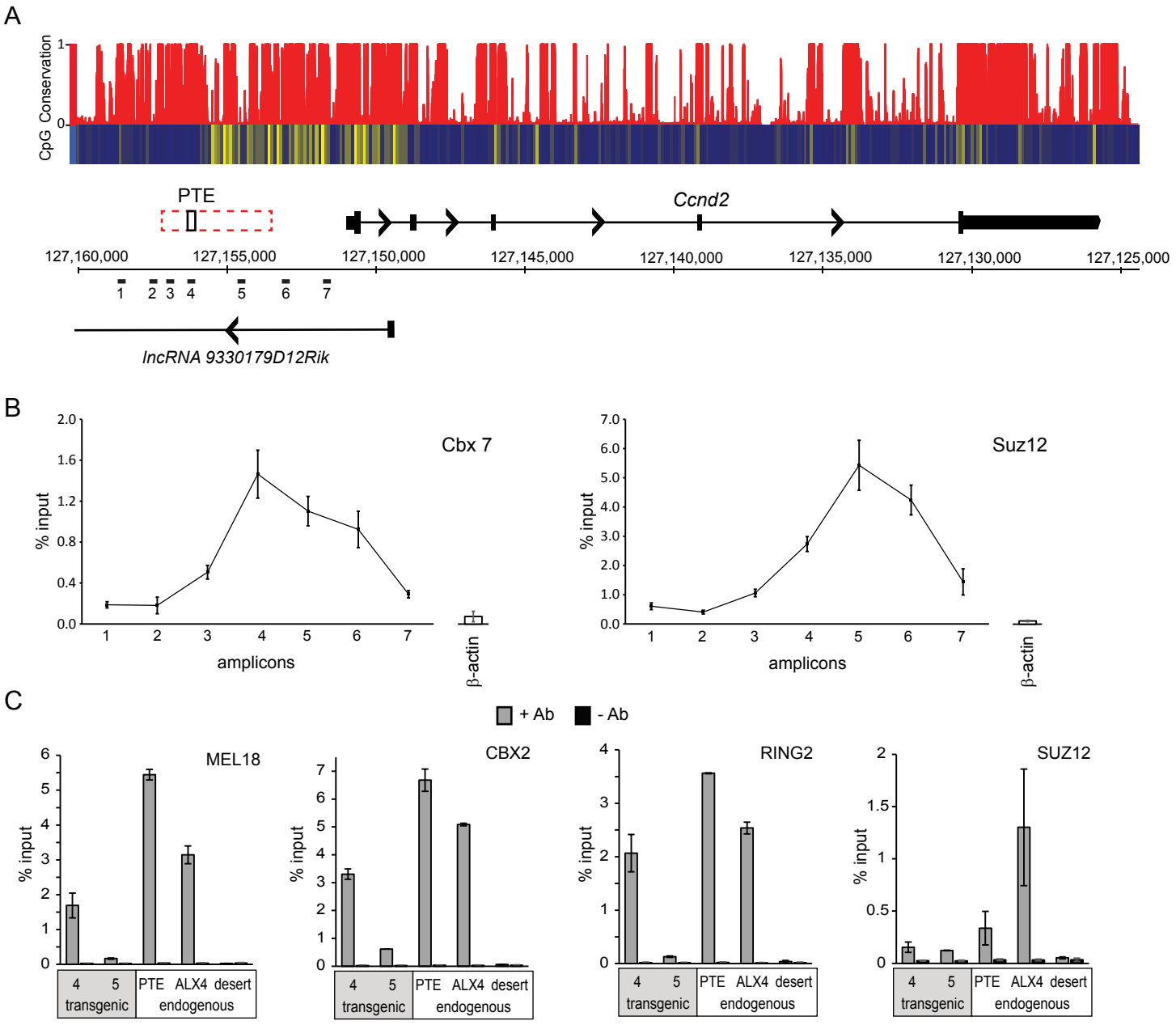
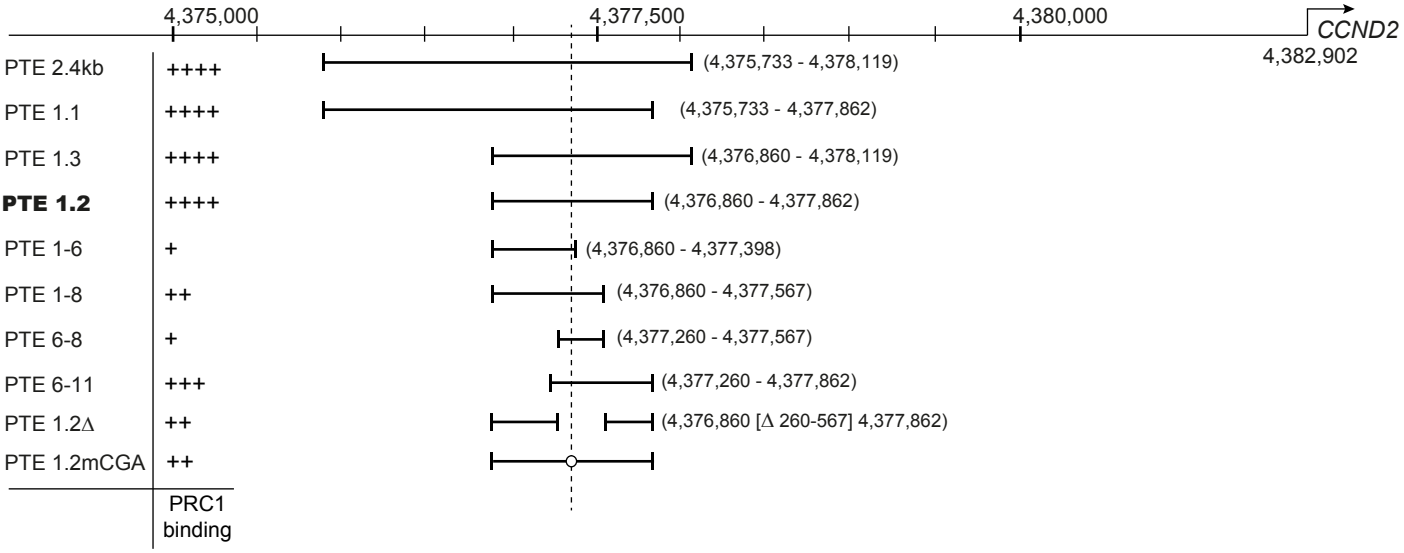
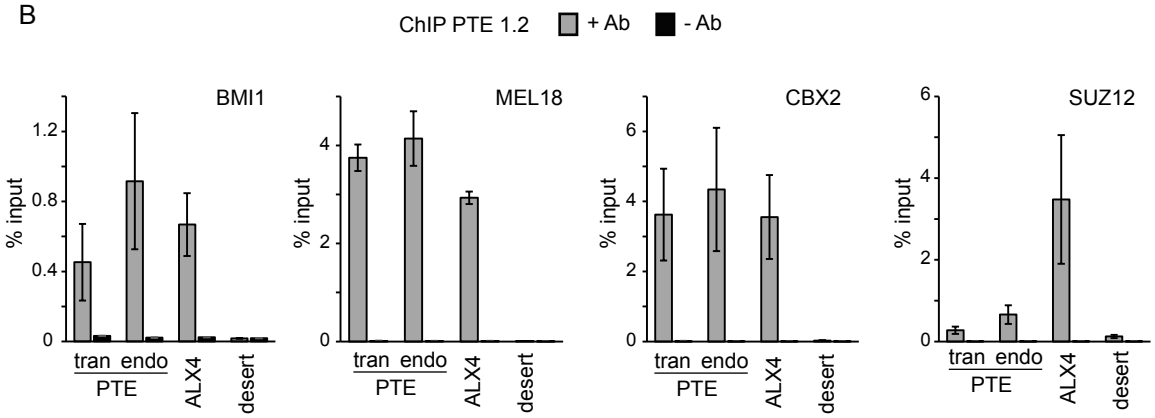

A
CpG Conservation
1
0
PTE
Ccnd2
127,160,000
127,155,000
127,150,000
127,145,000
127,140,000
127,135,000
127,130,000
127,125,000
1 2 3 4 5 6 7
lncRNA 9330179D12Rik
B
Cbx 7
% input
amplicons
β-actin
Suz12
% input
amplicons
β-actin
C
+ Ab
- Ab
MEL18
CBX2
RING2
SUZ12
% input
4 5 PTE ALX4 desert
transgenic
endogenous

Figure 5

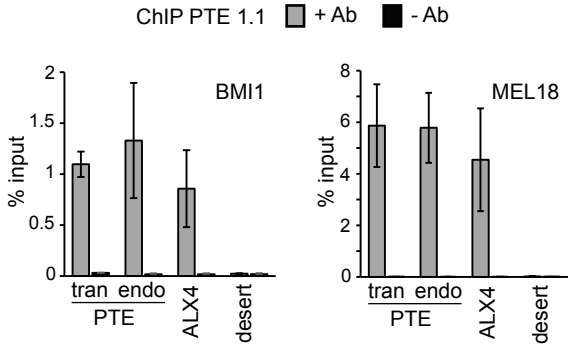
A



B



C



D

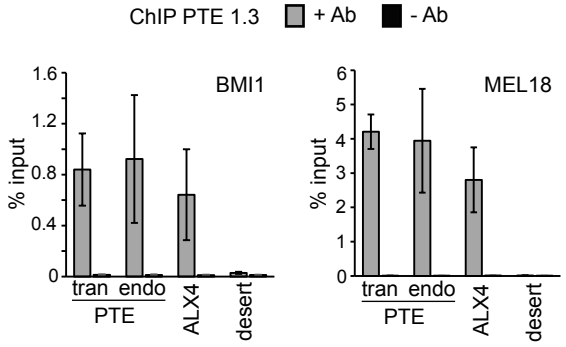
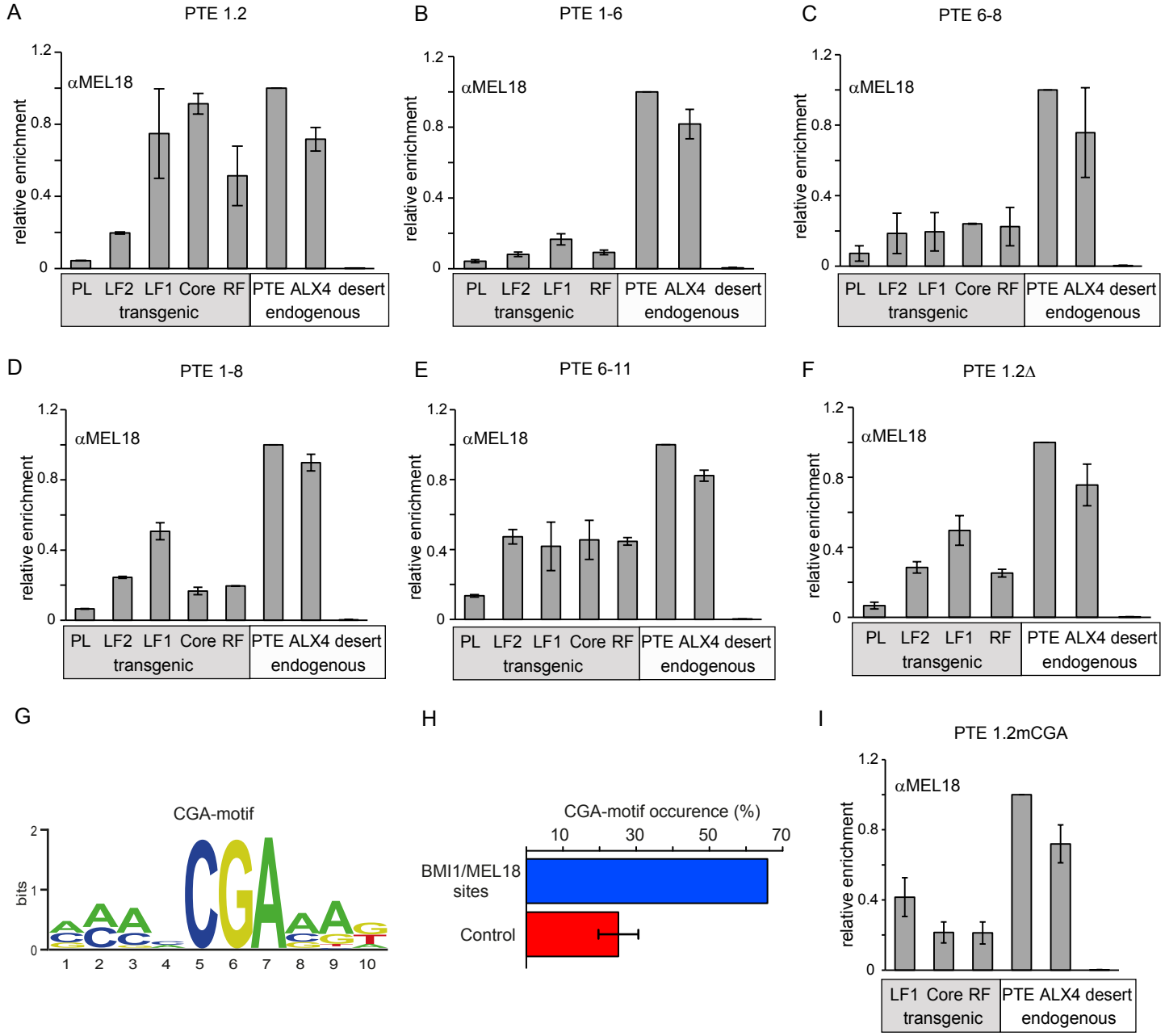


