Histone H3.3 is required for endogenous retroviral element silencing in embryonic stem cells

Simon J. Elsässer1,2*, Kyung-Min Noh3, Nichole Diaz3, C. David Allis3 & Laura A. Banaszynski3,4*

Transposable elements comprise roughly 40% of mammalian genomes1. They have an active role in genetic variation, adaptation and evolution through the duplication or deletion of genes or their regulatory elements2-4, and transposable elements themselves can act as alternative promoters for nearby genes, resulting in non-canonical regulation of transcription5-8. However, transposable element activity can lead to detrimental genome instability9, and hosts have evolved mechanisms to silence transposable element mobility appropriately10,11. Recent studies have demonstrated that a subset of transposable elements, endogenous retroviral elements (ERVs) containing long terminal repeats (LTRs), are silenced through trimethylation of histone H3 on lysine 9 (H3K9me3) by ESET (also known as SETDB1 or KMT1E)12, and a co-repressor complex containing KRAB-associated protein 1 (KAP1; also known as TRIM28)13 in mouse embryonic stem cells. Here we show that the replacement histone variant H3.3 is enriched at class I and class II ERVs, notably those of the early transposons (ETn)/MusD family and intracisternal A-type particles (IAPs). Deposition at a subset of these elements is dependent upon the H3.3 chaperone complex containing α-thalassaemia/mental retardation syndrome X-linked (ATRX)12 and death-domain-associated protein (DAXX)12-14. We demonstrate that recruitment of DAXX, H3.3 and KAP1 to ERVs is co-dependent and occurs upstream of ESET, linking H3.3 to ERV-associated H3K9me3. Importantly, H3K9me3 is reduced at ERVs upon H3.3 deletion, resulting in derepression and dysregulation of adjacent, endogenous genes, along with increased retrotransposition of IAPs. Our study identifies a unique heterochromatin state marked by the presence of both H3.3 and H3K9me3, and establishes an important role for H3.3 in control of ERV retrotransposition in embryonic stem cells.

Deposition of the histone variant H3.3 has been linked to regions of high nucleosome turnover and has been traditionally associated with gene activation. However, we and others have demonstrated that H3.3 is incorporated into both facultative and constitutive heterochromatin12,15,16. Here, we used chromatin immunoprecipitation followed by sequencing (ChIP-seq) to identify 79,532 regions of H3.3 enrichment across the entire mouse genome, including repetitive regions (see later and Methods for details of data analysis), and performed a hierarchical clustering of H3.3 with various chromatin modifications. Consistent with deposition at euchromatin and heterochromatin, we observe H3.3 associated with both active (for example, H3K4me3, H3K27ac, H3K4me1) and repressed (for example, H3K9me3, H3K27me3) chromatin states (Fig. 1a). While most H3.3 peaks localized to genic regions and intergenic regulatory regions such as enhancers15, 23% (18,606/79,532) intersected with H3K9me3 peaks indicative of heterochromatic regions. Of these, 59% (11,010/18,606) localized to interspersed repeats (longer than 1 kb) and only 9% (1,747/18,606) fell within genic regions (Fig. 1b). Sequential ChIP-seq (re-ChIP) demonstrated co-enrichment of H3.3 and H3K9me3 at these regions (Fig. 1c).

To identify repeat families that were associated with H3.3, we mapped our H3.3 ChIP-seq data to a comprehensive database of murine repetitive sequences17-19. Unbiased hierarchical clustering demonstrated a striking correlation between H3.3, H3K9me3 and H3.3-H3K9me3 re-ChIP over class I and II ERVs, as well as enrichment of known silencing factors KAP1 and ESET (Fig. 1d and Extended Data Fig. 1). Class III ERVs and non-LTR long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs) carry little H3.3 and H3K9me3 but higher levels of H3K9me2. However, the promoter/5′ untranslated region (UTR) of intact LINE1 elements are enriched with H3.3, H3K9me3, KAP1 and ESET (Fig. 1d and Extended Data Fig. 1), suggesting a related mechanism of repression. Analysing individual well-annotated integration sites of ERVs20-22, we found that IAP and ETn/MusD ERVs, the most active transposons in the mouse genome21-23, are significantly enriched in H3.3 and H3K9me3 (Extended Data Fig. 2a-c), with 94% of IAP and 53% of ETn ERVs enriched with both H3.3 and H3K9me3 (Extended Data Fig. 2d).

Repetitive regions provide a challenge to next-generation sequencing analysis due to the ambiguity arising from mapping short reads to non-unique sequences. Standard ChIP-seq alignments disregard reads that map to more than a single location in the genome, leaving gaps wherever the underlying sequence is non-unique (Fig. 1e). To include interspersed repeats, we allowed random assignment of ambiguously mappable reads to one of the best matches24 (Fig. 1e), effectively averaging counts over multiple occurrences of the same exact read match. As exemplified by ETn and IAP insertions downstream of the Vn4 transcription start site, H3K9me3 is broadly enriched over the non-unique ERV sequence, whereas H3.3 appears to be more confined over 3′ and 5′ regions of the repeats (Fig. 1e). Neither ChIP-seq using an antibody recognizing only the canonical H3 isoforms (H3.1/2) nor an antibody recognizing all H3 isoforms (total H3; H3.3 constitutes ~10% of total H3 in embryonic stem (ES) cells) show enrichment at the corresponding regions (Fig. 1e), and H3.3 enrichment was lost in ES cells lacking H3.3 (Extended Data Fig. 3). We were further able to detect both H3.3 and H3K9me3 in the uniquely mappable flanking sites of IAP and ETn ERVs, (Extended Data Fig. 4a, b). In addition to full ERVs, we found single (so-called ‘orphan’) LTRs to be enriched in both H3.3 and H3K9me3 (Extended Data Fig. 4c), suggesting that the LTR sequence itself is sufficient for the nucleation of H3.3 and heterochromatin factors.

