

CROSS: A cross-species chromosome substitution platform for dissecting chromosome architecture and activity

Wei Li

liwei@ioz.ac.cn

State Key Laboratory of Organ Regeneration and Reconstruction, Institute of Zoology, Chinese Academy of Sciences; University of Chinese Academy of Sciences <https://orcid.org/0000-0001-7864-404X>

Yihuan Mao

State Key Laboratory of Organ Regeneration and Reconstruction, Institute of Zoology, Chinese Academy of Sciences

Ning Yang

State Key Laboratory of Organ Regeneration and Reconstruction, Institute of Zoology, Chinese Academy of Sciences; Department of Research and Development, Beijing Arion Cancer Center

Yulong Zhao

State Key Laboratory of Organ Regeneration and Reconstruction, Institute of Zoology, Chinese Academy of Sciences

Libin Wang

School of Basic Medical Sciences, Peking University Stem Cell Research Center, Peking University Health Science Center

Kai Xu

Institute of Zoology, Chinese Academy of Sciences <https://orcid.org/0000-0002-5648-605X>

Yangcan Chen

Chinese Academy of Sciences

Yali Ding

State Key Laboratory of Organ Regeneration and Reconstruction, Institute of Zoology, Chinese Academy of Sciences

Bei Zhong

State Key Laboratory of Organ Regeneration and Reconstruction, Institute of Zoology, Chinese Academy of Sciences

Sinan Ma

State Key Laboratory of Organ Regeneration and Reconstruction, Institute of Zoology, Chinese Academy of Sciences

Jing Liu

Women' s Hospital, Evolutionary & Organismal Biology Research Center, School of Medicine, Zhejiang University <https://orcid.org/0000-0003-0420-1769>

Yu Zhang

MOE Laboratory of Biosystems Homeostasis & Protection and Zhejiang Provincial Key Laboratory for Cancer Molecular Cell Biology, Life Sciences Institute, Zhejiang University

Qingyang Ni

MOE Laboratory of Biosystems Homeostasis & Protection and Zhejiang Provincial Key Laboratory for Cancer Molecular Cell Biology, Life Sciences Institute, Zhejiang University

Guihai Feng

State Key Laboratory of Organ Regeneration and Reconstruction, Institute of Zoology, Chinese Academy of Sciences <https://orcid.org/0000-0002-8726-1734>

Qi Zhou

Zhejiang University <https://orcid.org/0000-0002-7419-2047>

Qi Zhou

Institute of Zoology, Chinese Academy of Science <https://orcid.org/0000-0002-6549-9362>

Article**Keywords:**

Posted Date: April 16th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-8995396/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

Abstract

The functional integrity of a mammalian chromosome is shaped by its long-term, co-evolution with species-specific nuclear environment. How chromosomes co-adapt with their native environment to define 3D architecture and transcriptional activity remain poorly understood, largely due to a lack of experimental models capable of systematically dissecting this co-evolution relationship. Here, we report a cross-species chromosome substitution (CROSS) method, a robust genomic engineering method that enables the stable, scarless replacement of host chromosomes with evolutionarily divergent orthologs. By integrating microcell-mediated chromosome transfer with CRISPR/Cas9, we imported the intact 158-Mb rat X chromosome into mouse embryonic stem cells, and subsequently achieved targeted substitution of its endogenous mouse counterpart, maintaining stability and integrity. Using this model, we found that rat-specific LINE1 and RatSatRep2 repeats failed to adequately recruit host SETDB1 in mouse cells, leading to localized erosion of H3K9me3 heterochromatin. This further triggered 3D structural remodeling, characterized by the de novo formation of topologically associating domain (TAD) boundaries that aberrantly activated adjacent genes—including *Rhox5*, the master regulator of the *Rhox* cluster—impairing cellular differentiation. Our method provides a powerful chromosome engineering platform for dissecting how genomic sequences and epigenetic mechanisms cooperate in regulating chromosome architecture and function, and for evaluating the structural and functional fidelity of large-scale synthetic or heterologous DNA across species.

Main

The 3D organization and function of mammalian genome are shaped by its long-term co-evolution with the nuclear environment. How this co-evolution dictates chromosome folding and transcriptional regulation, and ultimately governs cellular function is a central question. In simpler eukaryotes, large-scale genomic substitution has recently emerged as a transformative strategy for deciphering these sequence-dependent principles; notably, the integration of billion-year-diverged bacterial genomes into yeast demonstrated that foreign DNA can spontaneously adopt distinct chromatin archetypes even in the absence of host-sequence co-evolution¹. However, while these "top-down" evolutionary perturbations have redefined our understanding of nuclear compartmentalization in yeast, achieving the stable, scarless replacement of an entire endogenous chromosome remains a formidable challenge in mammalian systems. Current "bottom-up" synthetic approaches, such as the SynNICE system², provide powerful means to study de novo assembled DNA in vivo; yet, recapitulating the full structural complexity of a native mammalian chromosome—characterized by its dense landscape of repetitive elements^{3–7} and multi-layered 3D architecture^{8–10}—requires engineering methods capable of handling hundred-megabase-scale templates.

To bridge this gap, we developed a robust cross-species chromosome substitution (CROSS) platform. This method employs an "add-then-delete" strategy, integrating microcell-mediated chromosome transfer (MMCT)^{11–15} with CRISPR/Cas9^{16,17} to achieve the high-fidelity replacement of an endogenous chromosome with a heterologous ortholog. We demonstrate the utility of this platform by substituting

the mouse X chromosome with its 158-Mb rat ortholog in mouse embryonic stem cells (mESCs). The X chromosome serves as a high-stringency benchmark due to its requirement for chromosome-wide regulatory programs^{18–20} and its high degree of synteny (gene-order) and gene-content conservation^{4,18} (Fig. 1a, b and Extended Data Fig. 1a) contrasted by rapidly evolving non-coding sequences (Fig. 1c).

Our CROSS platform provided a "living stress test" to evaluate the structural and functional fidelity of large-scale exogenous DNA within a non-native nuclear environment. We showed that the substitution was stable and that the engineered mESCs maintained self-renewal and pluripotency. Beyond replacement, this system functioned as a discovery tool for identifying functional genomic elements sensitive to host regulatory machinery. By using the rat X chromosome as an evolutionary perturbation, we identified specific failures in host-dependent epigenetic recruitment. Species-specific LINE1 and RatSatRep2 repeats failed to adequately recruit the host methyltransferase SETDB1, leading to localized H3K9me3 erosion. This epigenetic incompatibility subsequently triggered local 3D structural remodeling, such as the de novo formation of topologically associating domain (TAD) boundaries, which were linked to the regulation of developmental loci including *Rhox5*.

By providing a streamlined workflow for whole-chromosome substitution, this study offers a powerful methodology for the functional annotation of non-coding sequences and establishes a critical framework for assessing the performance of large-scale heterologous or synthetic DNA in mammalian cells.

Results

A robust CROSS platform for scarless, intact, and cross-species chromosome substitution in mammalian cells

To achieve X chromosome substitution, a two-step process is required: First, transfer of the rat Chr.X from donor cells (rat ESC-A9 hybrid cells) to recipient cells (male mouse ESCs); Second, loss of endogenous mouse Chr.X (Fig. 1d). To enable the isolation of chromosome-transferred and subsequent chromosome-substituted clones, distinct selectable markers were introduced into male rat ESCs (hereafter referred to as rESCs) and male mouse ESCs (hereafter referred to as mESCs) by their X chromosomes through CRISPR/Cas9 system^{16,17,21,22} (Fig. 1d). The Chr.X of rESCs was labeled by puromycin resistant gene (Puro^R) and enhanced green fluorescent protein (EGFP) at the downstream of the X-linked housekeeping gene *Pgk1* (Fig. 1d, e and Extended Data Fig. 1b). To generate donor cells, the above rESCs were fused with A9 cells, yielding rESC-A9 hybrids. For recipient cells, while X-linked Hypoxanthine Phosphoribosyltransferase (*Hprt*) gene was simultaneously disrupted via targeted insertion of pEF1α-Thymidine kinase (TK)-DsRed-NeoR cassette into its coding sequence (Fig. 1d, e and Extended Data Fig. 1b). By using MMCT^{11,13,14}, rat Chr.X can be randomly transfer from donor cells to recipient cells. mESC clones with rat Chr.X (X^{tra} mESCs) were first isolated based on EGFP fluorescence and subsequently selected using HAT and puromycin (Fig. 1d). Previous research has demonstrated that

exogenous chromosomes exhibit instability within host cells and are susceptible to loss²³. Under puromycin treatment, clones were capable of maintaining rat Chr.X (Extended Data Fig. 1c). However, upon drug withdrawal, rat Chr.X underwent rapidly lost. After ten passages, proportion of GFP⁺ cells dropped below 20% (Extended Data Fig. 1c). Clones were further validated by polymerase chain reaction (PCR) amplification (Extended Data Fig. 1d, e), combined with karyotypic analysis (Extended Data Fig. 1f).

To derive mESC lines with X chromosome substitution, we next subjected the X^{tra} mESCs to ganciclovir (GCV) selection (Fig. 1d). This strategy exploited the spontaneous loss of one X chromosome in XX ESCs, thereby enriching for clones that had lost the endogenous mouse Chr.X (which carries the TK negative selection marker) (Fig. 1d, e). The resulting clones, designated X^{sub} mESCs, were isolated (Fig. 1f). PCR and karyotyping confirmed that these lines retained the rat Chr.X but lacked the mouse Chr.X (Fig. 1g, h). The transferred rat Chr.X remained stable during extended culture of Xsub mESCs (Fig. 1h and Extended Data Fig. 1c). Clones with possibly intact Chr.X were selected for whole-genome sequencing. Results showed that the mouse Chr.X and rat autosomes were absent, whereas the rat Chr.X was present and intact, indicating successful replacement of the Chr.X (Fig. 1i and Extended Data Fig. 1g). X^{sub} mESCs maintained morphology similar to ESCs (Fig. 1f), expressed ESC marker genes (Fig. 1j, k), and formed all three germ layers after teratoma differentiation (Fig. 1l). Collectively, these findings demonstrated the successful construction of a male mouse ESC line harboring only rat Chr.X.

High-fidelity conservation of 3D chromatin architecture on the substituted rat X chromosome

We next explored the 3D organization of the rat Chr.X in X^{sub} mESCs through high-throughput chromatin conformation capture (Hi-C). Hi-C contact maps indicated that the distal interactions of the rat Chr.X in X^{sub} mESCs were slightly downregulated when compared to rESCs, whereas the proximal interactions were slightly enhanced (Fig. 2a). Chromatin DNA folds at multiple scales, including compartment, TAD, chromatin loops, to build chromosomes⁸⁻¹⁰. Chromatin segregates into two compartments: compartment A (open, gene-active, and broadly euchromatic) and compartment B (closed, gene-inactive, and broadly heterochromatic)¹⁰. We first assessed Chr.X compartmentalization. The level of compartmentalization in both compartment A and compartment B decreased (Fig. 2b, c). Next, compartment changes were analyzed by plotting the E1 values, which revealed a bidirectional shift involving both compaction and decompaction (Fig. 2d-f). This analysis identified a pattern of global weakening, where the most definitive A and B compartments showed significant weaken strength (Fig. 2d). Furthermore, compartment switching events were observed in specific genomic regions (Fig. 2e). Correspondingly, 3D remodeling revealed that the chromatin underwent a moderate decompaction, resulting in a disorganized and more open architecture (Fig. 2g). This was further confirmed by the increased ratio of the surface area (S) to the volume (V) (Fig. 2h).

TAD organization of rat Chr.X was subsequently analyzed. Profiling insulation scores across TAD boundaries (± 50 kb) confirmed that boundaries were characterized by local minima (characteristic troughs), forming distinct troughs (Fig. 2i). Analysis of boundary dynamics revealed that gained boundaries exhibited a decrease in insulation score, whereas lost boundaries showed an increase; unchanged boundaries remained stable (Fig. 2i). Representative genomic regions exhibiting both gained and lost TADs were identified (Fig. 2j). The TAD landscape was largely stable, with 86% of boundaries remaining unchanged, while a small but detectable fraction of boundaries were gained (7.9%) (Fig. 2k). In summary, the TAD architecture remained predominantly conserved, with a subset of novel TADs emerging. As a control, the 3D structure of Chr.1 in X^{sub} mESCs showed slighter changes when compared to Chr.1 in mouse ESCs (Fig. 2f, k; and Extended Data Fig. 2a-f), reinforcing that the observed structural reorganization was more specific to the exogenous Chr.X.

