

ARTICLE

<https://doi.org/10.1038/s41467-018-08006-y>

OPEN

Transposable elements are regulated by context-specific patterns of chromatin marks in mouse embryonic stem cells

Jiangping He^{1,2,3,4}, Xiuling Fu⁵, Meng Zhang^{1,2,3,4,6}, Fangfang He⁵, Wenjuan Li^{2,6}, Mazid Md. Abdul^{1,2,3,4,6}, Jianguo Zhou^{1,2,3,6}, Li Sun⁵, Chen Chang⁵, Yuhao Li⁵, He Liu¹, Kaixin Wu¹, Isaac A. Babarinde⁵, Qiang Zhuang^{5,7}, Yui-Han Loh^{8,9}, Jiekai Chen ^{1,2,3}, Miguel A. Esteban ^{1,2,3,6} & Andrew P. Hutchins ⁵

The majority of mammalian genomes are devoted to transposable elements (TEs). Whilst TEs are increasingly recognized for their important biological functions, they are a potential danger to genomic stability and are carefully regulated by the epigenetic system. However, the full complexity of this regulatory system is not understood. Here, using mouse embryonic stem cells, we show that TEs are suppressed by heterochromatic marks like H3K9me3, and are also labelled by all major types of chromatin modification in complex patterns, including bivalent activatory and repressive marks. We identified 29 epigenetic modifiers that significantly deregulated at least one type of TE. The loss of *Setdb1*, *Ncor2*, *Rnf2*, *Kat5*, *Prmt5*, *Uhrf1*, and *Rrp8* caused widespread changes in TE expression and chromatin accessibility. These effects were context-specific, with different chromatin modifiers regulating the expression and chromatin accessibility of specific subsets of TEs. Our work reveals the complex patterns of epigenetic regulation of TEs.

¹Key Laboratory of Regenerative Biology of the Chinese Academy of Sciences and Guangdong Provincial Key Laboratory of Stem Cell and Regenerative Medicine, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, 510530 Guangzhou, China. ²Joint School of Life Sciences, Guangzhou Medical University and Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, 511436 Guangzhou, China. ³Guangzhou Regenerative Medicine and Health-Guangdong Laboratory (GRMH-GDL), 510530 Guangzhou, China. ⁴University of Chinese Academy of Sciences, 100049 Beijing, China. ⁵Department of Biology, Southern University of Science and Technology, 1088 Xueyuan Lu, 518055 Shenzhen, China. ⁶Laboratory of RNA, Chromatin, and Human Disease, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, 510530 Guangzhou, China. ⁷State Key Laboratory of Medicinal Chemical Biology and College of Life Sciences, Nankai University, 300071 Tianjin, China. ⁸Epigenetics and Cell Fates Laboratory, A*STAR Institute of Molecular and Cell Biology, 61 Biopolis Drive Proteos, Singapore 138673, Singapore. ⁹Department of Biological Sciences, National University of Singapore, Singapore 117543, Singapore. Correspondence and requests for materials should be addressed to A.P.H. (email: andrewh@sustc.edu.cn)

The modification of histones is an elaborate system to regulate gene expression, and provides an epigenetic landscape for the cell-type-specific interpretation of the genome. Yet, the major class of genomic elements in the cellular genome are not genes, but transposable elements (TEs), including endogenous retroviruses (ERVs)¹. TEs were originally thought of as genetic parasites with roles in human disease², but TEs are now understood to contribute to normal biological processes³. There are many examples of exapted TEs that have become host cell genes, such as the RAG enzymes which are essential for antibody and T cell receptor recombination, or the Syncytin genes which are involved in placental development^{3,4}. TEs can be transcribed to produce RNA, and have contributed to the evolution of long non-coding RNAs, microRNAs, and circular RNAs⁵. In addition, TEs often mimic host cell functions by incorporating cis-regulatory elements⁶, which can recruit transcription factors (TFs) to promote TE activity. This was initially observed for the TF repressor REST⁷, but has since been observed for a wide-range of TFs^{8–10}. TEs copy themselves in the genome and are co-opted to form new regulatory elements^{1,11,12}, and contribute to the rewiring of gene regulatory networks⁶. However, much of this data on exaptation of TEs is derived from genomic data and there is argument over how much is functional⁶.

TEs/ERVs are silenced by a range of molecular mechanisms, including heterochromatin formation^{13–15}, mRNA editing¹⁶, and DNA methylation¹⁷. DNA methylation is thought to be the dominant suppressive mechanism in somatic tissues¹⁷. However, DNA is globally demethylated during early embryonic development, and TEs are released from repression in a controlled, stage-specific manner^{18,19}. The TEs are then free to compete with the epigenetic suppression mechanisms to duplicate themselves and enter the germ line²⁰. Consequently, there is a delicate balance between the beneficial effects of TEs, and their deleterious effects on genome integrity^{6,21,22}.

TEs are suppressed in embryonic cells in a process that is well described for ERVs. Zinc-finger proteins (ZFPs) bind to specific sequences in ERVs²³, recruit the adaptor protein TRIM28/KAP1, and the histone H3K9me3 methyltransferase SETDB1 to silence TEs^{13,24–28}. In addition to H3K9me3, there are other modes of epigenetic suppression of TEs²⁹, such as the methylation of H4K20me3³⁰, H3K27me3³¹, and H4R3me2²⁰, the biotinylation and sumoylation of H2A, H3, and H4 histones^{32,33}, and the deposition of the histone variant H3.3³⁴. It is clear that the epigenetic system is regulating TEs^{4,29,35–38}, however, there are at least 1100 distinct types of TE, comprising millions of genomic copies, for which the epigenetic regulation is unclear.

Here, we reveal that TEs are marked by chromatin modifications in complex patterns. Of the 32 chromatin marks we explored, 22 were enriched on at least one TE type. We find evidence not only of repressive marks, but widespread marking of TEs by activatory marks, including bivalent marking of TEs by repressive H3K9me3, and activatory H3K27ac chromatin marks. When we knocked down a panel of chromatin regulators, 29 of 41 led to the deregulation of expression of at least one type of TE. The loss of *Setdb1*, *Ncor2*, *Rnf2*, *Kat5*, *Prmt5*, *Uhrf1*, and *Rrp8* led to widespread changes in the expression of TEs, and a corresponding change in chromatin accessibility. Finally, we explore the consequences of these observations by showing that the loss of chromatin modifiers induces a gene expression program similar to totipotent 2 cell (2C) embryos, characterized by the upregulation of MERVL TEs and 2C-specific genes. Overall, our data suggests that the chromatin modifying system manages the expression of TEs by context-specific deposition of chromatin marks that regulate the chromatin state and expression level of TEs.

Results

TEs are marked by overlapping patterns of chromatin marks.

To explore the chromatin code for the suppression of TEs, we uniformly reanalyzed mouse ESC ChIP/epigenetic-seq data for 32 chromatin marks, including histone methylation, acetylation, ubiquitination, and variant histones, along with DNA 5-methylcytosine (5mC), 5-hydroxymethylcytosine (5hmC), and Assay for Transposase-Accessible Chromatin (ATAC)-seq (Supplementary Data 1). Data was quality controlled by removing poorly correlating biological replicates (See Methods and Supplementary Figure 1a). A concern in the analysis of repeats is the use of multimapped reads. In our analysis, for multimapped reads only the highest scoring alignment was kept, and was randomly assigned to a TE copy. To reduce bias due to random distribution of reads across TE copies we considered TEs as metagenes that describe the entire set of TE copies across the genome. Below, we show some example genomic views, but it should be kept in mind that the analysis of individual TE copies can be ambiguous.

The ChIP-seq data comes from serum + LIF grown ESCs, which have relatively high levels of DNA methylation, particularly at Intracisternal A-particle (IAP) TEs^{38–40}. However, many TEs have demethylated DNA, and ESCs use DNA methylation-independent mechanisms to suppress TEs. Knockout of all three DNA methyltransferases led to upregulation of mostly IAP-family TEs only⁴⁰, and this upregulation is rapidly compensated for by small RNA and polycomb-mediated repression^{38,40}. Indeed, only 65/679 TE types had >2-fold enrichment of 5mC DNA methylation, and most of the methylated TEs were IAP or IAP LTRs (Long terminal-repeats), and a mixture of other ERVs, for example, BGLII_Mus (Fig. 1a, b, Supplementary Figure 1b and Supplementary Data 2).

The second major mechanism for the regulation of TEs in ESCs is the formation of heterochromatin, mediated by H3K9me3 and H4K20me3^{14,41,42}. We found that 87/679 TE types were >2-fold-enriched with H3K9me3, and 72/679 were marked by H4K20me3. H3K9me3 and H4K20me3 tended to co-occur, and they both marked 58/679 TE types (Supplementary Data 2). Examples are shown for the ERVK IAPLTR4_I, and IAPEY-int which were enriched for H3K9me3 and H4K20me3 (Fig. 1a, b).

