

# Targeting pericentric non-consecutive motifs for heterochromatin initiation

<https://doi.org/10.1038/s41586-024-07640-5>

Received: 2 August 2023

Accepted: 31 May 2024

Published online: 3 July 2024

 Check for updates

Runze Ma<sup>1,2,3,6</sup>, Yan Zhang<sup>1,2,6</sup>, Jing Zhang<sup>1,2</sup>, Pinqi Zhang<sup>1,2,3</sup>, Zeqi Liu<sup>1,2,3</sup>, Yiming Fan<sup>1</sup>, Hao-Tian Wang<sup>4</sup>, Zhuqiang Zhang<sup>1,2</sup> & Bing Zhu<sup>1,2,5</sup>✉

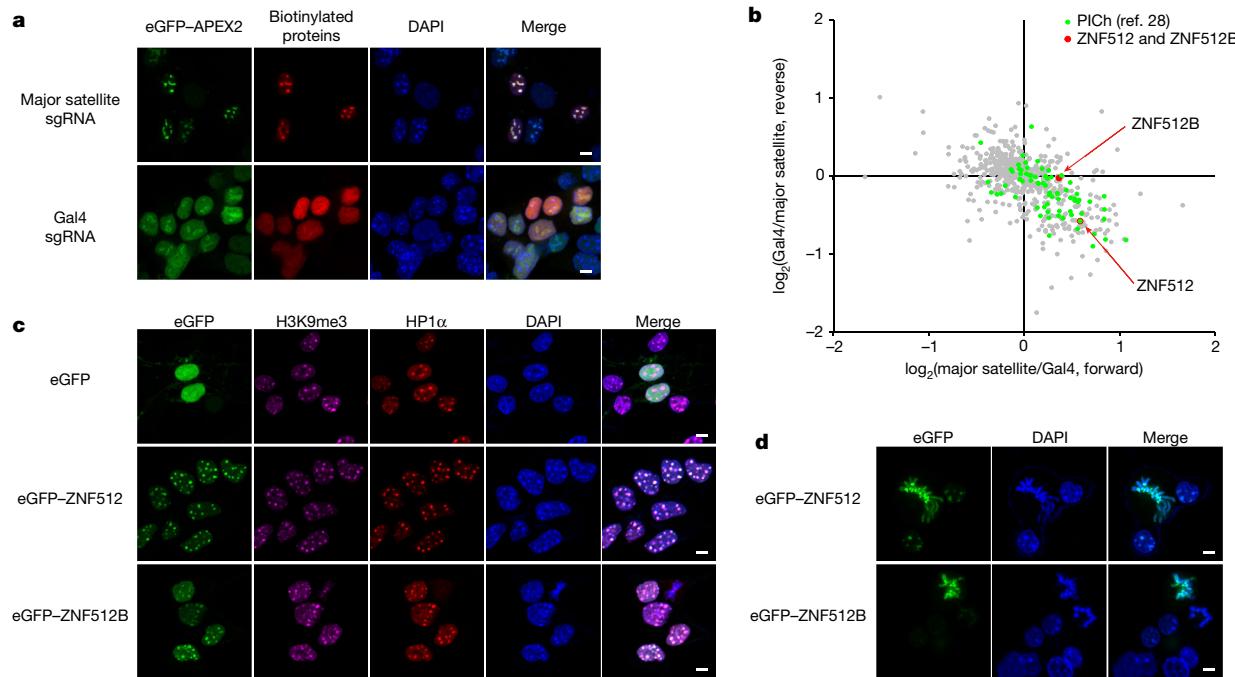
Pericentric heterochromatin is a critical component of chromosomes marked by histone H3 K9 (H3K9) methylation<sup>1–3</sup>. However, what recruits H3K9-specific histone methyltransferases to pericentric regions in vertebrates remains unclear<sup>4</sup>, as does why pericentric regions in different species share the same H3K9 methylation mark despite lacking highly conserved DNA sequences<sup>2,5</sup>. Here we show that zinc-finger proteins ZNF512 and ZNF512B specifically localize at pericentric regions through direct DNA binding. Notably, both ZNF512 and ZNF512B are sufficient to initiate de novo heterochromatin formation at ectopically targeted repetitive regions and pericentric regions, as they directly recruit SUV39H1 and SUV39H2 (SUV39H) to catalyse H3K9 methylation. SUV39H2 makes a greater contribution to H3K9 trimethylation, whereas SUV39H1 seems to contribute more to silencing, probably owing to its preferential association with HP1 proteins. ZNF512 and ZNF512B from different species can specifically target pericentric regions of other vertebrates, because the atypical long linker residues between the zinc-fingers of ZNF512 and ZNF512B offer flexibility in recognition of non-consecutively organized three-nucleotide triplets targeted by each zinc-finger. This study addresses two long-standing questions: how constitutive heterochromatin is initiated and how seemingly variable pericentric sequences are targeted by the same set of conserved machinery in vertebrates.

The term heterochromatin was first used by Emil Heitz in 1928 to refer to chromosome regions showing deep staining signals, indicating highly compacted chromatin<sup>6</sup>. Functionally, heterochromatin is related to the classic epigenetic phenomenon known as position-effect variegation, which was observed in 1930<sup>7</sup>. Despite the widespread existence of pericentric heterochromatin in eukaryotic species ranging from yeast to humans and its clear functional impact on gene silencing and genome stability<sup>1–3</sup>, identifying the critical molecular features of pericentric heterochromatin has been a long process. This was finally achieved with the milestone discovery that identified SUV39H homologues as histone H3 K9 (H3K9)-specific histone methyltransferases<sup>8</sup> and subsequent findings that confirmed the pericentric localization of SUV39H homologues<sup>9,10</sup>, H3K9 methylation as the molecular mark of pericentric heterochromatin<sup>8,11,12</sup>, and the binding platform of HP1 proteins<sup>9,10,13</sup>.

Despite a lack of significant sequence homology among pericentric sequences of eukaryotic species, H3K9 methylation catalysed by homologues of SUV39H is the principal pericentric heterochromatin mark in most eukaryotes, ranging from fission yeast to mammals<sup>1,2,5</sup>. A question arises as to what initiates heterochromatin formation and recruits SUV39H homologues to highly variable pericentric DNA sequences. In fission yeast *Schizosaccharomyces pombe*, which is the best-understood system, the RNA interference machinery recruits the SUV39H homologue Clr4 to pericentric repeats<sup>14–16</sup>. In addition, site-specific DNA-binding proteins Atf1 and Pcr1 recruit Clr4 to

another heterochromatin region, the mating type switch locus<sup>17</sup>. In vertebrates, however, the mechanisms governing de novo initiation of heterochromatin formation and direct recruitment of SUV39H homologues are less clear. In mammalian cells, transcripts from pericentric satellite repeats associate with HP1<sup>18</sup> and SUV39H homologues<sup>19–21</sup>, indicating a possible role of pericentric satellite RNA in stabilizing the pericentric localization of SUV39H homologues and HP1. Moreover, certain sequence-specific DNA-binding proteins, such as PAX3, PAX9 and FOXD3, localize to mouse major satellite repeats, and their depletion leads to a reduced H3K9 trimethylation (H3K9me3) signal at the major satellite regions and derepression of major satellite transcription<sup>22,23</sup>. However, all the aforementioned sequence-specific DNA-binding proteins are specifically expressed in certain developmental stages<sup>4</sup> (Extended Data Fig. 1a), whereas pericentric heterochromatin is known as constitutive heterochromatin because it exists in most cell types. Furthermore, the binding motifs of the above-mentioned sequence-specific DNA-binding proteins exist in mouse major satellite DNA sequences but not in human pericentric repeats (Extended Data Fig. 1b). To the best of our knowledge, direct evidence for a given factor that is sufficient to recruit SUV39H homologues to pericentric sequences or an ectopic locus and de novo initiate heterochromatin formation has yet to be shown. Ideally, such a protein or proteins should be ubiquitously expressed to play a part in ensuring constitutive heterochromatin fidelity in nearly all cell types.

<sup>1</sup>National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing, China. <sup>2</sup>Key Laboratory of Epigenetic Regulation and Intervention, Chinese Academy of Sciences, Beijing, China. <sup>3</sup>College of Life Sciences, University of Chinese Academy of Sciences, Beijing, China. <sup>4</sup>State Key Laboratory of Genetic Resources and Evolution/Key Laboratory of Healthy Aging Research of Yunnan Province, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China. <sup>5</sup>New Cornerstone Science Laboratory, Institute of Biophysics, Chinese Academy of Sciences, Beijing, China. <sup>6</sup>These authors contributed equally: Runze Ma, Yan Zhang. ✉e-mail: zhubing@ibp.ac.cn



**Fig. 1 | ZNF512 and ZNF512B specifically localize to pericentric regions.**

**a**, Representative cross-section showing the relative localization of eGFP–APEX2, biotinylated proteins and DAPI in major satellite-targeting cells or Gal4 sgRNA-expressing cells after proximity labelling. Scale bar, 5  $\mu$ m.

**b**, Scatter plot of the results of a SILAC-assisted proximity labelling assay to identify pericentric-heterochromatin-associated proteins in living mouse ES cells. Proteins are plotted by their SILAC ratios in the forward (x axis) and reverse (y axis) SILAC experiment. Green dots indicate proteins that were previously defined by the PICh method<sup>28</sup>. Zinc-finger proteins ZNF512 and

ZNF512B are shown as red dots. **c**, Representative fluorescence images showing the localization of eGFP, eGFP–ZNF512 and eGFP–ZNF512B in mouse ES cells during interphase. H3K9me3 and HP1 $\alpha$  were stained to indicate pericentric heterochromatin. Scale bar, 5  $\mu$ m. **d**, Representative fluorescence images showing the localization of eGFP–ZNF512 and eGFP–ZNF512B in mouse ES cells during M phase. Scale bar, 5  $\mu$ m. In **a,c,d**, data are representative of two independent biological repeats; for the results shown in **b**, the experiment was performed once.

## Pericentric-localized ZNF512 and ZNF512B

We suspected the existence of sequence-specific DNA-binding proteins that recognize pericentric repeats and directly interact with SUV39H homologues, which might have been overlooked in previous work using genetic approaches owing to redundancy. To identify such candidates, we leveraged the efficient targeting ability of the CRISPR–Cas9 system and the spatial location-specific labelling capacity of the APEX2-based proximity labelling system<sup>24–26</sup> to characterize the protein composition of pericentric heterochromatin (Extended Data Fig. 2a). With this approach, we successfully targeted APEX2 to DAPI-dense pericentric heterochromatin foci in mouse embryonic stem (ES) cell and detected specific localization of biotinylated proteins at the pericentric heterochromatin foci upon APEX2 activation (Fig. 1a).

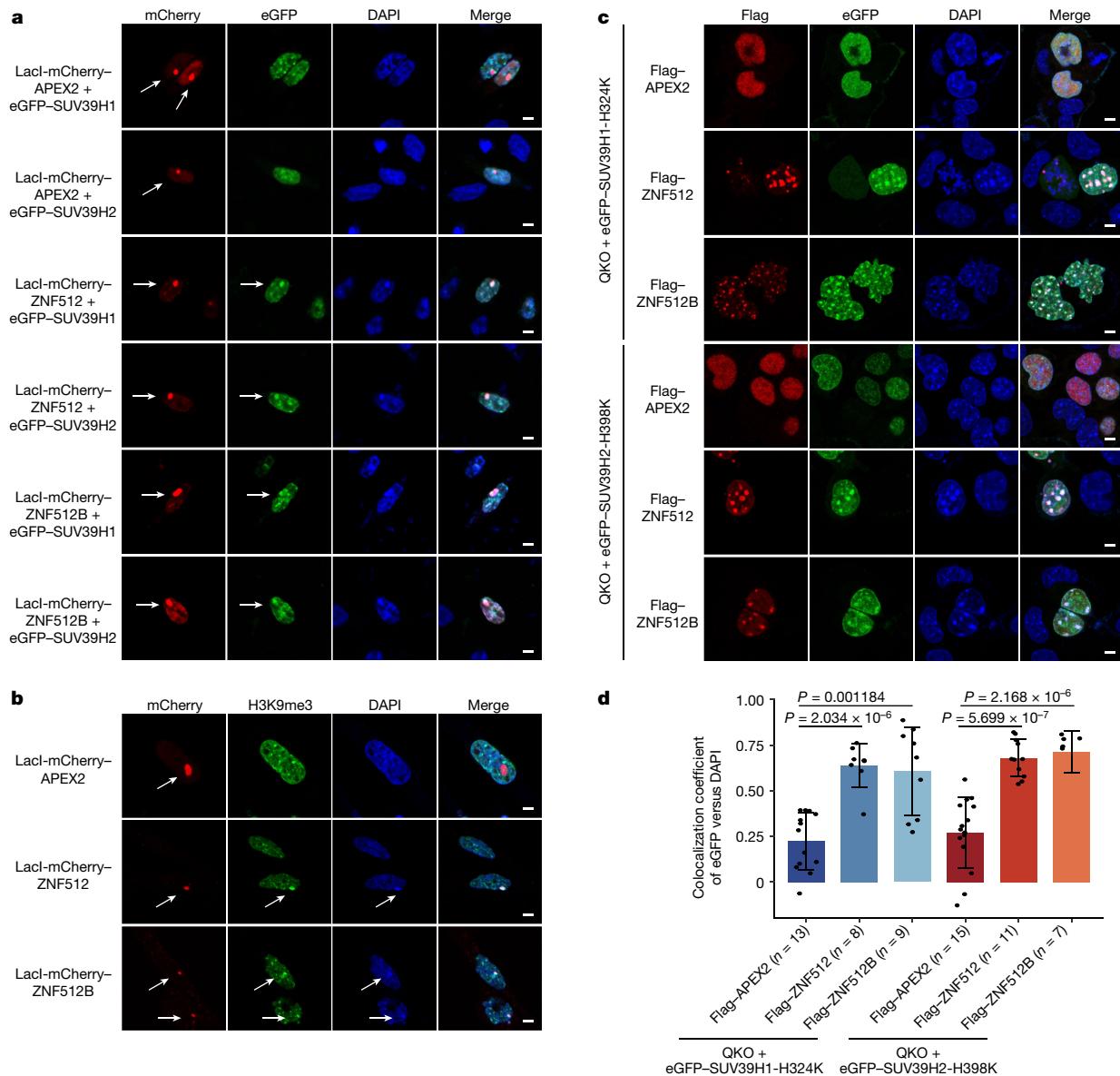
Next, we used the SILAC (stable isotope labelling with amino acids in cell culture) approach<sup>27</sup> and reciprocally labelled Gal4 single guide RNA (sgRNA)-expressing cells and major satellite-targeting cells with heavy-isotope-labelled lysine and arginine. Then, we activated APEX2, purified the biotinylated proteins, and performed quantitative mass spectrometry. Many of the proteins showing preferential association with pericentric heterochromatin overlapped with proteins previously found to associate with pericentric heterochromatin using the previously described PICh (proteomics of isolated chromatin segments) strategy<sup>28</sup> (Fig. 1b and Extended Data Fig. 2b). On the basis of our initial hypothesis, we chose candidate proteins that were likely to possess sequence-specific DNA-binding activities and focused on two zinc-finger proteins, ZNF512 and its parologue ZNF512B (also known as ZFP512 and ZFP512B, respectively). Both proteins showed preferential association with pericentric heterochromatin in our assay, and ZNF512 has also been reported to associate with pericentric region in

previous PICh experiments<sup>28</sup> (Fig. 1b). Moreover, these two proteins are expressed in most mouse cell types (Extended Data Fig. 1a). Indeed, enhanced green fluorescent protein (eGFP)-fused ZNF512, as previously reported<sup>29</sup>, and ZNF512B exhibited clear pericentric heterochromatin localization in mouse ES cells during interphase (Fig. 1c) and even M phase (Fig. 1d).

## Recruitment of SUV39H by ZNF512 and ZNF512B

To test whether ZNF512 and ZNF512B could initiate heterochromatin formation, we fused them with LacI and mCherry and introduced the fusion proteins into a CHO (Chinese hamster ovary)-derived reporter cell line harbouring 256 copies of LacO repeats in the genome<sup>30</sup> (Extended Data Fig. 3a). As a positive control, SUV39H1 or SUV39H2 tethered to LacO repeats successfully induced the enrichment of H3K9me3 and chromatin condensation (Extended Data Fig. 3b). Notably, the ectopic tethering of LacI-fused ZNF512 or ZNF512B was able to recruit SUV39H1 or SUV39H2 to the LacO repeats (Fig. 2a) and successfully de novo initiate H3K9me3 enrichment and heterochromatin formation (Fig. 2b).

Coimmunoprecipitation (Co-IP) experiments demonstrated that Flag-tagged ZNF512 or ZNF512B could pull down coexpressed haemagglutinin (HA)-tagged SUV39H in cells (Extended Data Fig. 3c). Moreover, purified recombinant Flag-tagged ZNF512 and ZNF512B exhibited direct protein–protein interactions with HA-tagged SUV39H proteins in vitro (Extended Data Fig. 3d). Then, we generated *Znf512* and *Znf512B* single- and double-knockout mouse ES cells (Extended Data Fig. 4a). Despite moderate derepression of the major satellite repeats (Extended Data Fig. 4b), H3K9me3 foci remained at pericentric heterochromatin (Extended Data Fig. 4c). Together, these results indicate that ZNF512 and ZNF512B are sufficient to initiate heterochromatin formation by



**Fig. 2 | ZNF512 and ZNF512B are sufficient to initiate heterochromatin formation through interaction with SUV39H.** **a**, Representative cross-section fluorescence image displaying the relative localization of LacI-mCherry-fused APEX2, ZNF512 or ZNF512B and eGFP-tagged SUV39H1 or SUV39H2 in CHO A03-1 cells. APEX2, ZNF512 and ZNF512B are indicated by mCherry fluorescence, and SUV39H1 and SUV39H2 are indicated by eGFP fluorescence. Scale bars, 5  $\mu$ m. **b**, Representative cross-section image showing the relative localization of LacI-mCherry-fused APEX2, ZNF512 or ZNF512B, H3K9me3 and DAPI in CHO A03-1 cells. APEX2, ZNF512 and ZNF512B are indicated by mCherry fluorescence.

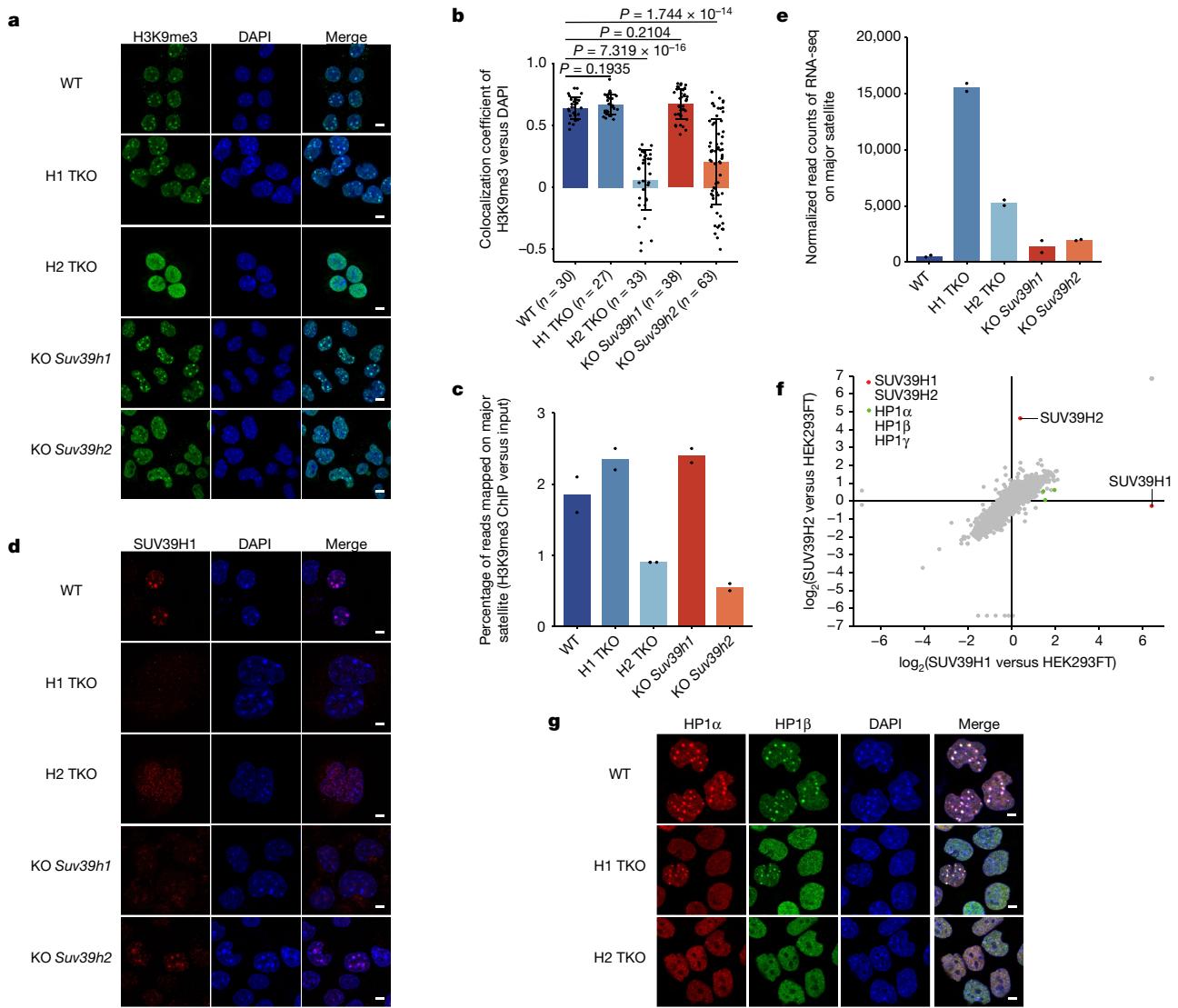
Condensed chromatin was stained with DAPI. Scale bars, 5  $\mu$ m. **c**, After overexpression of Flag-tagged APEX2, ZNF512 or ZNF512B, the localization of SUV39H1 or SUV39H2 with the catalytically dead mutant was visualized by eGFP fluorescence in the quadruple knockout (QKO) cell line. Scale bars, 5  $\mu$ m. **d**, Quantification of the colocalization coefficient of eGFP versus DAPI from images in **c**. Values for individual cells are indicated by dots. Data are mean  $\pm$  s.d. P-values were calculated using unpaired two-sided t-tests. In **a–c**, representative data from two independent biological repeats are shown.

directly recruiting SUV39H, but they do not seem to be essential for maintaining the molecular signature of heterochromatin, at least at the immunostaining level.