Figure 6


A
PTE 1.2
αMEL18
relative enrichment
1.2
0.8
0.4
0
PL LF2 LF1 Core RF
transgenic
PTE ALX4 desert
endogenous
B
PTE 1-6
αMEL18
relative enrichment
PL LF2 LF1 RF
transgenic
PTE ALX4 desert
endogenous
C
PTE 6-8
αMEL18
relative enrichment
PL LF2 LF1 Core RF
transgenic
PTE ALX4 desert
endogenous
D
PTE 1-8
αMEL18
relative enrichment
PL LF2 LF1 Core RF
transgenic
PTE ALX4 desert
endogenous
E
PTE 6-11
αMEL18
relative enrichment
PL LF2 LF1 Core RF
transgenic
PTE ALX4 desert
endogenous
F
PTE 1.2Δ
αMEL18
relative enrichment
PL LF2 LF1 RF
transgenic
PTE ALX4 desert
endogenous

G
CGA-motif
bits
2
1
0
1 2 3 4 5 6 7 8 9 10

H
CGA-motif occurence (%)
10 30 50 70
BMI1/MEL18
sites
Control

I
PTE 1.2mCGA
αMEL18
relative enrichment
1.2
0.8
0.4
0
LF1 Core RF
transgenic
PTE ALX4 desert
endogenous

Figure 7

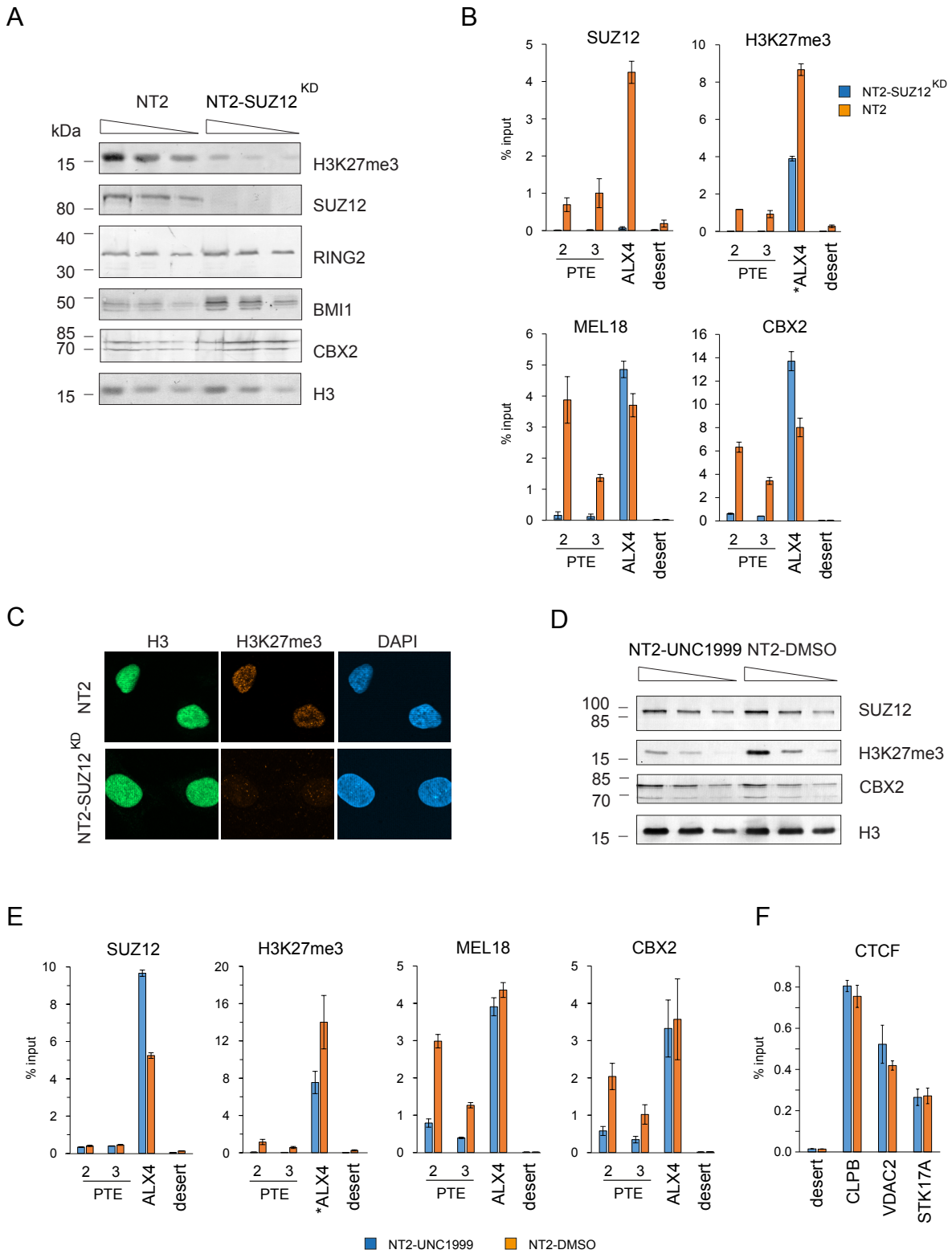

A
NT2
NT2-SUZ12 KD
kDa
15
80
40
30
50
85
70
15
H3K27me3
SUZ12
RING2
BMI1
CBX2
H3
B
SUZ12
H3K27me3
MEL18
CBX2
% input
NT2-SUZ12 KD
NT2
2 3
PTE
ALX4
*ALX4
desert
C
H3
H3K27me3
DAPI
NT2
NT2-SUZ12 KD
D
NT2-UNC1999 NT2-DMSO
100
85
15
85
70
15
SUZ12
H3K27me3
CBX2
H3
E
SUZ12
H3K27me3
MEL18
CBX2
% input
2 3
PTE
ALX4
*ALX4
desert
NT2-UNC1999
NT2-DMSO
F
CTCF
% input
desert
CLPB
VDAC2
STK17A

Figure 8

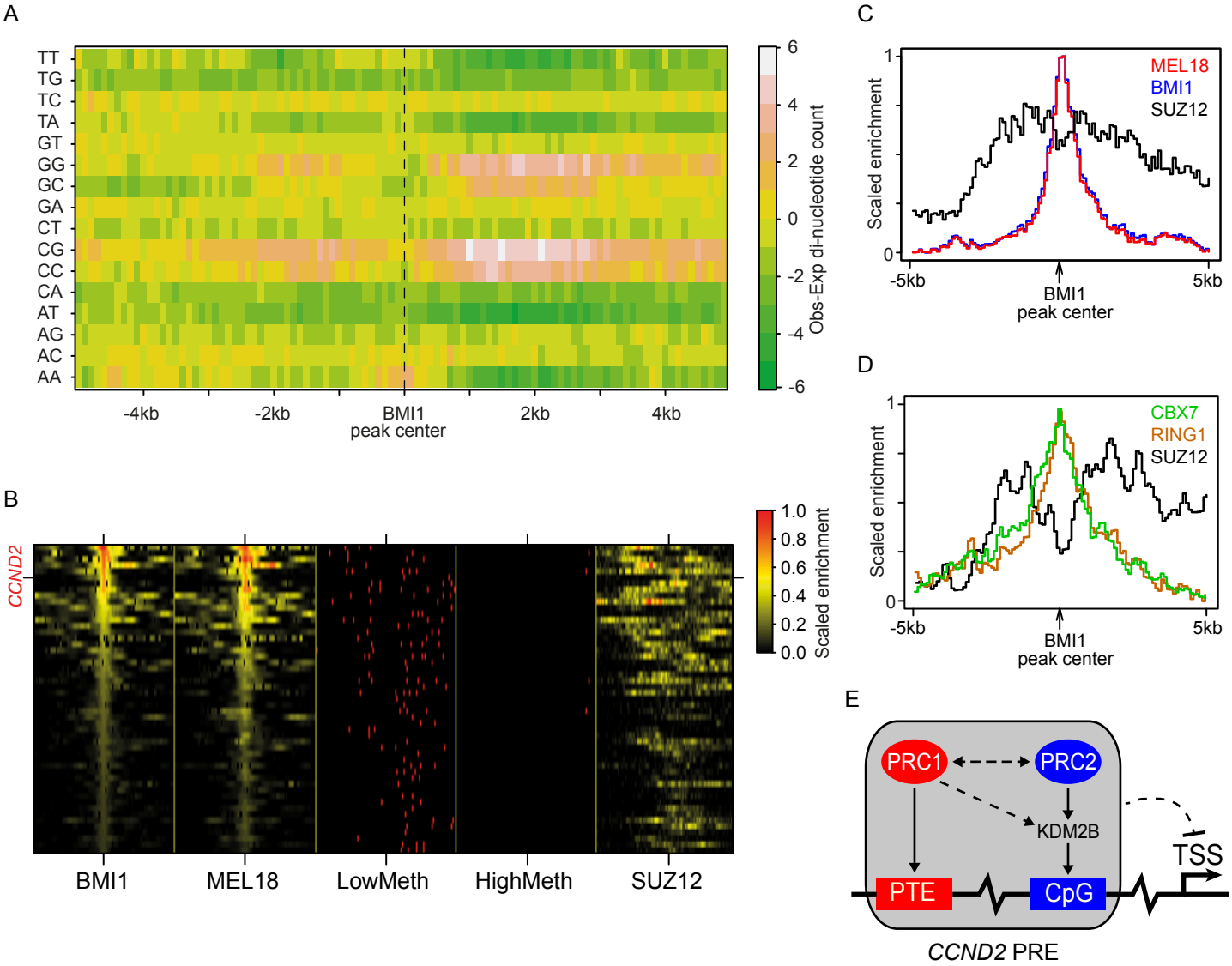

A
TT
TG
TC
TA
GT
GG
GC
GA
CT
CG
CC
CA
AT
AG
AC
AA
-4kb
-2kb
BMI1
peak center
2kb
4kb
6
4
2
0
-2
-4
-6
Obs-Exp di-nucleotide count
B
CCND2
1.0
0.8
0.6
0.4
0.2
0.0
Scaled enrichment
BMI1
MEL18
LowMeth
HighMeth
SUZ12
C
MEL18
BMI1
SUZ12
Scaled enrichment
1
0
-5kb
BMI1
peak center
5kb
D
CBX7
RING1
SUZ12
Scaled enrichment
1
0
-5kb
BMI1
peak center
5kb
E
PRC1
PRC2
KDM2B
TSS
PTE
CpG
CCND2 PRE

