An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment

This article has had an erroneous label (H3K27me3) removed from Figure 4C since Advance Online Publication.

Magnus D Lynch[1], Andrew JH Smith[1,2], Marco De Gobbi[1], Maria Flenley[1], Jim R Hughes[1], Douglas Vernimmen[1], Helena Ayyub[1], Jacqueline A Sharpe[1], Jacqueline A Sloane-Stanley[1], Linda Sutherland[3], Stephen Meek[3], Tom Burdon[3], Richard J Gibbons[1], David Garrick[1,4] and Douglas R Higgs[1,4,*

[1]MRC Haematology Unit, Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK, [2]Institute for Stem Cell Research, University of Edinburgh, Edinburgh, UK and [3]The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Scotland, UK

The role of DNA sequence in determining chromatin state is incompletely understood. We have previously demonstrated that large chromosomal segments from human cells recapitulate their native chromatin state in mouse cells, but the relative contribution of local sequences versus their genomic context remains unknown. In this study, we compare orthologous chromosomal regions for which the human locus establishes prominent sites of Polycomb complex recruitment in pluripotent stem cells, whereas the corresponding mouse locus does not. Using recombination-mediated cassette exchange at the mouse locus, we establish the primacy of local sequences in the encoding of chromatin state. We show that the signal for chromatin bivalency is redundantly encoded across a bivalent domain and that this reflects competition between Polycomb complex recruitment and transcriptional activation. Furthermore, our results suggest that a high density of unmethylated CpG dinucleotides is sufficient for vertebrate Polycomb recruitment. This model is supported by analysis of DNA methyltransferase-deficient embryonic stem cells.

*Corresponding author. MRC Molecular Haematology Unit, Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, Headington, Oxford OX3 9DS, UK. Tel.: + 44 01865 222393; Fax: + 44 01865 222424; E-mail: doug.higgs@imm.ox.ac.uk
[4]Senior authorship shared equally

Introduction

The mechanism by which gene expression and the associated chromatin states are encoded in primary DNA sequence is a fundamental question in molecular biology. Orthologous chromosomal regions from closely related species can exhibit differing patterns of transcription factor binding, histone modifications and transcriptional output in the same cell type. Many factors could play a role in this process including changes in DNA sequence, positioning within the nucleus, alterations in the epigenetic machinery and in the levels or modifications of the transcription factors.

Transfer of large chromosomal segments into the genome of another species has provided evidence for the primacy of DNA sequence in the encoding of gene expression and chromatin states. A comparison of transcription factor binding and H3K4me3 modification in hepatocytes from an aneuploid mouse carrying human chromosome 21 found that the human chromosome adopted the human rather than the mouse pattern of chromatin modifications (Wilson *et al*, 2008). Furthermore, we have previously reported that, following the replacement of ∼ 87 kilobases (kb) of the mouse α globin locus with a corresponding 120-kb region from the human genome, the human sequence adopts its native chromatin state in erythroid cells (Wallace *et al*, 2007). Thus, the information required for species-specific regulatory differences appear to be encoded in *cis*-acting DNA sequences.

An important outstanding question is whether this code is local or global (Coller and Kruglyak, 2008). Interspecies differences in the chromatin landscape could either reflect alterations restricted to the site of chromatin modifications or alternatively a synergistic interaction between regulatory elements distributed throughout the locus. This question is particularly pertinent to the establishment of bivalent chromatin domains, which have been defined as promoters marked by both the active H3K4me3 modification (mediated by Trithorax Group (TrxG) proteins) and the repressive H3K27me3 modification (mediated by Polycomb Group (PcG) proteins) (Azuara *et al*, 2006; Bernstein *et al*, 2006). Sites of H3K4 methylation are conserved between orthologous locations in human and mouse genomes; however, the underlying DNA sequences are often no more conserved than background (Bernstein *et al*, 2005). This implies either that the DNA elements directing H3K4 methylation represent only a small fraction of the underlying sequence or that they are influenced by distal flanking sequences. Sequences responsible for vertebrate PcG recruitment and modification by H3K27me3 are also incompletely characterized (Margueron and Reinberg, 2011). In support of a local model, CpG islands (CGIs; Ku *et al*, 2008; Mendenhall *et al*, 2010) and transcription factor binding sites (Barna *et al*, 2002; Caretti *et al*, 2004; Kim *et al*,

2009) have been implicated. In support of a global model, long non-coding RNA transcription mediates PcG recruitment *in cis* to the inactive X chromosome (Zhao *et al*, 2008), the vertebrate Kcnq1 (Pandey *et al*, 2008) and *INK4A* loci (Yap *et al*, 2010) and in Arabidopsis (Swiezewski *et al*, 2009; Heo and Sung, 2011). In addition, PcG recruitment *in trans* has been reported for the *HOTAIR* long non-coding RNA (Rinn *et al*, 2007).

The α globin genes provide a useful model for investigating the role of primary DNA sequences in the templating of chromatin states. In pluripotent embryonic stem (ES) cells, the human α globin locus contains prominent sites of PcG recruitment and chromatin bivalency whereas the corresponding mouse locus does not (Garrick *et al*, 2008). We have undertaken a comparative analysis of these loci to investigate the sequences encoding the bivalent chromatin state. We first confirm, by comparing these two loci within the same nucleus in a humanized mouse model, that *cis*-acting sequences are responsible for differential recruitment of PcG proteins. Next, to determine which sequences are responsible, we have used recombinase-mediated cassette exchange (RMCE) to insert various fragments of the human locus into the orthologous position in the mouse locus. We find that a 4-kb region of human sequence establishes a novel bivalent chromatin domain. Analysis of non-overlapping fragments shows that chromatin state is redundantly encoded across this 4 kb region. Using this model we provide evidence that, consistent with a recent report (Mendenhall *et al*, 2010), chromatin bivalency reflects competition between PcG recruitment and transcriptional activation at CGIs. These analyses highlight a correlation between density of unmethylated CpG dinucleotides and PcG recruitment and a causative relationship is supported by the finding of multiple sites of *de-novo* Polycomb repressive complex 2 (PRC2) recruitment at CpG-rich regions that are methylated in wild-type ES cells and lose methylation in Dnmt3a/b$^{-/-}$ ES cells.

Results

Local sequences are sufficient to encode chromatin bivalency

The α globin genes are similarly arranged in the human and mouse genomes (Figure 1A); however, the human α globin locus is associated with prominent sites of PcG recruitment and chromatin bivalency in pluripotent cells whereas the corresponding mouse locus is not (Garrick *et al*, 2008; Figure 1A; Supplementary Figure S1A–D). To confirm that *cis*-acting sequences are responsible for these differences in chromatin state, we analysed ES cells from a mouse model in which the entire mouse α globin locus is replaced with a syntenic region (~120 kb) from the human locus (Wallace *et al*, 2007; Figure 1B). Since only one mouse chromosome is

modified, species-specific real-time qPCR probes can be used to compare the chromatin profiles at the mouse and human loci within the same nucleus. There is a clear difference in chromatin state between the human and the mouse α globin genes: Cbx7, a component of the Polycomb repressive complex 1 (PRC1), Ezh2, a component of the PRC2 and the H3K4me3 histone modification are templated specifically to the human but not to the mouse locus (Figure 1C–E). Similarly, although the level of H3K27me3 is slightly above background levels at the mouse locus, the level at the human gene is considerably greater (Figure 1F). Thus, these differences in chromatin state between mouse and human α globin genes must be determined by *cis*-acting sequences rather than *trans*-acting factors which differ between human and mouse.

Next, we determined which sequences encode chromatin bivalency. Using RMCE (Figure 2A), we introduced test fragments from the human α globin cluster into the orthologous region of the mouse locus, which contains duplicated copies of the mouse α and θ genes in two homology blocks. We replaced the downstream homology block with an RMCE cassette by homologous recombination. Subsequently, using RMCE we introduced a 4-kb fragment of the α globin locus, including the human *HBA2* gene and flanking sequences, into this cassette. A linked *Hprt* selective marker was excised using Flp recombinase so that the human fragments were flanked only by *frt* and *lox511* sites in the mouse locus (Figure 2A). We also introduced DNA fragments encoding *FERD3L*, another short gene associated with a bivalent chromatin state in human ES cells (Supplementary Figure S1E) and as a negative control *HBB*, which does not recruit PcG or the H3K4me3 mark in human ES cells (Supplementary Figure S1F).

Both *HBA2* and *FERD3L* recruit PRC2 (Ezh2) and are modified by H3K27me3 at this ectopic location. The *HBA2* gene was also modified by H3K4me3 consistent with a bivalently modified chromatin domain. The levels of H3K4me3 modification observed were lower for *FERD3L* but still above background (Figure 2B and C). There was no recruitment of PRC2 or modification of chromatin with H3K27me3 or H3K4me3 for the negative control *HBB* fragment (Figure 2D). Thus, all of the sequences required for chromatin bivalency appear to be encoded locally in these short DNA fragments. ChIP with an antibody to unmodified histone H3 confirms that these differences do not reflect alterations in histone occupancy (Supplementary Figure S1G–I).