Of the remaining TEs, some had simple patterns of chromatin modifications, such as MERVL TEs that were enriched only for H3K56ac (Fig. 1b). In total 41/679 TE types were enriched with a single chromatin mark (Supplementary Data 2). The most common pattern was for multiple chromatin marks on the TEs, as 117/679 had 2 or more chromatin marks, and, at a relaxed enrichment threshold of >1.5 fold, 198/679 TE types were enriched with at least 2 chromatin marks (Supplementary Data 2). Examples include RLTRTN_Mm, which was enriched for 14 of the 32 chromatin marks we profiled (Fig. 1a, b). Interestingly, TEs were marked by chromatin patterns associated with other genomic features. Many TEs were enriched with chromatin marks indicative of enhancers, such as H3K4me1 which marked 23/679 TE types. For example, a 3.0 fold-enrichment of H3K4me1 was found on the ERVK RLTR9D2 (Supplementary Data 2). The promoter mark H3K4me3 was enriched at 14/679 TEs, including the LINE (long interspersed nuclear element) LIMd_Gf (7.2-fold-enriched), and the ERVK RLTRTN_Mm (2.9 fold-enriched) (Fig. 1a, b and Supplementary Data 2). ERVKs/ERVs were enriched with activatory chromatin marks, including histone acetylations, such as H3K27ac (RLTR13G >6.2-fold-enriched) and H2B120ac (MMERGLN_LTR >4.6 fold-enriched) (Fig. 1a, c). In total 75/679 TE types were marked by one or more acetylated histone mark and 42/679 by H3K27ac. To gain a global view of TE chromatin, we clustered the patterns of chromatin marks on TEs

environment, it is unclear if they are directly regulating the expression and chromatin state of these TEs.

Chromatin modifiers regulate type-specific TE expression. To explore how CMs control TE expression, we selected a panel of reader, writer, and eraser CMs⁵³ (Supplementary Figure 4a) to knockdown, based on the clusters we detected in Fig. 2a, b and Supplementary Figure 3. As cells can compensate for the derepression of TEs by activating alternative suppressive mechanisms^{38,40}, we chose to knockdown over a short time period of four days. This decision has two benefits. First, it allowed us to study those CMs that are lethal if knocked down for a prolonged time. Second, it should minimize changes in cell fate which would complicate the results, as the cells reconfigure their chromatin in response to differentiation. In total, we knocked down 41 CMs, and subjected them to RNA-seq. To rule out a change in cell fate we measured the expression of a panel of pluripotency and differentiation marker genes. Their expression indicated that most knockdowns had little influence on cell fate (Supplementary Figure 4b, c and Supplementary Data 4). Principal component analysis gave similar results, as the knockdown samples did not diverge from the control samples (Supplementary Figure 4d). Analyzing the number of differentially expressed genes came to a similar conclusion, as minimal numbers of genes were differentiated in each knockdown (Supplementary Figure 4e). There was, however, two notable exceptions. Knockdown of *Mcrs1* showed an upregulation of chromatin-related genes and a loss of germ cell character, and *Chd4*, which showed signs of differentiation towards the immune system (Supplementary Figure 4f). We include *Chd4* and *Mcrs1* in the analysis below, but it should be noted that these two knockdowns are differentiating, and interpretation of their effect on TEs is complicated. For all other knockdowns, there was little evidence of differentiation, and no loss of pluripotency. This data set provides a system to look at the short-term deregulation of TEs when CMs are lost, in the absence of major cell type changes.

We next explored if these CM knockdowns can deregulate TE expression. We aligned the RNA-seq data to LINE, SINE (short interspersed nuclear element), LTR (ERV), and DNA class TEs (See Methods). On average 1–5% of reads would map to TEs (Supplementary Figure 5a), and we observed higher percentages of reads mapping to TEs in the knockdown samples (most >2% versus 1.6% in the *shLuc* control), suggesting a higher level of TE-derived RNA (Supplementary Figure 5a). As expected, knockdown of *Setdb1* caused a robust upregulation of many types of TE (Fig. 3a, b). In addition, 29/41 other CM knockdowns significantly up or down-regulated at least 1 type of TE (Fig. 3a), and 8 knockdowns upregulated at least 1 type of TE >4 fold (Fig. 3b). We used RT-qPCR to validate a selection of TE expression changes in knockdowns with shRNAs against *Rnf2*, *Setdb1*, *Ncor2*, *Prmt5*, and *Ash2l* (Supplementary Figure 5b).

We selected eight CM knockdowns (and *shNcor1*, as a control) that deregulated 20 or more types of TE for further analysis. Western blotting for the pluripotency proteins OCT4 and NANOG confirmed that the eight selected knockdowns were not inducing differentiation (Supplementary Figure 5c). Knockdown of these eight CMs led to a variety of effects on the expression of TEs. Knockdown of *Setdb1* deregulated only a few TE types, but it led to the largest change in the magnitude TE expression. Conversely, knockdown of *Ncor2* or *Rnf2* deregulated more types of TE, but the magnitude of upregulation was reduced (Figs. 3a, b). There was surprisingly little overlap in the TE types that were upregulated, and most TE types were specific to a single knockdown (Supplementary Figure 5d, e).

TEs can span a wide-range of evolutionary time, and some TEs are active, or recently active, whilst others are inactive and

ancient. Importantly, young and old TEs are regulated differently¹². We measured the evolutionary age of TEs, and found that younger TEs tended to have higher enrichment of chromatin marks, although the correlation was modest (Supplementary Figure 1b). Younger TEs were more likely to be deregulated in a CM knockdown, although, again, the correlation was modest (Supplementary Figure 5f). These results suggest that younger TEs are more likely to be under active epigenetic regulation.

We next investigated the relationship between TE expression changes and chromatin marks. There was no global alteration in the levels of histone modifications, with the exception of the *Rnf2* knockdown, which reduced H2AK119ub and H3K27ac (Fig. 3c). This indicated that the changes in histone modifications were context-specific, as we saw previously with the overexpression of HDAC3 in ESCs which did not alter global H3K27ac levels, but did alter context-specific H3K27ac sites⁵⁴. There was a direct relationship between CMs and chromatin mark. For example, knockdown of *Setdb1* led to significant upregulation of TEs high in H3K9me3, whilst knockdown of the DNA methylation cofactor *Uhrf1* led to upregulation of TEs high in 5mC, particularly IAPs (Fig. 3d, e, and Supplementary Figure 5g). Knockdown of *Prmt5* has previously been demonstrated to upregulate LINE L1 and IAP elements in primordial germ cells, and ground state 2i + LIF ESCs²⁰, and we observe the same in serum + LIF grown ESCs (Fig. 3f, g, h). The LINE and IAP TEs upregulated in the *Prmt5* knockdown were enriched with H4R3me2, the catalytic target for PRMT5, indicating a link between PRMT5 and H4R3me2 at TEs (Figs. 1a, 3g, and Supplementary Figure 1b, d, e). In addition to these associations, we also found new links between CMs and chromatin marks. Knockdown of *Hdac2* and *Hdac5* upregulated TEs enriched with H3K56ac, H3K9me3, and H4K20me3, and knockdown of *Ncor2* upregulated TEs enriched with H3K9me3 and H3K56ac (Supplementary Figure 5g, h). NCOR2 is a co-repressor⁵⁴, so the knockdown of *Ncor2* activating TEs may seem unusual, but it is not uncommon for co-repressors to have activatory function. For example, the co-repressor SIN3A, a binding-partner for NCOR2, can function as a co-activator in ESCs⁵⁵. Overall, these observations indicate a complex code of CMs regulating sets of TEs by altering chromatin marks.

Chromatin modifiers alter chromatin accessibility at TEs. The loss of CMs leads to context-specific deregulation of TEs, which in many cases was associated with specific chromatin marks. However, whilst many CMs are specific epigenetic writers/erasers (e.g., SETDB1 writes H3K9me3, RNF2 writes H2AK119ub), they are also members of large multi-protein complexes involved in multiple regulations (e.g., RNF2 is part of polycomb repressor complex 1), or there is substantial redundancy (e.g., SETDB1 and SUV39H1 both target H3K9me3). Consequently, the number of potential chromatin mark changes caused by the loss of any single CM can be large. To gain an overview of changes in chromatin in response to the knockdown of a CM, we used ATAC-seq^{56,57}, to probe TE chromatin accessibility after knockdown. Of the 142,119 non-redundant peaks in the ATAC-seq data, we could detect changes in 15,076 loci (Supplementary Figure 6a, b). Chromatin that became accessible in each shRNA knockdown showed matching upregulation of genes within 10 kb of the ATAC-seq peak (Supplementary Figure 6c, d). The changes in chromatin accessibility were also reflecting chromatin marks. For example, ATAC peaks that became accessible when *Prmt5* was knocked down were enriched for H4R3me2. A similar pattern was observed for *Setdb1* knockdown and H3K9me3 (Supplementary Figure 6e). These results show that changes in accessibility are a suitable proxy for chromatin changes.