H3.3 deposition has been linked to dynamic chromatin regions with high levels of nucleosome turnover and DNA accessibility. As H3.3 enrichment at ETn and IAP ERVs was comparable to levels found at active promoters in ES cells (Extended Data Figs 2a and 5a; compare also to Rps12 enrichment in Fig. 1e), we tested whether ERVs were nucleosome-depleted in ES cells. Surprisingly, we found that ERVs...
Figure 1 | H3.3 is co-enriched with H3K9me3 at class I and II ERV-associated heterochromatin. a, Hierarchical (Spearman rank) clustering of H3.3 peaks on chromosome 1 with histone modifications associated with active (green) or repressed (red) chromatin states. Annotated genes and ERVs are shown. b, Venn diagram of H3.3 and H3K9me3 peaks demonstrating overlap at repetitive elements. c, ChIP-seq density heat maps for peaks classified as H3.3 only (n = 60,925), both H3.3 and H3K9me3 (n = 18,605), or H3K9me3 only (n = 54,204). Colour intensity represents normalized and globally scaled tag counts. d, ChIP-seq enrichment of H3.3 and heterochromatic histone modifications and factors mapped to the repetitive genome. Data are represented in a hierarchically (Spearman rank) clustered heat map of log2 fold enrichment (red) or depletion (blue) over a matched input. See Extended Data Fig. 1 for complete heat map. e, Genome browser ChIP-seq representations in ES cells. Read counts are normalized to total number of reads for each data set and exclude (‘unique’) or include (‘inclusive’) repetitive reads. MTA, MT subfamily A. f, ChIP-seq enrichment of H3.3 and H3K9me3 at various repeat regions in ES cells (ESCs) and NPCs. Data are represented as in d, g. Levels of co-enriched H3.3–H3K9me3 in control and ESET conditional knockout (cKO) ES cells. IAPez, IAP subfamily E; WT, wild type. ***p < 0.0001, one-sided Wilcoxon signed rank test. NS, not significant.

shown low DNA accessibility compared to promoters of highly expressed genes with comparable H3.3 enrichment, as measured by DNase and MNase digestion25, and showed no signs of transcription as judged by RNA polymerase (Pol) II occupancy12 (Extended Data Fig. 5a). Notably, we find that newly synthesized H3.3 (ref. 26) is rapidly incorporated at IAPs, despite the high levels of H3K9me3 and silent state (Extended Data Fig. 5b). Overall, our data suggest that a substantial fraction of H3.3 resides at ERVs in ES cells and constitutes a unique chromatin state fundamentally distinct from previously described combinations of histone variants and modifications.

Previous studies have demonstrated that silencing of ERVs via H3K9me3 is unique to the pluripotent or embryonic state, with adult somatic tissues showing dependence upon DNA methylation for ERV repression. Concomitant with loss of H3K9me3, H3.3 enrichment is responsible for H3K9me3 at all H3.3-containing classes of repeats (Fig. 1g and Extended Data Fig. 6c). SUV39h1/2 deletion resulted in a small decrease of H3K9me3 at IAP and ETn/MusD elements, but greatly decreased H3K9me3 at intact LINE elements, including their 5’ UTR (Extended Data Fig. 6c). In conclusion, the co-occurrence of H3.3 and H3K9me3 facilitated by ESET methyltransferase activity defines a novel class of heterochromatin that functions at ERVs and intact LINE 5’ ends.

The histone variant H3.3 is incorporated at distinct regions of chromatin by either the HIRA or ATRX–DAXX histone chaperone complexes12–14. We and others previously demonstrated that HIRA is responsible for H3.3 enrichment at genic regions, while the ATRX–DAXX complex facilitates H3.3 deposition at simple repeat regions such as telomeres12,13,15. Using ChIP-seq, we found that DAXX and ATRX were responsible for H3.3 incorporation at regions enriched with both H3.3 and H3K9me3, whereas HIRA facilitated deposition at genomic activity. Previous studies demonstrate that ESET has a critical role in the establishment of H3K9me3 at a large number of ERVs10, while SUV39h1/2 is involved in the maintenance and spreading of H3K9me3 at a subset of repeat elements27. To elucidate which methyltransferase was responsible for establishing H3.3/H3K9me3 including their 5’ ends.
Data are represented in a heat map of log2 fold enrichment (red) or depletion of chaperones and chaperone-dependent H3.3 deposition at repetitive regions. ERVs enriched with KAP1 and ESET, as well as telomeres (Fig. 2b). To demonstrate that both DAXX and ATRX co-occupied class I and II ERVs, we mapped genome-wide enrichment of KAP1 and found that almost half of the KAP1 peaks demonstrated substantial overlap in ES cells. d, Levels of H3.3 in control and KAP1 conditional knockout (cKO; top) and control and ESET cKO (bottom) ES cells. ***p < 0.0001, *p < 0.05, one-sided Wilcoxon signed rank test. NS, not significant. e, Immunoblotting of DAXX immunoprecipitated from wild-type or H3.3-null nuclear extracts showing co-immunoprecipitation with ATRX, H3.3, H3K9me3 and KAP1 independent of the H3.3 substrate. Asterisk denotes cross-reacting band.