Although ZNF512 and ZNF512B were capable of initiating de novo heterochromatin formation at ectopic loci (Fig. 2a,b), their deletion did not significantly affect H3K9me3 foci at pericentric heterochromatin (Extended Data Fig. 4c). This was probably because of self-reinforcement mechanisms of the constitutive heterochromatin histone modification machinery. Both SUV39H homologues and HP1 proteins are capable of recognizing H3K9me3 with their chromodomains<sup>9,13,31,32</sup>, and these activities and the interactions between SUV39H homologues and HP1 proteins probably reinforce heterochromatin maintenance<sup>33–35</sup>. In fission yeast, this reinforcement mechanism has

been shown to support heterochromatin maintenance in the absence of an initial recruitment mechanism<sup>16,32,36</sup>. The same principle is likely to hold true in vertebrates.

To test the above hypothesis, we generated quadruple-knockout (QKO) cells by knocking out *Suv39h1* and *Suv39h2* in *Znf512* and *Znf512b* double-knockout cells (Extended Data Fig. 4a). As expected, the QKO cells completely lost pericentric H3K9me3 foci (Extended Data Fig. 5a). We then overexpressed eGFP-tagged SUV39H1, SUV39H2, and their aromatic cage mutants (SUV39H1(W64A/Y67A) and SUV39H2(W139A/W142A), abbreviated as SUV39H1-2A and SUV39H2-2A), which disrupt the H3K9me3-binding ability of the chromodomains, as well as the catalytically dead mutants (SUV39H1-H324K, SUV39H2-H398K) in the QKO cells. Notably, the overexpressed wild-type (SUV39H1 and



**Fig. 3 | Distinct properties of SUV39H1 and SUV39H2.** **a**, Representative cross-section immunofluorescence image showing the relative localization of H3K9me3 modification and DAPI in wild-type (WT) and various knockout (KO) cells. Scale bars, 5  $\mu\text{m}$ . **b**, Quantification of the colocalization coefficient of H3K9me3 versus DAPI from images in **a**. Values for individual cells are indicated by dots. Data are mean  $\pm$  s.d. P values were calculated using unpaired two-sided t-tests. **c**, Bar plot showing the ratio of the percentage of reads mapped on mouse major satellite repeats between H3K9me3 ChIP samples and the corresponding input samples in indicated cell lines. Bars indicate the mean; dots indicate individual replicates. **d**, Representative cross-section image showing the relative localization of endogenous SUV39H1 and DAPI in wild-type and knockout cells. Scale bars, 5  $\mu\text{m}$ . **e**, Bar plot showing normalized read counts

from RNA sequencing of mouse major satellite repeats in indicated cell lines. Bars indicate the mean; dots indicate individual replicates. **f**, Scatter plots showing fold enrichments of HP1 determined by label-free quantification-based mass spectrometry (LFQ-MS) in HA-tagged SUV39H1 or HA-tagged SUV39H2 versus HEK293FT control. SUV39H1 and SUV39H2 are labelled with red dots. HP1 $\alpha$ , HP1 $\beta$  and HP1 $\gamma$  are indicated by green dots. **g**, Representative cross-section of immunofluorescence images showing the relative localization of endogenous HP1 $\alpha$ , HP1 $\beta$  and DAPI in wild-type and various knockout cells. Scale bars, 5  $\mu\text{m}$ . In **a**, **d**, **g**, representative data from two independent biological repeats are shown. Data in **f** and in **c** and **e** are from three and two technical replicates, respectively. In **b**, n = 30, 27, 33, 38 or 63 cells per cell line.

SUV39H2, as well as their chromodomain mutants, were recruited to the pericentric foci in the absence of ZNF512 and ZNF512B but not the catalytically dead mutants (Extended Data Fig. 5b). Simultaneous coexpression of ectopic ZNF512 or ZNF512B successfully recruited the catalytically dead SUV39H mutants to the pericentric foci (Fig. 2c,d).

These results indicate that ZNF512 and ZNF512B may be key factors responsible for the pericentric recruitment of SUV39H homologues, as they are required for the pericentric localization of SUV39H homologues in the absence of pericentric H3K9me3. However, we do not rule out the contribution of other potential recruiting factors, because overexpressed wild-type SUV39H proteins successfully localized to pericentric regions in the QKO cells (Extended Data Fig. 5b).

## Distinct properties of SUV39H proteins

While constructing the QKO cells, we obtained two intermediates: *Znf512*, *Znf512b* and *Suv39h1* triple-knockout (H1 TKO) cells and *Znf512*, *Znf512b* and *Suv39h2* triple-knockout (H2 TKO) cells. H3K9me3 was not so well colocalized with the DAPI-dense pericentric heterochromatin in H2 TKO cells, but this was not observed in H1 TKO cells (Fig. 3a,b). *Suv39h2* and *Suv39h1* single-knockout cells exhibited similar tendencies (Fig. 3a,b). The H3K9me3 chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) results were consistent with the immunostaining results (Fig. 3c), suggesting that in mouse ES cells, SUV39H2 is the main enzyme responsible for pericentric H3K9me3. To determine the subcellular localization of endogenous SUV39H1

in H2 TKO cells, we performed immunostaining experiments. Consistent with the delocalized H3K9me3 in H2 TKO cells (Fig. 3a–c), the endogenous SUV39H1 did not form focal distributions at pericentric regions (Fig. 3d). Notably, this was a ZNF512- and ZNF512B-dependent event, as the endogenous SUV39H1 remained at the pericentric foci in SUV39H2 single-knockout cells (Fig. 3d). By contrast, endogenous SUV39H2 localized at pericentric heterochromatin regions in H1 TKO cells (Extended Data Fig. 5c), consistent with the normal H3K9me3 signal at the DAPI-dense pericentric regions in H1 TKO cells (Fig. 3a–c).

Overexpressed active but not inactive SUV39H homologues localized to pericentric regions in the absence of *Znf512* and *Znf512b* (Extended Data Fig. 5b). Furthermore, overexpressed but not endogenous SUV39H1 localized to pericentric regions in the absence of *Znf512* and *Znf512b* (Fig. 3d and Extended Data Fig. 5b). These results collectively indicate that ZNF512- and ZNF512B-mediated direct recruitment and H3K9me3-dependent enrichment are key mechanisms underlying the pericentric localization of SUV39H homologues. Notably, despite SUV39H2 making a greater contribution to catalysis of pericentric H3K9me3 (Fig. 3a–c), the H1 TKO cells exhibited greater derepression effects on major satellites (Fig. 3e). To understand this, we performed label-free quantitative-based mass spectrometry (LFQ-MS) of the protein interactome of HA-tagged SUV39H1 and HA-tagged SUV39H2 and observed that HP1 proteins preferentially associated with SUV39H1 (Fig. 3f). In line with this observation, although H3K9m3 foci remained largely normal (Fig. 3a,b), pericentric enrichment of HP1 proteins was lost in most H1 TKO cells (Fig. 3g). This indicates that the presence of SUV39H1 probably has a stronger HP1 recruitment effect than H3K9me3.

The above results indicate that SUV39H1 and SUV39H2 possess distinct properties. SUV39H2 shows a greater preference for H3 tails and seems to be a more efficient enzyme<sup>37,38</sup>, whereas SUV39H1 probably has a stronger silencing effect owing to its greater affinity with HP1 proteins (Fig. 3f). The distinct properties of SUV39H1 and SUV39H2 have also been reported in early embryos<sup>39</sup>.

## Recognition of non-consecutive motifs

ZNF512 has three C2H2-type zinc-fingers, and ZNF512B has four (Extended Data Fig. 6a). A classic C2H2 zinc-finger comprises 28 residues, and structural studies have shown that the so-called zinc-fingerprint residues at positions –1, 3 and 6 of the C2H2 zinc-fingers are responsible for base-specific interactions with a three-nucleotide triplet<sup>40</sup>, determining the sequence recognition specificity. All three zinc-fingers (C1, C2 and C3) of ZNF512 and three of the four zinc-fingers (C1, C2 and C3) of ZNF512B have identical fingerprint residues (–1, S; 3, G; 6, Y) (Extended Data Fig. 6b). This indicates that all these zinc-fingers recognize the same three-nucleotide triplet, predicted to be TTC (Extended Data Fig. 6c) using a polynomial SVM-based algorithm<sup>41</sup> ([www.zf.princeton.edu](http://www.zf.princeton.edu)). We mutated the above-mentioned fingerprint residues from SGY to AAA for each zinc-finger or in combination. When all fingerprint residues were mutated, ZNF512 and ZNF512B did not localize to DAPI-dense foci, and instead accumulated in the nucleolus (Fig. 4a). Mutating the zinc-fingerprint residues of one or two but not all zinc-fingers had intermediate effects (Extended Data Fig. 6d,e). These results indicate that ZNF512 and ZNF512B indeed use their zinc-fingers to target pericentric sequences.

ZNF512 and ZNF512B are evolutionarily conserved proteins among vertebrates, and their C1, C2 and C3 zinc-fingers strictly contain identical fingerprint residues in vertebrates (Extended Data Fig. 7). This led us to question whether they could recognize pericentric repeats across different vertebrate species, which may seem counterintuitive owing to the significant variation in pericentric repeat sequences among different species<sup>2,5</sup> (Extended Data Fig. 1b). Surprisingly, ectopically expressed mouse ZNF512 and ZNF512B localized clearly to DAPI-dense pericentric heterochromatin regions in human HEK293FT cells, and

ZNF512 and ZNF512B orthologues from zebrafish and humans exhibited pericentric heterochromatin localization in mouse ES cells (Fig. 4b). These results indicate that seemingly non-conserved pericentric sequences can be targeted by the conserved ZNF512 and ZNF512B, probably through an unconventional mechanism.

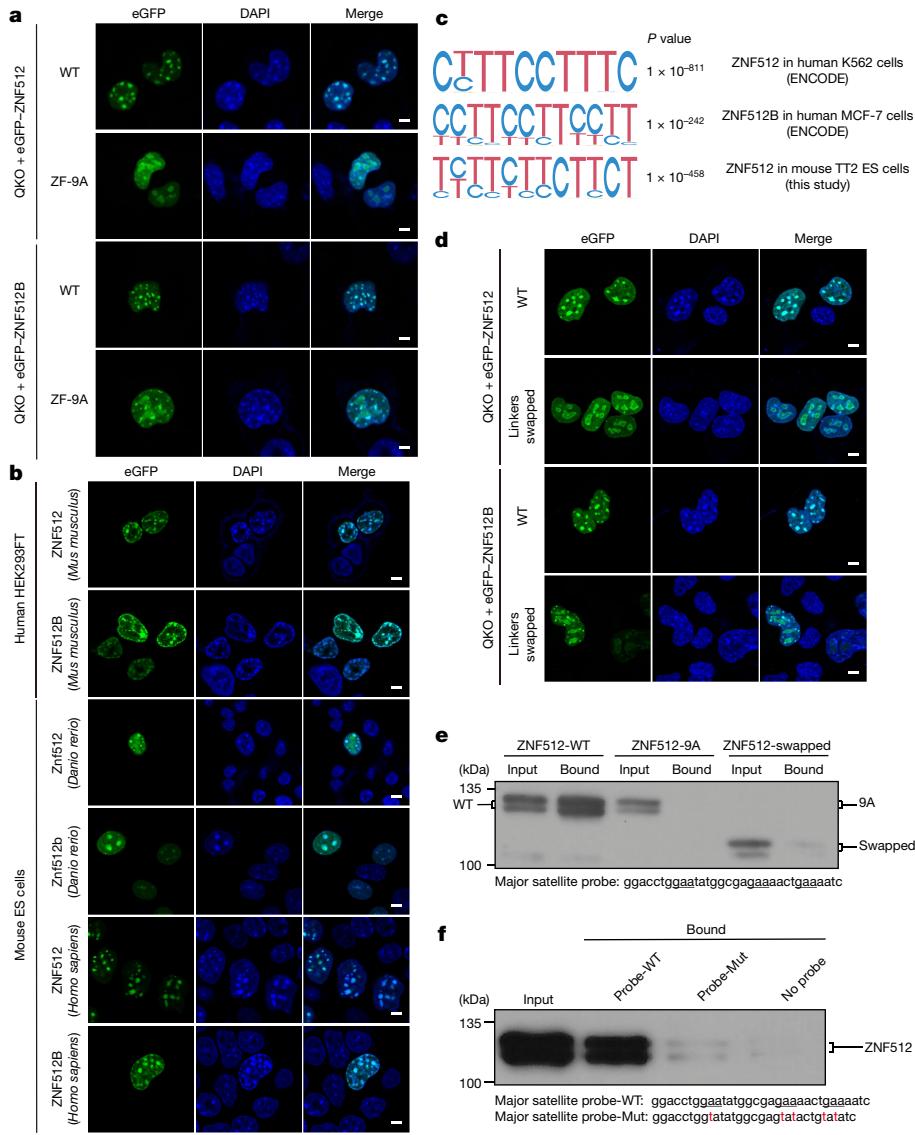
We called the consensus binding motifs of ZNF512 and ZNF512B homologues on the basis of our ChIP-seq data and the ENCODE data. In agreement with the predicted binding motif, the consensus binding motifs contain multiple TTC triplets (Fig. 4c). However, the mouse major satellite repeats did not strictly contain the ZNF512-binding motifs, although ZNF512 was indeed enriched at the pericentric repeats (Extended Data Fig. 8a). Nevertheless, multiple dispersed TTC triplets existed in the mouse major satellite repeats (Extended Data Fig. 1b), and the frequency of TTC triplets was enriched 3.3-fold in mouse major satellite repeats (Extended Data Fig. 8b). The human genome contains different types of pericentric repeat<sup>2,5</sup>. Notably, the frequency of TTC triplets was enriched 4.4-fold in human satellite III and 4.0-fold in human satellite II but not enriched in human satellite I (Extended Data Fig. 8c). Human ZNF512 and ZNF512B were preferentially enriched at satellite III and satellite II (Extended Data Fig. 8d).

Zinc-finger proteins typically contain a seven- or eight-residue spacer between two adjacent zinc-fingers, ensuring the spatial adjacency of consecutively organized zinc-fingers for recognizing consecutive DNA sequence motifs<sup>42</sup>. However, ZNF512 and ZNF512B have an atypical organization of their zinc-fingers, with long linkers separating them. In mouse ZNF512, the zinc-fingers are separated by 66 and 128 residues, and in mouse ZNF512B, the zinc-fingers are separated by 352, 66 and 130 residues (Extended Data Fig. 6a). AlphaFold predicted that the zinc-fingers of ZNF512 and ZNF512B would not be spatially close to each other (Extended Data Fig. 8e). Therefore, we suspected that the split zinc-fingers of ZNF512 and ZNF512B may not necessarily bind consecutive nine-base or 12-base DNA sequence motifs *in vivo*, as the longer linkers may allow a more flexible binding mode. To test this idea, we replaced the long linkers of ZNF512 and ZNF512B with standard eight-residue linkers from a typical zinc-finger protein, ZNF280C. AlphaFold predicted that the zinc-fingers of linker-swapped ZNF512 and ZNF512B would become spatially close to each other (Extended Data Fig. 8f). Linker-swapping abolished the pericentric localization of ZNF512 and ZNF512B (Fig. 4d and Extended Data Fig. 8g), confirming the importance of the long linkers in recognition of the dispersed TTC triplets within major satellite repeats. Moreover, unlike the wild-type proteins, linker-swapped ZNF512 and ZNF512B failed to restore the pericentric H3K9me3 foci and the pericentric localization of endogenous SUV39H1 in H2 TKO cells (Extended Data Fig. 8h,i).

We next explored the direct DNA-binding properties of ZNF512 and ZNF512B by DNA pull-down assay. Recombinant ZNF512 and ZNF512B proteins were pulled down by a DNA fragment corresponding to part of the mouse major satellite consensus sequence (Extended Data Fig. 8j). Fingerprint residue mutation (ZNF512-9A) or linker-swapping abolished the DNA-binding activity of ZNF512 (Fig. 4e). Moreover, the probe-binding activity of ZNF512 was greatly attenuated when the TTC triplets were mutated (Fig. 4f). We also note that the interaction seemed to be relatively weak, as only approximately 2% of input ZNF512 and ZNF512B was detected in the pull-down fraction (Extended Data Fig. 8j).

## A role of ZNF512 and ZNF512B at LINEs

The obvious next question was whether ZNF512 and ZNF512B facilitated H3K9me3 mediated by SUV39H homologues in other genomic regions. Among 22,761 SUV39H-dependent H3K9me3 peaks, 1,580 overlapped with ZNF512 peaks. These shared peaks were significantly enriched for major satellites, long terminal repeats (LTRs) and long interspersed nuclear elements (LINEs) (Fig. 5a). As LINEs are the major targets of SUV39H outside pericentric heterochromatin<sup>43</sup>, we focused our further analysis on LINEs. We observed that ZNF512 and H3K9me3



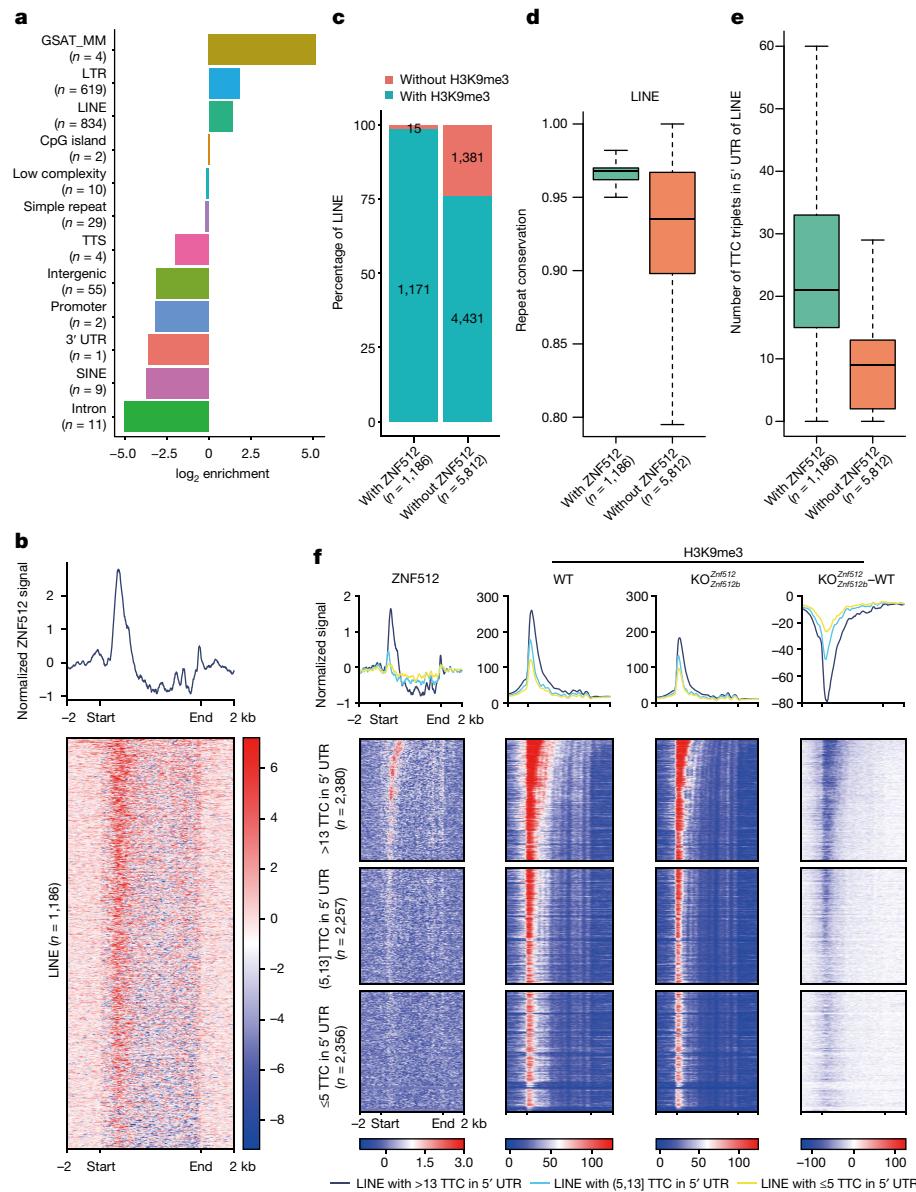
**Fig. 4 | Split zinc-fingers of ZNF512 and ZNF512B recognize non-consecutive TTC triplets in pericentric repeats.** **a**, Representative cross-section fluorescence images showing the relative localization of eGFP–ZNF512 and eGFP–ZNF512B (wild-type or zinc-fingerprint residue-mutated versions) and DAPI in *Znf512*, *Znf512b*, *Suv39h1* and *Suv39h2* QKO cells. Fingerprint residue-mutated versions: three zinc-fingers (C1, C2 and C3) of ZNF512 or three of four zinc-fingers (C1, C2 and C3) of ZNF512B mutated to AAA, abbreviated as ZF-9A. Scale bars, 5  $\mu$ m. **b**, Representative cross-section fluorescence image showing the relative localization of eGFP-tagged ZNF512 and ZNF512B from various species into cells from other species. Scale bars, 5  $\mu$ m. **c**, Top ZNF512-binding motif in human K562 cells, top ZNF512B-binding motif in human MCF-7 cells and top ZNF512-binding motif in mouse ES cells. *P* values were calculated using HOMER. **d**, Representative cross-section fluorescence images showing the

relative localization of eGFP–ZNF512 and eGFP–ZNF512B (wild-type or linker-swapped versions) and DAPI in *Znf512*, *Znf512b*, *Suv39h1* and *Suv39h2* QKO cells. Scale bars, 5  $\mu$ m. **e**, DNA pull-down with biotin-major satellite probe followed by western blot analysis for MBP tag using purified recombinant MBP-Flag-tagged ZNF512 (wild-type, 9A (fingerprint residue mutants), swapped (linkers swapped)). Mobilities of protein markers are indicated on the left. **f**, DNA pull-down with biotin-major satellite probe-WT, or biotin-major satellite probe-Mut, or no probe followed by western blot analysis for MBP tag using purified recombinant MBP-Flag-tagged ZNF512. The mobilities of protein markers are indicated on the left. In **a**, **b**, **d**–**f**, representative data from two independent biological repeats are shown. In **c**, peaks detected in both two technical replicates were used to detect motifs.

accumulation at LINE elements was concentrated at the 5' untranslated region (UTR) (Fig. 5b and Extended Data Fig. 9a), similar to the distribution pattern for SUV39H and H3K9me3 reported previously<sup>43</sup>. On the basis of ZNF512 and H3K9me3 occupancy, we observed that among LINEs with ZNF512 peaks, the ratio between LINEs with and without H3K9me3 was 78:1 (1,171:15). By contrast, among LINEs without ZNF512 peaks, the ratio between those with and without H3K9me3 was only 3.2:1 (4,431:1,381) (Fig. 5c). This indicates that LINEs with ZNF512 probably should be preferentially repressed. Indeed, 84% of LINEs with ZNF512 (993 out of 1,186) belonged to the evolutionarily younger subtype L1Md\_T (Extended Data Fig. 9b), and they exhibited

greater repeat conservation (Fig. 5d) and contained more TTC triplets at their 5' UTRs (Fig. 5e).