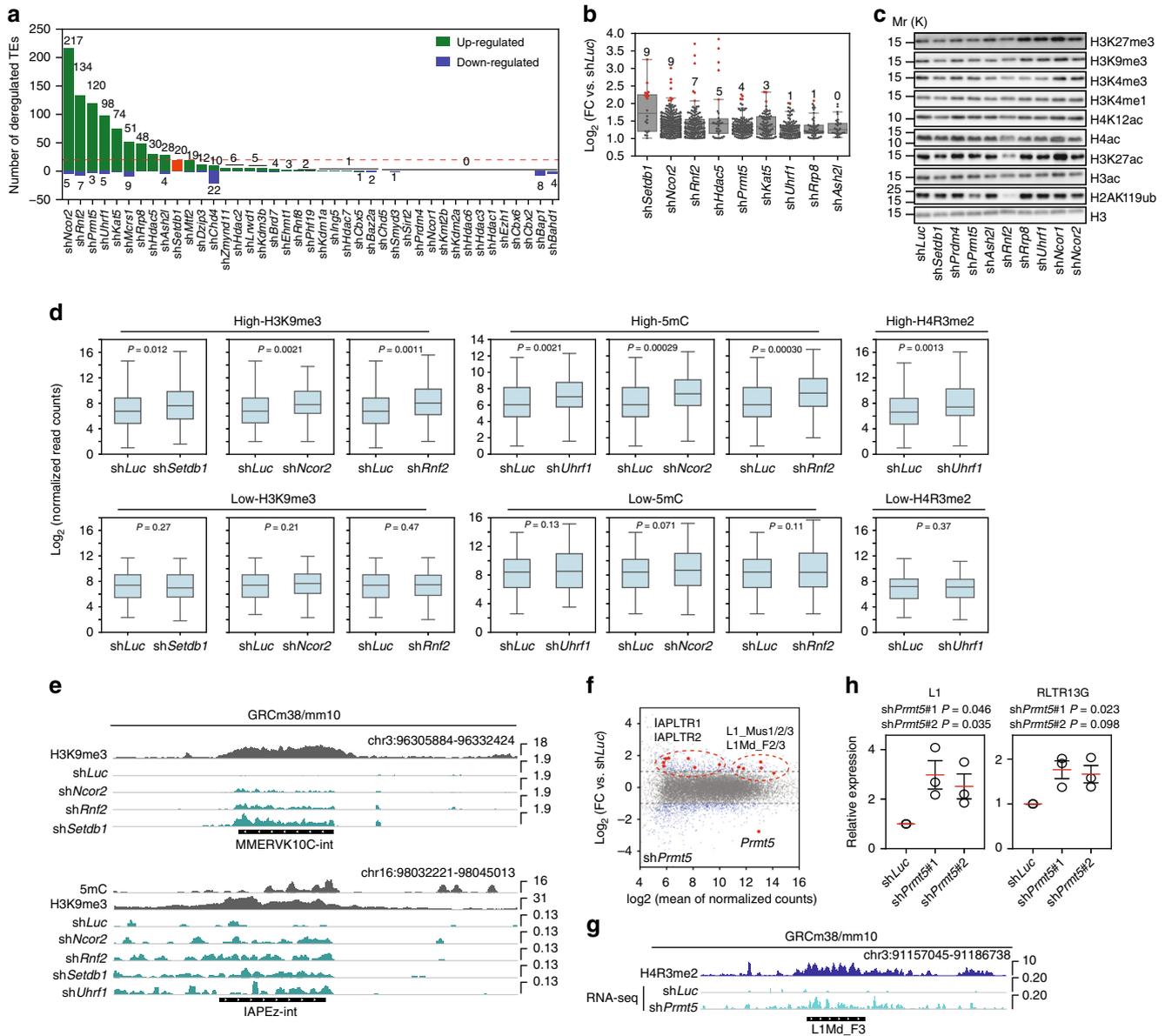


Fig. 3 Chromatin modifiers regulate the type-specific expression of TEs. **a** Number of significantly deregulated TE types for the indicated shRNA knockdown (P values < 0.05 , Benjamini & Hochberg corrected, absolute fold-change > 2 , relative to *shLuc*). See Supplementary Data 1 for detailed sample information. **b** Boxplots of \log_2 (fold-change (FC) versus *shLuc*) for all significantly deregulated TE types. In the boxplot, outliers represent the mean of a TE type. TE types significantly upregulated > 4 fold are indicated in red and the number is marked. **c** Western blot of chromatin marks, from the indicated shRNA knockdowns. Molecular weight is indicated as Mr (K). This experiment was performed twice with similar results. **d** Boxplots of the expression level for all TE types with either High or Low levels of the indicated chromatin mark. High was defined as those TE types with > 2 -fold enrichment of the indicated chromatin mark. Low sites were selected by ranking the TEs by fold enrichment, and taking TEs from the bottom of the list until it was the same length as the High list. **e** Genomic views of MMEVK10C-int and IAPEZ-int showing chromatin marks and RNA-seq in the indicated shRNA knockdown. **f** MA-scatter (log-fold-change, mean average expression) plot showing *shPrmt5* knockdown relative to *shLuc*. Blue dots are significantly different genes or TEs (P value < 0.05 , Benjamini & Hochberg corrected, absolute fold-change > 2 , relative to *shLuc*). Red dots indicate IAP or LINE1 elements, or the *Prmt5* gene. RNA-seq data was from 2 biological replicates. **g** Genome view of H4R3me2 ChIP-seq data and *shPrmt5* RNA-seq data for L1Md_F3 LINE1. **h** RT-qPCR expression of the LINE1 L1Md_F2, and RLTR13G, relative to *shLuc* control knockdown. P values are from a one-tailed Student's t -test. The circles are the mean of three technical replicates, the red bar is the mean of the three biological replicates, and the error bars are the standard error of the mean. Source data are provided as a Source Data file. Primers used are described in Supplementary Data 7. For the boxplots, the midline indicates the median, boxes indicate the upper and lower quartiles and the whiskers indicate 1.5*interquartile range

We measured the chromatin accessibility at TEs. 153 TE types showed an increase in chromatin accessibility (Fig. 4a, Supplementary Data 5). These changes were mainly specific to a single knockdown, as 108 TE types were uniquely upregulated, whilst 42 TE types were upregulated by more than one shRNA knockdown (Fig. 4a, b, and Supplementary Data 5). Importantly, opening of chromatin at TEs, was matched by significant upregulation of the

corresponding RNA expression (Fig. 4c). This pattern extended to individual types of TE, for example, the chromatin opened and RNA was expressed at MT2_Mm, when *Rnf2* was knocked down (Fig. 4d). Similarly, at LINE1 L1Md_F2 TEs, chromatin opened and expression was upregulated when *Prmt5* was knocked down (Fig. 4d, e). This suggests a direct relationship between L1Md_F2, and regulation of expression by PRMT5 through modification of