### Chromatin state is redundantly encoded throughout a bivalent domain

Domains of chromatin bivalency may extend over many kilobases (Ku *et al*, 2008). This could either reflect recruitment to an initial site followed by spreading along the chromosome

Figure 1 Differences in chromatin state at human and mouse α globin loci in humanized mouse ES cells. (**A**) The human α globin cluster is located close to the telomere (16p13.3), whereas the mouse cluster lies at an interstitial chromosomal position (11qA4). The positions of α globin genes (*HBA1/Hba1* and *HBA2/Hba2*) and other globin genes within the loci are indicated. The α globin genes in both species are arranged in duplicated homology blocks. In the human locus, these blocks contain only the α globin genes whereas in the mouse, the θ globin genes are also present. Published ChIP-seq data illustrate differences in the epigenetic status at human versus mouse α globin genes in human and mouse ES cells. To facilitate analysis of duplicated homology blocks, reads were remapped with Bowtie permitting up to two copies of a sequence in the genome. Read count is normalized to reads mapped per 10 million. (**B**) For one copy of mouse chromosome 11, the illustrated region was replaced with the 120-kb orthologous region of human chromosome 16 including the α globin genes (Wallace *et al*, 2007) permitting comparison of human and mouse globin loci within the same nucleus of pluripotent cells. (**C–F**) ChIP was performed with antibodies to (**C**) Ezh2, (**D**) Cbx7, (**E**) H3K4me3 and (**F**) H3K27me3 in these transgenic mouse ES cells bearing one wild-type allele and one humanized allele. Enrichment was quantified as percentage of input DNA and is plotted according to the position of species-specific qPCR amplicons relative to the human or mouse α globin gene. Positive and negative control points are shown on the right side of each panel (from left to right: 5′ mouse β globin gene, mouse β globin exon1, mouse β actin, mouse Gata6). Data are the result of at least two biological replicates ± s.d.

or recruitment by multiple redundant sequence elements throughout the domain. Initially, we assessed the relative contributions of the promoter and gene body sequences to chromatin state. To address this, the *HBA2* 4-kb test fragment was divided into three subfragments, which were separately integrated into the same genomic locus using RMCE (Figure 3). Fragment (I) is the original 4-kb fragment.