Taken together, no drastic alterations were observed in the overall 3D chromatin architecture of rat X chromosome in X^{sub} mESCs. However, reduction in compartment strength, a weakened segregation between compartments A and B, and moderate decompaction were evident on Chr.X. At a finer scale, while the majority of TAD boundaries in Chr.X were unchanged, the emergence of de novo TADs at specific loci was noted. This TAD stability aligns with the established role of CTCF binding sites in defining TAD boundaries and loops^{24–26}. As mentioned above, previous studies have demonstrated that heterochromatin is the primary factor driving the formation of chromosomal compartments^{27,28}. Besides, it has been reported that *Setdb1* knockdown/knockout resulted in H3K9me3 loss and gained ectopic binding of CTCF in mouse ESCs²⁹, suggested a potential role of H3K9me3 in prevent genomic CTCF binding, a core protein for TAD boundary formation. Thus, we extended our analysis to investigate the related epigenetic landscape of Chr.X, including heterochromatic histone modifications, to figure out the mechanisms of alterations in 3D genome organization.

Multi-omic characterization of H3K9me3 depletion and its association with local chromatin organization and transcriptional activity

We characterized the epigenetic landscapes in X^{sub} mESCs by analyzing multi-omics data from chromatin immunoprecipitation followed by sequencing (ChIP-Seq), Cleavage Under Targets and Tagmentation (CUT&Tag) assay, assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-Seq) and RNA-Seq (Fig. 3a and Extended Data Fig. 3a, b). Global erosion of the constitutive heterochromatin mark H3K9me3—the most pronounced change among all these epigenetic marks—characterized the substituted chromosome, as revealed by ChIP-seq and confirmed by CUT&Tag assay (Fig. 3a-c; Extended Data Fig. 3a). Notably, the longest of these H3K9me3 loss regions was a 3.2-Mb region (Extended Data Fig. 3c), indicating large-scale epigenetic reorganization. Meanwhile, we also observed focal increase in the facultative repressive mark H3K27me3 (Fig. 3a, b)—a potential compensatory response to the loss of global heterochromatin³⁰. Consistent with the observed changes

of A-B compartments mentioned above, the active marks H3K4me3 and H3K27ac, along with chromatin accessibility, exhibited concordant and overlapping patterns of focal upregulation (Fig. 3a, b). In contrast, all analyses showed less pronounced alterations on mouse autosomes (ranging from Chr.1 to Chr.19) in X^{sub} mESCs when compared to autosomes in mouse ESCs (Fig. 3b and Extended Data Fig. 3b). Notably, decompacted regions across Chr.X were closely associated with regions exhibiting pronounced loss of H3K9me3 (Fig. 3d). Similarly, de novo TAD boundaries formed on Chr.X in X^{sub} mESCs were also associated with substantial H3K9me3 loss (Fig. 3e, f). Multiple megabase-scale regions on Chr.X exhibited substantial H3K9me3 loss (Extended Data Fig. 3c, d), and displayed changes in chromatin 3D architecture (compartments and TADs) (Extended Data Fig. 3d). To summarize, our data demonstrated a profound epigenetic remodeling of Chr.X in X^{sub} mESCs when compared to Chr.X in rat ESCs, marked specifically by H3K9me3-dependent heterochromatin loss, which possibly contributed to changes in 3D organization of rat Chr.X.

Next, to assess the transcriptional impact of the chromosomal engineering, we performed pairwise correlation analysis of gene expression across samples. Importantly, X-linked gene expression in the X^{sub} mESCs most closely resembled that of the male rat ESCs ($R = 0.911$), exhibiting a higher correlation than X^{sub} mESCs vs. male mouse ESCs ($R = 0.759$) and mouse ESCs vs. rat ESCs ($R = 0.835$) (Fig. 3g and Extended Data Fig. 4a). The expression profile of key ESC marker genes was largely maintained (Fig. 1k). However, our data demonstrated a partial increase in transcriptional activity of rat Chr.X in X^{sub} mESCs (Fig. 3a, h). About 23% of the Chr.X-linked genes in X^{sub} mESCs were dysregulated (13% upregulated and 10% downregulated) compared to rat ESCs (Fig. 3h). This high degree of transcriptional conservation indicated that the rat Chr.X largely retained its native gene expression program with moderate disruption following its introduction into the mouse cellular environment. Meanwhile, 95% of the mouse autosome genes maintained their expression in the X^{sub} mESCs (Fig. 3h and Extended Data Fig. 4b). And the correlation patterns for autosomal genes were consistent across all sample pairs (Extended Data Fig. 4b). Then we assessed the X-chromosome-to-autosome (X: A) expression ratio to evaluate X chromosome dosage compensation. The X^{sub} mESCs exhibited an X: A ($X_{\text{rat}}: A_{\text{mouse}}$) ratio of coding genes comparable to male mouse ESCs, which was slightly higher than that of male rat ESCs (Fig. 3i and Extended Data Fig. 4c). This establishment of a mouse-like expression balance was mediated by a genome-wide upregulation of the rat Chr.X but not downregulation of autosomal genes, evidenced by an elevated X-derived RNA: total RNA ratio of X^{sub} mESCs compared to rat ESCs (Extended Data Fig. 4d) and the conservation of autosomal gene transcription mentioned above (Fig. 3h and Extended Data Fig. 4b). Besides, analysis of the Chr.X of X^{sub} mESCs identified specific pathway enrichments in differentially expressed genes (DEGs) (Extended Data Fig. 4e). GO analysis of autosomal DEGs indicated different pathways (Extended Data Fig. 4e). Overall, while autosomes maintained a high degree of expression conservation, the Chr.X globally conserved its expression pattern but underwent specific upregulation, resetting the X:A ratio to a level comparable to that in male mESCs.

To elucidate the mechanism underlying the transcriptional upregulation of rat X-linked genes in the mouse nuclear environment, we analyzed the chromatin landscape across gene bodies (± 3 kb from TSS to TES). We observed a pronounced global reduction of the heterochromatin mark H3K9me3, along with a focal increase in chromatin accessibility and H3K4me3 at transcription start sites (Extended Data Fig. 4f). Previous study suggested that novel TAD boundaries were related to proximal gene upregulation^{31,32}. Building on the above finding that the loss of H3K9me3 was associated with the formation of new TAD boundaries (Fig. 3e, f), we asked whether this structural remodeling contributed to gene upregulation. Importantly, genes located near these de novo TAD boundaries exhibited a significant tendency for transcriptional activation (Fig. 3j). Thus, our data supported a model in which H3K9me3 erosion facilitates the formation of new chromatin TAD boundaries, which promotes the upregulation of boundary-proximal genes.

Taken together, our results delineated the epigenetic fate of the rat Chr.X within a mouse nuclear environment. the chromosome underwent a profound epigenetic remodeling, characterized by a global loss of H3K9me3 and a consequent collapse of its native 3D architecture. While overall gene expression was broadly conserved, it underwent specific upregulation on the Chr.X-linked genes, resetting the X:A ratio. Critically, we established that the formation of new TAD boundaries at sites of H3K9me3 depletion provided a mechanistic link, explaining the upregulation of adjacent genes. The central role of H3K9me3 erosion prompted a question: what leads to the dissolution of this repressive mark in a foreign nuclear environment?

Defective SETDB1 recruitment at species-specific repeats as the molecular basis of heterochromatin erosion

We first mapped the genomic regions (Fig. 4a) and their top enriched motifs (Extended Data Fig. 5a) with reduced H3K9me3 and revealed that H3K9me3 loss were predominantly localized to distal intergenic sequences. This finding, together with the established role of H3K9me3 in silencing repetitive DNA, TE composition within these depleted regions was further examined. Our findings revealed a significant enrichment of specific TE subfamilies, including LINE: LINE1 (L1), Unknown: RatSatRep2, LTR: ERVK, and LTR: ERV1, among the H3K9me3-lost regions on Chr.X (Fig. 4b). Comparative genomic analysis of repetitive elements revealed that LINEs are highly abundant in both the rat and mouse genomes and are further enriched on their Chr.X (Fig. 4b, c). Notably, the rat Chr.X exhibits a substantial proportion of rat-specific LINEs (Fig. 4c). Previous studies have reported the evolutionary arms race between TEs and host defense mechanisms^{33–38}. Thus, we hypothesized that the rat-specific TE sequences may no longer be efficiently recognized by the mouse heterochromatin machinery. To test this, the evolutionary conservation of those top TE types within H3K9me3-lost regions were analyzed. Importantly, rat-specific TEs constituted the majority of these analyzed TE types, with RatSatRep2 being exclusively rat-specific (100%) (Fig. 4d). Consistent with this, species-specific LINEs displayed significantly sharper decreases in H3K9me3 levels than conserved LINEs (Fig. 4e). And the completely species-specific RatSatRep2 elements also exhibited substantial H3K9me3 reduction (Fig. 4e). Phylogenetic analysis of full-length L1

retrotransposons from mouse and rat further demonstrated species-specific clustering, with distinct clades corresponding to mouse and rat L1 sequences (Fig. 4f). Within each species, L1s derived from the Chr.X did not form separate clusters but were intermixed with autosomal L1s (Fig. 4f). Both the total number and the Chr.X-linked count of full-length L1s were higher in the rat genome compared to the mouse genome (Fig. 4f). Further analysis revealed that those young L1s emerging after rat-mouse divergence showed high H3K9me3 in their native state and suffered from greater H3K9me3 loss in X^{sub} mESCs (Fig. 4g). This heterochromatin breakdown was accompanied by the aberrant transcriptional activation of the Chr.X-linked TEs (Extended Data Fig. 5b). This indicated preferential epigenetic erosion of recently evolved repetitive elements.

To further investigate the mechanism of this recognition failure, the key H3K9 methyltransferase SET Domain Bifurcated Histone Lysine Methyltransferase 1 (SETDB1) occupancy was examined by ChIP-seq. SETDB1 showed markedly reduced binding at H3K9me3-lost regions in Chr.X of X^{sub} mESCs (Fig. 4h), despite unchanged expression level of *Setdb1*, *Suv39h2*, *Kdm4a*, and *Kdm4c* (Extended Data Fig. 5c). Within the H3K9me3-lost region, rat-specific L1s showed sharper decreases in SETDB1 levels than conserved L1s (Fig. 4i). These results indicated impaired recruitment of the H3K9me3 silencing machinery at these regions.

In summary, our data traced the origin of H3K9me3 erosion to an evolutionary incompatibility: the sequence divergence of rat-specific transposable elements compromises their targeting by the mouse heterochromatin machinery. This defect manifests through impaired SETDB1 recruitment, ultimately leading to large-scale heterochromatin loss of the rat X chromosome in a mouse cellular environment.

CROSS-mediated discovery of functional TAD boundaries at the *Rhox5* locus regulating cellular differentiation

Next, the developmental potential of X^{sub} mESCs was assessed using both in vivo and in vitro assays. Although X^{sub} mESCs expressed core ESC marker genes (Fig. 1j, k) and formed teratomas containing all three germ layers (Fig. 1l), they exhibited profound defects in differentiation capacity (Fig. 5a-d and Extended Data Fig. 6a-d). Diploid blastocyst injection and somatic cell nuclear transfer (SCNT) assays showed that X^{sub} mESCs contributed only to pre-implantation embryos but not to post-implantation stages (Fig. 5a-c and Extended Data Fig. 6a, b). Consistently, X^{sub} -mESC-derived teratomas retained abundant KI67⁺ undifferentiated cells (Extended Data Fig. 6c), indicating impaired terminal differentiation. In vitro, X^{sub} mESCs formed embryoid bodies (EBs) with delayed silencing of pluripotency genes (Fig. 5d). Given the in vivo developmental arrest around implantation, we further examined the naive-to-primed transition using an ESCs to epiblast-like cells (EpiLCs) differentiation model in vitro, which revealed a significant delay in the downregulation of naive pluripotency genes (Extended Data Fig. 6d).