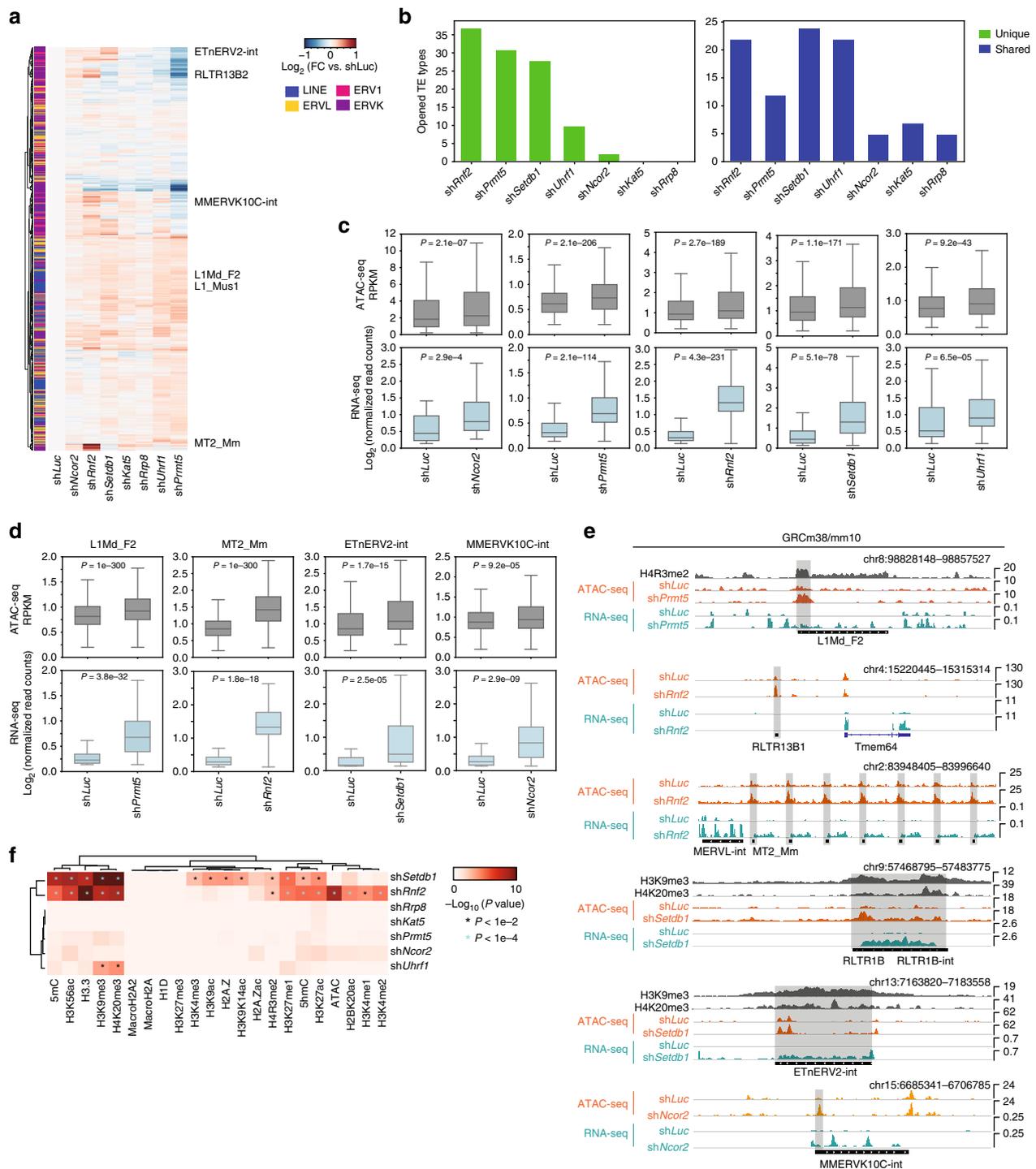


Fig. 4 Loss of CMs opens chromatin at specific TEs. **a** Heatmap of the fold-changes in chromatin accessibility (measured by ATAC-seq) at TE types. The list of TE types are those that have 20% more or 20% less open or closed chromatin. Selected TE types are indicated, and the full table of all TE chromatin changes is in Supplementary Data 5. **b** Bar chart of the number of TE types that acquire open chromatin in a unique shRNA knockdown (left), or in more than one shRNA knockdown (right). **c** Boxplots showing the relationship between chromatin (ATAC-seq, top rows) and expression (RNA-seq, bottom rows). The boxplots contain all TE types that acquire open chromatin, for the selected shRNA knockdowns. P values were derived from a Mann-Whitney U test. **d** Boxplots of the levels of ATAC-seq (top rows, dark gray) and RNA-seq (bottom rows, light gray) for selected TE families for the indicated CM knockdown, for the indicated type of TE. P values were from a Mann-Whitney U test. **e** Example genome views of individual genomic TEs showing specific chromatin marks, expression (RNA-seq), and chromatin accessibility (ATAC-seq). The examples show TEs that acquire open chromatin and expression in the indicated shRNA knockdown. Examples include the LINE1 elements (L1Md_T, L1Md_A), ERVL (MERVL-int), MT2_Mm, ERVKs (RLTR13B1, ETnERV2-int, and MMERVK10C-int). **f** Enrichment of chromatin marks at TE types that become open in response to the indicated shRNA knockdown. Significance is derived from a Fisher exact test, $P < 0.01$ and is indicated with an asterisk. For the boxplots, the midline indicates the median, boxes indicate the upper and lower quartiles and the whiskers indicate 1.5*interquartile range

H4R3me2 (Supplementary Figure 1d). A similar relationship was observed for the H3K9me3 enriched ETnERV2-int which acquired open chromatin and RNA expression when *Setdb1* was knocked down (Fig. 4d, e). To explore these relationships, we measured the correlation between the chromatin marks and changes in accessibility (Fig. 4f). The clearest relationships were between *shSetdb1* and heterochromatin marks H3K9me3 and H4K20me3, and other repressive marks, H3.3 and 5mC (Fig. 4f). Knockdown of *Rnf2* also correlated with repressive epigenetic marks (Fig. 4f). Overall, these results demonstrate that the loss of CMs leads to the alteration of chromatin at TEs, and a matching upregulation of the expression of TEs. There is a relationship between the CMs target, and the chromatin modifications marking the TEs, although due to the nature of CMs as members of multi-protein complexes, a direct relationship was not always observable.

Knockdown of CMs activates early embryonic gene expression.

TEs are expressed and their chromatin is regulated in a phase-specific pattern in the early embryo^{18,58}. TE expression has become a useful marker for early embryonic gene expression

programs in ESCs, as MERVL expression marks a small sub-population of 2C-like cells, that have totipotent properties²⁶, a capability ESCs lack. Knockdown of CMs can increase the fraction of these cells^{15,59}, but the relationship between CM, chromatin, MERVLs, and the 2C gene expression program remains unclear.

To shed light on the regulatory network controlling 2C-like cells, we interrogated our shRNA knockdowns for activation of a 2C-like gene expression program. Downregulation of *Rnf2*, *Brd7*, and *Hdac5* led to the upregulation of MERVLs, and a corresponding increase in 2C-stage embryonic genes, such as *Zscan4c* (Fig. 5a, b). To assess the type of embryonic gene program being activated we measured the expression of the significantly upregulated genes in each of the knockdowns, against RNA-seq data from a range of embryonic stages⁶⁰. Genes upregulated in the *Rnf2*, *Hdac5*, and *Brd7* knockdowns were significantly highly expressed in 2C embryos, but not in other embryonic stages (Fig. 5c). Knockdown of *Setdb1* did not have this effect (Fig. 5c). This pattern could be confirmed by measuring the expression level of the top 200 upregulated genes (versus ESCs) in 2C embryos (Fig. 5d). A characteristic of the 2C stage is

