Endogenous retroviruses drive species-specific germline transcriptomes in mammals

Akihiko Sakashita, So Maezawa, Kazuki Takahashi, Kris G. Alavattam, Masashi Yukawa, Yueh-Chiang Hu, Shohei Kojima, Nicholas F. Parrish, Artem Barski, Mihaela Pavlicev and Satoshi H. Namekawa

Gene regulation in the germline ensures the production of high-quality gametes, long-term maintenance of the species and speciation. Male germline transcriptomes undergo dynamic changes after the mitosis-to-meiosis transition and have been subject to evolutionary divergence among mammals. However, the mechanisms underlying germline regulatory divergence remain undetermined. Here, we show that endogenous retroviruses (ERVs) influence species-specific germline transcriptomes. After the mitosis-to-meiosis transition in male mice, specific ERVs function as active enhancers to drive germline genes, including a mouse-specific gene set, and bear binding motifs for critical regulators of spermatogenesis, such as A-MYB. This raises the possibility that a genome-wide transposition of ERVs rewired germline gene expression in a species-specific manner. Of note, independently evolved ERVs are associated with the expression of human-specific germline genes, demonstrating the prevalence of ERV-driven mechanisms in mammals. Together, we propose that ERVs fine-tune species-specific transcriptomes in the mammalian germline.

The testis has the most diverse, complex and rapidly evolving transcriptome of all the organs in mammals. Furthermore, the testis expresses the largest number of transcription factors (TFs) of all mammalian organs. These qualities are due, in part, to specific and dynamic bursts in the expression of thousands of germ-line genes after the mitosis-to-meiosis transition. This transition occurs when germ cells have completed mitotic proliferation and have entered into meiosis, an essential process in the preparation of haploid gametes. Notably, a wide variety of species-specific transcripts have been identified in the later stages of spermatogenesis, giving rise to morphologically and functionally diverse gametes in mammals. However, the mechanisms that enable the rapid evolution of species-specific germline transcriptomes remain to be determined.

In this study, we identify a mechanism that underlies germline regulatory divergence. We report that many rapidly evolved cis-regulatory elements—in particular, active enhancers—are derived from certain types of endogenous retrovirus (ERV). ERVs are the remnants of retroviruses that have integrated into the germ-line genome. Transposable elements (TEs) with long terminal repeats (LTRs), a feature shared by ERVs and exogenous retroviruses, constitute ~10% of mammalian genomes. Other classes of TEs, which together account for 40–50% of a given mammalian genome, include other retrotransposons such as long and short interspersed nuclear elements (LINEs and SINEs), as well as DNA transposons.

TEs have long been considered genetic threats because transposition can be deleterious by, for example, disrupting the exons of protein-coding genes. On the other hand, the geneticist Barbara McClintock, the discoverer of TEs, proposed in 1950 that TEs function as gene regulatory elements. Studies in the past decade, long after McClintock’s proposal, have indeed established that TEs can impact host genomes by introducing gene regulatory elements, including promoters and enhancers. Many interspersed ERVs have lost the information necessary to encode the proteins that support autonomous transposition (for example, pol), however, their LTRs retain the ability to recruit TFs and regulate gene expression in host genomes.

Endogenous retroviruses (ERVs) are the remnants of retroviruses that have integrated into the germ-line genome. Transposable elements (TEs) with long terminal repeats (LTRs), a feature shared by ERVs and exogenous retroviruses, constitute ~10% of mammalian genomes. Other classes of TEs, which together account for 40–50% of a given mammalian genome, include other retrotransposons such as long and short interspersed nuclear elements (LINEs and SINEs), as well as DNA transposons.

In the germline, in which mutations due to transposition are potentially heritable, TE mobility is tightly controlled. The germline draws on several TE-suppression mechanisms, including DNA methylation, H3K9 methylation and PIWI-interacting RNA (piRNA). Yet, despite these silencing mechanisms, recent studies have revealed regulatory functions for TEs in male meiosis, including post-transcriptional regulation of mRNA and long noncoding RNAs (lncRNAs) via the piRNA pathway, and promoter functions that drive the expression of lncRNAs. However, at the mitosis-to-meiosis transition, when dynamic reorganization of three-dimensional (3D) chromatin and the epigenome takes place, cis-regulatory functions for TEs remain undetermined.

Here, we use an unbiased, genome-wide approach to identify ERVs that are within accessible chromatin and expressed after the mitosis-to-meiosis transition. We show that ERVs function as species-specific enhancers in the germline. These enhancers drive the expression of evolutionarily novel germline genes after the...
mitosis-to-meiosis transition, thereby defining the species specificity of germline transcriptomes in mammals. We also demonstrate the prevalence of ERV-driven germline genes in humans, and we propose a model whereby ERVs fine-tune species-specific transcriptomes in mammalian germlines.

**Results**

**Dynamic expression of repetitive elements during mouse spermatogenesis.** To understand the dynamics of repetitive element expression in spermatogenesis, we analyzed the transcriptomes of four representative stages of spermatogenesis: THY1+ undifferentiated spermatogonia (which contain spermatogonial stem cells and progenitor cells), KIT+ differentiating spermatogonia, pachytene spermatocytes (PSs) in the midst of meiosis and postmeiotic round spermatids (RSs)26,27 (Fig. 1a). To define regions of interest, we used a RepeatMasker annotation, a unique genomic annotation for interspersed repetitive loci, that specifies the best-matched class of repetitive elements for a given locus, and which does not have redundant annotation (Methods and Fig. 1b). In this way, we filtered to 1,755,061 ‘high confidence’ loci (Fig. 1c). Applying this ‘best match’ TE annotation set to our RNA-seq processing pipeline (Extended Data Fig. 1a), we detected the expression of individual TE copies in the four representative stages of spermatogenesis (Extended Data Fig. 1b). Unambiguously expressed TE loci make up a small fraction (less than 3%) of all copies of a given class in the genome (Extended Data Fig. 1b). Yet, notably, the majority of detected TEs were differentially expressed during each transition of spermatogenesis (Fig. 1d and Supplementary Data 1); in particular, 89.0% (18,552/20,853) of expressed TEs were differentially expressed at the KIT+ spermatogonia-to-PS transition (the mitosis-to-meiosis transition). LINE, SINE and LTR TEs comprised the major classes of differentially expressed TEs (Fig. 1e). Next, we sought to examine the relationships between stage-to-stage changes in TE expression and stage-to-stage changes in TE-adjacent gene expression. TE expression changes did not correlate with gene expression changes in the THY1+–KIT+ transition (Fig. 1f). However, when we analyzed the KIT+–PS transition, we noted a positive correlation between TE expression changes and changes in adjacent gene expression, and the same was true for the PS–RS transition (Fig. 1f). Next, we examined the distance between the TEs and the transcription start sites of their adjacent genes. Our analyses of the KIT+–PS transition revealed that, even when separated by 50–100 kb, TE and adjacent gene expression levels change together (Fig. 1g). This observation raised the possibility that gene transcription in the mitosis-to-meiosis transition is influenced by some portion of TEs in a long-range manner, leading us to interrogate the functions of TEs as enhancers.

A subset of ERVs have enhancer-like features in late spermatogenesis. Among the major classes of expressed TEs (LINE, SINE and ERV LTR), ERV LTRs bear TF-binding sites and are known to function as gene regulatory elements in other settings14,15. Therefore, we suspected that ERVs function as gene regulatory elements after the mitosis-to-meiosis transition. Open, accessible chromatin is a prominent feature of functioning gene regulatory elements. Accordingly, to determine the sites of accessible chromatin in PSs, we analyzed previously published ATAC-seq ( assay for transposase-accessible chromatin using sequencing) data15. Although we found that most ERV loci evince closed chromatin genome-wide (Extended Data Fig. 2a), we found that numerous types of ERV were significantly enriched in the open, accessible chromatin in PSs (Fig. 2a). Interestingly, the majority of ERVs in accessible chromatin come from the ERVK family, one of the three major families that comprise ERVs (ERV1 family, 14 types; ERVK family, 39 types; ERVL family, six types; Fig. 2a,b). In analyzing multiple stages of spermatogenesis, we noticed that several types of accessible ERV were specific to PSs or both PSs and RSs (Fig. 2b), suggesting such ERVs possess specific functions in meiosis and subsequent stages of spermatogenesis.

Given that ERVs are interspersed throughout the genome, we hypothesized that ERVs function as enhancers that drive the expression of spermatogenesis-specific genes. To test this hypothesis, we analyzed the chromatin immunoprecipitation sequencing (ChIP-seq) signal enrichment for H3K27 acetylation (H3K27ac), a marker of active enhancers, in PSs29,30. Through the thresholding of H3K27ac enrichment and accessible chromatin at individual ERV loci (see Methods), we defined a category of ERVs said to be ‘enhancer-like’ in PSs (ERV1, 116 enhancer-like loci; ERVK, 970 enhancer-like loci; ERVL, 56 enhancer-like loci; Fig. 2c and Supplementary Data 2). Two major ERVK subfamilies, RMER17 (445 loci) and RLTR10 (249 loci), were highly represented among ERVK loci bearing significantly enriched H3K27ac and accessible chromatin (Fig. 2d). Notably, H3K27ac was highly enriched on RLTR10 in comparison to RMER17 (Fig. 2e). Curiously, RLTR10C, a type of RLTR10, was frequently adjacent to MMERVK10C, which has full viral elements flanked by two RLTR10C loci14 and is supposed to be transcribed19.1 in the germline23. However, the overlap between enhancer-like RLTR10C and MMERVK10C is largely coincidental (Extended Data Fig. 2b), suggesting that enhancer-like RLTR10C is a solo LTR that has lost flanking viral elements.

Intriguingly, in PSs and RSs, we noted that the establishment of H3K27ac and open chromatin at autosomal RTRLR10 loci was associated with the transcriptional upregulation of adjacent genes (Fig. 2f). Average tag density analyses revealed significant H3K27ac enrichment within enhancer-like ERVs in PSs and RSs (Fig. 2g). In support of its putative gene regulatory status, low levels of RNA-seq signal were detected at enhancer-like ERVs (Fig. 2g). Enhancer-like ERVs were also enriched for the active mark H3K4me3 in PSs and RSs (Fig. 2g). H3K4me3 peaks at enhancer-like ERVs were located far from promoters (≥10 kb; Extended Data Fig. 3). Consistent with this, the majority of enhancer-like ERV-adjacent genes are located far from promoters (≥5–500 kb away from enhancer-like ERVs (Fig. 2h). Such H3K4me3 localization patterns, together with the low levels of mRNA transcription, comprise a known feature of tissue-specific enhancers16. As a control, we noted that the repressive mark H3K27me3 did not accumulate on enhancer-like ERVs (Fig. 2g). A previous study suggested that ERVs function as enhancers in placenta and testes24. Our results corroborate this notion: specific subsets of ERVs gain the features of active enhancers in late spermatogenesis.