To further investigate the potential contribution of the aforementioned epigenetic remodeling to the differentiation repair, we focused on genes near de novo TAD boundaries associated with H3K9me3 loss

on Chr.X. We ranked these genes by expression level and selected the top 10 candidates (Fig. 5e) for further functional validation. The majority of them were upregulated in the X^{sub} mESCs compared to controls (Fig. 5e). Among them, we focused on *Rhox5* (reproductive homeobox 5), a gene thought to be related to survival of male germ cell and tumor progress^{39,40}. This gene was highly expressed in our mouse ESCs but silenced upon differentiation (data not shown). Concomitantly, *Rhox5* failed to silence properly during differentiation, maintaining aberrantly high expression in X^{sub} mESC-derived EpiLCs, teratomas and EBs (Fig. 5f and Extended Data Fig. 6e). In X^{sub} mESCs, the locus surrounding *Rhox5* exhibited substantial epigenetic alterations: a prominent H3K9me3 domain immediately downstream was eroded, coinciding with the formation of a new TAD boundary, a local decrease in CTCF binding, and a gain of H3K4me3 (Fig. 5g). We found that the newly formed TAD boundaries were enriched with rat-specific TEs, specifically L1VL4a, RatSatRep2, and RNIAP1bLTR (Fig. 5g).

To test whether *Rhox5* dysregulation contribute to differentiation delay, we performed knockout (KO) of *Rhox5* in X^{sub} mESCs using CRISPR/Cas9 system (Fig. 5h and Extended Data Fig. 6f). *Rhox5* KO of X^{sub} mESCs led to larger EBs with improved lineage gene expression and decreased ESC marker gene expression (Fig. 5i and Extended Data Fig. 6g); improved naïve-to-primed differentiation in vitro (Fig. 5j); and reduced KI67⁺ cells after teratoma differentiation (Extended Data Fig. 6h). Importantly, *Rhox5* KO enabled development to E6.5 in SCNT assays (Fig. 5k). However, embryos did not survive to later stages in both SCNT and diploid blastocyst injection assays (Extended Data Fig. 6i and data not shown), indicating that *Rhox5* dysregulation was an important, but not the sole, contributor to the developmental failure.

The case of *Rhox5* demonstrated how the epigenetic instability of the rat Chr.X—initiated by TE-related H3K9me3 loss and 3D remodeling—propagates to disrupt key developmental genes. This connects a chromosome-scale epigenetic incompatibility to an observed developmental arrest, revealing of the important role of a single, X-linked gene during development. Our model thus provides a platform to identify X-linked dosage-sensitive genes involved in developmental regulation and investigate X chromosome dosage compensation evolution.

Discussion

In this study, we established CROSS as a high-fidelity platform for the intact and scarless substitution of hundred-megabase-scale chromosomes in mammalian cells. By replacing the mouse X chromosome with its rat ortholog, we demonstrated that the exogenous rat Chr.X is stably maintained in mouse ESCs in the absence of mouse Chr.X. The resulting mouse ESCs harboring only rat Chr.X (X^{sub} mESCs) could be stably maintained in culture, expressed key pluripotency markers, and retained the capacity to differentiate into all three germ layers. This system directly tests how a chromosome's native sequence governs its architecture and function within a heterologous nuclear environment.

A key technical challenge in large-scale mammalian genome engineering is maintaining the integrity of the epigenetic and structural landscape. Our findings directly linked the evolutionary divergence of

species-specific transposable elements (L1/RatSatRep2) to a failure in recruiting the histone methyltransferase SETDB1, resulting in pervasive H3K9me3 loss. This provides direct experimental support for the model that TEs act as genomic platforms for heterochromatin machinery, whose recognition is shaped by sequence-host co-evolution^{41–44}. Notably, a recent study reported that synthetic megabase human DNA delivered into mouse embryos also lacked H3K9me3 enrichment². This shared lack of heterochromatin in both cross-species systems suggests that H3K9me3 deposition on foreign DNA is not an automatic outcome of chromatinization and likely depends on co-evolved, sequence-specific recognition mechanisms. At the level of 3D organization, the rat X chromosome maintained broadly conserved architecture, consistent with previous reports suggesting the deterministic role of DNA sequence features like LINE/SINE distribution in establishing compartment identity⁴⁵ and CTCF-binding site distribution in TAD boundary formation by CTCF/cohesin-mediated loop extrusion mechanism^{24–26}. Nevertheless, However, heterochromatin erosion mediated localized structural changes, including reduced compartmentalization and the emergence of de novo TAD boundaries. These alterations were consistent with studies suggesting that *Setdb1* knockdown/knockout results in H3K9me3 loss and ectopic binding of CTCF binding, and thereby influence TAD and chromatin loop landscapes^{29,46,47}. Thus, we suggested that beyond the deterministic role of DNA sequence, the co-evolution of DNA and trans acting heterochromatin machinery dynamically shapes chromatin organization.

Besides, our system offers a discovery tool to systematically identify functional non-coding elements that regulate epigenetic patterning and 3D genome folding. The characterization of the *Rhox5* locus—where H3K9me3 loss and de novo TAD boundary formation correlate with differentiation defects—exemplifies the platform's capacity to reveal functional dependencies between 3D organization and developmental phenotypes. Unlike traditional knockout screens, the "evolutionary perturbation" approach offered by CROSS allows for the identification of regulatory elements within their native chromosomal context, providing insights into the functional requirements of the non-coding genome.

In parallel with very recent progress in high-fidelity chromosome transfer and engineering pipelines⁴⁸, our work addressed a distinct and important question: the functional outcome of cross-species chromosomal replacement. While existing methods enable the physical transfer of intact chromosomes, our model demonstrated that epigenetic compatibility—especially at species-specific repeats—constitutes a newly identified, requirement for stable chromosomal function.

In conclusion, the CROSS platform bridges the gap between small-scale gene editing and whole-genome synthesis. It provides a standardized workflow for probing genomic compatibility and offers a critical framework for the functional validation of large-scale synthetic DNA. As the field moves toward the de novo design of mammalian chromosomes, the principles of repeat-mediated epigenetic stability identified here will likely inform the engineering of functional and robust synthetic genomes.

Methods

Animal care and use

All animal experiments were approved by the Institutional Animal Care and Use Committee of the Institute of Zoology, Chinese Academy of Sciences (IOZ, CAS). The C57BL/6, B6D2F1 and CD-1 mice and Dark Agouti (DA) rats were obtained from Beijing Vital River Laboratories. All the mice used were raised in the SPF environment of the Laboratory Animal Center of the IOZ, CAS.

ESC derivation and culture

Mouse and rat ESCs were derived from embryos at E3.5 and cultured on feeder cells in 2i/LIF medium⁴⁹ plus 5% knockout serum replacement. The 2i/LIF medium consists of N2B27 medium (DMEM/F12, Neurobasal, N2, B27, GlutaMAX, β -Mercaptoethanol, 2% Bovine Serum Albumin, 10 mg/mL Insulin, and Penicillin-Streptomycin) supplemented with 1 μ M MEK inhibitor PD0325901 (Stemgent), 3 μ M GSK3b inhibitor CHIR99021 (Stemgent), and 10^3 units/mL mLIF (Millipore) for mouse ESCs and X^{sub} mESCs, or rLIF (Millipore) for rat ESCs.

Culture of other cell types

The A9 cells were purchased from the Cell Bank of the Chinese Academy of Sciences, and cultured in DMEM medium supplemented with 10% fetal bovine serum (FBS) and penicillin/streptomycin. rES-A9 hybrid cells were also cultured in the same medium. Feeder cells, mouse embryonic fibroblasts inactivated by treatment with 10 μ g/ml mitomycin C, were cultured in the above medium.

Genetic engineering of chromosome donor cells and recipient cells

To enable site-specific DNA cleavage and integration, we used the CRISPR-Cas9 system to generate double-strand breaks at designated genomic loci. Single-guide RNAs (sgRNAs) were designed using Cas-Designer (<http://www.rgenome.net/cas-designer/>) and Cas-OFFinder (<http://www.rgenome.net/cas-offinder/>). SgRNAs with a score higher than 60 and minimal predicted off-target activity in the genome were selected. Each sgRNA spacer sequences was cloned into a BsaI-digested pUC19-U6-sgRNA vector by ligation of annealed oligonucleotides. SgRNA spacer sequences were listed in Supplementary Table 4.

Donor cell lines. To engineer the donor cell line, we inserted an pEF1 α -EGFP-PuroR cassette downstream of the *Pgk1* gene on the X chromosome of DA rat ESCs (Fig. 1d, e). The engineered rat ESCs were further fused with A9 cell lines to produce rES-A9 hybrid cells (see *Cell fusion* section), which served as Chr.X donor cells in MMCT assay.

Recipient cell lines. We used a male C57BL/6 mESC line that is competent for diploid aggregation and tetraploid complementation and carries a random ERT2-Cre-ERT2-DsRed integration. On this parental line, we knocked out the *Hprt* gene in X chromosome and inserted an pEF1 α -TK-DsRed-NeoR cassette at the same locus in one step of this ESC line (Fig. 1d, e). For each targeting experiment, we co-delivered a pCAG-SpCas9-EGFP plasmid, the relevant U6-sgRNA plasmid, and the appropriate homology-directed repair (HDR) donor plasmid. The transfection was carried out using the electroporation transfection instrument Neon transfection system (Invitrogen). 48–72 hours after electroporation, GFP- or DsRed-positive cells were sorted by flow cytometry, respectively. Following drug selection, single-cell clones were isolated, expanded, and genotyped to confirm correct modifications.

Cell fusion

Cell fusion was performed using Polyethylene Glycol 1500 (PEG 1500, ROCHE). The A cells and B cells to be fused were digested, terminated, and washed with PBS. Then, the cells were resuspended in PBS and counted. In brief, taking 10^8 rat ESCs and 2×10^7 A9 cells as an example. The mixed cell suspension was centrifuged: at a speed of 1000–1500 rpm for 2–3 minutes. Remove the supernatant completely. Place the centrifuge tube in a 37 °C water bath to maintain the fusion process at 37 °C. Slowly and continuously add 1 mL of 50% PEG 1500 solution preheated to 37 °C to the centrifuge tube within 1 minute, and continue to stir. Continue stirring the cells for 1–2 minutes. Slowly and continuously add 1 mL of preheated KnockOut™ DMEM (Gibco) medium at 37 °C to the centrifuge tube and continue to stir. Add 3 mL of preheated KnockOut™ DMEM medium at 37 °C slowly and continuously, and continue to stir. Slowly add 10 mL of preheated KnockOut™ DMEM medium at 37 °C, and the cell suspension will be incubated at 37 °C for 5 minutes. Centrifuge, discard the supernatant. Resuspend the cell clumps in fresh culture medium, and after 24 hours of culture, add the drug sieve culture medium.

Microcell-mediated chromosome transfer (MMCT)

The MMCT procedures were primarily adapted from previously established protocols^{12,50} with modifications. In brief, we first screened several candidate cell lines (rat ESC-A9 hybrids harboring rat Chr.X) for micronucleation efficiency and selected efficient donors for subsequent MMCT experiments. Donor cell lines were treated with 0.04–0.05 μ g/mL colcemid for 48 h to form micronuclei. After trypsinization and washing, cells were resuspended in KnockOut™ DMEM containing 10 μ g/mL cytochalasin B and subjected to density gradient centrifugation in Percoll: KnockOut™ DMEM (1: 1) at $16,000 \times g$ for 60–75 min. The microcell enriched fraction was collected, washed, and sequentially filtered through 8 μ m and 5 μ m membranes to isolate microcells containing one to several chromosomes. After additional washing and counting, purified microcells were fused with recipient cells using PEG 1500 as described in the *Cell Fusion* section.

Screening and identification of chromosome-transfer cell lines (X^{tra} mESC lines)

48 hours after MMCT, cells dual positive for GFP and RFP were enriched by fluorescence activated cell sorting (FACS). Puromycin and HAT were then added to the culture medium for selection. After 1–2 weeks, the clones were picked and expanded. Genomic DNA was extracted from each single clone cell line, and successful transfer of the rat X chromosome and the absence of contaminating other chromosomes (including rat autosomes) were verified by genotyping, karyotyping, and whole-genome sequencing. The primer sequences used for genotyping were listed in Supplementary Tables 1 and 3.