Fragment (II) and Fragment (III) extend from the start of Fragment (I) to the transcriptional start site (TSS) and exon 2, respectively. Fragment (IV) extends from the TSS to the end of the original 4-kb fragment. The nucleotide composition of these various fragments and the *HBB* and *FERD3L* fragments is illustrated (Supplementary Figure S2A–F). For each fragment, three independently derived cell lines were analysed.
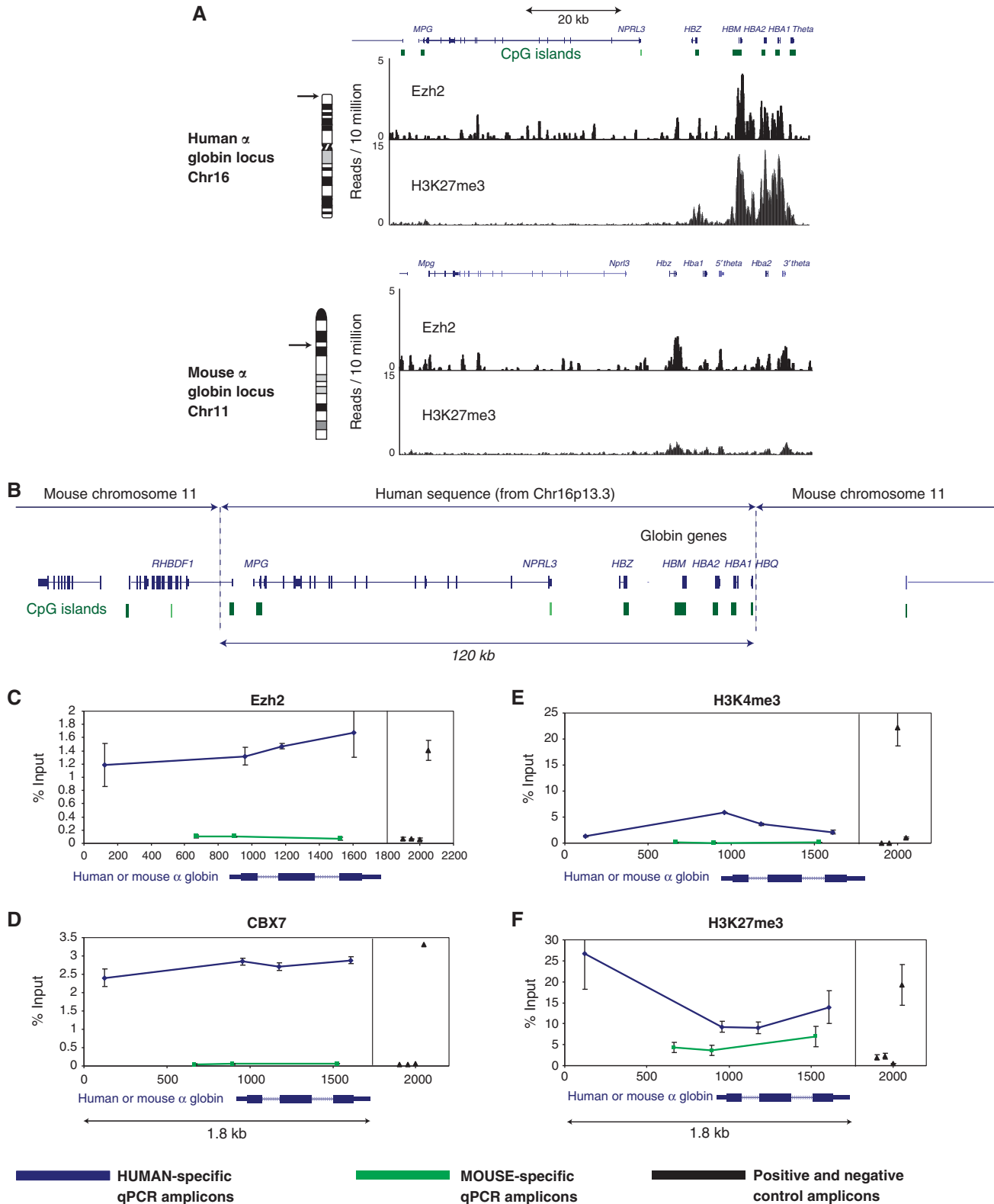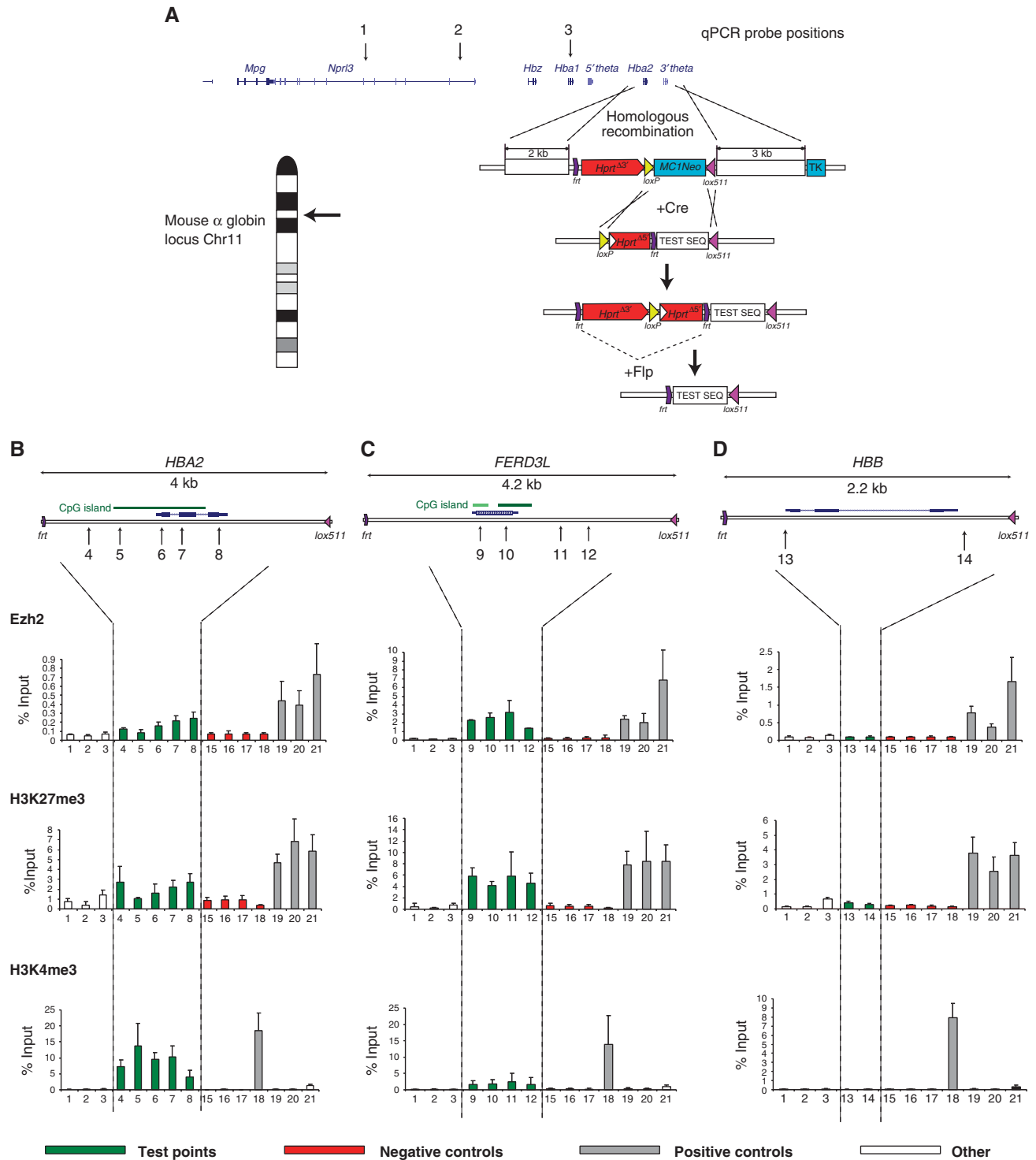
Figure 2 Establishment of a novel bivalent chromatin domain in the mouse α globin locus. (**A**) An RMCE cassette was targeted to the wild-type mouse α globin locus deleting the 3′ homology block containing the *Hba2* and *3′ theta* genes. The position of qPCR primers located within the mouse α globin locus upstream of the RMCE exchange site (1–2) and in exon 1 of the mouse α globin gene (3) are indicated. (**B–D**) DNA fragments encoding the human (**A**) *HBA2*, (**B**) *FERD3L* and (**C**) *HBB* genes were separately integrated into this locus using the RMCE system. Following Flp-mediated excision of the *Hprt* selective marker gene, ChIP was performed with antibodies to Ezh2, H3K4me3 and H3K27me3 and enrichment quantified with species-specific qPCR probes. The positions of qPCR probes within the tested fragments are indicated (numbered 4–14). Other numbered probes are control points in the mouse genome as follows: (15) 5′ β globin; (16) β globin exon 1; (17) 3′ β globin; (18) β actin promoter; (19) *HoxB7*; (20) *HoxC5* and (21) *Gata6* promoter. Test probes are shown in green along with mouse genomic negative control points (red) and genomic positive control points (grey). Other points are shown as unfilled bars. Enrichment was quantified as percentage of input DNA. Data are shown for three separately derived and analysed cell lines ± s.d.
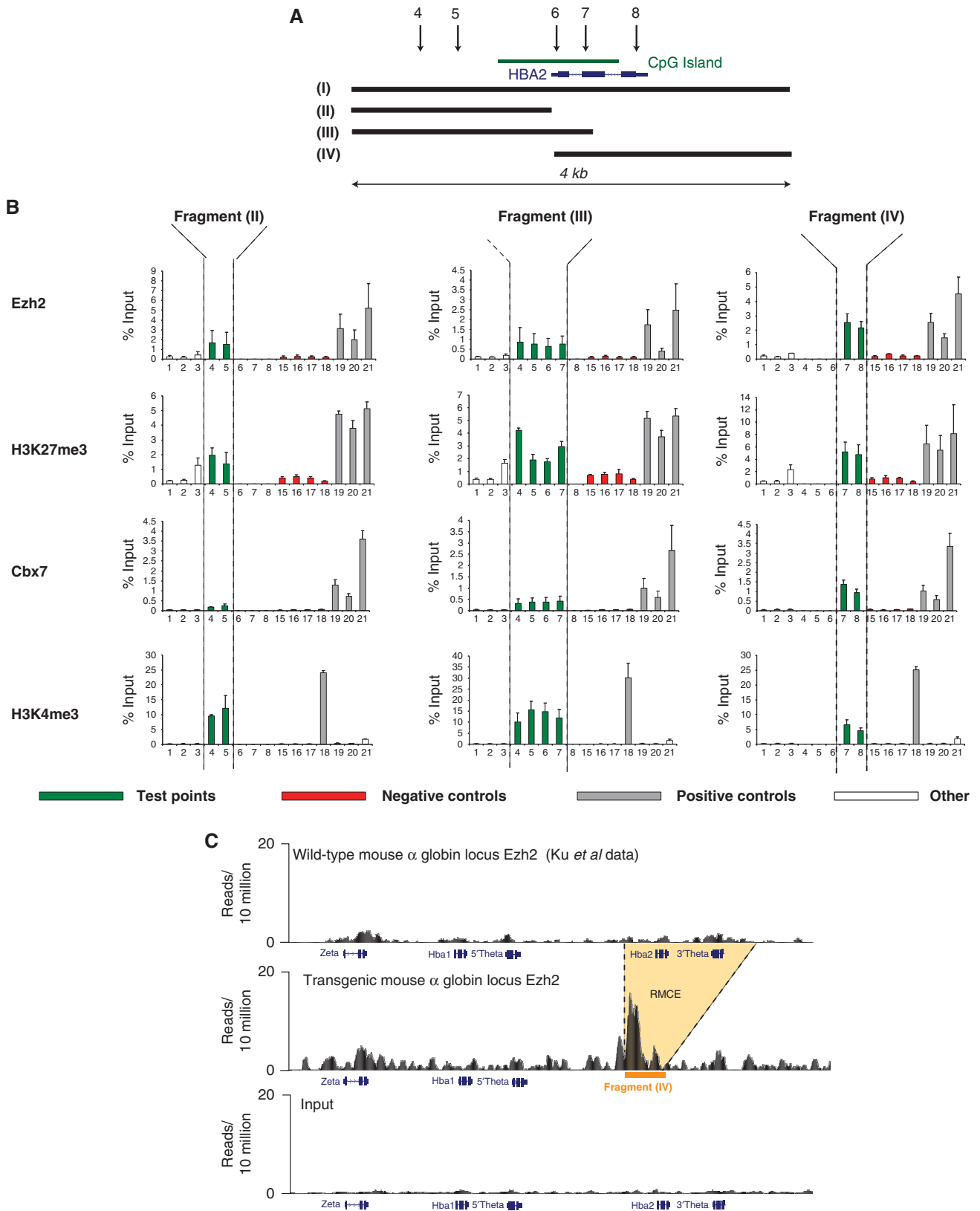
Figure 3 Chromatin state is redundantly encoded in the 4-kb *HBA2* fragment. (**A**) The indicated subfragments were separately integrated into the mouse α globin locus using the RMCE system. The positions of tested subfragments relative to the original 4-kb *HBA2* fragment are shown. (**B**) Following Flp-mediated excision of the *Hprt* selective marker gene, ChIP was performed with antibodies to Ezh2, Cbx7, H3K4me3 and H3K27me3 and enrichment was quantified with species-specific qPCR probes. qPCR probes 1–3 are located within the endogenous mouse α globin locus as indicated in Figure 1. Other probe positions are identical to Figure 2. Enrichment was quantified as percentage of input DNA. Data are the shown for three separately derived and analysed cell lines ± s.d. (**C**) ChIP-seq was performed with an antibody to Ezh2 for a cell line with Fragment (IV) inserted into the mouse α globin locus (orange bar). Reads were mapped to the transgenic locus (middle track). For comparison, ChIP-seq data are shown for Ezh2 at the wild-type mouse α globin locus (upper track; data from Ku *et al*). An input track is also displayed (lower track). ChIP-seq enrichment is normalized to mapped reads per 10 million.
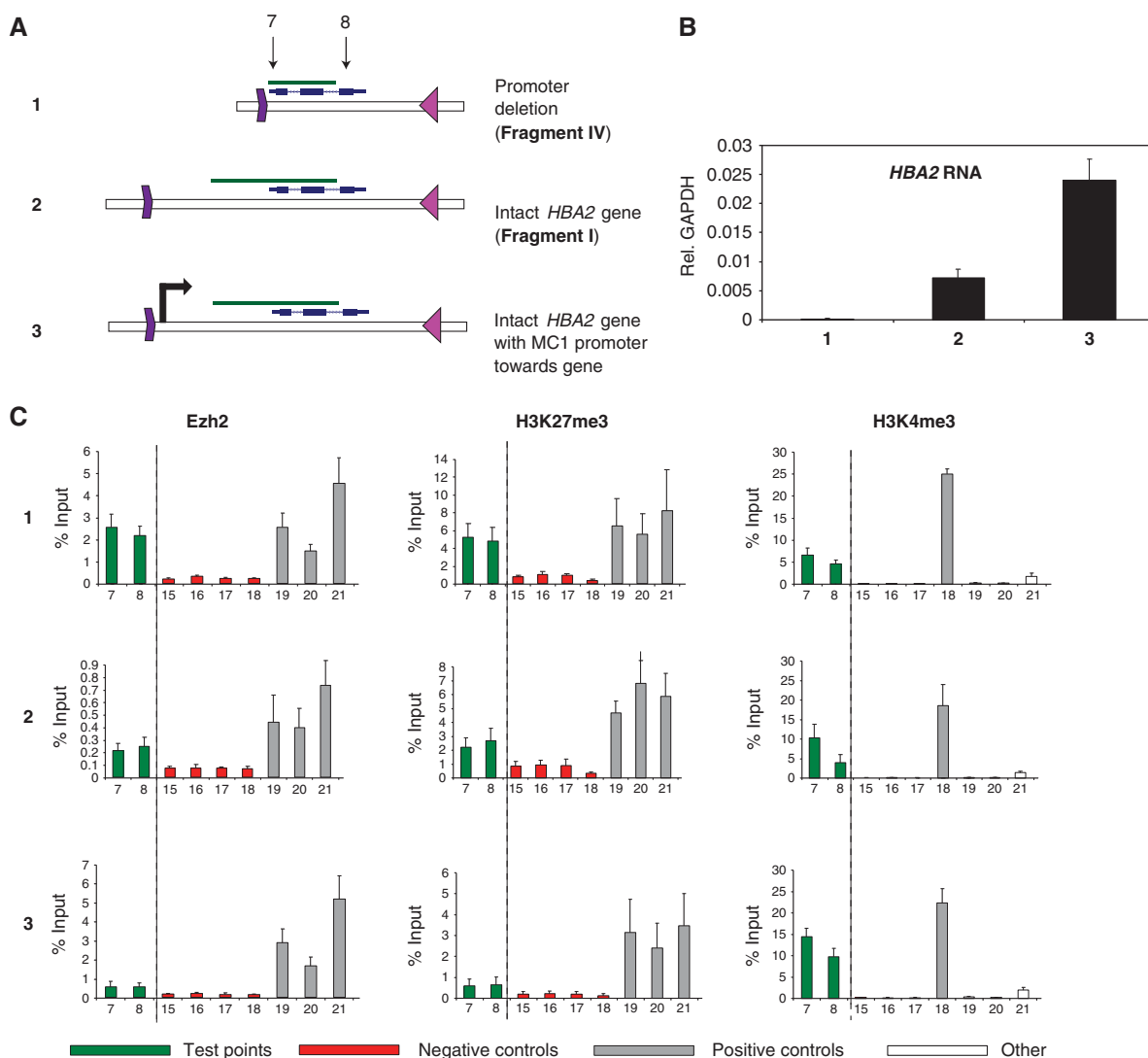
Figure 4 Chromatin bivalency at the *HBA2* gene reflects competition between Polycomb recruitment and transcriptional activation. (**A**) Fragments tested: (1) *HBA2* gene with promoter deletion (Fragment IV in Figure 3). (2) Intact *HBA2* gene (Fragment I in Figure 3). (3) Intact *HBA2* gene with an ~200-bp fragment encoding the MC1 promoter inserted upstream of the full 4-kb *HBA2* fragment with orientation towards the gene. (**B**) Expression of spliced *HBA2* RNA was quantified with a species-specific qPCR probe. Expression is quantified relative to mouse *Gapdh*. For each fragment tested, data are shown for three separately derived and analysed cell lines ± s.d. (**C**) ChIP was performed with antibodies to H3K4me3, Ezh2 and H3K27me3. qPCR amplicons present in all of the tested fragments are indicated (green). Enrichment was quantified as percentage of input DNA. Data are shown for three separately derived and analysed cell lines ± s.d.