Figure 3 | H3.3 is required for the maintenance of H3K9me3 at specific class I and II ERVs. a, Levels of H3K9me3 and total H3 in control and H3.3 knockout (KO) ES cells. ****p < 0.0001, one-sided Wilcoxon signed rank test. NS, not significant. b, ChiP-seq analysis at repeats demonstrated that both DAXX and ATRX co-occupied class I and II ERVs enriched with KAP1 and ESET, as well as telomeres (Fig. 2b). To understand further the relationship between the corepressor KAP1 and ATRX–DAXX-dependent H3.3 deposition at ERVs, we mapped genome-wide enrichment of KAP1 and found that almost half (13,730/29,185) of the KAP1 peaks coincided with shared H3.3/ H3K9me3 peaks (Fig. 2c). We therefore wanted to determine whether KAP1 had a role in targeting H3.3 deposition via recruitment of ATRX–DAXX. Indeed, H3.3 enrichment was reduced at IAP ERVs in the absence of KAP1 but was independent of ESET (Fig. 2d and Extended Data Fig. 7c–e), suggesting a novel role for KAP1 in recruitment of ATRX–DAXX.

To determine whether KAP1 and ATRX–DAXX associated biochemically, we prepared nuclear extracts from ES cells. We found that DAXX co-immunoprecipitated its known complex member ATRX as well as its substrate H3K9me3 (Fig. 2e). Of note, DAXX-associated histone was enriched with H3K9me3 (Fig. 2e). In addition, DAXX co-immunoprecipitated KAP1 (Fig. 2e), suggesting that DAXX–ATRX and KAP1 can form a biochemical complex. HIRA was not co-immunoprecipitated KAP1 (Fig. 2e), suggesting that DAXX–ATRX and KAP1 had a role in targeting H3.3 deposition via recruitment of ATRX–DAXX. Indeed, H3.3 enrichment was reduced at IAP ERVs in the absence of KAP1 but was independent of ESET (Fig. 2d and Extended Data Fig. 7c–e), suggesting a novel role for KAP1 in recruitment of ATRX–DAXX.
consequence of reduced DAXX protein stability in the absence of H3.3. Together, these data suggest that H3.3, DAXX and KAP1 are cooperative in their function related to ERV silencing (Fig. 2g). Intriguingly, while H3.3 enhances KAP1 and DAXX recruitment to ETn/MusD elements (Fig. 2f and Extended Data Fig. 7g), the variant remains enriched at these elements in the absence of the corepresser complex (Fig. 2b, d and Extended Data Fig. 7c–e).

As we observed a positive correlation between H3K9me3 and H3.3 enrichment at IAP integration sites (Extended Data Figs 4a and 8a), we next tested whether there was a functional link between H3.3 deposition and H3K9me3 establishment at specific subclasses of ERVs. Although global levels of H3K9me3 were relatively unaffected by the loss of H3.3 (Extended Data Fig. 8b), we found that H3K9me3 was restricted specifically to peaks enriched with both H3.3 and H3K9me3, concomitant with a reduction of KAP1 occupancy (Extended Data Fig. 8c). Indeed, H3K9me3 levels were reduced up to 50% at IAP, ETn and MusD repeats in the absence of H3.3 (Fig. 3a). Importantly, nucleosome density was not reduced, as evidenced by the overall maintenance of total H3 (Fig. 3a). Intriguingly, H3K9me3 levels were reduced at ETn/MusD elements in the absence of DAXX, ATRX, KAP1 or ESET (Extended Data Fig. 8d–h), whereas H3.3 enrichment at these elements was independent of the corepresser complex (Fig. 2 and Extended Data Fig. 7), suggesting a multifaceted mechanism in which both H3.3 deposition and corepresser complex recruitment contribute to ERV silencing.

Intriguingly, ERVs retained H3.3 to a larger extent than other regions in ES cells RNA interference (RNAi)-depleted of H3.3 (H3.3 knockout16; Extended Data Fig. 9a–c), suggesting they may act as ‘sinks’ for the remaining low levels of H3.3 present in H3.3-knockdown ES cells. Furthermore, exogenously expressed H3.3, but not H3.2, in both H3.3-knockout and H3.3-knockdown ES cells was focally enriched at IAP ERVs (Extended Data Fig. 9d–f). Importantly, exogenous expression of H3.3, but not H3.2, was able partially to rescue the loss of H3K9me3 at specific repeat regions (Fig. 3b). Together, these data suggest a direct and variant-specific role for H3.3 in establishing H3K9me3 chromatin at a subset of ERVs that cannot be compensated by the canonical H3.1/2 isoforms.

As H3K9me3 is known to be required for silencing of ERVs26, we tested whether loss of H3.3 would cause a derepression of ERVs concomitant with a reduction of H3K9me3 levels. RNA-sequencing (RNA-seq) demonstrated a moderate increase in global transcripts from IAPs, but not ETn/MusD ERVs (Fig. 4a). Since ERVs have recently been shown to control expression of nearby genes26, we next tested whether endogenous genes that were deregulated in H3.3-knockout ES cells were proximal to ERVs. While the majority of ERVs are neutral to neighbouring genes, a number of genes in the vicinity of ERVs were highly upregulated (Fig. 4b and Extended Data Fig. 10a), including the known gene Cyp2b23 and a new putative chimaeric transcript originating from a MusD element within the Aass gene (Fig. 4c). Notably, the same set of transcripts was upregulated in H3.3-depleted ES cells, albeit at a lower level (Fig. 4b), suggesting that the remaining H3.3 is partially functional in maintaining silent ERVs.

We hypothesized that ERV desilencing should result in increased ERV mobility. Paired-end sequencing of genomic DNA identified 80 non-annotated IAP integration sites unique to H3.3-knockout ES cells, and only 17 unique to wild-type ES cells (Fig. 4d and Extended Data Fig. 10b). As derepressed IAPs have been shown to cause chromosome rearrangements, we analysed H3.3-knockout ES cells for increased genome instability. Indeed, karyotypic analysis of H3.3-knockout ES cells showed a number of chromosomal abnormalities not observed in the wild-type control (Extended Data Fig. 10c). Despite these observations, we cannot exclude that genomic instability in H3.3-knockout ES cells might result from a loss of function unrelated to retrotransposon silencing27,30.