Screening and identification of chromosome-substituted cell lines (X^{sub} mESC lines)

To obtain subclones that had lost the endogenous mouse X chromosome (and its linked HSV-TK negative selection marker) from the established X^{tra} mESC lines, cells were treated with ganciclovir (GCV). GCV-resistant clones were picked and expanded. Initial genotypic screening for loss of the mouse X chromosome was conducted using locus-species PCR (primers listed in Supplementary Table 2). Finally, putative positive clones were then subjected to karyotype analysis, and the complete replacement of the mouse X chromosome by the rat X chromosome was confirmed by whole-genome sequencing.

Karyotyping analysis

Treat the cells with 0.05 $\mu\text{g/mL}$ colcemid for 2 to 6 hours. After trypsin digestion, wash the cells with phosphate buffered saline (PBS). Resuspend the cells in 0.075 M KCl hypotonic solution and incubate at 37°C for 30 minutes. Then, fix the cells on a 3:1 (volume ratio) methanol-acetic acid fixative at 4°C for 30 minutes. The cell suspension is dropped onto a pre-cooled glass slide under the influence of gravity. The cells were stained using Giemsa Stain.

Immunofluorescence staining

Immunofluorescence staining was performed as previously described⁵¹. Fix the cells with 4% paraformaldehyde for 10 minutes, then permeabilize and block the cells with 0.5% Triton X-100 containing 2% BSA for 1 hour. Add the primary antibodies at 4°C for overnight incubation. Incubate with the secondary antibody at room temperature for 1 hour. DNA was stained with Hoechst 33342 for 10 minutes. Imaging was performed using a two-photon confocal microscope (Leica, TCS Sp8). Antibodies: anti-OCT4 (ab19857, Abcam); anti-NANOG (ab80892, Abcam); anti-SSEA1 (ab16285, Abcam); anti-Ki67 (ab16667, Abcam).

Formation of teratomas

Teratoma assay was performed as previously published⁵¹. Briefly, 0.05% pancreatic trypsin was used to digest and collect ESCs. Then 10^7 cells were then subcutaneously injected into both sides of the abdomen of SCID mice. After 21 days or one month, the tumors were excised and subjected to histological analysis through fixation and H&E staining or immunofluorescence staining.

Formation of Embryoid bodies (EBs)

First, the ESCs are digested into single cells using 0.05% pancreatic trypsin. Subsequently, the ESCs were suspended and cultured in the N2B27 medium. After approximately 8 days, the differentiated EB spheres were picked out.

Oocyte collection

Oocyte collection was performed as previously described⁵². Eight-week-old female mice were superovulated by consecutive injection of 7.5 IU human chorionic gonadotropin (hCG) and 7.5 IU pregnant mare's serum gonadotropin (PMSG). MII oocytes were harvested from ovarian oviduct 13–15 hours after the injections. Based on the instructions, derived oocytes were cultured in M16 medium after being washed with M2 medium. Hyaluronidase was used to remove the cumulus cells. Oocytes were then cultured in M2 medium at 37°C with 5% CO₂.

Somatic cell nuclear transfer (SCNT)

SCNT was conducted follow protocols of the published article⁵³. In brief, inject ESCs into enucleated oocytes to form embryos. The embryos are incubated in M16 medium for 1 hour, then activated in the CZB medium containing 5 mg/mL Cytochalasin B (Abcam) and 10 mM SrCl₂ (Sigma) for 5 to 6 hours. Subsequently, the embryos are transferred to M16 medium and cultured for 18 to 22 hours at 5% CO₂ and 37°C. Then they are transferred to the KSOM solution and cultured for 3.5 days at 5% CO₂ and 37°C.

Diploid blastocyst injection

This assay was carried out following the published article⁵⁴. 3.5 days after the mating of superovulated female CD-1 mice with male CD-1 mice, blastocysts were taken from their uteri. The ESCs were digested with trypsin and 12 to 15 cells were microinjected into each blastocyst. After culturing for 1 to 4 hours, the treated embryos were transferred to the oviducts of pseudo-pregnant CD-1 mice at 0.5 days post-conception. Identification of the chimeras were based on GFP and coat color.

Quantitative real-time PCR

Total RNA was extracted using TRIzol® Reagent (Life Technologies) and further purified with the PureLink® RNA Mini Kit (Invitrogen). Genomic DNA was removed by treatment with DNase I (TIANGEN). Subsequently, the RNA was reverse-transcribed into cDNA using the ReverTra Ace® qPCR RT Master Mix with gDNA Remover Kit. The real-time fluorescent quantitative PCR system contained SYBR qPCR Mix and 50 · ROX reference dye. The fluorescence quantitative PCR experiment was performed using the QuantStudio 6 Pro Real-Time PCR System (Thermo Fisher Scientific). The reference gene was *Gapdh*. The primer sequences used were listed in Supplementary Table 5.

CRISPR/Cas9-mediated deletion of *Rhox5*

The design and construction of CRISPR/Cas9 system was described in the above *Genetic engineering of chromosome donor cells and recipient cells* section. Briefly, we co-delivered a pCAG-SpCas9-EGFP plasmid and the relevant U6-sgRNA plasmid into X^{sub} mESCs to knockout *Rhox5* gene. SgRNA spacer sequences were listed in Supplementary Table 4.

Reference Genome Construction

Reference genomes for *Mus musculus* (mm39) and *Rattus norvegicus* (rn7) were downloaded from the Ensembl database (Release 106). A combined mm39-rn7 reference genome was generated by merging the sequences and annotations from both assemblies to facilitate integrated cross-species analysis.

Whole genome sequencing

Genomic DNA was extracted using the E.Z.N.A.® MicroElute Genomic DNA Kit (Omega Bio-tek) according to the manufacturer's instructions. DNA quality was assessed by agarose gel electrophoresis and quantified using a Qubit® 3.0 Fluorometer. Library construction and high-throughput sequencing were performed on Illumina NovaSeq 6000 platform by Annoroad Gene Technology (Beijing) Co., Ltd.

Raw reads underwent quality assessment using FastQC (v0.11.5) (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Adapter sequences and low-quality bases were subsequently trimmed using trim_galore (v0.6.6) (<https://github.com/FelixKrueger/>

TrimGalore). The cleaned reads were aligned to the above reference genome (mm39-rn7) using BWA-MEM (v0.7.17)⁵⁵ with default parameters, followed by duplicate marking and removal with GATK (v4.1.8)⁵⁶. The resulting BAM files were used for copy number variation (CNV) analysis with SCCNV (v1.0.2)⁵⁷, generating visualizations of genomic segments' copy number states.

Hi-C library construction and data analysis.

Hi-C libraries were constructed using standard protocol by Annoroad and were sequenced on Illumina platforms. Hi-C data were processed using HiC-Pro (v3.1.0)⁵⁸ to generate raw contact matrices at 10-kb and 100-kb genomic bin resolutions. Systematic biases in the contact matrices were corrected using the 'hicCorrectMatrix' function from HiCExplorer (v3.6)⁵⁹. The corrected 100 kb matrix was utilized for global structural assessment. Contact matrix visualization and saddle plots were generated with cooltools (v0.7.1)⁶⁰, and three-dimensional (3D) genome structures were inferred using miniMDS⁶¹.

Chromatin compartment (A/B) analysis was performed using cscoreTool (v1.1)⁶². A compartment switch was defined as a genomic bin where the absolute difference in compartment score between samples exceeded 0.3.

Topologically Associating Domains (TADs) were identified at 10-kb resolution by calculating the insulation score with cooltools⁶⁰. Dynamic TAD boundaries were classified as: *Gain boundary*, a genomic locus identified as a TAD boundary exclusively in the X^{sub} mESC group; *Lost boundary*, a genomic locus identified as a TAD boundary exclusively in the control (ctrl) group; *Maintained boundary*, a TAD boundary called in both groups. In all analyses, the ctrl was rat ESCs when assessing the rat X chromosome, and mouse ESCs when assessing autosomes. TADs were visualized using pyGenomeTracks⁶³, and insulation scores distribution plot were generated with deepTools⁶⁴.

ChIP-seq library preparation

Chromatin immunoprecipitation was performed using the SimpleChIP® Plus Sonication Chromatin IP Kit (Cell Signaling Technology, 56383) following the manufacturer's protocol. Briefly, cells were cross-linked with formaldehyde, lysed, and chromatin was fragmented by sonication. Immunoprecipitation was carried out with target-specific antibodies. Antibodies: anti-H3K9me3 (ab8898, Abcam), anti-H3K27me3 (9733s, Cell Signaling technology), anti-H3K27ac (ab4729, Abcam), anti-H3K4me3 (ab8580, Abcam), anti-CTCF (3418S, Cell Signaling technology), anti-SETDB1 (11231-1-AP, Proteintech). After reversal of cross-links, DNA was purified using the kit's spin columns. Libraries were constructed and sequenced by Annoroad on Illumina NovaSeq platform.

ATAC-seq library preparation

ATAC-seq libraries were prepared by Annoroad using their standard protocol, based on hyperactive Tn5 transposase⁶⁵. In brief, 50,000 viable cells were used for transposase-mediated tagmentation, followed by PCR amplification with indexed primers.

ChIP-seq and ATAC-seq data analysis

ChIP-seq and ATAC-seq libraries were sequenced on the Illumina NovaSeq platform by Annoroad. ChIP-seq and ATAC-seq datasets were processed through an identical pipeline. Following quality control with FastQC (v0.11.5), adapter sequences were removed using trim_galore (v0.6.6). The processed reads were aligned to the reference genome with Bowtie2 (v2.3.5)⁶⁶ under the parameters “--sensitive --end-to-end”, and PCR duplicates were eliminated using a custom script. Read coverage was normalized to Reads Per Kilobase per Million mapped reads (RPKM) using the ‘bamCoverage’ utility from deepTools (v3.5.1)⁶⁴. The bigwig files were visualized using pyGenomeTracks (v3.6)⁶³. The mean read count distribution plot was generated using deeptools. Peak calling was executed with MACS2 (v2.1.2)⁶⁷ with the parameters “-p 0.01 --nomodel --nolambda”. All peaks were segmented into 1-kb windows using the sliding window method. These regions were categorized as follows: *gain regions*, peaks present in the X^{sub} mESC group but absent in the control group (ctrl), exhibiting an RPM fold-change (X^{sub} mESC/ctrl) ≥ 4 ; *lost regions*, peaks present in the ctrl group but absent in the X^{sub} mESC group, exhibiting an RPM fold-change (ctrl/ X^{sub} mESC) ≥ 4 ; *maintained regions*, peaks called in both X^{sub} mESC and ctrl group. In all analyses, the ctrl was rat ESCs when assessing the rat X chromosome, and mouse ESCs when assessing autosomes.

Motif enrichment

de novo motif discovery within specific peaks was performed using the MEME suite (v5.5.2)⁶⁸.

CUT&Tag library preparation and data processing

CUT&Tag libraries were prepared using the Hyperactive Universal CUT&Tag Assay Kit for Illumina Pro (Vazyme, TD904) following the manufacturer’s instructions. Cells were bound to ConA beads, permeabilized with digitonin, and incubated with primary antibodies (anti-H3K9me3, ab8898, Abcam) and secondary antibodies. Protein A/G-Tn5 transposase was then targeted to antibody-bound regions for tagmentation. DNA was extracted, amplified with indexed primers, and purified using DNA Clean Beads. Libraries were quantified and sequenced on the Illumina NovaSeq platform.

After removing the adapter sequences with trim_galore, CUT&Tag reads were mapped to the reference genome (mm39-rn7) using bowtie2⁶⁶ with the parameters “--local --very-sensitive --no-mixed --no-discordant”. The redundant reads were removed using our own script. The mean read count distribution plot was generated using deeptools⁶⁴.

RNA-Seq library construction and data analysis

Total RNA was extracted as described in the *Quantitative real-time PCR* section. Libraries were quantified and sequenced on Illumina platforms. Raw RNA-seq reads underwent quality assessment using FastQC. Adapter sequences and low-quality bases were trimmed with trim_galore. The resulting high-quality reads were aligned to the composite mm39-rn7 genome using STAR (v2.7.1a)⁶⁹. Uniquely mapped reads

were isolated using a custom script for subsequent quantification. Transcript abundance was estimated with StringTie (v2.0)⁷⁰ and normalized to Fragments Per Kilobase of transcript per Million mapped reads (FPKM). Differential gene expression analysis was conducted using the DESeq2 package (v1.38.0)⁷¹ in R. Genes demonstrating an absolute log₂ fold change > 1 with p-value < 0.05 were designated as differentially expressed. Functional enrichment analysis for Gene Ontology (GO) terms was performed on these gene sets using the DAVID⁷² bioinformatics database.