Remarkably, all three of the test fragments became modified by H3K4me3 and H3K27me3 and recruited PRC1 and PRC2 complexes, albeit to varying degrees (Figure 3B). Therefore, it appears that the signal for chromatin bivalency is redundantly encoded throughout Fragment (I). Surprisingly, the greatest magnitude of PcG recruitment and H3K27me3 modification was observed not for the largest Fragment (I) but for a smaller Fragment (IV) which lacks the gene promoter. This fragment also had the lowest level of H3K4me3 modification. To compare the degree of PcG recruitment at the newly inserted human fragments with the flanking mouse locus, we performed ChIP-seq in transgenic ES cells containing Fragment (IV) using an antibody to Ezh2 (Figure 3C, lower track). The corresponding region in the unmodified genomic locus is illustrated for comparison (Figure 3C, upper track). This shows that the bivalent chromatin state is templated by sequences within the fragment rather than acquiring modification that has spread from

flanking sequences. We also note that there is not marked spreading of the Ezh2 signal into adjacent chromatin.

Chromatin bivalency reflects competition between PcG recruitment and transcriptional activation

Since PcG recruitment is greater when the promoter sequences are deleted it seemed possible that chromatin bivalency reflects a competition between the recruitment of PcG and activating proteins at the promoter. If so, the presence of additional activating sequences should shift the balance towards an active epigenetic state. To test this, we generated another transgenic cell line in which an ~200-bp fragment encoding an MC1 promoter, which is constitutively active in mouse ES cells (Thomas and Capecchi, 1987), was inserted upstream of the 4-kb *HBA2* fragment (Figure 4A), and the resulting construct was integrated by RMCE and compared with cell lines carrying Fragment (I) or Fragment (IV).

Quantification of spliced *HBA2* cDNA in these three cell lines revealed a very low level of expression when the promoter had been deleted, higher expression for the *HBA2* gene with intact promoter and higher still for the fragment containing the MC1 promoter (Figure 4B). It should be noted that the level of expression observed even for this fragment is several orders of magnitude lower than for *HBA2* in an erythroid cell. Next, we compared active and repressive histone modifications and PRC2 recruitment for all three of these RMCE-modified cell lines. There is an inverse relationship between H3K27me3 and H3K4me3 (Figure 4C) with the highest level of PRC2 recruitment and lowest level of H3K4me3 seen when the activating sequences associated with the promoter were deleted. Complete clearing of PRC2 and H3K27me3 occurred in the presence of the MC1 promoter and an intermediate level of H3K4me3 and H3K27me3 was observed for the wild-type construct.

CGI erosion during mammalian evolution is associated with loss of chromatin bivalency

A clear difference in primary sequence between the human and mouse α globin genes is the presence of prominent CGIs in the human but not in the mouse locus; this reflects erosion of the CGI in the mouse lineage compared with a common mammalian ancestor (Antequera and Bird, 1993). To investigate whether this association between CGI erosion and loss of PcG recruitment is a general phenomenon, we identified 2088 peaks of H3K27me3 in human ES cells that were associated with elevated CpG density ($\geqslant 6\%$ in a 500-bp window). These were compared with their corresponding genomic intervals (using the UCSC liftOver tool) in mouse and rat ES cells.

Peaks are ranked according to CpG density (in a 500-bp window) in the mouse genome (Figure 5A). Each line displays a peak of H3K27me3 in human ES cells and the chromatin state of corresponding genomic regions in mouse ES cells. CpG density is generally conserved between corresponding regions in the human and rodent genomes and this is associated with conservation of H3K27me3 recruitment in ES cells. However, numerous examples were identified for which CGI erosion between human and mouse is associated with diminution of the H3K27me3 histone modification (Figure 5A, below dashed line). This is confirmed by a pileup analysis comparing enrichment of H3K27me3 at mouse regions associated with conserved ($>3\%$; green) and eroded ($\leqslant 3\%$; blue) CpG density (Figure 5B). The correlation is also apparent when the data are presented as a scatter plot (Supplementary Figure S3). To confirm that this is not a species-specific phenomenon, a map of H3K27me3 was generated for rat ES cells and an analogous pileup analysis performed with similar findings (Figure 5C), although it is noted that the ratio of signal to background for this data set is inferior to the human and mouse data sets. Finally, three loci are illustrated for which CGI erosion in the mouse compared with human genome is associated with loss of the bivalent chromatin state in ES cells (Figure 5D–F).

De-novo PRC2 recruitment to CpG-rich sequences in DNA methyltransferase-deficient ES cells

Most sites of PcG recruitment in ES cells are associated with an elevated density of CpG dinucleotides (Ku *et al*, 2008). CGIs associated with actively expressed promoters are marked exclusively by H3K4me3 (Mikkelsen *et al*, 2007) and results presented here and elsewhere (Mendenhall *et al*, 2010) suggest that the binding of activating factors reduces PcG recruitment. However, this does not explain why some CGIs recruit neither PcG nor the H3K4me3 modification. We have found that DNA methylation can explain a proportion of these CGIs. An example is the CGI present in body of the human *RHBDF1* gene. In human ES cells (Figure 6A), the promoter is marked by H3K4me3 but the intragenic CGI is not modified by either H3K4me3 or H3K27me3. As the gene is silenced during erythropoiesis, the promoter recruits H3K27me3 but the intragenic CGI remains unmarked. We have previously reported that this intragenic CGI is methylated in multiple adult somatic tissues including erythroid cell lines (Vyas *et al*, 1992). This was confirmed for human ES cells by inspection of published genome-wide bisulphite sequencing data (Figure 6A; Lister *et al*, 2009). A genome-wide comparison of DNA methylation and the H3K4me3 and H3K27me3 histone modifications confirms that most methylated CGIs are unmarked by both histone modifications in human ES cells (Figure 6B–E). Conversely, inspection of chromatin state at unmethylated ($<5\%$ methylated) CGIs (Supplementary Figure S4A) reveals most unmethylated CGIs to be modified by either the H3K27me3 or H3K4me3 mark in human ES cells; moreover, there is a positive correlation with CGI size so that essentially all unmethylated CGIs >1 kb in size are modified by one or other of these marks (Supplementary Figure S4B–E).