To examine the functional significance of the above observation, we categorized the full-length (greater than 6 kb) LINEs into different groups according to the number of TTC triplets within their 5' UTRs. The group with the highest number of TTC triplets exhibited the highest ZNF512 binding, the highest H3K9me3 occupancy and the greatest reduction in H3K9me3 signal upon knockout of *Znf512* and *Znf512b* (Fig. 5f). These results support a role of ZNF512 and ZNF512B in facilitating SUV39H-dependent H3K9me3 outside pericentric repeats. Given that SETDB1 and HUSH have a well-established role in silencing



**Fig. 5 | ZNF512 and ZNF512B facilitate SUV39H-dependent H3K9me3 formation at LINEs.** **a**, Enrichment of various genomic elements in SUV39H-dependent H3K9me3 peaks bound by ZNF512. *n* indicates the number of SUV39H-dependent H3K9me3 peaks annotated to each genomic element. **b**, Profile (top) and heatmap (bottom) showing normalized ZNF512 signals around ZNF512-occupied LINEs (*n* = 1,186). The x axis represents the distance from the start or end of the LINE in kilobases. **c**, Percentage of H3K9me3 modification in LINEs with ZNF512 binding (*n* = 1,186) and those without ZNF512 binding (*n* = 5,812). **d**, Box plots showing repeat conservation of LINEs with ZNF512 binding (*n* = 1,186) and those without ZNF512 binding (*n* = 5,812). **e**, Box plots showing the number of TTC triplets in the 5' UTR of LINEs with ZNF512 binding (*n* = 1,186) and those without ZNF512 binding (*n* = 5,811).

In box plots (d,e) centre line denotes median; box denotes 25th and 75th percentiles; and whiskers denote 1.5× interquartile range. **f**, Profile (top) and heatmap (bottom) showing ZNF512 signals, H3K9me3 signals in wild-type or *Znf512* and *Znf512b* double-knockout mouse ES cells around LINEs. The subtraction of H3K9me3 in *Znf512* and *Znf512b* double-knockout from wild-type mouse ES cells is also shown. The x axis represents the distance from the start or end of the LINE in kilobases. Heatmaps are sorted by the number of TTC triplets in the 5' UTR of LINEs. LINEs were classified on the basis of the number of TTC triplets in their 5' UTR. Dark blue indicates LINEs with more than 13 TTC triplets (*n* = 2,380) and yellow those with fewer than five TTC triplets (*n* = 2,356); other LINEs (*n* = 2,257) are shown in light blue. Data are from two independent technical replicates. GSAT\_MM, mouse major satellite repeats.

LINEs<sup>44,45</sup>, it is interesting to speculate about potential cross-talk between the above two mechanisms, because the H3K9me3-binding activity of HUSH complex subunit MPP8 may not be able to distinguish H3K9me3 catalysed by SUV39H proteins and SETDB1.

## Discussion

On the basis of our findings and prior knowledge, we propose the following model. In vertebrates, ZNF512 and ZNF512B homologues recognize repetitive and non-consecutive TTC sequences in pericentric

repeats through their split zinc-fingers. Subsequently, they recruit SUV39H homologues to initiate heterochromatin formation. Once heterochromatin has been initiated, further mechanisms, such as H3K9me3, HP1, SUV4-20, DNA methylation and major satellite RNA, participate in its maintenance<sup>5,9,11,13,18–21,28,46–49</sup>. In specific cell types, other DNA-binding proteins such as PAX3, PAX9 and FOXD3 may contribute to heterochromatin silencing<sup>22,23</sup>. Given that deterioration of heterochromatin is a hallmark of ageing<sup>50</sup>, an interesting future direction would be to investigate whether the above mechanisms contribute to ageing-associated heterochromatin decay. The recognition of

non-consecutive DNA sequences by the split zinc-fingers of ZNF512 and ZNF512B is of importance, as it shows that the target motif of a DNA-binding protein does not necessarily have to be a consecutive sequence. This discovery may lead to the identification of more such examples and could aid bioinformaticians in developing new algorithms to calculate DNA-binding motifs, taking the non-consecutive regulatory information into consideration.

In conclusion, our study addresses two long-standing questions: the initiation of pericentric constitutive heterochromatin formation; and the targeting of variable pericentric sequences by conserved machinery in vertebrates.

## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-024-07640-5>.

1. Allshire, R. C. & Madhani, H. D. Ten principles of heterochromatin formation and function. *Nat. Rev. Mol. Cell Biol.* **19**, 229–244 (2018).
2. Janssen, A., Colmenares, S. U. & Karpen, G. H. Heterochromatin: guardian of the genome. *Annu. Rev. Cell Dev. Biol.* **34**, 265–288 (2018).
3. Grewal, S. I. S. & Jia, S. Heterochromatin revisited. *Nat. Rev. Genet.* **8**, 35–46 (2007).
4. Becker, J. S., Nicetto, D. & Zaret, K. S. H3K9me3-dependent heterochromatin: barrier to cell fate changes. *Trends Genet.* **32**, 29–41 (2016).
5. Saksouk, N., Simboeck, E. & Dejardin, J. Constitutive heterochromatin formation and transcription in mammals. *Epigenetics Chromatin* **8**, 3 (2015).
6. Heitz, E. *Das heterochromatin der moose* (Bornträger, 1928).
7. Muller, H. J. Types of visible variations induced by x-rays in *Drosophila*. *J. Genet.* **22**, 299–334 (1930).
8. Rea, S. et al. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**, 593–599 (2000).
9. Lachner, M., O'Carroll, D., Rea, S., Mechler, K. & Jenuwein, T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**, 116–120 (2001).
10. Nakayama, J., Rice, J. C., Strahl, B. D., Allis, C. D. & Grewal, S. I. S. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* **292**, 110–113 (2001).
11. Peters, A. H. F. M. et al. Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* **107**, 323–337 (2001).
12. Peters, A. H. F. M. et al. Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. *Mol. Cell* **12**, 1577–1589 (2003).
13. Bannister, A. J. et al. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410**, 120–124 (2001).
14. Volpe, T. A. et al. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297**, 1833–1837 (2002).
15. Verdel, A. et al. RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* **303**, 672–676 (2004).
16. Hall, I. M. et al. Establishment and maintenance of a heterochromatin domain. *Science* **297**, 2232–2237 (2002).
17. Jia, S. T., Noma, K. & Grewal, S. I. S. RNAi-independent heterochromatin nucleation by the stress-activated ATF/CREB family proteins. *Science* **304**, 1971–1976 (2004).
18. Maisond, C. et al. SUMOylation promotes de novo targeting of HP1alpha to pericentric heterochromatin. *Nat. Genet.* **43**, 220–227 (2011).
19. Velazquez Camacho, O. et al. Major satellite repeat RNA stabilize heterochromatin retention of Suv39h enzymes by RNA-nucleosome association and RNA:DNA hybrid formation. *eLife* **6**, e25293 (2017).
20. Johnson, W. L. et al. RNA-dependent stabilization of SUV39H1 at constitutive heterochromatin. *eLife* **6**, e25299 (2017).
21. Shirai, A. et al. Impact of nucleic acid and methylated H3K9 binding activities of Suv39h1 on its heterochromatin assembly. *eLife* **6**, e25317 (2017).
22. Bulut-Karslioglu, A. et al. A transcription factor-based mechanism for mouse heterochromatin formation. *Nat. Struct. Mol. Biol.* **19**, 1023–1030 (2012).
23. Puri, D. et al. Foxd3 controls heterochromatin-mediated repression of repeat elements and 2-cell state transcription. *EMBO Rep.* **22**, e53180 (2021).
24. Gao, X. D. et al. C-BERST: defining subnuclear proteomic landscapes at genomic elements with dCas9-APEX2. *Nat. Methods* **15**, 433–436 (2018).
25. Myers, S. A. et al. Discovery of proteins associated with a predefined genomic locus via dCas9-APEX-mediated proximity labeling. *Nat. Methods* **15**, 437–439 (2018).
26. Qiu, W. Q. et al. Determination of local chromatin interactions using a combined CRISPR and peroxidase APEX2 system. *Nucleic Acids Res.* **47**, e52 (2019).
27. Ong, S. E., Foster, L. J. & Mann, M. Mass spectrometric-based approaches in quantitative proteomics. *Methods* **29**, 124–130 (2003).
28. Saksouk, N. et al. Redundant mechanisms to form silent chromatin at pericentromeric regions rely on BEND3 and DNA methylation. *Mol. Cell* **56**, 580–594 (2014).
29. Schmidtmann, E., Anton, T., Rombaut, P., Herzog, F. & Leonhardt, H. Determination of local chromatin composition by CasID. *Nucleus* **7**, 476–484 (2016).
30. Robinett, C. C. et al. In vivo localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition. *J. Cell Biol.* **135**, 1685–1700 (1996).
31. Wang, T. et al. Crystal structure of the human SUV39H1 chromodomain and its recognition of histone H3K9me2/3. *PLoS ONE* **7**, e52977 (2012).
32. Zhang, K., Mosch, K., Fischle, W. & Grewal, S. I. S. Roles of the Clr4 methyltransferase complex in nucleation, spreading and maintenance of heterochromatin. *Nat. Struct. Mol. Biol.* **15**, 381–388 (2008).
33. McCarthy, R. L., Zhang, J. & Zaret, K. S. Diverse heterochromatin states restricting cell identity and reprogramming. *Trends Biochem. Sci.* **48**, 513–526 (2023).
34. Grewal, S. I. S. The molecular basis of heterochromatin assembly and epigenetic inheritance. *Mol. Cell* **83**, 1767–1785 (2023).
35. Wang, C., Zhu, B. & Xiong, J. Recruitment and reinforcement: maintaining epigenetic silencing. *Sci. China Life Sci.* **61**, 515–522 (2018).
36. Ragunathan, K., Jih, G. & Moazed, D. Epigenetic inheritance uncoupled from sequence-specific recruitment. *Science* **348**, 1258699 (2015).
37. Kudithipudi, S., Schuhmacher, M. K., Kebede, A. F. & Jeltsch, A. The SUV39H1 protein lysine methyltransferase methylates chromatin proteins involved in heterochromatin formation and VDJ recombination. *ACS Chem. Biol.* **12**, 958–968 (2017).
38. Schuhmacher, M. K., Kudithipudi, S., Kusevic, D., Weirich, S. & Jeltsch, A. Activity and specificity of the human SUV39H2 protein lysine methyltransferase. *Biochim. Biophys. Acta* **1849**, 55–63 (2015).
39. Burton, A. et al. Heterochromatin establishment during early mammalian development is regulated by pericentromeric RNA and characterized by non-repressive H3K9me3. *Nat. Cell Biol.* **22**, 767–778 (2020).
40. Yang, P., Wang, Y. X. & Macfarlan, T. S. The role of KRAB-ZFPs in transposable element repression and mammalian evolution. *Trends Genet.* **33**, 871–881 (2017).
41. Persikov, A. V. & Singh, M. De novo prediction of DNA-binding specificities for Cys2His2 zinc finger proteins. *Nucleic Acids Res.* **42**, 97–108 (2014).
42. Wolfe, S. A., Nekludova, L. & Pabo, C. O. DNA recognition by Cys2His2 zinc finger proteins. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 183–212 (2000).
43. 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).
44. Liu, N. et al. Selective silencing of euchromatic L1s revealed by genome-wide screens for L1 regulators. *Nature* **553**, 228–232 (2018).
45. Seczynska, M. & Lehrer, P. J. The sound of silence: mechanisms and implications of HUSH complex function. *Trends Genet.* **39**, 251–267 (2023).
46. Schotta, G. et al. A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Gene Dev.* **18**, 1251–1262 (2004).
47. Kourmouli, N. et al. Heterochromatin and tri-methylated lysine 20 of histone H4 in animals. *J. Cell Sci.* **117**, 2491–2501 (2004).
48. Lehrhertz, B. et al. Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Curr. Biol.* **13**, 1192–1200 (2003).
49. Maison, C. et al. Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an RNA component. *Nat. Genet.* **30**, 329–334 (2002).
50. Booth, L. N. & Brunet, A. The aging epigenome. *Mol. Cell* **62**, 728–744 (2016).

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2024

## Article

---

## Methods

Grammatical errors were corrected with ChatGPT.

## Cell culture

Mouse ESTT2 cells<sup>51</sup> were cultured on 0.1% gelatin-coated tissue culture plates in the presence of 1,000 U ml<sup>-1</sup> leukaemia inhibitory factor (LIF) (Novoprotein) with DMEM (ThermoFisher Scientific) supplemented with 10% KnockOut Serum Replacement (ThermoFisher Scientific), 1% fetal bovine serum (FBS; Hyclone), 1 mM L-glutamine (Sangon Biotech), 100 mM non-essential amino acids (Sigma-Aldrich), 0.55 mM β-mercaptoethanol (Gibco) and 1% penicillin–streptomycin (Sangon Biotech). HEK293FT cells were cultured in Dulbecco's modified Eagle medium (ThermoFisher Scientific) supplemented with 10% FBS (Biological Industries) and 1% penicillin–streptomycin (Sangon Biotech). CHO A03-1 cells<sup>30</sup> carrying 256 copies of LacO operators were cultured in F-12 Ham's medium (Gibco) supplemented with 10% FBS (ExCell Bio) and 1% penicillin–streptomycin (Sangon Biotech). *Drosophila* S2 cells were cultured with Schneider's Drosophila medium (Gibco) supplemented with 10% FBS (Sigma-Aldrich) at 28 °C. Identity of the TT2, HEK293FT, CHO A03-1 and *Drosophila* S2 cell lines were frequently checked by their morphological features. Cell lines were not tested for mycoplasma contamination.

## **Generation of overexpression cell lines**

The CAG-NLS-Flag-dCas9-NLS plasmid was transfected using Lipofectamine 3000 (ThermoFisher Scientific) following the manufacturer's instructions. After plasmid transfection, puromycin (Gibco) was added to a final concentration of  $1\text{ }\mu\text{g ml}^{-1}$  to select for dCas9 overexpressed cells. These cells were then grown in 96-well plates to obtain single colonies. The expression level of dCas9 was monitored through anti-Flag immunoblotting. The cell line with the highest expression level of dCas9 was expanded and used for the next transfection. The UBC-MCP-NLS-eGFP-Flag-NLS-APEX2 plasmid was then transfected using Lipofectamine 3000 (ThermoFisher Scientific) following the manufacturer's instructions. After transfection, cells with a lower level of eGFP signal were selected using BD FACS Aria III cell sorter and analysed with BD FACS Diva software v.8.0.1 (Supplementary Fig. 2) and grown in 96-well plates to obtain single colonies. The cell line with higher dCas9 expression and lower APEX2 expression was confirmed by anti-Flag immunoblotting. This cell line was expanded and used for subsequent sgRNA transfection. For stable sgRNA expression, the U6-major satellite sgRNA or -Gal4 sgRNA-2 $\times$ MS plasmid was transfected into cells expressing dCas9 and APEX2 with Lipofectamine 3000 (ThermoFisher Scientific) following the manufacturer's instructions. The major satellite sgRNA sequence was 5'-ggcaagaaaaactgaaaatca-3', and the Gal4 sgRNA sequence was 5'-gaacgactagttaggcgtt-3'. Transfected cells were selected using hygromycin (Amresco) and grown in 96-well plates to obtain single colonies. Cell lines with the highest levels of sgRNA were monitored using reverse transcription with quantitative PCR (RT-qPCR), which was performed on an Applied Biosystems 7500 Fast system v.2.3.

For transient expression of eGFP-tagged or mCherry-tagged proteins, CHO A03-1 cells were seeded on plates overnight. Plasmids were transfected into cells with Lipofectamine 3000 (ThermoFisher Scientific) according to the manufacturer's instructions. At 24 h after transfection, cells were seeded on glass coverslips for imaging. For transient expression of eGFP-tagged or Flag-tagged proteins, the *Znf512*, *Znf512b*, *Suv39h1* and *Suv39h2* QKO cell line or *Znf512*, *Znf512b* and *Suv39h2* TKO cell line (H2 THO) was transfected with Lipofectamine 3000 (Thermo-Fisher Scientific) according to the manufacturer's instructions. At 24 h after transfection, cells were seeded on glass coverslips for imaging.

## CRISPR–Cas9-mediated gene editing

KO cell lines were generated using CRISPR–Cas9-mediated genome editing with two sgRNA-expressing plasmids and a donor plasmid

containing homologous recombination arms, T2A sequence, antibiotic resistance gene, stop codons and SV40 poly(A) signal. To knock out *Znf512*, the sgRNA sequences used were 5'-tagggcccagtgtccataa-3' and 5'-cttgcactttccagccacga-3', and the neomycin resistance gene was used to select positive cells. To knock out *Znf512b*, the sgRNA sequences were 5'-aaaagtggacctggaaagg-3' and 5'-tggcctcaactaccacacca-3', and the blasticidin S deaminase gene was used to select positive cells. To knock out *Suv39h1*, the sgRNA sequences were 5'-attctcttagaaacaccaa-3' and 5'-ggttgttaagtgtatgtt-3', and either the blasticidin S deaminase gene or the puromycin N-acetyltransferase gene was used to select positive cells. For *Suv39h2* KO, the sgRNA sequences were 5'-gtatcaagtggaaactaggca-3' and 5'-tgcatctttagaactagcaa-3', and the bleomycin resistance gene was used to select positive cells. The two sgRNA plasmids and one donor plasmid were cotransfected into TT2 mouse ES cells with Lipofectamine 3000 (Life Technologies) following the manufacturer's instructions. After 4 days of transfection, cells were selected using the corresponding antibiotics and then seeded into 96-well plates to obtain single colonies. Two weeks later, cell clones with one colony per well were picked and verified by genotyping PCR. Further validation was done by RT-qPCR and RNA sequencing. The specific primers used for genotyping PCR were as follows. For KO *Znf512* genotyping PCR: KO genotyping F1: 5'-tgtcttcatagggtcatgtcaccc-3'; KO genotyping R1: 5'-ccggccacagtgcgtatcc-3'; KO genotyping F2: 5'-gaccgttcctcggtttacgg-3'; KO genotyping R2: 5'-tggagagagagtcgtgtcagg-3'; KO genotyping F3: 5'-tcagttgttagctgtgtc-3'; KO genotyping R3: 5'-agtccccctccctcagaaacc-3'; KO genotyping F4: 5'-gtaacatctggccagatctgag-3'; KO genotyping R4: 5'-ctcagcacaaatttttgcgttc-3'. For KO *Znf512b* genotyping PCR: KO genotyping F1: 5'-ggttgttcctggcaatgacgg-3'; KO genotyping R1: 5'-ctcatttccgcgtcgcacatacaag-3'; KO genotyping F2: 5'-cttgcattatgtgtgtatcc-3'; KO genotyping R2: 5'-ggccagttcatctctgtcaatctcc-3'; KO genotyping F3: 5'-gtatctcgtataggctggcctccc-3'; KO genotyping R3: 5'-cgctcagggtcttcctggggatgg-3'. For KO *Suv39h1* genotyping PCR: KO genotyping F1: 5'-gttgaagaagggtgtctgtgaagc-3'; KO genotyping R1: 5'-tcatttccgcgtcgcacatacaag-3' or 5'-ccaggaggcccttcatgttc-3'; KO genotyping F2: 5'-cttgcattatgtgtgtatcc-3'; KO genotyping R2: 5'-tgcacacagggtcacacacaaatcc-3'; KO genotyping F3: 5'-ggagtattacctgtttaagtggcgtg-3'; KO genotyping R3: 5'-aagatcgacagatcgtagcgatgcc-3'. For KO *Suv39h2* genotyping PCR: KO genotyping F1: 5'-ggatccggaaatgtatccactgg-3'; KO genotyping R1: 5'-aggcgtcccgaaatgtgtggacac-3'; KO genotyping F2: 5'-cctggacgtgtacccggatgg-3'; KO genotyping R2: 5'-ccctgtaaaatctgtcaactcc-3'; KO genotyping F3: 5'-aggatggccagattctacaacacc-3'; KO genotyping R3: 5'-aagatcgacagatgtgtgtcc-3'.

