

Potential role of KRAB-ZFP binding and transcriptional states on DNA methylation of retroelements in human male germ cells

Kei Fukuda^{1*}, Yoshinori Makino², Satoru Kaneko³, Chikako Shimura¹, Yuki Okada², Kenji Ichiyangi⁴, Yoichi Shinkai^{1*}

¹Cellular Memory Laboratory, RIKEN Cluster for Pioneering Research, Wako, Japan; ²Laboratory of Pathology and Development, Institute for Quantitative Biosciences, The University of Tokyo, Tokyo, Japan; ³Department of Obstetrics and Gynecology, Ichikawa General Hospital, Tokyo Dental College, Ichikawa, Japan; ⁴Laboratory of Genome and Epigenome Dynamics, Department of Animal Sciences, Graduate School of Bioagricultural Sciences, Nagoya University, Nagoya, Japan

Abstract DNA methylation, repressive histone modifications, and PIWI-interacting RNAs are essential for controlling retroelement silencing in mammalian germ lines. Dysregulation of retroelement silencing is associated with male sterility. Although retroelement silencing mechanisms have been extensively studied in mouse germ cells, little progress has been made in humans. Here, we show that the Krüppel-associated box domain zinc finger proteins are associated with DNA methylation of retroelements in human primordial germ cells. Further, we show that the hominoid-specific retroelement SINE-VNTR-*Alus* (SVA) is subjected to transcription-directed de novo DNA methylation during human spermatogenesis. The degree of de novo DNA methylation in SVAs varies among human individuals, which confers significant inter-individual epigenetic variation in sperm. Collectively, our results highlight potential molecular mechanisms for the regulation of retroelements in human male germ cells.

*For correspondence: kei.fukuda@riken.jp (KF); yshinkai@riken.jp (YS)

Competing interest: The authors declare that no competing interests exist.

Funding: See page 18

Preprinted: 19 May 2021

Received: 06 January 2022

Accepted: 21 March 2022

Published: 22 March 2022

Reviewing Editor: Deborah Bourc'his, Institut Curie, France

© Copyright Fukuda *et al.* This article is distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and redistribution provided that the original author and source are credited.

Editor's evaluation

Retrotransposons undergo massive reprogramming of their methylation states during germ cell development, but some elements are immune to this remodeling. This manuscript explores the contribution of binding motifs for KRAB-Zinc Finger Proteins (KZFPs) and position towards genes to explain the variable methylation dynamics of different retrotransposon families, namely L1, SVA and LTR12, as well as potential inter-individual variation during male germ cell development in humans, using an integrative analyses of available sequencing datasets. By bringing insights into the complex regulation of retrotransposons, it could be of particular interest to the epigenetics community.

Introduction

Transposable elements comprise more than 40% of most extant mammalian genomes (*Lander et al., 2001*). Among these, certain types of transposable elements called retroelements, including short/long interspersed elements (SINEs/LINEs) and hominoid-specific retrotransposons SINE-VNTR-*Alus* (SVA) are active in humans and can be transposed (*Huang et al., 2012; Maksakova et al., 2013; Ostertag et al., 2003*). As retrotransposons cause genome instability, insertional mutagenesis, or

transcriptional perturbation and are often deleterious to host species (*O'Donnell and Burns, 2010*), multiple defense mechanisms have evolved against transposition. The first line of defense is transcriptional silencing of integrated retroelements by chromatin modifications, such as DNA methylation and histone H3 lysine 9 (H3K9) methylation (*Fukuda and Shinkai, 2020; Goodier, 2016*). Most retroelement families are bound by Krüppel-associated box domain zinc finger proteins (KRAB-ZFPs), which coevolved to recognize specific retroelement families (*Imbeault et al., 2017; Jacobs et al., 2014; Wolf et al., 2015*). KRAB-ZFPs repress retroelements by recruiting KAP1/TRIM28 (*Sripathy et al., 2006*) and other repressive epigenetic modifiers (*Schultz et al., 2002; Schultz et al., 2001*).

Restricting retroelements is especially important for germ cells, because only germ cells transmit genetic information to the next generation. During embryonic development, primordial germ cells (PGCs) undergo epigenetic reprogramming, characterized by DNA demethylation and global resetting of histone marks in mice and humans (*Gkoutela et al., 2015; Guo et al., 2015; Kobayashi et al., 2013; Seisenberger et al., 2012; Seki et al., 2007; Tang et al., 2015*). A subset of young retroelements resists this global DNA demethylation event in PGCs, which may be required for retroelement silencing (*Gkoutela et al., 2015; Guo et al., 2015; Kobayashi et al., 2013; Seisenberger et al., 2012; Seki et al., 2007; Tang et al., 2015*). H3K9 trimethylation mediated by SETDB1 is enriched in DNA demethylation-resistant retroelements in mouse PGCs (*Liu et al., 2014*). As SETDB1 regulates DNA methylation of a subset of retroelements (*Matsui et al., 2010; Rowe et al., 2013*), and it is recruited to the retroelements via interaction with KRAB-ZFPs, it has been hypothesized that SETDB1/KRAB-ZFPs may contribute to DNA demethylation resistance in PGCs.

In contrast to the extensive DNA hypomethylation in PGCs, the genomic DNA of sperm is highly methylated in both humans and mice (*Hammoud et al., 2014; Kobayashi et al., 2013; Molaro et al., 2011; Okae et al., 2014*). Retroelements are also subjected to de novo DNA methylation during spermatogenesis in mice via the PIWI/piRNA pathway (*Aravin et al., 2008; Inoue et al., 2017*). Epigenetic alterations in retroelements and dysfunction of retroelement silencing pathways in male germ cells are associated with male sterility linked to azoospermia (*Aravin et al., 2007; Bourc'his and Bestor, 2004; Carmell et al., 2007; Heyn et al., 2012; Urduingio et al., 2015*). In addition, epigenetic alterations of retroelements in male germ cells can be potentially transmitted to the next generation with phenotypic consequences (*Daxinger et al., 2016; Rakyan et al., 2003*). Therefore, deciphering the regulatory mechanisms of retroelements in germ cells contributes to the understanding of sterility and transgenerational epigenetic inheritance. Extensive studies have been conducted to understand DNA methylation mechanisms in mouse spermatogenesis; however, limited progress has been achieved in humans.

In this study, we aimed to clarify the regulatory mechanisms of DNA methylation of retroelements in human germ cells and performed an integrative analysis using three sets of previously reported data, which included whole-genome bisulfite sequencing (WGBS) data for human PGCs (hPGCs) and sperm, the transcriptome of human male germ cells, and comprehensive human KRAB-ZFPs ChIP-exo data.

Results

Transposable elements showing DNA demethylation resistance in hPGCs

To learn more about the factors that contribute to DNA demethylation resistance in hPGCs, we reanalyzed publicly available WGBS data for male hPGCs (*Guo et al., 2015*). The global erasure of DNA methylation is mostly complete at 19 weeks of gestation (*Figure 1A*), therefore, we analyzed the DNA methylation status of full-length transposable elements (a copy whose length is 90% or more of the length of the consensus sequence of each subtype, listed in *Supplementary file 1*) in male hPGCs at 19 weeks of gestation to identify retroelements that showed resistance to demethylation. We generally focused on the retroelement types that had been analyzed for the DNA methylation status more than 30 copies. Among the retroelements we analyzed, the primate-specific retroelement families L1PA, SVA, and LTR12 showed high levels of DNA methylation (*Figure 1B*). In the SVA family, SVA_A, which emerged 13–14 million years ago (Mya) and is the oldest SVA type, showed the highest DNA methylation levels relative to other SVA types. This includes the currently active SVA_E/F (*Figure 1C*). In the L1 family, L1PA3–5, which emerged 12–20 Mya and is moderately young, showed

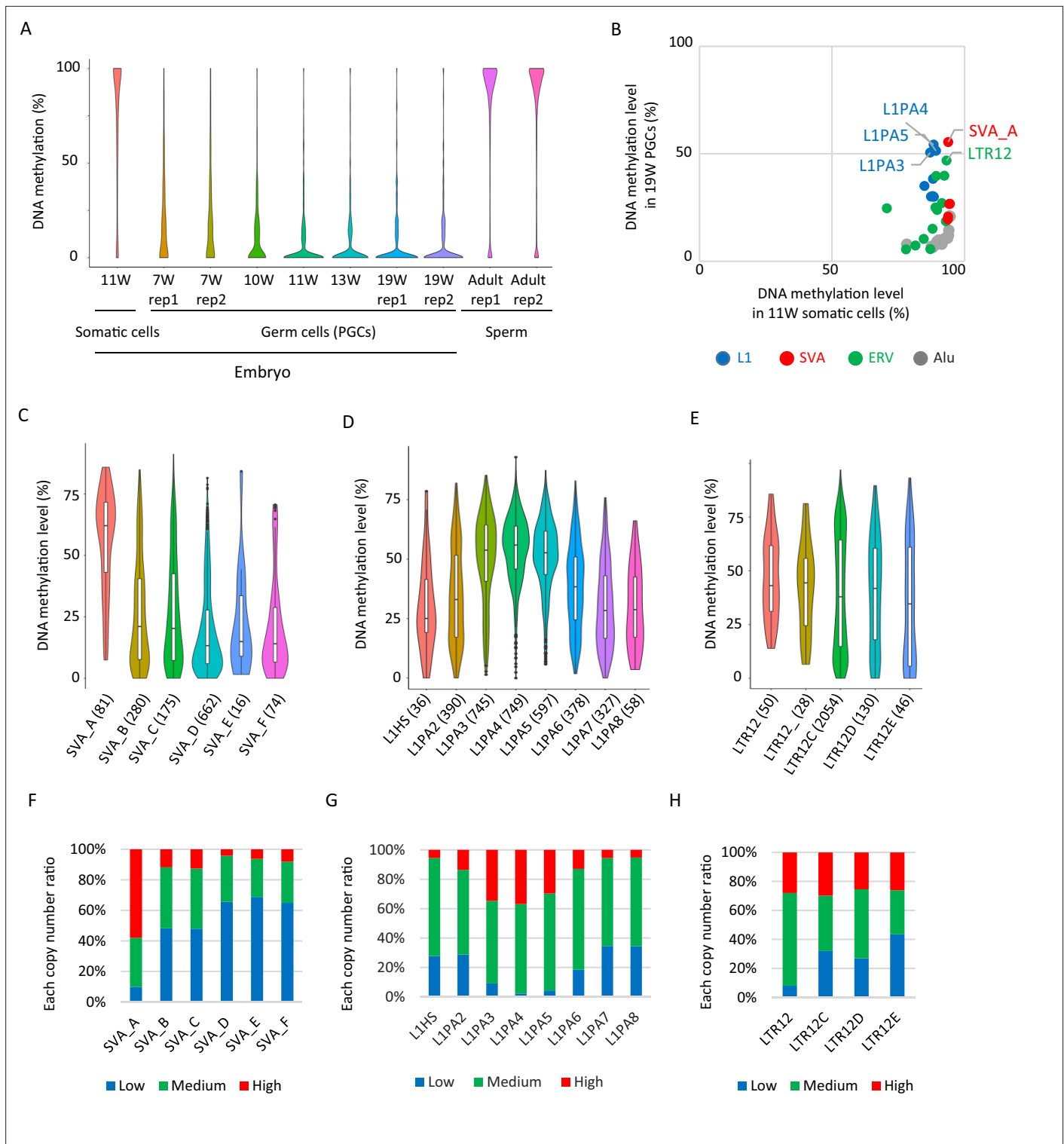


Figure 1. Retroelements showing DNA demethylation resistance. **(A)** Violin plots showing DNA methylation levels of each CpG site during human male germ-cell development. DNA demethylation was almost completed at 19 weeks of gestation. **(B)** Scatter plots showing average DNA methylation level of each retroelement type between somatic cells and male human primordial germ cells (hPGCs) at 19 weeks of gestation. Only full-length copies were used for this analysis, and retroelement types with ≥ 30 full-length copies were shown. Each plot was colored according to its retroelement family (red: SVA, blue: L1, green: LTR, gray: Alu). **(C–E)** Violin plots showing DNA methylation level of each retroelement type in hPGCs at 19 weeks of gestation. p-Value was calculated by Tukey’s test and was described in **Supplementary file 2**. The number in parentheses was analyzed copy number. **(F–H)** Bar Figure 1 continued on next page

Figure 1 continued

graphs showing the fraction of 'low', 'medium', and 'high' methylated class of each retroelement type in male hPGCs at 19 weeks of gestation. The retroelement copies used in these figures were same as those in (C-E).

The online version of this article includes the following source data and figure supplement(s) for figure 1:

Source data 1. Raw data of graphs in **Figure 1**.