The anticorrelation between DNA methylation and PcG recruitment in ES cells could reflect either (i) inhibition of PcG recruitment by DNA methylation, (ii) inhibition of DNA methylation by H3K4me3 and/or H3K27me3 or (iii) a confounding factor, such as the restriction of DNA methylation to non-promoter CGIs in human ES cells. To distinguish these models, we performed genome-wide ChIP-seq for H3K27me3 in mouse ES cells deficient for Dnmt3a and Dnmt3b methyltransferases and in wild-type mouse ES cells. The general pattern of H3K27me3 recruitment at positive (Figure 7A; Supplementary Figure S5A) and negative (Figure 7B; Supplementary Figure S5B) control regions is conserved between wild-type and knockout cell lines, although there appears to be somewhat greater spreading of the signal in knockout cells. However, numerous *de-novo* sites of H3K27me3 recruitment were observed at CpG-rich regions in the knockout cells (Figure 7C–J; Supplementary Figure S5C and D). A number of these sites were confirmed by qPCR (Supplementary Figure S5E). We hypothesized that DNA associated with these regions is methylated in wild-type ES cells and loses methylation in Dnmt3a/b$^{-/-}$ ES cells and this was confirmed by bisulphite sequencing (Supplementary Figure S6). These results demonstrate that loss of DNA methylation leads to *de-novo* sites of PcG recruitment at CpG-rich sequences, suggesting that PcG recruitment is the default state for such sequences in the absence of binding of transcriptional activators.

Discussion

Studies in which large chromosomal segments are introduced into the cells of another species (Wallace *et al*, 2007; Wilson *et al*, 2008) suggest that differences in chromatin state and
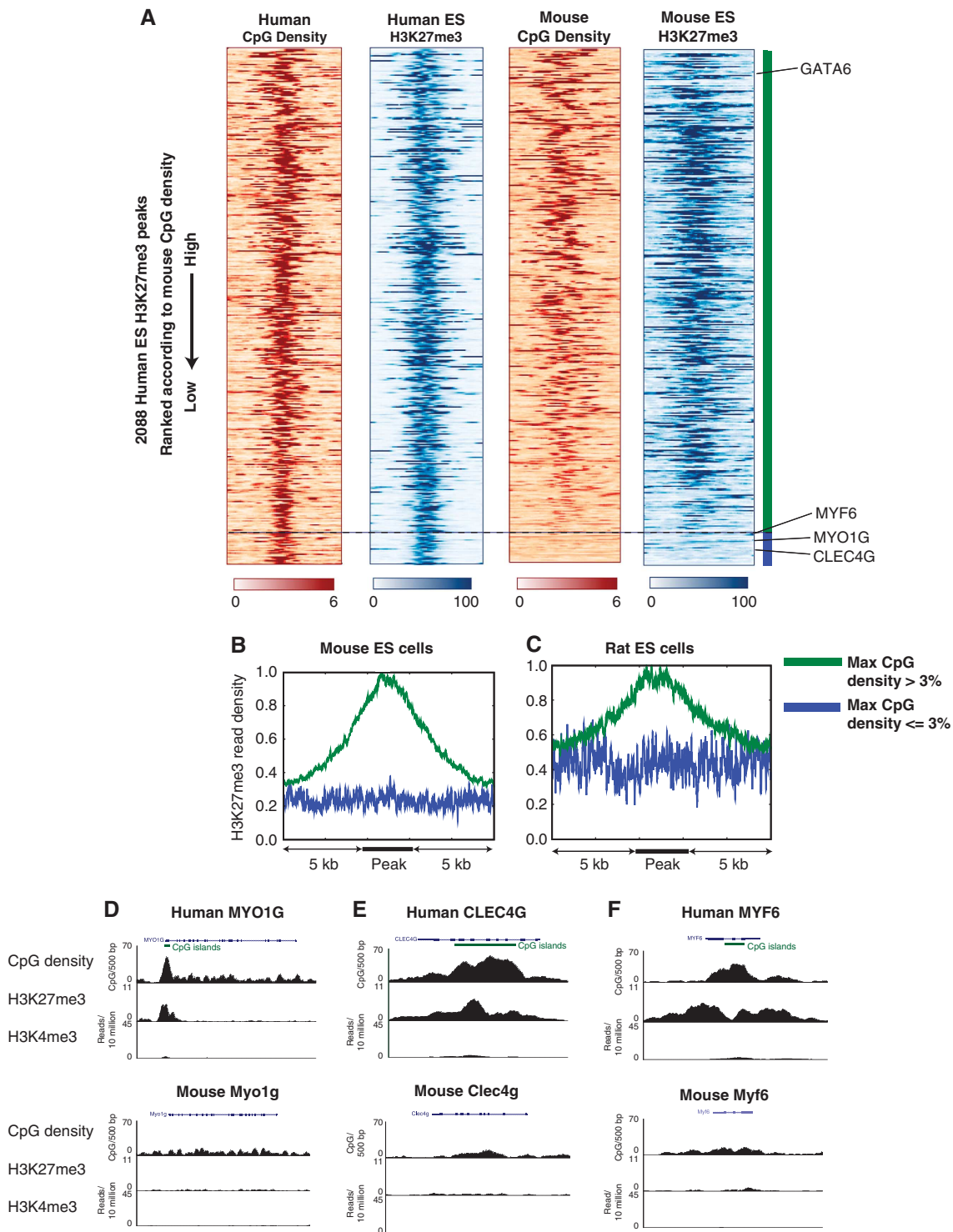
Figure 5 CpG island erosion during mammalian evolution is associated with loss of chromatin bivalency. (**A**) In all, 2088 peaks of H3K27me3 enrichment associated with a CpG density of 6% or greater (500 bp window) were identified in human ES cells (Ernst *et al*, 2011). Data for each peak are plotted on a single line with flanking regions of 5 kb (*X* axis). The corresponding genomic intervals in the mouse genome were identified (UCSC liftOver tool). CpG density in a 500-bp window (percent) and H3K27me3 read density in a 500-bp window (percentage of maximal enrichment) for mouse ES cells are plotted adjacent to the corresponding human peaks (Ku *et al*, 2008). Peaks are displayed in order of maximum CpG density in the mouse genomic interval. CpG erosion in the mouse genome is associated with a diminution of H3K27me3 recruitment in mouse ES cells. The locations of selected genes are indicated. (**B**) Pileup analysis of the same data comparing mouse genomic regions associated with a maximum CpG density (500 bp window) of >3% (green) to those with <3% (blue). (**C**) A map of H3K27me3 was generated for rat ES cells and an analogous pileup analysis was performed. (**D**, **E**) Selected examples of CpG island erosion between human and mouse genomes at the (**D**) *MYO1G*, (**E**) *CLEC4G* and (**F**) *MYF6* genes associated with loss of chromatin bivalency. ChIP-seq data are displayed as mapped reads per 10 million.
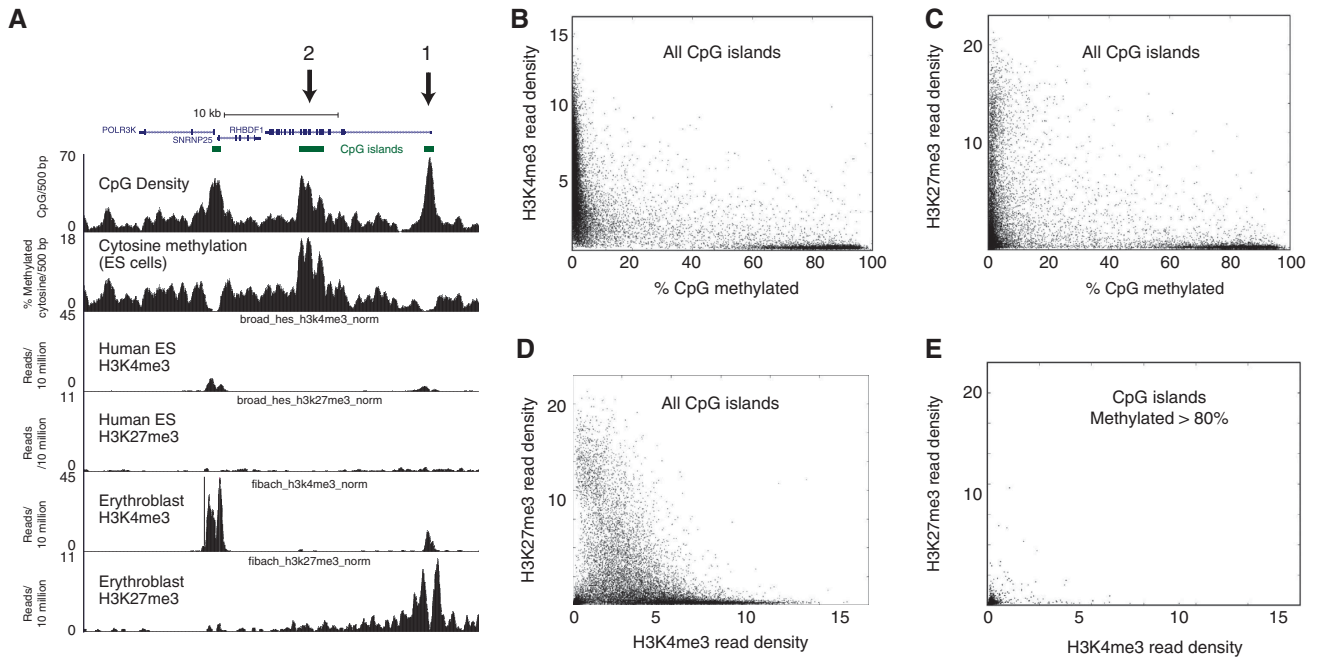
Figure 6 DNA methylation prevents the PcG-associated H3K27me3 histone modification in pluripotent cells. (**A**) Chromatin state (Ernst *et al*, 2011) and DNA methylation (Lister *et al*, 2009) at the human *RHBDF1* gene in human ES cells and erythroid cell types. ChIP-seq data are normalized to mapped reads per 10 million. The same scales are employed for ES and erythroid cell types. (**B, C**) For human H1 ES cells, DNA methylation was quantified at all CpG islands in the human genome (Lister *et al*, 2009) and compared with density of either (**B**) H3K4me3 or (**C**) H3K27me3 histone modifications at the corresponding CpG islands (Ernst *et al*, 2011). (**D, E**) H3K27me3 read density is plotted against H3K4me3 read density for (**D**) all CpG islands and (**E**) CpG islands with >80% methylated CpG dinucleotides. Read density is displayed as reads per 10 million per 500 bp window.