During male meiosis, the sex chromosomes undergo a tightly coordinated process of transcriptional inactivation known as ‘meiotic sex chromosomes inactivation’ (MSCI); perhaps counterintuitively, it is in this context that the accessibility of sex chromosome-associated chromatin increases25 and many active enhancers are established26. A representative track view demonstrates that, on the PS X chromosome, the establishment of H3K27ac and open chromatin at RTRLR10 loci in PSs correlates with activation of transcripts that escape postmeiotic silencing in RSs (Extended Data Fig. 4a). Of note, enhancer-like ERVs were enriched on the X chromosome (Extended Data Fig. 4b,c), although H3K27ac enrichment at enhancer-like ERV loci is comparable between the sex chromosomes and autosomes (Extended Data Fig. 4d), and enhancer-like ERVs on the sex chromosomes are preferentially located in intergenic regions (Extended Data Fig. 4e). The establishment of H3K27ac on the silent X chromosome in meiosis and subsequent escape gene activation in RSs is regulated by RNFL8, a DNA damage response factor26. Therefore, on chromosome X, enhancer-like ERVs are regulated downstream of RNFL8.

To further define the functions of ERVs as enhancers, we tested the hypothesis that genes adjacent to enhancer-like ERVs evince preferential expression relative to non-adjacent genes after the mitosis-to-meiosis transition. To this end, we identified 1,452 genes that are adjacent to enhancer-like ERVs in PSs. Importantly, these
1,452 genes were highly expressed in PSs in comparison to other stages analyzed in this study: THY1+, undifferentiated spermatogonia; KIT+, differentiating spermatogonia; PS, pachytene spermatocytes; RS, round spermatids. Distances between TEs and the transcription start sites of their adjacent genes are shown. *P < 0.05; **P < 0.001; NS, not significant; Mann–Whitney U-tests. Central bars represent medians, the boxes encompass 50% of the data points, and the whiskers indicate 90% of the data points. Data for c–g are available as source data.

ERVs are known to carry binding sites for TFs and, therefore, bear the potential to rewire transcriptomes via transposition. To determine the TF-binding sites present in enhancer-like ERVs, we performed motif analyses. In enhancer-like RLTR10 loci we identified TF motifs such as binding sites for A-MYB (also known as MYBL1), a male germline-specific TF that drives spermatogenesis-related gene expression from meiotic prophase onward. In line with this finding, the consensus sequence of RLTR10B, which is listed in the Dfam database, contains two A-MYB binding motifs (Extended Data Fig. 4f). A-MYB-binding sites were not observed in RMER17, another major ERVK subfamily constituting enhancer-like ERVs, nor were they observed in ERV1 (Fig. 3d). However, A-MYB-binding sites were also detected.
Identification of enhancer-like ERVs in meiosis. **a**, Scatter plots depicting observed ERV copy numbers in regions of accessible chromatin (within ATAC peak regions; y axis) versus the expected prevalence of ERV loci throughout the mouse genome (x axis) in the following ERV families in PSS: ERV1, ERVK, and ERVL. Each dot represents a single type of ERV within a subfamily; red diamonds represent ERV types that exhibit significant enrichment in ERV copy numbers in regions of accessible chromatin (≥2-fold observed/expected enrichment; P < 0.05, binomial test; see Methods). **b**, Heatmaps depict log2(fold enrichment) of ERV copies in ATAC peak regions relative to genomic prevalence. ERV loci that are accessible in PSS are shown. MΦ, macrophage; ESC, embryonic stem cell; MEF, mouse embryonic fibroblast. **c**, Average tag density plots around enhancer-like ERVs. **d**, Pie chart indicating the relative abundances of enhancer-like ERVs. **e**, Relative H3K27ac enrichment at enhancer-like RMER17 and RLTR10 loci in PSS. **f**, Distance distribution of enhancer-like ERVs. **g**, Track views of an enhancer-like ERV locus. **h**, Bar chart depicts the regional distribution of genes adjacent to enhancer-like ERVs: proximal adjacency, ±5 kb; distal adjacency, up to ±1 Mb. Numbers of genes are shown above bars. Data for **a**, **b** and **d** are available as source data.
in a set of all enhancer-like ERVKs that excluded RLTR10B and RMER17 (other ERVKs); Fig. 3d). Importantly, A-MYB-binding sites were not detected in non-enhancer-like RLTR10, suggesting a specific function for A-MYB in the regulation of enhancers. In support of our motif analyses, A-MYB ChiP-seq peaks from whole testis tissue overlapped with enhancer-like ERV loci—specifically, RLTR10B loci—in intergenic regions, both on autosomes and the X chromosome (Fig. 2f and Extended Data Fig. 4a). Consistent with this, a recent study demonstrated that A-MYB binds to RLTR10B. In addition to A-MYB, we detected binding sites for other TFs. In evaluating motifs associated with (1) RLTR10, (2) RMER17, (3) other ERVKs and (4) ERV1, we detected binding sites for the following TFs: NFB, TBP, RFX4, RFX1, ZBTB7A, SOX5, GFI1, YY1 and PKNOX2 (Fig. 3d). Furthermore, the expression of these TFs was highly upregulated in PSs as compared to all other genes (5,461 preferentially expressed genes divided by all 22,661 RefSeq genes in the genome), this association (381 genes divided by 1,452 genes) is statistically significant (P = 0.0270, hypergeometric test). C, Bar chart depicting the statistical significance of GO terms for genes adjacent to enhancer-like ERVs. D, HOMER motif analyses for putative transcription-factor-binding sites in the following enhancer-like ERVs: RLTR10; RMER17; a set of all enhancer-like ERVKs excluding RLTR10B and RMER17; ERV1 loci. E, Heatmap depicting the expression levels of representative transcription factors in stages of spermatogenesis. F, Model, where enhancer-like ERVs act as activators of germline genes. Data for a and e are available as source data.

A-MYB acts on ERV enhancers to activate adjacent germline genes. We sought to test the possibility that binding of A-MYB to enhancer-like ERVs enables activation of adjacent genes in late spermatogenesis. In support of this hypothesis, we observed a significant overlap between enhancer-like ERVs and A-MYB-binding sites throughout the genome (443/1,122, 39.5%; Fig. 4a). We analyzed previously published RNA-seq data from the testes of A-myb mutants (Myb<sup>1<sup>post<sup>37</sup></sup></sup>) at postnatal day 14 (P14)7 (Fig. 4b). Consistent with the reported role of A-MYB in the activation of late spermatogenesis genes, 1,705 genes were differentially expressed, and most of them were downregulated upon the loss of A-MYB (Fig. 4b). Importantly, we observed a significant overlap of ERV-adjacent genes and genes differentially expressed in A-myb mutants: 103 genes out of the set of 381 highly expressed ERV-adjacent, mitosis-to-meiosis genes—many of them found among the downregulated genes of A-myb mutants. Of note, A-MYB binds the central regions of enhancer-like ERV’s adjacent to the 103 genes that are differentially expressed in A-myb mutants (n = 134 loci; Fig. 4c), suggesting that A-MYB functions at enhancer-like ERV loci.

To determine whether A-MYB acts on enhancer-like ERVs to activate genes, we performed luciferase reporter assays in HEK293T cells to measure the activity of enhancer-like ERVs as regulatory elements under conditions where A-MYB is expressed. In these experiments, we tested the activity of two independent enhancer-like RLTR10B loci with forward and reverse orientations (Fig. 4d). Reverse orientations for both enhancer-like RLTR10B loci exhibited stronger activities (up to 566-fold), confirming the activity of RLTR10B as a gene regulatory element (Fig. 4d). Such a result indicates that A-MYB acts on RLTR10B to activate target genes.

To test the in vivo function of A-MYB in the activation of enhancer-like ERVs, we performed ultra-low-input native ChIP-seq for H3K27ac using small numbers of A-myb mutant sperm.
PSs—an experimental necessity because A-myb mutant PSs fail to complete meiosis and are thus available in limited quantities. Representative track views demonstrate that H3K27ac was significantly reduced at enhancer-like ERVs at an autosomal locus and at an X-chromosomal locus (Fig. 4g). We noted that, in A-myb mutant PSs, the establishment of H3K27ac was largely impaired at enhancer-like ERVs throughout the genome (Fig. 4h). Taken together, these data support a function for A-MYB in the establishment of enhancer-like ERVs.

**ERV enhancers function to activate adjacent germline genes.** To confirm the activation of germline genes adjacent to enhancer-like ERVs, we performed CRISPR activation (CRISPRa) experiments using embryonic stem cells (ESCs) in which meiotic enhancer-like ERVs and germline genes are not active. We generated doxycycline (Dox)-inducible CRISPRa ESCs (J1 ESCs harboring a Dox-inducible dCas9-VPR transgene; Extended Data Fig. 5a). Using the CRISPRa ESCs, we activated a representative enhancer-like RLTR10B locus adjacent to the Tdrd3 gene by introducing two guide RNAs (gRNAs) within a 1-kb region of the A-MYB-binding site (Fig. 5a).

Upon Dox induction and gRNA treatment, expression of Tdrd3 was induced (Fig. 5a); upon additional expression of A-MYB, Tdrd3 expression was enhanced (Fig. 5a). Based on the functional validation of an individual RLTR10B locus, we sought to understand the functions of multiple RLTR10B loci via CRISPRa of the RLTR10B2 consensus sequence, which shares high homology with other RLTR10B subtypes. We therefore transduced the cells with a lentiviral construct containing five gRNAs that target the consensus sequence of RLTR10B (Fig. 5b). In the Dox+;A-MYB+ model, we observed a significant increase in cell death (Fig. 5c). Principal component analysis (PCA) of our RNA-seq samples confirmed that the Dox+;A-MYB+ model deviated from global gene expression profiles derived from the control (Dox−) model compared to conditions with CRISPRa or A-MYB expression only (Fig. 5d). In accordance with our PCA data, RNA-seq analysis revealed ectopic gene expression on both local and global scales. Zscan2, an ERV enhancer-adjacent gene, was activated upon induction of CRISPRa and A-MYB expression (Fig. 5c). Genome-wide, we noted significant upregulation of genes adjacent to enhancer-like ERVs (Fig. 5f), particularly ERV–adjacent differentially expressed genes observed in A-MYB mutants (Fig. 5f). Further, simultaneous induction of CRISPRa and A-MYB was essential for the activation of the Zscan2 locus.
expression exacerbates abnormal gene expression in comparison to CRISPRa or A-MYB-expression-only conditions (Fig. 5f and Supplementary Data 4).