The cell line with AviTag knocked in at the amino terminus of ZNF512 was generated by CRISPR–Cas9-mediated genome editing. One sgRNA-expressing plasmid and a donor plasmid containing homologous recombination arms and AviTag sequence were transfected into the BirA-expressing TT2 cell line with Lipofectamine 3000 (ThermoFisher Scientific) following the manufacturer's instructions. The sgRNA sequence used was 5'-ttgggtgaaattgttaggcg-3'. Three days after transfection, cells were seeded into 96-well plates by a limited dilution method. After two weeks, cell clones with one colony per well were verified by genotyping PCR. The following primers were used for KI genotyping PCR: KI genotyping F1: 5'-cgaggctcagaaaatcaatggc-3'; KI genotyping R1: 5'-gccacgcaggaaataccactaactg-3'; KI genotyping F2: 5'-acagtctttaaggggagaggcc-3'; KI genotyping R2: 5'-tgagecctcgaaatgtcggtcagg-3'; KI genotyping F3: 5'-ggcggaaagttaggttgcagaacagg-3'; KI genotyping R3: 5'-gtacaacaaacaaaggcaaggaccgcg-3'.

## SILAC labelling

SILAC labelling was performed as described previously<sup>52</sup> with a few modifications. In brief, mouse ES cells (TT2) were cultured on dishes coated with 0.1% gelatin (Millipore). The culture medium contained DMEM for SILAC (ThermoFisher Scientific), 10% dialysed FBS (Life Technologies),

1 mM L-glutamine (Sangon Biotech), 100 mM non-essential amino acids (Sigma-Aldrich), 0.55 mM  $\beta$ -mercaptoethanol (Gibco), 1% penicillin-streptomycin (Sangon Biotech), 1,000 U ml<sup>-1</sup> leukaemia inhibitory factor (Novoprotein), 91  $\mu$ g ml<sup>-1</sup> L-lysine: HCl (for heavy medium, K8, Cambridge Isotope Laboratories, CNLM-291-0.5; for light medium, K0, Sigma, L8662) and 89  $\mu$ g ml<sup>-1</sup> L-arginine: HCl (for heavy medium, R10, Cambridge Isotope Laboratories, CNLM-539-0.5; for light medium, R0, Sigma, A8094). In the forward experiment, major satellite DNA-targeting cells were grown in heavy SILAC medium (K8 + R10), and Gal4 sgRNA-expressing cells were grown in light SILAC medium (K0 + R0). In the reverse experiment, Gal4 sgRNA-expressing cells were grown in heavy SILAC medium (K8 + R10), whereas major satellite-targeting cells were grown in light SILAC medium (K0 + R0). TT2 cells were grown for ten passages to ensure sufficient incorporation of the isotopes. To evaluate the efficiency of SILAC labelling, histones from cells cultured in K8 + R10 mediums were extracted using acid extraction<sup>53</sup>. The extracted histones were then separated by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Subsequently, histone proteins were stained with Coomassie Brilliant Blue, and gel slices were excised for in-gel tryptic digestion and mass spectrometry analysis.

#### Proximity labelling and nuclear extract preparation

Proximity labelling was performed as described previously<sup>54</sup> with a few modifications. TT2 cells expressing APEX2 were incubated with 500  $\mu$ M biotin phenol (BP) in complete medium for 30 min at 37 °C. The BP-containing medium was prewarmed to 37 °C to facilitate the dissolution of BP. Then, an appropriate amount of H<sub>2</sub>O<sub>2</sub> was added to the BP solution directly, followed by agitation briefly, to achieve a final concentration of 1 mM. After incubation of cells at room temperature for 1 min, the labelling solution was rapidly aspirated, and the cells were washed three times with a quencher solution. The quencher solution was prepared freshly and composed of Dulbecco's PBS (DPBS) buffer supplemented with 10 mM sodium ascorbate (Sangon Biotech), 5 mM Trolox (Acros Organics) and 10 mM sodium azide (GenStar). The cells were scraped from the bottom of the well with fresh quencher solution and then pelleted by centrifugation for 10 min at 3,000g and 4 °C. Nuclear extracts were prepared as described previously<sup>50</sup> with a few modifications. Cell pellets were lysed by gentle pipetting in buffer A (20 mM Tris-HCl, pH 8.0, 1.5 mM MgCl<sub>2</sub>, 10 mM KCl, 0.5 mM DTT and 10% glycerol) and buffer C (20 mM Tris-HCl, pH 8.0, 420 mM NaCl, 1.5 mM MgCl<sub>2</sub>, 1 mM EDTA and 10% glycerol), supplemented with protease inhibitor cocktail (Roche), 1 mM phenylmethylsulfonyl fluoride, 10 mM sodium azide, 10 mM sodium ascorbate, and 5 mM Trolox. The samples were flash-frozen and stored at -80 °C.

#### Enrichment of biotinylated proteins

Biotinylated cell lysate samples in buffer C were quantified. To enrich biotinylated proteins, a 40  $\mu$ l aliquot of 50% streptavidin agarose bead slurry (Millipore) was used for each sample. Each aliquot of beads was washed twice with 1 ml of buffer C supplemented with 0.1% NP-40. Then, 1.5 mg of heavy cell lysate sample and 1.5 mg of light cell lysate sample were incubated with 40  $\mu$ l of streptavidin agarose beads at 4 °C overnight on a rotator. The remaining cell lysate was saved for gel and western blot analyses. Beads were then pelleted into a purification column, and the flowthrough was collected for subsequent analysis. To remove non-specific binders, beads were washed with a series of buffers: twice with 1 ml buffer C, once with 1 ml 1 M KCl, once with 1 ml 0.1 M Na<sub>2</sub>CO<sub>3</sub>, once with 1 ml 2 M urea in 10 mM Tris-HCl (pH 8.0) and twice with 1 ml buffer C. Washing buffers were kept on ice throughout the procedure. Beads were then collected by centrifuging at 2,000 rpm and 4 °C. Biotinylated proteins were eluted from the beads by boiling in 50  $\mu$ l of protein loading buffer (10 mM Tris-HCl, pH 6.8, 1% SDS) supplemented with 2 mM biotin and 20 mM DTT for 10 min. Beads were vortexed briefly and spun down at

12,000 rpm for 10 min. Elutes were collected and stored at -80 °C for electrophoresis.

#### Trypsin digestion, liquid chromatography coupled with tandem mass spectrometry, and data analysis

Proteins separated on SDS-PAGE were stained by Coomassie Brilliant Blue and then cut into 1 cm pieces for subsequent in-gel trypsin digestion. Liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS) experiments were performed on a Q Exactive mass spectrometer (Thermo Scientific) equipped with an Easy n-LC 1000 high-performance liquid chromatography (HPLC) system (Thermo Scientific). The peptides were loaded onto a 100  $\mu$ m i.d.  $\times$  2 cm fused silica trap column packed in-house with reversed-phase silica (Reprosil-Pur C18 AQ, 5  $\mu$ m, Dr. Maisch GmbH) and then separated on a 75  $\mu$ m i.d.  $\times$  20 cm C18 column packed with reversed-phase silica (Reprosil-Pur C18 AQ, 3  $\mu$ m, Dr. Maisch GmbH). The peptides bound on the column were eluted with a 75 min linear gradient. Solvent A consisted of 0.1% formic acid in H<sub>2</sub>O, and solvent B consisted of 0.1% formic acid in acetonitrile solution. The segmented gradient was 4–12% B, 5 min; 12–22% B, 50 min; 22–32% B, 12 min; 32–90% B, 1 min; 90% B, 7 min at a flow rate of 300 nl min<sup>-1</sup>. We performed the MS analysis with a Q Exactive mass spectrometer (Thermo Scientific). The MS data were acquired in data-dependent acquisition mode at a high resolution of 70,000 (200 m/z) across a mass range of 300–1,600 m/z. The target value was 3  $\times$  10<sup>6</sup> with a maximum injection time of 60 ms. The top 20 precursor ions were selected from each MS full scan with an isolation width of 2 m/z for fragmentation in the HCD collision cell with a normalized collision energy of 27%. Subsequently, MS/MS spectra were acquired at a resolution of 17,500 at 200 m/z. The target value was 5  $\times$  10<sup>4</sup> with a maximum injection time of 80 ms. The dynamic exclusion time was 40 s. The nano electrospray ion source settings were as follows: spray voltage 2.0 kV; no sheath gas flow; heated capillary temperature 320 °C. The raw data from the Q Exactive were analysed with Proteome Discovery v.2.2.0.388 using the Sequest HT search engine for protein identification and Percolator for false discovery rate (FDR) analysis. The UniProt mouse protein database (updated October 2017) was used to search data from the mouse sample. Some important searching parameters were set as follows: trypsin was selected as the enzyme, and two missed cleavages were allowed for searching; the mass tolerance of precursor was set to 10 ppm, and the product ion tolerance was 0.02 Da; the cysteine carbamidomethylation was specified as fixed modifications and the methionine oxidation as variable modifications; and KORO and K8R10 were selected as light and heavy labels, respectively. FDRs were analysed with Percolator, and FDR < 1% was set for protein identification. The peptide confidence was set as high for the peptide filter. Protein label-free quantification was also performed with Proteome Discovery v.2.2.0.388 using the areas of identified peptides. Only unique and razor peptides of proteins were selected for protein relative quantification. MS data are provided in the source data for Fig. 1. For SILAC experiments, protein identification was subject to a two-peptide cut-off. SILAC datasets were then filtered to exclude proteins with heavy versus light ratios detected in only one orientation. The values of the heavy versus light ratio and light versus heavy ratio were normalized to the Gaussian distribution for graphing.

#### Trypsin digestion, LFQ-MS and data analysis

Proteins separated by SDS-PAGE were stained with Coomassie Brilliant Blue, and the gel was cut into 1 cm pieces. In-gel trypsin digestion was performed, followed by nanoLC-MS/MS experiments on an Orbitrap Exploris 480 (Thermo Scientific) equipped with an Easy n-LC 1200 HPLC system (Thermo Scientific). The peptides were loaded on to a 100  $\mu$ m i.d.  $\times$  2 cm fused silica trap column packed in-house with reversed-phase silica (Reprosil-Pur C18 AQ, 5  $\mu$ m, Dr. Maisch GmbH) and then separated on a 75  $\mu$ m i.d.  $\times$  25 cm C18 column packed with reversed-phase silica (Reprosil-Pur C18 AQ, 1.9  $\mu$ m, Dr. Maisch GmbH). The peptides bound

# Article

on the column were eluted with a 73 min linear gradient. Solvent A consisted of 0.1% formic acid in H<sub>2</sub>O solution, and solvent B consisted of 80% acetonitrile and 0.1% formic acid. The segmented gradient was as follows: 4–9% B, 3 min; 9–20% B, 22 min; 20–30% B, 20 min; 30–40% B, 15 min; 40–95% B, 3 min; 95% B, 10 min at a flow rate of 300 nl min<sup>-1</sup>. The MS analysis used an Orbitrap Exploris 480 mass spectrometer with the FAIMS Pro interface (Thermo Scientific). FAIMS separations were performed with two compensation voltages (~45 and ~65). Using data-dependent acquisition mode, the MS data were acquired at a high resolution of 60,000 ( $200\text{ m/z}$ ) across a mass range of 350–1,500  $m/z$ . The target value was set to  $3 \times 10^6$  with a maximum injection time of 22 ms. Data-dependent mode was selected as the cycle time mode was set to 2 s. The precursor ions were selected from each MS full scan with an isolation width of 1.6  $m/z$  for fragmentation in the ion routing multipole with a normalized collision energy of 28%. Subsequently, MS/MS spectra were acquired at a resolution of 15,000 at  $200\text{ m/z}$ . The target value was set to  $7.5 \times 10^4$  with a maximum injection time of 22 ms. The dynamic exclusion time was 40 s. The nano electrospray ion source settings were as follows: the spray voltage was set to 2.0 kV with no sheath gas flow, and the heated capillary temperature was 320 °C. Raw data from Orbitrap Exploris 480 were analysed with Proteome Discovery v.2.4.1.15 using the Sequest HT search engine for protein identification. The UniProt human protein database (updated September 2018) was used for searching data from samples. Some important searching parameters were set as follows: trypsin was selected as the enzyme, and two missed cleavages were allowed for searching; the mass tolerance of the precursor was set to 10 ppm, and the product ion tolerance was 0.02 Da; the cysteine carbamidomethylation was specified as fixed modifications, whereas the methionine oxidation was chosen to be variable modifications. We performed FDR analysis with Percolator, and FDR < 1% was set for protein identification. Protein label free quantification was also performed on Proteome Discovery using the areas of identified peptides. Only unique and razor peptides of proteins were selected for protein relative quantification. MS data are provided in the source data for Fig. 3. Protein identification was subject to a two-peptide cut-off. The values of the SUV39H1 versus HEK293FT and SUV39H2 versus HEK293FT ratios were normalized to the Gaussian distribution for graphing.

## RNA preparation, RT-qPCR and RNA sequencing

Total RNA was extracted using TRIzol (Invitrogen) according to the user guide. Subsequently, 500 ng of RNA was reverse-transcribed to cDNA using the HiScript II Q RT SuperMix for qPCR (+gDNA wiper) (Vazyme). We performed the RT-qPCR analysis with an Applied Biosystems 7500 Fast system v.2.3 and Microsoft Excel, using the KAPA SYBR FAST Universal qPCR kit (KAPA Biosystems). Expression levels of different sgRNAs and gene expression fold changes were normalized to *Gapdh* transcripts, which served as an internal standard. RNA libraries were prepared at Berry Genomics and then sequenced on a NovaSeq 6000 at the same facility. If major satellite RNA was detected in the RNA sample, genomic DNA was digested with Turbo DNase (Ambion) according to the manufacturer's instructions, followed by clean-up with TRIzol again. The following primers were used for quantification by qPCR. Major satellite forward (F): 5'-gacgacttggaaaaatgacgaaatc-3'; major satellite reverse (R): 5'-catatccaggcttcgtgtgc-3'; *Gapdh* F: 5'-gcctggagaaacctgcaag-3'; *Gapdh* R: 5'-cgccatcgaaagggttggagag-3'; major satellite sgRNA F: 5'-ggcaagaaaactgaaatcgtttt-3'; major satellite sgRNA R: 5'-aaagcacccactcggtgccacttgg-3'; Gal4 sgRNA F: 5'-gaacgacttaggttggcggtgtgtttt-3'; Gal4 sgRNA R: 5'-aaagcacccactcggtgccacttgg-3'. RT-qPCR data of major satellite RNA in different knock-out cell lines are provided in the source data for Extended Data Fig. 4.

## Immunofluorescence and imaging data analysis

Cells were cultured on a coverslip coated with gelatin. After 24 h, cells were washed three times with DPBS briefly and then fixed in a 4%

paraformaldehyde solution (Beyotime Biotechnology) for 15 min at room temperature. Fixed cells were washed three times with DPBS then permeabilized with 0.5% Triton X-100 in DPBS for 15 min and washed with DPBS three times. Cells were then blocked with 10% FBS in DPBS supplemented with 0.1% Triton X-100 for 1 h at room temperature. Primary antibody staining was performed in 10% FBS in DPBS supplemented with 0.1% Triton X-100 overnight at 4 °C, using the following concentrations: H3K9me3 (1:1,000, Abcam, ab8898), SUV39H1 (1:1,000, CST, 8729), SUV39H2 (1:1,000, HUABIO, ET7108-37), Flag (1:1,000, Sigma-Aldrich, F3165), HP1α (1:1,000, Millipore, MAB3584) and HP1β (1:800, Cell Signaling Technology, 8676). Cells were then washed three times with DPBS supplemented with 0.1% Triton X-100 and incubated with AlexaFluor 647 donkey anti-rabbit IgG(H+L) (1:1,000, Life Technologies, A31573) or AlexaFluor 488 goat anti-rabbit IgG(H+L) (1:1,000, Invitrogen, A11008) or goat anti-rabbit IgG(H+L) AlexaFluor plus 555 (1:1,000, Invitrogen, A32732) or goat anti-mouse IgG(H+L) AlexaFluor plus 555 (1:1,000, Invitrogen, A32727), raised in 10% FBS in DPBS supplemented with 0.1% Triton X-100 for 1 h at room temperature. All subsequent steps were protected from light. Cells were washed with DPBS supplemented with 0.1% Triton X-100 three times then counterstained with 1 µg ml<sup>-1</sup> DAPI (Sigma-Aldrich) in DPBS for 3 min. After another three washes with DPBS, the coverslip was sealed with nail polish; then, the cells were imaged with a Zeiss LSM700 confocal laser scanning microscope, and data were collected using ZEN 2012 black edition software. Images were quantitatively analysed using ZEN 2012 blue edition and ImageJ (v.2.0.0-rc-69/1.52p) software, and colocalization was analysed using the Coloc 2 method. Pearson's R values (above threshold) for all regions of interest were analysed using GraphPad Prism 9. Data of the colocalisation coefficient are provided in the source data for Figs. 2 and 3.

For eGFP or mCherry staining, the primary antibody staining and secondary antibody staining steps were omitted. After fixation, permeation and blocking, cells were stained directly with 1 µg ml<sup>-1</sup> DAPI in DPBS.

For biotin staining, after fixation, permeation and blocking, cells were stained with Cy3-conjugated streptavidin (1:150, Sangon Biotech, D111115) at room temperature for 1 h. After washing with DPBS supplemented with 0.1% Triton X-100 three times, cells were stained with 1 µg ml<sup>-1</sup> DAPI in DPBS.

## Co-IP

For *in vivo* interaction studies, HEK293FT cells were transiently transfected with plasmids encoding Flag-tagged ZNF512, Flag-tagged ZNF512B, HA-tagged SUV39H1 and HA-tagged SUV39H2 using polyethyleneimine as the transfection reagent. After 48 h of transfection, cells were lysed in buffer A and buffer C. The cell nuclear extracts were subjected to immunoprecipitation with anti-Flag M2 affinity gel (Sigma-Aldrich) under 150 mM NaCl conditions overnight at 4 °C. The beads were then extensively washed with buffer C supplemented with 0.1% NP-40, BC500 (20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 1 mM DTT, 1 mM EDTA, 10% glycerol) supplemented with 0.1% NP-40, and BC100 (20 mM Tris-HCl, pH 8.0, 100 mM NaCl, 1 mM DTT, 1 mM EDTA, 10% glycerol). The bound proteins were eluted using Flag peptide in BC100. Elutes were then separated by SDS-PAGE and analysed with antibodies to detect the Flag tag and HA tag.

## Western blotting

Western blotting was performed using standard procedures. In brief, samples were separated using SDS-PAGE and transferred on to polyvinylidene fluoride membranes (Millipore). Membranes were blocked with 5% milk in Tris-buffered saline with 0.1% Tween 20 (TBST) and incubated with primary antibodies at the following concentrations: anti-Flag (1:1,000, Abmart, TT0003), anti-HA (1:1,000, CST, 3724), anti-MBP (1:7,500, Proteintech, 66003-1-Ig) for 1 h at room temperature or overnight at 4 °C. Membranes were washed three times with TBST

and then incubated with HRP-conjugated AffiniPure donkey anti-mouse IgG(H+L) (1:10,000, Proteintech, SA00001-8) or HRP-conjugated AffiniPure donkey anti-rabbit IgG(H+L) (1:10,000, Proteintech, SA00001-9) for 1 h at room temperature. Membranes were washed three times with TBST and imaged. Uncropped scans of blots are shown in Supplementary Fig. 1.

### Recombinant protein expression and purification

DNA fragments encoding Flag-tagged ZNF512, Flag-tagged ZNF512B, HA-tagged SUV39H1, HA-tagged SUV39H2 and HA-tagged GFP-APEX2 were cloned into the pET30a-MBP plasmid. The recombinant proteins were induced by isopropyl-beta-D-thiogalactopyranoside in *Escherichia coli* at 18 °C overnight. Cells were then collected by centrifugation at 3,000 rpm and resuspended in DPBS. After centrifugation, cells were resuspended in lysis buffer (20 mM Tris-HCl, pH 8.0, 300 mM NaCl, 10% glycerol) and sonicated. The supernatants were incubated with Dextrin beads (Smart-Lifesciences) at 4 °C for 2 h. The dextrin beads were then washed with 300 mM NaCl lysis buffer (20 mM Tris-HCl, pH 8.0, 300 mM NaCl, 10% glycerol) supplemented with 0.1% NP-40, 500 mM NaCl lysis buffer (20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 10% glycerol) supplemented with 0.1% NP-40, and 750 mM NaCl lysis buffer (20 mM Tris-HCl, pH 8.0, 750 mM NaCl, 10% glycerol) supplemented with 0.1% NP-40, respectively. Recombinant proteins were eluted with 300 mM NaCl lysis buffer supplemented with 10 mM maltose.

### Pull-down

For the protein pull-down assay, MBP-Flag-tagged ZNF512 or MBP-Flag-tagged ZNF512B, with MBP-HA-tagged SUV39H1, MBP-HA-tagged SUV39H2 or MBP-HA-tagged GFP-APEX2, were incubated with HA beads (Smart-Lifesciences) in 500 µl of binding buffer (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.5 mM DTT, 1.5 mM MgCl<sub>2</sub>, 0.2 mM EDTA, 10% glycerol) at 4 °C overnight. The beads were washed with a series of buffers to remove non-specific binders: once with 1 ml DPBS supplemented with 0.1% NP-40, twice with 1 ml buffer C supplemented with 0.1% NP-40, once with 1 ml BC500 supplemented with 0.1% NP-40 and twice with 1 ml BC100 (20 mM Tris-HCl, pH 8.0, 100 mM NaCl, 1 mM DTT, 1 mM EDTA, 10% glycerol). Then, proteins were eluted using HA peptide in BC100, separated by SDS-PAGE and stained with Coomassie Brilliant Blue. Uncropped scans of gels are shown in Supplementary Fig. 1.

### Biotin-DNA pull-down

Biotin single-stranded DNA oligos were synthesized by Sangon Biotech. Double-stranded DNA oligos were annealed in the annealing buffer (10 mM Tris-HCl, pH 7.5, 1 mM EDTA, 100 mM NaCl) from single-stranded complementary oligos. Complementary single-stranded DNAs were mixed in equal molar amounts then heated to 100 °C for 10 min and cooled to room temperature gradually.