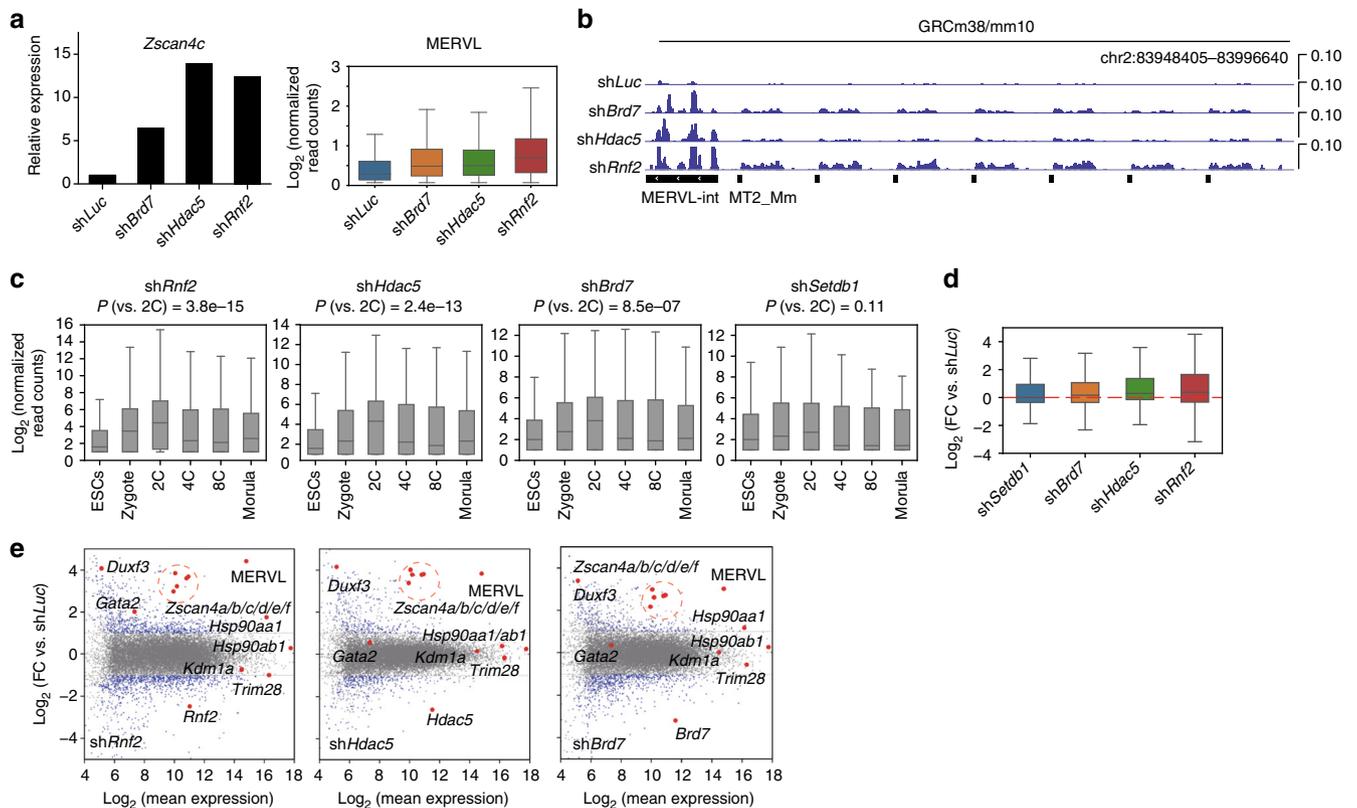


Fig. 5 Knockdown of chromatin modifiers activates a 2C-like gene expression program. **a** RNA-seq expression level of the 2C-related gene *Zscan4c* in the indicated knockdowns (left), and a boxplot showing the level of MERVL TEs in the same knockdown RNA-seq data. Bar chart values indicate the mean of the two biological replicates (*shHdac5* or *shRnf2*), or a biological singlicate (*shBrd7*), relative to *shLuc* control knockdown. The boxplot shows the Log₂ normalized read count mapping to the total set of MERVL TEs, from the same RNA-seq data. **b** Example genome view showing the RNA-seq pileup data and activation of MERVL-int and MT₂Mm TE expression in *Brd7*, *Hdac5*, and *Rnf2* shRNA knockdowns. **c** Boxplot of RNA-seq expression for genes upregulated in the indicated knockdowns. For the sets of genes significantly upregulated in the indicated CM knockdown, their expression was measured in the indicated embryonic stages and in ESCs (RNA-seq data was taken from⁶⁰). *P* values are from a Mann-Whitney *U* test. The *x*-axis indicates the overall level of expression of the knockdown-specific genes, in those cell types. **d** Upregulation of a 2C-related gene signature by the indicated CM knockdowns. A 2C-related gene signature was generated by taking the fold-change of all 2C genes versus ESCs and taking the top 200 fold-change upregulated genes. The expression of those top 200 2C genes was then measured in the indicated knockdowns, relative to *shLuc* control. **e** MA-scatter plots showing the expression of all genes and TEs in the *shRnf2*, *shHdac5*, and *shBrd7* knockdowns. Labels and red spots indicate the expression of the shRNA target gene, the MERVL-int TE, and selected 2C-related genes (*Zscan4a/b/c/d/e/f*-cluster, *Duxf3*, *Hsp90aa1*, *Hsp90ab1*, and *Gata2*). Note that the effect is not related to deregulation of *Trim28* or *Kdm1a*, two known MERVL regulators. For the boxplots in this figure, the midline indicates the median, boxes indicate the upper and lower quartiles and the whiskers indicate 1.5*interquartile range

the splicing of MERVL TEs into genes on the same strand orientation⁶¹, a pattern we observed only in the *Brd7*, *Hdac5*, and *Rnf2* knockdowns (Supplementary Figure 7a, b). These results showed that the 2C gene expression program was only seen in the sh*Brd7*, sh*Hdac5* and sh*Rnf2* knockdowns (Fig. 5d).

We next sought to understand how these knockdowns were activating a 2C gene expression program. Our analysis indicates that MERVLs are enriched for the activatory H3K56ac chromatin mark, but are not enriched for H2AK119ub (Fig. 1a and Supplementary Data 2). This is interesting as RNF2 catalyses the deposition of H2AK119ub, and BRD7 and HDAC5 can both target histone acetylation. These observations suggest two pathways for the activation of the 2C gene expression program; one direct, and one indirect, pathway. The direct pathway involved the acetylation of histones, as knockdown of the histone deacetylase *Hdac5*, and knockdown of the histone acetylation binder *Brd7* could upregulate MERVLs and 2C genes (Fig. 5a, c). Conversely, upregulation of MERVLs in the *Rnf2* knockdown, suggests an indirect pathway, as MERVLs are not enriched with H2AK119ub, but downregulation of H2AK119ub is a requirement for 2C-like cells⁶². The TF GATA2 has been implicated in driving cells to a 2C-fate⁶¹, and in our data only knockdown of *Rnf2* could activate *Gata2* (Fig. 5e). Interrogating ChIP-seq data for RNF2 and H2AK119ub⁶³, indicated that they are both enriched at the *Gata2* locus (Supplementary Figure 7c). These results suggest that 2C-like cells can emerge by the manipulation of two pathways, one pathway involving H3K56ac, HDAC5, and BRD7, and an indirect pathway involving H2AK119ub, RNF2, and *Gata2*. These pathways complement other factors known to activate the 2C-like state, including *Chaf1a/b*, *Chd5*, *Rif1*, *Kdm1a*, *Ehmt2*, *Yy1*, *Rybp*, and miR-34a^{15,26,28,61,64}. These results suggest that there are multiple convergent pathways that can drive cells to a 2C state, although whether all of these factors are converging on the same pathway or act independently remains unclear.

Discussion

TEs are the single most prevalent genomic element, more numerous than coding genes or other regulatory units. Although potentially dangerous for genome integrity they are nonetheless a pool of potential regulatory sequences, chromatin control elements, and protein coding genes, and they can function as a substrate for evolutionary processes^{3,9,21}. During early embryonic development there is extensive chromatin remodeling⁵⁸, and the activation of TE expression¹⁸. Consequently, a careful balance is maintained between the potential benefits of TEs, versus their deleterious aspects, and this control is carried out by the epigenetic regulatory system.

Heterochromatin formation is a major mechanism for the suppression of TEs, and we find that many TEs are marked by heterochromatic histone modifications. However, data suggests that, at least for LINEs and IAP-family TEs, TEs are repressed by the loss of activatory marks, rather than through the gain of repressive heterochromatin¹⁹. Our data indicate that TEs are marked by many activatory marks, like histone acetylations. We also observed bivalent marking of TE chromatin by H3K9me3 and H3K27ac, which may help to explain why we only observed a limited number of downregulated TEs in response to CM knockdown. This was surprising, as ESCs express over 200 TE types at levels comparable to many genes. Bivalent marking may mean that these TEs are kept in partially activated and repressed states.

TEs are increasingly implicated in a wide-range of biological functions in embryonic cell types. It has long been known that there is reverse transcriptase activity (indicative of TE activity), in developing mouse embryos⁶⁵, and HERVK-derived viral-like

particles have been observed in normal human embryos⁶⁶. TEs are expressed in a stage-specific pattern in both human and mouse embryos^{18,19}, and their expression in ESCs is associated with alternative properties. For example, MERVL expression marks the 2C stage of embryonic development as well as 2C-like cells in ESCs²⁶, and the inhibition of MERVLs, or MERVL-containing lincRNAs, impairs the transition from the zygote to a 4C embryo stage^{67,68}. LINE1 RNA is transiently required for 2C stage embryos to exit to the 4C stage, and in ESCs LINE1 is required for self-renewal, and chromatin remodeling^{43,44}. It is increasingly clear that TEs are involved in normal biological processes.

In conclusion, our study reveals the complexity of epigenetic regulation of TEs in ESCs. TEs are marked not only by repressive chromatin marks, but are also widely marked by activatory chromatin marks, including patterns of simultaneous marking by activatory and repressive marks. We show that the loss of CMs caused specific changes in TE expression and chromatin accessibility, and these effects are not limited to the formation of heterochromatin or DNA methylation, but encompass a large fraction of known chromatin modifications. This work suggests that the regulation of TEs is a function of overlapping epigenetic pathways, and that a major task of the epigenetic system is to manage the expression of TEs.

Methods

Cell lines and cell culture medium. All experiments were performed in mouse E14 ESCs (RRID:CVCL_C320), to avoid any differences in genetic background that may influence the number of TEs. Mouse ESCs were grown in serum + LIF medium: DMEM-HIGH GLUCOSE (SH30022.01, HyClone), 15% (v/v) fetal bovine serum (S1580-500, Biowest), 1% (v/v) L-Glutamine (35050-061, GIBCO), 1% (v/v) sodium pyruvate (25-000-CIR, Corning), 1% (v/v) non-essential amino acids (25-025-CI, Corning), 0.5% (v/v) penicillin-streptomycin solution (SV30010, HyClone), 0.1 mM 2-mercaptoethanol (21985-023, GIBCO), and 1000 U/ml mLIF (ESG1107, Millipore).