To understand the functional significance of a representative enhancer-like ERV in an in vivo model for spermatogenesis, we performed CRISPR deletion for a representative enhancer-like ERV in mouse spermatogenesis. We generated a mouse line in which an enhancer-like RLTR10B upstream of the gene Zfy2, a Y chromosome-linked gene, was deleted. In this mouse model, we performed CRISPR deletion targeting a single RLTR10B locus (highlighted in red). Top: schematic for CRISPRa of a single enhancer-like ERV in mouse spermatogenesis. We generated a mouse line in which an enhancer-like RLTR10B upstream of the gene Zfy2, a Y chromosome-linked gene, was deleted. In this mouse model, we performed CRISPR deletion targeting a single RLTR10B locus (highlighted in red). Top: schematic for CRISPRa of a single RLTR10B locus (highlighted in red). Top: schematic for CRISPRa of a single RLTR10B locus (highlighted in red). Top: schematic for CRISPRa of a single RLTR10B locus (highlighted in red). Top: schematic for CRISPRa of a single RLTR10B locus (highlighted in red).
testis morphology was not affected (Extended Data Fig. 5c). This result is consistent with an independent study showing that deletion of Zfy2 is compatible with normal spermatogenesis. We analyzed P28 testes because, at this timepoint, spermatogenesis has progressed to the RS stage and Zfy2 is highly expressed. We conclude that RLTR10B can function as a bona fide enhancer that activates adjacent germline genes, and that A-MYB acts on RLTR10B to activate ERV enhancers. We hereafter refer to enhancer-like ERVs as ‘ERV enhancers’.

Rodent-specific ERV enhancers regulate species-specific gene expression. Meiotic spermatocytes and postmeiotic spermatids manifest high levels of transcriptomic diversity across mammalian species. Therefore, we reasoned that rodent-specific ERV enhancers may drive the expression of newly evolved genes, thereby conferring a species-specific form of transcriptomic diversity in late spermatogenesis. To test this possibility, we sought to determine the degree of sequence diversity of ERV-adjacent genes in mammals. Notably, a subset of ERV-adjacent genes found in mice do not have unambiguous homologs in other mammals that we examined—including another rodent, rat (48/381, 12.6%; Fig. 6a). Furthermore, many ERV-associated genes with homologs among mammals are poorly conserved, which raises the possibility of divergent functions in mouse (Fig. 6a). These results suggest that genes close to ERV enhancers are evolutionarily new in mice and/or rapidly evolved among mammals. Thus, ERV enhancers in mice are likely to regulate mouse-specific or evolutionarily diverged genes.

To determine the evolutionary traits of young ERV-adjacent genes, we examined the genomic distributions of ERV enhancers in mammalian species. Of the ERV enhancers in mice, specific types are found only in rodents, and one of these ERV-adjacent loci, RLTR10C, has no counterparts outside of mice (Fig. 6b). ERV enhancers with counterparts in rats displayed varied copy numbers (Fig. 6b). To test the conservation of ERV integration in rats and mice, we compared the genomic distributions of ERV enhancers and found that, for the most part, the genomic distributions and integration of their ERV enhancers differ (Extended Data Fig. 6a,b).

Subsets of ERVK and ERV1 are associated with meiotic gene expression in humans. To investigate species-specific functions of ERVs in other mammalian species, we analyzed human spermatogenesis. In particular, we sought to determine whether human-specific ERVs have enhancer-like features in spermatogenesis. To this end, we analyzed H3K27ac ChIP-seq data from human testes deposited in ENCODE. We found that MER57E3, a type of ERV1, and LTR5B, a type of ERVK, are enriched with H3K27ac and occupy a location adjacent to transcripts in human PSs (Fig. 7a). To evaluate the genome-wide features of ERVs in human testes, we examined the enrichment of H3K27ac on each type of ERV in the following ERV families: ERV1, ERVK and ERVL. We found a subset of human ERV types that are highly enriched with H3K27ac (>2-fold enrichment) of this subset, MER57E3 exhibited the highest levels of H3K27ac (Fig. 7b). Among 66 enhancer-like MER57E3 loci, 52 were found within the first introns of zinc finger (ZF) genes (Extended Data Fig. 7). These findings raise the possibility that a majority of MER57E3 enhancers were amplified as part of gene duplication events. Importantly, among 52 ZF genes, 47 contained Krüppel-associated box (KRAB) domains, enabling us to categorize these genes as KRAB-ZF genes. KRAB-ZF proteins bind ERVs and evolve to regulate host genomes, which draws an interesting co-evolutionary link between KRAB-ZF genes and ERVs. Motif analyses of human enhancer-like ERVs revealed that ERV1s and ERV5s contain binding sites for A-MYB (Fig. 7c). In each family, we further identified representative types and individual loci of enhancer-like ERVs (Fig. 7d and Supplementary Data 5). These results suggest that, in humans, in addition to ERV1s, ERV1s act as enhancers through A-MYB-dependent mechanisms. In support of this notion, we confirmed that A-MYB is highly expressed in both mouse and human spermatocytes through immunofluorescence analyses of testis sections (Extended Data Fig. 8).

Next, we sought to test the hypothesis that genes adjacent to H3K27ac-enriched ERV loci are associated with active genes after the mitosis-to-meiosis transition (that is, active in PSs compared to KIT+ spermatogonia) in humans. We found that, although genes adjacent to H3K27ac-enriched ERV1s did not manifest significant gene expression changes after the mitosis-to-meiosis transition (Fig. 7e), genes adjacent to MER57E3s were significantly activated in PSs (Fig. 7e). Notably, genes adjacent to H3K27ac-enriched ERV1s tended to be associated with genes activated after the mitosis-to-meiosis transition compared to other genes in the human genome, while genes adjacent to H3K27ac-enriched ERV1s did not show such an association (Fig. 7e). These results suggest that a subset of ERVKs and ERV1s act as enhancers to activate meiotic genes in humans.

Fig. 6 | Genes adjacent to rodent enhancer-like ERVKs are less conserved across species. a, Heatmap of sequence identity percentages for 381 mouse ERV-adjacent genes across six other species. Mouse-specific genes were significantly enriched in enhancer-like ERV-adjacent genes in comparison to a randomly picked background set of genes (see Methods); **P < 0.01, Fisher’s exact test. b, Phylogenetic tree and heatmap depicting the abundance of selected enhancer-like ERVK types across seven species. Data for a and b are available as source data.
Fig. 7 | Enhancer-like human ERVKs and ERV1s are associated with meiotic gene expression. a, Representative track view of H3K27ac ChIP-seq and RNA-seq signals in human testes and two spermatogenic cell populations: KIT+ and PS. Red and blue highlights indicate enhancer-like ERV1 and ERVK loci that overlap with H3K27ac deposition. b, Beeswarm plots of H3K27ac enrichment on each type of ERV in the following families: ERV1, ERVK and ERVL in human testes. Significantly enriched types of ERV element are defined as those with values ≥1 log2(observed/expected) (see Methods) and are highlighted in red circles. c, HOMER motif analyses of enhancer-like human ERV elements for putative TF-binding sites. d, Pie charts showing the numbers and representative types of enhancer-like ERV loci. e, RNA-seq analyses: cumulative distribution plots of log2(fold change) between KIT+ and PSs for the expression of genes adjacent to the enhancer-like elements ERV1, ERVK and ERVL, all with respect to other expressed genes (black). *P < 0.05; **P < 0.01; NS, not significant; Kolmogorov–Smirnov test. f, Heatmap of sequence identity percentages for 138 human enhancer-like ERV-adjacent genes across six other species. Human- or primate-specific genes were significantly enriched in ‘enhancer-like ERV adjacent’ genes in comparison to a randomly picked background set of genes (see Methods): **P < 0.001, Fisher’s exact test. g, Phylogenetic tree and a heatmap depicting the abundance of selected enhancer-like ERV copies in respective genomes. Data for b, f and g are available as source data.
Notably, a subset of ERV-adjacent genes in humans do not have unambiguous homologs in the other mammals that we examined and may thus be specific to humans and/or primates (61/138 genes, 44.2%; Fig. 7f). ERVs that are enhancer-like in humans are specific to the primate lineage (Fig. 7g), rather than being shared with other mammals. Together, our results support the concept that ERV-driven miotic enhancers are a general feature of mammals, and we propose that ERV enhancers represent a general mechanism for the divergence of transcriptomes during mammalian late spermatogenesis.

Discussion

We have identified a novel function for ERVs as species-specific enhancers in the germline—a function distinct from the reported functions of ERVLs as promoters that drive IncRNA expression in spermatogenesis. Curiously, over 15% of all oocyte transcripts start at LTR promoters that belong to the ERVL family; these ERVL promoters function during the oocyte-to-embryo transition. After fertilization, ERV enhancers are derepressed and expressed in preimplantation embryos, an essential event in early development. Together, our results further expand the repertoire of ERV functions, by showing that ERVs are also rapidly evolving enhancers in the germline.

For the most part, both the expression and chromatin state of TEs are reprogrammed at the mitosis-to-meiosis transition. Indeed, ERV enhancers also exhibit low levels of transcription in a meiosis-specific manner (Fig. 2g). One possible explanation is that enhancer RNA may be expressed at enhancer-like ERV loci. Although our analyses focus on the genes adjacent to ERV loci as targets of ERV enhancers, there are probably many more target genes because long-distance chromatin interactions were found throughout the genome in spermatogenesis. Thus, further investigation is warranted to identify the full repertoire of genes regulated by ERV enhancers.