X:A ratio: StringTie⁷⁰-derived gene expression tables were merged into a unified matrix, retaining annotated protein-coding genes on autosomes and the X chromosome; genes in the top 5% of maximal expression were excluded. Two gene sets were used: all expressed protein-coding genes and one-to-one mouse–rat homologous protein-coding genes obtained from Ensembl BioMart⁷³. X:A ratio was assessed using both expression-based and read-level approaches. For expression-based analysis, genes with FPKM > 2 were retained, and a bootstrapping strategy⁷⁴ was applied to account for the smaller number of X-linked genes: in each iteration, a random autosomal subset equal in size to the X-linked gene set was sampled, and median expression values were compared over 10,000 iterations to derive X:A ratios and confidence intervals. For a gene-annotation–independent estimate, chromosome-wise read counts and lengths were extracted using samtools⁷⁵ idxstats, and a length-normalized X:A read-density ratio was computed. Statistical analyses and visualization were performed in R using ggplot2⁷⁶, with group comparisons conducted using the Wilcoxon rank-sum test and results displayed as boxplots with jittered points.

Transposable element (TE) expression analysis

Genomic annotations for repetitive elements were obtained from the UCSC Genome Browser for the respective mm39 and rn7 assemblies. TE expression was quantified using the SQuIRE⁷⁷ software (v0.9.9) under default parameters. Differentially expressed TEs were identified using the same thresholds as for genes: $|\log_2\text{FoldChange}| > 1$ and p-value < 0.05.

Evolutionary analysis of mouse and rat TEs

The evolutionary ages of TEs were estimated based on RepeatMasker output files acquired from the UCSC Genome Browser for the mouse (mm39) and rat (rn7) genomes. For each repeat element locus, the "milliDiv" value (representing the percentage divergence from the consensus sequence) was extracted directly from the output. The Jukes-Cantor method was applied using species-specific substitution rates: 5.2×10^{-9} per base per year for mouse and 5.5×10^{-9} per base per year for rat^{4,78}. The evolutionary age (in years) was then calculated using the formula: age = divergence rate / substitution rate, yielding an estimated age for each repetitive element locus⁷⁹.

Phylogenetic analysis of mouse and rat full length LINE1. Full-length LINE1 sequences (> 6.5 kb) were extracted from the telomere-to-telomere (T2T) mouse genome and the rat genome (RN8) using RepeatMasker (v4.1.1)⁸⁰ based on the Repbase repeat database. Sequences were subsequently clustered using CD-HIT (v4.8.1)⁸¹ with the parameters “-d 0 -c 0.8 -aL 0.80” to reduce redundancy and define representative sequences. Multiple sequence alignment of the clustered sequences was performed with MAFFT (v7.475)⁸² under default settings. A phylogenetic tree was constructed from the aligned sequences using FastTree (v2.1.11)⁸³ with default parameters. Finally, the resulting tree was visualized and annotated using the iTOL platform⁸⁴.

Statistics and reproducibility

Statistical analyses were performed in R using ggplot2⁷⁶ or Graphpad Prism. Levels of significance were calculated using the two-tailed Student's t test. In all figures: *, p value < 0.05; **, p value < 0.01, ***, p value < 0.001; ****, p value < 0.0001. The number of independent biological replicates and statistical approach for each experiment were described in the figure legends or the *Methods* section.

Data availability

The sequencing data are available at Genome Sequence Archive (GSA) of China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences (<https://ngdc.cnbc.ac.cn/gsa/>) under accession CRA035509. Previously published raw data used in this paper is available at Gene Expression Omnibus under accessions GSE44150, GSE97966⁸⁵, GSE178701⁸⁶, GSE220805⁸⁷, and GSE90516⁸⁸.

Declarations

Acknowledgments

This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDC0200000 to W.L. and XDA0510400); the National Key Research and Development Program (2022YFA1103600 to L.W. and 2019YFA0903800); the National Natural Science Foundation of China (32225030 and 82488301 to W.L.); the CAS Project for Young Scientists in Basic Research (YSBR-012 to W.L.); the Beijing Natural Science Foundation (Z230011 to W.L.); Initiative Scientific Research Program, Institute of Zoology, Chinese Academy of Sciences (2023IOZ0101). We acknowledge Runze Ma, Xiaohua Shen and Bing Zhu for their helpful suggestions. We thank Lingyun Wei, Youjia Shu, Qing Meng, Peipei Long, Ming Ge, Weiyu Jin, Shiwen Li, Xili Zhu, Hua Qin and Xia Yang for their technical assistance.

Author contributions

W.L., Q.Z., Y.M. and L.W. conceived and designed the study. Y.M., L.W., Y.Z., K.X., B.Z., and S.M. performed the experiments; N.Y., Y.M., W.L., L.W., G.F., Q.Z., Y.D., J.L., Y.Z. and Q.N. analyzed the data; W.L. and Q.Z. supervised the research; W.L., Y.M., Y.Z., N.Y., Q.Z., Y.C., and Q.Z. wrote the manuscript, assisted by the other authors.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Wei Li.

References

1. Meneu L et al (2025) Sequence-dependent activity and compartmentalization of foreign DNA in a eukaryotic nucleus. *Science* 387. 10.1126/science.adm9466
2. Liu Y et al (2025) De novo assembly and delivery of synthetic megabase-scale human DNA into mouse early embryos. *Nat Methods* 22:1686–1697. 10.1038/s41592-025-02746-8
3. Venter JC et al (2001) The Sequence of the Human Genome. *Science* 291:1304–1351. 10.1126/science.1058040
4. Gibbs RA et al (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* 428:493–521. 10.1038/nature02426
5. Nurk S et al (2022) The complete sequence of a human genome. *Science* 376:44–53. 10.1126/science.abj6987
6. Venter JC et al (2001) The sequence of the human genome. *Science* 291:1304–1351. 10.1126/science.1058040
7. Waterston RH et al (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature* 420:520–562. 10.1038/nature01262
8. Dixon JR et al (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485:376–380. 10.1038/nature11082
9. Nora EP et al (2012) Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 485:381–385. 10.1038/nature11049
10. Lieberman-Aiden vBN, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *science*
11. Fournier RE, Ruddle FH (1977) Microcell-mediated transfer of murine chromosomes into mouse, Chinese hamster, and human somatic cells. *Proceedings of the National Academy of Sciences* 74, 319–323. 10.1073/pnas.74.1.319

12. Koi M, Shimizu M, Morita H, Yamada H, Oshimura M (1989) Construction of mouse A9 clones containing a single human chromosome tagged with neomycin-resistance gene via microcell fusion. *Japanese J cancer research: Gann* 80:413–418. 10.1111/j.1349-7006.1989.tb02329.x
13. Tomizuka K et al (1997) Functional expression and germline transmission of a human chromosome fragment in chimaeric mice. *Nat Genet* 16:133–143. 10.1038/ng0697-133
14. Diana Hernandez PJM, Joanne E, Martin VLJ, Tybulewicz, Elizabeth MC (1999) Fisher. Transchromosomal mouse ESCs and chimeric mice that containing free segregating segments of human chromosome 21. *Hum Mol Genet*
15. O'Doherty A et al (2005) An aneuploid mouse strain carrying human chromosome 21 with Down syndrome phenotypes. *Science* 309:2033–2037. 10.1126/science.1114535
16. Jinek M et al (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337:816–821. 10.1126/science.1225829
17. Cong L et al (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* 339:819–823. 10.1126/science.1231143
18. ohno S (1966) *Sex Chromosomes and Sex-Linked Genes*. Springer-, Berlin Heidelberg
19. Payer B, Lee JT (2008) X Chromosome Dosage Compensation: How Mammals Keep the Balance. *Annu Rev Genet* 42:733–772. 10.1146/annurev.genet.42.110807.091711
20. Graves JAM (2016) Evolution of vertebrate sex chromosomes and dosage compensation. *Nat Rev Genet* 17:33–46. 10.1038/nrg.2015.2
21. Hsu PD, Lander ES, Zhang F (2014) Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell* 157:1262–1278
22. O'Connell MR et al (2014) Programmable RNA recognition and cleavage by CRISPR/Cas9. *Nature* 516:263–266. 10.1038/nature13769
23. Liu Y et al (2025) De novo assembly and delivery of synthetic megabase-scale human DNA into mouse early embryos. *Nat Methods* 22:1686–1697. 10.1038/s41592-025-02746-8
24. Schwarzer W et al (2017) Two independent modes of chromatin organization revealed by cohesin removal. *Nature* 551:51–56. 10.1038/nature24281
25. Nora EP et al (2017) Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* 169:930–944e922. 10.1016/j.cell.2017.05.004
26. Zuin J et al (2013) Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proceedings of the National Academy of Sciences* 111, 996–1001. 10.1073/pnas.1317788111
27. Falk M et al (2019) Heterochromatin drives compartmentalization of inverted and conventional nuclei. *Nature* 570:395–399. 10.1038/s41586-019-1275-3
28. Nuebler J, Fudenberg G, Imakaev M, Abdennur N, Mirny LA (2018) Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proc Natl Acad Sci U S A* 115:E6697–e6706. 10.1073/pnas.1717730115

29. Tam PLF, Cheung MF, Chan LY, Leung D (2024) Cell-type differential targeting of SETDB1 prevents aberrant CTCF binding, chromatin looping, and cis-regulatory interactions. *Nat Commun* 15. 10.1038/s41467-023-44578-0
30. Montavon T et al (2021) Complete loss of H3K9 methylation dissolves mouse heterochromatin organization. *Nat Commun* 12. 10.1038/s41467-021-24532-8
31. Zhang Y et al (2019) Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells. *Nat Genet* 51:1380–1388. 10.1038/s41588-019-0479-7
32. Hong Y et al (2024) SAFB restricts contact domain boundaries associated with L1 chimeric transcription. *Mol Cell* 84:1637–1650e1610. 10.1016/j.molcel.2024.03.021
33. Thomas JH, Schneider S (2011) Coevolution of retroelements and tandem zinc finger genes. *Genome Res* 21:1800–1812. 10.1101/gr.121749.111
34. Pontis J et al (2019) Hominoid-Specific Transposable Elements and KZFPs Facilitate Human Embryonic Genome Activation and Control Transcription in Naive Human ESCs. *Cell Stem Cell*. 10.1016/j.stem.2019.03.012
35. Jacobs FM et al (2014) An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature* 516:242–245. 10.1038/nature13760
36. Iwasaki YW, Shoji K, Nakagawa S, Miyoshi T, Tomari Y (2025) Transposon–host arms race: a saga of genome evolution. *Trends Genet* 41:369–389. 10.1016/j.tig.2025.01.009
37. Imbeault M, Helleboid PY, Trono D (2017) KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature* 543:550–554. 10.1038/nature21683
38. Bruno M, Mahgoub M, Macfarlan TS (2019) The Arms Race Between KRAB–Zinc Finger Proteins and Endogenous Retroelements and Its Impact on Mammals. *Annu Rev Genet* 53:393–416. 10.1146/annurev-genet-112618-043717
39. Hu Z, Shanker S, MacLean JA 2nd, Ackerman SL, Wilkinson MF (2008) The RHOX5 homeodomain protein mediates transcriptional repression of the netrin-1 receptor gene *Unc5c*. *J Biol Chem* 283:3866–3876. 10.1074/jbc.M706717200
40. Wilkinson MF, Kleeman J, Richards J, MacLeod CL (1990) A novel oncofetal gene is expressed in a stage-specific manner in murine embryonic development. *Dev Biol* 141:451–455
41. Haws S, Simandi Z, Barnett R, Phillips-Cremins (2022) J. 3D genome, on repeat: Higher-order folding principles of the heterochromatinized repetitive genome. *Cell* 185:2690–2707. 10.1016/j.cell.2022.06.052
42. Lawson HA, Liang Y, Wang T (2023) Transposable elements in mammalian chromatin organization. *Nat Rev Genet* 24:712–723. 10.1038/s41576-023-00609-6
43. Musselman CA, Lalonde ME, Côté J, Kutateladze TG (2012) Perceiving the epigenetic landscape through histone readers. *Nat Struct Mol Biol* 19:1218–1227. 10.1038/nsmb.2436