Figure supplement 1. Mappability of whole-genome bisulfite sequencing (WGBS) reads on transposable elements.

higher methylation levels than the older (L1PA5–8) and younger L1 types, including the currently active L1 (L1HS) (**Figure 1D**). LTR12 (also known as HERV9 LTR) is not currently active and all LTR12 types are highly methylated (**Figure 1E**). Therefore, it appears that young but inactive L1PA, SVA, and LTR12 types are resistant to DNA demethylation in hPGCs. Because 100 bp short-read NGS data did not map efficiently onto the currently active L1 transposon, L1HS (**Figure 1—figure supplement 1A**), and DNA methylation of only about 10% of full-length L1HS copies could be analyzed (**Figure 1—figure supplement 1B**), it is possible that a subset of L1HS is resistant to DNA demethylation. Epigenome analysis using long-read sequence technology, such as nanopore sequencing, may provide an answer to this question (**Ewing et al., 2020**). Even though some retroelement types showed relatively high DNA methylation levels in hPGCs, the DNA methylation levels of each retroelement type were highly variable among full-length copies (**Figure 1C–E**), which prompted us to try to identify potential DNA sequences required for DNA demethylation resistance. To this end, we classified each retroelement copy according to DNA methylation levels as follows: low <20%, 20% ≤ medium < 60%, and high ≥60%. Using this classification, we determined that both the 'high' and 'low' classes of copies exist in highly methylated retroelement types in hPGCs, such as SVA_A, L1PA3, and LTR12C (**Figure 1F–H**).

The presence of ZNF28- and ZNF257-binding motifs are correlated with demethylation resistance in SVA_A

KRAB-ZFPs are important factors for retroelement silencing. Their activity is mediated by the recruitment of KAP1 and SETDB1, which induces retroelement DNA methylation (**Matsui et al., 2010**). To investigate whether KRAB-ZFPs could be involved in the DNA demethylation resistance of SVAs, we reanalyzed the binding peak data of 250 KRAB-ZFPs identified by ChIP-exo using exogenously tagged KRAB-ZFPs in human HEK293T cells (**Helleboid et al., 2019; Imbeault et al., 2017**). We observed that the ZNF257 and ZNF28 peaks overlapped more frequently with highly methylated SVA_A copies than with lowly methylated copies (**Figure 2A**). Because peaks of ZNF611 and ZNF91, which interact with SVAs in human embryonic stem cells (hESCs) (**Haring et al., 2021; Jacobs et al., 2014**), were observed in both lowly and highly methylated SVA_A copies (**Figure 2A**), it is unlikely that these two KRAB-ZFPs contribute to the differences in DNA methylation among SVA_A copies. Of the 'high' SVA_A elements, 63.8% and 44.7% were bound by ZNF257 or ZNF28, respectively. However, no 'low' SVA_A showed binding (**Figure 2A**), and both 'high' and 'medium' SVA_A copies significantly overlapped with the ZNF257- or ZNF28-binding peaks (**Figure 2—figure supplement 1A**). The frequency of overlap with the ZNF257/28 peaks and the enrichment of ZNF257/28 in SVA_A were positively correlated with DNA methylation (**Figure 2B and C**), and both ZNF257 and ZNF28 showed the highest enrichment of SVA_A when SVA family members were compared (**Figure 2D**).

It is possible that the correlation between the DNA demethylation resistance of SVA_A and the binding potential of specific KRAB-ZNFs based on ChIP-exo mapping data in HEK293T cells could result from differences in read mappability. To determine the likelihood of this, we calculated the mappability of each transposon copy by virtually creating reads from the retroelements and mapping them onto the genome. Although highly methylated SVA_A copies showed greater mappability than those that were lowly methylated (**Figure 2—figure supplement 1B**), the correlation between SVA_A DNA methylation levels and enrichment for ZNF28/257 was observed even when only SVA_A copies with similar mappability (50–70%) were used for analysis (**Figure 2—figure supplement 1C**). Therefore, we concluded that the enrichment of ZNF28/257 in SVA_A in HEK293T cells is correlated with SVA_A DNA methylation levels in hPGCs.

The SVA element has a region containing variable-number tandem repeats (VNTRs) in the middle segment. SVA_A contains one type of VNTR (VNTR1), whereas the other SVA classes possess two

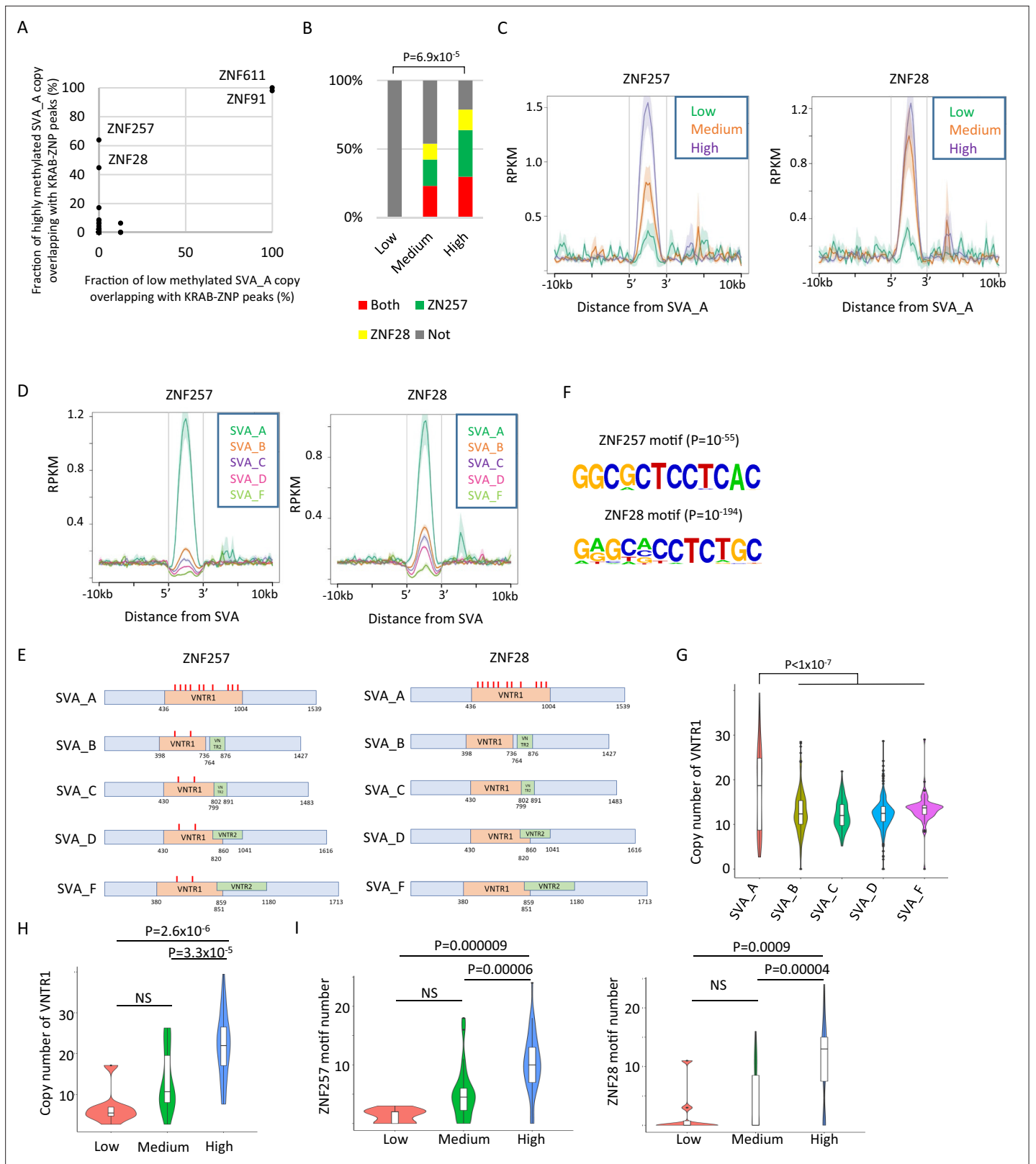


Figure 2. Identification of Krüppel-associated box domain zinc finger proteins (KRAB-ZFPs) associated with DNA demethylation resistance in SINE-VNTR-Alus (SVAs). **(A)** Scatter plots showing the fraction of low-methylated or highly methylated SVA_A copies which overlaps of KRAB-ZFP peaks. ZNF257 and ZNF28 peaks were more frequently overlapped with ‘high’ methylated SVA_A than ‘low’ methylated SVA_A. For this analysis, publicly available ChIP-exo data from 250 human KRAB-ZFPs in HEK293T cells were used. **(B)** Bar graphs showing the fraction of SVA_A copies with ZNF257 and ZNF28 peaks. **(C)** RPKM tracks for ZNF257 and ZNF28. **(D)** RPKM tracks for SVA_A variants. **(E)** Schematic diagrams of SVA_A, SVA_B, SVA_C, SVA_D, and SVA_F variants, showing the location of VNTR1 and VNTR2 regions with coordinates. **(F)** Motif logos for ZNF257 motif ($P=10^{-55}$) and ZNF28 motif ($P=10^{-194}$). **(G)** Violin plot showing the copy number of VNTR1 for SVA_A, SVA_B, SVA_C, SVA_D, and SVA_F. A p-value of $P < 1 \times 10^{-7}$ is indicated. **(H)** Violin plot showing the copy number of VNTR1 for Low, Medium, and High methylation levels. P-values: $P=2.6 \times 10^{-6}$ (Low vs High), $P=3.3 \times 10^{-5}$ (Medium vs High), NS (Low vs Medium). **(I)** Violin plot showing the ZNF257 motif number for Low, Medium, and High methylation levels. P-values: $P=0.000009$ (Low vs High), $P=0.00006$ (Medium vs High), NS (Low vs Medium). **(J)** Violin plot showing the ZNF28 motif number for Low, Medium, and High methylation levels. P-values: $P=0.0009$ (Low vs High), $P=0.00004$ (Medium vs High), NS (Low vs Medium).

Figure 2 continued

ZNF28 peaks. SVA_A copies were classified by DNA methylation levels in 19 W human primordial germ cells (hPGCs) (N = 8, 26, 47 in low, medium, and high, respectively). p-Value was calculated by chi-square test. (C) Enrichment of ZNF257 and ZNF28 on SVA_A classified by DNA methylation levels in male hPGCs at 19 weeks of gestation. (D) Enrichment of ZNF257 and ZNF28 on each SVA subtype. (E) Position of ZNF257- and ZNF28-binding motifs along SVA consensus sequences. VNTR1 and VNTR2 are composed of multiple copy number of tandem repeats, and the copy number of these number tandem repeats (VNTRs) is highly variable among SVA copies. Both ZNF257- and ZNF28-binding motifs were found within VNTR1 of SVAs. (F) Sequence logo of ZNF257- and ZNF28-binding motifs. (G) Violin plots showing copy number of VNTR1 of each SVA subtype. (H) Violin plots showing VNTR1 copy number of SVA_A classified by its DNA methylation status in male hPGCs at 19 weeks of gestation. (I) Violin plots showing the number of ZNF257 and ZNF28 motifs in SVA_A classified by DNA methylation status in male hPGCs at 19 weeks of gestation. p-Value was calculated by Tukey's test.

The online version of this article includes the following source data and figure supplement(s) for figure 2:

Source data 1. Raw data of graphs in **Figure 2**.

Figure supplement 1. Correlation between ZNF257/ZNF28-binding motifs and DNA methylation levels in SVA_A copies.

Figure supplement 2. Expression of Krüppel-associated box domain zinc finger proteins (KRAB-ZFP) genes in somatic cells and human primordial germ cells (hPGCs).

types of VNTRs (VNTR1 and VNTR2) (**Figure 2E**). The ZNF257- and ZNF28-binding motifs, which were predicted by HOMER (*Heinz et al., 2010; Figure 2F*), are in VNTR1 (**Figure 2E, Figure 2—figure supplement 1D**). The number of ZNF257- and ZNF28-binding motifs within SVAs was the highest in SVA_A (**Figure 2E**) and was most strongly correlated with the copy number of VNTR1 in SVA_A out of all the SVA classes (**Figure 2G**). The VNTR1 copy number was also highly variable among SVA_A copies (**Figure 2G**), and DNA methylation of SVA_A was positively correlated with the VNTR1 copy number (**Figure 2H, Figure 2—figure supplement 1E**) and the number of ZNF257/28 motifs (**Figure 2I, Figure 2—figure supplement 1F**). We also confirmed that DNA methylation levels within the VNTR were correlated with ZNF257 or ZNF28 association (**Figure 2—figure supplement 1G, H**). These results indicate that a high number of ZNF257- and ZNF28-binding motifs within the VNTR increases the enrichment of KRAB-ZFPs. This might contribute to maintaining SVA_A DNA methylation during hPGC development. We confirmed the RNA expression of ZNF257 and ZNF28 in hPGCs by reanalysis of single-cell RNA-seq data from hPGCs and neighboring somatic cells (*Guo et al., 2015; Figure 2—figure supplement 2A, B*). However, there was no direct evidence for ZNF28/257 protein expression and its binding to SVAs in hPGCs, which warrants further studies.

The presence of the ZNF649-binding motif is correlated with demethylation resistance in L1s

We also analyzed the correlation between KRAB-ZFP-binding motifs and the DNA methylation status of L1s and LTR12s in hPGCs. Consistent with previous reports that ZNF649 and ZNF93 bind L1s (*Cosby et al., 2019; Jacobs et al., 2014*), ZNF649 and ZNF93 peaks were frequently found in L1PA2–6 and L1PA3–6, respectively (**Figure 3A**), and these two KRAB-ZFPs were enriched at the 5' terminus of the L1 sequences (**Figure 3B**). The frequency of L1 copies overlapping with ZNF649 and ZNF93 peaks was correlated with the DNA methylation levels of L1s in hPGCs (**Figure 3C, Figure 3—figure supplement 1A**). Because read mappability in L1 (L1PA4) was similar across the different DNA methylation groups (**Figure 3—figure supplement 1B**), ZNF649 and ZNF93 are candidate factors for the DNA demethylation resistance of these L1s. As was the case for SVA_A, the presence of ZNF649- or ZNF93-binding motifs (**Figure 3D**) was also correlated with DNA methylation levels (**Figure 3E**).