Protein MBP-Flag-ZNF512-WT, MBP-Flag-ZNF512-9A, MBP-Flag-ZNF512-swap, or MBP-Flag-ZNF512B-WT was incubated with biotin-double-stranded major satellite probe-WT or biotin-double-stranded major satellite probe-Mut (mutated), streptavidin agarose bead slurry (Millipore) in DNA pull-down buffer (20 mM HEPES, pH 8.0, 150 mM NaCl, 1 mM DTT, 0.5% NP-40, 2 mM MgCl<sub>2</sub>, 10% glycerol) for 1 h. Then, the streptavidin agarose beads were washed six times with DNA pull-down buffer. Bead-bound protein was extracted using SDS loading buffer and analysed by SDS-PAGE and western blotting using anti-MBP antibody and HRP-conjugated AffiniPure donkey anti-mouse IgG(H+L). Uncropped scans of blots are shown in Supplementary Fig. 1.

Single-stranded DNA sequences were as follows: bio-major satellite probe-WT F: 5'-biotin-ggacctggatatggcgagaaaactgaaaaatc-3'; bio-major satellite probe-WTR: 5'-biotin-gatttcagggttctcgccatattccagggtcc-3'; bio-major satellite probe-Mut F: 5'-biotin-ggacctggatatggcgagtatactgtatatac-3'; bio-major satellite probe-Mut R: 5'-biotin-gatatacagttatactcgccatataccagggtcc-3'.

### ChIP-seq and library preparation

ChIP experiments were performed as described previously<sup>52</sup> with a few modifications. TT2 mouse ES cells were fixed with a 1% formaldehyde solution (Sigma) for 10 min at room temperature. Then, the cross-linking reaction was stopped by adding glycine to a final concentration of 125 mM for 5 min at room temperature. After being washed with DPBS three times, cells were scraped from dishes thoroughly with a cell scraper and transferred into a 15 ml tube. After centrifugation, cells were resuspended in lysis buffer 1 (50 mM HEPES, pH 7.9, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% IGEPAL CA-630 and 0.25% Triton X-100) and incubated for 10 min on ice. After centrifugation, the cells were resuspended in lysis buffer 2 (10 mM Tris-HCl, pH 8.0, 200 mM NaCl, 1 mM EDTA and 0.5 mM EGTA) and incubated at room temperature for 10 min. The chromatin fraction was resuspended in lysis buffer 3 (10 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% sodium deoxycholate, 0.5% sodium N-lauroylsarcosine and 0.1% SDS) and sonicated into 200–400 base-pair fragments (Covaris M220).

*Drosophila* S2 cells were crosslinked with a 1% formaldehyde solution (Sigma) for 10 min at room temperature and quenched by adding 2 M glycine to a final concentration of 0.125 M. The *Drosophila* S2 cells were washed with ice-cold DPBS three times and then resuspended with buffer A (10 mM HEPES, pH 7.6, 10 mM EDTA, 0.5 mM EGTA, 0.25% Triton X-100) and incubated for 10 min on ice. After centrifugation, the pellet was resuspended with buffer B (10 mM HEPES, pH 7.6, 1 mM EDTA, 0.5 mM EGTA, 200 mM NaCl, 0.01% Triton X-100) and incubated for 10 min on ice. After centrifugation, the chromatin was resuspended in sonication buffer (20 mM Tris-HCl, pH 8.0, 140 mM NaCl, 1 mM EDTA and 1% SDS) and sonicated into fragments of 200–400 base pairs (Covaris M220).

For the ChIP, chromatin from TT2 mouse ES cells and *Drosophila* S2 cells was mixed at a ratio of 30:1 and incubated with H3K9me3 antibody (Abcam, ab8898) overnight at 4 °C. Then the antibody–chromatin complexes were incubated with Protein A Dynabeads (ThermoFisher Scientific) for 1 h at 4 °C. The Protein A Dynabeads were sequentially washed with the following buffers: once with FA lysis buffer (150 mM NaCl, 1% Triton X-100, 50 mM HEPES, pH 7.9, 2 mM EDTA and 0.5% sodium deoxycholate), twice with FA high-salt buffer (500 mM NaCl, 1% Triton X-100, 50 mM HEPES, pH 7.9, 2 mM EDTA and 0.5% sodium deoxycholate), once with LiCl buffer (250 mM LiCl, 0.5% IGEPAL CA-630, 0.5% sodium deoxycholate, 1 mM EDTA and 10 mM Tris-HCl, pH 8.0) and twice with TE buffer (10 mM Tris-HCl, pH 8.0, and 1 mM EDTA). The enriched chromatin was eluted with 250 µl freshly prepared elution buffer (1% SDS and 100 mM NaHCO<sub>3</sub>) for 15 min at room temperature with periodic gentle vortexing. The elution step was repeated once, and the two eluted fractions were combined.

For AVI-tagged ZNF512 ChIP, BirA-expressing TT2 mouse ES cells were fixed with a 1% formaldehyde solution (Sigma) and lysed with lysis buffer 1, lysis buffer 2 and lysis buffer 3. After being sonicated into 200–400 base-pair fragments (Covaris M220), chromatin fragments were incubated with Dynabeads M-280 Streptavidin (Life Technologies) in FA lysis buffer supplemented with Triton X-100 to a final concentration of 1%. After incubation overnight at 4 °C, the protein-bound beads were washed sequentially with the following buffers: 2% SDS once, FA high-salt buffer twice, LiCl buffer (250 mM LiCl, 1% IGEPAL CA-630, 0.5% sodium deoxycholate, 1 mM EDTA and 10 mM Tris-HCl, pH 8.0) once and TE buffer twice. Protein complexes were eluted in 300 µl of elution buffer (50 mM Tris-HCl, pH 8.0, 1% SDS, and 10 mM EDTA) at 65 °C with agitation at 800 rpm overnight. The elution step was repeated once, and the two eluted fractions were combined.

Then, input samples and the eluted chromatin were reverse-crosslinked and treated with RNase A (Sigma-Aldrich) and proteinase K (Biosharp). DNA was extracted using phenol chloroform and precipitated with isopropanol. A KAPA Hyper pre kit (KAPA Biosystems) and NEBNext multiplex oligos for Illumina (NEB) were used to prepare

# Article

the libraries for sequencing according to the manufacturer's protocol. DNA libraries were subjected to size selection using DNA Clean Beads (Vazyme) and PCR amplification (12 cycles for input DNA libraries and 13 cycles for eluted DNA libraries). Libraries were sequenced on a NovaSeq 6000 at Berry Genomics.

## RNA-sequencing analysis

First, adapters and low-quality bases with Phred quality score less than 20 were removed using Trimmomatic (v.0.36)<sup>55</sup>. Then, the paired-end reads were aligned to the annotated mouse transcripts (mm10 Genome vM15 release) using HISAT2 (v.2.1.0)<sup>56</sup>. For quantification, featureCounts (v.1.6.3)<sup>57</sup> was used to calculate read counts for both genes and repeats, taking into account multimapping reads as well. Library size factors were computed using DESeq2 (v.1.40.2)<sup>58</sup> on the basis of the gene expression data when assessing the expression levels of major satellite repeats. RNA-sequencing data of major satellite RNA in different knockout cell lines are provided in the source data for Fig. 3.

## ChIP-seq analysis

Sequencing reads were first trimmed to eliminate adapters and low-quality sequences using Trim Galore (v.0.6.7, [https://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/)) with default parameters. The trimmed reads were then aligned to the reference genomes: mm10 for ZNF512 ChIP-seq in TT2 mouse ES cells, mm10 + dm6 for H3K9me3 ChIP-seq generated in this study, hg38 for ZNF512 ChIP-seq in human K562 cells and ZNF512B ChIP-seq in human MCF-7 cells. Alignments were performed using Bowtie2 (v.2.3.4.4)<sup>59</sup>. To quantify the enrichment of ZNF512 or ZNF512B on human pericentric satellite repeats, ZNF512 or ZNF512B ChIP-seq data from human K562 and MCF-7 cells were also mapped to the human genome CHM13. PCR duplicates were removed using SAMtools (v.1.3.1)<sup>60</sup>. For visualization in IGV, the final bam files were converted to wig files using igvtools (v.2.3.98) and subsequently converted to bigwig files using the wigToBigWig script available on the UCSC Genome Browser website (<http://hgdownload.soe.ucsc.edu/admin/exe/>). Count matrices and heatmaps were generated using the ComputeMatrix and plotHeatmap tools from deepTools (v.3.2.1)<sup>61</sup>, respectively. When plotting heatmaps, regions with z score greater than 50 were considered to be outliers and removed. Peaks were identified using MACS2 (v.2.1.1)<sup>62</sup>. In the case of ZNF512 ChIP-seq in TT2 mouse ES cells, peaks detected in control cells (expressing BirA protein but lacking the AviTag, designated the BirA cell line in this study) were excluded, and peaks overlapping with the blacklist<sup>63</sup> were also removed. The binding motif of ZNF512 or ZNF512B was determined using the findMotifGenome.pl function of HOMER<sup>64</sup> with default settings, on the basis of the common peaks shared by two replicates of ZNF512 or ZNF512B ChIP-seq. Bedtools (v.2.27.1) software was used for intersection calculations. For ChIP-seq data, average signals from replicates were used. Data for ZNF512 and ZNF512B signals on human satellites III, II and I are provided in the source data for Extended Data Fig. 8.

For the differential analysis of H3K9me3 peaks, we used R package DiffBind (v.3.2.0)<sup>65</sup>. An H3K9me3 peak was considered to be SUV39H-dependent if the FDR was less than 0.01 and the fold change between wild-type and *Suv39h1*, *Suv39h2* double-knockout cell lines exceeded 2. H3K9me3 ChIP-seq data on major satellite in different KO cell lines are provided in the source data for Fig. 3.

For the analysis of LINEs, only those longer than 6 kb were considered. The repeat conservation was calculated using the following formula: 1 – divergence (%) – deletion (%) – insertion (%).

The consensus sequence of mouse major satellite repeats used in this study was GGACCTGGAATATGGCGAGAAAATGAAATCACGG AAAATGAGAAATACACACTTGTGAAATATGGCGAGAAAATCTG AAAAAGGTGAAATTAGAAATGTCCACTGTAGGACGTGGAATATGG CAAGAAAATGAAATCATGGAAAATGAGAACATCCACTTGACGAC TTAATAATGACGAAATCACTAAAAACGTGAAAATGAGAAATGCAC ACTGAA.

## Evolutionary analysis using the maximum likelihood method

Multiple zinc-finger domain sequence alignments were performed using MUSCLE. The evolutionary history was inferred using the maximum likelihood method and the JTT matrix-based model with 100 bootstraps. The evolutionary analyses were conducted in MEGA11 (ref. 66).

## Protein structure prediction

Protein structures were predicted by AlphaFold-Multimer (v.2.2.0) with the monomer-ptm model. The protein structure files with the highest scores were selected for visualization using PyMOL (v.2.5.2).

## Normalization

For ZNF512 ChIP-seq, public ZNF512B ChIP-seq, the sequencing data were normalized to 10 million mapped and properly paired fragments. For H3K9me3 ChIP-seq in TT2 mouse ES cells generated in this study, to facilitate comparison between samples, *Drosophila* chromatin was spiked in, and the ChIP signal was scaled by multiplying the read counts by a normalization factor,  $\alpha$ .

$$\alpha = \frac{\text{input}_{\text{dm6}}}{\text{ChIP}_{\text{dm6}} \times \text{input}_{\text{mm10}}} \times 1 \times 10^{-8}$$

## Statistics and reproducibility

Statistical analyses were performed in R v.3.4.0. At least two biological replicates were used in each experiment unless otherwise stated. P values less than 0.05 were considered to indicate statistical significance. Data are presented as mean  $\pm$  s.d. Unpaired two-sided t-tests were used to calculate the P values indicated in the figure legends, for assessment of the statistical significance of differences between groups. In box plots, the centre line represents the median, the box limits show the upper and lower quartiles, and the whiskers represent 1.5 $\times$  the interquartile range. Pearson correlation coefficients were calculated to assess correlation. No statistical methods were used to predetermine sample sizes in this study. Data collection and analysis were not performed blind to the conditions of the experiments.

There were two replicates for ZNF512 ChIP-seq in TT2 mouse ES cells, H3K9me3 ChIP-seq in TT2 mouse ES cells, public ZNF512 ChIP-seq in human K562 cells and public ZNF512B ChIP-seq in human MCF-7 cells. The reproducibility of replicates generated in this study is shown in Extended Data Fig. 10. In the reproducibility analysis, for ZNF512 ChIP-seq, common peaks of two replicates detected in wild-type ES cells were used, removing peaks also detected in control cells; for H3K9me3 ChIP-seq, common peaks shared by two replicates detected in wild-type ES cells were used.

## Reporting summary

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

## Data availability

The raw sequencing data reported in this paper have been deposited to the Genome Sequence Archive of the National Genomics Data Center, China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences (GSA: CRA011890) and are publicly accessible at <https://ngdc.cncb.ac.cn/gsa>. The processed data have been deposited in BIG OMIX under accession number OMIX004553 and are publicly accessible at <https://ngdc.cncb.ac.cn/omix>. Accession codes of the public data in ENCODE used in this study are as follows: ZNF512 ChIP-seq in human K562 cells, ENCSR696URR; ZNF512B ChIP-seq in human MCF-7 cells, ENCSR761LRR. The expression atlas data for *Pax3*, *Pax9*, *Foxd3*, *Zfp512* and *Zfp512b* used in Extended Data Fig. 1 were downloaded from <https://www.ebi.ac.uk/gxa/home>. We used mouse

genome version mm10 and human genome versions hg38 and CHM13 in this study. Source data for Figs. 1b, 2d, 3b,c,e,f and Extended Data Figs. 2b, 4b and 8d are provided with this paper. The gel for Extended Data Fig. 3d and uncropped film scans for Fig. 4e,f and Extended Data Figs. 3c and 8j are presented in Supplementary Fig. 1.

## Code availability

The software used to analyse these data is listed in the Methods and is all publicly available.

51. Tachibana, M. et al. G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. *Genes Dev.* **16**, 1779–1791 (2002).
52. Xiong, J. et al. Cooperative action between SALL4A and TET proteins in stepwise oxidation of 5-methylcytosine. *Mol. Cell* **64**, 913–925 (2016).
53. Xu, M., Chen, S. & Zhu, B. Investigating the cell cycle-associated dynamics of histone modifications using quantitative mass spectrometry. *Methods Enzymol.* **512**, 29–55 (2012).
54. Hung, V. et al. Spatially resolved proteomic mapping in living cells with the engineered peroxidase APEX2. *Nat. Protoc.* **11**, 456–475 (2016).
55. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
56. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
57. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
58. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
59. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
60. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
61. Ramirez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
62. Zhang, Y. et al. Model-based analysis of ChIP-seq (MACS). *Genome Biol.* **9**, R137 (2008).
63. Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE blacklist: identification of problematic regions of the genome. *Sci. Rep.* **9**, 9354 (2019).
64. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
65. Ross-Innes, C. S. et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**, 389–393 (2012).
66. Tamura, K., Stecher, G. & Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis version 11. *Mol. Biol. Evol.* **38**, 3022–3027 (2021).

**Acknowledgements** We thank G. Li and L.-L. Du for helpful discussion; F. Suo, Q. Zhang and H. Zhang for assistance with protein structure prediction using AlphaFold; J. Jia and S. Meng for assistance with FACS experiments; and J. Wang and M. Zhang from the Laboratory of Proteomics of the Institute of Biophysics for assistance with mass spectrometry. This work was supported by grants from the China Natural Science Foundation (32288102, 32130021 and 921513302), the National Key R&D Programme of China (2018YFE0203302), the Chinese Academy of Sciences (XDB39010100 and JZHKYPT-2021-05), and the New Cornerstone Science Laboratory. Z.Z. is supported by the Youth Innovation Promotion Association (2017133) of the Chinese Academy of Sciences.

**Author contributions** B.Z. designed and supervised the project. R.M. designed and performed most of the experiments. Y.Z. performed most of the bioinformatics analysis. J.Z. participated in performing experiments during revision. Y.F. performed experiments for Extended Data Fig. 3c. H.-T.W. helped to prepare Extended Data Figs. 6b, 7 and 9b. P.Z., Z.L. and Z.Z. assisted with data analysis. R.M. and B.Z. wrote the manuscript, and Y.Z. participated in writing the methods and figure legends. All the authors read and commented on the manuscript.

**Competing interests** The authors declare no competing interests.

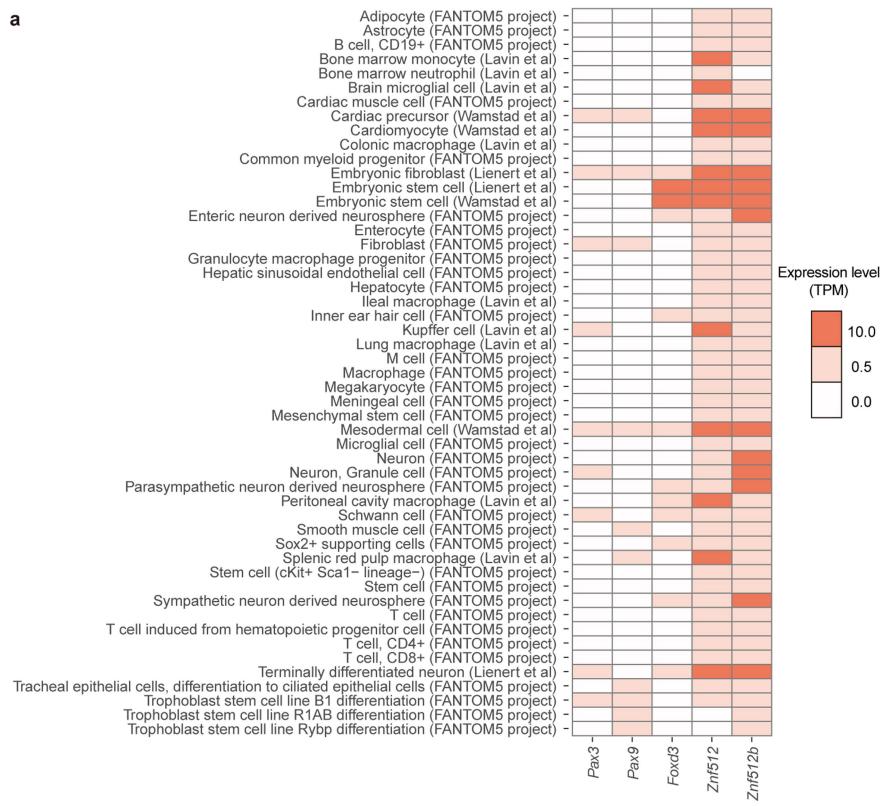
## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41586-024-07640-5>.

**Correspondence and requests for materials** should be addressed to Bing Zhu.

**Peer review information** *Nature* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.

**b****Major satellite consensus sequence in mouse pericentric regions**

**FOXD3**  
**PAX3**  
**PAX3/PAX9**  
**TAAAGTGTGATTTCTCATTTTCACCTTTTCAGTTTCCTGCCATAATTCACGTCC**

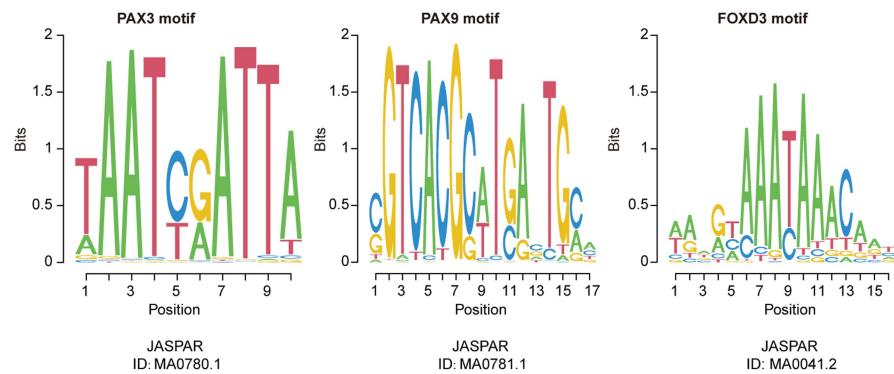
**FOX3**  
**TCAAGTGGATGTTCTCATTTTCATGATTTCAGTTTCCTGCCATAATTCACGTCC**

**PAX3**  
**TACAGTGGACATTTCTAAATTTCACCTTTTCAGTTTCCTGCCATAATTCACGTCC**

**PAX3/PAX9**  
**TAAAGTGTGATTTCTCATTTCCGTGATTTCAGTTTCCTGCCATAATTCAGGTCC**

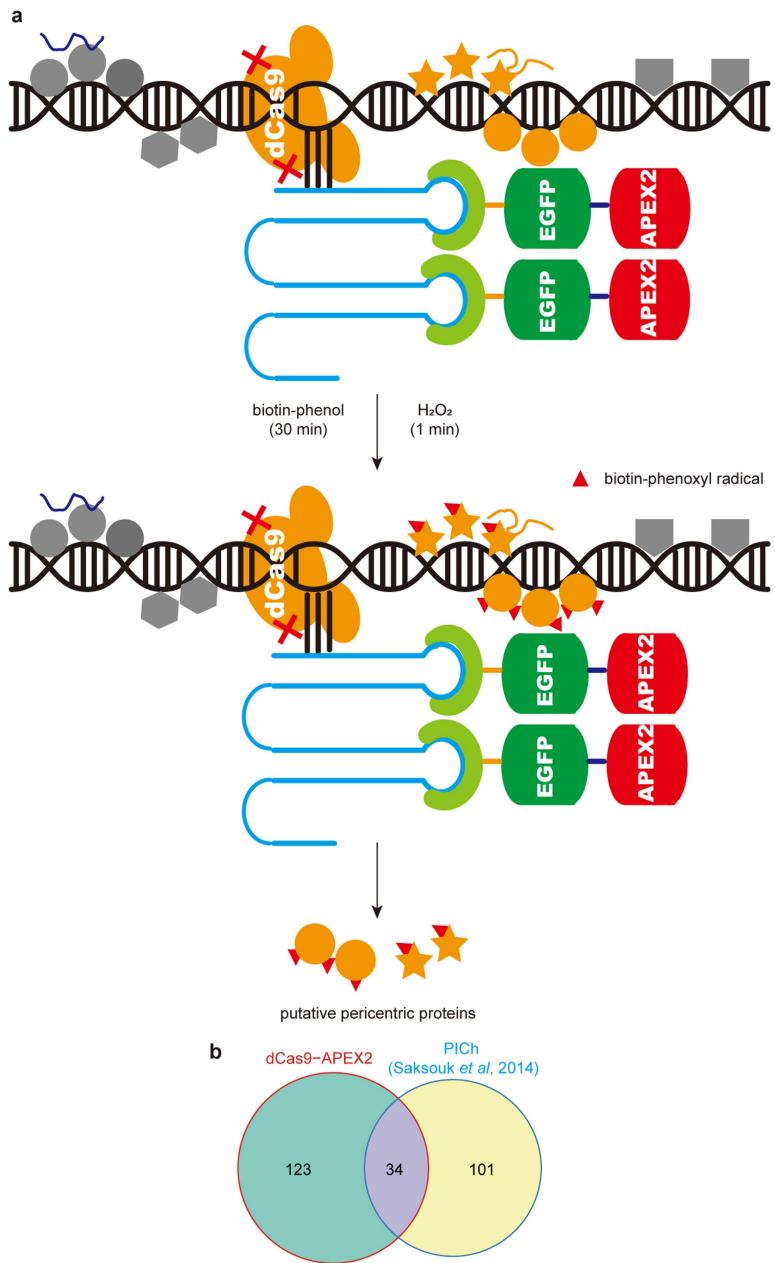
**Three types of satellite repeats in human pericentric regions**satellite III (**ATTC**)<sub>n</sub> + or - (**A****T**/**G****T****C****G****G****T****G**)satellite II (**ATTC****ATTC****G**)<sub>n</sub> + (**ATG**)<sub>1</sub> or <sub>2</sub>

satellite I (ACATAAAATATCGAAAGT) and (ACCCAAAATAGTAGTATTATATACTGT)