44. Risca VI, Denny SK, Straight AF, Greenleaf WJ (2017) Variable chromatin structure revealed by in situ spatially correlated DNA cleavage mapping. *Nature* 541:237–241. 10.1038/nature20781
45. Lu JY et al (2021) Homotypic clustering of L1 and B1/Alu repeats compartmentalizes the 3D genome. *Cell Res.* 10.1038/s41422-020-00466-6
46. Jiang Y et al (2017) The methyltransferase SETDB1 regulates a large neuron-specific topological chromatin domain. *Nat Genet* 49:1239–1250. 10.1038/ng.3906
47. Gualdrini F et al (2022) H3K9 trimethylation in active chromatin restricts the usage of functional CTCF sites in SINE B2 repeats. *Genes Dev* 36:414–432. 10.1101/gad.349282.121
48. Petris G et al (2025) High-fidelity human chromosome transfer and elimination. *Science* 390:1038–1043. 10.1126/science.adv9797
49. Silva J et al (2008) Promotion of Reprogramming to Ground State Pluripotency by Signal Inhibition. *PLoS Biol* 6:2237–2247. 10.1371/journal.pbio.0060253
50. Paulis M (2011) *Mammalian Chromosome Engineering: Methods and Protocols* (ed Gyula Hadlaczký) 57–67. Humana
51. Zhao XY et al (2009) iPS cells produce viable mice through tetraploid complementation. *Nature* 461:86–90. 10.1038/nature08267
52. Wan H et al (2013) Parthenogenetic haploid embryonic stem cells produce fertile mice. *Cell Res* 23:1330–1333. 10.1038/cr.2013.126
53. Zhou Q et al (2003) Generation of fertile cloned rats by regulating oocyte activation. *Science* 302:1179. 10.1126/science.1088313
54. Li W et al (2012) Androgenetic haploid embryonic stem cells produce live transgenic mice. *Nature* 490:407–411. 10.1038/nature11435
55. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760. 10.1093/bioinformatics/btp324
56. McKenna A et al (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20:1297–1303. 10.1101/gr.107524.110
57. Dong X, Zhang L, Hao XX, Wang T, Vijg JSCCNV (2020) A Software Tool for Identifying Copy Number Variation From Single-Cell Whole-Genome Sequencing. *Front Genet* 11, doi:ARTN 50544110.3389/fgene.505441 (2020)
58. Servant N et al (2015) HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol* 16, doi:ARTN 25910.1186/s13059-015-0831-x
59. Wolff J et al (2018) Galaxy HiCExplorer: a web server for reproducible Hi-C data analysis, quality control and visualization. *Nucleic Acids Res* 46:W11–W16. 10.1093/nar/gky504
60. Abdennur N et al (2024) Cooltools: Enabling high-resolution Hi-C analysis in Python. *Plos Comput Biol* 20, doi:ARTN e101206710.1371/journal.pcbi.1012067
61. Rieber L, Mahony S (2017) miniMDS: 3D structural inference from high-resolution Hi-C data. *Bioinformatics* 33:l261–l266. 10.1093/bioinformatics/btx271

62. Zheng XB, Zheng YX (2018) CscoreTool: fast Hi-C compartment analysis at high resolution. *Bioinformatics* 34:1568–1570. 10.1093/bioinformatics/btx802
63. Lopez-Delisle L et al (2021) pyGenomeTracks: reproducible plots for multivariate genomic datasets. *Bioinformatics* 37:422–423. 10.1093/bioinformatics/btaa692
64. Ramirez F, Dundar F, Diehl S, Gruning BA, Manke T (2014) deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* 42:W187–191. 10.1093/nar/gku365
65. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ (2015) ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol* 109. .29.21–21.29.29
66. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10. doi:ARTN R10.1186/gb-2009-10-3-r25
67. Zhang Y et al (2008) Model-based Analysis of ChIP-Seq (MACS). *Genome Biol* 9. doi:ARTN R13710.1186/gb-2008-9-9-r137
68. Bailey TL, Johnson J, Grant CE, Noble WS The MEME Suite
69. Dobin A et al (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21. 10.1093/bioinformatics/bts635
70. Pertea M et al (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* 33:290–295. 10.1038/nbt.3122
71. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15, doi:ARTN 55010.1186/s13059-014-0550-8
72. Sherman BT et al (2022) DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res* 50:W216–W221. 10.1093/nar/gkac194
73. Kinsella RJ et al (2011) Ensembl BioMart: a hub for data retrieval across taxonomic space. *Database (Oxford)* bar030. 10.1093/database/bar030 (2011)
74. Borensztein M et al (2017) Xist-dependent imprinted X inactivation and the early developmental consequences of its failure. *Nat Struct Mol Biol* 24:226–233. 10.1038/nsmb.3365
75. Danecek P et al (2021) Twelve years of SAMtools and BCFtools. *GigaScience* 10. 10.1093/gigascience/giab008
76. Ginestet C (2011) ggplot2: Elegant Graphics for Data Analysis. *Journal of the Royal Statistical Society Series A: Statistics in Society* 174, 245–246. 10.1111/j.1467-985X.2010.00676_9.x %J Journal of the Royal Statistical Society Series A: Statistics in Society
77. Yang WR, Ardeljan D, Pacyna CN, Payer LM, Burns KH (2019) SQuIRE reveals locus-specific regulation of interspersed repeat expression. *Nucleic Acids Res* 47, doi:ARTN e2710.1093/nar/gky1301
78. Huchon D et al (2002) Rodent phylogeny and a timescale for the evolution of Glires: evidence from an extensive taxon sampling using three nuclear genes. *Mol Biol Evol* 19:1053–1065. 10.1093/oxfordjournals.molbev.a004164

79. Yu HW et al (2022) Dynamic reprogramming of H3K9me3 at hominoid-specific retrotransposons during human preimplantation development. *Cell Stem Cell* 29:1031–. 10.1016/j.stem.2022.06.006
80. Bailly-Bechet M, Haudry A, Lerat E (2014) One code to find them all: a perl tool to conveniently parse RepeatMasker output files. *Mobile DNA-Uk* 5, doi:Artn 1310.1186/1759-8753-5-13
81. Li WZ, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22:1658–1659. 10.1093/bioinformatics/btl158
82. Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30:3059–3066. 10.1093/nar/gkf436
83. Liu XY, Deng YD, Ni YF, Li ZH (2015) FastTree: A Hardware KD-Tree Construction Acceleration Engine for Real-Time Ray Tracing. *Des Aut Test Europe*, 1595–1598
84. Letunic I, Bork P (2021) Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* 49:W293–W296. 10.1093/nar/gkab301
85. Li TD et al (2017) Rat embryonic stem cells produce fertile offspring through tetraploid complementation. *Proc Natl Acad Sci U S A* 114:11974–11979. 10.1073/pnas.1708710114
86. Oikawa M et al (2022) Functional primordial germ cell-like cells from pluripotent stem cells in rats. *Science* 376:176–179. 10.1126/science.abl4412
87. Iwatsuki K et al (2023) Rat post-implantation epiblast-derived pluripotent stem cells produce functional germ cells. *Cell Rep Methods* 3:100542. 10.1016/j.crmeth.2023.100542
88. Werner RJ et al (2017) Sex chromosomes drive gene expression and regulatory dimorphisms in mouse embryonic stem cells. *Biol Sex Differ* 8:28. 10.1186/s13293-017-0150-x

Figures

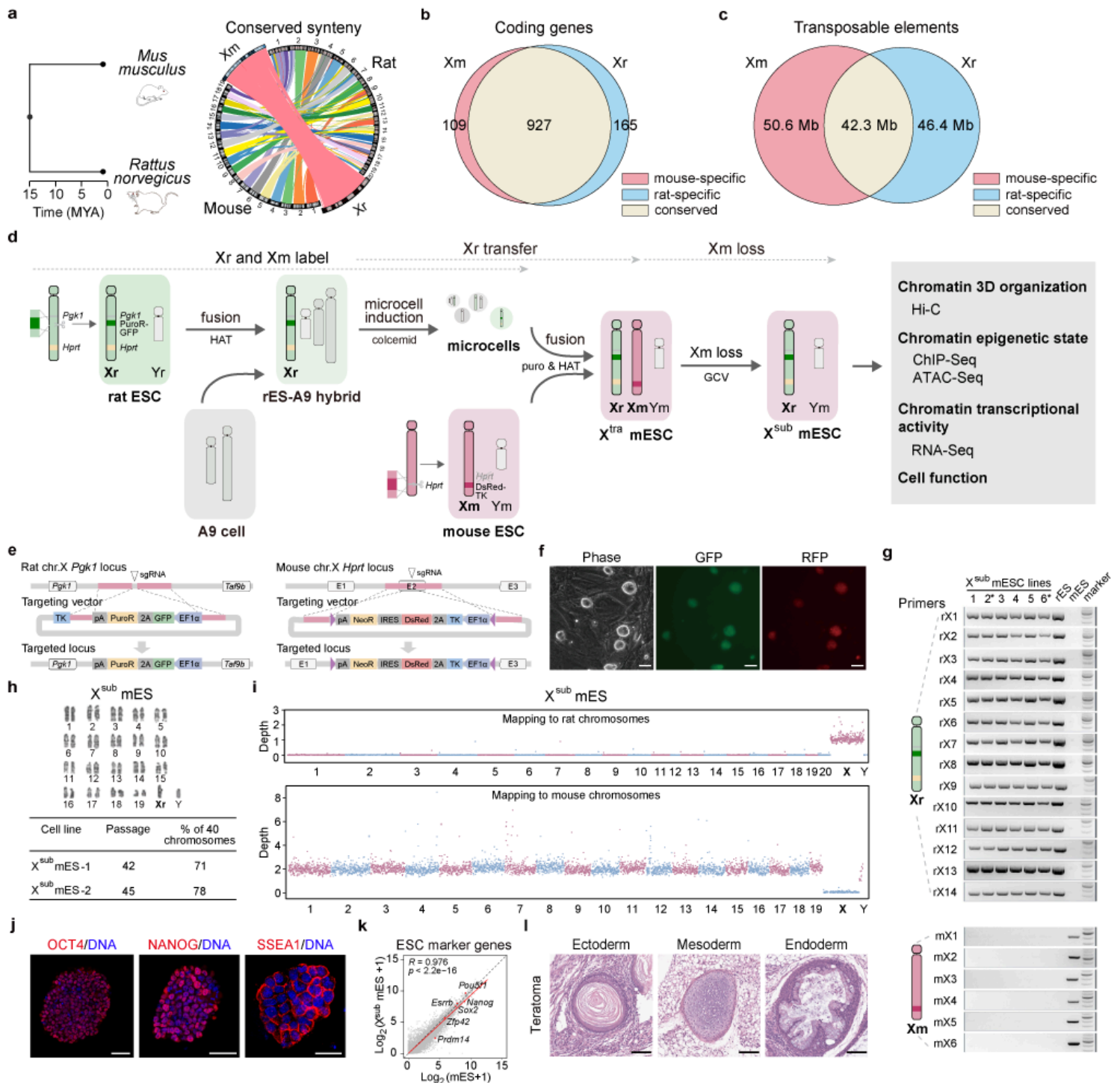


Figure 1

Overview of the CROSS platform for intact chromosome transfer and substitution in mESCs. **a**, evolutionary distance (*Left*) and conserved synteny (*Right*) between rat and mouse genomes. Orthologous chromosomal segments are shown based on mouse chromosomes. Xm, mouse X chromosome (Chr.X); Xr, rat Chr.X. **b**, **c**, Overlap of protein-coding genes (**b**) and transposable elements (**c**) between Xm and Xr. **d**, Schematic of rat Chr.X replacement in mouse ESCs. Engineered rat Chr.X (with PuroR-EGFP) was transferred via microcells into mESCs bearing edited mouse Chr.X (with TK-DsRed-NeoR). Puro and HAT selection first isolated X-transferred mESCs (X^{tra} mESCs), then GCV selection

isolated X-substituted mESCs (X^{sub} mESCs) losing mouse Chr.X. **e**, Design of CRISPR/Cas9-mediated engineering on rat (*left*) and mouse (*right*) Chr.X. **f**, Morphology of X^{sub} mESCs in bright-field, GFP, and RFP channels. Scale bar, 50 μm . **g**, Genotyping validation of X^{sub} mESCs using primers (listed in Supplementary Table 1 and 2) specific across rat and mouse Chr.X. **h**, Representative karyotype of X^{sub} mESCs showing retention of the rat Chr.X and the absence of mouse Chr.X. **i**, Whole-genome sequencing analysis of X^{sub} mESCs ($n = 2$). **j**, Immunofluorescence of ESC marker proteins in X^{sub} mESCs. Scale bar, 70 μm . **k**, Scatter plot comparing autosomal gene expression between mESCs ($n = 2$) and X^{sub} mESCs ($n=4$). ESC marker genes were highlighted in red. Linear regression line was shown. P values were calculated by two-sided one-sample t -test. **l**, Teratomas sections (day 30) derived from X^{sub} mESCs. Scale bar, 100 μm .