transcriptional output largely reflect differences in primary DNA sequence. Consequently, comparative genomics provides a powerful strategy for deciphering this regulatory code (Waterston *et al*, 2002). Sequence conservation across multiple organisms is predictive of functional elements (Pennacchio and Rubin, 2001; Hughes *et al*, 2005). However, not all functionally important sequences exhibit a high degree of conservation (Ruvinsky and Ruvkun, 2003; Bernstein *et al*, 2005), and at present the analysis of conserved sequence cannot predict which functional elements are required for a given effect. To date, the function of a particular sequence is usually addressed by transgenic experiments. However, the analysis of expression and chromatin state in such experiments is often confounded by copy number variation and position effects. Here, we have analysed elements required for the recruitment of a specific bivalent chromatin signature (H3K27me3 and H3K4me3) using RMCE, which allows the analysis of all tested fragments at single copy in a defined chromosomal location. This provides a powerful system for identifying the precise *cis* elements involved in chromatin templating and how far the effect may spread beyond the primary elements.

Bivalent domains are thought to result from the interplay between repressive (PcG/H3K27me3) and activating (TrxG/H3K4me3) pathways and play an important role in marking the promoters of key developmental regulators in multipotent cells (Boyer *et al*, 2006; Lee *et al*, 2006). The identity of *cis*-acting sequences responsible for PcG recruitment and the establishment of chromatin bivalency in mammals is controversial. In support of local sequences, a 1.8-kb region between human *HOXD11* and *HOXD12* was demonstrated to recruit PRC1 and PRC2 (Woo *et al*, 2010) and CpG-rich

sequences were reported to recruit PcG components to a transgenic BAC (Mendenhall *et al*, 2010). Conversely, there is evidence that genomic context plays an important role in PcG recruitment. A 3-kb element located adjacent to the mouse MafB recruited only PRC1 and not PRC2 components when assayed in mammalian cells (Sing *et al*, 2009). In addition, a number of reports have identified non-coding RNAs encoded *in cis* or *in trans* that are required for the recruitment of PcG (Rinn *et al*, 2007; Pandey *et al*, 2008; Zhao *et al*, 2008; Yap *et al*, 2010).

In this study, we were initially struck by the observation that human locus encoding α globin (*HBA*) contains prominent sites of PcG recruitment and chromatin bivalency in pluripotent cells whereas the orthologous mouse locus does not. We first confirmed that this is due to *cis* sequence differences by comparing these loci within the same nucleus. Furthermore, we found that a relatively small 4 kb DNA fragment containing the human *HBA2* gene was sufficient to recreate a novel site of chromatin bivalency when inserted into the corresponding position in the mouse locus. This observation is not specific to the α globin genes since another 4 kb region containing the human *FERD3L* gene also established chromatin bivalency. Thus, for these examples, chromatin bivalency is encoded by local sequences. Nevertheless, some domains of chromatin bivalency extend over many kilobases (Ku *et al*, 2008). Here, by analysing the *HBA2* gene, we have shown that this could be explained by redundant recruitment to multiple sequence elements as opposed to recruitment of chromatin modifying enzymes to a single site and subsequent spreading. Redundant encoding of a chromatin state may confer robustness to gene regulation in the face of single base substitutions.
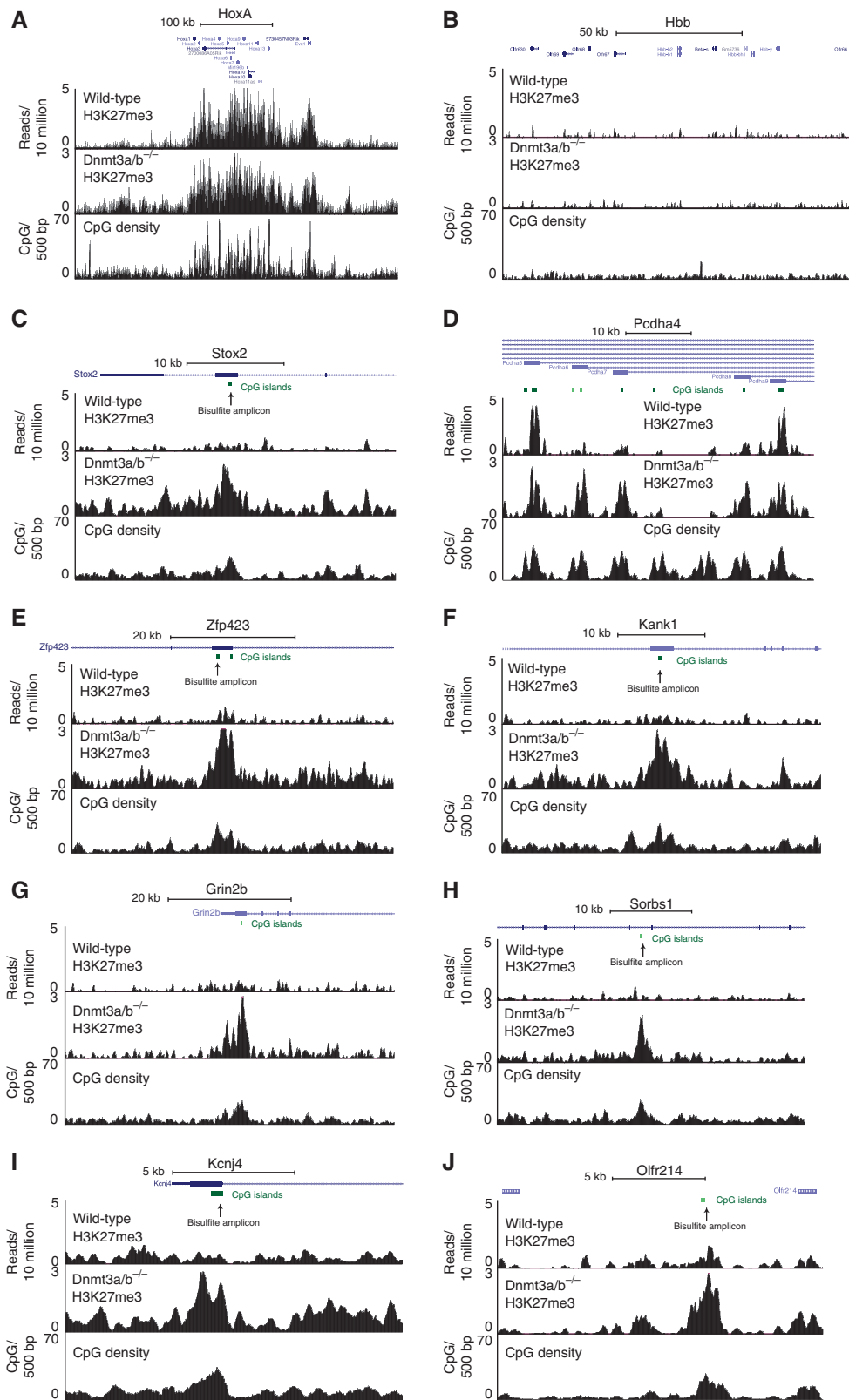
Figure 7 *De-novo* genomic sites of H3K27me3 methylation in Dnmt3a/b$^{-/-}$ ES cells. Genome-wide maps of H3K27me3 were generated for Dnmt3a/b$^{-/-}$ and wild-type ES (E14-TG2a) cell lines. For each locus, the profile of H3K27me3 (reads/10 million) is displayed for wild-type and mutant cells and compared with CpG density (CpGs/500 bp). (**A**) Positive control region (Hox A locus). (**B**) Negative control region (Hbb locus). (**C–J**) Examples of *de-novo* PRC2 recruitment at intragenic CpG islands that are methylated in wild-type cells and become demethylated in mutant cells. The locations of bisulphite amplicons displayed in Supplementary Figure S6 are indicated.