Reanalysis of single-cell RNA-seq data for hPGCs and neighboring somatic cells (*Guo et al., 2015*) showed that both ZNF649 and ZNF93 were expressed more in hPGCs than in neighboring somatic cells (**Figure 2—figure supplement 2A, B**). Because the correlation between the presence of binding motifs and DNA methylation levels was stronger in ZNF649 than in ZNF93 (**Figure 3E**), we investigated ZNF649 in more detail. The ZNF649-binding motif was located at the 5' UTR of L1s (**Figure 3F**), consistent with the enrichment of ZNF649 in the 5' UTR (**Figure 3B**). The enrichment of ZNF649 in L1s was decreased in L1PA2 and abrogated in L1HS (**Figure 3B**). Along with the decreased ZNF649 enrichment, a base substitution at the fifth position of the ZNF649-binding site was observed in the consensus sequences of L1HS (**Figure 3F**). Because the fifth position of the ZNF649-binding site (T) is conserved in highly methylated L1 copies (**Figure 3G**), a T in this position may be required for ZNF649 to bind to L1. Although highly methylated L1 copies had two mismatches within the ZNF649-binding motif, one at the third position (T→G) and one at the sixth position (A→T) (**Figure 3G**), a minor fraction

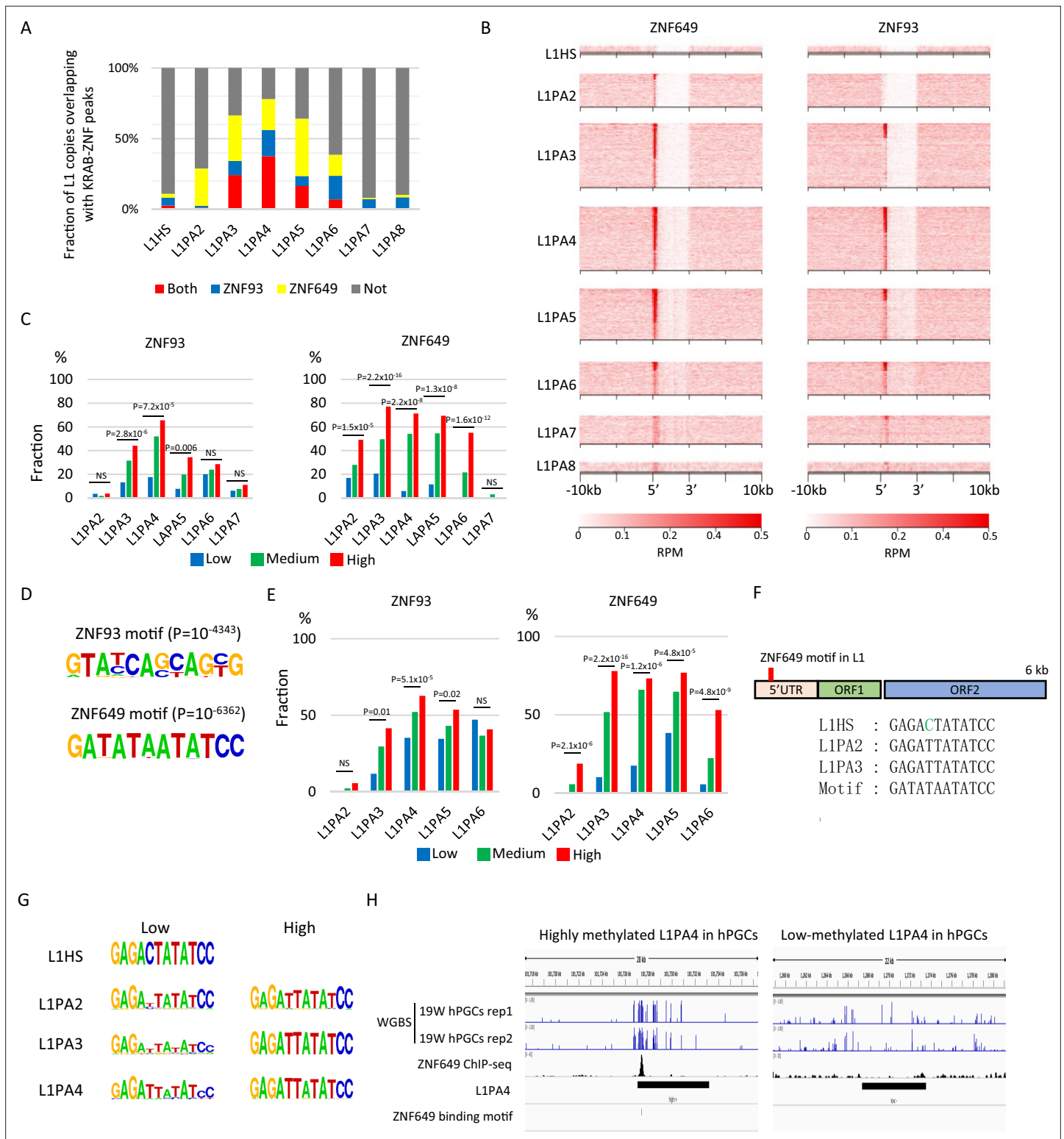


Figure 3. Identification of Krüppel-associated box domain zinc finger proteins (KRAB-ZFPs) associated with DNA demethylation resistance in L1. **(A)** Bar graphs showing the fraction of full-length L1 copies with ZNF93 and ZNF649 peaks. **(B)** Heatmaps showing enrichment of ZNF649 and ZNF93 along full-length L1 copies. ZNF649 binds 5' regions of L1PA2–PA8, while ZNF93 binds the 5' regions of L1PA3–PA8. **(C)** Bar graphs showing the fraction of L1 copies with ZNF93 and ZNF649 peaks. L1 copies were classified by their type and DNA methylation levels in male human primordial germ cells (hPGCs) at 19 weeks of gestation. **(D)** Sequence logo of ZNF93- and ZNF649-binding motifs. **(E)** Bar graphs showing the fraction of L1 copies with ZNF93- and ZNF649-binding motifs. The presence of ZNF93- and ZNF649-binding motifs was correlated with higher DNA methylation of L1 in male hPGCs at

Figure 3 continued on next page

Figure 3 continued

19 weeks of gestation (L1PA2 and -PA6 were not significant for ZNF93). p-Value was calculated by Hypothesis Testing for the Difference in the Population Proportions using a function of `prop.test` by R. (F) Comparison of sequences of ZNF649-binding sites among L1 types. L1HS lost the ZNF649 motif by a base substitution. (G) Comparison of sequences at ZNF649-binding sites between low- and high-methylated L1. (H) Representative view of correlation between DNA methylation of L1PA4 in hPGCs and ZNF649-binding peak.

The online version of this article includes the following source data and figure supplement(s) for figure 3:

Source data 1. Raw data of graphs in **Figure 3**.

Figure supplement 1. Enrichment of ZNF649/93 on L1PA4.

Figure supplement 2. ZNF850 association is correlated with DNA demethylation resistance of LTR12 family in human primordial germ cells (hPGCs).

of the ZNF649-binding motif had the same base composition at these sites (**Figure 3D**). Thus, these two mismatches may not abrogate ZNF649 binding. We also confirmed high DNA methylation in the ZNF649-binding motifs at individual loci (**Figure 3H**).

The presence of the ZNF850-binding motif is correlated with demethylation resistance in LTR12C

For the LTR12C family, we found that ZNF850 more frequently overlapped with highly methylated LTR12C/D/E copies than lowly methylated ones when we analyzed the binding peak data for 250 KRAB-ZFPs (*Helleboid et al., 2019; Imbeault et al., 2017; Figure 3—figure supplement 2A, B*). We focused on LTR12C because it had the highest analyzable copy number (LTR12C: 2054; LTR12D: 130; LTR12E: 46 copies). The ZNF850-binding motif was more frequently found in highly methylated LTR12C copies than in lowly methylated copies (**Figure 3—figure supplement 2C**). Two high-confidence binding motifs (q-value < 0.01) were identified at the 5' portion of LTR12C consensus sequences (**Figure 3—figure supplement 2D**), which was consistent with ZNF850 enrichment in the 5' portion of LTR12C (**Figure 3—figure supplement 2E**). Lowly methylated LTR12C copies contained mismatches more frequently at the eighth and tenth positions of the first and second predicted binding sites along LTR12C, respectively (**Figure 3—figure supplement 2F**). An example of highly methylated LTR12C loci with a ZNF850 peak is shown in **Figure 3—figure supplement 2G**. These data suggest that KRAB-ZNFs prevent DNA demethylation during male germ-cell development.

The mode of DNA methylation acquisition during spermatogenesis varies depending on retroelement type

To investigate whether lowly methylated retroelements in hPGCs acquire DNA methylation during spermatogenesis, we analyzed publicly available human sperm WGBS data from two donors (*Hammoud et al., 2014*). The two donors were of similar age (donor 1 – 32 and donor 2 – 37), and both were white Caucasians. The dynamics of DNA methylation in retroelements during spermatogenesis vary depending on retroelement type and individual characteristics. Most L1 copies acquired DNA methylation during spermatogenesis in both individuals, whereas LTR12C copies maintained their DNA methylation status in hPGCs during spermatogenesis (**Figure 4A**). A substantial difference between individuals was observed in the SVAs. The majority of SVA copies acquired DNA methylation during spermatogenesis in sperm donor 1, but not in donor 2 (**Figure 4A**). To evaluate these trends more efficiently, we classified retroelement copies based on DNA methylation levels in sperm (common high: > 60% in both donors; high and low: > 60% in donor 1 and < 20% in donor 2; common low: < 20% in both donors). The majority of lowly methylated L1 copies in hPGCs were highly methylated in sperm cells from both donors (**Figure 4B**). In contrast, most LTR12C/D copies maintained their PGC DNA methylation status during spermatogenesis (**Figure 4C**). Among the SVA types, SVA_A showed high levels of DNA methylation in both sperm donors, whereas other SVA types showed variable DNA methylation levels when both sperm donors were compared (**Figure 4D**), especially in SVA copies that had low DNA methylation levels in hPGCs (**Figure 4E**).

The degree of DNA methylation acquisition during spermatogenesis varies among SVA copies

Although the DNA methylation status of SVAs was highly variable between the sperm donors, a subset of SVA copies acquired DNA methylation or maintained a low methylation state during

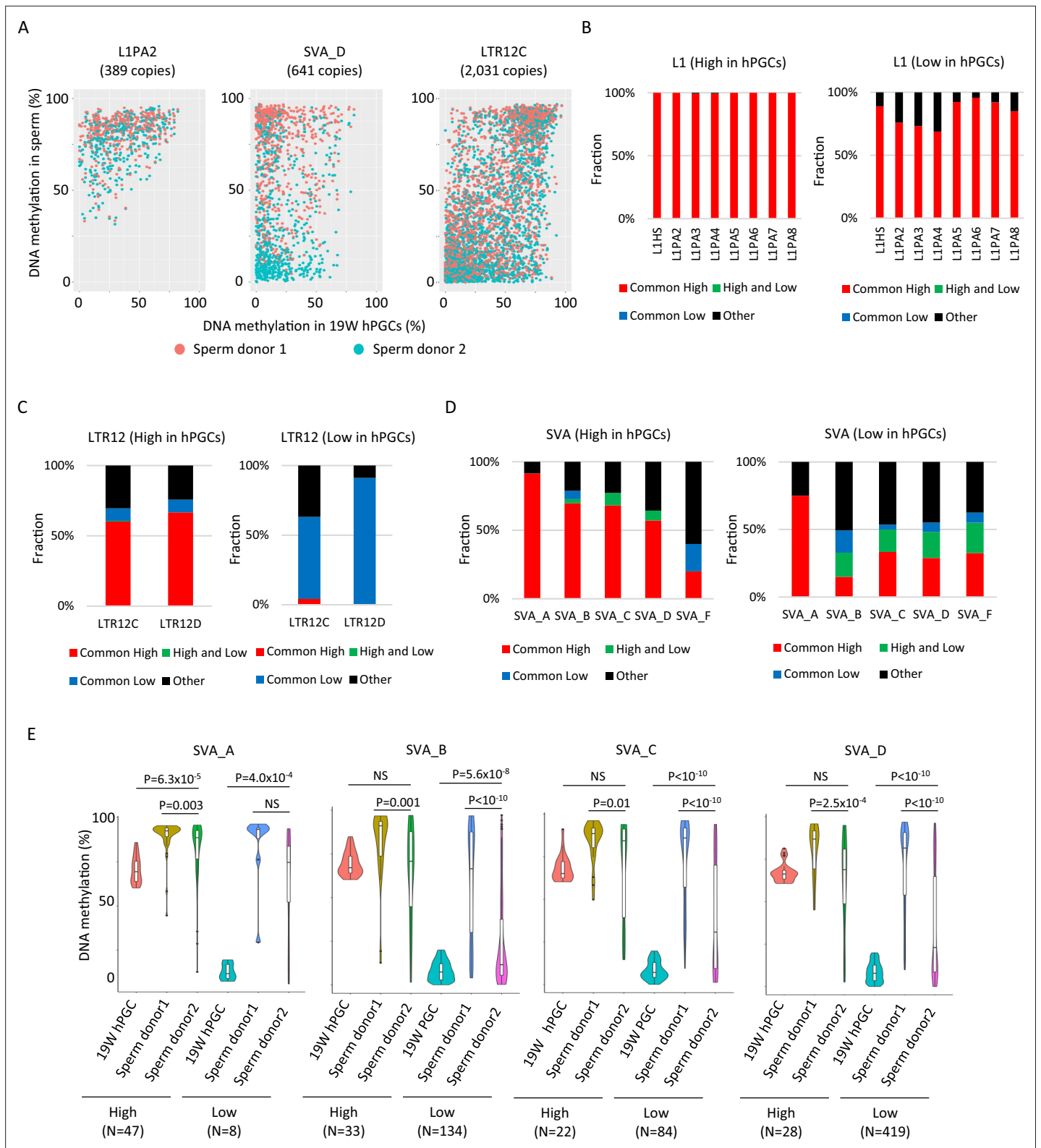


Figure 4. DNA methylation dynamics of retroelements during human spermatogenesis. **(A)** Scatter plots showing DNA methylation levels of each retroelement copy in male human primordial germ cells (hPGCs) at 19 weeks of gestation and sperm. Whole-genome bisulfite sequencing (WGBS) data from two sperm donors (*Hammoud et al., 2014*) were used for this analysis. Donor 1 and donor 2 were colored by orange and cyan, respectively. **(B–D)** Bar graphs showing the fraction of groups determined by DNA methylation patterns in two sperm donors in L1 **(B)**, LTR12 **(C)**, and SINE-VNTR-*Alus* (SVA) **(D)**. ‘Other’ indicates groups except for common high, common low, and high and low, such as low methylated in donor 1 and mediumly