We have uncovered an unexpected role for the histone variant H3.3 in the establishment of heterochromatin. We demonstrate a hierarchy for deposition of H3.3, favouring DAXX–ATRX-mediated chromatin assembly at ERVs over transcription-associated deposition. We propose a model in which H3.3-containing chromatin facilitates the recruitment of KAP1 to ERVs, which in turn recruits DAXX–ATRX for the maintenance of H3.3 chromatin, thus creating a positive feedback or propagation loop. This mechanism acts synergistically with ESET-mediated H3K9me3 in maintaining a silent chromatin state at ERVs. Our data also indicate an H3.3-independent function of DAXX–ATRX in maintaining H3K9me3, possibly related to an architectural role in a larger corepresser complex with KAP1 and ESET. Our findings solidify an emerging understanding of the importance of the histone variant H3.3 in the establishment of silenced chromatin states and in maintenance of genome stability.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 23 June 2014; accepted 19 February 2015.
Published online 4 May 2015.

1. Waterston, R. H. et al. Initial sequencing and comparative analysis of the mouse genome. Nature 420, 520–562 (2002).
2. Longo, M. S., Carone, D. M., Green, E. D., O’Neill, M. J. & O’Neill, R. J. Distinct retroelement classes define evolutionary breakpoints demarcating sites of evolutionary novelty. BMC Genomics 10, 334 (2009).

3. Lee, J., Han, K., Meyer, T. J., Kim, H. S. & Batzer, M. A. Chromosomal inversions between human and chimpanzee lineages caused by retrotransposons. PLoS ONE 3, e4047 (2008).

4. Feschotte, C. & Gilbert, C. Endogenous viruses: insights into viral evolution and impact on host biology. Nature Rev. Genet. 13, 283–296 (2012).

5. Karimi, M. M. et al. DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs. Cell Stem Cell 8, 675–687 (2011).

6. Rowe, H. M. et al. TRIM28 repression of retrotransposon-based enhancers is necessary to preserve transcriptional dynamics in embryonic stem cells. Genome Res. 23, 452–461 (2013).

7. Robberecht, C., Voet, T., Zamani Esteki, M., Nowakowska, B. A. & Vermeesch, J. R. Nonallelic homologous recombination between retrotransposable elements is a driver of de novo unbalanced translocations. Genome Res. 23, 411–418 (2013).

8. Maksakova, I. A., Mager, D. L. & Reiss, D. Keeping active endogenous retroviral-like repetitive elements from high-throughput sequence data. Genome Biol. 11, 678–691 (2010).

9. Lewis, P. W., Elsaesser, S. J., Noh, K. M., Stadler, S. C. & Allis, C. D. Daxx is an essential telomeric chromatin in embryonic stem cells. Proc. Natl Acad. Sci. USA 107, 14075–14080 (2010).

10. Wong, L. H. et al. Histone H3.3 incorporation provides a unique and functionally essential telomeric chromatin in embryonic stem cells. Genome Res. 19, 404–414 (2009).

11. Banaszynski, L. A. et al. Hira-dependent histone H3.3 deposition facilitates PRC2 recruitment at developmental loci in ES cells. Cell 155, 107–120 (2013).

12. Mikkelsen, T. S. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448, 553–560 (2007).

13. Day, D. S., Luquette, L. J., Park, P. J. & Kharchenko, P. V. Estimating enrichment of repetitive elements from high-throughput sequence data. Genome Biol. 11, R69 (2010).

14. Frey, A., Listovsky, T., Guilbaud, G., Sarkies, P. & Sale, J. E. Histone H3.3 is required to maintain replication fork progression after UV damage. Curr. Biol. 24, 2195–2201 (2014).

Acknowledgements This work was supported by the Rockefeller University Fund and the Tri-Institutional Stem Cell Initiative. S.J.E. acknowledges funding from EMBO ALTF 1232-2011 and the Cambridge University Henrietta Smith Fund. We thank D. Pickett, D. Trono and D. Shinkai for cell lines. We thank A. Goldberg and C. Li for technical assistance and members of the Allis laboratory for helpful discussions.

Author Contributions S.J.E. and L.A.B. contributed equally to this project. S.J.E. and L.A.B. conceived, performed and interpreted most experiments with guidance from C.D.A. K.-M.N. and N.D contributed to experiments. S.J.E. and L.A.B. wrote the manuscript with input from all authors.

Author Information The ChiP-seq and RNA-seq data sets have been deposited in the Gene Expression Omnibus under accession number GSE59189. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.J.E (simon.elsasser@scilifelab.se), C.D.A (alliscd@rockefeller.edu) or L.A.B. (laura.banaszynski@utsouthwestern.edu).
METHODS

ChIP-seq analysis. ChIP-seq libraries were prepared according to the Illumina protocol and sequenced with HiSeq 2000. Raw reads in FASTQ format were aligned to the mouse genome version mm9 with Bowtie34 using ‘-m 1 --best’ protocol and sequenced with HiSeq 2000. Raw reads in FASTQ format were extracted from normalized wig files using the java-genomics-toolkit wigmath.Scale function (as a reference, 17.5 Mio mapped reads for a fragment size of 150 kbp with 1 average genome-wide read density of 1 (for mm9). Figures of these continuous tag counts over selected genomic intervals were created in the IGV browser (The Broad Institute).