We have also demonstrated that ERV enhancer activation is regulated by A-MYB. Curiously, the reverse orientation of two RLTR10B loci performed better in the luciferase assay compared to the forward orientation. Because the orientation of enhancer-like ERVs is randomly integrated with respect to adjacent genes, parsing how enhancer-like ERVs interact with target genes in the 3D chromatin environment is an important future undertaking. In humans, A-MYB-binding sites are found in ERVKs and ERVs. Therefore, we postulate that retrotransposition of ERVs provides new binding sites for key transcription factors, which, in turn, function as newly evolved cis regulatory elements for many genes. Importantly, we have also shown that A-MYB is associated with super-enhancers to drive the expression of key germline genes. Therefore, an A-MYB-dependent mechanism appears to lie at the heart of two distinct enhancer types: (1) super-enhancers, which drive robust activation of germline genes; (2) ERV-driven, rapidly evolving enhancers, which fine-tune the expression of species-specific germline genes. Together, these findings raise the important follow-up question of how A-MYB-binding sites on ERVs are protected from activation prior to the mitosis-to-meiosis transition. One intriguing possibility involves the function of KRAB-ZF proteins, a family of proteins that has co-evolved with ERVs to suppress ERV expression, the consequence of an evolutionary arms race between ERVs and the host genome.

Another important aspect of enhancer-like ERVs is species-specific gene regulation. In trophoblast stem cells, RLTR13DS, which comprises a mouse-specific ERVK type, has enhancer functions to establish a regulatory network specific to trophoblast stem cells, and the same study predicted the existence of enhancer-like ERVs in testes and ESCs. Curiously, the placenta is known to be a fast-evolving organ in which many ERVs have been co-opted. It is intriguing to speculate that ERVs are drivers of species-specific transcriptomes in rapidly evolving organs such as the testis and placenta, although mechanisms underlying intrinsic ERV activity in testes and placenta remain undetermined. Because ERV-based molecular mechanisms expose nuclei to risks of transposition and mutagenesis, their presence and, indeed, apparent importance in germline development are highly enigmatic. If KRAB-ZF proteins are involved in the control of such mechanisms, then it will be crucial to determine the crosstalk between KRAB-ZF proteins and other means of epigenetic silencing, such as DNA methylation and the piRNA pathway, to understand the precise control of both TE silencing and vital TE activities in the germline.

Online content

Any methods, additional references, Nature Research 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/s41594-020-0487-4.

Received: 7 May 2019; Accepted: 10 July 2020;
Published online: 7 September 2020

References

1. Ramskold, D., Wang, E. T., Burge, C. B. & Sandberg, R. An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. PLoS Comput. Biol. 5, e1000598 (2009).
2. Brawand, D. et al. The evolution of gene expression levels in mammalian organs. Nature 478, 343–348 (2011).
3. Soumillon, M. et al. Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. Cell Rep. 3, 2179–2190 (2013).
4. Lambert, S. A. et al. The human transcription factors. Cell 172, 659–665 (2018).
5. Shima, J. E., McLean, D. J., McCaryy, J. R. & Griswold, M. D. The murine testicular transcriptome: characterizing gene expression in the testis during the progression of spermatogenesis. Biol. Reprod. 71, 319–330 (2004).
6. Namekawa, S. H. et al. Postmeiotic sex chromatin in the male germline of mice. Curr. Biol. 16, 660–667 (2006).
7. Hasegawa, K. et al. SCML2 establishes the male germline epigenome through regulation of histone H2A ubiquitination. Dev. Cell 32, 574–588 (2015).
8. Sin, H. S., Kartashov, A. V., Hasegawa, K., Barski, A. & Namekawa, S. H. Pooled chromatin and bivalent domains facilitate the mitosis-to-meiosis transition in the male germline. BMC Biol. 13, 53 (2015).
9. Lesch, B. I., Silber, S. J., McCaryy, J. R. & Page, D. C. Parallel evolution of male germline epigenetic poising and somatic development in animals. Nat. Genet. 48, 888–894 (2016).
10. Waterston, R. H. et al. Initial sequencing and comparative analysis of the mouse genome. Nature 420, 520–562 (2002).
11. Lander, E. S. et al. Initial sequencing and analysis of the human genome. Nature 409, 860–921 (2001).
12. McClintock, B. The origin and behavior of mutable loci in maize. Proc. Natl Acad. Sci. USA 36, 344–355 (1950).
13. Meyer, T. J., Rosenkrantz, J. L., Carbone, L. & Chavez, S. L. Endogenous retroviruses: with us and against us. Nat. Rev. Genet. 17, 429–451 (2015).
14. Rebollo, R., Romanish, M. T. & Mager, D. L. Transposable elements: from conflicts to benefits. Nat. Rev. Genet. 46, 21–42 (2012).
15. Friedli, M. & Trono, D. The developmental control of transposable elements and the evolution of higher species. Annu. Rev. Cell Dev. Biol. 31, 429–451 (2015).
16. Choung, E. B., Elde, N. C. & Feschotte, C. Regulatory activities of transposable elements: from conflicts to benefits. Nat. Rev. Genet. 18, 71–86 (2017).
17. Garcia-Pérez, J. L., Widmann, T. J. & Adams, J. R. The impact of transposable elements on mammalian development. Development 143, 4101–4114 (2016).
18. Thompson, P. J., Macfarlan, T. S. & Lorincz, M. C. Long terminal repeats: from parasitic elements to building blocks of the transcriptional regulatory repertoire. Mol. Cell 62, 766–776 (2016).
19. Zamudio, N. & Bourc’his, D. Transposable elements in the mammalian germline: a comfortable niche or a deadly trap? Heredity (Edinb.) 105, 92–104 (2010).
20. Crichton, J. H., Dunican, D. S., MacLennan, M., Meehan, R. R. & Adams, I. R. Defending the genome from the enemy within: mechanisms of retrotransposon suppression in the mouse germline. Cell. Mol. Life Sci. 71, 1581–1605 (2014).
21. Ku, H. Y. & Lin, H. PIWI proteins and their interactors in piRNA biogenesis, germline development and gene expression. Natl Sci. Rev. 1, 205–218 (2014).
22. Winckelb, T., Cheng, E. C., Zhou, M. & Lin, H. Retrotransposons and pseudogenes regulate mRNAs and IncRNAs via the piRNA pathway in the germline. Genome Res. 25, 368–380 (2015).

976 NATURE STRUCTURAL & MOLECULAR BIOLOGY | VOL 27 | OCTOBER 2020 | 967–977 | www.nature.com/nsmb
23. Davis, M. P. et al. Transposon-driven transcription is a conserved feature of vertebrate spermatogenesis and transcript evolution. EMBO Rep. 18, 1231–1247 (2017).
24. Maezawa, S., Yukawa, M., Alavattam, K. G., Barski, A. & Namekawa, S. H. Dynamic reorganization of open chromatin underlies diverse transcriptomes during spermatogenesis. Nucleic Acids Res. 46, 593–608 (2018).
25. Alavattam, K. G. et al. Attenuated chromatin compartmentalization in mouse spermatogenesis. EMBO Rep. 16, 175–184 (2019).
26. Patel, L. et al. Dynamic reorganization of the genome shapes the recombination landscape in meiotic prophase. Nat. Struct. Mol. Biol. 26, 164–174 (2019).
27. Wang, Y. et al. Reprogramming of meiotic chromatin architecture during spermatogenesis. Mol. Cell 73, 547–561 (2019).
28. Maezawa, S. et al. Polycomb protein SCML2 facilitates H3K27me3 to establish bivalent domains in the male germline. Proc. Natl Acad. Sci. USA 115, 4957–4962 (2018).
29. Adams, S. R. et al. RNF8 and SCML2 cooperate to regulate ubiquitination and H3K27 acetylation for escape gene activation on the sex chromosomes. PLoS Genet. 14, e1007233 (2018).
30. Maezawa, S. et al. Super-enhancer switching drives a burst in gene expression at the mitosis-to-meiosis transition. Nat. Struct. Mol. Biol. https://doi.org/10.1038/s41594-020-04883-3 (2020).
31. Reichmann, J. et al. Microarray analysis of LTR retrotransposon silencing identifies Hdac1 as a regulator of retrotransposon expression in mouse embryonic stem cells. PLoS Comput. Biol. 8, e1002486 (2012).
32. Ollinger, R. et al. Deletion of the pluripotency-associated Tex19.1 gene causes activation of endogenous retroviruses and defective spermatogenesis in mice. PLoS Genet. 4, e1000199 (2008).
33. Russ, B. E. et al. Regulation of H3K4me3 at transcriptional enhancers characterizes acquisition of virus-specific CDb8+ T cell-lineage-specific function. Cell Rep. 21, 3624–3636 (2017).
34. Chuong, E. B., Rumi, M. A., Soares, M. J. & Baker, J. C. Endogenous retroviruses function as species-specific enhancer elements in the placenta. Nat. Genet. 45, 325–329 (2013).
35. Sin, H. S. et al. RNF8 regulates active epigenetic modifications and escape gene activation from inactive sex chromosomes in post-miotic spermatids. Genes Dev. 26, 2737–2748 (2012).
36. Bolcun-Filas, E. et al. A-MYB (MYBL1) transcription factor is a master regulator of male meiosis. Development 138, 3319–3330 (2011).
37. Li, X. Z. et al. An ancient transcription factor initiates the burst of piRNA production during early meiosis in mouse testes. Mol. Cell 50, 67–81 (2013).
38. Hubley, R. et al. The Dfam database of repetitive DNA families. Nucleic Acids Res. 44, D81–D89 (2016).
39. Isabel, L. et al. Trim33 binds and silences a class of young endogenous retroviruses in the mouse testis; a novel component of the arms race between retrotransposons and the host genome. PLoS Genet. 11, e1005693 (2015).
40. Brind’Amour, J. et al. An ultra-low-native captive ChIP-seq protocol for genome-wide profiling of rare cell populations. Nat. Commun. 6, 6033 (2015).
41. Nakasuji, T. et al. Complementary critical functions of Zfy1 and Zfy2 in mouse spermatogenesis and reproduction. PLoS Genet. 13, e1006578 (2017).
42. McCarthy, J. R. Toward a more precise and informative nomenclature describing fetal and neonatal male germ cells in rodents. Biol. Reprod. 89, 47 (2013).
43. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74 (2012).
44. Ecco, G. et al. Transposable elements and their KRAB-ZFP controllers regulate gene expression in adult tissues. Dev. Cell 36, 611–623 (2016).
45. Imbeault, M., Helleboid, P. Y. & Trono, D. KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. Nature 543, 550–554 (2017).
46. Peaston, A. E. et al. Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. Dev. Cell 7, 597–606 (2004).
47. Veselovska, L. et al. Deep sequencing and de novo assembly of the mouse oocyte transcriptome define the contribution of transcription to the DNA methylation landscape. Genome Biol. 16, 209 (2015).
48. Franke, V. et al. Long terminal repeats power evolution of genes and gene expression programs in mammalian oocytes and zygotes. Genome Res. 27, 1384–1394 (2017).
49. Bogutz, A. et al. Evolution of imprinting via lineage-specific insertion of retroviral promoters. Nat. Genet. 49, 941–945 (2017).
50. De Iaco, A. et al. DUX-family transcription factors regulate zygotic genome activation in placental mammals. Nat. Genet. 49, 925–934 (2017).
51. Kim, T. K. et al. Widespread transcription at neuronal activity-regulated enhancers. Nature 465, 182–187 (2010).
52. Wang, M. et al. Single-cell RNA sequencing analysis reveals sequential cell fate transition during human spermatogenesis. Cell Stem Cell 23, 599–614 (2018).
53. Dunn-Fletcher, C. E. et al. Anthropoid primate-specific retroviral element THE1B controls expression of CRH in placenta and alters gestation length. PLoS Biol. 16, e2006337 (2018).
54. Chuong, E. B. The placenta goes viral: retroviruses control gene expression in pregnancy. PLoS Biol. 16, e3000028 (2018).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
© The Author(s), under exclusive licence to Springer Nature America, Inc. 2020
Methods