Figure S1

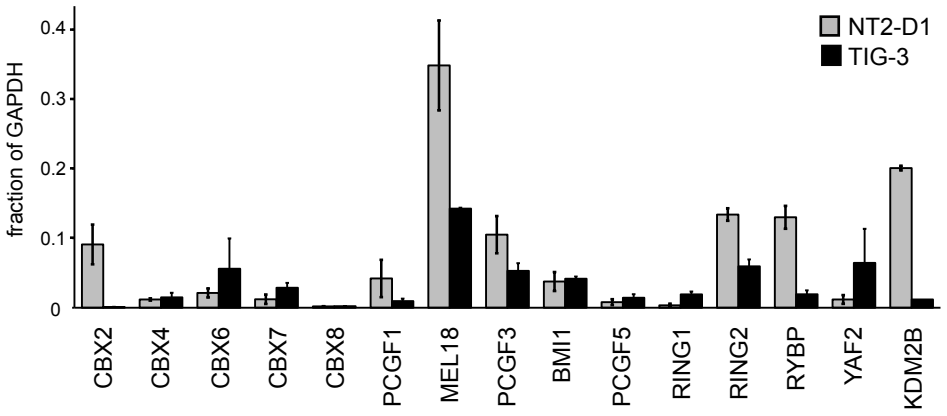
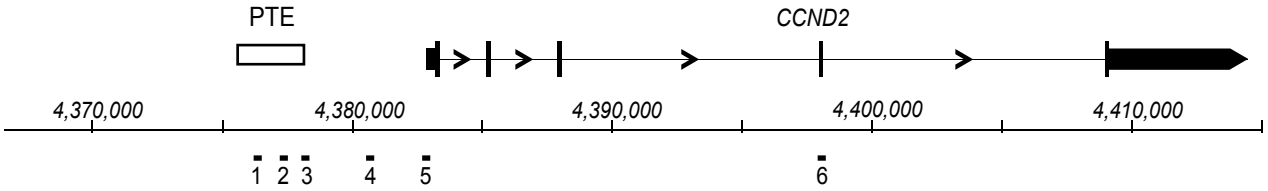

NT2-D1
TIG-3
fraction of GAPDH
0.4
0.3
0.2
0.1
0
CBX2
CBX4
CBX6
CBX7
CBX8
PCGF1
MEL18
PCGF3
BMI1
PCGF5
RING1
RING2
RYBP
YAF2
KDM2B

Figure S2

A


PTE
CCND2
4,370,000
4,380,000
4,390,000
4,400,000
4,410,000
1 2 3 4 5
6

B

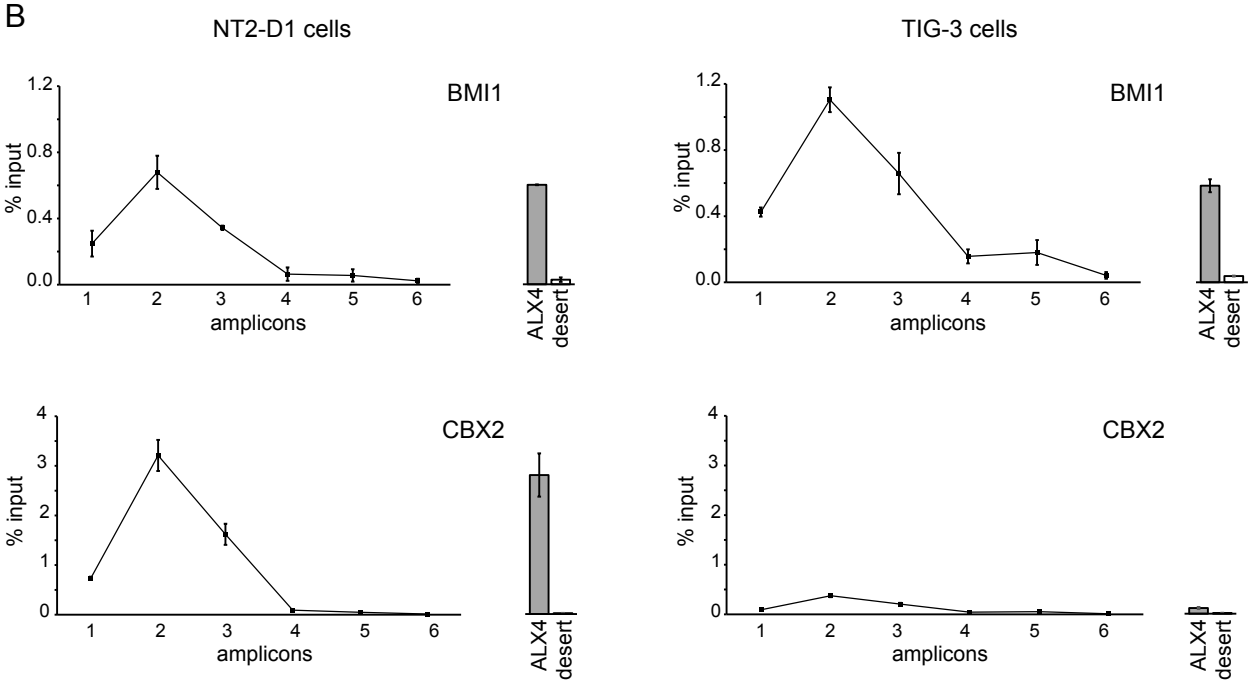
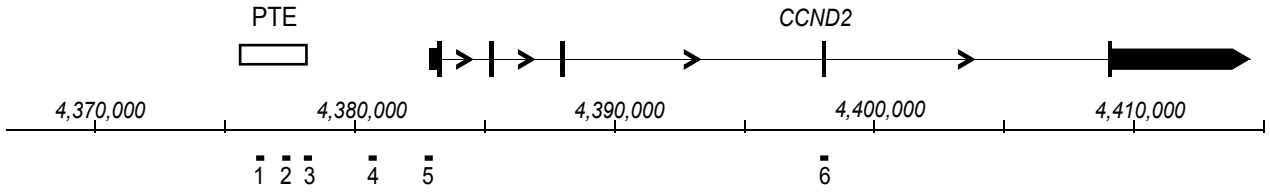
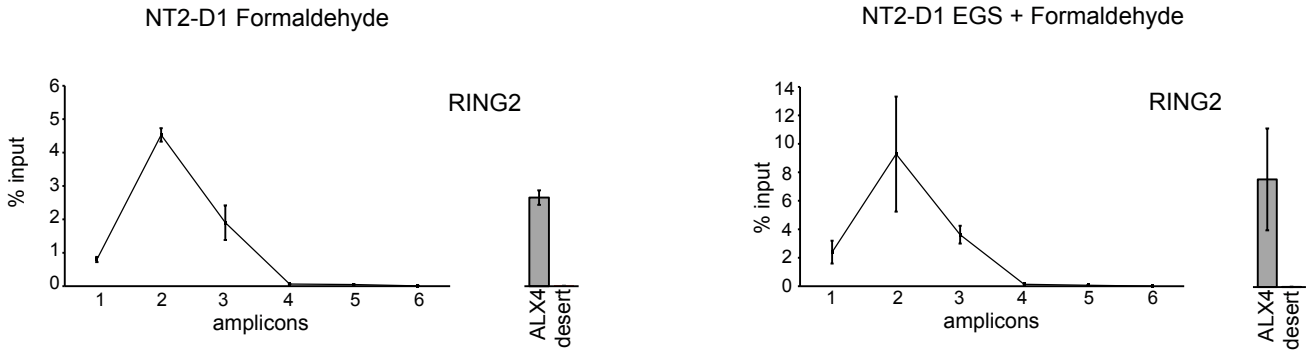

NT2-D1 cells
TIG-3 cells
BMI1
CBX2
% input
amplicons
ALX4
desert

Figure S3

A



B



C

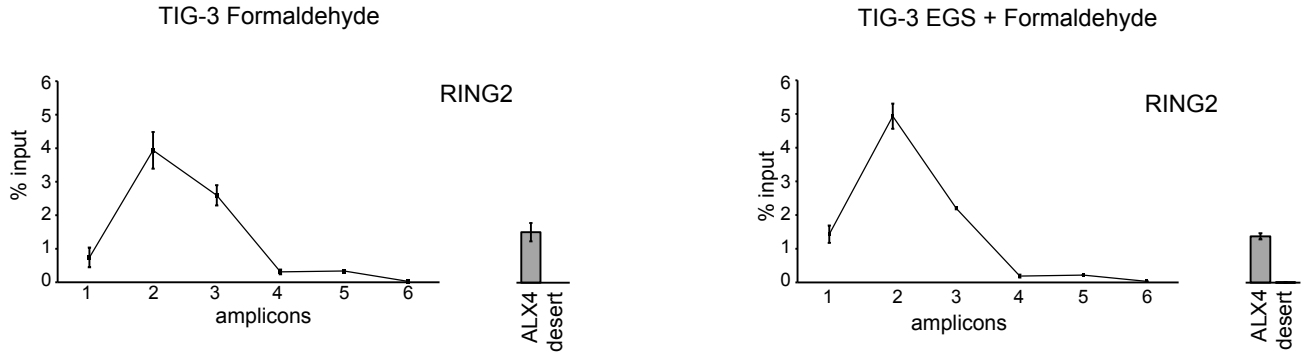
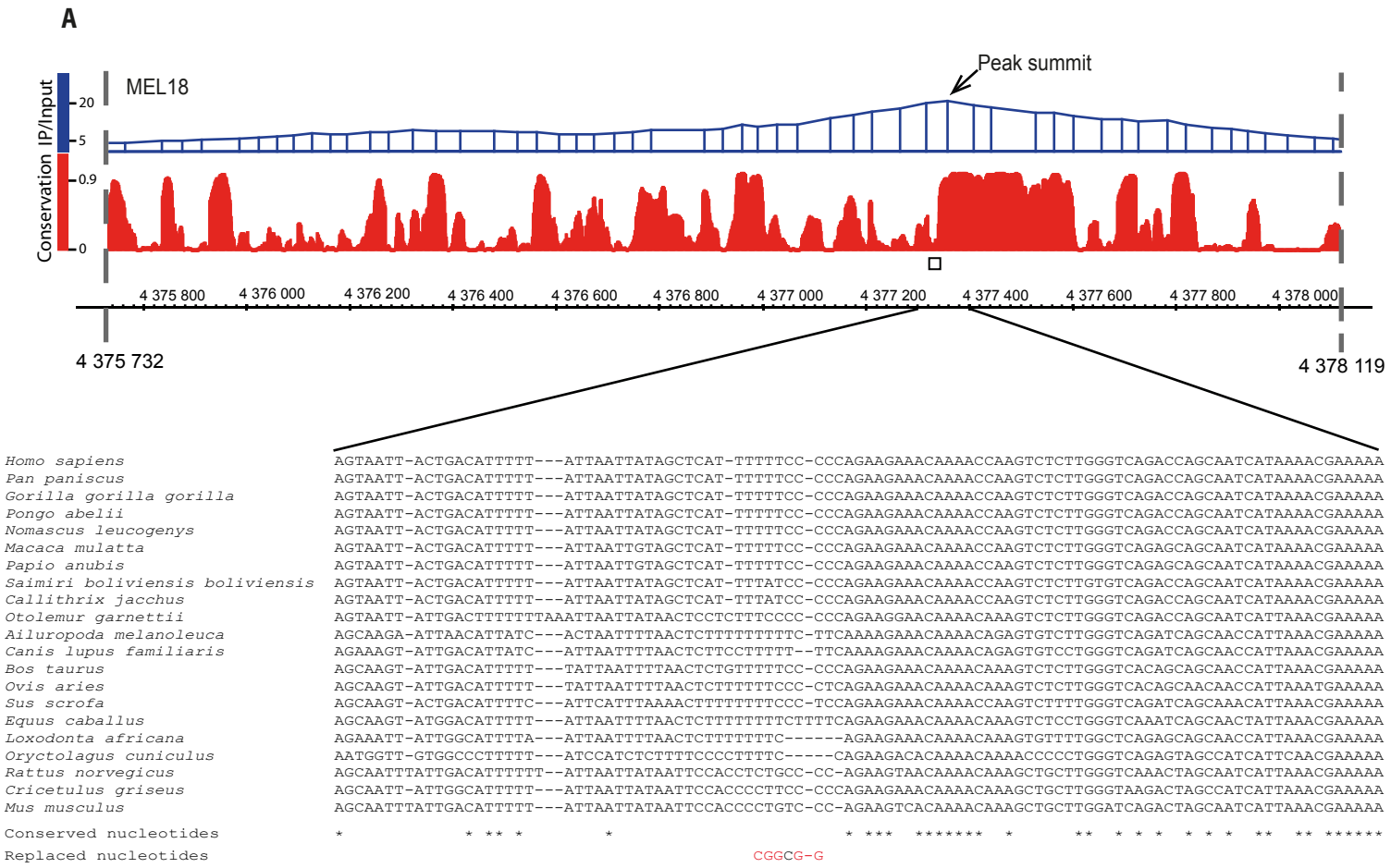