**Extended Data Fig. 1 | The spatiotemporal expression pattern and binding sites of pericentric sequence-specific DNA binding proteins.** (a) The expression atlas of *Pax3*, *Pax9*, *Foxd3*, *Znf512*, and *Znf512b* in S1 experiments. TPM: transcript per million. (b) upper: major satellite consensus sequence in mouse pericentric regions and three types of satellite repeats in human

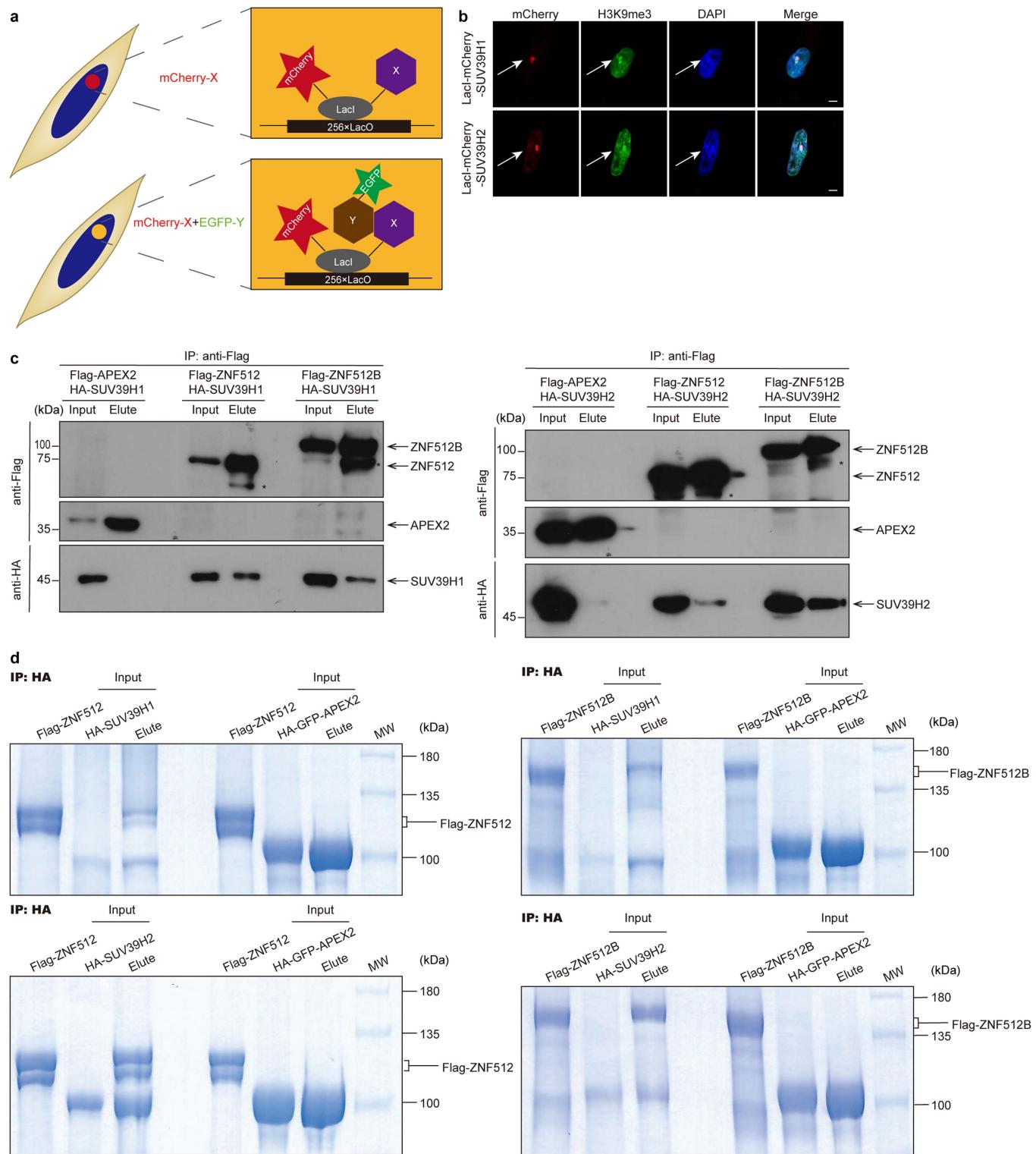
pericentric regions. The binding sites of PAX3 and PAX9 in major satellite consensus sequence were highlighted in orange, and the binding site of FOXD3 in major satellite consensus sequence was highlighted in blue. TTC triples in mouse major satellite consensus sequence and human satellite repeats were highlighted in bold. lower: binding motifs of PAX3, PAX9 and FOXD3.



**Extended Data Fig. 2 | Workflow and the overlap analysis to characterize the protein composition of pericentric heterochromatin.** (a) Schematic diagram of the approach to map specific genomic site-associated proteome based on proximity labeling. (b) Overlap between proteins identified at

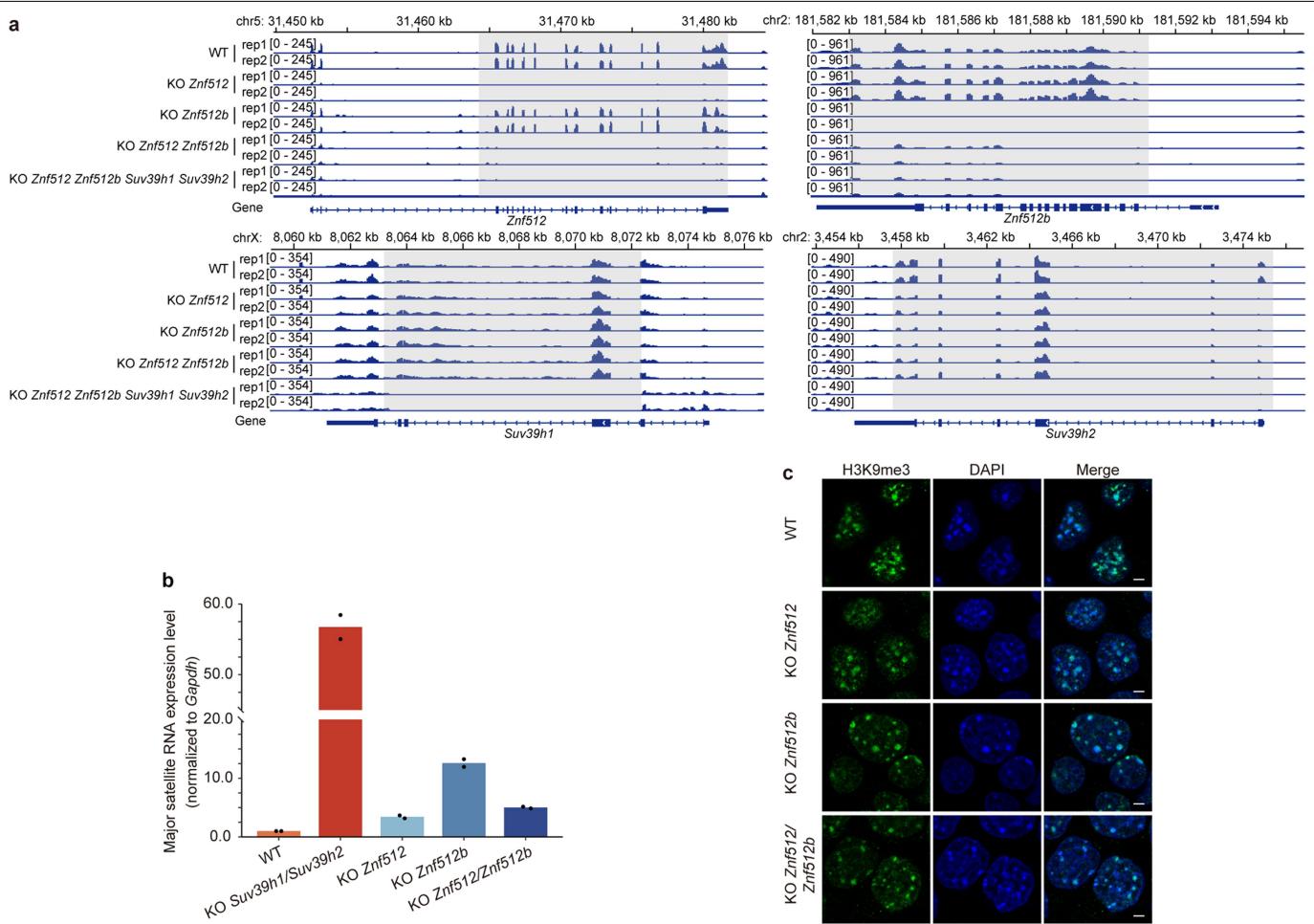
pericentric heterochromatin by dCas9-APEX2 and candidates identified by PICh. A full list of pericentric heterochromatin associated proteins was included in Source Data for Extended Data Fig. 2.

# Article



**Extended Data Fig. 3 | ZNF512 and ZNF512B can interact with SUV39H1 and SUV39H2 in vivo and in vitro.** (a) upper: A schematic diagram showing the LacO-Lacl targeting system for testing whether a factor can establish H3K9me3 modified heterochromatin de novo. lower: A schematic diagram showing the LacO-Lacl targeting system for detecting protein-protein interaction. (b) Validation of the LacO-Lacl reporter system for testing whether a factor can establish H3K9me3 modified heterochromatin de novo. A representative cross-section immunofluorescence image showing the relative localization of LacI-mCherry fused SUV39H1 or SUV39H2, H3K9me3, and DAPI in CHO A03-1 cells. Condensed chromatin was stained by DAPI. Scale bar, 5 μm. (c) Co-IP and

western blot analyzing the interaction between ZNF512 or ZNF512B and SUV39H1 or SUV39H2 in vivo. The mobilities of protein markers were indicated on the left of each panel. The asterisk refers to the degradation band of proteins. (d) MBP-Flag-ZNF512 or MBP-Flag-ZNF512B and MBP-HA-SUV39H1 or MBP-HA-SUV39H2 or MBP-HA-GFP-APEX2 were pulled down by HA beads. Bound proteins were eluted by HA peptide, separated on SDS-PAGE gels, and stained with Coomassie Brilliant Blue. The mobilities of protein markers were indicated on the right of each panel. For panels b, c, and d, representative data from two independent biological repeats are shown.

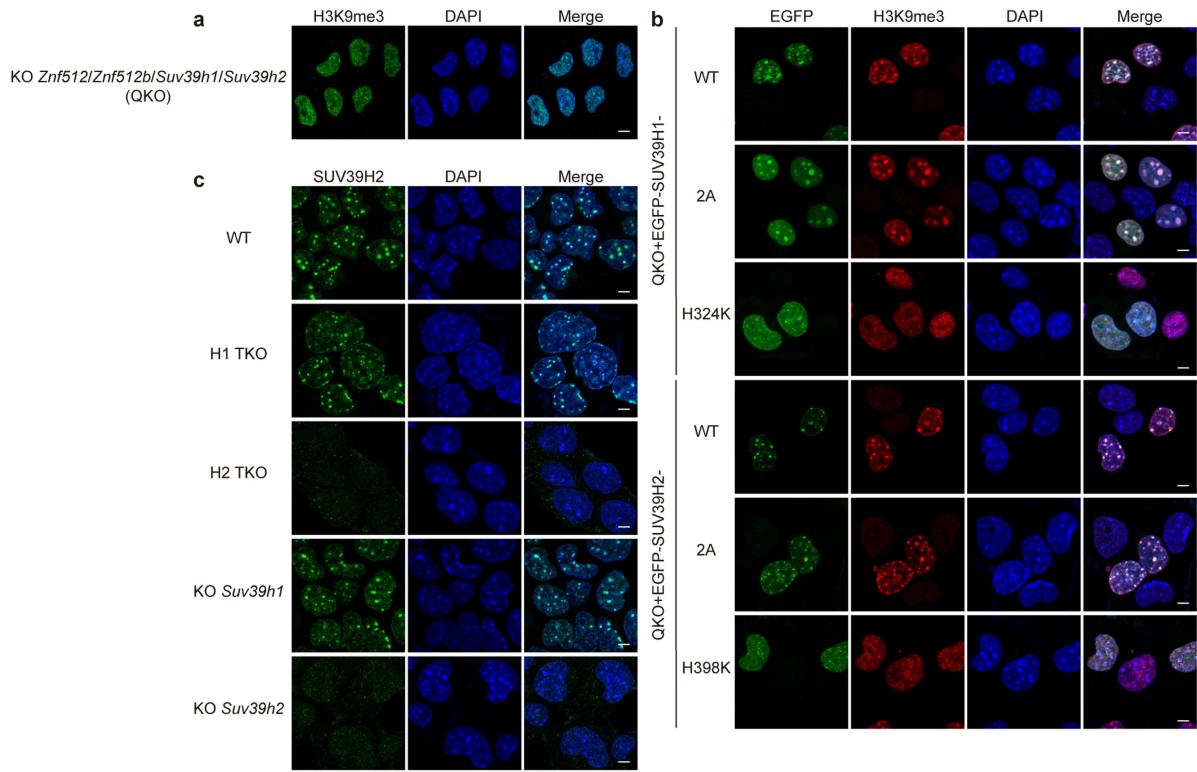


#### Extended Data Fig. 4 | ZNF512 and ZNF512B do not appear to be essential for maintaining the molecular signature of pericentric heterochromatin.

(a) Transcriptome browser tracks representing the knockout of *Znf512*, *Znf512b*, *Suv39h1* and *Suv39h2* in wild type cell, *Znf512* single KO cell, *Znf512b* single KO cell, *Znf512* and *Znf512b* double knockout cell, and QKO cell. (b) RT-qPCR showing the relative expression level of major satellite RNA in wild-type and

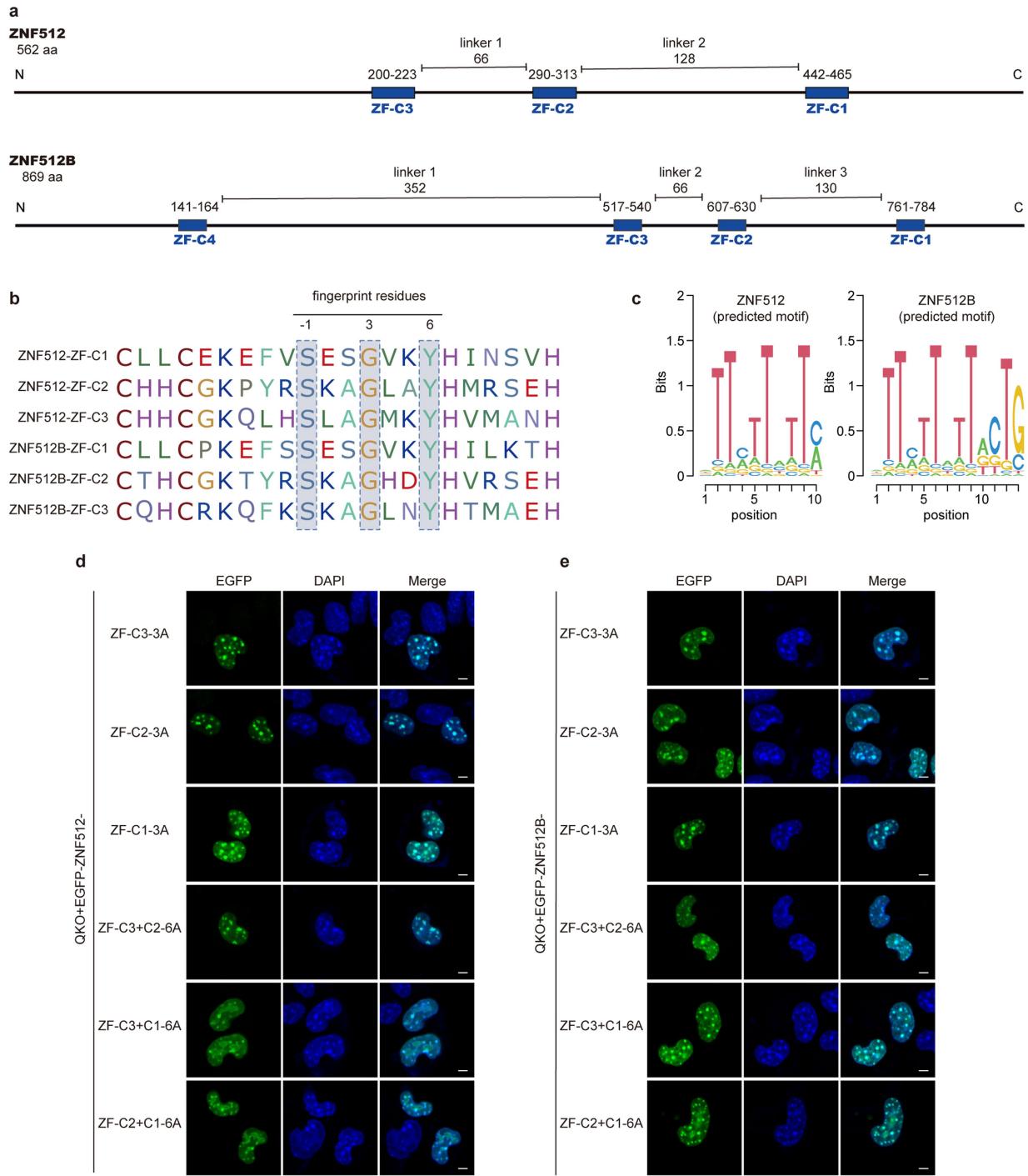
different KO cells. The expression level was normalized to *Gapdh*. Bars indicate the mean; dots indicate individual replicate. (c) A representative cross-section image showing the localization of H3K9me3 in wild type, *Znf512* single KO, *Znf512b* single KO, *Znf512* and *Znf512b* double KO cells. Scale bar, 2 μm. For panel b, data are from two independent biological replicates. For panel c, representative data from four independent biological repeats are shown.

# Article



**Extended Data Fig. 5 | The localization of H3K9me3 and mutated EGFP-SUV39H in QKO cells and the localization of endogenous SUV39H2 in H1 TKO cells.** (a) A representative cross-section image showing the localization of H3K9me3 in *Znf512*, *Znf512b*, *Suv39h1* and *Suv39h2* quadruple KO (QKO) cells. Scale bar, 5  $\mu$ m. (b) The localization of EGFP-SUV39H1, EGFP-SUV39H2 wild type or mutated versions and H3K9me3 were analyzed in *Znf512*, *Znf512b*, *Suv39h1*, *Suv39h2* quadruple KO (QKO) cell line. SUV39H1-W64A/Y67A, the aromatic

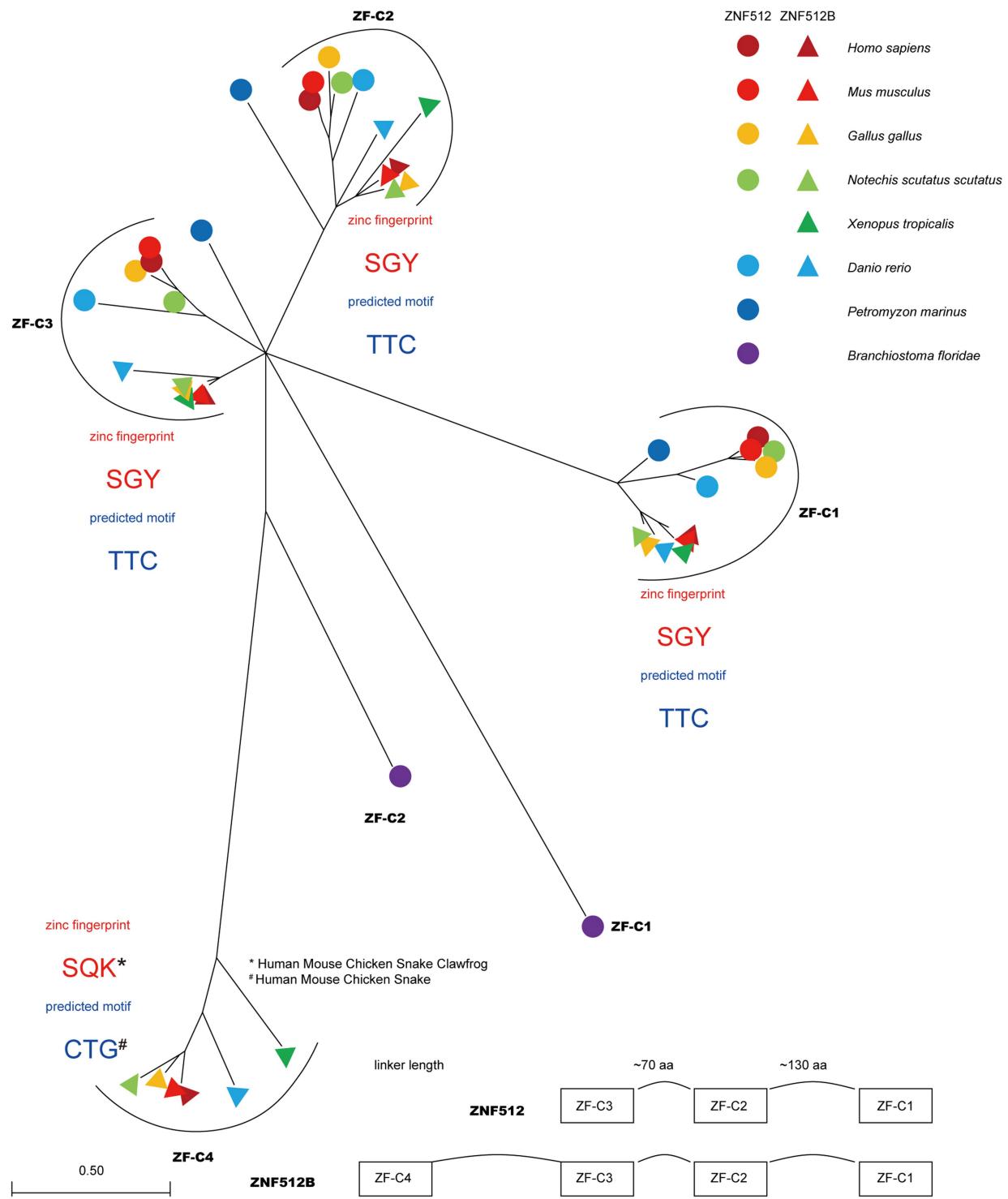
cage mutant, was abbreviated as SUV39H1-2A, and SUV39H2-W139A/W142A, was abbreviated as SUV39H2-2A. SUV39H1-H324K and SUV39H2-H398K were catalytically dead mutants. Scale bar, 5  $\mu$ m. (c) A representative cross-section image showing the localization of endogenous SUV39H2 and DAPI in wild-type and different KO cells. Scale bar, 5  $\mu$ m. For panel a, b, and c, representative data from two independent biological repeats are shown.