Repetitive genome ChIP-seq analysis. The current build of repetitive repeat sequences was downloaded from Repbase (http://www.girinst.org/repbase/) and filtered for Mus musculus sequences. A Bowtie index was created with Bowtie-build. Raw ChIP-seq FASTQ reads were mapped to the repetitive sequence database using Bowtie.”--best” and “--k 1” options. A table of mapped short read counts per repetitive element were extracted from bam file using SAMTools idxstats function. Further analysis was performed with R and visualized as heat maps using GENE-E. Mapped read counts were expressed as a fraction of total mapped repetitive reads for each sample. For enrichment analysis, normalized read counts of ChIP samples were divided by normalized read counts of a matched input sample and expressed as log, fold enrichment. In addition, the following quality controls were performed: read distribution across the repetitive sequence was inspected using IGV genome browser for each repeat family to confirm coverage of the whole repetitive sequence. To avoid over- or underestimating fold enrichments due to low sequence representation, repetitive sequences with consistently less than ~100 mapped reads per sample or control were excluded from analysis. Peak calling. Peaks were called for H3.3, H3K9me3 and total H3 ChIP-seq data from control C7BL/6J ES cells16, including non-unique reads. MACS ChIP-seq peak finding was performed against a matched input using cut-off values “--pval 1e-6 --mfold 10.50.” 79,532, 72,811 and 29,189 peaks were called for H3.3, H3K9me3 and KAP1, respectively. For total H3, only 966 peaks were called with the same parameters.

Enrichment analysis over H3.3 peaks. For Fig. 1a, enrichment of H3.3 and histone modiﬁcations over H3.3 peaks were calculated as follows: average ChIP-seq read densities over the peak interval defined in the MACS41 bed output file were extracted from normalized wig ﬁles using the java-genomics-toolkit ngs.IntervalStats function (T. Palpant; http://palpant.us/java-genomics-toolkit/). ChIP-seq enrichment for each interval was normalized subsequently, dividing the mean read density of the ChIP-seq sample by the corresponding density of the matched input sample. Data were visualized in a heat map as log, fold enrichment over input and clustered with GENE-E (The Broad Institute).

Enrichment analysis over repetitive and unique genomic regions. For Figs 1g, 2d, f, 3a and Extended Data Figs 2a–c, 5, 6, 8d, 9c, f, i intervals were derived from following sources: Transcription start sites (TSSs) of ~2 000 highly active genes previously shown to be enriched in H3.3 were deﬁned as intervals from ~1 kb to +1 kb around their annotated TSS. H3K27me3-containing promoters (K27/prom) were previously characterized12. Curated sets of IAP, IAPd, RLTR10, ETn, MusD, ERGLN, ERVK10C6 and L1MM were previously described16. Additional intact IAP elements were identiﬁed using the BLAT function of UCSC and combined with the existing IAP data set. Intact LINE L1MM promoters’/5′ UTRs were identiﬁed in the reference genome using BLAT with the RepBase sequence L1MM_F_I1.

Enrichments over these intervals were calculated as described earlier from normalized ChIP and input wig ﬁles using wigmATH.IntervalStats. Log, fold enrichments over individual intervals were summarized using R in boxplots (Tucker-box-and-whisker plots using R boxplot defaults). Speciﬁcally, the box indicates median, as well as upper and lower quartiles of the data. Whiskers extend to 1.5 times the interquartile range (IQR). Outliers are not shown. Signiﬁcance levels were calculated using Wilcoxon tests: not signiﬁcant, P > 0.05; * P < 0.05; ** P < 0.01; *** P < 0.001; **** P < 0.0001. Peak proﬁle heat maps. For Figs 1c, 2a and Extended Data Figs 8c, 9e, peak proﬁle heat maps were calculated using ngspovot over a 5-kb window around the MACS peak centres (parameters: -SC global -10 L 2500 -MQ 0 -RB 0.05) from BAM ﬁles using the inclusive mapping procedure. Data sets are normalized to total mapped read counts and all maps are represented on the same global scale.

Analysis regions flanking repetitive elements. For Extended Data Fig. 4, proﬁles were calculated from uniquely mapped reads only, that is, non-unique reads and duplicates were discarded before calculating the genomic using IGVTools count function (see earlier). Proﬁles over flanking regions were aggregated using the sitepro function from the CEAS suite36 with the following modiﬁcations: proﬁles were not centred over the element but instead separately collected for the 3′ and 5′ flanking regions. The mean of the proﬁles in two, 5′ and 3′, 500
bp windows was extracted for each interval as an approximation of enrichment over the central, repetitive, interval. Profiles were either visualized as heat maps (using GENE-E), or averaged into a single plot (CEAS sitepro). Wig files were normalized to a global average of 1, thus the ordinate of the profile plot represents fold enrichment/depletion over a random genome-wide distribution of reads.

**RNA-seq preparation and analysis.** RNA was isolated using QIAGEN RNeasy. Libraries were prepared according to the Illumina TruSeq protocol and were sequenced on the HiSeq 2000. Resulting reads (101 nucleotides) were aligned to the mouse genome (mm9) using TopHat\(^\text{39}\). Gene expression level measured as FPKM was determined by the maximum likelihood estimation method implemented in the Cufflinks software package with annotated transcripts as references. Differential expression was analysed using the Student’s \( t \)-test in the program Cuffdiff\(^\text{40}\) with \( P \) values corrected for multiple testing.