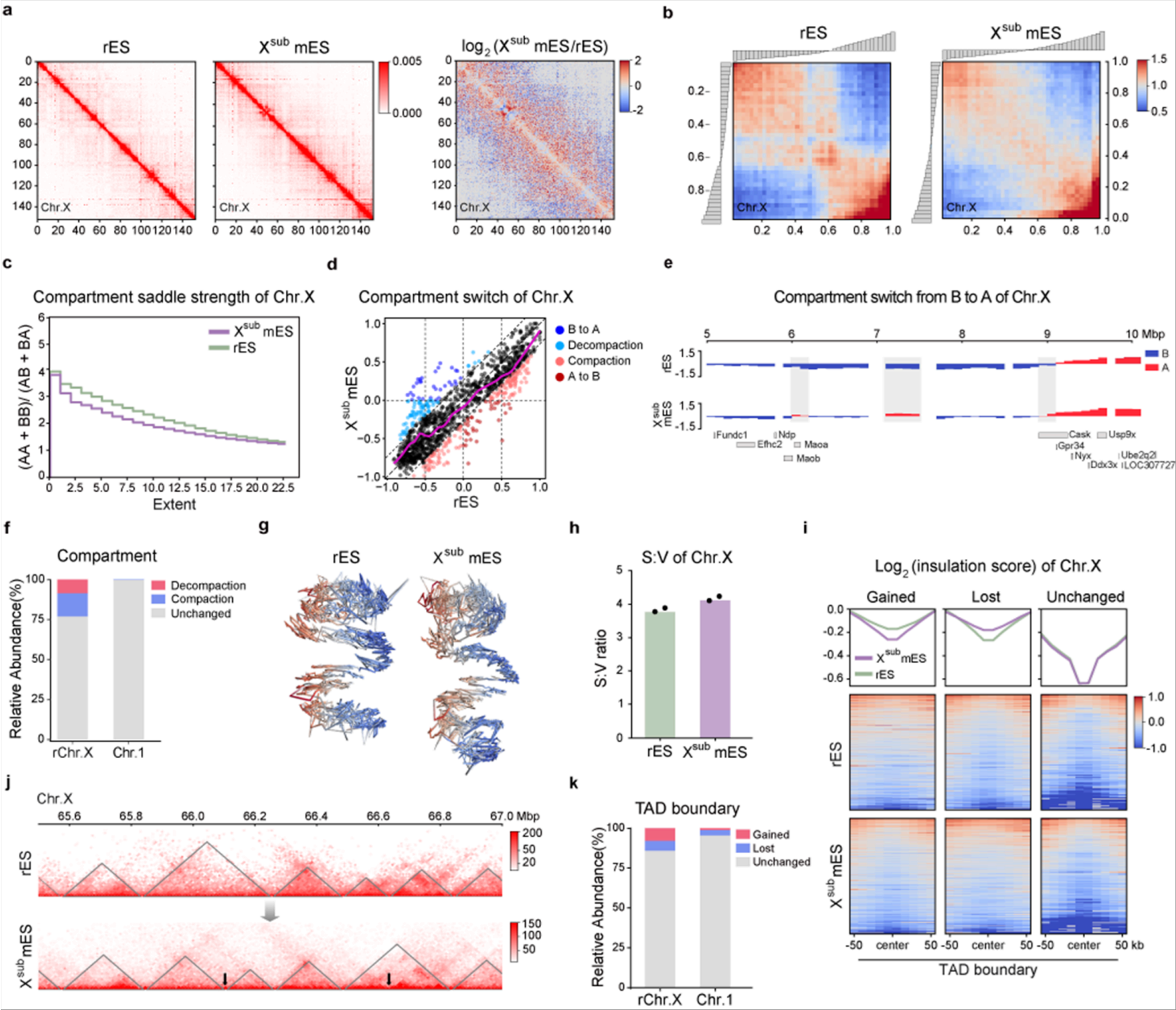


Figure 2

3D architecture of rat X chromosome in the X^{sub} mESCs. **a**, Hi-C contact matrices of the Chr.X in rESCs (*left*, 2 biological replicates), X^{sub} mESCs (*middle*, 2 biological replicates), and $\log_2(X^{\text{sub}} \text{ mESC/rESC})$ (*right*) at 100-kb resolution. The intensity of each pixel reflects the normalized contact frequency between pairs of loci. **b**, Saddle plots of Chr.X compartmentalization strength, calculated as the ratio of observed to expected contact frequencies between A and B compartments (i.e., $(AA+BB)/(AB+BA)$). Regions are grouped by their eigenvector value (active, high; inactive, low). Red indicates enrichment (A-A, B-B interactions), and blue indicates depletion (A-B interactions) of contacts. The bar charts on the left and top indicate the distribution of signal intensity. **c**, compartment saddle strength curve of Chr.X, based on (**b**). **d**, Compartment changes in the Chr.X of X^{sub} mESCs relative to rESCs. The dashed line marks regions unchanged; the pink curve represented the fitted trend line. **e**, Representative genome browser tracks showing compartment switch of Chr.X in X^{sub} mESCs V.S. rESCs. **f**, Compartment changes in rat Chr.X (compared to Chr.X of rESCs) and Chr.1 (compared to Chr.1 of mESCs) in X^{sub} mESCs. **g**, 3D structural models of the Chr.X. Blue and pink denote compartments B and A, respectively. **h**, Surface Area: Volume (S: V) ratio of Chr.X in rESCs and X^{sub} mESCs. **i**, Alterations in topologically associating domains (TADs) of the X chromosome in X^{sub} mESCs. Insulation score of TAD boundaries, including gained, lost, unchanged in X^{sub} mESCs compared to rESCs. **j**, A genome browser screenshot shows altered TADs across 65.5-67 Mb on Chr.X. **k**, Changes in TAD boundaries of the rat Chr.X (compared to Chr.X of rESCs) and Chr.1 (compared to Chr.1 of mESCs) in X^{sub} mESCs. rES, male rat ESCs; mES, male mouse ESCs

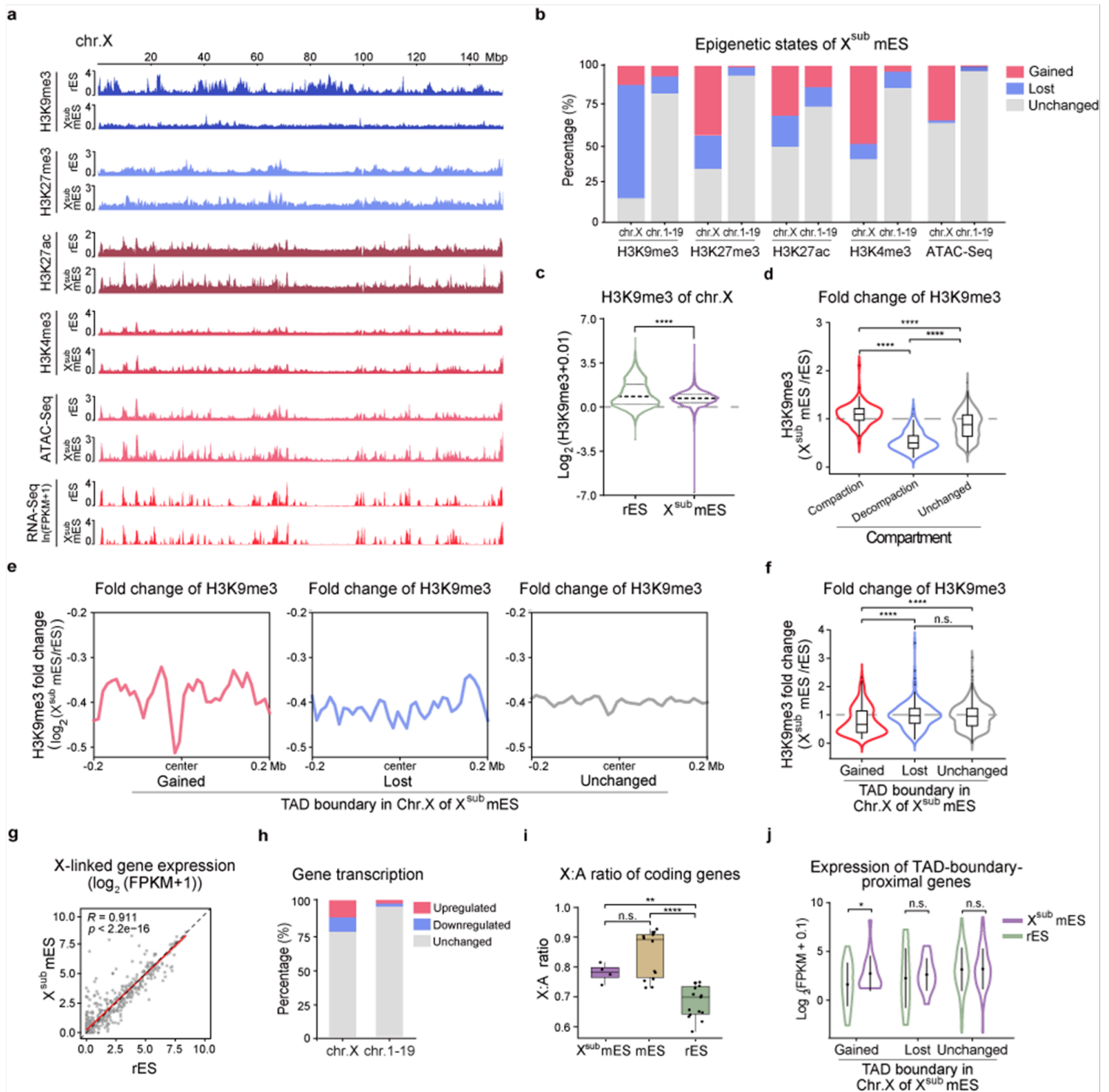


Figure 3

Multi-omic characterization of localized epigenetic and structural divergence on the substituted chromosome. **a**, Genome browser tracks showing enrichment profiles of histone marks, accessibility (ATAC-seq), and transcription (RNA-seq) on Chr.X in rat ESCs and X^{sub} mESCs. **b**, Dynamic of related signals on the rat Chr.X (compared to Chr.X of rESCs) and autosomes (compared to autosomes of mESCs) in X^{sub} mESCs. **c**, Violin plots comparing the H3K9me3 signal intensities across the entire Chr.X in rat ESCs and X^{sub} mESCs ($n = 2$ biologically independent samples per group). Dashed line indicates

the median value. **d**, Correlation between H3K9me3 fold change and chromatin compartment changes of the Chr.X in X^{sub} mESCs. **e**, H3K9me3 fold change in regions near gained, lost, unchanged TAD boundaries in Chr.X of X^{sub} mESCs, respectively. **f**, Correlation between H3K9me3 fold change and TAD boundary changes of the Chr.X of X^{sub} mESCs. **g**, Pairwise correlation of Chr.X-linked gene expression across X^{sub} mESCs (n=4) vs. rat ESCs (n=2). **h**, Statistical analysis of gene expression changes on the Chr.X (compared to rat ESCs) and autosomes (compared to mESCs) in X^{sub} mESCs, respectively. Pink, blue, gray indicated up-regulated, down-regulated and unchanged genes, respectively. **i**, Box plot showing the X-to-autosome (X: A) expression ratio of coding genes in X^{sub} mESCs (n=4), male mouse ESCs (n=13), and male rat ESCs (n=15). Data for X^{sub} mESCs were from this study, while mouse and rat ESCs were combined from this study and public datasets (GSE44150, GSE97966, GSE178701, GSE220805, and GSE90516). *P* values were calculated using Wilcoxon rank-sum test. **j**, Correlation between TAD boundary changes and nearby gene expression on Chr.X of X^{sub} mESCs. 2 biological replicates of each group for **a-j**, except **g** and **i**. Statistical significance was assessed using an unpaired two-sided *t*-test.