Animals. Mice were maintained and used according to the guidelines of the Institutional Animal Care and Use Committee (protocol no. IACUC/2018-0040) at Cincinnati Children’s Hospital Medical Center. A *Myb*+/− (MybΔVPR) mouse, which was generated by N-ethyl-N-nitrosourea (ENU)-induced mutations on the C57BL/6J background, have been reported previously[19]. Through mating between male and female *A*−/− heterozygotes[20], *A*−/− male mice were born at the expected ratios according to Mendel’s Law. For the genotyping of *A*−/− mice, PCR was carried out using specific primer sets[21] (Supplementary Data 6).

Methods for construction of single-guide RNAs (sgRNAs) and the production of animals have been described previously[22]. In short, we targeted each side of the Zfy2-associated ERV with two chemically modified sgRNAs (IDT) according to on- and off-target scores generated via the web tool CRISPOR (http://crispor.tefor.net/). The target sequences were AAAGTTGAAAGCTTCGGG and AATAGTTCTGACTCTGGT for the upstream sites and CACTAGTCCATACCCAAAACA and TTGTCTCAGAATGGACT for the downstream sites. To form ribonucleoprotein complexes (RNPs), sgRNAs (25 nmol/l) each were mixed with Cas9 protein (IDT; 200 nmol/l) in Opti-MEM medium (ThermoFisher) and incubated at 37 °C for 15 min. Zygotes from superovulated female mice on the C57BL/6 genetic background were electroporated with 7.5 pmol of RNPs on ice using a Genome Editor electroporator (BEX; 30 V, 1-ms width, five pulses with 1-s intervals). Two minutes after electroporation, zygotes were moved into 500 ml of cold M2 medium (Sigma), warmed to room temperature, and then transferred into the oviductal ampulla of pseudopregnant CD-1 females. pups were born and genotyped by PCR and Sanger sequencing. Animals were housed in a controlled environment with a 12-h light/12-h dark cycle, with free access to food and water and a standard chow diet. All animal procedures were carried out in accordance with the Institutional Animal Care and Use Committee-approved protocol of Cincinnati Children's Hospital Medical Center.

Cell lines. Wild-type J1 male ESCs derived from male agouti 129Sv/SyJae embryos have been described previously[23]. Human HEK293T cells were obtained from ATCC (CRL-11268). CRISPRa ESC and RLR10B2-targeting CRISPRa ESC cell lines were generated in this study. Because these cells were easily distinguished based on colony morphologies, cell lines were authenticated by microscopic inspection. CRISPRa ESCs and RLR10B2-targeting CRISPRa ESCs were further authenticated by genotyping using specific primer sets (Supplementary Data 6). None of the cell lines were tested for mycoplasma contamination.

Cell culture. The ESCs were cultured in ESC medium (15% FBS, 25 mM HEPES, 1X GlutaMAX, 1X MEM non-essential amino acids solution, 1X penicillin/streptomycin, 1X 100 µM β-mercaptoethanol in DMEM high glucose (4.5 g/l)−1 containing 2i (1 µM PD0325901, LC Laboratories; and 3µg/ml CHIR99021, LC Laboratories) and LIF (1.300 U/ml, in-house) on mouse embryonic fibroblast cell lines seeded in each well of a 24-well plate coated with 0.2% gelatin under feeder-free conditions. HEK293T cells (CRL-11268, ATCC) were cultured in DMEM high glucose supplemented with (10% FBS, 1X penicillin/streptomycin, 1X mM sodium pyruvate, 1X MEM non-essential amino acids solution, 1X 100 µM β-mercaptoethanol in DMEM high glucose (4.5 g/l)−1 containing 2i, in-house) on cell culture plates coated with 0.2% gelatin and containing ESC medium supplemented with 2i and LIF. The ESCs were cultured in ESC medium with ESC medium supplemented with 1 µg/ml LIF. Dox. After 24 h of Dox induction, we transduced the cells with 500 ng of A-MYB expression vector (PGK-A-MYB plasmid) and Lipofectamine 3000 transfection reagent (Thermo Fisher) according to the manufacturer's instructions. At day 3, the adherent cells in each well were lysed for RNA extraction.

Crispra: a representative enhancer-like ERV locus. To perform functional evaluations of a representative enhancer-like ERV locus (Tdir3-ERV) via CRISPRa, we used CRISPOR (http://crispor.tefor.net/) to design two sgRNAs for the two loci of regions flanking Tdir3-ERV A-MYB expression vectors (PGK-A-MYB plasmid) was transfected with Lipofectamine 3000 transfection reagent (Thermo Fisher) according to the manufacturer's instructions. At day 3, the adherent cells in each well were lysed for RNA extraction.

RNA extraction and RT-qPCR. Total RNA was isolated using an RNeasy Plus Mini Kit (Qiagen). First-strand CDNA synthesis was performed using 200 ng of total RNA with SuperScript IV reverse transcriptase and oligo-dT (20) primer. Nluc was driven by a minimal promoter and was quantified with the ΔΔCT method and normalized to Hprt expression.

RNA-seq. We extracted total RNA from the ESCs in each well of a 24-well plate using an RNeasy Plus Mini Kit (Qiagen) with genomic DNA elimination. RNA-seq library preparation was carried out using a TrueSeq Stranded mRNA Library Prep Kit (Illumina) following the manufacturer's instructions. Indexed libraries were pooled and sequenced using Illumina Novaseq 6000 sequencer (paired end, 150 bp). Two independent biological replicates were generated for each sample.

Dual-luciferase reporter assays. Dual-luciferase reporter assays were performed in which the activity of regulatory elements was indicated by the expression of Nanoluc luciferase (Nluc). Nluc was driven by a minimal promoter and normalized to the expression of control firefly luciferase (Fluc), in turn driven by a PKG promoter. For the construction of A-MYB expression vectors, the full-length of a PKG promoter sequence was amplified via PCR using KOD Xpert Hot Start DNA polymerase (Sigma-Aldrich) and a specific primer set containing BstGl and AsiI recognition sites at the 5' ends (Supplementary Data 6). The PCR product was inserted into the BstGl and AsiI recognition sites of MG221561 plasmid (Origene), which bears the sequence for pA for A-MYB (GenBank: NC000019.13). To construct pNL3.2 reporter, DNA fragments in both orientations of two tandemly arranged regulatory elements (REs) were amplified from pA-mybmut A-mybmut and pA-mybmut A-mybmut using primers (Supplementary Data 6). DNA polymerase (Sigma-Aldrich) and a specific primer set containing BsrGI and AslI were used to construct the specific primer set. The constructed RLR10B2-targeting sgRNA expression plasmid (pDsRed reporter and blasticidin S resistance genes), pSPAX2 (12260, Addgene) packaging vector and pMD2.G (12259, Addgene) viral envelope expressing vector. The vector was constructed at a ratio of 0.377 (sgRNA/multi-vector) 0.247 (pMD2.G vector) using transfection reagent (Thermo Fisher). After 24 h of transfection, cells were treated with 10 µM forkosilin (F9171, Sigma-Aldrich). Viral supernatants were collected 48 h following transfection and concentrated using a Lenti-X concentrator (631231, Clontech). Virus titer was measured using a Lentix-GoStix Plus system (631280, Clontech), then stored at −80 °C.

One day before transduction, 1 × 10⁵ CRISPRa ESCs were seeded onto a 60-mm dish coated with 0.2% gelatin. For the viral infection of CRISPRa ESCs, concentrated sgRNA lentiviral particles (2 × 10⁵ IPFU) were used with 8 µg/ml of polybrene (TR-1003-G, Millipore) and 1/100 diluted ViralPlus transduction enhancer (G698, abm). Following transduction, cells were allowed to expand for four days in ESC medium with 2i and LIF. To enrich for DsRed-positive cells, cells were sorted with a fluorescence activated cell sorting (FACS) instrument (SH8005 cell sorter, Sony; 100-µm microfluidic sorting chip) four days after selection in ESC medium containing 2i, LIF, 200 µg/ml hygromycin B (InvivoGen) and 20 µg/ml blasticidin S (Gibco). We termed the newly established cell line RLR10B2-targeting CRISPRa ES cells. The cell line was maintained in ESC medium containing 2i, LIF, 200 µg/ml hygromycin B (InvivoGen) and 20 µg/ml blasticidin S (Gibco).

To evaluate the roles of enhancer-like ERVs in the expression of adjacent genes, we seeded 2 × 10⁵ RLR10B2-targeting CRISPRa ESCs into each well of a 24-well plate. The wells were washed twice with PBS and LIF. The following day, the cell line was expanded in ESC medium with ESC medium supplemented with 1 µg/ml LIF. Dox. After 24 h of Dox induction, we transduced the cells with 500 ng of A-MYB expression vector (PGK-A-MYB plasmid) and Lipofectamine 3000 transfection reagent (Thermo Fisher). The cell line was expanded in ESC medium with ESC medium supplemented with 1 µg/ml LIF. Dox. After 24 h of Dox induction, the adherent cells in each well were lysed for RNA extraction.