**De novo mapping of unannotated ERVs.** Genomic DNA from H3.3 wild-type and KO1 ES cells was sheared to an average of 500 bp. Illumina paired-end sequencing was performed with 50 bp read lengths. ERVs were mapped to the reference genome in a two-step procedure. First, all reads were mapped to a genome consisting of all RepBase sequences belonging to specific ERV class (for example, IAPs), using Bowtie2. Next, unpaired read pairs (where one mate matched an ERV sequence but the other could not be aligned) were extracted using samtools and mapped to the mm9 reference genome using Bowtie (allowing only for uniquely mappable reads). This strategy allowed us to anchor each ERV integration site with up to 10 uniquely mappable reads. This strategy allowed us to anchor each ERV integration site with up to 10 uniquely mappable reads on either side of the repetitive sequence. Plus/minus-strand-specific wig coverage tracks were created using IGVTools, extending reads to 500 bp. We took advantage of the fact that left-hand anchor reads mapped exclusively to the plus strand and right-hand anchor reads to the minus strand. Thus, while existing ERVs were demarcated by a plus peak on the left and minus peaks on the right of the repeat sequence, non-annotated integration sites were characterized by a plus peak directly overlapping with a minus peak at the insertion site. Plus and minus peaks were identified separately using the FindOutlierRegion of Java genomics toolkit on split plus and minus wig files. Peak intervals were then intersected to find overlapping plus/minus peaks. Wild-type and KO1 ES cell peaks were intersected and new integration sites were only called if a plus/minus peak did not overlap with a minus or plus peak in the respective control data set. IAP integration sites were validated by genotyping, using primer pairs spanning a ~300 bp region between the IAP LTR and the unique flanking region.

31. Sarma, K. et al. ATRX directs binding of PRC2 to Xist RNA and Polycomb targets. *Cell* **159**, 869–883 (2014).
32. Bilodeau, S., Kagey, M. H., Frampton, G. M., Rahl, P. B. & Young, R. A. SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state. *Genes Dev.* **23**, 2484–2489 (2009).
33. Das, P. P. et al. Distinct and combinatorial functions of Jmjd2b/Kdm4b and Jmjd2c/Kdm4c in mouse embryonic stem cell identity. *Mol. Cell* **53**, 32–48 (2014).
34. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
35. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
36. Liu, T. Use model-based analysis of ChIP-Seq (MACS) to analyze short reads generated by sequencing protein–DNA interactions in embryonic stem cells. *Methods Mol. Biol.* **1150**, 81–95 (2014).
37. Shen, L., Shao, N., Liu, X. & Nestler, E. ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC Genomics* **15**, 284 (2014).
38. Shin, H., Liu, T., Manrai, A. K. & Liu, X. S. CEAS: cis-regulatory element annotation system. *Bioinformatics* **25**, 2605–2606 (2009).
39. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
40. Trapnell, C. et al. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature Biotechnol.* **31**, 46–53 (2013).
Extended Data Figure 1 | H3.3 and H3K9me3 correlate within the mouse repetitive ES cell genome. Related to Fig. 1. Hierarchically (Spearman rank) clustered heat map showing occupancy of histone H3.3 and known heterochromatic histone modification and factors over a comprehensive set of mouse repetitive sequences (see Methods for details). Published data sets used are listed in the Methods section. Data are represented as log2 fold enrichment over matched inputs for each ChIP data set. Repeats with less then 0.01% abundance are omitted.
H3.3 ChIP (FA)

H3K9me3 ChIP (native)

FA crosslinked Input

native Input

H3.3-H3K9m3 Re-ChIP (FA)

Read density over input

normalized input

Distribution of H3.3 and H3K9me3 peaks amongst interspersed repeats
Extended Data Figure 2 | H3.3 and H3K9me3 co-occupy class I and II ERVs. Related to Fig. 1. 

a, Direct comparison of H3.3 enrichment at genic and repetitive sites. Box plot (top) showing enrichment of H3.3 over sets of intervals either representing genic or repetitive elements annotated in the reference genome, using inclusive read mapping. H3.3 ChIP was performed using an H3.3 antibody and formaldehyde (FA) crosslinking in H3.3 wild-type (WT) cell line. H3.3 enrichment is shown as standardized ChIP-seq read density divided by the standardized input read density on a per-interval basis. The width of the box is proportional to the number of intervals in each group. TSS, transcription start sites of highly active genes; K27pro, bivalent promoters. Box plot (bottom) shows the input read density (standardized by scaling to a genome-wide mean of 1), confirming the even representation of unique and repetitive sequences resulting from the inclusive mapping procedure (see Methods for details). Result of one-sided Wilcoxon rank sum test against a set of randomly selected genomic intervals (shuffled) is indicated (***(P < 0.0001). 

b, H3K9me3 enrichment at genic and repetitive sites. H3K9me3 ChIP was performed using MNase digestion under native conditions. Box plot (top) showing enrichment of H3K9me3 over sets of intervals either representing genic or repetitive elements analogous to a. Box plot (bottom) shows the input read density analogous to a. Result of one-sided Wilcoxon rank sum test against a set of randomly selected genomic intervals (shuffled) is indicated (***(P < 0.0001. 

c, Sequential H3.3 and H3K9me3 (re)-ChIP at genic and repetitive sites. Boxplots showing enrichment of Re-ChIP inclusive read mapping relative to an input control. Result of one-sided Wilcoxon rank sum test against a set of randomly selected genomic intervals (shuffled) is indicated (***(P < 0.0001). 