Our results suggest that the bivalent chromatin state reflects a competitive equilibrium between the recruitment of PcG to CpG-rich sequences and gene activation associated with recruitment of TrxG complexes that contain H3K4 methyltransferases. Deletion of promoter sequences from the *HBA2* gene increases PcG recruitment and H3K27me3 modification relative to the wild-type sequence, with corresponding diminution of H3K4 methylation. Conversely, the addition of a constitutively active promoter increases the level of H3K4me3 in association with a reduction in PcG recruitment. These findings are consistent with a study of randomly integrated transgenic BACs in mouse ES cells which found that the deletion of activating motifs from a housekeeping CGI promoter led to PcG recruitment (Mendenhall *et al*, 2010). A limitation of that study is that the site of genomic integration and copy number is not controlled; this is important since the presence of a transgene in multiple copies is sufficient to initiate PcG silencing in *Drosophila* (Pal-Bhadra *et al*, 1997).

Our results are also informative regarding the sequences responsible for recruitment of H3K4 methyltransferases to bivalent domains. Genome-wide studies have established that most sites of PcG recruitment in ES cells are also associated with at least a low level of the H3K4me3 modification (Mikkelsen *et al*, 2007) and it has been suggested that this 'pre-marks' the gene for activation. Recruitment of the hSet1 complex to unmethylated CGIs via the Cfp1 protein appears to play a role in this process (Thomson *et al*, 2010). On the other hand, there is a correlation between the magnitude of the H3K4me3 modification and the transcriptional output from bivalently marked promoters (Adli *et al*, 2010; De Gobbi *et al*, 2011). Our results are consistent with a hybrid model in which CpG-rich sequences are sufficient for basal levels of H3K4me3 but this is boosted by activating sequences in the promoter. Of interest, it appears that a relatively low level of transcription (as observed for the wild-type *HBA2* fragment) is associated with a substantial level of H3K4me3 modification, whereas a higher level of transcription is associated with clearing of PcG proteins.

The nature of the genomic signals for vertebrate PcG recruitment is a central question in the field of epigenetics. We, and others, have previously proposed a role for CpG-rich sequences in PcG recruitment (Garrick *et al*, 2008; Ku *et al*, 2008; Mendenhall *et al*, 2010). Consistent with this, the most striking sequence differences between the human and mouse α globin loci is the presence of prominent CGIs in the human but not in the mouse, in which the corresponding CGIs have become eroded. Here, we have shown that the association between erosion of CpG dinucleotide density and loss of PcG recruitment is a general phenomenon in mammalian evolution.

In the light of these observations, we revisited the genome-wide relationship between CGIs and PcG binding. It has been proposed that PcG recruitment is the default state for CGIs that do not recruit activating complexes (Ku *et al*, 2008). However, a major limitation of this hypothesis is the existence of CGIs that recruit neither PcG nor the active H3K4me3 modification. Consistent with previous reports of antagonism between PcG recruitment and DNA methylation in differentiated cell types (Lindroth *et al*, 2008; Puschendorf *et al*, 2008; Wu *et al*, 2010) we found that, in ES cells, the majority of methylated CGIs are marked by neither H3K27me3 nor

H3K4me3 and conversely that most unmethylated CGIs are modified by either the H3K27me3 or the H3K4me3 mark in human ES cells (Supplementary Figure S4).

These results suggested that, in the absence of binding of transcriptional activators, PcG recruitment is the default state for a genomic region containing a high density of unmethylated CpG dinucleotides. To test this hypothesis, we generated genome-wide maps of H3K27me3 in both Dnmt3a/b$^{-/-}$ and wild-type ES cells. Remarkably, we observed numerous examples of *de-novo* recruitment of the PcG-associated H3K27me3 mark at CpG-rich sites that lose DNA methylation. Taken together, these findings strongly suggest that a high density of unmethylated CpG dinucleotides is sufficient for vertebrate PcG recruitment. The mechanism by which this genomic signal is recognized remains to be determined.

Materials and methods

Targeting of RMCE cassette to the mouse α globin locus

A targeting vector for the mouse α globin locus was assembled in pNTFlox (a gift from J Hughes). In this vector, the floxed selection markers were replaced with an RMCE acceptor cassette (*frt/Hprt$^{-\Delta 3'}$/loxP/MC1neo/lox511*) created by modification of a previously described chromosome engineering cassette (Wallace *et al*, 2007) and flanked by homology arms designed to delete the 3′α–θ homology block in the mouse α globin locus. E14-TG2a.IV mouse ES cells, which are hypoxanthine phosphoribosyl transferase deficient (HPRT$^-$), were cultured and gene targeting by homologous recombination was performed as previously described (Wallace *et al*, 2007). Correctly targeted clones were identified by Southern blot with *Hind*III and *Bgl*II digests. Further details of the constructs used are available on request.

Recombinase-mediated cassette exchange

Test sequences were cloned into the *Asc*I site of a plasmid containing an RMCE donor cassette (*loxP/Hprt$^{-\Delta 5'}$/frt/AscI/lox511*) created by modification of a previously described chromosome engineering cassette (Wallace *et al*, 2007). In all, 75 μg of each RMCE donor plasmid was co-electroporated with 25 μg of pCAGGS-Cre-IRESpuro plasmid into an ES cell line containing the correctly integrated RMCE acceptor cassette. Clones by which Cre recombination had correctly reconstituted a functional *Hprt* selective marker were recovered by selection for HPRT$^+$ cells as previously described (Wallace *et al*, 2007). Finally, clones with a confirmed exchange event were electroporated with 25 μg of Flp(o) a mouse codon-optimized Flp recombinase (Raymond and Soriano, 2007), grown non-selectively for 6 days then plated out at 10^4 cells per 10 cm plate in medium supplemented with 10 μM 6-thioguanine as previously described (Wallace *et al*, 2007) in order to derive cells with the *Hprt* selection marker deleted.

Chromatin immunoprecipitation

Chromatin immunoprecipitation was performed with the Millipore ChIP Assay Kit (Millipore, 17-295). Briefly, ES cells were crosslinked with 1% formaldehyde in PBS for 10 min at 37°C. Chromatin was prepared according to the Millipore protocol and sonicated to an average size of 500–1000 bp using a Diagenode Bioruptor. Chromatin fragments were immunoprecipitated with antibodies to H3K4me3 (Millipore, 05-745R), H3K27me3 (Millipore, 07-449), Ezh2 (Abnova, pAB0649) or Cbx7 (Santa Cruz, P-15 sc70232). Immunoprecipitated DNA was either analysed by real-time qPCR or prepared for ChIP sequencing according to standard Illumina protocols. Enrichment was quantified by real-time qPCR as a percentage of input DNA with Taqman probes specific to the human and mouse α globin loci (Anguita *et al*, 2004). Primers and probes employed in this study are detailed in Supplementary Table S1.

Rat ES cells were expanded in feeder-free conditions on laminin-coated tissue culture plates (10 μg/ml; Sigma, L2020) in a modified 2i inhibitor medium based on published protocols (Buehr *et al*, 2008; Meek *et al*, 2010) and chromatin was prepared for sequencing as described above.

DNA and RNA analysis

RNA was prepared with TRI reagent (Sigma) and quantified relative to mouse *Gapdh* with RT-qPCR primers specific to the spliced human α globin transcript (Anguita *et al*, 2004). Bisulphite conversion of genomic DNA was performed with the EZ DNA Methylation-Gold kit (Zymo Research, D5005) and methylation was quantified by cloning into pGEM-T Easy (Promega) and sequencing.

Bioinformatic analysis

To investigate the relationship between CGI erosion and chromatin bivalency, publically available ChIP-seq data sets for H3K27me3 in human (Ernst *et al*, 2011) and mouse (Mikkelsen *et al*, 2007) ES cells were analysed with custom Python scripts. Peaks in human ES cells were identified with a sliding window of 500 bp and moving increment of 50 bp. Peaks separated by 1 kb or less were merged and peaks associated with enrichment on an input DNA track (Ernst *et al*, 2011) were eliminated. For peaks associated with a CpG density of 6% or greater in a 500-bp window, the corresponding mouse genomic regions were identified using the UCSC liftOver tool. Density of H3K27me3 and CpG dinucleotides was plotted for these genomic regions in human and mouse ES cells. Finally, a pileup analysis was performed to compare the genomic regions in mouse ES cells associated with ⩽3 versus >3% CpG density. An identical pileup analysis was also performed for H3K27me3 in rat ES cells.

To quantify histone modifications and DNA methylation at CGIs in human ES cells, publically available data sets for H3K27me3 (Ernst *et al*, 2011), H3K4me3 (Ernst *et al*, 2011) and high-coverage bisulphite sequencing (Lister *et al*, 2009) were analysed. For each annotated CGI (UCSC definition, hg18), CpG methylation was quantified as a fraction by dividing the total number of methylated cytosines by the total number of unmethylated cytosines in sequencing reads that mapped to that CGI. The density of histone modifications was quantified by taking the maximum read density in a sliding 500 bp window at any position within the CGI and flanking 1 kb regions (to account for nucleosome depletion at a subset of CGIs). Sex chromosomes and CGIs to which reads could not be mapped (UCSC mappability track) were excluded from the analysis.