RNA-seq library preparation was carried out using a TrueSeq Stranded mRNA Library Prep Kit (Illumina) following the manufacturer's instructions. Indexed libraries were pooled and sequenced using Illumina NovaSeq 6000 sequencer (pair-end, 150 bp). Two independent biological replicates were generated for each sample.
To measure the activity of the above-prepared REs, we used the Nano-Glo dual-luciferase reporter assay system (Promega). A total of 5 × 10⁴ HEK293T cells were seeded into each well of a tissue culture-treated 96-well solid white polystyrene microplate (Corning) one day before transfection. Transfection transfections were performed with Lipofectamine 3000 transfection reagent (Thermo Fisher) following the manufacturer’s instructions. The cells in all wells were co-transfected with the following: 30 ng of plG4.53 (Promega), a transfection control reporter; 30 ng of plN3.2 with or without the REs, an experimental reporter; 40 ng of A-MYB expression vector (MG222161 or PGK-A-MYB plasmids). Three replicates were used for each condition. After 48 h of transfection, dual-luciferase reporter measurements were performed using a luminometer (Promega). Luciferase activity in each well was measured using a Synergy H1 hybrid multi-mode microplate reader (BioTek) with a 1 s integration time.

Native ChIP and sequencing. For native ChIP-seq of PS from wild-type and A-myb mutant mice, we performed testicular cell suspensions from 2-week-old male mice aged 8–12 weeks. We isolated A-myb mutant PS using the small-scale STA-PUT method as described in a previous report. Briefly, a pair of testes from one mouse, wild-type or mutant, underwent digestion by treatments with collagenase, trypsin and DNAse I. The cells were isolated and suspended in Krebs-Ringer bicarbonate buffer containing 0.5% BSA, and a 4% cell suspension was loaded into a gradient of Krebs-Ringer bicarbonate buffer containing 2% and 4% BSA; the gradient was generated through the use of a gradient maker (WVR, GM-100). The cell suspension was allowed to settle for 3 h at 4 °C before fractions were collected. Purity was confirmed by nuclear staining of a sample aliquot of each collected fraction with Hoechst 33342 via fluorescence microscopy. Greater than 90% purity was confirmed for each purification. To collect wild-type PS, we used an optimized quick sorting method. After preparing testicular cell suspensions, cells were stained with Vybrant DyeCycler Violet Stain (DCV, Thermo Fisher) for 30 min at 37 °C (2 µl DCV per 2 × 10⁵ cells). DCV staining patterns for testis cell types were detected and sorted via flow cytometry (SH800S Cell Sorter, SONY; a 100-µm microfluidic sorting chip was used). Approximately 5 × 10⁵ to 7.5 × 10⁵ cells were used for one native ChIP-seq experiment.

The protocol for native ChIP was adapted from a previous report with minor modifications. Briefly, isolated PSs were suspended in 20 µl of Nuclei EZ lysis buffer (Sigma-Aldrich) and digested chromatin with 2 µl of micrococcal nuclease (MNase, NEB) at 37 °C for 5 min. The MNase digestion was halted with the addition of 10% vol/vol 100 mM EDTA. Chromatin was completely solubilized with the addition of 10% vol/vol detergent solution (1% Triton X-100 and 1% sodium deoxycholate) with gentle inversion at 4 °C for 1 h. After solubilization of the chromatin, 10% of total chromatin was removed for use as an input control. The ChIP reaction was performed using 1 µg of rabbit anti-H3K27ac polyclonal antibody (ab4729, Abcam) conjugated to Dynabeads Protein A/G (1:1) magnetic beads (Life Technologies) overnight at 4 °C with gentle inversion and agitation. To remove non-specific binding interactions, magnetic beads bound to antibody–chromatin complexes were washed three times with a low-salt wash buffer (150 mM NaCl) and then twice with a high-salt buffer (500 mM NaCl). The chromatin was eluted from magnetic beads with elution buffer (100 mM NaHCO₃, and 1% sodium dodecyl sulfaide in ddH₂O) while shaking at 65 °C for 1 h. Immunoprecipitated DNA was isolated and purified by phenol–chloroform extraction and ethanol precipitation.

ChIP-seq library preparation was carried out using a NEBNext Ultra II DNA Library Prep Kit (NEB) following the manufacturer’s instructions. Indexed libraries were pooled and sequenced using an Illumina NextSeq 500 sequencer (paired-end, 150 bp). Two independent biological replicates were generated for each sample.

RNA-seq analyses. Raw RNA-seq reads were aligned to either the mouse (GRCm38/mm10) or human (GRCh38/hg38) genomes using bowtie2 (version 2.3.0.1) with default settings; the reads were filtered to remove alignments mapped to multiple locations by calling grip with the -v option. Using SeqMonk (Barbraham Bioinformatics), we calculated Pearson correlation coefficients between 1-kb bins of biological replicates. Peak calling for ATAC- and ChIP-seq data was performed using MACS (version 1.4.2) with default arguments (we used a cutoff of P ≤ 10⁻¹⁰). We normalized aligned ChIP-seq reads in enhancer-like ERV loci to RPKM, and relative ChIP-seq enrichments were calculated by input vs. background peak calling scripts.

To detect enhancer-like ERVs, we obtained RepeatMasker track annotations (GRCm38/mm10) from the UCSC Genome Browser (genome.ucsc.edu). First, to identify accessible ERV loci in the PS stage, we determined overrepresented ERV families through comparisons of the observed copy numbers of ERV families overlapping MACS-defined ATAC-seq peak regions versus the expected background. The expected background was estimated by randomly generating and calculating numbers of background genomic regions equal to the numbers of ATAC-seq peak regions. We computed the numbers of overlapping ERV copies within ATAC-seq peak regions (observed) and background genomic prevalence (expected) using custom shell scripts that call the BEDTools suite (version 2.26.0) functions intersect, sort, and calculate. Spearman’s rank correlation coefficient (P = 0.05, binomial test) were defined as ‘accessible’ ERVs in PSs (Fig. 2a,b). Accessible ERVs in PSs were further filtered to require ≥1.5 H3K27ac enrichment relative to input control and defined as ‘miose-specific enhancer-like ERVs’ (n = 1,122, Fig. 3c and Supplementary Data 2). The program ngs.plot was used to draw tag density plots and heatmaps for read enrichment.
Evaluation of sequence similarities across mammalian species. We sought to calculate sequence similarities and detect orthologous genes adjacent to mouse and human enhancer-like ERVs across the following mammalian species: rat, rabbit, marmoset, gorilla, and chimpanzee. To do so, we applied a list of mouse and human ERV adjacent genes to BioMart\textsuperscript{33} to compute sequence similarities across the species. For these analyses, we made use of NCBI ReSeq genes.

Histology and immunofluorescence analyses. Wild-type C57BL/6J male mice (three independent mice, 90–120 days old) were used for the immunofluorescence and histological analyses. To prepare testicular paraffin blocks, testes were fixed with 4% paraformaldehyde (PFA) overnight at 4°C. Testes were dehydrated and embedded in paraffin. For histological analyses, 5-µm-thick paraffin sections were deparaffinized and autoclaved in target retrieval solution (Dako) for 10 min at 121 °C. Sections were blocked with blocking One Histo (Nacalai) for 1 h at room temperature and then incubated with anti-γH2AX (05-636-AF647, Millipore) and anti-MYBL1 (A-MYB; NB1-90171, Novus Biologicals) primary antibodies overnight at 4°C. The resulting signals were detected by secondary antibodies conjugated to fluorophores (Thermo Fisher Scientific, Biotium or Jackson ImmunoResearch). Sections were deparaffinized and autoclaved in target retrieval solution (DAKO) for 10 min at 121 °C. Sections were blocked with Blocking One Histo (Nacalai) for 1 h at room temperature and then incubated with anti-γH2AX (05-636-AF647, Millipore) and anti-MYBL1 (A-MYB; NB1-90171, Novus Biologicals) primary antibodies overnight at 4°C. The resulting signals were detected by secondary antibodies conjugated to fluorophores (Thermo Fisher Scientific, Biotium or Jackson ImmunoResearch). Sections were counterstained with 4,6-diamidino-2-phenylindole. Images were obtained with a TiE fluorescence microscope (Nikon) and processed with NIS-Elements (Nikon) and ImageJ (National Institutes of Health)\textsuperscript{34}.

Statistics. Statistical methods and P values for each plot are listed in the figure legends and/or in the Methods. In brief, all grouped data are represented as mean ± s.e.m. All box-and-whisker plots are represented as center lines (median), box limits (interquartile range; 25th and 75th percentiles) and whiskers (≥90% of the data points), unless otherwise stated. Statistical significances for pairwise comparisons were determined using two-sided Mann–Whitney U tests, unpaired t-tests and chi-square tests with Yates’s correction. All quantitative analyses, excluding Extended Data Fig. 5a, are represented as the mean ± s.e.m. of three to four biological replicates. Fisher’s exact test and the hypergeometric test were used for the detection of significantly enriched GO terms, genes and loci compared with backgrounds. Differentially expressed genes and TE copies were determined using the DESeq\textsuperscript{35} package. Next-generation sequencing data (RNA-seq, ATAC-seq and ChIP-seq) were based on two independent replicates. For all experiments, no statistical methods were used to preclude sample size. Experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessments.

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

Data availability The H3K27ac ChIP-seq data reported in this study are described in the accompanying study by Maezawa et al.\textsuperscript{29} and are deposited to the Gene Expression Omnibus (GEO) under accession code GSE141273. H3K27ac natively ChIP-seq data in WT and A-myb mutant PsVs and RNA-seq data in CRISPRa ES cells are deposited under accession code GSE142173. All other next-generation sequencing datasets used in this study are publicly available and referenced in Supplementary Data 7–9, 34–36. Source data are provided with this paper.