Chromatin modifier knockdowns. CMs were knocked down in serum + LIF medium using pLKO.puro lentiviral system as previously described^{27,41}. Briefly, pLKO.puro lentiviruses were added to ESCs and the medium was changed after 8 h. Cells were selected with 2 µg/ml puromycin (Gibco) for 2 days, and harvested after a further 2 days. Sequences for the shRNAs are in Supplementary Data 6.

Western blot. Histone proteins were extracted by acid precipitation. Briefly, cells from a 6-well dish were scraped, washed with PBS and resuspended in extraction buffer (PBS, 0.5% Triton-X-100 (9002-93-1, Sigma-Aldrich), 5 mM sodium butyrate (B5887, Sigma-Aldrich)), and allowed to extract for 20 mins. The extract was centrifuged at 14,000× g for 20 mins and the supernatant was discarded. The pellet was resuspended in 0.2 N HCl and left at 4 °C overnight. The mixture was vortexed, and the pH was neutralized with 1 M Tris-HCl pH 8.0. Western blots were performed using typical laboratory procedures with the antibodies: anti-H3K9me3 (1:1000; ab8898, abcam), anti-H3K4me3 (1:1000; ab8580, abcam), anti-H3K4me1 (1:1000; ab8895, abcam), anti-H3K27me3 (1:1000; 07-449, Millipore), anti-H3ac (1:1000; 06-599, Millipore), anti-H4K12ac (1:1000; ab46983, abcam), anti-H4ac (1:1000; 06-886, Millipore), anti-H3K27ac (1:1000; ab4729, abcam), anti-H2AK119ub (1:1000; 8240, Cell Signaling Technology), anti-H3 (1:1000; ab1971, abcam), anti-OCT4 (1:10,000; SC-8628, Santa Cruz), anti-NANOG (1:1000; A300-397A, Bethyl), and anti-GAPDH (1:10,000; MAB374, Millipore). Uncropped raw images of the Western blot membranes are in Supplementary Figure 8.

RNA-seq and RT-qPCR experimental protocols. RNA-seq was performed as described⁶⁹. Briefly, a 6-well dish of ESCs was scraped and lysed in TRIzol (RN190-200, Invitrogen). RNA was extracted according to the manufacturer's instructions, and 1 µg was submitted for sequencing. All RNA-seq was sequenced to a depth of 50 M tags, and 150 bp paired-end. For RT-qPCR RNA was extracted using TRIzol (RN190-200, Invitrogen), cDNA was prepared with reverse transcriptase kit (RR36B, Takara), and RT-qPCR was performed with SYBR Green Master kit (RR820B, Takara). RT-qPCR primers are described in Supplementary Data 7.

ATAC-seq and native-ChIP-qPCR experimental protocols. ATAC-seq was performed as described^{57,70}. A total of ~50,000 cells were washed once with 50 µl of cold PBS and resuspended in 50 µl lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl₂, 0.2% (v/v) IGEPAL CA-630). The suspension of nuclei was then centrifuged for 10 min at 500 × g at 4 °C, followed by the addition of 50 µl

transposition reaction mix (25 μ l TD buffer, 2.5 μ l Tn5 transposase and 22.5 μ l nuclease-free H₂O) of Nextera DNA library Preparation Kit (96 samples) (FC-121-1031, Illumina). Samples were then PCR amplified and incubated at 37 °C for 30 min. DNA was isolated using a MinElute Kit (Qiagen). ATAC-seq libraries were first subjected to five cycles of pre-amplification. To determine the suitable number of cycles required for the second round of PCR, the library was assessed by quantitative PCR⁷⁰, and the library was then PCR amplified for the appropriate number of cycles. Libraries were purified with a Qiaquick PCR (Qiagen) column and were quantified using a KAPA Library Quantification kit (KK4824, Kapa Biosystems) according to the manufacturer's instructions. Library integrity was checked by gel electrophoresis. Finally, the ATAC library was sequenced on a NextSeq 500 using a NextSeq 500 High Output Kit v2 (150 cycles) (FC-404-2002, Illumina) according to the manufacturer's instructions.

Native-ChIP-qPCR was performed as described⁷¹. Ten million cells were centrifuged and resuspended in Buffer 1 (0.32 M Sucrose, 60 mM KCl, 15 mM NaCl, 5 mM MgCl₂, 0.1 mM EGTA, 15 mM Tris-HCl pH 7.5, 0.5 mM DTT, 0.1 mM PMSF, 1:1000 protease inhibitor cocktail (Sigma-Aldrich)), along with 2 ml of Buffer1 + 0.1% IGEPAL. The resulting 4 ml of nuclei were layered on top of 8 ml of Buffer 3 (Same as Buffer 1, but with 1.2 M Sucrose) and centrifuged at 10,000 \times g for 20 min at 4 °C unbraked. Nuclei were resuspended in Buffer A (0.34 M sucrose, 15 mM Hepes [pH 7.4], 15 mM NaCl, 60 mM KCl, 4 mM MgCl₂, 1 mM DTT, 0.1 mM PMSF, 1:1000 protease inhibitor cocktail (Sigma-Aldrich)) and digested for 10 mins at 37 °C with MNase (Affymetrix) in buffer A supplemented with 3 mM CaCl₂. The reaction was stopped with 5 mM EGTA, centrifuged at 13,500 \times g for 10 min, and chromatin resuspended in (10 mM EDTA [pH 8.0], 1 mM PMSF, 1:1000 protease inhibitor cocktail) and rotated at 4 °C for 2–4 h. The mixture was adjusted to 500 mM NaCl, allowed to rotate for another 45 min and then centrifuged at 13,500 \times g for 10 mins. Chromatin supernatant was diluted to 100 ng/ μ l with buffer B (20 mM Tris [pH 8.0], 5 mM EDTA, 500 mM NaCl, 0.2% Tween 20) and incubated for 20 mins at 4 °C with 60 μ l protein G beads (GE Healthcare). Antibody was added and rotated overnight at 4 °C. A volume of 100 μ l of Protein G beads were added and the sample was rotated for a further 3 h at 4 °C. The beads were washed three times with buffer B, and once with buffer B without Tween 20. The DNA was eluted with 300 μ l of elution buffer (20 mM Tris [pH 7.5], 20 mM EDTA, 0.5% SDS, 500 μ g/ml Proteinase K) and incubated for 4 h at 56 °C. The resulting samples were extracted with phenol-chloroform followed by purification using a Qiagen MinElute columns, according to the manufacturer's instructions. Antibodies used: anti-H3K9me3 (5 μ g; ab8898, abcam), anti-H3K27ac (5 μ g; ab4729, abcam). qPCR was performed with SYBR Green Master kit (RR820B, Takara). Primer pairs are described in Supplementary Data 7.

RNA-seq and TE RNA-seq computational analysis. As a quality control step the RNA-seq data was analyzed for normal gene expression, using the method described^{60,69}. Briefly, reads were aligned to the genome using bowtie2⁷² and RSEM⁷³, with the settings ‘—bowtie2—bowtie2-sensitivity-level very_sensitive—no-bam-output—estimate-rspd’ using an index built against the ENSEMBL v81 transcriptome, and normalized using EDASeq⁷⁴ (v2.4.1) (which = ‘full’). Samples that did not closely correlate with other replicate samples, clustered closer to unrelated samples, or were outliers (R^2 correlation < 0.6), were deleted from the analysis.

For the analysis of TEs, reads were mapped to the mouse genome (mm10) using the STAR aligner⁷⁵, with the options:—readFilesCommand zcat—outFilterMultimapNmax 100—winAnchorMultimapNmax 100—outMultimapperOrder Random—runRNGseed 777—outSAMmultNmax 1—outSAMtype BAM—outFilterType BySJout—alignSJoverhangMin 8—alignSJDBoverhangMin 1—outFilterMismatchNmax 999—alignIntronMin 20—alignIntronMax 1000000—alignMatesGapMax 1000000. Tags mapping to genes and TEs were counted using Tetranscripts⁷⁶ with the option:—mode multi. GTF files for gene annotation were downloaded from GENCODE (M9). GTF files for TE annotations were downloaded from the Tetranscripts website (mm10_rmsk_TE.gtf.gz). For the quantification of genes and TEs, samples were normalized using EDASeq⁷⁴ (v2.4.1) (which = ‘full’). The GC content for TEs was generated by taking the average for all TEs. TEs were normalized at the same time as the genes. This has the effect of suppressing the magnitude of the fold-changes, however the fold-changes reported here are similar to those reported in other studies that used knockdowns⁴⁰. Differential expression was called using DESeq2⁷⁵, with a P value < 0.05 (Benjamini-Hochberg-corrected) and an absolute fold-change > 2. Gene ontology analysis was performed using goseq (v3.2.5)⁷⁷. Other analysis was performed using gbase⁷⁸.