Figure 4 continued on next page

Figure 4 continued

methylated in donor 2. Bar graphs were also separated by DNA methylation levels (high or low) in male hPGCs at 19 weeks of gestation. (E) Violin plots showing DNA methylation levels of SVA copies in male hPGCs at 19 weeks of gestation, sperm donor 1, and sperm donor 2. The violin plots were also separated by DNA methylation levels of SVA copies in male hPGCs at 19 weeks of gestation. Although hypomethylated SVA copies in male hPGCs at 19 weeks of gestation acquired DNA methylation during spermatogenesis, the degree of DNA methylation increase was significantly different between sperm donors. p-Value was calculated by Dunnett's test.

The online version of this article includes the following source data for figure 4:

Source data 1. Raw data of graphs in **Figure 4**.

spermatogenesis in both sperm donors (**Figure 4D**). It is possible to get insight for mechanisms of de novo DNA methylation of SVAs during spermatogenesis by comparing SVAs that acquired DNA methylation ('low' in hPGCs and 'common high' in sperm) to SVAs that maintained hypomethylation ('low' in hPGCs and 'common low' in sperm) in both sperm donors. Phylogenetic analysis of 'common low' and 'common high' SVA copies (SVA_B and _D) showed that these two classes were not genetically separated (**Figure 5A**), indicating that the acquisition of DNA methylation in SVAs during spermatogenesis is not genetically determined.

The presence of transcription-directed retroelement silencing mechanisms, such as the PIWI/piRNA pathway (**Watanabe et al., 2018**), prompted us to investigate the correlation between the genomic distribution of SVA copies and DNA methylation. Approximately half of the SVA_B–F copies were inserted into the gene body, and most of them were in the antisense direction (**Figure 5B**). 'Common high' SVA_B–F copies were enriched in the gene body in the antisense direction, while 'common low' SVA_B–F copies were depleted from the gene body (**Figure 5B**). Reanalysis of publicly available single-cell RNA-seq data in human testes (**Sohni et al., 2019**) revealed that genes with 'common high' SVA_B–F copies in the antisense orientation showed greater expression in spermatogonial stem cells relative to genes with 'common low' (**Figure 5C**). Therefore, SVAs located in actively transcribed regions in the antisense orientation are efficiently subjected to de novo DNA methylation during spermatogenesis. The expression of genes with 'high and low' SVA_B–F copies in the antisense direction was higher in spermatogonial stem cells than the expression of other randomly extracted genes and genes with 'common low' SVA_B–F copies. However, the expression of these genes was lower than the expression of genes with 'common high' SVA_B–F copies (**Figure 5C**). Approximately half of the 'high and low' SVA_B–F copies were located in non-genic regions, but RNA-seq reads from previously reported undifferentiated spermatogonia (**Tan et al., 2020**) mapped more frequently around the non-genic 'high and low' SVA_B–F copies than the 'common low' B–F copies (**Figure 5D**). Therefore, non-genic 'high and low' SVA_B–F copies are frequently inserted in transcribed regions during spermatogenesis. These results implicate the possibility that SVAs acquire DNA methylation during DNA methylation via transcription-directed machinery, and that the effectiveness of de novo DNA methylation varies among individuals.

SVAs are a potential source of inter-individual epigenetic variation in sperm

Inter-individual variation in DNA methylation in SVAs was also observed when another set of publicly available sperm WGBS data from three Japanese donors was analyzed (**Okae et al., 2014; Figure 6A**). For additional validation, we performed amplicon sequencing (amplicon-seq) of bisulfite PCR products for SVAs on sperm from five Japanese donors (**Figure 6B**). Our amplicon-seq yielded approximately 1.7–2.2 M read pairs and measured the DNA methylation level of over 90% of the full-length SVA_B–F copies (minimum read depth of CpG ≥ 5 , analyzed CpG number ≥ 10) (**Figure 6C**). Again, 'high and low' SVA_B–F copies showed variations in DNA methylation among donors (**Figure 6D**). Thus, inter-individual variation in SVA methylation in sperm is a common phenomenon and is not ethnically specific.

To estimate the impact of SVAs on inter-individual epigenetic variations in sperm, we identified differentially methylated regions (DMRs) in two sperm donors, as shown in **Figures 4 and 5; Hammoud et al., 2014**. Although the DNA methylation profiles between the two donors were highly correlated (**Figure 6E**), 2008 regions were identified as DMRs (donor 1 < donor 2: 332, donor 2 < donor 1: 1676). Of the 1676 donor 1-specific methylated DMRs, 772 (46.1%) overlapped with SVAs (**Figure 6F**). We also observed differential DNA methylation among individuals in SVA-associated DMRs in our

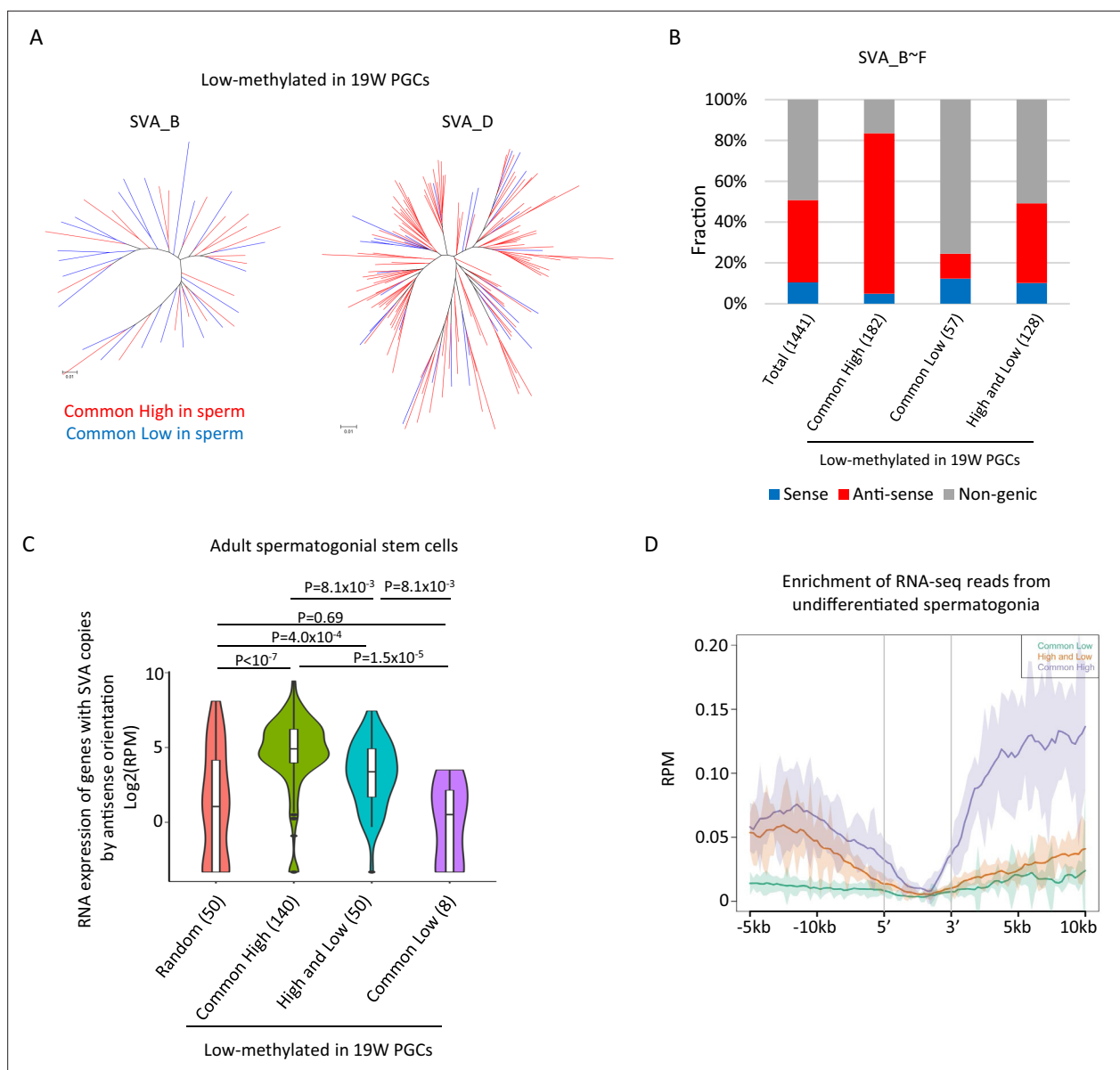


Figure 5. Transcription-associated regulation of DNA methylation of SINE-VNTR-Alus (SVA) during spermatogenesis. **(A)** Phylogenetic analysis of SVA_B (left) and SVA_D (right) copies low methylated in male human primordial germ cells (hPGCs) at 19 weeks of gestation. SVA copies highly methylated by both sperm donors were colored by red, while those hypomethylated by both sperm donors were colored by blue. **(B)** Bar graphs showing the fraction of SVA_B–F copies inserted in a gene body. SVA copies were classified by DNA methylation patterns in two sperm donors. Only low-methylated SVA copies in male hPGCs at 19 weeks of gestation were used for this analysis. The number in parentheses represents analyzed copy number. **(C)** Violin plots showing the expression of genes in adult spermatogonial stem cells 2 (Sohni et al., 2019). Genes were classified according to the DNA methylation status of SVAs inserted in them in the antisense direction. p-Value was calculated by Tukey's test. **(D)** Enrichment of RNA-seq reads from undifferentiated spermatogonia (Tan et al., 2020) around non-genic SVAs. Only low-methylated SVA copies male hPGCs at 19 weeks of gestation were used for the analysis, and SVA copies were classified by DNA methylation patterns in two sperm donors (common low, high, and low and common high).

The online version of this article includes the following source data for figure 5:

Source data 1. Raw data of graphs in Figure 5.

amplicon-seq data (Figure 6G). Therefore, SVAs significantly contribute to inter-individual variations in the sperm epigenome. In contrast to the inter-individual epigenetic variation of SVAs in sperm, a reanalysis of WGBS data of adult skeletal muscle from 15 individuals and of helper CD4-positive T cells from 18 individuals, which was deposited in the International Human Epigenome Consortium (IHEC) portal, showed high DNA methylation of SVAs in all individuals (Bujold et al., 2016; Figure 6—figure