d, Co-occupancy of H3.3 and H3K9me3 at specific classes of ERVs. H3.3 and H3K9me3 peak intervals were independently intersected with annotated ERVs and their co-occurrences within the same ERV were evaluated. L1Md_F (full) is a subset of L1Md_F, comprising only full length repeats (>5 kb). All pie charts include total number of intervals for each family that had none, (at least) one H3.3 peak (H3.3 only), or H3K9me3 peak(s) (H3K9me3 only), or at least one of each (H3.3+H3K9me3).
Extended Data Figure 3 | Generation of H3.3-isoform-specific antibodies. Related to Fig. 1. a, Schematic of amino acid sequence differences for the canonical histones H3.1 and H3.2 versus the histone variant H3.3. H3.3 differs from H3.2 or H3.1 at only 4 or 5 amino acids, positions 31, 87, 89, 90 and 96, as indicated. b, Immunoblot against recombinant histones using the final purified antibody (Millipore 09-838), confirming specificity of the H3.3-isoform-specific antibody. c, ChIP-qPCR analysis of H3.3 enrichment at various repeat regions in control and H3.3-knockout ES cells. Error bars represent s.d. from one experiment (n = 3). d, ChIP-seq enrichment of H3.3 at repetitive regions of the mouse genome in control and H3.3-knockout ES cells. Data are represented in a heat map of log_2 fold enrichment (red) or depletion (blue) over a matched input.
Extended Data Figure 4 | H3.3 is enriched in regions flanking ERVs and orphan LTRs. Related to Fig. 1. a, ChIP-seq density heat maps for unique sites flanking full-length IAP ERVs (n = 800) rank ordered by H3K9me3 enrichment. Colour intensity represents normalized and globally scaled tag counts. b, H3.3 (top) and H3K9me3 (bottom) enrichment over regions flanking IAP, ERVK10C, ETn ERVs and L1 elements. H3.3 ChIP-seq was performed with FA crosslinking, H3K9me3 ChIP-seq under native conditions. Average profiles were aligned and aggregated at the 5′ and 3′ boundaries of hundreds of annotated elements from standardized unique read count coverage tracks. The profiles are directional with the 5′ ends on the left and 3′ end on the right. c, H3.3 (top) and H3K9me3 (bottom) enrichment over regions flanking single (so-called orphan) IAP LTRs, ~500 bp. Orphan LTRs are the result of a recombination event between two LTRs—usually the 3′ and 5′ LTRs of the same ERV—effectively deleting the internal coding sequence. Approximately 600 full-length LTRs (~500 bp) enriched in H3.3 and H3K9me3 were identified in the mouse genome and aggregated for the profiles.
Extended Data Figure 5 | H3.3 at IAPs is not associated with transcription, DNase I or MNase sensitivity. Related to Fig. 1. a, Direct comparison of chromatin properties at TSSs of highly expressed genes and IAP ERVs. Box plots showing (from left to right) comparable enrichment of H3.3; DNase I sensitivity; MNase sensitivity; elongating RNAP2 occupancy. MNase data sets are from a recent study, showing H3.3 localizing to MNase hypersensitive regions such as active promoters25. In this study, MNase sensitivity was assessed by sequencing nucleosomes released under mild (‘short’) or extensive (‘long’) MNase digestion conditions; MNase hypersensitive sites were shown to be specifically enriched by mild MNase digestion, whereas long digestion released chromatin more evenly25. b, Comparison of kinetics of H3.3 incorporation26 at the TSS of highly expressed genes and IAP ERVs; as control, a randomized set of intervals is shown.
### a) H3.3-HA-ChIP (Goldberg. et. al)

|        | TSS | IAP | ETn | MusD | L1Md_A | L1Md_F | 5'UTR | shuffled |
|--------|-----|-----|-----|------|--------|--------|-------|----------|
| Read density over input |     |     |     |      |        |        |       |          |
| ESC NPC | ESC NPC | ESC NPC | ESC NPC | ESC NPC | ESC NPC | ESC NPC | ESC NPC | ESC NPC |

### b) H3K9me3

|        | TSS | IAP | ETn | MusD | L1Md_A | L1Md_F | 5'UTR | shuffled |
|--------|-----|-----|-----|------|--------|--------|-------|----------|
| Read density over input |     |     |     |      |        |        |       |          |
| ESC NPC | ESC NPC | ESC NPC | ESC NPC | ESC NPC | ESC NPC | ESC NPC | ESC NPC | ESC NPC |

### c) ESET

|        | IAP | ETn | MusD | L1Md_A | L1Md_F | 5'UTR | shuffled |
|--------|-----|-----|------|--------|--------|-------|----------|
| Read density over input |     |     |      |        |        |       |          |
| ESET WT KO | ESET WT KO | ESET WT KO | ESET WT KO | ESET WT KO | ESET WT KO | ESET WT KO | ESET WT KO |