The age of TEs was calculated as in^{8,11}. Briefly, the average divergence between the TE repeats and the consensus (milliDiv from RepeatMasker UCSC track) was calculated for each TE type, and the Jukes-Cantor method with a substitution rate of 4.5e-9 per base per year was applied. This gave an estimate for TE age in years. The results were broadly consistent with other estimates of TE age^{8,11}.

The list of 2C-related genes was generated by calculating the fold-change of 2C genes versus ESC genes, and taking the top 200 2C-associated genes. RNA-seq data was from⁶⁰. A second approach was used to generate a 2C-associated gene signature, by taking the top 200 upregulated genes in each CM knockdown, and then measuring the fold-change of those genes in the 2C sample, relative to the ESCs.

MERVL-associated genes were defined as those genes that have a MERVL (either MERVL-int or solo-LTR MT2_Mm) within 10 kb up- or downstream of the gene body. A gene was annotated as the same orientation (same ori.) if it had a MERVL in the same strand flanking it, or in the opposite orientation (opposite ori.) if it had a MERVL element in the opposite strand, as described in⁶¹.

Mapping statistics and details of all the samples used in this study are contained in Supplementary Data 1.

ChIP-seq and ATAC-seq computational analysis. ChIP-seq data for histone modifications, variant histones, DNA methylation, chromatin accessibility, and a selection of TFs and chromatin modifiers (Supplementary Data 1), were downloaded from GEO/SRA and the reads were aligned to the mouse genome (mm10) using bowtie2 (v2.2.5)⁷² with the options: -p 20—very-sensitive—end-to-end—no-unal. To analyze repetitive sequences, only the best alignment is reported for multimapped reads, if more than one equivalent best alignment was found, then one random alignment was reported. Reads mapping to mitochondrial DNA or unassigned sequences were discarded. For pair-end sequence data, only concordantly aligned pairs were kept. Alignment bam files were transformed into read coverage files (bigwig format) using deepTools⁷⁹ with the RPKM (Reads Per Kilobase per Million mapped reads) normalization method. For analysis of TEs, coordinates and annotations of TEs were downloaded from the UCSC Genome Browser (GRCm38/mm10) version of RepeatMasker. TEs shorter than 300 bp or those TE types with less than 50 copies were deleted from the analysis. For TE enrichment analysis, as many ChIP and ATAC signals were not evenly distributed across TEs, we divided the TEs into evenly spaced 500 bp bins. The coverage signal on each bin was extracted using deepTools, and the bin with the maximum signal was used as the observed value. For the expected background value, random genomic regions of the same number as the TE type and the same size as the TE were taken, the coverage signal was measured, and the maximal bin was taken as the randomly expected background. Enrichment was calculated by taking the observed over the expected enrichment, and is expressed as log₂(fold-enrichment). Quality control was performed by measuring the reads per kilobase of TE per million library sequence reads (RPKM) at all TEs, and then combining replicates where the Pearson correlation was >0.6 (Supplementary Figure 1a). Outlier samples that did not correlate closely with other ChIP-seq data samples of the same type were deleted from the analysis. The full data set of ChIP-seq experiments used in this study is detailed in Supplementary Data 1. Heatmaps and pileups were generated using deepTools⁷⁹, motif discovery was performed using MEME⁸⁰.

Quantification and statistical analysis. No statistical methods were used to predetermine sample size, and experiments were not randomized. The investigators were not blinded to the allocation of experiments, nor the outcome. The number of biological replicates are described in the figure legends, or in Supplementary Data 1. Samples were considered statistically significant at $P < 0.05$, unless indicated. The statistical test used is indicated in the figure legend.

Reporting Summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The datasets supporting the conclusions of this article are available in the Gene Expression Omnibus (GEO), under accession number GSE108091. The authors declare that the data supporting the findings of this study are available within the article and its Supplementary Information files, or from the corresponding author upon reasonable request. A Reporting Summary for this Article is available as a Supplementary Information file. The Source Data underlying Figs. 1d, 3h, and Supplementary Figs 2h and 5b are provided as a Source Data file.

Received: 25 May 2018 Accepted: 11 December 2018

Published online: 03 January 2019

References

- Thompson, P. J., Macfarlan, T. S. & Lorincz, M. C. Long terminal repeats: from parasitic elements to building blocks of the transcriptional regulatory repertoire. *Mol. Cell* **62**, 766–776 (2016).
- Callinan, P. A. & Batzer, M. A. Retrotransposable elements and human disease. *Genome Dyn.* **1**, 104–115 (2006).
- Shapiro, J. A. Exploring the read-write genome: mobile DNA and mammalian adaptation. *Crit. Rev. Biochem. Mol. Biol.* **52**, 1–17 (2017).
- Hutchins, A. P. & Pei, D. Transposable elements at the center of the crossroads between embryogenesis, embryonic stem cells, reprogramming, and long non-coding RNAs. *Sci. Bull.* **60**, 1722–1733 (2015).
- Johnson, R. & Guigo, R. The RIDL hypothesis: transposable elements as functional domains of long noncoding RNAs. *RNA* **20**, 959–976 (2014).