Figure S4


A
MEL18
Peak summit
IP/Input
Conservation
20
5
0.9
0
4 375 800
4 376 000
4 376 200
4 376 400
4 376 600
4 376 800
4 377 000
4 377 200
4 377 400
4 377 600
4 377 800
4 378 000
4 375 732
4 378 119
Homo sapiens
Pan paniscus
Gorilla gorilla gorilla
Pongo abelii
Nomascus leucogenys
Macaca mulatta
Papio anubis
Saimiri boliviensis boliviensis
Callithrix jacchus
Otolemur garnettii
Ailuropoda melanoleuca
Canis lupus familiaris
Bos taurus
Ovis aries
Sus scrofa
Equus caballus
Loxodonta africana
Oryctolagus cuniculus
Rattus norvegicus
Cricetulus griseus
Mus musculus
Conserved nucleotides
Replaced nucleotides
CGGCG-G

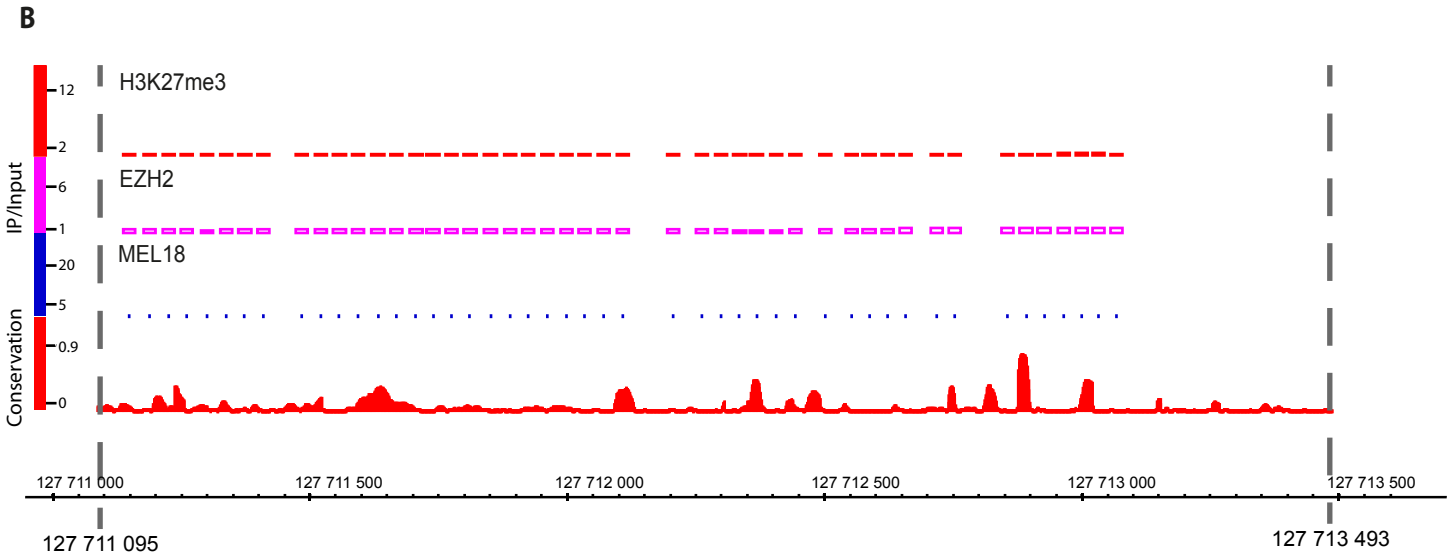
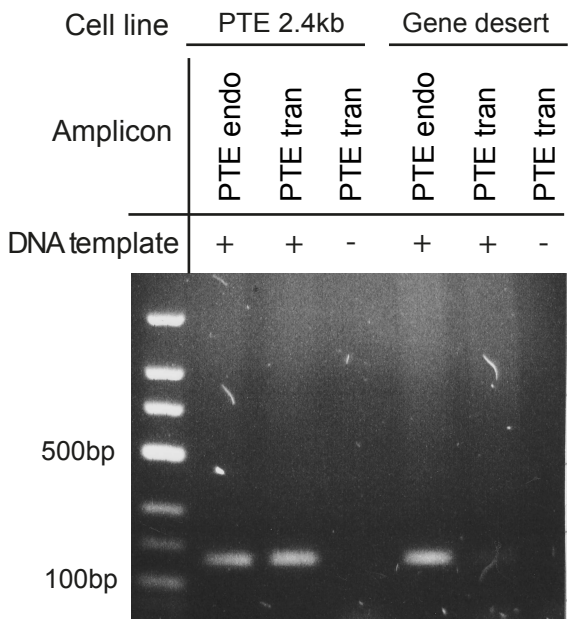
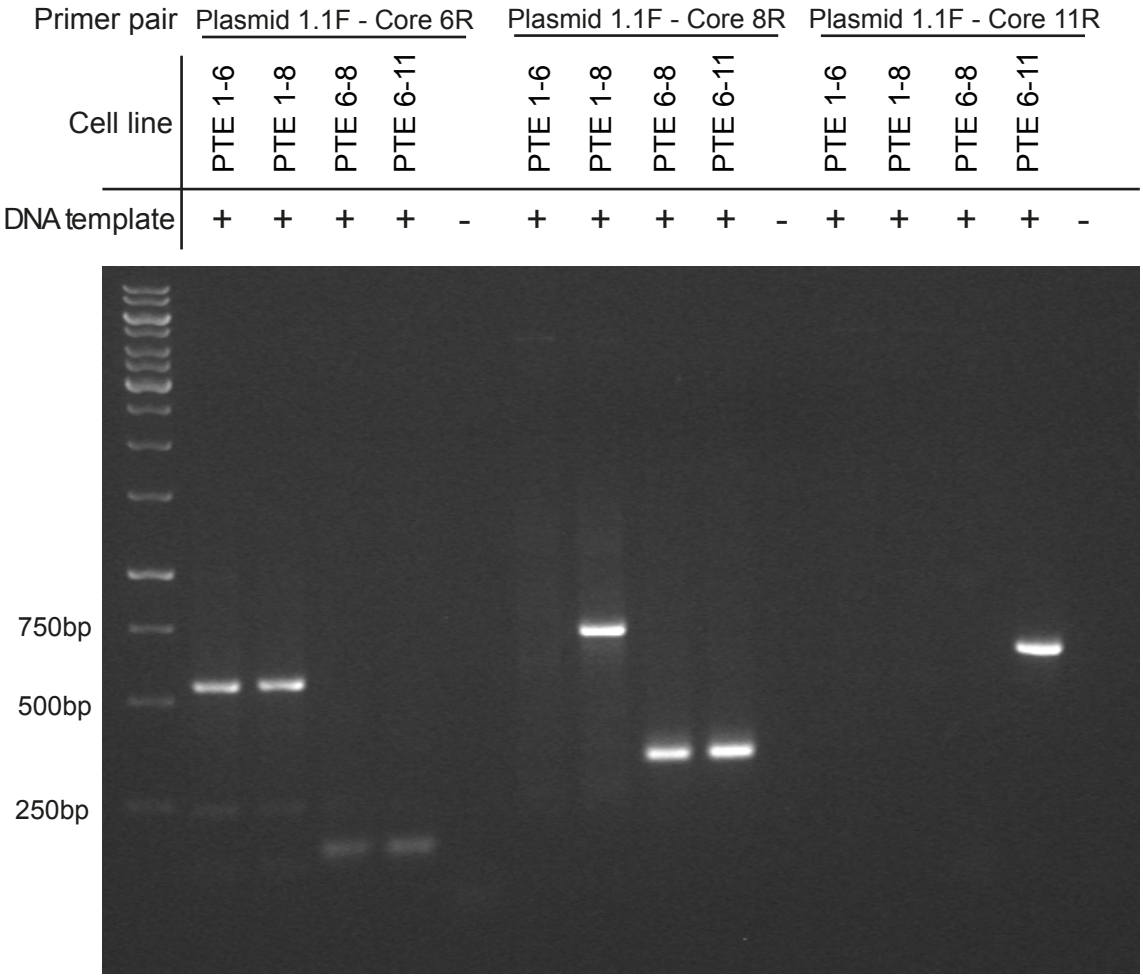

B
H3K27me3
EZH2
MEL18
IP/Input
Conservation
12
2
6
1
20
5
0.9
0
127 711 000
127 711 500
127 712 000
127 712 500
127 713 000
127 713 500
127 711 095
127 713 493

Figure S5

A



B

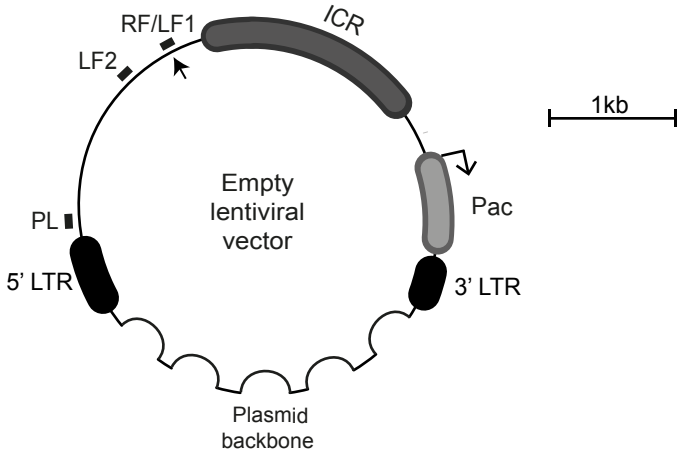


C

Primer pair	Cell line			
	PTE 1-6	PTE 1-8	PTE 6-8	PTE 6-11
Plasmid 1.1F - Core 6R	552bp	552bp	152bp	152bp
Plasmid 1.1F - Core 8R	-	730bp	330bp	330bp
Plasmid 1.1F - Core 11R	-	-	-	625bp

Figure S6

A


RF/LF1
ICR
LF2
1kb
Empty
lentiviral
vector
Pac
PL
5' LTR
3' LTR
Plasmid
backbone

B

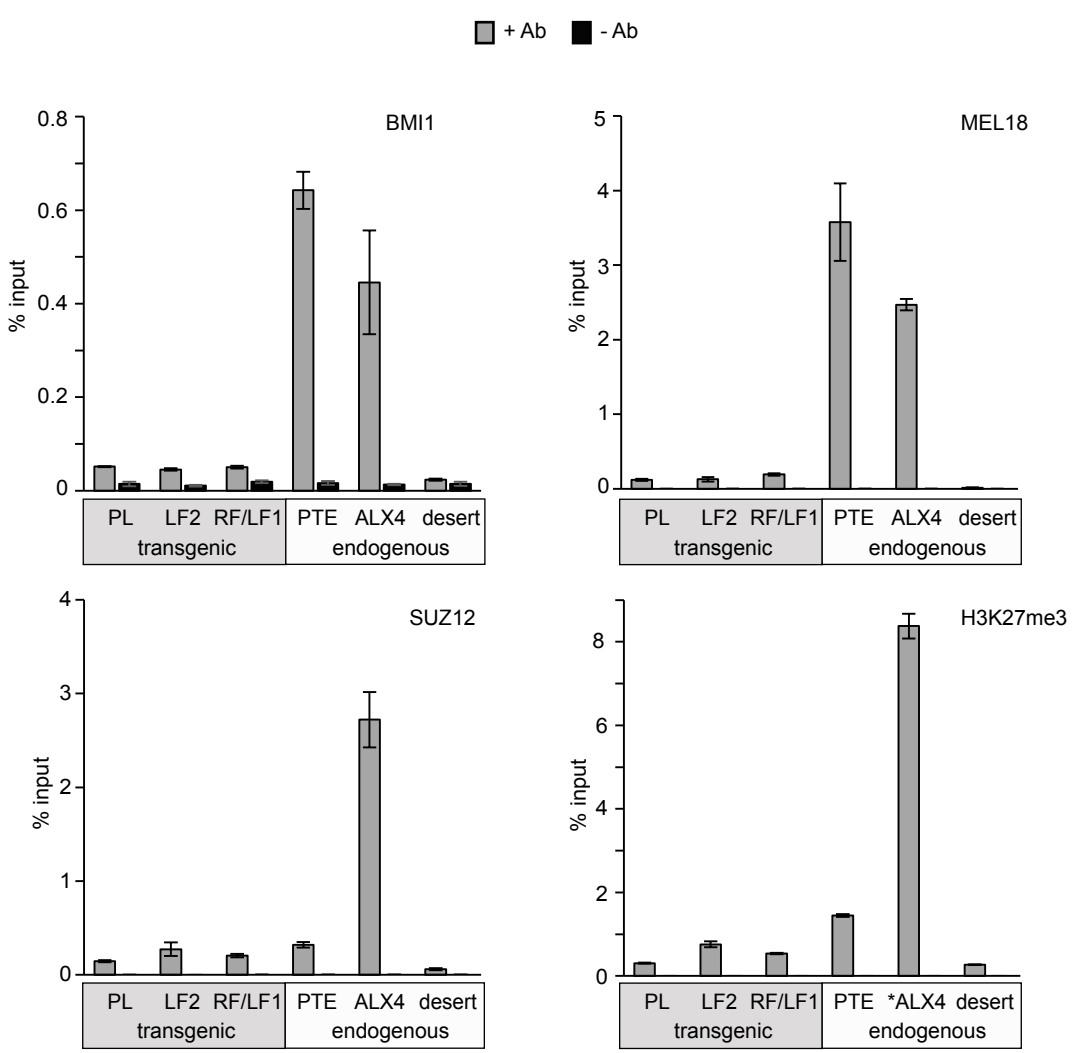

+ Ab
- Ab
BMI1
MEL18
SUZ12
H3K27me3
% input
PL
LF2
RF/LF1
PTE
ALX4
*ALX4
desert
transgenic
endogenous