**Extended Data Fig. 6 | ZNF512 and ZNF512B utilize their zinc fingers to target the pericentric sequences.** (a) Schematic diagram of ZNF512 and ZNF512B protein sequences. Zinc finger domains and linkers' lengths were labeled. (b) Multi-sequence alignments of zinc finger domains in mouse ZNF512 and ZNF512B. The zinc fingerprint residues at positions -1, 3, 6 of each C2H2 zinc finger were highlighted. (c) The binding motifs of mouse ZNF512 and ZNF512B were predicted by a polynomial SVM-based algorithm. (d) Representative cross-section fluorescence images showing the relative localization of EGFP-ZNF512 with one or two zinc fingerprints mutated and DAPI in *Znf512*, *Znf512b*, *Suv39h1*, *Suv39h2* quadruple KO (QKO) cells. Zinc fingerprint residues mutated versions,

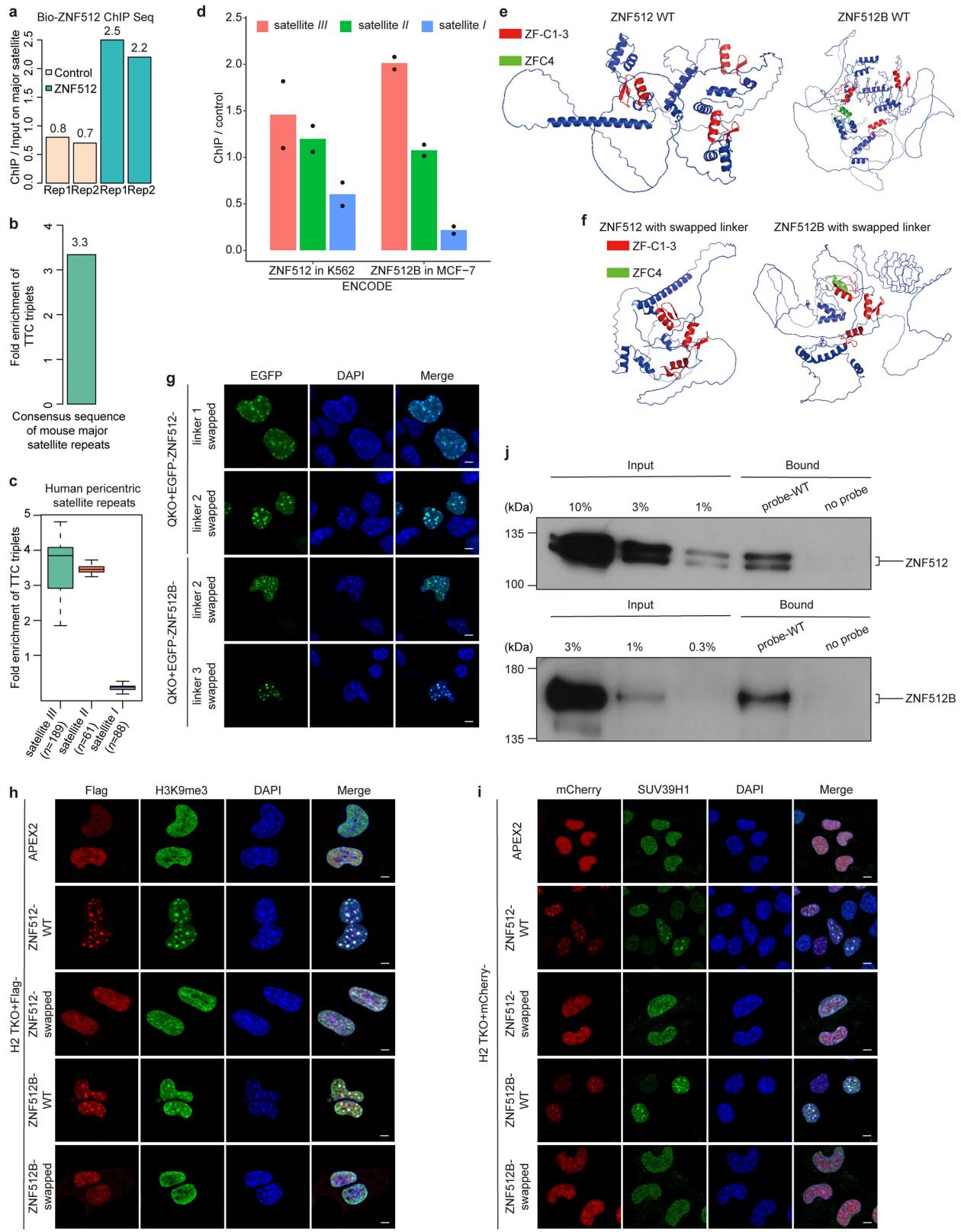
in which one zinc finger of ZNF512 mutated to AAA, were abbreviated as ZF-3A. Zinc fingerprint residues mutated versions, which two zinc fingers of ZNF512 mutated to AAA, were abbreviated as ZF-6A. Scale bar, 5  $\mu$ m. (e) Representative cross-section fluorescence images showing the relative localization of EGFP-ZNF512B with one or two zinc fingerprints mutated and DAPI in *Znf512*, *Znf512b*, *Suv39h1*, *Suv39h2* quadruple KO (QKO) cells. Zinc fingerprint residues mutated versions, which one zinc finger of ZNF512B mutated to AAA, were abbreviated as ZF-3A. Zinc fingerprint residues mutated versions, which two zinc fingers of ZNF512B mutated to AAA, were abbreviated as ZF-6A. Scale bar, 5  $\mu$ m. Data are representative of three (d) or one (e) independent biological experiments.

# Article



**Extended Data Fig. 7 | Evolution of the zinc finger domains and the length of linkers in ZNF512 and ZNF512B.** Phylogenetic tree showing the evolutionary relationship of zinc finger domains of ZNF512 and ZNF512B in humans, mice, chicken, mainland tiger snake, western clawed frog, zebrafish, sea lamprey, and

Florida lancelet. Zinc fingerprints were highlighted in red, and predicted binding motifs were highlighted in blue. The length of linkers in ZNF512 and ZNF512B was also evolutionarily conserved among vertebrates.

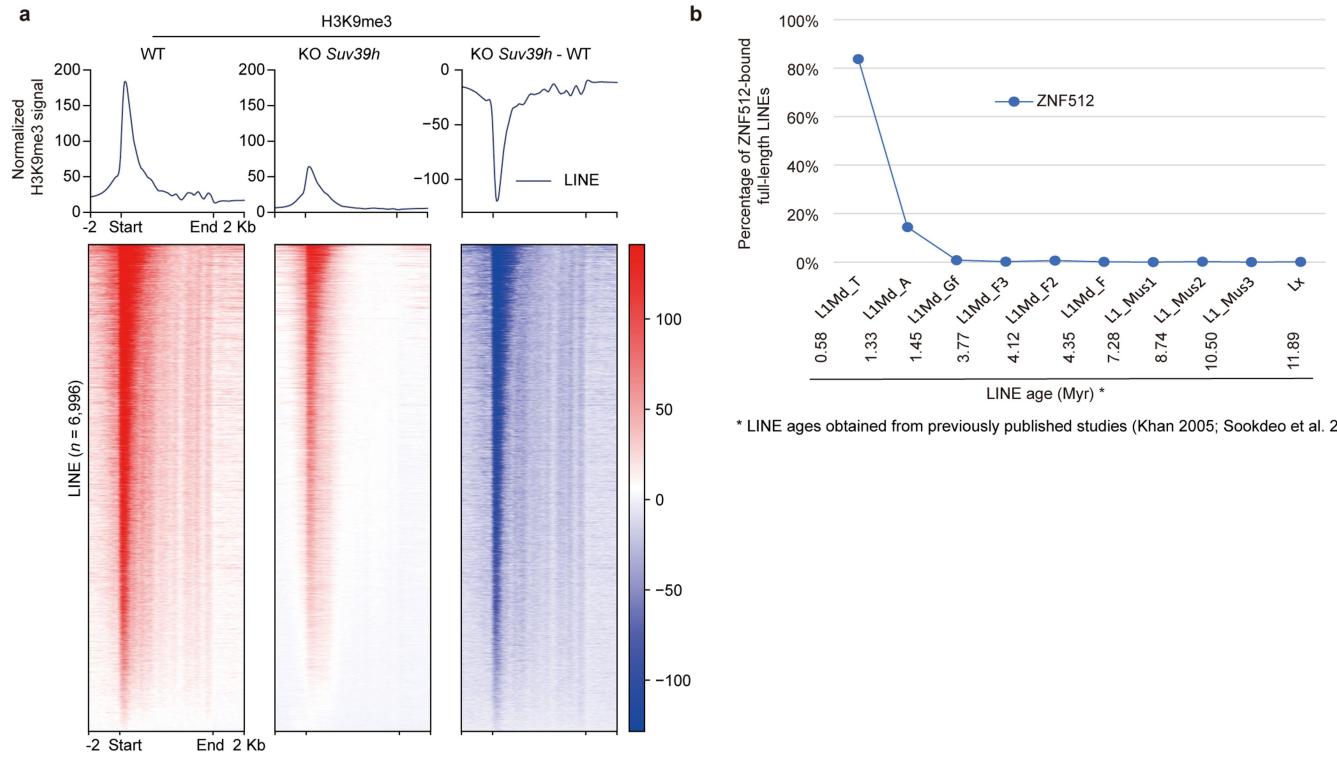


**Extended Data Fig. 8** | See next page for caption.

# Article

**Extended Data Fig. 8 | The long linkers play an important role in ZNF512 and ZNF512B for targeting the pericentric sequences.** (a) Bar plots showing the ratio of the percentage of reads mapped on mouse major satellite repeats between biotin ChIPed samples and the corresponding input samples in indicated cell lines. (b) Bar plot showing the fold enrichment of TTC triplets in the consensus sequence of mouse major satellite repeats. (c) Box plots showing the fold enrichment of TTC triplets in human pericentric satellite III/II/I. Centre line, median; box, 25<sup>th</sup> and 75<sup>th</sup> percentiles; whiskers, 1.5× the interquartile range (IQR). Numbers of satellites:  $n = 189$  (satellite III),  $n = 61$  (satellite II),  $n = 88$  (satellite I). (d) Bar plots showing the ratio of the percentage of reads mapped on human pericentric satellite repeats between ZNF512 or ZNF512B ChIPed samples and the corresponding control samples in indicated cell lines. Bars indicate the mean; dots indicate individual replicate. (e, f) Predicted structures of wild-type or linkers-swapped ZNF512 or ZNF512B proteins using AlphaFold.

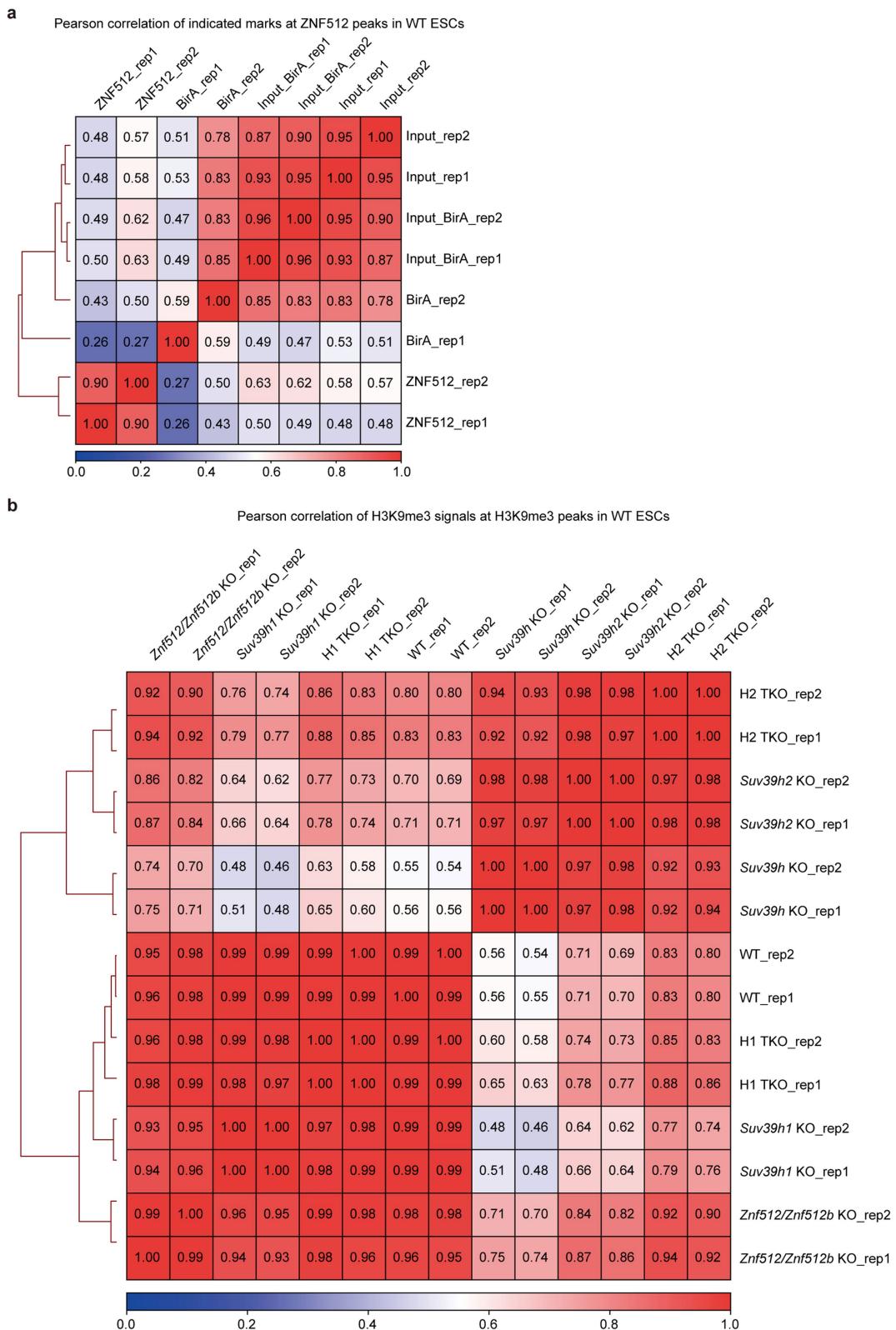
Zinc finger domains were highlighted in red color. The fourth zinc finger domain of ZNF512B was colored green. (g) Representative cross-section fluorescence images showing the relative localization of EGFP-ZNF512 or EGFP-ZNF512B linker-swapped versions in QKO cells. Scale bar, 5 μm. (h) After overexpressing Flag-tagged APEX2, ZNF512 or ZNF512B wild type or linker-swapped version in *Znf512*, *Znf512b*, *Suv39h2* triple KO (H2 TKO) cells, H3K9me3 modification at pericentric regions was visualized. Scale bar, 5 μm. (i) After overexpressing mCherry-tagged APEX2, ZNF512 or ZNF512B wild type or linker-swapped version in H2 TKO cells, endogenous SUV39H1 at pericentric regions was visualized. Scale bar, 5 μm. (j) DNA pull-down with biotin-major satellite probe or no probe followed by western blot analysis for MBP tag using purified recombinant MBP-Flag-tagged ZNF512 or ZNF512B. The mobilities of protein markers were indicated on the left of the panel. Data are representative of two (h, i and j) or one (g) independent biological experiments.



**Extended Data Fig. 9 | H3K9me3 ChIP-seq signals around LINEs.** (a) Profile (top) and heatmap (bottom) showing H3K9me3 signals in WT and *Suv39h1* and *Suv39h2* double knockout mESCs around LINEs ( $n = 6,996$ ). The difference in H3K9me3 between *Suv39h1* and *Suv39h2* double knockout and WT mESCs was also shown. The x-axis represented the distance from the start or end of LINE in

kilobases (Kb). Heatmaps were sorted by H3K9me3 signal. The average signals of two replicates were shown. (b) Percentage of ZNF512-bound full-length LINEs per subfamily in mouse ESCs arranged from the youngest to the oldest subfamily. LINE ages obtained from previously published studies. (Myr) Million years.

# Article



**Extended Data Fig. 10 | Reproducibility of replicates in this study.**

(a) Overall similarity of ZNF512 signals within ZNF512 peaks in WT mouse ES cells ( $n = 3,218$ ). Pearson correlation coefficients were indicated. (b) Overall

similarity of H3K9me3 signals within H3K9me3 peaks in WT mouse ES cells ( $n = 29,982$ ). Pearson correlation coefficients were indicated.

Corresponding author(s): Bing Zhu

Last updated by author(s): May 5, 2024

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection	ZEN 2012 black edition, nanoLC-Q EXACTIVE, nanoLC-Orbitrap Exploris 480, NovaSeq 6000, 7500 Software v2.3, BD FACS Diva software v8.0.1
Data analysis	ZEN 2012 blue edition, GraphPad Prism 9, ImageJ (version 2.0.0-rc-69/1.52p), Microsoft Excel, MEGA11, PyMOL (version 2.5.2), Alphafold-Multimer (version 2.2.0), Trimmomatic v0.36, Hisat2 v2.1.0, featureCounts v1.6.3, DESeq2 v1.40.2, R v3.4.0, Trim Galore v0.6.7, Bowtie2 v2.3.4.4, SAMtools v.1.3.1, igvtools v2.3.98, wigToBigWig, deepTools v3.2.1, MACS2 v2.1.1, HOMER, Bedtools v2.27.1, DiffBind v3.2.0

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The raw sequencing data reported in this paper have been deposited to the Genome Sequence Archive in National Genomics Data Center, China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences (GSA: CRA011890) that are publicly accessible at <https://ngdc.cncb.ac.cn/gsa>. The

processed data have been deposited in BIG OMIX under accession number OMIX004553 that are publicly accessible at <https://ngdc.cncb.ac.cn/omix>. Expression atlas of Pax3, Pax9, Foxd3, Znf512, and Znf512b used in Extended Data Fig. 1 were downloaded from <https://www.ebi.ac.uk/gxa/home>. ZNF512 ChIP-seq data in human K562 cells, and ZNF512B ChIP-seq data in human MCF-7 cells were downloaded from ENCODE project, and the accession numbers are ENCSR696URR and ENCSR761LRR, respectively. Source data for Fig. 1b, Fig. 2d, Fig. 3b, 3c, 3e, 3f, Extended Data Fig. 2b, Extended Data Fig. 4b, Extended Data Fig. 8d are provided. Gel for Extended Data Fig. 3d and uncropped film scans for Fig. 4e, 4f, Extended Data Fig. 3c, Extended Data Fig. 8j are presented in Supplementary Figure 1. We used mouse genome version mm10 and human genome version hg38 and version CHM13.

## Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender

N/A

Population characteristics

N/A

Recruitment

N/A

Ethics oversight

N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences

Behavioural & social sciences

Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

No statistical methods were used to predetermine the sample size. Sample sizes were determined based on previous experience to obtain statistical significance and reproducibility. Sample sizes were reported in the figure legends.

Data exclusions

When plotting Heatmaps, regions with z-scores >50 were considered as outliers and removed. In Fig. 6f and Extended Data Fig. 9a, the profile above the heatmap shows the mean value. When a data point is extremely high, it will abnormally elevate the mean value and cannot reflect the real distribution shown in the heatmap. In this study, we use z-score to define outliers and a threshold of 50 was used to remove as few LINEs as possible. In Fig. 5f, 4 of 6,997 LINEs were removed. In Extended Data Fig. 9a, 2 of 6,998 LINEs were removed.