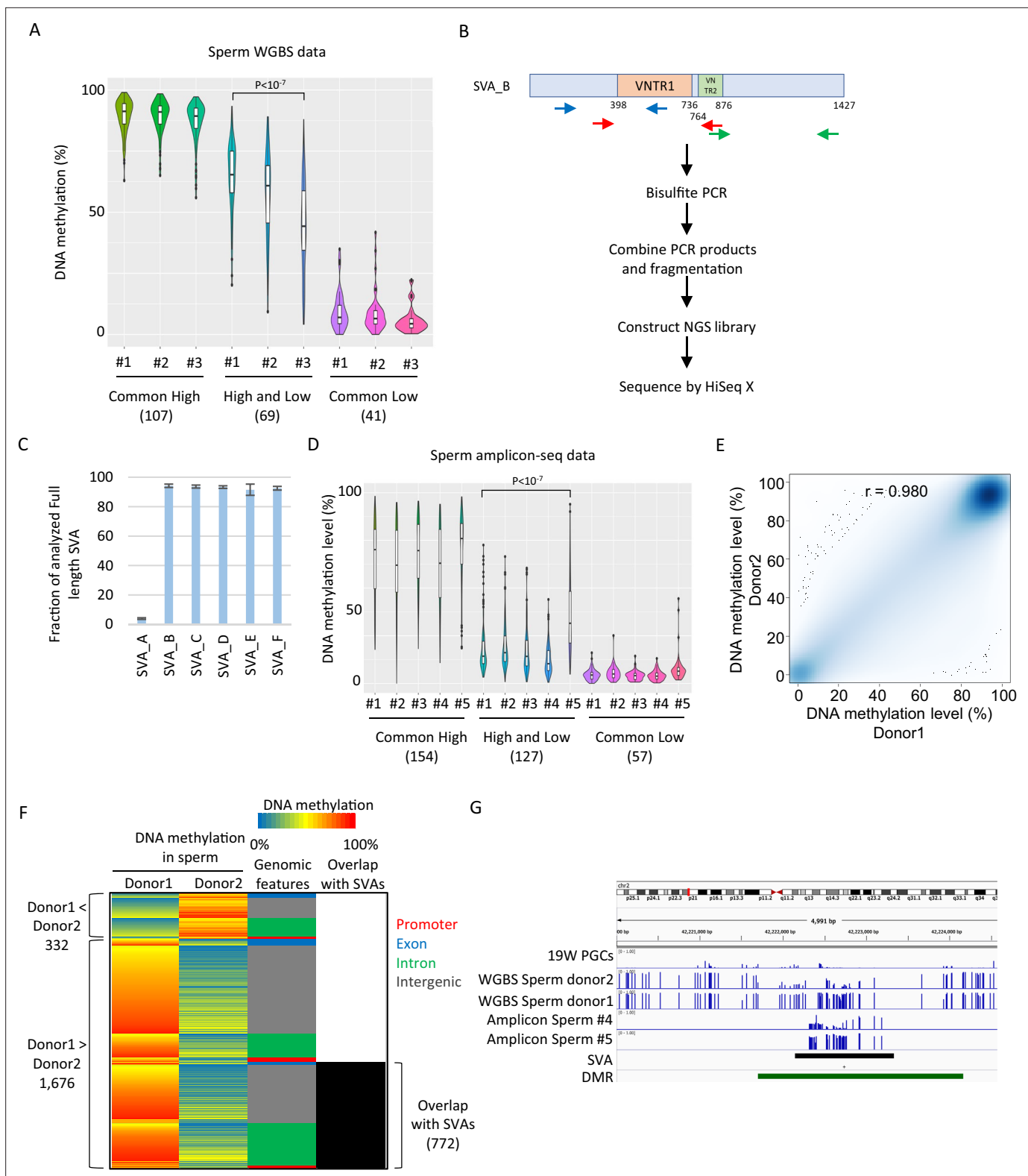


Figure 6. SINE-VNTR-*Alus* (SVAs) constitute a major source of inter-individual epigenetic variations in sperm. **(A)** Violin plots showing DNA methylation of SVA copies in previously reported three sperm donors (#1–#3) (Okae et al., 2014). Only low-methylated SVA copies in male human primordial germ cells (hPGCs) at 19 weeks of gestation were used for the analysis. SVA copies were classified by DNA methylation levels of two sperm donors from Hammoud et al., 2014. Donor #1 showed significantly higher DNA methylation levels in ‘high and low’ SVA copies than other sperm donors. p-Value was calculated by Tukey’s test. **(B)** Scheme of amplicon sequencing (amplicon-seq) for analyzing SVA methylation. **(C)** Bar plots showing the

Figure 6 continued on next page

Figure 6 continued

fraction of analyzed full-length SVA copies by amplicon-seq. (D) Violin plots showing DNA methylation levels of SVA copies in five sperm donors from amplicon-seq. Only low-methylated SVA copies in male hPGCs at 19 weeks of gestation were used for the analysis. SVA copies were classified by DNA methylation levels of two sperm donors from Hammoud et al. Donor #5 showed significantly higher DNA methylation levels in 'high and low' SVA copies than other sperm donors. p-Value was calculated by Tukey's test. (E) Scatter plot showing the DNA methylation between sperm donor 1 and sperm donor 2 from Hammoud et al. DNA methylation levels between these two donors were highly correlated. (F) Heatmap showing DNA methylation levels, genomic distribution, and overlap with SVAs of differentially methylated regions (DMRs). (G) Representative view of DMRs overlapping with SVA. Black and green boxes represent SVA and DMR, respectively.

The online version of this article includes the following source data and figure supplement(s) for figure 6:

Source data 1. Raw data of graphs in **Figure 6**.

Figure supplement 1. Inter-individual epigenetic variation in SINE-VNTR-*Alus* (SVAs) was correlated with testosterone levels in bloods.

supplement 1A, B). Thus, inter-individual variations in DNA methylation of SVAs in sperm are essentially canceled during development.

Finally, to investigate whether inter-individual DNA methylation variations are associated with physiological or disease conditions, we reanalyzed publicly available WGBS data from five healthy donors and six oligozoospermic patients (European Nucleotide Archive [ENA] under the accession number PRJEB34432) (Leitão et al., 2020). The disease condition was not associated with inter-individual DNA methylation variations in SVAs, because both healthy donors and oligozoospermic patients showed inter-individual variations of DNA methylation in 'high and low' SVA_{B-F} copies (**Figure 6—figure supplement 1C**). On the other hand, a comparison of various physiological conditions between highly methylated individuals and lowly methylated ones (median methylation levels of 'high and low' > 50% vs. < 50%) revealed that blood testosterone levels were significantly higher in lowly methylated individuals than in highly methylated ones (**Figure 6—figure supplement 1D**). However, prolactin, follicle stimulating hormone, luteinizing hormone (LH), sex hormone-binding globulin blood levels, and age were not significantly different between the two groups (**Figure 6—figure supplement 1D**). Although further validation of this correlation is required, DNA methylation of SVAs in sperm may be associated with physiological conditions.

Discussion

In this study, we showed that the binding potential of KRAB-ZFPs correlates with retroelement DNA demethylation resistance in hPGCs. Furthermore, we found that de novo DNA methylation patterns in spermatogenesis vary among the L1, LTR, and SVA retroelements. In addition, we ascertained that the SVAs located in transcription-active regions in the antisense orientation are prone to methylation during spermatogenesis, which implies that the transcription-directed DNA methylation machinery might contribute to de novo DNA methylation of SVAs in male germ cells. Notably, the extent of de novo DNA methylation of SVAs in male germ cells is variable among human individuals, with SVAs being a major source of epigenetic variation in sperms.

We showed that DNA demethylation resistance in hPGCs frequently occurred in moderately young retroelements such as L1PA, SVA_A, and LTR12, but not in currently active retroelements. Because we targeted full-length transposons, our analysis was biased toward relatively young transposons. Thus, it is possible that some fragmented older transposons may also be resistant to DNA demethylation in hPGCs. A subset of LTR transposons, including LTR12, function as enhancers (Deniz et al., 2020). It was recently reported that LTR5s, which are Hominidae-specific LTR-type transposons and hypomethylated in hPGCs (DNA methylation levels < 10%), can function as enhancers to promote hPGC differentiation (Xiang et al., 2022). Therefore, in the case of LTR12C, maintaining DNA methylation might be beneficial for hPGC development because it suppresses inappropriate activation of transposon-embedded enhancer function.

In addition, KRAB-ZFP binding potentially contributed to the DNA demethylation resistance of L1s and SVAs. ZNF257/28, ZNF649/ZNF93, and ZNF850 were associated with the DNA demethylation resistance of SVAs, L1s, and LTR12Cs, respectively (**Figure 7**). In hPGCs, multiple KRAB-ZFPs were correlated with DNA demethylation resistance in the same retroelements, which may contribute to more robust or cooperative retroelement suppression. Although ZNF91 reportedly binds to the VNTR in SVAs and silences SVA expression in embryonic stem cells (ESCs) (Haring et al., 2021; Jacobs

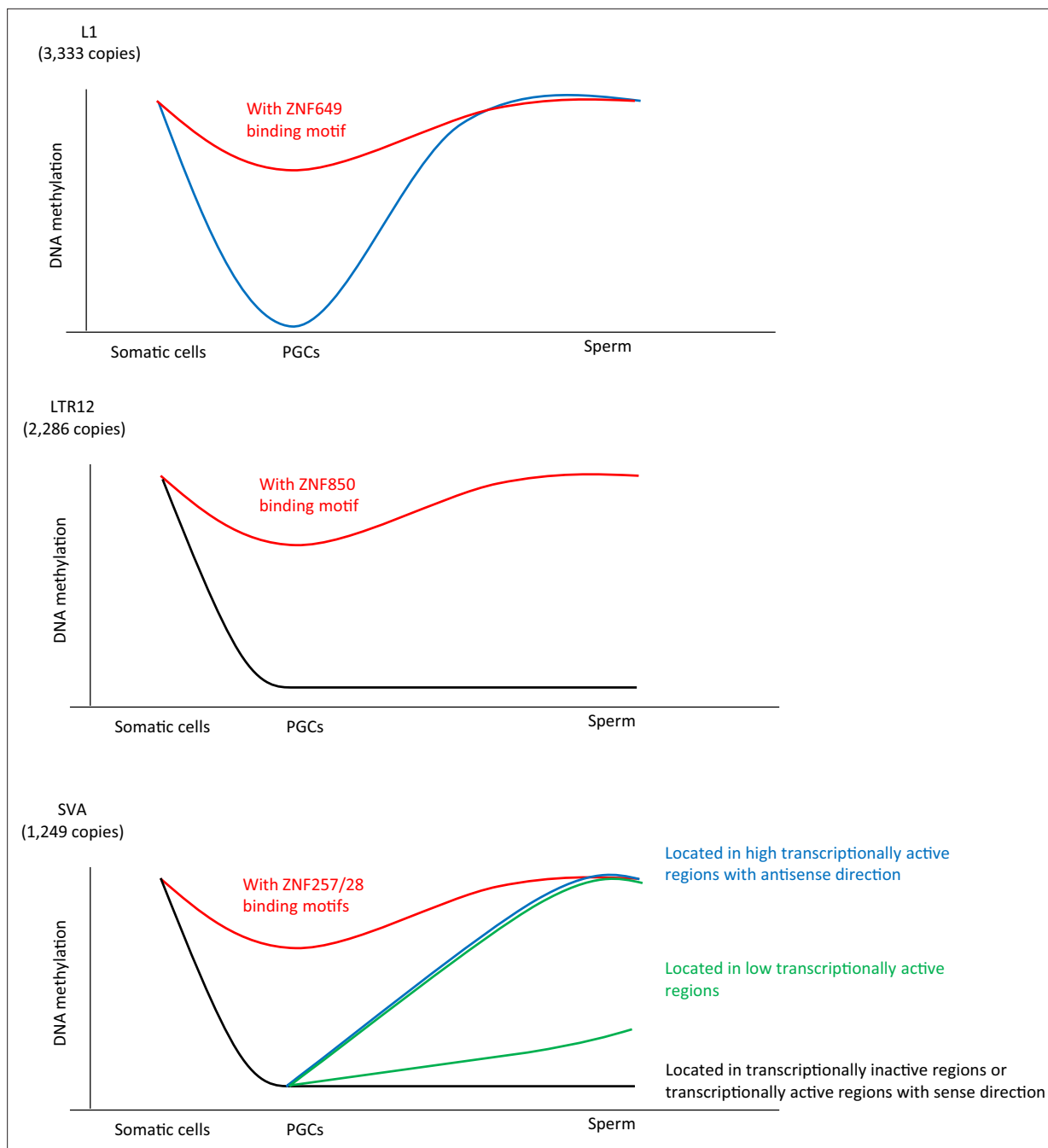


Figure 7. Summary of this study. Our data demonstrated an association between KRAB-ZFP-binding motifs and the DNA demethylation resistance of L1, SVA, and LTR12. ZNF649, ZNF257/28, and ZNF850 were associated with DNA demethylation resistance of L1, SVA, and LTR12C, respectively. The dynamics of DNA methylation during spermatogenesis are largely different among retroelement types. The majority of L1 copies acquired DNA methylation during spermatogenesis, whereas the DNA methylation status of LTR12 in human primordial germ cells (hPGCs) tended to be maintained during spermatogenesis. The mode of DNA methylation changes in SINE-VNTR-*Alus* (SVAs) during spermatogenesis largely differs between copies and individuals. SVA copies located in highly transcriptionally active regions acquire DNA methylation during spermatogenesis, while those located in transcriptionally inactive regions maintain a hypomethylated state during spermatogenesis. In contrast, the degree of DNA methylation in sperm in SVA copies located in low transcriptionally active regions was highly variable among the individuals. These results suggest that SVAs may be methylated by transcription-directed DNA methylation mechanisms during spermatogenesis, and their activity varies among individuals.

et al., 2014), the DNA demethylation resistance of SVAs did not correlate with ZNF91 binding, indicating that a different KRAB-ZFP set is used to suppress SVAs in human PGCs and ESCs. Both ZNF257 and ZNF28 bound to VNTR1 (**Figure 2E**), and high copy numbers of VNTR1 were correlated with high ZNF257 and ZNF28 enrichment and DNA methylation (**Figure 2H1**). The reduction in VNTR1 copy number after SVA_B emergence (**Figure 2G**) may have been necessary for SVAs to escape silencing mechanisms in hPGCs. The human genome encodes at least 350 KRAB-ZFPs, and not all KRAB-ZFPs were included in the ChIP-seq dataset used in this study (100 copies remained unmapped). Thus, the involvement of other KRAB-ZFPs in DNA demethylation resistance of retroelements in hPGCs is possible. Although we observed a strong correlation between KRAB-ZFPs and DNA demethylation resistance, direct evidence for this correlation remains elusive because of the limited availability of human fetal gonads and of high-specificity antibodies for KRAB-ZFPs. Because they can function as in vitro derivation systems, PGC-like cells (PGCLCs) may be a promising model for investigating the biology of PGCs. Although successful establishment of human PGCLCs has been reported (*Sasaki et al., 2015*), sufficient DNA demethylation has not been observed in human PGCLCs (*von Meyenn et al., 2016*). Thus, the currently available human PGCLCs are not suitable models for investigating the mechanisms of DNA demethylation resistance. Optimizing the derivation conditions for human PGCLCs will aid in our understanding of retroelement silencing in PGCs.