Figure S7

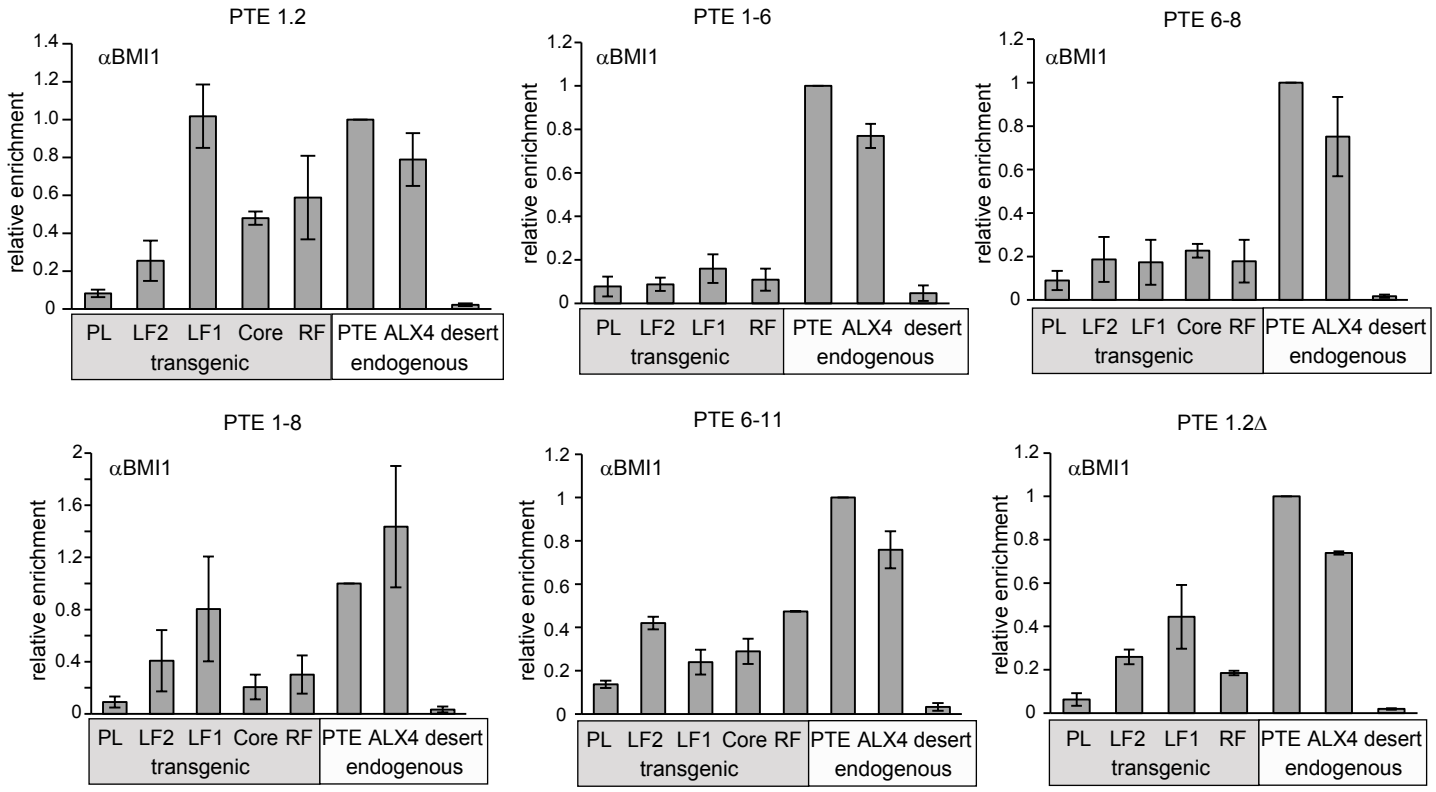
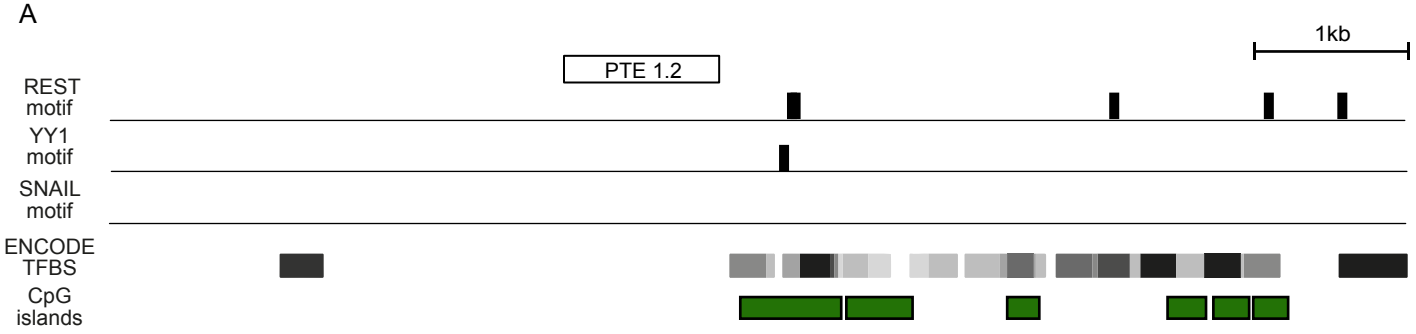
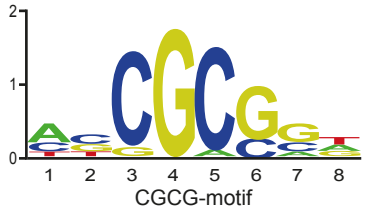

PTE 1.2
αBMI1
relative enrichment
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0
PL LF2 LF1 Core RF PTE ALX4 desert
transgenic
endogenous
PTE 1-6
αBMI1
relative enrichment
1.2
1
0.8
0.6
0.4
0.2
0
PL LF2 LF1 RF PTE ALX4 desert
transgenic
endogenous
PTE 6-8
αBMI1
relative enrichment
1.2
1
0.8
0.6
0.4
0.2
0
PL LF2 LF1 Core RF PTE ALX4 desert
transgenic
endogenous
PTE 1-8
αBMI1
relative enrichment
2
1.6
1.2
0.8
0.4
0
PL LF2 LF1 Core RF PTE ALX4 desert
transgenic
endogenous
PTE 6-11
αBMI1
relative enrichment
1.2
1
0.8
0.6
0.4
0.2
0
PL LF2 LF1 Core RF PTE ALX4 desert
transgenic
endogenous
PTE 1.2Δ
αBMI1
relative enrichment
1.2
1
0.8
0.6
0.4
0.2
0
PL LF2 LF1 RF PTE ALX4 desert
transgenic
endogenous

Figure S8

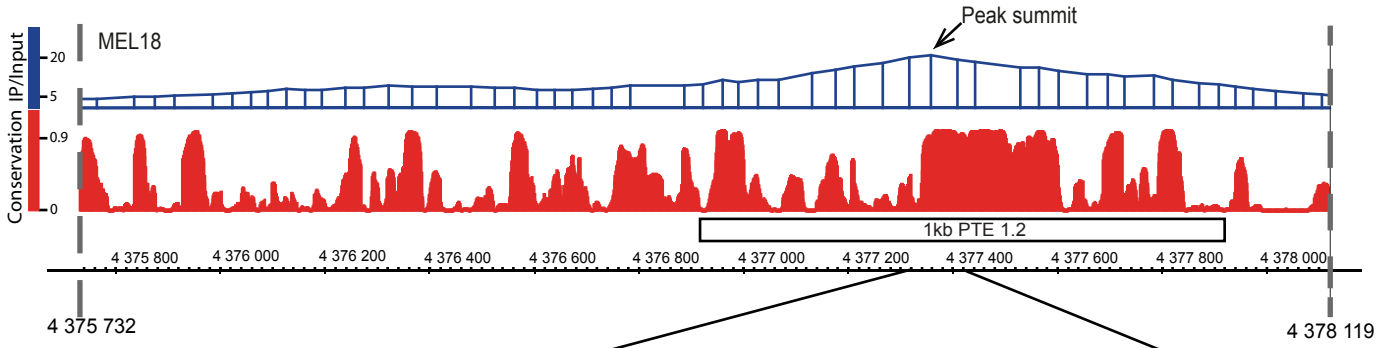
A



B



C




```
Homo sapiens                CCCAGAAGAAACAAAACCAAGTCTCTTGGGTCAGACCAGCAATCATAAAACGAAAAAGAAATAAATAATAAGTTAAAAA-TAGAGGGAATGTGTTAAAAAGAAGGAAAAA
Pan paniscus                CCCAGAAGAAACAAAACCAAGTCTCTTGGGTCAGACCAGCAATCATAAAACGAAAAAGAAATAAATAATAAGTTAAAAA-TAGAGGGAATGTGTTAAAAAGAAGGAAAAA
Gorilla gorilla gorilla     CCCAGAAGAAACAAAACCAAGTCTCTTGGGTCAGACCAGCAATCATAAAACGAAAAAGAAATAAATAATAAGTTAAAAA-TAGAGGGAATGTGTTAAAAAGAAGGAAAAA
Pongo abelii                CCCAGAAGAAACAAAACCAAGTCTCTTGGGTCAGACCAGCAATCATAAAACGAAAAAGAAATAAATAATAAGTTAAAAA-TAGAGGGAATGTGTTAAAAAGAAGGAAAAA
Nomascus leucogenys         CCCAGAAGAAACAAAACCAAGTCTCTTGGGTCAGACCAGCAATCATAAAACGAAAAAGAAATAAATAATAAGTTAAAAA-TAGAGGGAATCTGTTAAAAAGAAGGAAAAA
Macaca mulatta              CCCAGAAGAAACAAAACCAAGTCTCTTGGGTCAGAGCAGCAATCATAAAACGAAAAAGAAATAAATAATAAGTTAAAAA-TAGAGGGAATGTGTTAAAAAGAAGGAAAAA
Papio anubis                CCCAGAAGAAACAAAACCAAGTCTCTTGGGTCAGAGCAGCAATCATAAAACGAAAAAGAAATAAATAATAAGTTAAAAA-TAGAGGGAATGTGTTAAAAAGAAGGAAAAA
Saimiri boliviensis boliviensis CCCAGAAGAAACAAAACCAAGTCTCTTGTGTCAGACCAGCAATCATAAAACGAAAAAGAAATAAATAATAAGTTAAAAA-TAGAAGGAATGTGTTAAAAAGAAGGAAAAA
Callithrix jacchus          CCCAGAAGAAACAAAACCAAGTCTCTTGGGTCAGACCAGCAATCATAAAACGAAAAAGAAATAAATAATAAGTTAAAAA-TAGAAGGAATATGTTAAAAAGAAGGAAAAA
Otolemur garnettii          CCCAGAAGGAACAAAACAAAGTCTCTTGGGTCAGACCAGCAATCATTAAACGAAAAAAAAATAAATAGTAAGTTAAAAA-TAGAGAGAAAGTGTTAAAAAGAAGAAAGAA
Ailuropoda melanoleuca      TTCAAAAGAAACAAAACAGAGTGTCTTGGGTCAGATCAGCAACCATTAAACGAAAAAGAA-TAAATAATTAGTTAAAAA-TAGAGGGAATGCACTAAAACGAATGAAGAA
Canis lupus familiaris      TTCAAAAGAAACAAAACAGAGTGTCCTGGGTCAGATCAGCAACCATTAAACGAAAAAGAAATAAATATTAAGTTAAAAAATAAAGGAAATGTACTAAAACGAATGAAGAA
Bos taurus                  CCCAGAAGAAACAAAACAAAGTCTCTTGGGTCACAGCAGCAACCATTAAACGAAAAAGAAAGAAATAATGAGTTAAAAA-TAGAGGGAATGTGTTCAAACGAAGGAAGAA
Ovis aries                  CTCAGAAGAAACAAAACAAAGTCTCTTGGGTCACAGCAACAACCATTAAATGAAAAAGAAAGAAATAATGAGTTAAAAA-TAGAGGGAATGTGTTCAAACGAAGGAAGAA
Sus scrofa                  TCCAGAAGAAACAAAACCAAGTCTTTTGGGTCAGATCAGCAAACATTAAACGAAAAAGAAAGAAATAAAAAGTTAAAAA-GAGAGGGAATATGTTCAAACGAAGGAAGAA
Equus caballus              TTCAGAAGAAACAAAACAAAGTCTCCTGGGTCAAATCAGCAACTATTAAACGAAAAAGAAATAAATAATAAGCTGAAAA-TAGAGGGAACGTGTTAAAACGAAGGAAGAA
Loxodonta africana          ---AGAAGAAACAAAACAAAGTGTTTTGGCTCAGAGCAGCAACCATTAAACGAAAAAGAAATAAATAATAAGTTAAAAA-TAGAAGGAATGTGTTAAAACGAAGGAGGAA
Oryctolagus cuniculus       --CAGAAGACACAAAACAAAACCCCCTGGGTCAGAGTAGCCATCATTCAACGAAAAACAAATCAATAATGAGTTAAAAA-TAGAGAGGATGTGTTCAAACGAAGGAAGAA
Rattus norvegicus           CC-AGAAGTAACAAAACAAAGCTGCTTGGGTCAAACTAGCAATCATTAAACGAAAAAGAACTAAATAGTAAGCTAAAAA-TAGACA--ATGTGTTCAAATGAAGCAAGAA
Cricetulus griseus          CCCAGAAGAAACAAAACAAAGCTGCTTGGGTAAGACTAGCCATCATTAAACGAAAAAGAACTAAATAGTAAGTTAAAAA-TAGAGAGAATGTGCTCAAGCGAAGCAAGAA
Mus musculus                CC-AGAAGTCACAAAACAAAGCTGCTTGGATCAGACTAGCAATCATTAAACGAAAAAGAACTAAATAATAAGTTAAAAA-TAGAGA--ATGTGTTCAAAGGAAGCAAGAA
```