Code availability Source code for all software and tools used in this study, with documentation, examples and additional information, is available at following URLs: https://github.com/Genomelmmunobiology/Sakashita_et_al_2020 (best match TE annotation set), https://github.com/alexdojin/STAR (STAR RNA-seq aligner), http://crispor.tororf/crispor, http://daechwanykdlab.github.io/hisa2 (HiSA2), https://cellbasesoftware/stringtie (StringTie), https://bitbucket.org/bedtools/bedtools.git (BedTools), https://biocoredata.org/packages/release/bioc/html/DESeq2.html (DESeq2), https://david.ncifcrf.gov/summary.jsp (DAVID), https://software.broadinstitute.org/morpheus (Morpheus), https://software.broadinstitute.org/software/igv/igvtools (IGV/igvWeb), https://bitbucket.org/bigsourceforge.net/bowtie2 (bowtie2), https://www.bioinformatics.babraham.ac.uk/projects/#seqmonk (SeqMonk), https://github.com/tazouli/MACS (MACS), https://github.com/shenlab-sina/ngsplot (ngsplot), https://cran.r-project.org/web/packages/gplots/index.html (gplots), https://github.com/tidyverse/ggplot2 (ggplot2), https://cran.r-project.org/web/packages/chromomap/vignettes/chromomap.html (chromoMap), http://bomer.ucsd.edu/homer (HOMER), http://great.stanford.edu/public/html (GREAT), https://useast.ensembl.org/info/datasets (biomart/index.html (BioMart), https://imagej.net/Fiji/Downloads (Fiji, ImageJ), https://systems.crumpl.uc.edu/hypergeometric (Hypergeometric P value calculator).

References
56. Yuan, C. L. & Hu, Y. C. A transgenic core facility's experience in genome editing revolution. Adv. Exp. Med. Biol. 1016, 75–90 (2017).
57. Haeneler, M. et al. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. Genome Biol. 17, 148 (2016).
58. Li, E., Bestor, T. H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell 69, 915–926 (1992).
59. Chavez, A. et al. Highly enriched Cas9-mediated transcriptional programming. Nat. Methods 12, 326–328 (2015).
60. Wang, W. et al. Chromosomal transposition of PiggyBac in mouse embryonic stem cells. Proc. Natl Acad. Sci. USA 105, 9290–9295 (2008).
61. Kim, D., Langmead, B. & Salzer, S. L. HiSA: a fast spliced aligner with low memory requirements. Nat. Methods 12, 357–360 (2015).
62. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169 (2015).
63. Pertea, M., Kim, D., Pertea, G. M., Leek, J. T. & Salzer, S. L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat. Protoc. 11, 1650–1667 (2016).
64. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparison genomic features. Bioinformatics 26, 841–842 (2010).
65. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and determination of RNA-seq data with DESeq2. Genome Biol. 15, 550 (2014).
66. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4, 44–57 (2009).
67. Robinson, J. T. et al. Integrative genomics viewer. Nat. Biotechnol. 29, 24–26 (2011).
68. Langmead, B. & Salzer, S. L. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359 (2012).
69. Feng, J., Liu, T. & Zhang, Y. Using MACS to identify peaks from ChIP-Seq data. Curr. Protocols Bioinformatics 34, 2.14.1–2.14.14 (2011).
70. Shen, L., Shao, N., Liu, X. & Nesterl, E. ngs.plot: quick mining and visualization of next-generation sequencing data by integrating genomic databases. BMC Genomics 15, 284 (2014).
71. Heinz, S. et al. Simple combinations of lineage-determination transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell 38, 576–589 (2010).
72. McLean, C. Y. et al. GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. 28, 495–501 (2010).
73. Kinsella, R. J. et al. Ensembl BioMarts: a hub for data retrieval across taxonomic space. Database 2011, bar030 (2011).
74. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to Image: 25 years of image analysis. Nat. Methods 9, 671–675 (2012).
75. Jung, Y. H. et al. Chromatin states in mouse sperm correlate with embryonic and adult regulatory landscapes. Cell Rep. 18, 1366–1382 (2017).
76. Lavin, Y. et al. Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. Cell 159, 1312–1326 (2014).
77. Guo, J. et al. Chromatin and single-cell RNA-Seq profiling reveal dynamic signaling and metabolic transitions during human spermatogonial stem cell development. Cell Stem Cell 21, 533–546 (2017).
78. Li, D. et al. Chromatin accessibility dynamics during iPSC reprogramming. Cell Stem Cell 21, 819–833 (2017).
79. He, S. et al. Hemi-methylated CpG sites connect DNA demethylation during somatic cell reprogramming. Cell Discov. 5, 11 (2019).
80. Cao, S. et al. Chromatin accessibility dynamics during chemical induction of pluripotency. Cell Stem Cell 22, 529–542 (2018).
Acknowledgements

We thank M. Weirauch and members of the Namekawa laboratory for discussion and helpful comments regarding the manuscript, the CCHMC Research Flow Cytometry Core for sharing FACS equipment (supported by NIH S10OD023410), X. Li at the University of Rochester Medical Center for sharing A-myb mutant mice, the laboratory of B. Bernstein at Massachusetts General Hospital for providing human testis H3K27ac ChIP-seq data (ENCRR136ZQZ, ENCODE) and the Transgenic Animal and Genome Editing Core at CCHMC for generating the Zfy2 enhancer-deletion mice. We acknowledge the following funding sources: Lalor Foundation Postdoctoral Fellowship and JSPS Overseas Research Fellowship to A.S.; the Research Project Grant by the Azabu University Research Services Division, Ministry of Education, Culture, Sports, Science and Technology (MEXT) Supported Program for the Private University Research Branding Project (2016–2019), a Grant-in-Aid for Research Activity Start-up (19K21196) and the Uehara Memorial Foundation Research Incentive Grant (2018) to S.M.; an Albert J. Ryan Fellowship to K.G.A.; National Institute of Health (NIH) grant DP2 GM119134 to A.B.; a March of Dimes Prematurity Research Centre Collaborative Grant (#22-FY14-470) to M.P.; NIH R01 GM122776 grant to S.H.N.

Author contributions

The manuscript was written by A.S., K.G.A. and S.H.N., with critical feedback from all other authors. A.S. and S.H.N. designed the study. S.M. performed crosslinking ChIP-seq experiments and A.S. performed native ChIP-seq experiments. A.S. analyzed A-myb mutant mice with the help of K.T. A.S. and K.T. performed CRISPRa experiments. A.S. performed immunostaining and dual-luciferase reporter assays. Y.-C.H. supervised the generation of the Zfy2 enhancer-deletion mice. A.S., K.G.A., M.Y., S.K., N.F.P., A.B., M.P. and S.H.N. designed and interpreted the computational analyses. A.S. performed the majority of computational analyses. S.H.N. supervised the project.

Competing interests

A.B. is a cofounder of Datirium, LLC.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41594-020-0487-4.

Supplementary information is available for this paper at https://doi.org/10.1038/s41594-020-0487-4.

Correspondence and requests for materials should be addressed to S.H.N.

Peer review information Beth Moorefield was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Reprints and permissions information is available at www.nature.com/reprints.
Extended Data Fig. 1 | Analysis of repetitive element expression during mouse spermatogenesis. a, The RNA-seq pipeline for comprehensive quantification of TE copies. The flowchart indicates the various RNA-seq and data analysis processes that comprise the pipeline. Round-corner rectangles, input files; rectangles, output files; diamond, branch condition. The specific tools used are highlighted in red. b, The proportion of expressed and unexpressed copies of repetitive elements in each class during spermatogenesis. Of note, nearly half of rRNA genes are expressed in spermatogenic differentiation following the KIT+ spermatogonia stage.
Extended Data Fig. 2 | ATAC-seq read enrichment at representative enhancer-like ERV loci and 5,000 randomly selected repetitive element loci.

**a**, Heatmap depicts RPKM-normalized ATAC-seq reads at enhancer-like RLTR10 and RMER17 loci (n = 694), and 5,000 randomly selected repetitive element loci in representative stages of spermatogenesis. 

**b**, Top: Venn diagram shows the intersection between total copy numbers of MMERVK10C loci (green) and total copy numbers of all RLTR10C loci (pink). Bottom: Venn diagram shows the intersection between total copy numbers of MMERVK10C loci (green) and total copy numbers of enhancer-like RLTR10C loci (red).
Extended Data Fig. 3 | H3K4me3 enrichment at enhancer-like ERVs loci. **a**, Average tag density plots and heatmaps show H3K27ac and H3K4me3 enrichments around enhancer-like ERVs (±1 kb around ±5 kb of ERVs) in PS. **b**, Scatter plot depicts H3K4me3 enrichments at enhancer-like ERV loci in PS. X-axis indicates relative distance of enhancer-like ERV loci from TSS of nearest genes. Y-axis indicates relative H3K4me3 enrichments at individual enhancer-like ERV loci. Red line shows a regression line.
Extended Data Fig. 4 | The genomic features of enhancer-like ERVs in meiosis. a, Representative track views show H3K27ac ChIP-seq, ATAC-seq, RNA-seq, and A-MYB ChIP-seq signals on chromosome X. The red highlight indicates an enhancer-like ERV locus. b, Pie charts indicate the distributions of enhancer-like ERVs on autosomes and sex chromosome. c, Top: Bar chart depicts the numbers of enhancer-like ERVs on each chromosome. Bottom: Chromosome map shows the distribution of enhancer-like ERVs throughout the mouse genome. Values for H3K27ac enrichment represent log2 fold enrichment of H3K27ac signal relative to input. d, Box-and-whisker plots show relative H3K27ac enrichment at enhancer-like ERV loci on autosomes and sex chromosomes. Values: log2 fold enrichment of H3K27ac signal relative to input. Central bars represent medians, the boxes encompass 50% of the data points, and the error bars indicate 90% of the data points. We detected no statistical difference in H3K27ac enrichment at autosome enhancer-like ERVs vs. sex chromosome enhancer-like ERVs: P = 0.307, Mann-Whitney U test. e, Bar chart shows enhancer-like ERVs distribution across genomic entities (intergenic, intronic, etc.) in autosomes versus the sex chromosomes: P = 3.6 × 10^{-5}, Chi-square test with Yates’s correction. f, The consensus sequence of RLTR10B, listed in the Dfam database, contains two A-MYB binding motifs (GGCAGTT).