Additionally, we showed that the mode of DNA methylation acquisition during spermatogenesis was very different among retroelement types. The majority of L1 copies acquired DNA methylation during spermatogenesis, whereas LTR12 maintained its DNA methylation status in hPGCs during spermatogenesis (**Figure 7**). L1HS, which both ZNF93 and ZNF649 were unable to bind, also acquired DNA methylation during spermatogenesis (**Figure 4B**), suggesting the involvement of other factors in the de novo DNA methylation of L1 during spermatogenesis. The PIWI-piRNA pathway is responsible for the DNA methylation of L1 transposons in mouse male germ cells (*Aravin et al., 2007; Carmell et al., 2007; Kojima-Kita et al., 2016; Manakov et al., 2015; Shoji et al., 2009*). The PIWI-piRNA pathway may also be functional in humans, because mutations in genes associated with the PIWI-piRNA pathway are linked to human male infertility (*Arafat et al., 2017; Gu et al., 2010*). Moreover, the majority of putative piRNAs that mapped to transposons at gestational week 20 are derived from L1 (*Reznik et al., 2019*). Therefore, the PIWI-piRNA pathway is a candidate pathway for L1 silencing in human male germ cells.

Our data showed that the acquisition of DNA methylation of SVAs during spermatogenesis correlated with the inserted regions and not with the nucleotide sequence. SVAs inserted in transcriptionally active regions in the antisense direction are efficiently targeted for de novo DNA methylation during spermatogenesis. In mouse spermatogenesis, MIWI2 binds piRNAs and is recruited to the nascent transcribed regions that are complementary to piRNAs (*Watanabe et al., 2018*). Subsequently, MIWI2-interacting protein SPOCD1, which forms a complex with DNMT3A and DNMT3L, and potentially with a rodent-specific DNA methyltransferase DNMT3C (*Barau et al., 2016*), induces DNA methylation on transposons (*Zoch et al., 2020*). Therefore, one possible mechanism for the de novo DNA methylation of SVAs during spermatogenesis is that the MIWI2/SVA-derived piRNA complex targets nascent transcripts with antisense SVAs and induces DNA methylation. There are also other transcription-directed repetitive element silencing mechanisms, such as those involving the HUSH complex, which repress L1s and SVAs (*Fukuda et al., 2018; Liu et al., 2018; Robbez-Masson et al., 2018*). The HUSH complex targets young full-length L1s located within the introns of actively transcribed genes (*Fukuda and Shinkai, 2020; Liu et al., 2018*). In addition to the HUSH complex, efficient pericentromeric heterochromatin formation requires the transcription of pericentromeric satellite repeats, which stabilize SUV39H pericentromeric localization (*Johnson et al., 2017; Shirai et al., 2017; Velazquez Camacho et al., 2017*). Because SUV39H is also associated with retroelement silencing (*Bulut-Karslioglu et al., 2014*), both the HUSH complex and SUV39H are candidate factors associated with the transcription-directed DNA methylation of SVAs in human male germ cells. In eukaryotes, gene bodies are the most conserved targets of DNA methylation. Gene body DNA methylation levels are often correlated with transcriptional levels (*Teissandier and Bourc'his, 2017*). This is because of the interaction between the elongating RNA polymerase II and SETD2, which results in H3K36me3. H3K36me3 participates in the de novo methylation of DNA by recruiting DNMT3 enzymes via their chromatin reading PWWP domains (*Baubec et al., 2015; Shirane et al., 2020*). Furthermore, antisense RNAs embedded within protein-coding genes are selectively silenced

by H3K36 methyltransferase SET2 in *Saccharomyces cerevisiae* (Venkatesh et al., 2016). Thus, the machinery for gene body DNA methylation regulated by SETD2 is also a candidate for the de novo DNA methylation of SVAs during spermatogenesis.

The mechanism underlying inter-individual epigenetic variations in SVA in human sperm is unknown. In addition to genetic differences among individuals, both intrinsic and extrinsic environmental differences may contribute to inter-individual variations in SVAs. Our data indicate that in sperm, the degree of DNA methylation of SVAs located in genomic regions with low transcriptional activity varies among individuals. Thus, the effectiveness of transcription-directed de novo DNA methylation in male human germ cells may vary among individuals. Previous studies have shown that hypermethylation of the *PIWIL2* and *TDRD1* promoter regions, which are involved in the PIWI-piRNA pathway, is associated with abnormal DNA methylation and male infertility in humans (Heyn et al., 2012). Therefore, the effectiveness of the PIWI-piRNA pathway may vary among individuals and contribute to the epigenetic variation of SVAs in male germ cells. SVAs function as enhancers (Gianfrancesco et al., 2017), alter the chromatin state near the insertion site (Fukuda et al., 2017), and are associated with Fukuyama-type congenital muscular dystrophy and Lynch syndrome (Ostertag et al., 2003; Payer and Burns, 2019). Therefore, differences in SVA regulation among individuals may induce changes in gene regulation in male germ cells, alter the risk of genome instability, and affect the incidence of disease among individuals.

Materials and methods

Semen collection

Ejaculates were provided by patients who visited the Reproduction Center of the Ichikawa General Hospital, Tokyo Dental College. All study participants were briefed about the aims of the study and the parameters to be measured, and consent was obtained. The study was approved by the ethics committees of RIKEN, Tokyo University, and Ichikawa General Hospital. Sperm concentration and motility were measured using a computer-assisted image analyzer (C-Men, Compix, Cranberry Township, PA). Human semen was diluted twice with saline, layered on 5.0 mL of 20 mM HEPES buffered 90% isotonic Percoll (Cytiba, Uppsala, Sweden), and centrifuged at $400 \times g$ for 22 min. The sperm in the sediment was recovered to yield 0.2 mL, and then introduced to the bottom of 2.0 mL of Hanks' solution to facilitate swim-up. The motile sperm in the upper layer were recovered.

Preparation of SVA amplicon-seq

Genomic DNA was subjected to bisulfite-mediated C to U conversion using the MethylCode Bisulfite Conversion Kit (ThermoFisher Scientific), and then used as a template for PCR for 35 cycles with EpiTaq (Takara) using the following primers: SVA_1_Fw TTATTGTAATTTTTTTGTTTGATTTTTTTTGT TTAG. SVA_1_Rv AAAAAAACTCCTCACATCCCAAAC SVA_2_Fw TTAATGTTGTTTAGGTTGGAGTGT AGTG SVA_2_Rv CAAAAAACTCCTCACTTCCCAATA. SVA_3_Fw TTTGGGAGGTGTATTTAATAGTTT ATTGAGAA SVA_3_Rv TAAACAAAAATCTCTAATTTTCCTAAACAAAAAAC. The PCR products from three sets of primers were combined, purified using a MinElute PCR Purification Kit (QIAGEN), and fragmented using Picoruptor (Diagenode) for 10 cycles of 30 s on and 30 s off. Then, the amplicon-seq library was constructed using KAPA LTP Library Preparation Kits (KAPA BIOSYSTEMS) and SeqCap Adapter Kit A (Roche). The amplicon-seq libraries were sequenced on a HiSeq X platform (Illumina).

WGBS and amplicon-seq analysis

We used the hg19 version of the human genome for NGS analysis because the predicted KRAB-ZFP peaks were derived from this version. Using the newest version of the human genome (GRCh38) did not significantly affect the conclusions. The following publicly available WGBS data were used in this study: hPGCs (SRP050499), sperm (SRP028572, ERP117337, JGAS00000000006), and adult tissues (IHEC data portal). For the IHEC data, we used processed data for our analysis.

- Quality control, read mapping, and DNA methylation calculation

Low-quality bases and adaptor sequences were trimmed using Trim Galore version 0.3.7 (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). For WGBS data from hPGCs, the first nine bases were further trimmed. The trimmed reads were aligned to the hg19 genome using Bismark

v0.14.1, with paired-end and non-directional mapping parameters (--non_directional) (*Krueger and Andrews, 2011*). The unmapped reads after paired-end mapping were re-aligned to the same reference in single-end mode. We validated that this mapping mode only reported uniquely mapped reads. The methylation level of each CpG site was calculated as follows: (number of methylated reads/number of total reads). Only CpG sites with at least five reads were used for all analyses. Only nearly full-length retroelements, whose length is 90% or more of the length of the consensus sequence of each subtype, were used for DNA methylation analysis of retroelements. We also included solo-LTR transposons in the DNA methylation analysis if they also possessed more than 90% of the consensus LTR sequence. Retroelement information was obtained from the UCSC Genome Browser (<http://genome.ucsc.edu/>). For the DNA methylation analysis of retroelements, we used retroelements containing at least 10 CpG sites with a read depth of at least five reads. The methylation level of each retroelement copy was calculated by averaging the methylation levels of CpG sites within the copy.

Classification of retroelement copy according to DNA methylation levels.

Retroelement copies were classified according to their DNA methylation levels as follows: low < 20%, 20% ≤ medium < 60%, high ≥ 60%.

- Association of KRAB-ZFP peaks, binding motifs, and retroelements.

We obtained the peak regions of 250 KRAB-ZFP, which were previously reported (*Helleboid et al., 2019; Imbeault et al., 2017*), from the gene expression omnibus GSE78099 and GSE120539. Overlap of the KRAB-ZFP peak and retroelement copy was investigated using bedtools v2.15.0 (*Quinlan and Hall, 2010*). The binding motif of each KRAB-ZFP was predicted by the findMotifsGenome.pl program in Homer v4.8.3 (*Heinz et al., 2010*). The KRAB-ZFP-binding motifs along retroelement copies were searched using FIMO (*Grant et al., 2011*). We used predicted motif sites with a q-value of 0.00005 or less for ZNF257/ZNF28/ZNF850 and with a q-value of 0.05 or less for ZNF93 and ZNF649 in this study.

- DMR identification

DMR candidates were identified using the 'Commet' command in BisulFighter (*Saito et al., 2014*). To enhance the confidence of DMR call, we calculated the average methylation levels of the candidates using CpG sites with ≥5 reads in both sperm donors, and among the candidates, those containing ≥10 successive analyzable CpG sites and showing a ≥40% methylation difference were determined as DMRs.

Phylogenetic analysis of retroelement copies

The evolutionary history was inferred using the maximum likelihood method based on the Tamura-Nei model (*Tamura and Nei, 1993*). The initial tree(s) for the heuristic search were obtained by applying the neighbor-joining method to a matrix of pairwise distances estimated using the maximum composite likelihood approach. The tree was drawn to scale, with branch lengths measured as the number of substitutions per site. There were 10,153 positions in the final dataset. Evolutionary analyses were conducted using MEGA6 (*Tamura et al., 2013*).

Calculation of read mappability of each retroelement copy

We generated 100 bp reads from each position along the retroelement copy and aligned the simulated reads to the human genome using Bowtie with -m 1 or Bismark. Then, the mappability of each copy was calculated by dividing the number of properly mapped reads by the total number of reads derived from each copy.

RNA-seq analysis

We reanalyzed previously reported single-cell RNA-seq data from the testes (*Sohni et al., 2019*), hPGCs, and somatic cells next to hPGCs (*Guo et al., 2015*). Read count data of genes and cell type annotation of each cell were obtained from the Gene Expression Omnibus under accession numbers GSE124263 and GSE63818. Reads per million mapped reads (RPM) for the genes were calculated for each cell. We used the average RPM of spermatogonial stem cells 2 (*Figure 5C*). We also reanalyzed previously reported RNA-seq data from undifferentiated spermatogonia (*Tan et al., 2020*), which was deposited in the Gene Expression Omnibus under accession number GSE144085. Low-quality bases and adaptor sequences were trimmed using Trim Galore version 0.3.7. Then, trimmed reads

were aligned to the hg19 genome using Bowtie v0.12.7 with -m 1 to remove multiple mapped reads. Enrichment of RNA-seq reads around SVAs was visualized using ngsplot (*Shen et al., 2014*).

ChIP-seq analysis

We reanalyzed previously reported KRAB-ZFP ChIP-exo data (*Helleboid et al., 2019; Imbeault et al., 2017*), which were deposited in the Sequence Read Archive SRP070561 and SRP162756. Low-quality bases and adaptor sequences were trimmed using the Trim Galore version 0.3.7. Then, trimmed reads were aligned to the hg19 genome using Bowtie v0.12.7 with -m 1 to avoid multiple mapped reads. Enrichment of ChIP-exo reads around retroelements was visualized using ngsplot (*Shen et al., 2014*).

Visualization of NGS data

The Integrative Genomics Viewer (*Robinson et al., 2011*) was used to visualize the NGS data. Enrichment of RNA-seq reads and KRAB-ZFPs was visualized using ngsplot (*Shen et al., 2014*). Scatter plot and violin plot analyses were performed using the ggplot2 package in R.

Data access

All reads from amplicon-seq in this study have been submitted to the Gene Expression Omnibus under accession number GSE174562.