Figure S9

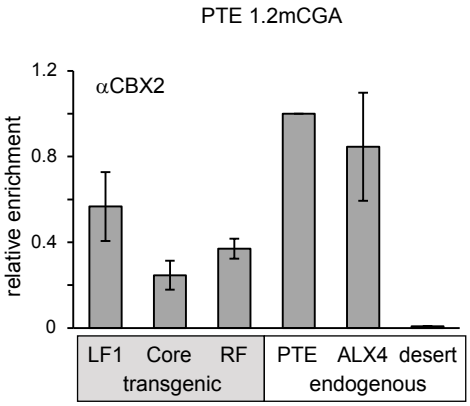
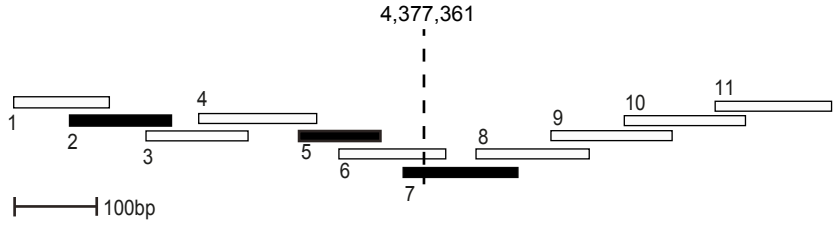

PTE 1.2mCGA
αCBX2
relative enrichment
1.2
0.8
0.4
0
LF1
Core
RF
transgenic
PTE
ALX4
desert
endogenous

Figure S10

A



B

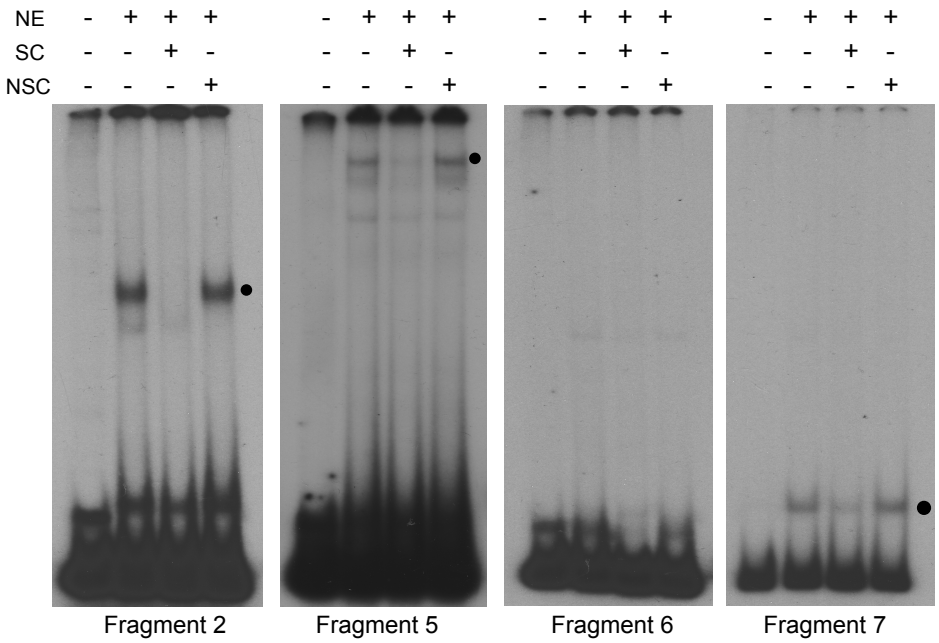


Figure S11

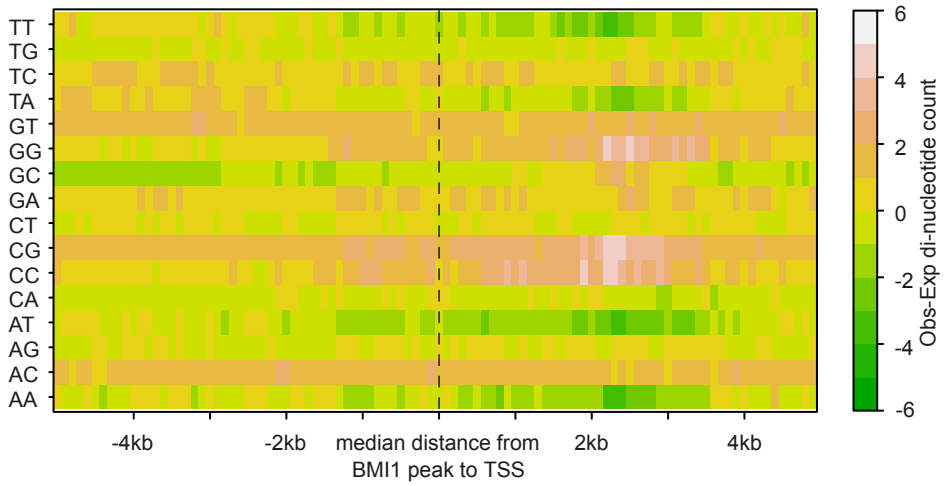

TT
TG
TC
TA
GT
GG
GC
GA
CT
CG
CC
CA
AT
AG
AC
AA
-4kb
-2kb
median distance from
BMI1 peak to TSS
2kb
4kb
6
4
2
0
-2
-4
-6
Obs-Exp di-nucleotide count

Figure S12

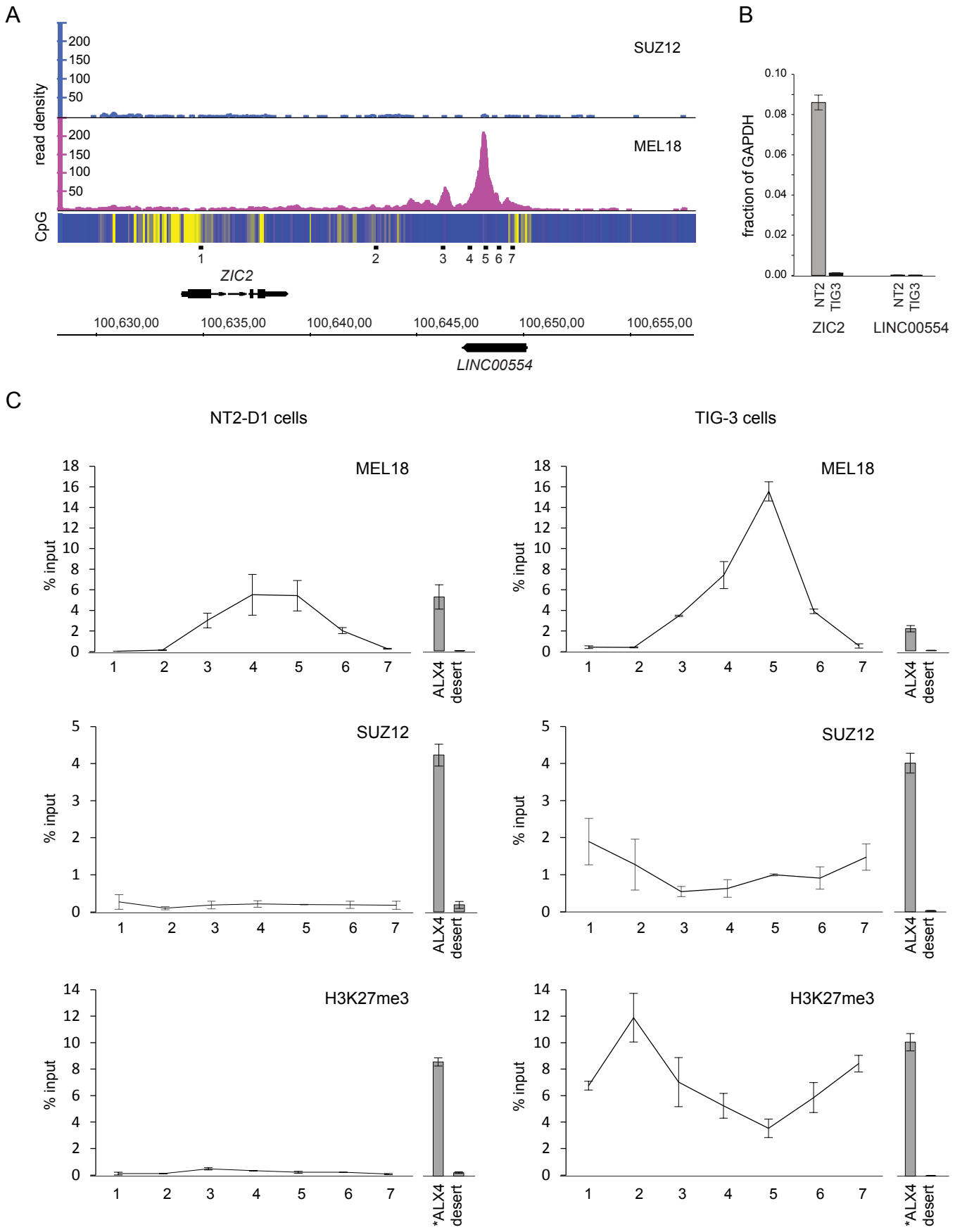

A
SUZ12
MEL18
read density
CpG
ZIC2
LINC00554
100,630,00
100,635,00
100,640,00
100,645,00
100,650,00
100,655,00
B
fraction of GAPDH
NT2
TIG3
ZIC2
LINC00554
C
NT2-D1 cells
TIG-3 cells
MEL18
SUZ12
H3K27me3
% input
ALX4
desert
*ALX4