6. Chuong, E. B., Elde, N. C. & Feschotte, C. Regulatory activities of transposable elements: from conflicts to benefits. *Nat. Rev. Genet.* **18**, 71–86 (2017).
7. Johnson, R. et al. Identification of the REST regulon reveals extensive transposable element-mediated binding site duplication. *Nucleic Acids Res.* **34**, 3862–3877 (2006).
8. Bourque, G. et al. Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res.* **18**, 1752–1762 (2008).
9. Kunarso, G. et al. Transposable elements have wired the core regulatory network of human embryonic stem cells. *Nat. Genet.* **42**, 631–634 (2010).
10. Sun, X. et al. Transcription factor profiling reveals molecular choreography and key regulators of human retrotransposon expression. *Proc. Natl Acad. Sci. USA* **115**, E5526 (2018).
11. Jacques, P. E., Jeyakani, J. & Bourque, G. The majority of primate-specific regulatory sequences are derived from transposable elements. *PLoS Genet.* **9**, e1003504 (2013).
12. Simonti, C. N., Pavlicev, M. & Capra, J. A. Transposable element exaptation into regulatory regions is rare, influenced by evolutionary age, and subject to pleiotropic constraints. *Mol. Biol. Evol.* **34**, 2856–2869 (2017).
13. Leung, D. C. et al. Lysine methyltransferase G9a is required for de novo DNA methylation and the establishment, but not the maintenance, of proviral silencing. *Proc. Natl Acad. Sci. USA* **108**, 5718–5723 (2011).
14. Bulut-Karslioglu, A. et al. Suv39h-dependent H3K9me3 marks intact retrotransposons and silences LINE elements in mouse embryonic stem cells. *Mol. Cell* **55**, 277–290 (2014).
15. Li, P. et al. Rif1 promotes a repressive chromatin state to safeguard against endogenous retrovirus activation. *Nucleic Acids Res.* **45**, 12723–12738 (2017).
16. Esnault, C. et al. APOBEC3G cytidine deaminase inhibits retrotransposition of endogenous retroviruses. *Nature* **433**, 430 (2005).
17. Reik, W. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* **447**, 425–432 (2007).
18. Goke, J. et al. Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. *Cell. Stem. Cell.* **16**, 135–141 (2015).
19. Fadloun, A. et al. Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. *Nat. Struct.* **20**, 332 (2013).
20. Kim, S. et al. PRMT5 protects genomic integrity during global DNA demethylation in primordial germ cells and preimplantation embryos. *Mol. Cell* **56**, 564–579 (2014).
21. Frank, J. A. & Feschotte, C. Co-option of endogenous viral sequences for host cell function. *Curr. Opin. Virol.* **25**, 81–89 (2017).
22. Schlesinger, S. & Goff, S. P. Retroviral transcriptional regulation and embryonic stem cells: war and peace. *Mol. Cell. Biol.* **35**, 770–777 (2015).
23. Najafabadi, H. S. et al. C2H2 zinc finger proteins greatly expand the human regulatory lexicon. *Nat. Biotechnol.* **33**, 555–562 (2015).
24. Yang, B. X. et al. Systematic identification of factors for provirus silencing in embryonic stem cells. *Cell* **163**, 230–245 (2015).
25. 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).
26. Macfarlan, T. S. et al. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487**, 57–63 (2012).
27. Castro-Diaz, N. et al. Evolutionarily dynamic L1 regulation in embryonic stem cells. *Genes Dev.* **28**, 1397–1409 (2014).
28. Maksakova, I. A. et al. Distinct roles of KAP1, HP1 and G9a/GLP in silencing of the two-cell-specific retrotransposon MERV1 in mouse ES cells. *Epigenetics & chromatin* **6**, 15 (2013).
29. 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).
30. Mikkelsen, T. S. et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–560 (2007).
31. Leeb, M. et al. Polycomb complexes act redundantly to repress genomic repeats and genes. *Genes & Dev.* **24**, 265–276 (2010).
32. Chew, Y. C. et al. Biotinylation of histones represses transposable elements in human and mouse cells and cell lines and in *Drosophila melanogaster*. *J. Nutr.* **138**, 2316–2322 (2008).
33. Pestinger, V., Wijeratne, S. S., Rodriguez-Melendez, R. & Zempleni, J. Novel histone biotinylation marks are enriched in repeat regions and participate in repression of transcriptionally competent genes. *J. Nutr. Biochem.* **22**, 328–333 (2011).
34. Elsassser, S. J., Noh, K. M., Diaz, N., Allis, C. D. & Banaszynski, L. A. Histone H3.3 is required for endogenous retroviral element silencing in embryonic stem cells. *Nature* **522**, 240–244 (2015).
35. Lisch, D. & Benetzen, J. L. Transposable element origins of epigenetic gene regulation. *Curr. Opin. Plant Biol.* **14**, 156–161 (2011).
36. Yin, Y., Zhou, L. & Yuan, S. Enigma of retrotransposon biology in mammalian early embryos and embryonic stem cells. *Stem Cells Int.* **2018**, 6239245 (2018).
37. Leung, D. C. & Lorincz, M. C. Silencing of endogenous retroviruses: when and why do histone marks predominate? *Trends Biochem. Sci.* **37**, 127–133 (2012).
38. Walter, M., Teissandier, A., Perez-Palacios, R. & Bourc'his, D. An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. *eLife* **5**, e11418 (2016).
39. Deniz, Ö., de la Rica, L., Cheng, K. C. L., Spensberger, D. & Branco, M. R. SETDB1 prevents TET2-dependent activation of IAP retroelements in naïve embryonic stem cells. *Genome Biol.* **19**, 6 (2018).
40. Berrens, R. V. et al. An endosRNA-based repression mechanism counteracts transposon activation during global DNA demethylation in embryonic stem cells. *Cell. Stem. Cell.* **21**, 694–703.e697 (2017).
41. Rowe, H. M. et al. KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* **463**, 237–240 (2010).
42. Matsui, T. et al. Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature* **464**, 927–931 (2010).
43. Jachowicz, J. W. et al. LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. *Nat. Genet.* **49**, 1502–1510 (2017).
44. Percharde, M. et al. A LINE1-nucleolin partnership regulates early development and ESC identity. *Cell* **174**, 391–405.e319 (2018).
45. Sundaram, V. & Wang, T. Transposable element mediated innovation in gene regulatory landscapes of cells: re-visiting the “gene-battery” model. *Bioessays* **40**, 1700155 (2017).
46. Schoorlemmer, J., Perez-Palacios, R., Climent, M., Guallar, D. & Muniesa, P. Regulation of mouse retroelement MuERV-L/MERVL expression by REX1 and epigenetic control of stem cell potency. *Front. Oncol.* **4**, 14 (2014).
47. Guo, G. & Smith, A. A genome-wide screen in EpiSCs identifies Nr5a nuclear receptors as potent inducers of ground state pluripotency. *Development* **137**, 3185–3192 (2010).
48. Habibi, E. et al. Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. *Cell. Stem. Cell.* **13**, 360–369 (2013).
49. Leung, D. et al. Regulation of DNA methylation turnover at LTR retrotransposons and imprinted loci by the histone methyltransferase Setdb1. *Proc. Natl Acad. Sci. USA* **111**, 6690–6695 (2014).
50. Baubec, T., Ivánek, R., Lienert, F. & Schübeler, D. Methylation-dependent and -independent genomic targeting principles of the MBD protein family. *Cell* **153**, 480–492 (2013).
51. Zuin, J. et al. Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc. Natl Acad. Sci. USA* **111**, 996–1001 (2014).
52. He, Q. et al. The Daxx/Atrx complex protects tandem repetitive elements during DNA hypomethylation by promoting H3K9 trimethylation. *Cell. Stem. Cell.* **17**, 273–286 (2015).
53. Medvedeva, Y. A. et al. EpiFactors: a comprehensive database of human epigenetic factors and complexes. *Database* **2015**, bav067 (2015).
54. Zhuang, Q. et al. NCoR/SMRT co-repressors cooperate with c-MYC to create an epigenetic barrier to somatic cell reprogramming. *Nat. Cell Biol.* **20**, 400–412 (2018).
55. Saunders, A. et al. The SIN3A/HDAC corepressor complex functionally cooperates with NANOG to promote pluripotency. *Cell Rep.* **18**, 1713–1726 (2017).
56. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
57. Li, D. et al. Chromatin accessibility dynamics during iPSC reprogramming. *Cell. Stem. Cell.* **21**, 819–833.e816 (2017).
58. Wu, J. et al. The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature* **534**, 652 (2016).
59. Hendrickson, P. G. et al. Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERV1/HERVL retrotransposons. *Nat. Genet.* **49**, 925–934 (2017).
60. Hutchins, A. P. et al. Models of global gene expression define major domains of cell type and tissue identity. *Nucleic Acids Res.* **45**, 2354–2367 (2017).
61. Choi, Y. J. et al. Deficiency of microRNA miR-34a expands cell fate potential in pluripotent stem cells. *Science* **355**, 6325 (2017).
62. Rodriguez-Terrones, D. et al. A molecular roadmap for the emergence of early-embryonic-like cells in culture. *Nat. Genet.* **50**, 106–119 (2018).
63. Li, H. et al. RNA helicase DDX5 inhibits reprogramming to pluripotency by miRNA-based repression of RYBP and its PRC1-dependent and -independent functions. *Cell. Stem. Cell.* **20**, 462–477.e466 (2017).
64. Hisada, K. et al. RYBP represses endogenous retroviruses and preimplantation- and germ line-specific genes in mouse embryonic stem cells. *Mol. Cell Biol.* **32**, 1139–1149 (2012).
65. Pittoggi, C. et al. Role of endogenous reverse transcriptase in murine early embryo development. *Mol. Reprod. Dev.* **66**, 225–236 (2003).

66. Grow, E. J. et al. Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature* **522**, 221–225 (2015).
67. Kigami, D., Minami, N., Takayama, H. & Imai, H. MuERV-L is one of the earliest transcribed genes in mouse one-cell embryos. *Biol. Reprod.* **68**, 651–654 (2003).
68. Wang, J. et al. A novel long intergenic noncoding RNA indispensable for the cleavage of mouse two-cell embryos. *EMBO Rep.* **17**, 1452 (2016).
69. Hutchins, A. P., Takahashi, Y. & Miranda-Saavedra, D. Genomic analysis of LPS-stimulated myeloid cells identifies a common pro-inflammatory response but divergent IL-10 anti-inflammatory responses. *Sci. Rep.* **5**, 9100 (2015).
70. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* **109**, 21–29 (2015).
71. Hasson, D. et al. The octamer is the major form of CENP-A nucleosomes at human centromeres. *Nat. Struct. Mol. Biol.* **20**, 687–695 (2013).
72. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
73. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinforma.* **12**, 323 (2011).
74. Risso, D., Schwartz, K., Sherlock, G. & Dudoit, S. GC-content normalization for RNA-Seq data. *BMC Bioinforma.* **12**, 480 (2011).
75. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
76. Jin, Y., Tam, O. H., Paniagua, E. & Hammell, M. TETranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. *Bioinformatics* **31**, 3593–3599 (2015).
77. Young, M. D., Wakefield, M. J., Smyth, G. K. & Oshlack, A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* **11**, R14 (2010).
78. Hutchins, A. P., Jauch, R., Dyla, M. & Miranda-Saavedra, D. glbase: a framework for combining, analyzing and displaying heterogeneous genomic and high-throughput sequencing data. *Cell Regen.* **3**, 1 (2014).
79. Ramirez, F., Dundar, F., Diehl, S., Gruning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–W191 (2014).
80. Bailey, T. L. et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009).

Acknowledgements

We thank Ralf Jauch for discussions. This work was supported by the National Natural Science Foundation of China (31801217, 31850410463, 31850410486, 31671537,

31501185), Shenzhen Peacock plan, the Shenzhen Science and Technology Innovation Committee general program (JCYJ20170307110638890), and the Science and Technology Planning Project of Guangdong Province (2017B030314056).

Author contributions

J.H. performed the bioinformatic analysis, and prepared the manuscript. X.F. did the ATAC-seq, RNA-seq, M.Z., W.L., M.M.A., J.Z. performed the knockdown panel RNA-seq. H.L., W.K., F.H., C.C. performed experiments, L.S., Y.L., I.A.B., helped with the bioinformatics. A.P.H. conceived the project, supervised and funded the study, and prepared the manuscript with assistance from M.A.E, C.J, Q.Z., I.A.B, and Y.-H.L.

Additional information

Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41467-018-08006-y>.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019