ChIP-sequencing data from this study have been deposited with the GEO database (accession number GSE27580).

Conflict of interest

The authors declare that they have no conflict of interest.

References

Adli M, Zhu J, Bernstein BE (2010) Genome-wide chromatin maps derived from limited numbers of hematopoietic progenitors. *Nat Methods* **7:** 615–618

Anguita E, Hughes J, Heyworth C, Blobel GA, Wood WG, Higgs DR (2004) Globin gene activation during haemopoiesis is driven by protein complexes nucleated by GATA-1 and GATA-2. *EMBO J* **23:** 2841–2852

Antequera F, Bird A (1993) Number of CpG islands and genes in human and mouse. *Proc Natl Acad Sci USA* **90:** 11995–11999

Azuara V, Perry P, Sauer S, Spivakov M, Jorgensen HF, John RM, Gouti M, Casanova M, Warnes G, Merkenschlager M, Fisher AG (2006) Chromatin signatures of pluripotent cell lines. *Nat Cell Biol* **8:** 532–538

Barna M, Merghoub T, Costoya JA, Ruggero D, Branford M, Bergia A, Samori B, Pandolfi PP (2002) Plzf mediates transcriptional repression of HoxD gene expression through chromatin remodeling. *Dev Cell* **3:** 499–510

Bernstein BE, Kamal M, Lindblad-Toh K, Bekiranov S, Bailey DK, Huebert DJ, McMahon S, Karlsson EK, Kulbokas III EJ, Gingeras TR, Schreiber SL, Lander ES (2005) Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120:** 169–181

Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, Jaenisch R, Wagschal A, Feil R, Schreiber SL, Lander ES (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125:** 315–326

Boyer LA, Plath K, Zeitlinger J, Brambrink T, Medeiros LA, Lee TI, Levine SS, Wernig M, Tajonar A, Ray MK, Bell GW, Otte AP, Vidal M, Gifford DK, Young RA, Jaenisch R (2006) Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441:** 349–353

Buehr M, Meek S, Blair K, Yang J, Ure J, Silva J, McLay R, Hall J, Ying QL, Smith A (2008) Capture of authentic embryonic stem cells from rat blastocysts. *Cell* **135:** 1287–1298

Caretti G, Di Padova M, Micales B, Lyons GE, Sartorelli V (2004) The Polycomb Ezh2 methyltransferase regulates muscle gene expression and skeletal muscle differentiation. *Genes Dev* **18:** 2627–2638

Coller HA, Kruglyak L (2008) Genetics. It's the sequence, stupid!. *Science* **322:** 380–381

De Gobbi M, Garrick D, Lynch M, Vernimmen D, Hughes JR, Goardon N, Luc S, Lower KM, Sloane-Stanley JA, Pina C, Soneji S, Renella R, Enver T, Taylor S, Jacobsen SE, Vyas P, Gibbons RJ, Higgs DR (2011) Generation of bivalent chromatin domains during cell fate decisions. *Epigenetics Chromatin* **4:** 9

Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, Ku M, Durham T, Kellis M, Bernstein BE (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473:** 43–49

Garrick D, De Gobbi M, Samara V, Rugless M, Holland M, Ayyub H, Lower K, Sloane-Stanley J, Gray N, Koch C, Dunham I, Higgs DR (2008) The role of the polycomb complex in silencing alpha-globin gene expression in nonerythroid cells. *Blood* **112:** 3889–3899

Heo JB, Sung S (2011) Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science* **331:** 76–79

Hughes JR, Cheng JF, Ventress N, Prabhakar S, Clark K, Anguita E, De Gobbi M, de Jong P, Rubin E, Higgs DR (2005) Annotation of cis-regulatory elements by identification, subclassification, and functional assessment of multispecies conserved sequences. *Proc Natl Acad Sci USA* **102:** 9830–9835

Kim H, Kang K, Kim J (2009) AEBP2 as a potential targeting protein for Polycomb Repression Complex PRC2. *Nucleic Acids Res* **37:** 2940–2950

Ku M, Koche RP, Rheinbay E, Mendenhall EM, Endoh M, Mikkelsen TS, Presser A, Nusbaum C, Xie X, Chi AS, Adli M, Kasif S, Ptaszek LM, Cowan CA, Lander ES, Koseki H, Bernstein BE (2008) Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet* **4:** e1000242

Lee TI, Jenner RG, Boyer LA, Guenther MG, Levine SS, Kumar RM, Chevalier B, Johnstone SE, Cole MF, Isono K, Koseki H, Fuchikami T, Abe K, Murray HL, Zucker JP, Yuan B, Bell GW, Herbolsheimer E, Hannett NM, Sun K *et al* (2006) Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125:** 301–313

Lindroth AM, Park YJ, McLean CM, Dokshin GA, Persson JM, Herman H, Pasini D, Miro X, Donohoe ME, Lee JT, Helin K, Soloway PD (2008) Antagonism between DNA and H3K27 methylation at the imprinted Rasgrf1 locus. *PLoS Genet* **4:** e1000145

Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, Edsall L, Antosiewicz-Bourget J, Stewart R, Ruotti V, Millar AH, Thomson JA, Ren B, Ecker JR (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462:** 315–322

Margueron R, Reinberg D (2011) The Polycomb complex PRC2 and its mark in life. *Nature* **469:** 343–349

Meek S, Buehr M, Sutherland L, Thomson A, Mullins JJ, Smith AJ, Burdon T (2010) Efficient gene targeting by homologous recombination in rat embryonic stem cells. *PLoS One* **5:** e14225

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 Genet* **6:** e1001244

Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, Lee W, Mendenhall E, O'Donovan A, Presser A, Russ C, Xie X, Meissner A, Wernig M, Jaenisch R, Nusbaum C *et al* (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448:** 553–560

Pal-Bhadra M, Bhadra U, Birchler JA (1997) Cosuppression in Drosophila: gene silencing of Alcohol dehydrogenase by white-Adh transgenes is Polycomb dependent. *Cell* **90:** 479–490

Pandey RR, Mondal T, Mohammad F, Enroth S, Redrup L, Komorowski J, Nagano T, Mancini-Dinardo D, Kanduri C (2008) Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol Cell* **32:** 232–246

Pennacchio LA, Rubin EM (2001) Genomic strategies to identify mammalian regulatory sequences. *Nat Rev Genet* **2:** 100–109

Puschendorf M, Terranova R, Boutsma E, Mao X, Isono K, Brykczynska U, Kolb C, Otte AP, Koseki H, Orkin SH, van Lohuizen M, Peters AH (2008) PRC1 and Suv39h specify parental asymmetry at constitutive heterochromatin in early mouse embryos. *Nat Genet* **40:** 411–420

Raymond CS, Soriano P (2007) High-efficiency FLP and PhiC31 site-specific recombination in mammalian cells. *PLoS One* **2:** e162

Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E, Chang HY (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129:** 1311–1323

Ruvinsky I, Ruvkun G (2003) Functional tests of enhancer conservation between distantly related species. *Development* **130:** 5133–5142

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

Swiezewski S, Liu F, Magusin A, Dean C (2009) Cold-induced silencing by long antisense transcripts of an Arabidopsis Polycomb target. *Nature* **462:** 799–802

Thomas KR, Capecchi MR (1987) Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* **51:** 503–512

Thomson JP, Skene PJ, Selfridge J, Clouaire T, Guy J, Webb S, Kerr AR, Deaton A, Andrews R, James KD, Turner DJ, Illingworth R, Bird A (2010) CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* **464:** 1082–1086

Vyas P, Vickers MA, Simmons DL, Ayyub H, Craddock CF, Higgs DR (1992) Cis-acting sequences regulating expression of the human alpha-globin cluster lie within constitutively open chromatin. *Cell* **69:** 781–793

Wallace HA, Marques-Kranc F, Richardson M, Luna-Crespo F, Sharpe JA, Hughes J, Wood WG, Higgs DR, Smith AJ (2007) Manipulating the mouse genome to engineer precise functional syntenic replacements with human sequence. *Cell* **128:** 197–209

Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, Antonarakis SE, Attwood J, Baertsch R, Bailey J, Barlow K, Beck S, Berry E, Birren B, Bloom T, Bork P *et al* (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature* **420:** 520–562

Wilson MD, Barbosa-Morais NL, Schmidt D, Conboy CM, Vanes L, Tybulewicz VL, Fisher EM, Tavare S, Odom DT (2008) Species-specific transcription in mice carrying human chromosome 21. *Science* **322:** 434–438

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 H, Coskun V, Tao J, Xie W, Ge W, Yoshikawa K, Li E, Zhang Y, Sun YE (2010) Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes. *Science* **329:** 444–448

Yap KL, Li S, Munoz-Cabello AM, Raguz S, Zeng L, Mujtaba S, Gil J, Walsh MJ, Zhou MM (2010) Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. *Mol Cell* **38:** 662–674

Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT (2008) Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322:** 750–756