Table S1. Screening for MEL18 sites with weak binding of SUZ12

Coordinates are in genomic release Hg19

Chromosome	Mel18/Su(z)12	Mel18 peak Center	Mel18 peak Length	Mel18 signal	Suz12 signal	Distance to closest TSS	Strand	Closest gene
chr13	53.05882353	100648150	2750	180.4	3.4	14125	+	ZIC2
chr12	12.4	4377800	3100	111.6	9	5101	+	CCND2
chr4	12.34090909	94751750	3750	108.6	8.8	1673	+	ATOH1
chr7	11.13207547	20833050	7750	118	10.6	6542	-	SP8
chr7	7.704761905	26897200	3050	161.8	21	7141	-	SKAP2
chr1	6.715686275	51442250	6450	137	20.4	6609	+	CDKN2C
chr2	6.357142857	74741200	1400	106.8	16.8	395	+	TLX2
chr1	6.107843137	51436050	4700	124.6	20.4	409	+	CDKN2C
chr1	5.952380952	63791950	3450	100	16.8	3221	+	FOXD3
chr1	5.25751073	110611450	2850	245	46.6	1872	-	ALX3
chr1	5.082508251	47697250	4600	308	60.6	1807	-	TAL1
chr19	4.94214876	6589000	4400	119.6	24.2	2163	-	CD70
chr2	4.617886179	66665550	9950	113.6	24.6	3019	+	MEIS1
chr3	4.586206897	169378000	4350	159.6	34.8	3563	-	MECOM
chr16	4.28125	73099850	6200	109.6	25.6	7316	-	ZFHX3
chr1	3.703030303	18960000	4700	122.2	33	2501	+	PAX7
chr1	3.553488372	75598700	7600	152.8	43	1866	+	LHX8
chr10	3.335	101288000	4750	133.4	40	4689	+	NKX2-3
chr15	3.27173913	60294300	13700	120.4	36.8	2120	+	FOXB1
chr2	3.200757576	63275950	7750	169	52.8	294	-	LOC100132215
chr2	3.192307692	176997250	3300	116.2	36.4	2783	+	HOXD8
chr7	3.186666667	35295150	14300	143.4	45	1439	-	TBX20
chr3	2.945848375	128210950	3150	163.2	55.4	1080	-	GATA2
chr3	2.873563218	169384650	6600	100	34.8	3087	-	MECOM
chr4	2.808290155	111560650	10400	108.4	38.6	2629	-	PITX2
chr3	2.7890625	147130500	10000	142.8	51.2	3320	+	ZIC1
chr18	2.743478261	19747750	4050	126.2	46	1653	+	GATA6
chr7	2.715053763	96633450	10450	101	37.2	1839	+	DLX6
chr2	2.626198083	45157700	21100	164.4	62.6	11336	+	SIX3
chr13	2.574257426	79171750	9300	104	40.4	5945	-	POU4F1
chr1	2.52398524	47909250	3650	136.8	54.2	7562	+	FOXD2
chr9	2.465116279	79636450	2650	106	43	1880	+	FOXB2
chr1	2.457142857	44874650	5100	103.2	42	3691	+	RNF220
chr8	2.454954955	25906700	5900	109	44.4	4060	-	EBF2
chr15	2.433333333	96885300	20050	102.2	42	8732	+	NR2F2
chr3	2.384955752	157813700	4550	107.8	45.2	10252	-	SHOX2
chr3	2.312	147140850	8850	115.6	50	13670	+	ZIC1
chr7	2.293548387	1267700	7650	284.4	124	4953	+	UNCX
chr8	2.231454006	11558050	10350	150.4	67.4	3666	+	GATA4
chr7	2.176691729	121947450	18450	115.8	53.2	2885	-	FEZF1
chr2	2.168831169	105474500	10850	167	77	2532	+	POU3F3
chr3	2.139240506	27766350	5200	101.4	47.4	2565	-	EOMES
chr1	2.123563218	46956500	2750	147.8	69.6	16167	+	DMBX1
chr2	2.023255814	223154200	22150	139.2	68.8	8665	+	CCDC140

chr1	2.003194888	165323150	6550	125.4	62.6	2328	-	LMX1A
chr2	1.988636364	63287350	9100	105	52.8	9414	+	OTX1
chr11	1.988095238	44327100	2900	167	84	4616	-	ALX4
chr4	1.934328358	4863100	3450	129.6	67	1709	+	MSX1
chr2	1.632596685	172958050	19400	118.2	72.4	7843	+	DLX1
chr7	1.591549296	27146500	14150	180.8	113.6	4106	-	HOXA2
chr15	1.563888889	76632500	12800	112.6	72	3354	+	ISL2
chr2	1.549618321	175204150	19350	121.8	78.6	4330	+	SP9
chr7	1.49339207	27155050	4100	135.6	90.8	4164	-	HOXA3
chr13	1.455801105	112724300	6800	105.4	72.4	2388	+	SOX1
chr2	1.303896104	105468850	5800	100.4	77	916	-	LOC100506421
chr2	1.287917738	73146400	4050	100.2	77.8	1797	+	EMX1
chr1	1.238721805	50890250	9900	131.8	106.4	1131	-	DMRTA2
chr7	1.201663202	27223650	7550	115.6	96.2	1185	-	HOXA11
chr7	1.193661972	27153900	5350	135.6	113.6	5314	-	HOXA3
chr12	1.169133192	54423750	6350	110.6	94.6	1557	+	HOXC6
chr7	1.072765073	27179550	9100	103.2	96.2	432	+	HOXA-AS3
chr7	1.064516129	27233600	4500	112.2	105.4	6125	-	HOXA13
chr21	1.018302829	38083100	10600	122.4	120.2	11110	+	SIM2
chr7	0.988709677	1279350	10700	122.6	124	6697	+	UNCX
chr21	0.945091514	38064900	12250	113.6	120.2	7090	+	SIM2
chr7	0.881756757	27245750	7700	104.4	118.4	5711	+	HOTTIP

Table S2. The sequence and name of PCR primers for cloning.

Primer name	Sequence
PRE Spacer Forward (126277048)	GATCACGAGACTAGCCCCTGGAGATGTGTCCATG
PRE Spacer Reverse (126279445)	AAGCCAACTCTCAGCCTGGTGAAACTGTGAGATAATG
PRE (Cyclin D2) F 4245993	GATCACGAGACTAGCCAGAATATCTTCCTTGACACC
PRE (Cyclin D2) R 4248379	AAGCCAACTCTCAGCTTCATGCGCGACATATTCCTCC
Mut. PRE (1.1) R'	CCGCGGAAATGAGCTATAATTAATAAAAATGT
Mut. PRE Seq. (1.2) 1 F'	AGCTCATTTCCGCGGCCAGAAGAAACAAAA
PRE EcoRV F	CAAAATTCATCCGATATCGCT
Plas -CMV eGFP R	GCCACTGTGCTGGATTTCATGCGCGACATATTCCT
pLenti-CMV eGFP + pmlI F	GATCACGAGACTAGCGCTGAGAGTTGGCTTCACGTGCTAGACCCAGCTTTC
pLenti-CMV eGFP + pmlI R	GAAAGCTGGGTCTAGCACGTGAAGCCAACTCTCAGCGCTAGTCTCGTGATC
Core 1Kb PRE F	GATCACGAGACTAGCGCTTCAATTGGCTGCAACTG
Core 1Kb PRE R	AAGCCAACTCTCAGCAGAATCTTCTAAACTGCAAGGG
Cloning Core PRE 6 F	GATCACGAGACTAGCGCAAGCACTCTTAAATACAGTT
Cloning Core PRE 6 R	AAGCCAACTCTCAGCTATGATTGCTGGTCTGACCC
Cloning Core PRE 8 R	AAGCCAACTCTCAGCGATTTGAAGTGGAATTATTTGTT
Cloning 6R - 6 to 8	GTGTCTGAAAATGTGTTTTTGGTGCCTGATTTGATGG
Cloning 8F - 6 to 8	CACATTTTCAGACACCCCTA
CGA mut AAA Cloning R	ATCATAAAAAAAAAAAGAAATAAATAAT
CGA mut AAA Cloning F	TTTTTTTTTATGATTGCTGGTCTGA
ms PTE Cloning 1 F	GATCACGAGACTAGCTAGGGGTCCGAGCTAGTGT
ms PTE Cloning 1 R	TCACCCACGGATTTTGCTGT
ms PTE Cloning 2 F	AAAATCCGTGGGTGAAATGTTC
ms PTE Cloning 2.1 R	AAGCCAACTCTCAGCGGGAAAGTTAAAGCGCGTCC

The nucleotides in red are the 15nt homology sequence required for In-Fusion cloning.
Underlined nucleotides indicate introduced restriction sites.

Table S3. The correspondence between PCR primers and transgenic constructs.

construct	Primer name
gene desert	PRE Spacer Forward (126277048)
	PRE Spacer Reverse (126279445)
PTE 2.4	PRE (Cyclin D2) F 4245993
	Mut. PRE (1.1) R'
	Mut. PRE Seq. (1.2) 1 F'
	PRE (Cyclin D2) R 4248379
PTE 1.1	PRE (Cyclin D2) F 4245993
	Mut. PRE (1.1) R'
PTE 1.2	Core 1Kb PRE F
	Core 1Kb PRE R
PTE 1.3	Core 1Kb PRE F
	PRE (Cyclin D2) R 4248379
PTE 1-6	Core 1Kb PRE F
	Cloning Core PRE 6 R
PTE 1-8	Core 1Kb PRE F
	Cloning Core PRE 8 R
PTE 6-8	Cloning Core PRE 6 F
	Cloning Core PRE 8 R
PTE 6-11	Cloning Core PRE 6 F
	Core 1Kb PRE R
PTE 1.2Δ	Core 1Kb PRE F
	Cloning 6R - 6 to 8
	Cloning 8F - 6 to 8
	Core 1Kb PRE R
PTE 1.2 CGA mut	Core 1Kb PRE F
	CGA mut AAA Cloning R
	CGA mut AAA Cloning F
	Core 1Kb PRE R
Mouse PTE	ms PTE Cloning 1 F
	ms PTE Cloning 1 R
	ms PTE Cloning 2 F
	ms PTE Cloning 2.1 R

Table S4. List of antibodies used for ChIP analysis, western blot and immunofluorescence.

Antibody	Host	Suppliers	Cat. No.	ChIP concentration	Western blot dilutions	IF dilutions
H3	Rabbit	Abcam	ab1791		1/100000	1/100
H2AK119Ub	Rabbit	Cell Signaling	8240	1:1000		
H3K27me3	Mouse	Abcam	ab 6002	0.3 µg		
H3K27me3	Rabbit	Millipore	07-449		1/4000	
H3K27me3	Mouse	Diagenode	MAb-181-050			1/100
BMI1	Mouse	Millipore	05-637	5 µg	1/500	
MEL18	Rabbit	Abcam	ab 16651 (discontinued 2015)	5 µg		
MEL18	Mouse	Abcam	ab77255	1µg	1/1000	
SUZ12	Rabbit	Cell Signaling	3737	1:100	1/2000	1/100
CBX2	Rabbit	Abcam	ab 80044	5 µg	1/500	
RING2	Mouse	Gift from H. Koseki		40ul	1/500	
Cbx7	Rabbit	Abcam	ab21873	5ug		
Kdm2B	Rabbit	Gift from R. Klose		5ul		
CTCF	Rabbit	Cell Signaling	2899	2µl		

Table S5. The PCR primers used for ChIP-qPCR, RT-qPCR and genotyping.