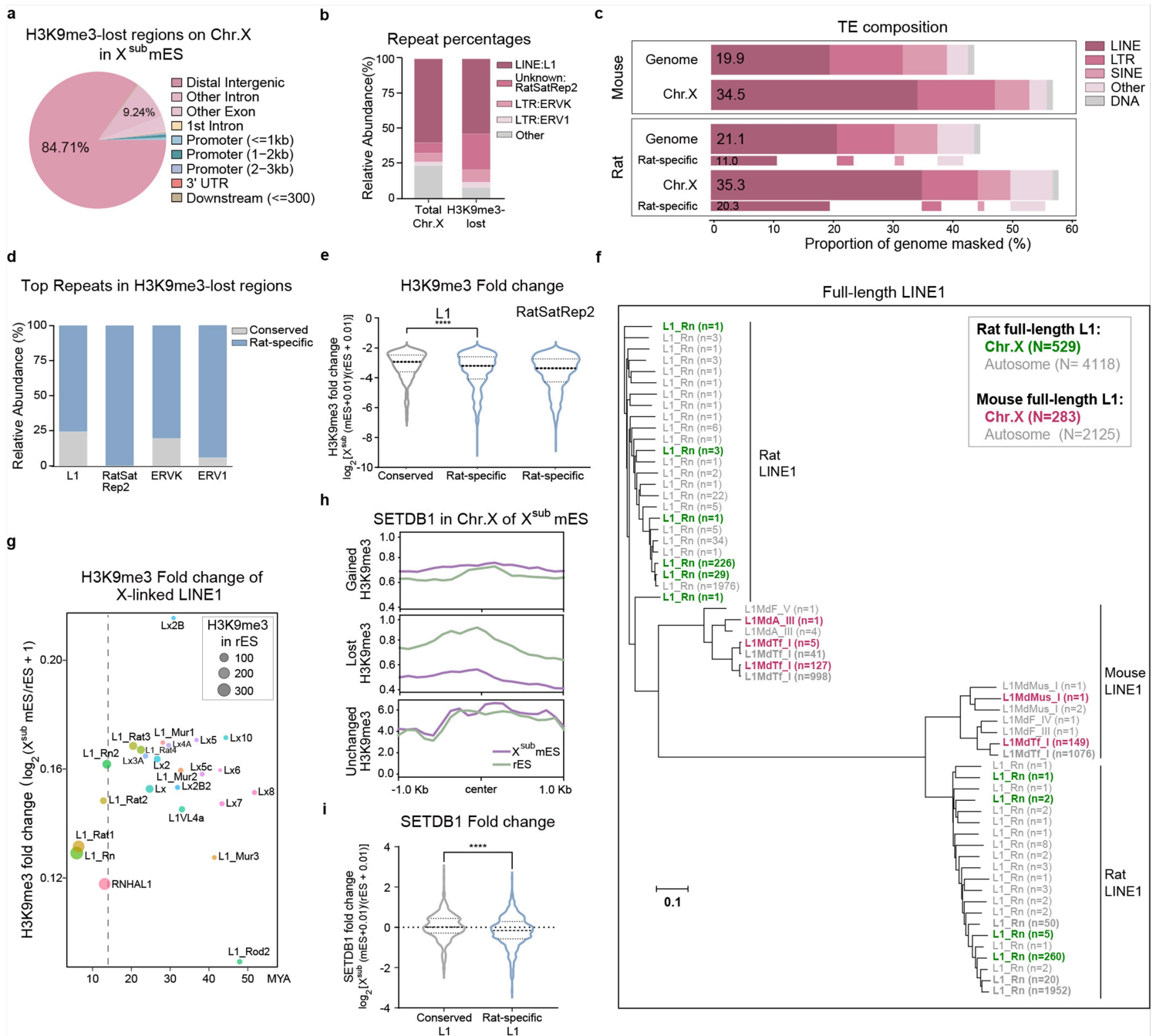


Figure 4

Molecular basis of heterochromatin erosion linked to defective SETDB1 recruitment at species-specific repeats. **a**, Pie chart showing the genomic distribution of H3K9me3-lost regions on the Chr.X in X^{sub} mESCs. **b**, Bar chart showing the proportion of repeat types within H3K9me3-lost regions on the Chr.X in X^{sub} mESCs and within all repeats across Chr.X. LTR, long terminal repeat; SINE, short interspersed nuclear element; LINE, long interspersed nuclear element; DNA, DNA transposon; Others, other types of repeats. **c**, The comparative analysis of transposable elements (TEs) composition in the whole genome and the Chr.X of mouse and rat. The proportion of rat-specific TE was highlighted. **d**, Conservation analysis of top repeat types suffering from H3K9me3 loss on the Chr.X. **e**, Fold change of H3K9me3 peak intensity (X^{sub} mESCs vs rat ESCs) between conserved and rat-specific LINE1s (L1s), and rat-

specific RatSatRep2 on the Chr.X. **f**, Phylogenetic analysis of full-length L1s in rat and mouse. Colors denoted genomic origin: grey, mouse or rat autosomes; pink, mouse Chr.X; green, rat Chr.X. Branch labels indicated L1 subfamily and copy number (n). Scale bar, 0.1 substitutions per site. Summary counts of full-length L1s per genomic compartment were shown. **g**, Fold change of H3K9me3 on X-linked L1 subfamilies, plotted against their evolutionary age (million year ago, MYA). Dashed line marked rat-mouse divergence. Point sizes indicated basal H3K9me3 level in rat ESCs. **h**, SETDB1 binding at sites of gained, lost, and unchanged H3K9me3 on Chr.X of X^{sub} mESCs (n = 2) and rESCs (n = 2), respectively. **i**, Fold change of SETDB1 (X^{sub} mESCs vs rat ESCs) between conserved and rat-specific L1s on the H3K9me3-lost region of Chr.X. Statistical significance was assessed by unpaired two-sided *t*-test.

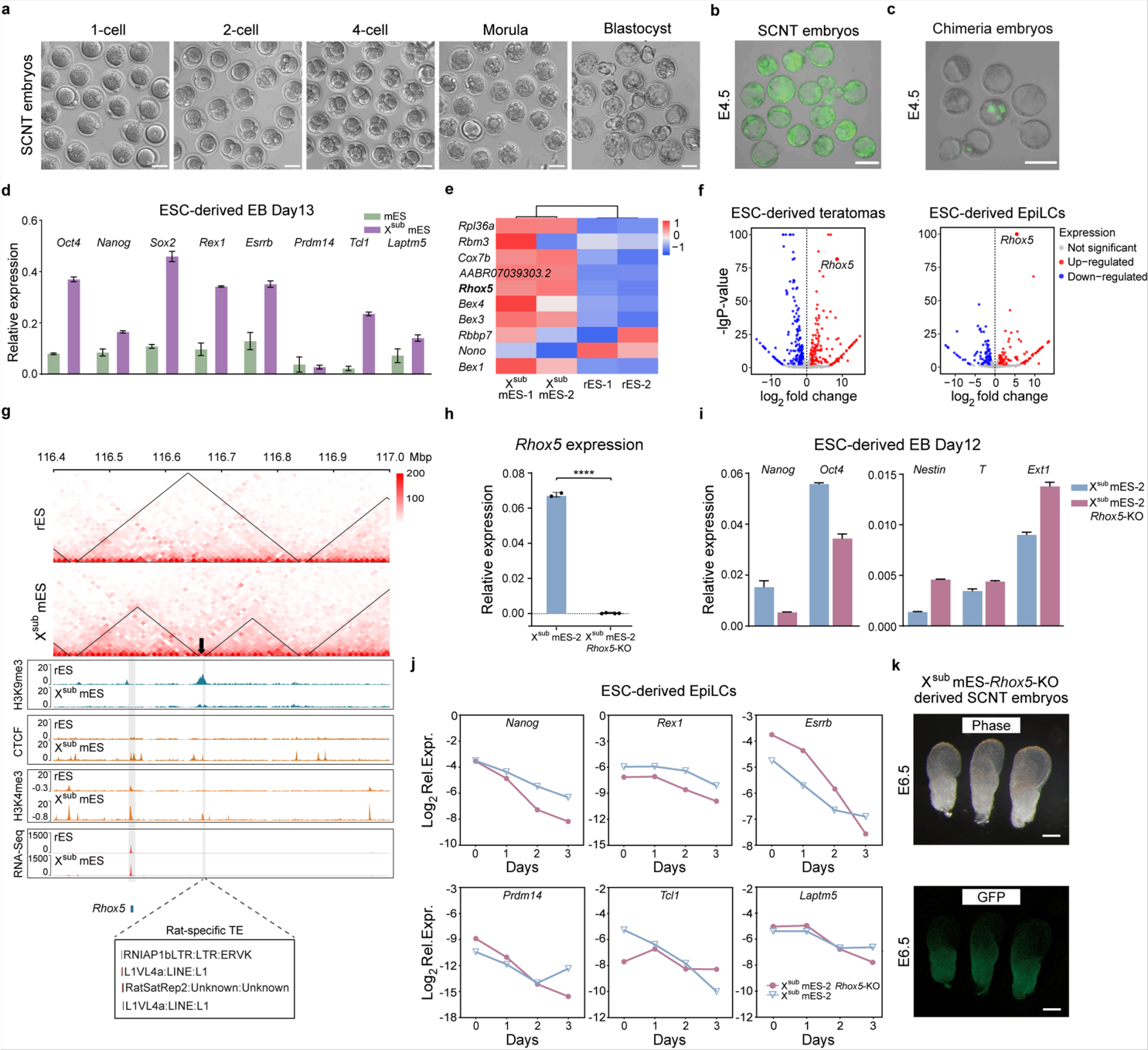


Figure 5

CROSS-mediated discovery of functionally active TAD boundaries at the *Rhox5* locus linked to cellular differentiation. **a**, SCNT embryos of X^{sub} mESCs across several developmental stages. Scale bar, 100 μm . **b**, SCNT embryos derived from X^{sub} mESCs at E4.5 (embryonic day 4.5). Scale bar, 70 μm . **c**, X^{sub} -mESC (GFP⁺ cells)-containing chimeric embryos at E4.5 after diploid blastocyst injection. Scale bar, 100 μm . **d**, Expression of ESC maker genes in mESC- and X^{sub} mESC-derived EBs. **e**, Heatmap showing expression of the top 10 most highly expressed genes near those de novo TAD boundaries on the Chr.X in X^{sub} mESCs. **f**, Volcano plots comparing X-linked gene expression. The *left* and *right* panels show differential expression in EpiLCs and teratomas, respectively, derived from mESCs versus X^{sub} mESCs. $n = 2$ for each group and P value calculated by *Wald* test. The *Rhox5* was labeled. **g**, Changes in TAD, epigenetic marks and transcription around the *Rhox5* locus of X^{sub} mESCs, alongside gene and rat-specific TE annotations. Two biological replicates used for each group, except in the RNA-seq where X^{sub} mESCs had four replicates. **h**, Comparing *Rhox5* mRNA expression between X^{sub} mESCs and *Rhox5*-knockout X^{sub} mESCs in RT-qPCR assay. **i**, Expression of related genes in X^{sub} -mESC-derived EBs and *Rhox5*-knockout- X^{sub} -mESC-derived EBs. **j**, Temporal expression dynamics of naïve pluripotency genes in EpiLCs derived from X^{sub} mESCs and *Rhox5*-knockout X^{sub} mESCs, respectively. **k**, Developmental analysis of SCNT embryos derived from *Rhox5*-knockout X^{sub} mESCs at E6.5. Scale bar, 100 μm . EB, embryoid body; SCNT, somatic cell nuclear transfer; EpiLC, epiblast-like cells.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementarytables.pdf](#)
- [Extendeddatafigures.pdf](#)