TGTTGGACGCGCAGATGGGCAGTTGCAGATGGCAGTTGG
CTACTGCTGGCACCACACATACATAGCGAGTGAAGTTCTTTCG
CAAGATGAGAAGTAAACCAATCAAGTACGAGAAGAAGGTAAACCA
ATACAGATGACACGCTCCTTCAGGCTATGGAGCAGCAACC
GTTTGGGCTGCGGTCTTTCCGCTCCTACATCAAGCCTCCTCCCAA
TAAACGTGTGCAAGAGGATCCTGTGAGCTGCTTCTCTCCGAGC
AGTCGACGCAGAGCAGAAG

**GGCAGTT** A-MYB binding motifs
Extended Data Fig. 5 | The generation of CRISPRa embryonic stem cell lines, and the evaluation of CRISPR-deletion mice. 

**a,** qRT-PCR analyses of CRISPRa embryonic stem (ES) cells show expression level changes of the dCas9-VPR transgene 24 h after doxycycline (Dox) induction. Expression levels were normalized to the endogenous housekeeping gene Hprt. Upon addition of Dox, all ES cell clones evinced overt dCas9-VPR mRNA expression. Because clone #6 exhibited the highest upregulation of dCas9-VPR transcript, we restricted further experiments to clone #6. 

**b,** Representative image of CRISPRa ES cell colonies at day 4 after transduction with the sgRNA lentiviral construct. We validated the degree of sgRNA expression through observations of the red fluorescent reporter protein DsRed. Scale bar, 200 μm. 

**c,** Testis sections from wild-type (WT; left) and Zfy2 enhancer-deletion mice (right) at postnatal day 28 (P28). The sections were stained with hematoxylin and eosin. Scale bars, 100 μm. In our observations of Zfy2 enhancer-deletion samples, we noted no gross changes to testis morphology; however, we observed multinucleated cells (arrowheads).
Extended Data Fig. 6 | The synteny of mouse meiosis-specific enhancer-like ERVs in rats and other placental mammals. a, Pie charts indicate the genomic distribution of enhancer-like ERVs in the following genomes: mouse (mm10) and rat (rn6). Between the two species, genomic feature enrichment statistically differs: *** P < 0.001, Chi-square test with Yates's correction. b, Representative track views show evolutionary conservation in regions adjacent to enhancer-like ERVs across several placental mammals. Red highlights indicate enhancer-like ERV loci; such loci exhibit low levels of conservation across placental mammals, including rats, a species closely related to mice.
Extended Data Fig. 7 | MER57E3 is enriched in KRAB-ZF-encoding genes that have rapidly evolved in primates or humans. a. Representative track views show H3K27ac ChIP-seq enrichment for whole, adult human testis tissue and RNA-seq signal in human KIT+ and PS. Red highlights indicate enhancer-like MER57E3s that overlap high levels of H3K27ac deposition. b. Pie charts indicate the genomic distribution of enhancer-like MER57E3 loci in the human genome (hg38). Most enhancer-like MER57E3s are located within the first intronic regions of KRAB-ZF-encoding genes.
Extended Data Fig. 8 | A-MYB is highly expressed in both mouse and human spermatocytes. **a**, Testis sections from mice at 12 weeks of age immunostained with antibodies raised against A-MYB (red) and γH2AX (green), and counterstained with DAPI (gray). The Roman numerals indicate stages of the seminiferous epithelium cycle. Scale bars, 20 µm. **b**, Representative testis sections from humans at 29-to-65 years of age immunohistochemically stained with an antibody raised against A-MYB (brown), counterstained with hematoxylin. Images of human testis sections were sourced and adapted from the Human Protein Atlas (www.proteinatlas.org/ENSG00000185697-MYBL1/tissue/testis). Scale bars, 20 µm.
Reporting Summary

Nature Research 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 Research policies, see Authors & Referees 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

Date were collected with GNU wget and SRA Toolkit 'prefetch' from the following web databases: NCBI Gene Expression Omnibus (GEO) and ENCODE: Encyclopedia of DNA Elements.

Data analysis

Source code for all software and tools used in this study with documentation, examples and additional information, is available at following URLs:

- https://github.com/GenomeImmunobiology/Sakashita_et_al_2020 (best-match TE annotation set)
- https://github.com/alexdoxin/STAR (STAR RNA-Seq aligner)
- http://crispor.tefor.net (CRISPOR)
- http://daehwankimlab.github.io/hsat2 (HSAT2)
- https://ccb.jhu.edu/software/stringtie (StringTie)
- https://htseq.readthedocs.io/en/master/HTSeq)
- https://bedtools.readthedocs.io/en/astest/content/installation.html (BEDTools)
- https://bioconductor.org/packages/release/bioc/html/DESeq2.html (DESeq2)
- https://david.ncifcrf.gov/summary.jsp (DAVID)
- https://software.broadinstitute.org/morpheus (Morpheus)
- https://software.broadinstitute.org/software/lgvgtools (LGVTools)
- http://bostie.bio.sourceforge.net/bostie2 (bostie2)
- https://www.bioinformatics.babraham.ac.uk/projects/seqmonk (SeqMonk)
- https://github.com/taoliu/MACS (MACS)
- https://github.com/shenab-sina/ngsplot (ngsplot)
- https://rdr.io/cran/gplots (gplots)
- https://github.com/sidyverse/ggplot2 (ggplot2)
- https://cran.r-project.org/web/packages/chromomap/vignettes/chromomap.html#install-chromomap (chromoMap)
- http://homer.ucsd.edu/homer (HOMER)
- http://great.stanford.edu/public/html (GREAT)
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 list of figures that have associated raw data
- A description of any restrictions on data availability

All raw sequence data used in this study are listed in Supplementary DataSet 7.

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

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 sample sizes.

Data exclusions
No data were excluded from analyses.

Replication
We confirmed consistent results between three-to-four independent biological replicates for immunofluorescence analysis, dual-luciferase reporter assays, CRISPRa experiments evaluated by qPCR, and the Zfy2 CRISPR-deletion mouse analysis by qPCR. For Nextgen sequencing analysis, we obtained all biological or technical replicate samples from web databases (above mentioned) and confirmed consistent results between respective replicates. We also confirmed consistent results between two independent biological replicates for H3k27ac native ChIP-seq experiments based on Pearson’s correlation coefficient of each peak by SeqMonk.

Randomization
The experiments were not randomized.

Blinding
The experiments were not 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
  - Antibodies
  - Eukaryotic cell lines
  - Palaeontology
  - Animals and other organisms
  - Human research participants
  - Clinical data

Methods
n/a
- Involved in the study
  - ChIP-seq
  - Flow cytometry
  - MRI-based neuroimaging

Antibodies

The methods section of the manuscript contains information on all antibodies.

Anti-phospho Histone H2A.X (Ser139) Antibody, clone JBW301, Alexa Fluor® 647 (05-636-AF647, Milipore: https://www.emdmillipore.com/US/en/product/Anti-phospho-Histone-H2A.X-Ser139-Antibody-clone-JBW301-Alexa
Fluor-647,MM_NF-05-636-AF647)

A-Myb Antibody (NBP1-90171, NOVUS BIOLOGICALS; https://www.novusbio.com/products/a-myb-antibody_nbp1-90171)
Validation

Antibodies used for immunofluorescence analysis were validated by manufacturers.
- Anti-phospho Histone H2A.X (Ser139) Antibody, http://www.emdmillipore.com/US/en/product/Anti-phospho-Histone-H2A.X-Ser139-Antibody-clone-JBW301-Alexa-Fluor-647, MM_NF-05-636-AF647
- A-Myb Antibody, https://www.novusbio.com/products/a-myb-antibody_nb1-90171

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s) Mouse J1 ES cells were obtained from Dr. Yuya Ogawa (Cincinnati Children's Hospital Medical Center, Cincinnati, USA). Human HEK293T cells were obtained from ATCC (CRL-11268). CRISPRa ES cell lines have been generated in the Namekawa laboratory.

Authentication Since these cells were easily distinguished based on colony morphologies, cell lines have been authenticated by microscopic inspection. CRISPRa ES cells were authenticated by the presence of functional transgenes (Details were described in Methods section).

Mycoplasma contamination None of the cell lines used have been tested.

Commonly misidentified lines (See ICLAC register) Name any commonly misidentified cell lines used in the study and provide a rationale for their use.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals Wild-type C57BL/6J male mice (three independent mice, at 90-120 days of age) were used for the immunofluorescence analysis and histological analysis. Wild-type and A-myb mutant male mice (two independent mice, at 8-12 weeks of age) were used for the native ChIP-seq analysis.

Wild animals Provide details on animals observed in or captured in the field; report species, sex and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.

Field-collected samples For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.

Ethics oversight Mice were maintained and used according to the guidelines of the Institutional Animal Care and Use Committee (protocol no. IACUC2018-0040) at Cincinnati Children’s Hospital Medical Center. Note that full information on the approval of the study protocol must also be provided in the manuscript.

ChIP-seq

Data deposition

[ X ] Confirm that both raw and final processed data have been deposited in a public database such as GEO.

[ X ] 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.

H3K27ac ChIP-seq data reported in this study are deposited to the Gene Expression Omnibus (GEO) under the accession number GSE130652 and GSE142173. Other RNA- and ChIP-seq data datasets used in this study were obtained from public databases. All raw sequencing files from databases are listed in Supplementary DataSet 7.

Files in database submission

Provide a list of all files available in the database submission.

Genome browser session (e.g. UCSC)

Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.

Methodology

Replicates

Describe the experimental replicates, specifying number, type and replicate agreement.

Sequencing depth

Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.

Antibodies

Antibody against H3K27ac for native ChIP-seq experiments was obtained from Abcam: ab4729.

Peak calling parameters

Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.
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

Please see the Methods section; "Native ChIP and sequencing" and "CRISPRa: RLTR10B2 consensus sequence"

Instrument

SH800S Cell Sorter with 4 laser, SONY; a 100-μm microfluidic sorting chip was used.

Software

SH800 Software

Cell population abundance

For isolation of pachytene spermatocytes (PS), High-purity PS were approximately 5% of the total population. Purified PS were authenticated by microscopic inspection. For isolation of RLTR10B2 targeting CRISPRa ES cells, approximately 54% of the total cell population exhibited robust DsRed expression.

Gating strategy

FSC/SSC gates of the starting cell population were used to distinguish between viable cells and cell debris. Doublet cells were removed using FSC-A/ FSC-W gating. For isolation of PS, DyeCycle Violet (DCV) blue/DCV red gates were used to distinguish each testicular cell type. For isolation of RLTR10B2 targeting CRISPRa ES cells, gaiting in the histogram (DsRed-A/Counts) was determined by the clear separation between DsRed positive cells and negative cells.

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