Replication

Experiments were reproduced as stated in the manuscript and appropriate positive and negative control were used. The following figure panels show representative data from at least two independent experiment that showed similar results: Fig. 1a, 1c, 1d, 2a, 2b, 2c, 3a, 3c, 3d, 3e, 3f, 3g, 4a, 4b, 4d, 4e, 4f, 5f, Extended data Fig. 3b, 3c, 3d, 4a, 4b, 4c, 5a, 5b, 5c, 6d, 8a, 8i, 8j, 9a. For ChIP-seq, multiple replications show good reproducibility. The experiments in Extended data Fig. 6e, 8g were performed once, but appropriate positive and negative control were used. The experiments in Fig. 1b were performed once, but SILAC approach and reciprocally labeling was used.

Randomization

Randomization is not applicable to this study because we did not perform any experiments where there are treatment and control groups that would necessitate randomization between the subjects.

Blinding

Blinding is not applicable to this study because we did not perform any experiments where there are treatment and control groups that would necessitate blinding.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

## Methods

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Antibodies

### Antibodies used

Cy3-conjugated streptavidin, Sangon Biotech, D111115, immunofluorescence (1:150); anti-Flag, Sigma, F3165, immunofluorescence (1:200); anti-H3K9me3, Abcam, ab8898, immunofluorescence (1:1,000), ChIP (3 µg/ChIP); anti-SUV39H1, Cell Signaling Technology, 8729, immunofluorescence (1:1,000); anti-SUV39H2, HUABIO, ET7108-37, immunofluorescence (1:1,000); anti-HP1α, Millipore, MAB3584, immunofluorescence (1:1,000); anti-HP1β, Cell Signaling Technology, 8676, immunofluorescence (1:800); anti-Flag, Abmart, TT0003, western blot (1:1,000); anti-HA, Cell Signaling Technology, 3724, western blot (1:1,000); anti-MBP, Proteintech, 66003-1-Ig, western blot (1:7,500); HRP-conjugated affinipure donkey anti-mouse IgG(H+L), Proteintech, SA00001-8, 1:10,000; HRP-conjugated affinipure donkey anti-rabbit IgG(H+L), Proteintech, SA00001-9, 1:10,000; Alexa Fluor 647 donkey anti-rabbit IgG(H+L), Life technologies, A31573, 1:1,000; Alexa Fluor 488 goat anti-rabbit IgG(H+L), Invitrogen, A11008, 1:1,000; Goat anti-rabbit IgG(H+L) Alexa Fluor plus 555, Invitrogen, A32732, 1:1,000; Goat anti-mouse IgG(H+L) Alexa Fluor plus 555, Invitrogen, A32727, 1:1,000.

### Validation

The specificities of Cy3-conjugated streptavidin, Flag antibody, HP1α antibody and HP1β antibody for immunofluorescence was shown by negative control in the Figure 1a, Figure 2c and Extended Data Figure 8h, Figure 3g, and Figure 3g. Antibody against H3K9me3 was validated with Znf512/Znf512b/Suv39h1/Suv39h2 QKO mESCs by immunofluorescence in Extended Data Figure 5a. Antibody against SUV39H1 was validated with Suv39h1 KO and Znf512/Znf512b/Suv39h1 TKO mESCs by immunofluorescence in Figure 3d. Antibody against SUV39H2 was validated with Suv39h2 KO and Znf512/Znf512b/Suv39h2 TKO mESCs by immunofluorescence in Extended Data Figure 5c. Specificities of Flag antibody and HA antibody for western blot analysis were shown by negative controls in the Extended Data Figures 3c. Additionally, most commercial antibodies have been tested for specificity by their respective suppliers. Flag (Sigma, F3165): validation for all, all other information can be found at [https://www.sigmal Aldrich.cn/CN/zh/product/sigma/f3165?utm\\_campaign=LR4\\_760\\_PPR%202%20-20%20China&utm\\_medium=cpc&utm\\_source=bing&utm\\_content=sigma/f3165&utm\\_term=sigma%20f3165&msclkid=dba0f891e21b102950ad9ebe8d32c6fc](https://www.sigmal Aldrich.cn/CN/zh/product/sigma/f3165?utm_campaign=LR4_760_PPR%202%20-20%20China&utm_medium=cpc&utm_source=bing&utm_content=sigma/f3165&utm_term=sigma%20f3165&msclkid=dba0f891e21b102950ad9ebe8d32c6fc); H3K9me3 (Abcam, ab8898): validation for Mouse, Cow, Human, all other information can be found at <https://www.abcam.cn/products/primary-antibodies/histone-h3-tri-methyl-k9-antibody-chip-grade-ab8898.html>; SUV39H1 (CST, 8729): validation for Human, Mouse, Rat, Monkey, all other information can be found at <https://www.cellsignal.cn/products/primary-antibodies/suv39h1-d11b6-rabbit-mab/8729>; SUV39H2 (HUABIO, ET7108-37): validation for Human, Mouse, Rat, all other information can be found at <http://staging.www.huabio.cn/product/SUV39H2-antibody-ET7108-37>; HP1α (Millipore, MAB3584): validation for Human, Mouse, all other information can be found at [https://www.merckmillipore.com/CN/zh/product/Anti-Heterochromatin-Protein-1-Antibody-clone-2HP-1H5,MM\\_NF-MAB3584?ReferrerURL=https%3A%2F%2Fcdn.bing.com%2F&bd=1#overview](https://www.merckmillipore.com/CN/zh/product/Anti-Heterochromatin-Protein-1-Antibody-clone-2HP-1H5,MM_NF-MAB3584?ReferrerURL=https%3A%2F%2Fcdn.bing.com%2F&bd=1#overview); HP1β (CST, 8676): validation for Human, Mouse, Rat, Monkey, all other information can be found at <https://www.cellsignal.cn/products/primary-antibodies/hp1b-d2f2-xp-rabbit-mab/8676>; Flag (Abmart, TT0003): validation for all, all other information can be found at <http://www.ab-mart.com.cn/page.aspx?node=%2077%20&id=%2031790>; HA (CST, 3724): validation for all, all other information can be found at <https://www.cellsignal.cn/products/primary-antibodies/ha-tag-c29f4-rabbit-mab/3724>; MBP (Proteintech, 66003-1-Ig): validation for Human, Mouse, Rat, all other information can be found at <https://www.ptgcn.com/Products/MBP-Tag-Antibody-66003-1-Ig.htm>.

## Eukaryotic cell lines

Policy information about [cell lines](#) and [Sex and Gender in Research](#)

### Cell line source(s)

TT2 mouse ESCs was gift from Yoichi Shinkai, Kyoto University, PMID: 12130538. HEK 293-FT cell was obtained from the National Institute of Biological Sciences, Beijing, which were originally obtained from ATCC (PTA-5077). CHO A03-1 cell was gift from Guohong Li, Institute of Biophysics, which were originally obtained from Andrew S. Belmont, University of Illinois, PMID: 8991083. Drosophila S2 cell was gift from Rongwen Xi, National Institute of Biological Sciences, Beijing, PMID: 29158494.

### Authentication

Identity of the TT2, HEK 293-FT, CHO A03-1 and Drosophila S2 cell lines were frequently checked by their morphological features.

### Mycoplasma contamination

The cell lines were not tested for mycoplasma contamination.

### Commonly misidentified lines (See [ICLAC](#) register)

No commonly misidentified cell lines were used.

# ChIP-seq

## Data deposition

- Confirm that both raw and final processed data have been deposited in a public database such as [GEO](#).
- Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

### Data access links

*May remain private before publication.*

The raw data have been deposited in BIG GSA under accession number CRA011890 that are publicly accessible at <https://ngdc.cncb.ac.cn/gsa>. The processed data have been deposited in BIG OMIX under accession number OMIX004553 that are publicly accessible at <https://ngdc.cncb.ac.cn/omix>.

### Files in database submission

```
BirA-ChIPSeq-TT2-1.bw
BirA-ChIPSeq-TT2-2.bw
H3K9me3-2reps-q3.broadPeak
H3K9me3-Suv39h1KO-1.bw
H3K9me3-Suv39h1KO-2.bw
H3K9me3-Suv39h2KO-1.bw
H3K9me3-Suv39h2KO-2.bw
H3K9me3-Suv39hdn-1.bw
H3K9me3-Suv39hdn-2.bw
H3K9me3-TT2-1.bw
H3K9me3-TT2-2.bw
H3K9me3-Znf512dn-1.bw
H3K9me3-Znf512dn-2.bw
H3K9me3-Znf512Znf512bSuv39h1TKO-1.bw
H3K9me3-Znf512Znf512bSuv39h1TKO-2.bw
H3K9me3-Znf512Znf512bSuv39h2TKO-1.bw
H3K9me3-Znf512Znf512bSuv39h2TKO-2.bw
Input-BirA-ChIPSeq-TT2-1.bw
Input-BirA-ChIPSeq-TT2-2.bw
Input-ZNF512-ChIPSeq-TT2-1.bw
Input-ZNF512-ChIPSeq-TT2-2.bw
ZNF512-ChIPSeq-TT2-1.bw
ZNF512-ChIPSeq-TT2-2.bw
ZNF512_2reps-q2-BirA-blacklist.narrowPeak
BirA-ChIPSeq-TT2-1.R1.fastq.gz
BirA-ChIPSeq-TT2-1.R2.fastq.gz
BirA-ChIPSeq-TT2-2.R1.fastq.gz
BirA-ChIPSeq-TT2-2.R2.fastq.gz
H3K9me3-ChIPSeq-Suv39h1h2dn-1.R1.fastq.gz
H3K9me3-ChIPSeq-Suv39h1h2dn-1.R2.fastq.gz
H3K9me3-ChIPSeq-Suv39h1h2dn-2.R1.fastq.gz
H3K9me3-ChIPSeq-Suv39h1h2dn-2.R2.fastq.gz
H3K9me3-ChIPSeq-Suv39h1KO-1.R1.fastq.gz
H3K9me3-ChIPSeq-Suv39h1KO-1.R2.fastq.gz
H3K9me3-ChIPSeq-Suv39h1KO-2.R1.fastq.gz
H3K9me3-ChIPSeq-Suv39h1KO-2.R2.fastq.gz
H3K9me3-ChIPSeq-Suv39h2KO-1.R1.fastq.gz
H3K9me3-ChIPSeq-Suv39h2KO-1.R2.fastq.gz
H3K9me3-ChIPSeq-Suv39h2KO-2.R1.fastq.gz
H3K9me3-ChIPSeq-Suv39h2KO-2.R2.fastq.gz
H3K9me3-ChIPSeq-TT2-1.R1.fastq.gz
H3K9me3-ChIPSeq-TT2-1.R2.fastq.gz
H3K9me3-ChIPSeq-TT2-2.R1.fastq.gz
H3K9me3-ChIPSeq-TT2-2.R2.fastq.gz
H3K9me3-ChIPSeq-Znf512dn-1.R1.fastq.gz
H3K9me3-ChIPSeq-Znf512dn-1.R2.fastq.gz
H3K9me3-ChIPSeq-Znf512dn-2.R1.fastq.gz
H3K9me3-ChIPSeq-Znf512dn-2.R2.fastq.gz
H3K9me3-ChIPSeq-Znf512Znf512bSuv39h1TKO-1.R1.fastq.gz
H3K9me3-ChIPSeq-Znf512Znf512bSuv39h1TKO-1.R2.fastq.gz
H3K9me3-ChIPSeq-Znf512Znf512bSuv39h1TKO-2.R1.fastq.gz
H3K9me3-ChIPSeq-Znf512Znf512bSuv39h2TKO-1.R1.fastq.gz
H3K9me3-ChIPSeq-Znf512Znf512bSuv39h2TKO-1.R2.fastq.gz
H3K9me3-ChIPSeq-Znf512Znf512bSuv39h2TKO-2.R1.fastq.gz
H3K9me3-ChIPSeq-Znf512Znf512bSuv39h2TKO-2.R2.fastq.gz
Input-BirA-ChIPSeq-TT2-1.R1.fastq.gz
Input-BirA-ChIPSeq-TT2-1.R2.fastq.gz
Input-BirA-ChIPSeq-TT2-2.R1.fastq.gz
Input-BirA-ChIPSeq-TT2-2.R2.fastq.gz
Input-H3K9me3-ChIPSeq-Suv39h1h2dn-1.R1.fastq.gz
Input-H3K9me3-ChIPSeq-Suv39h1h2dn-1.R2.fastq.gz
```

Input-H3K9me3-ChIPSeq-Suv39h1h2dn-2.R1.fastq.gz  
 Input-H3K9me3-ChIPSeq-Suv39h1h2dn-2.R2.fastq.gz  
 Input-H3K9me3-ChIPSeq-Suv39h1KO-1.R1.fastq.gz  
 Input-H3K9me3-ChIPSeq-Suv39h1KO-1.R2.fastq.gz  
 Input-H3K9me3-ChIPSeq-Suv39h1KO-2.R1.fastq.gz  
 Input-H3K9me3-ChIPSeq-Suv39h1KO-2.R2.fastq.gz  
 Input-H3K9me3-ChIPSeq-Suv39h2KO-1.R1.fastq.gz  
 Input-H3K9me3-ChIPSeq-Suv39h2KO-1.R2.fastq.gz  
 Input-H3K9me3-ChIPSeq-Suv39h2KO-2.R1.fastq.gz  
 Input-H3K9me3-ChIPSeq-Suv39h2KO-2.R2.fastq.gz  
 Input-H3K9me3-ChIPSeq-TT2-1.R1.fastq.gz  
 Input-H3K9me3-ChIPSeq-TT2-1.R2.fastq.gz  
 Input-H3K9me3-ChIPSeq-TT2-2.R1.fastq.gz  
 Input-H3K9me3-ChIPSeq-TT2-2.R2.fastq.gz  
 Input-H3K9me3-ChIPSeq-Znf512dn-1.R1.fastq.gz  
 Input-H3K9me3-ChIPSeq-Znf512dn-1.R2.fastq.gz  
 Input-H3K9me3-ChIPSeq-Znf512dn-2.R1.fastq.gz  
 Input-H3K9me3-ChIPSeq-Znf512dn-2.R2.fastq.gz  
 Input-H3K9me3-ChIPSeq-Znf512Znf512bSuv39h1TKO-1.R1.fastq.gz  
 Input-H3K9me3-ChIPSeq-Znf512Znf512bSuv39h1TKO-1.R2.fastq.gz  
 Input-H3K9me3-ChIPSeq-Znf512Znf512bSuv39h1TKO-2.R1.fastq.gz  
 Input-H3K9me3-ChIPSeq-Znf512Znf512bSuv39h1TKO-2.R2.fastq.gz  
 Input-H3K9me3-ChIPSeq-Znf512Znf512bSuv39h2TKO-1.R1.fastq.gz  
 Input-H3K9me3-ChIPSeq-Znf512Znf512bSuv39h2TKO-1.R2.fastq.gz  
 Input-H3K9me3-ChIPSeq-Znf512Znf512bSuv39h2TKO-2.R1.fastq.gz  
 Input-H3K9me3-ChIPSeq-Znf512Znf512bSuv39h2TKO-2.R2.fastq.gz  
 Input-Znf512-ChIPSeq-TT2-1.R1.fastq.gz  
 Input-Znf512-ChIPSeq-TT2-1.R2.fastq.gz  
 Input-Znf512-ChIPSeq-TT2-2.R1.fastq.gz  
 Input-Znf512-ChIPSeq-TT2-2.R2.fastq.gz  
 Znf512-ChIPSeq-TT2-1.R1.fastq.gz  
 Znf512-ChIPSeq-TT2-1.R2.fastq.gz  
 Znf512-ChIPSeq-TT2-2.R1.fastq.gz  
 Znf512-ChIPSeq-TT2-2.R2.fastq.gz

Genome browser session  
(e.g. [UCSC](#))

no longer applicable

## Methodology

Replicates

All ChIP-seq experiments have two technical replicates.

Sequencing depth

All the ChIP-seq reads are 150bp paired-ended.  
 sample name total reads mapped reads  
 BirA-ChIPSeq-TT2-1 26,085,757 9,834,675  
 BirA-ChIPSeq-TT2-2 42,598,116 37,991,668  
 H3K9me3-ChIPSeq-Suv39h1h2dn-1 35,190,652 33,736,341  
 H3K9me3-ChIPSeq-Suv39h1h2dn-2 26,647,146 25,575,257  
 H3K9me3-ChIPSeq-Suv39h1KO-1 41,688,380 40,453,466  
 H3K9me3-ChIPSeq-Suv39h1KO-2 27,790,942 26,984,990  
 H3K9me3-ChIPSeq-Suv39h2KO-1 30,405,812 29,346,997  
 H3K9me3-ChIPSeq-Suv39h2KO-2 21,196,859 20,434,048  
 H3K9me3-ChIPSeq-TT2-1 27,335,232 26,464,412  
 H3K9me3-ChIPSeq-TT2-2 30,448,130 29,543,171  
 H3K9me3-ChIPSeq-Znf512dn-1 28,107,853 27,148,767  
 H3K9me3-ChIPSeq-Znf512dn-2 28,672,671 27,790,369  
 H3K9me3-ChIPSeq-Znf512Znf512bSuv39h1TKO-1 33,357,133 32,378,963  
 H3K9me3-ChIPSeq-Znf512Znf512bSuv39h1TKO-2 34,273,751 33,148,398  
 H3K9me3-ChIPSeq-Znf512Znf512bSuv39h2TKO-1 33,208,523 32,163,507  
 H3K9me3-ChIPSeq-Znf512Znf512bSuv39h2TKO-2 30,020,265 29,083,478  
 Input-BirA-ChIPSeq-TT2-1 28,058,044 27,191,219  
 Input-BirA-ChIPSeq-TT2-2 19,046,207 18,422,904  
 Input-H3K9me3-ChIPSeq-Suv39h1h2dn-1 24,771,270 23,878,619  
 Input-H3K9me3-ChIPSeq-Suv39h1h2dn-2 19,199,735 18,578,499  
 Input-H3K9me3-ChIPSeq-Suv39h1KO-1 27,972,599 26,155,783  
 Input-H3K9me3-ChIPSeq-Suv39h1KO-2 26,904,388 25,283,032  
 Input-H3K9me3-ChIPSeq-Suv39h2KO-1 30,522,263 29,321,830  
 Input-H3K9me3-ChIPSeq-Suv39h2KO-2 23,411,933 22,592,813  
 Input-H3K9me3-ChIPSeq-TT2-1 22,274,964 21,400,403  
 Input-H3K9me3-ChIPSeq-TT2-2 23,671,634 22,811,130  
 Input-H3K9me3-ChIPSeq-Znf512dh-1 25,434,304 24,598,686  
 Input-H3K9me3-ChIPSeq-Znf512dn-2 22,570,517 21,762,974  
 Input-H3K9me3-ChIPSeq-Znf512Znf512bSuv39h1TKO-1 25,457,137 24,517,844  
 Input-H3K9me3-ChIPSeq-Znf512Znf512bSuv39h1TKO-2 25,309,121 24,395,787  
 Input-H3K9me3-ChIPSeq-Znf512Znf512bSuv39h2TKO-1 29,957,091 27,442,037  
 Input-H3K9me3-ChIPSeq-Znf512Znf512bSuv39h2TKO-2 23,587,151 21,790,280

Input-Znf512-ChIPSeq-TT2-1 30,077,027 28,946,198  
 Input-Znf512-ChIPSeq-TT2-2 22,979,177 21,689,151  
 Znf512-ChIPSeq-TT2-1 27,716,873 14,839,089  
 Znf512-ChIPSeq-TT2-2 39,972,572 32,756,078

#### Antibodies

rabbit anti-H3K9me3, Abcam, ab8898.

#### Peak calling parameters

"macs2 callpeak -t sample.bam -c input.bam -f BAMPE -g mm -q 0.01" for ZNF512 ChIP-seq in ESCs.  
 "macs2 callpeak -t sample.bam -c input.bam -f BAMPE -g hs -q 0.01" for ZNF512 and ZNF512B ChIP-seq in human cell lines.  
 "macs2 callpeak -t sample.bam -c input.bam -f BAMPE -g mm -q 1e-3 --broad --broad-cutoff 0.01" for H3K9me3 ChIP-seq in mESCs.

#### Data quality

Sequencing reads were first trimmed to eliminate adapters and low-quality sequences using Trim Galore (v0.6.7) with default parameters. The trimmed reads were then aligned to the reference genomes: mm10 for ZNF512 ChIP-seq in TT2 mouse ESCs, mm10 +dm6 for H3K9me3 ChIP-seq generated in this study, hg38 for ZNF512 ChIP-seq in human K562 cells and ZNF512B ChIP-seq in human MCF-7 cells. Alignments were performed using Bowtie2 (v2.3.4.4). To quantify the enrichment of ZNF512 or ZNF512B on human pericentric satellite repeats, ZNF512 or ZNF512B ChIP-seq data from human K562 and MCF-7 cells were also mapped to the human genome CHM13. PCR duplicates were removed using SAMtools (v1.3.1). For visualization in IGV, the final bam files were converted to wig files using igvtools (v2.3.98) and subsequently converted to bigwig files using the wigToBigWig script available on the UCSC Genome Browser website. Peaks were identified using MACS2 (v2.1.1). In the case of ZNF512 ChIP-seq in TT2 mouse ESCs, peaks detected in control cells (expressing BirA protein but lacking the AviTag, we designated this cell line BirA in this study) were excluded, and peaks overlapping with the blacklist were also removed.

#### Software

R v3.4.0  
 Trim Galore v0.6.7  
 Bowtie2 v2.3.4.4  
 SAMtools v.1.3.1  
 igvtools v2.3.98  
 wigToBigWig  
 deepTools v3.2.1  
 MACS2 v2.1.1  
 HOMER  
 Bedtools v2.27.1  
 DiffBind v3.2.0

## Flow Cytometry

### Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

### Methodology

#### Sample preparation

Cells were trypsinized, resuspended in culture media and sorted on BD FACS Aria III. No staining involved.

#### Instrument

BD FACS Aria III

#### Software

BD FACS Diva software v8.0.1 for collection.

#### Cell population abundance

For APEX2 over-expression, TT2 cells were transfected with EGFP-encoding vectors and resulting cell population of 4.59% EGFP lower-level cells were FACS purified to 62.8% EGFP lower-level cells.

#### Gating strategy

Cells were gated for live/dead and double exclusion using FSC and SSC channels. For APEX2 over-expression, cells were gated for presence of EGFP signal.

- Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.