Primer name	Sequence
PRE 1.1 internal F	CATGCATAGGTGCAGACTTTGT
PRE 1.1 internal R	GTGTGCTAGACAATCCAAGCC
Seq. Homolg. BmiI PRE 1 F	AATTATAGCTCATTTTTTCCCC
Seq. Homolg. BmiI PRE 1 R	TTCCTTCTTTTTAACACATTCC
PRE qPCR 1.2 F	ATGTAATAGCCATCGTATTAACC
Genomic PRE qPCR 1.2 R	TTGACTTTTACACCCGGAGTTT
CCND2 2.2 F	TCCGCTCAAAGGTGGCTTAG
CCND2 2.2 R	TTAGTCACGTTGGGCTTCCC
CCND2 promoter 2 F	GGAGGAGAGCTAACTGCCC
CCND2 promoter 2 R	CGGTCCTCCCCTTAAAACTG
CCND2 F	TTTAAGTTTGCCATGTACCCAC
CCND2 R	TCAGGGCATCACAAGTGAGC
Mut. PRE qPCR 1 F'	AATTATAGCTCATTTCCGCGG
Mut. PRE qPCR 1 R'	TTCCTTCTTTTTAACAGCTGGG
pLenti Plasmid qPCR 1.2 R'	GGGAGTAAAGCCAACTCTCA
Spacer qPCR F	AGCTAACAGTGACAATCACCTT
Spacer qPCR R	AGGAGGAGATATAGTTGCAAAG
Spacer qPCR 1.2 F'	CCCGTGCTTTCAGAAGTGTGT
Genomic Spacer qPCR 1.2 R'	CCAGCCTGGCAAAGCTGTT
ALX4-PRE1.1	CTAACTAACAGTGTACACATAC
ALX4-PRE1.2	CAATCGACTTGCATGATCTTG
ALX4 H3K27me3 2.1 F	CAACCCCGAGACAGGAACAA
ALX4 H3K27me3 2.2 R	AGGCGAGGTGATGCTCAAAA
pLenti Plasmid qPCR 1.1 F'	TTTTATCGATCACGAGACTAGC
plenti - CMV CpG qPCR F	CGCTGAGAGTTGGCTTCAC
CpG qPCR R	GAAATCCCGATCACTGGTCG
CpG qPCR F	CGCGGAGAAGGCGGGACT
plenti - CMV CpG qPCR R2	AAGAAAGCTGGGTCTAGCAC
Plas 5'LTR prox F	GAACAGGGACCTGAAAGCGA
Plas 5'LTR prox R	TTTGGCGTACTCACCAGTCG
Plasmid Prox 5' PRE F	TATCGTTTCAGACCCACCTCC
Plasmid Prox 5' PRE R	CCGTCGAGATCCGTTCACTAA
Core PRE 2 R	GTAAAATGTCAAGACTATCCTC
Core PRE 6 R	TATGATTGCTGGTCTGACCC
Core PRE 6 F	GCAAGCACTCTTAAATACAGTT
Core PRE 8 F	GGGAATGTGTTAAAAAGAAGGA
Core PRE 11 F	GTGGGTTTAGAAAATAATGACAC
CGA mut AAA Cloning R	ATCATAAAAAAAAAAAGAAATAAATAAT
msPRE 1.1.2 F	AAGCGTTGATCTGGACCATA
msPRE 1.1.2 R	CTCGGGGGCATATTTGATGA
msPRE 1.1.1 F	TCTCAAGTGCCTCAGATCAC
msPRE 1.1.1 R	ACTACCATAACTAGCACACAC
msPRE oyh1 F	AACTTTTAGCCTTTCCAAACCC
msPRE oyh1 R	GCTTCAAAATGCAAAACTCAACC
msPRE F	GATTATGAGCAACACTTTCTTGC
msPRE R	CCCCTGTCCCAGAAGTCAC
msPRE 2.1 F	CCTCGTGGCAGAGGATAAGC

msPRE 2.1 R	GGAAGACGGGAAGTTTGGGT
msPRE 2.2 F	GGCGTCAGAGCTGTACTGAG
msPRE 2.2 R	CTTCACTGGAGATGCGGAGG
msPRE 2.3v2 F	CTGTTGGTAGAGGGGGACAC
msPRE 2.3v2 R	CCCCTTCCTCCCTCCTTTAA
mBActin F	GCTACAGCTTCACCACCACA
mBActin R	TCTCCAGGGAGGAAGAGGAT
ZIC2-exp1.1	ATGTGCTCAACGGGCAGATG
ZIC2-exp1.2	GCCGTACTGGTTGTGGAGTTG
LINC00554-4.1	CGCTGCTCCGATATTGATCC
LINC00554-4.2	CCCCTACCCTGGCTCATTAG
LINC00554-2.1	CACTTCTTCCTTTGCTGGGT
LINC00554-2.2	TGTGTGTCTGAAGTTGAACTACC
LINC00554-5.1	ATCTCACATAGGGTGCCATTG
LINC00554-5.2	ACAATGTTGGAGGCACAAGAG
LINC00554-1.1	CTTCCTTTCCTAGTGCCAGACAAC
LINC00554-1.2	GGGAAATGCGCGAAACAAACAA
LINC00554-6.1	AGAGCGCGTTGAGCTATTTCT
LINC00554-6.2	CTGGCGTTTGTTGGCTGTTG
LINC00554-exp1.1	TCTGGAAGTGAAGCAGGACGA
LINC00554-exp1.2	GGTGCGACCTAGGTACCCTTTAT

Table S6. The correspondence between combination of PCR primers and amplicons.

Amplicon	Primer name
human CCND2 (1)	PRE 1.1 internal F
	PRE 1.1 internal R
human CCND2 (2)	Seq. Homolg. BmiI PRE 1 F
	Seq. Homolg. BmiI PRE 1 R
human CCND2 (3)	PRE qPCR 1.2 F
	Genomic PRE qPCR 1.2 R
human CCND2 (4)	CCND2 2.2 F
	CCND2 2.2 R
human CCND2 (5)	CCND2 promoter 2 F
	CCND2 promoter 2 R
human CCND2 (6)	CCND2 F
	CCND2 R
Transgenic PTE	Mut. PRE qPCR 1 F'
	Seq. Homolg. BmiI PRE 1 R
Desert	Spacer qPCR F
	Spacer qPCR R
ALX4	ALX4-PRE1.1
	ALX4-PRE1.2
*ALX4	ALX4 H3K27me3 2.1 F
	ALX4 H3K27me3 2.2 R
PL	Plas 5'LTR prox F
	Plas 5'LTR prox R
LF2	Plasmid Prox 5' PRE F
	Plasmid Prox 5' PRE R
RF LF1	pLenti Plasmid qPCR 1.1 F'
	plenti - CMV CpG qPCR R2
LF1 (PTE1.2/PTE1.2Δ)	pLenti Plasmid qPCR 1.1 F'
	Core PRE 2 R
LF1 (PTE1-6/PTE1-8)	pLenti Plasmid qPCR 1.1 F'
	Core PRE 2 R
LF1 (PTE6-8/PTE6-11)	pLenti Plasmid qPCR 1.1 F'
	Core PRE 6 R
RF (PTE1.2-CMV/PTE1.2Δ/PTE6-11)	Core PRE 11 F
	plenti - CMV CpG qPCR R2
RF (PTE1-6)	Core PRE 6 F
	plenti - CMV CpG qPCR R2
RF (PTE1-8/PTE6-8)	Core PRE 8 F
	plenti - CMV CpG qPCR R2
Transgenic PTE mCGA	Mut. PRE qPCR 1 F'
	CGA mut AAA Cloning R
mouse Ccnd2 (1)	msPRE 1.1.2 F
	msPRE 1.1.2 R
mouse Ccnd2 (2)	msPRE 1.1.1 F
	msPRE 1.1.1 R
mouse Ccnd2 (3)	msPRE oyh1 F
	msPRE oyh1 R

mouse Ccnd2 (4)	msPRE F
	msPRE R
mouse Ccnd2 (5)	msPRE 2.1 F
	msPRE 2.1 R
mouse Ccnd2 (6)	msPRE 2.2 F
	msPRE 2.2 R
mouse Ccnd2 (7)	msPRE 2.3v2 F
	msPRE 2.3v2 R
β-actin	mBActin F
	mBActin R
ZIC2 (1)	ZIC2-exp1.1
	ZIC2-exp1.2
ZIC2 (2)	LINC00554-4.1
	LINC00554-4.2
ZIC2 (3)	LINC00554-2.1
	LINC00554-2.2
ZIC2 (4)	LINC00554-5.1
	LINC00554-5.2
ZIC2 (5)	LINC00554-1.1
	LINC00554-1.2
ZIC2 (6)	LINC00554-6.1
	LINC00554-6.2
ZIC2 (7)	LINC00554-exp1.1
	LINC00554-exp1.2

Table S7. The PCR primers used for RT-qPCR analysis.

Primer name	Amplicon	Sequence
GAPDH-2.1	GAPDH	AAAACCTGCCAAATATGATG
GAPDH-2.2		TGTTGAAGTCAGAGGAGACC
Cbx2 F iso.1	Cbx2	GAGCTCCATCGTGCACTACA
Cbx2 R iso.1		CCTTTCTGCGTCCTCACCTT
Cbx4 F	Cbx4	TAGTACTTGCCGCTCTTGCC
Cbx4 R		GGGCATCAGTACGAGCTCAA
Cbx6 F	Cbx6	GACGCTCTCGCTGAACTTCT
Cbx6 R		ACCGCATCATCCTGAACCTG
Cbx7 F	Cbx7	AGGTACTTGTGGGCCTTTCG
Cbx7 R		CTCTGCTTCTCCCTGACGTG
Cbx8 F	Cbx8	GCCAGCTGGATCACACTGTG
Cbx8 R		CCTCACAGAGGCCCTTAGG
NSPC1/PCGF1 F	PCGF1	GCAAGCCAGGCACCAGCTT
NSPC1/PCGF1 R		ACCTCCAAACTAGCAAGTACT
Mel18/PCGF2 F	Mel18	TGGCACCTTCGTAGAATTCG
Mel18/PCGF2 R		CAACGGCTCCAATGAGGAC
PCGF3 F	PCGF3	GAGCATCTGCCTGGAGTGTA
PCGF3 R		CGTTAAAGGATGAAAGGTTGA
BmiI/PCGF4 F	BmiI	CCAATGGCTCTAATGAAGATAG
BmiI/PCGF4 R		GTTCTGGTCAAAGAATTCAATG
PCGF5 F	PCGF5	TGATACTTCAAAAGCTGACAAAC
PCGF5 R		TACATTGTCCCCTGATTGCC
RYBP F	RYBP	TGGTATTTTTCTTGGTAACACTAGG
RYBP R		GCTGGTGGCACAACAAGTG
huRING1_ex3_F	RING1	CGAAAGAAGCTGGTGTCCAA
huRING1_ex3_R		GTCTTGATGGGCCTCGTATT
Ring1B/RNF2 F	RING2	CTGCAGCGAGGCAAGAAACA
Ring1B/RNF2 R		TGGTCCGTTTGTTACTAGGG
KDM2B F	KDM2B	ACACCCTCCAGGAGTGCCT
KDM2B R		TCTCACTGAGTTTGAACTGAAG
(2) YAF2 F	YAF2	AATGAGATTCTCCATTCAATGA
(2) YAF2 R		CCTGCATCTAGTGCTGCTTC
CCND2 F	CCND2	TTTAAGTTTGCCATGTACCCAC
CCND2 R		TCAGGGCATCACAAGTGAGC
ZIC2-exp1.1	ZIC2 (1)	ATGTGCTCAACGGGCAGATG
ZIC2-exp1.2		GCCGTACTGGTTGTGGAGTTG
LINC00554-6.1	ZIC2 (6)	AGAGCGCGTTGAGCTATTTCT
LINC00554-6.2		CTGGCGTTTGTTGGCTGTTG

Table S8. PCR primers to produce and label the fragments for EMSA

Primer name	Fragment	Sequence
Core PRE 1 F	1	GCTTCAATTGGCTGCAACTG
Core PRE 1 R		AATTTTGTCCAAATTCAAACATTC
Core PRE 2 F	2	GGAAAGTGGACATTTTTAAGGG
Core PRE 2 R		GTAAAATGTCAAGACTATCCTC
Core PRE 3 F	3	ATTCTGGTGAGGATAGTCTTG
Core PRE 3 R		GCAAAATGTCCTTTAAATTTAG
Core PRE 4 F	4	TACTTAACTGTCTGAAATGAAGG
Core PRE 4 R		TTGATGGACACTTGGTTTGTCT
Core PRE 5 F	5	TCTACCTTTATGCAGACAAACC
Core PRE 5 R		GTAATTACTGTCATGATTAGCC
Core PRE 6 F	6	GCAAGCACTCTTAAATACAGTT
Core PRE 6 R		TATGATTGCTGGTCTGACCC
Core PRE 7 F	7	TCCCCCAGAAGAAACAAAACC
Core PRE 7 R		GGTGACACAGTTTGGTTATGAA
Core PRE 8 F	8	GGGAATGTGTTAAAAAGAAGGA
Core PRE 8 R		GATTTGAAGTGGAATTATTTGTT
Core PRE 9 F	9	GACAGAAAGCTTCTCTCCTAC
Core PRE 9 R		TGTTTGGGAGAGAAATACTGTG
Core PRE 10 F	10	CTTTACCTCAAAGAGGTAAAAG
Core PRE 10 R		GGGGTTATAGCTGAAATGTG
Core PRE 11 F	11	GTGGGTTTAGAAAATAATGACAC
Core PRE 11 R		AGAATCTTCTAAACTGCAAGGG