### SUV39h1/2

|        | IAP | ETn | MusD | L1Md_A | L1Md_F | 5'UTR | shuffled |
|--------|-----|-----|------|--------|--------|-------|----------|
| Read density over input |     |     |      |        |        |       |          |
| SUV39h1/2 WT DKO | SUV39h1/2 WT DKO | SUV39h1/2 WT DKO | SUV39h1/2 WT DKO | SUV39h1/2 WT DKO | SUV39h1/2 WT DKO | SUV39h1/2 WT DKO | SUV39h1/2 WT DKO |
Extended Data Figure 6 | H3.3 and ESET-dependent H3K9me3 enrichment at IAPs is lost upon differentiation. Related to Fig. 1. a, b, Comparison of H3.3 (a) and H3K9me3 (b) enrichment at the TSS of highly expressed genes and various repeat classes in ES cells and NPCs using inclusive read mapping. H3.3 ChIP was performed using a genomic knock-in tagged H3.3B–HA and FA crosslinking12. H3K9me3 ChIP was performed using FA crosslinking17. Enrichment is shown as standardized ChIP-seq read density divided by the standardized input read density on a per interval basis. Result of one-sided Wilcoxon signed rank test (NPCs versus ES cells) are shown (***, P < 0.0001; ***, P < 0.0005; ** P < 0.005; * P < 0.05; no annotation = not significant). c, Levels of H3K9me3 enrichment in control and ESET-knockout ES cells (top) or control and SUV39h1/2-knockout ES cells27 (bottom) at various repeat classes. DKO, double knockout. Data are represented as in a and b.
Extended Data Figure 7 | Contribution of DAXX, ATRX, KAP1 and ESET to H3.3 enrichment at ERVs. Related to Fig. 2. a–d, ChIP-qPCR analysis of H3.3 enrichment at various repeat regions in control and ATRX-knockout (a), DAXX-knockout (b), KAP1-knockout (c) and ESET-knockout (d) ES cells. Error bars represent s.d. from one experiment (n = 3). Data are representative of two independent ChIP experiments. e, ChIP-seq enrichment of KAP1 and H3.3 in control and KAP1-knockout ES cells at repetitive regions of the mouse genome. Data are represented in a heat map of log2 fold enrichment (red) or depletion (blue) over a matched input. f, Loss of H3.3 reduces nuclear DAXX levels. Immunoblot from whole-cell extracts (WCE) or nuclear extracts (NE) in the presence and absence of H3.3. Asterisk denotes cross-reacting band. g, ChIP-seq enrichment of KAP1 and DAXX in control and H3.3-knockout ES cells. Data are represented as in e. Note the different colour scale used for KAP1 and DAXX.
Extended Data Figure 8 | Effects of H3.3 and corepressor complex depletion on H3K9me3 heterochromatin. Related to Fig. 3. a, Positive correlation of H3.3 and H3K9me3 at IAP ERVs. H3.3 ChIP-seq enrichment at 800 unique IAP flanking regions (see Fig. 1e) was binned into three groups by their H3K9me3 ChIP-seq enrichment (low, medium and high). Wilcoxon rank sum test (****P < 0.0001). b, Immunoblot from ES cell whole-cell lysates in the presence and absence of H3.3. c, H3.3, H3K9me3 and KAP1 ChIP-seq density heat maps for peaks classified as H3.3 only (n = 60,925), both H3.3 and H3K9me3 (n = 18,605), or H3K9me3 only (n = 54,204) in control and H3.3-knockout ES cell. Five-kilobase intervals around peak centres are shown. Colour intensity represents normalized and globally scaled tag counts. d, Levels of H3K9me3 at IAP, ETn, MusD ERVs and LINE elements in control and KAP1-knockout ES cells (top) and control and H3.3-knockout ES cells (bottom). Box plots show enrichment over matched input. e–h, ChIP-qPCR analysis of H3K9me3 at various repeat regions in control and KAP1-knockout (e), ESET-knockout (f), ATRX-knockout (g) and DAXX-knockout (h) ES cells. Error bars represent s.d. from one experiment (n = 3). Data are representative of two independent ChIP experiments.
Extended Data Figure 9 | Global effects of H3.3 depletion. Related to Fig. 3. 

a. H3.3 transcript levels in control, H3.3-knockdown and H3.3-knockout ES cells. Data are represented as mean expression relative to Gapdh ± s.d. (n = 3). 
b. Relative gain/loss upon H3.3 knockdown of H3K9me3, H3.3 and total H3 are shown over a section of chromosome 10 containing the highly transcribed Rps12 gene and several ERVs. Gain/loss tracks are calculated by subtracting the respective control from H3.3 KD1 tracks, both standardized to a global mean of 1. Note that H3.3 ChIP-seq data in KD1 cells represents the remaining 10% H3.3. The global loss of H3.3 is not directly apparent from the track due to the necessary normalization of the data. The H3.3 difference track thus does not indicate the global loss of H3.3 but merely represent the relative redistribution of the remaining H3.3 from active genes (Rps12) towards repetitive sequences. 
c. Levels of H3.3 and H3 and IAP, ETn, MusD, and the TSS of highly expressed genes in control, H3.3-knockdown and H3.3-knockout ES cells. Box plots show enrichment over matched input. 
d. Incorporation of exogenous, constitutively expressed, H3.3 and H3.2 added back into H3.3-knockdown or H3.3-knockout ES cells. H3.2 cannot substitute for H3.3 at repetitive ERVs but is efficiently incorporated at sites of active transcription. ChIP-seq was performed on lentivirally integrated H3.3–haemagglutinin (HA) and H3.2–HA in H3.3 KD1 and H3.3 KO1. 
e. ChIP-seq density heat maps for peaks classified as enriched with both H3.3 and H3K9me3 (n = 18,605) or H3.3 only (n = 60,925). Colour intensity represents tag counts scaled and normalized globally. Five-kilobase intervals around peak centres are shown. 
f. Quantification of H3.3–HA and H3.2–HA add-back in H3.3-knockout enrichment at low and highly expressed genes, as well as the TSS (±1 kb) of the latter, IAP, ETn and MusD ERVs, and full-length LINE elements and their 5′ promoter regions. Data are represented as enrichment over input.
Extended Data Figure 10 | ERV reactivation upregulates adjacent genes and may be linked to unbalanced chromosomal translocations. Related to Fig. 4. 

a, Repetitive elements associated with genes in Fig. 4b. Elements that were found either within or nearby the transcription unit are listed and the closest distance of an ERV to an exon is given (accounting for the possibility that ERVs could initiate a partial transcript from an alternative start site).

b, Newly annotated sites of IAP integration in wild-type and H3.3 KO1 are indicated on karyogram. c, Karyotype analysis of wild-type and H3.3-knockout ES cells. Abnormal karyotype is indicated by arrows. All analysed cells in H3.3 KO1 had a small reciprocal translocation between chromosomes 2q and 6q and an unbalanced translocation between chromosomes 6 and 17 resulting in partial gain of chromosomal segment 6qD to 6qG and partial loss of chromosomal segment 17qE2 to 17qE5. Approximately 45% of the cells had chromosomal breaks or gaps (1–2 per cell). Approximately 45% of the H3.3 KO2 ES cells had a duplication of the segment 8qC to 8qD resulting in partial gain of this segment.