Acknowledgements

We thank the staff of the Support Unit for Bio-Material Analysis (BMA) at the RIKEN Center for Brain Science (CBS) Research Resources Division (RRD) for NGS library construction. This research was supported by the Special Postdoctoral Researcher (SPDR) Program of RIKEN to KF, the Japan Ministry of Education, Culture, Sports, Science, and Technology Grant-in-Aid for Scientific Research (18H05530, 18H03991) to YS, and RIKEN internal research funds to YS.

Additional information

Funding

Funder	Grant reference number	Author
Japan Society for the Promotion of Science	18H05530	Yoichi Shinkai
Japan Society for the Promotion of Science	18H03991	Yoichi Shinkai
RIKEN	SPDR	Kei Fukuda

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Kei Fukuda, Conceptualization, Data curation, Formal analysis, Investigation, Validation, Writing - original draft, Writing - review and editing; Yoshinori Makino, Satoru Kaneko, Yuki Okada, Resources; Chikako Shimura, Investigation; Kenji Ichiyanagi, Conceptualization; Yoichi Shinkai, Conceptualization, Funding acquisition, Resources, Supervision, Writing - original draft, Writing - review and editing

Author ORCIDs

Kei Fukuda  <http://orcid.org/0000-0002-4111-9409>

Yoichi Shinkai  <http://orcid.org/0000-0002-6051-2484>

Ethics

Human subjects: This study was approved by the ethics committees of RIKEN, Tokyo University, and Ichikawa General Hospital. All study participants were briefed about the aims of the study and the parameters to be measured, and consent was obtained.

Decision letter and Author responseDecision letter <https://doi.org/10.7554/eLife.76822.sa1>Author response <https://doi.org/10.7554/eLife.76822.sa2>**Additional files****Supplementary files**

- Supplementary file 1. List of full-length L1/SVA/LTR12.
- Supplementary file 2. Statistical analysis of DNA methylation levels of repeat elements in human primordial germ cells (PGCs) (related to **Figure 1C–E**).
- Transparent reporting form

Data availability

All reads from amplicon-seq in this study have been submitted to the Gene Expression Omnibus under accession number GSE174562.

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Fukuda K, Shinkai Y	2021	Amplicon-seq of SVA methylation in human sperm	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE174562	NCBI Gene Expression Omnibus, GSE174562

The following previously published datasets were used:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Imbeault M, Helleboid PY, Trono D	2017	ChIP-exo of human KRAB-ZNFs transduced in HEK 293T cells and KAP1 in hES H1 cells	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE78099	NCBI Gene Expression Omnibus, GSE78099
Guo F, Guo H, Li L, Tang F	2015	The Transcriptome and DNA Methylome Landscapes of Human Primordial Germ Cells	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE63818	NCBI Gene Expression Omnibus, GSE63818
Low DH, Hammoud SS	2014	Chromatin and Transcription Transitions of Mammalian Adult Germline Stem Cells and Spermatogenesis	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE49624	NCBI Gene Expression Omnibus, GSE49624
Okabe H, Chiba H, Hiura H, Hamada H, Sato A, Utsunomiya T, Kikuchi H, Yoshida H, Tanaka A, Suyama M, Arima T	2014	Genome-Wide Analysis of DNA Methylation Dynamics during Early Human Development	https://humandbs.biosciencedbc.jp/hum0009-v1	the Japanese Genotype-phenotype Archive, JGAS000000000006
Sohni A, Tan K, Song H, Burow D, Wilkinson MF	2018	Neonatal and adult human testis defined at the single-cell level	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE124263	NCBI Gene Expression Omnibus, GSE124263

References

- Arafat M**, Har-Vardi I, Harlev A, Levitas E, Zeadna A, Abofoul-Azab M, Dyomin V, Sheffield VC, Lunenfeld E, Huleihel M, Parvari R. 2017. Mutation in TDRD9 causes non-obstructive azoospermia in infertile men. *Journal of Medical Genetics* **54**:633–639. DOI: <https://doi.org/10.1136/jmedgenet-2017-104514>, PMID: 28536242
- Aravin AA**, Sachidanandam R, Girard A, Fejes-Toth K, Hannon GJ. 2007. Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science (New York, N.Y.)* **316**:744–747. DOI: <https://doi.org/10.1126/science.1142612>, PMID: 17446352

- Aravin AA**, Sachidanandam R, Bourc'his D, Schaefer C, Pezic D, Toth KF, Bestor T, Hannon GJ. 2008. A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Molecular Cell* **31**:785–799. DOI: <https://doi.org/10.1016/j.molcel.2008.09.003>, PMID: 18922463
- Barau J**, Teissandier A, Zamudio N, Roy S, Nalesso V, Héroult Y, Guillou F, Bourc'his D. 2016. The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. *Science (New York, N.Y.)* **354**:909–912. DOI: <https://doi.org/10.1126/science.aah5143>, PMID: 27856912
- Baubec T**, Colombo DF, Wirbelauer C, Schmidt J, Burger L, Krebs AR, Akalin A, Schübeler D. 2015. Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature* **520**:243–247. DOI: <https://doi.org/10.1038/nature14176>, PMID: 25607372
- Bourc'his D**, Bestor TH. 2004. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* **431**:96–99. DOI: <https://doi.org/10.1038/nature02886>, PMID: 15318244
- Bujold D**, Morais DA, Gauthier C, Côté C, Caron M, Kwan T, Chen KC, Laperle J, Markovits AN, Pastinen T, Caron B, Veilleux A, Jacques PÉ, Bourque G. 2016. The International Human Epigenome Consortium Data Portal. *Cell Systems* **3**:496–499. DOI: <https://doi.org/10.1016/j.cels.2016.10.019>, PMID: 27863956
- Bulut-Karslioglu A**, De La Rosa-Velázquez IA, Ramirez F, Barenboim M, Onishi-Seebacher M, Arand J, Galán C, Winter GE, Engist B, Gerle B, O'Sullivan RJ, Martens JHA, Walter J, Manke T, Lachner M, Jenuwein T. 2014. Suv39h-dependent H3K9me3 marks intact retrotransposons and silences LINE elements in mouse embryonic stem cells. *Molecular Cell* **55**:277–290. DOI: <https://doi.org/10.1016/j.molcel.2014.05.029>, PMID: 24981170
- Carmell MA**, Girard A, van de Kant HJG, Bourc'his D, Bestor TH, de Rooij DG, Hannon GJ. 2007. MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Developmental Cell* **12**:503–514. DOI: <https://doi.org/10.1016/j.devcel.2007.03.001>, PMID: 17395546
- Cosby RL**, Chang NC, Feschotte C. 2019. Host-transposon interactions: conflict, cooperation, and cooption. *Genes & Development* **33**:1098–1116. DOI: <https://doi.org/10.1101/gad.327312.119>, PMID: 31481535
- Daxinger L**, Oey H, Isbel L, Whitelaw NC, Youngson NA, Spurling A, Vonk KKD, Whitelaw E. 2016. Hypomethylation of ERVs in the sperm of mice haploinsufficient for the histone methyltransferase Setdb1 correlates with a paternal effect on phenotype. *Scientific Reports* **6**:25004. DOI: <https://doi.org/10.1038/srep25004>, PMID: 27112447
- Deniz Ö**, Ahmed M, Todd CD, Rio-Machin A, Dawson MA, Branco MR. 2020. Endogenous retroviruses are a source of enhancers with oncogenic potential in acute myeloid leukaemia. *Nature Communications* **11**:3506. DOI: <https://doi.org/10.1038/s41467-020-17206-4>, PMID: 32665538
- Ewing AD**, Smits N, Sanchez-Luque FJ, Faivre J, Brennan PM, Richardson SR, Cheetham SW, Faulkner GJ. 2020. Nanopore Sequencing Enables Comprehensive Transposable Element Epigenomic Profiling. *Molecular Cell* **80**:915–928. DOI: <https://doi.org/10.1016/j.molcel.2020.10.024>, PMID: 33186547
- Fukuda K**, Inoguchi Y, Ichiyanagi K, Ichiyanagi T, Go Y, Nagano M, Yanagawa Y, Takaesu N, Ohkawa Y, Imai H, Sasaki H. 2017. Evolution of the sperm methylome of primates is associated with retrotransposon insertions and genome instability. *Human Molecular Genetics* **26**:3508–3519. DOI: <https://doi.org/10.1093/hmg/ddx236>, PMID: 28637190
- Fukuda K**, Okuda A, Yusa K, Shinkai Y. 2018. A CRISPR knockout screen identifies SETDB1-target retroelement silencing factors in embryonic stem cells. *Genome Research* **28**:846–858. DOI: <https://doi.org/10.1101/gr.227280.117>, PMID: 29728365
- Fukuda K**, Shinkai Y. 2020. SETDB1-Mediated Silencing of Retroelements. *Viruses* **12**:E596. DOI: <https://doi.org/10.3390/v12060596>, PMID: 32486217
- Gianfrancesco O**, Bubb VJ, Quinn JP. 2017. SVA retrotransposons as potential modulators of neuropeptide gene expression. *Neuropeptides* **64**:3–7. DOI: <https://doi.org/10.1016/j.npep.2016.09.006>, PMID: 27743609
- Gkountela S**, Zhang KX, Shafiq TA, Liao W-W, Hargan-Calvopiña J, Chen P-Y, Clark AT. 2015. DNA Demethylation Dynamics in the Human Prenatal Germline. *Cell* **161**:1425–1436. DOI: <https://doi.org/10.1016/j.cell.2015.05.012>, PMID: 26004067
- Goodier JL**. 2016. Restricting retrotransposons: a review. *Mobile DNA* **7**:16. DOI: <https://doi.org/10.1186/s13100-016-0070-z>, PMID: 27525044
- Grant CE**, Bailey TL, Noble WS. 2011. FIMO: scanning for occurrences of a given motif. *Bioinformatics (Oxford, England)* **27**:1017–1018. DOI: <https://doi.org/10.1093/bioinformatics/btr064>, PMID: 21330290
- Gu A**, Ji G, Shi X, Long Y, Xia Y, Song L, Wang S, Wang X. 2010. Genetic variants in Piwi-interacting RNA pathway genes confer susceptibility to spermatogenic failure in a Chinese population. *Human Reproduction (Oxford, England)* **25**:2955–2961. DOI: <https://doi.org/10.1093/humrep/deq274>, PMID: 20940137
- Guo F**, Yan L, Guo H, Li L, Hu B, Zhao Y, Yong J, Hu Y, Wang X, Wei Y, Wang W, Li R, Yan J, Zhi X, Zhang Y, Jin H, Zhang W, Hou Y, Zhu P, Li J, et al. 2015. The Transcriptome and DNA Methylome Landscapes of Human Primordial Germ Cells. *Cell* **161**:1437–1452. DOI: <https://doi.org/10.1016/j.cell.2015.05.015>, PMID: 26046443
- Hammoud SS**, Low DHP, Yi C, Carrell DT, Guccione E, Cairns BR. 2014. Chromatin and transcription transitions of mammalian adult germline stem cells and spermatogenesis. *Cell Stem Cell* **15**:239–253. DOI: <https://doi.org/10.1016/j.stem.2014.04.006>, PMID: 24835570
- Haring NL**, van Bree EJ, Jordaan WS, Roels JRE, Sotomayor GC, Hey TM, White FTG, Galland MD, Smidt MP, Jacobs FMJ. 2021. ZNF91 deletion in human embryonic stem cells leads to ectopic activation of SVA retrotransposons and up-regulation of KRAB zinc finger gene clusters. *Genome Research* **31**:551–563. DOI: <https://doi.org/10.1101/gr.265348.120>, PMID: 33722937
- Heinz S**, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for

- macrophage and B cell identities. *Molecular Cell* **38**:576–589. DOI: <https://doi.org/10.1016/j.molcel.2010.05.004>, PMID: 20513432
- Helleboid P-Y**, Heusel M, Duc J, Piot C, Thorball CW, Coluccio A, Pontis J, Imbeault M, Turelli P, Aebersold R, Trono D. 2019. The interactome of KRAB zinc finger proteins reveals the evolutionary history of their functional diversification. *The EMBO Journal* **38**:e101220. DOI: <https://doi.org/10.15252/embj.2018101220>, PMID: 31403225
- Heyn H**, Ferreira HJ, Bassas L, Bonache S, Sayols S, Sandoval J, Esteller M, Larriba S. 2012. Epigenetic disruption of the PIWI pathway in human spermatogenic disorders. *PLOS ONE* **7**:e47892. DOI: <https://doi.org/10.1371/journal.pone.0047892>, PMID: 23112866
- Huang CRL**, Burns KH, Boeke JD. 2012. Active transposition in genomes. *Annual Review of Genetics* **46**:651–675. DOI: <https://doi.org/10.1146/annurev-genet-110711-155616>, PMID: 23145912
- Imbeault M**, Helleboid PY, Trono D. 2017. KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature* **543**:550–554. DOI: <https://doi.org/10.1038/nature21683>, PMID: 28273063
- Inoue K**, Ichihyanagi K, Fukuda K, Glinka M, Sasaki H. 2017. Switching of dominant retrotransposon silencing strategies from posttranscriptional to transcriptional mechanisms during male germ-cell development in mice. *PLOS Genetics* **13**:e1006926. DOI: <https://doi.org/10.1371/journal.pgen.1006926>, PMID: 28749988
- Jacobs FMJ**, Greenberg D, Nguyen N, Haeussler M, Ewing AD, Katzman S, Paten B, Salama SR, Haussler D. 2014. An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature* **516**:242–245. DOI: <https://doi.org/10.1038/nature13760>, PMID: 25274305
- Johnson WL**, Yewdell WT, Bell JC, McNulty SM, Duda Z, O'Neill RJ, Sullivan BA, Straight AF. 2017. RNA-dependent stabilization of SUV39H1 at constitutive heterochromatin. *eLife* **6**:e25299. DOI: <https://doi.org/10.7554/eLife.25299>, PMID: 28760200
- Kobayashi H**, Sakurai T, Miura F, Imai M, Mochiduki K, Yanagisawa E, Sakashita A, Wakai T, Suzuki Y, Ito T, Matsui Y, Kono T. 2013. High-resolution DNA methylome analysis of primordial germ cells identifies gender-specific reprogramming in mice. *Genome Research* **23**:616–627. DOI: <https://doi.org/10.1101/gr.148023.112>, PMID: 23410886
- Kojima-Kita K**, Kuramochi-Miyagawa S, Nagamori I, Ogonuki N, Ogura A, Hasuwa H, Akazawa T, Inoue N, Nakano T. 2016. MIWI2 as an Effector of DNA Methylation and Gene Silencing in Embryonic Male Germ Cells. *Cell Reports* **16**:2819–2828. DOI: <https://doi.org/10.1016/j.celrep.2016.08.027>, PMID: 27626653
- Krueger F**, Andrews SR. 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics (Oxford, England)* **27**:1571–1572. DOI: <https://doi.org/10.1093/bioinformatics/btr167>, PMID: 21493656
- Lander ES**, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**:860–921. DOI: <https://doi.org/10.1038/35057062>, PMID: 11237011
- Leitão E**, Di Persio S, Laurentino S, Wöste M, Dugas M, Kliesch S, Neuhaus N, Horsthemke B. 2020. The sperm epigenome does not display recurrent epimutations in patients with severely impaired spermatogenesis. *Clinical Epigenetics* **12**:61. DOI: <https://doi.org/10.1186/s13148-020-00854-0>, PMID: 32375885
- Liu S**, Brind'Amour J, Karimi MM, Shirane K, Bogutz A, Lefebvre L, Sasaki H, Shinkai Y, Lorincz MC. 2014. Setdb1 is required for germline development and silencing of H3K9me3-marked endogenous retroviruses in primordial germ cells. *Genes & Development* **28**:2041–2055. DOI: <https://doi.org/10.1101/gad.244848.114>, PMID: 25228647
- Liu N**, Lee CH, Swigut T, Grow E, Gu B, Bassik MC, Wysocka J. 2018. Selective silencing of euchromatic L1s revealed by genome-wide screens for L1 regulators. *Nature* **553**:228–232. DOI: <https://doi.org/10.1038/nature25179>, PMID: 29211708
- Maksakova IA**, Thompson PJ, Goyal P, Jones SJ, Singh PB, Karimi MM, Lorincz MC. 2013. Distinct roles of KAP1, HP1 and G9a/GLP in silencing of the two-cell-specific retrotransposon MERVL in mouse ES cells. *Epigenetics & Chromatin* **6**:15. DOI: <https://doi.org/10.1186/1756-8935-6-15>, PMID: 23735015
- Manakov SA**, Pezic D, Marinov GK, Pastor WA, Sachidanandam R, Aravin AA. 2015. MIWI2 and MILI Have Differential Effects on piRNA Biogenesis and DNA Methylation. *Cell Reports* **12**:1234–1243. DOI: <https://doi.org/10.1016/j.celrep.2015.07.036>, PMID: 26279574
- Matsui T**, Leung D, Miyashita H, Maksakova IA, Miyachi H, Kimura H, Tachibana M, Lorincz MC, Shinkai Y. 2010. Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature* **464**:927–931. DOI: <https://doi.org/10.1038/nature08858>, PMID: 20164836
- Molaro A**, Hodges E, Fang F, Song Q, McCombie WR, Hannon GJ, Smith AD. 2011. Sperm methylation profiles reveal features of epigenetic inheritance and evolution in primates. *Cell* **146**:1029–1041. DOI: <https://doi.org/10.1016/j.cell.2011.08.016>, PMID: 21925323
- Okae H**, Chiba H, Hiura H, Hamada H, Sato A, Utsunomiya T, Kikuchi H, Yoshida H, Tanaka A, Suyama M, Arima T. 2014. Genome-wide analysis of DNA methylation dynamics during early human development. *PLOS Genetics* **10**:e1004868. DOI: <https://doi.org/10.1371/journal.pgen.1004868>, PMID: 25501653
- Ostertag EM**, Goodier JL, Zhang Y, Kazazian HH. 2003. SVA elements are nonautonomous retrotransposons that cause disease in humans. *American Journal of Human Genetics* **73**:1444–1451. DOI: <https://doi.org/10.1086/380207>, PMID: 14628287
- O'Donnell KA**, Burns KH. 2010. Mobilizing diversity: transposable element insertions in genetic variation and disease. *Mobile DNA* **1**:21. DOI: <https://doi.org/10.1186/1759-8753-1-21>, PMID: 20813032

- Payer LM**, Burns KH. 2019. Transposable elements in human genetic disease. *Nature Reviews. Genetics* **20**:760–772. DOI: <https://doi.org/10.1038/s41576-019-0165-8>, PMID: 31515540
- Quinlan AR**, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)* **26**:841–842. DOI: <https://doi.org/10.1093/bioinformatics/btq033>, PMID: 20110278
- Rakyan VK**, Chong S, Champ ME, Cuthbert PC, Morgan HD, Luu KVK, Whitelaw E. 2003. Transgenerational inheritance of epigenetic states at the murine Axin(Fu) allele occurs after maternal and paternal transmission. *PNAS* **100**:2538–2543. DOI: <https://doi.org/10.1073/pnas.0436776100>, PMID: 12601169
- Reznik B**, Cincotta SA, Jaszczak RG, Mateo LJ, Shen J, Cao M, Baskin L, Ye P, An W, Laird DJ. 2019. Heterogeneity of transposon expression and activation of the repressive network in human fetal germ cells. *Development (Cambridge, England)* **146**:12. DOI: <https://doi.org/10.1242/dev.171157>, PMID: 30658985
- Robbez-Masson L**, Tie CHC, Conde L, Tunbak H, Husovsky C, Tchasovnikarova IA, Timms RT, Herrero J, Lehner PJ, Rowe HM. 2018. The HUSH complex cooperates with TRIM28 to repress young retrotransposons and new genes. *Genome Research* **28**:836–845. DOI: <https://doi.org/10.1101/gr.228171.117>, PMID: 29728366
- Robinson JT**, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nature Biotechnology* **29**:24–26. DOI: <https://doi.org/10.1038/nbt.1754>, PMID: 21221095
- Rowe HM**, Friedli M, Offner S, Verp S, Mesnard D, Marquis J, Aktas T, Trono D. 2013. De novo DNA methylation of endogenous retroviruses is shaped by KRAB-ZFPs/KAP1 and ESET. *Development (Cambridge, England)* **140**:519–529. DOI: <https://doi.org/10.1242/dev.087585>, PMID: 23293284
- Saito Y**, Tsuji J, Mituyama T. 2014. Bisulfighter: accurate detection of methylated cytosines and differentially methylated regions. *Nucleic Acids Research* **42**:e45. DOI: <https://doi.org/10.1093/nar/gkt1373>, PMID: 24423865
- Sasaki K**, Yokobayashi S, Nakamura T, Okamoto I, Yabuta Y, Kurimoto K, Ohta H, Moritoki Y, Iwatani C, Tsuchiya H, Nakamura S, Sekiguchi K, Sakuma T, Yamamoto T, Mori T, Woltjen K, Nakagawa M, Yamamoto T, Takahashi K, Yamanaka S, et al. 2015. Robust In Vitro Induction of Human Germ Cell Fate from Pluripotent Stem Cells. *Cell Stem Cell* **17**:178–194. DOI: <https://doi.org/10.1016/j.stem.2015.06.014>, PMID: 26189426
- Schultz DC**, Friedman JR, Rauscher FJ. 2001. Targeting histone deacetylase complexes via KRAB-zinc finger proteins: the PHD and bromodomains of KAP-1 form a cooperative unit that recruits a novel isoform of the Mi-2alpha subunit of NuRD. *Genes & Development* **15**:428–443. DOI: <https://doi.org/10.1101/gad.869501>, PMID: 11230151
- Schultz DC**, Ayyanathan K, Negorev D, Maul GG, Rauscher FJ. 2002. SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes & Development* **16**:919–932. DOI: <https://doi.org/10.1101/gad.973302>, PMID: 11959841
- Seisenberger S**, Andrews S, Krueger F, Arand J, Walter J, Santos F, Popp C, Thienpont B, Dean W, Reik W. 2012. The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. *Molecular Cell* **48**:849–862. DOI: <https://doi.org/10.1016/j.molcel.2012.11.001>, PMID: 23219530
- Seki Y**, Yamaji M, Yabuta Y, Sano M, Shigetani M, Matsui Y, Saga Y, Tachibana M, Shinkai Y, Saitou M. 2007. Cellular dynamics associated with the genome-wide epigenetic reprogramming in migrating primordial germ cells in mice. *Development (Cambridge, England)* **134**:2627–2638. DOI: <https://doi.org/10.1242/dev.005611>, PMID: 17567665
- Shen L**, Shao N, Liu X, Nestler E. 2014. ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC Genomics* **15**:284. DOI: <https://doi.org/10.1186/1471-2164-15-284>, PMID: 24735413
- Shirai A**, Kawaguchi T, Shimojo H, Muramatsu D, Ishida-Yonetani M, Nishimura Y, Kimura H, Nakayama J-I, Shinkai Y. 2017. Impact of nucleic acid and methylated H3K9 binding activities of Suv39h1 on its heterochromatin assembly. *eLife* **6**:e25317. DOI: <https://doi.org/10.7554/eLife.25317>, PMID: 28760201
- Shirane K**, Miura F, Ito T, Lorincz MC. 2020. NSD1-deposited H3K36me2 directs de novo methylation in the mouse male germline and counteracts Polycomb-associated silencing. *Nature Genetics* **52**:1088–1098. DOI: <https://doi.org/10.1038/s41588-020-0689-z>, PMID: 32929285
- Shoji M**, Tanaka T, Hosokawa M, Reuter M, Stark A, Kato Y, Kondoh G, Okawa K, Chujo T, Suzuki T, Hata K, Martin SL, Noce T, Kuramochi-Miyagawa S, Nakano T, Sasaki H, Pillai RS, Nakatsuji N, Chuma S. 2009. The TDRD9-MIWI2 complex is essential for piRNA-mediated retrotransposon silencing in the mouse male germline. *Developmental Cell* **17**:775–787. DOI: <https://doi.org/10.1016/j.devcel.2009.10.012>, PMID: 20059948
- Sohni A**, Tan K, Song H-W, Burrow D, de Rooij DG, Laurent L, Hsieh T-C, Rabah R, Hammoud SS, Vicini E, Wilkinson MF. 2019. The Neonatal and Adult Human Testis Defined at the Single-Cell Level. *Cell Reports* **26**:1501–1517. DOI: <https://doi.org/10.1016/j.celrep.2019.01.045>, PMID: 30726734
- Sripathy SP**, Stevens J, Schultz DC. 2006. The KAP1 corepressor functions to coordinate the assembly of de novo HP1-demarcated microenvironments of heterochromatin required for KRAB zinc finger protein-mediated transcriptional repression. *Molecular and Cellular Biology* **26**:8623–8638. DOI: <https://doi.org/10.1128/MCB.00487-06>, PMID: 16954381
- Tamura K**, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution* **10**:512–526. DOI: <https://doi.org/10.1093/oxfordjournals.molbev.a040023>, PMID: 8336541
- Tamura K**, Stecher G, Peterson D, Filipksi A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution* **30**:2725–2729. DOI: <https://doi.org/10.1093/molbev/mst197>, PMID: 24132122

- Tan K**, Song HW, Thompson M, Munyoki S, Sukhwani M, Hsieh TC, Orwig KE, Wilkinson MF. 2020. Transcriptome profiling reveals signaling conditions dictating human spermatogonia fate in vitro. *PNAS* **117**:17832–17841. DOI: <https://doi.org/10.1073/pnas.2000362117>, PMID: 32661178
- Tang WWC**, Dietmann S, Irie N, Leitch HG, Floros VI, Bradshaw CR, Hackett JA, Chinnery PF, Surani MA. 2015. A Unique Gene Regulatory Network Resets the Human Germline Epigenome for Development. *Cell* **161**:1453–1467. DOI: <https://doi.org/10.1016/j.cell.2015.04.053>, PMID: 26046444
- Teissandier A**, Bourc'his D. 2017. Gene body DNA methylation conspires with H3K36me3 to preclude aberrant transcription. *The EMBO Journal* **36**:1471–1473. DOI: <https://doi.org/10.15252/embj.201796812>, PMID: 28442531
- Urduingio RG**, Bayón GF, Dmitrijeva M, Toraño EG, Bravo C, Fraga MF, Bassas L, Larriba S, Fernández AF. 2015. Aberrant DNA methylation patterns of spermatozoa in men with unexplained infertility. *Human Reproduction (Oxford, England)* **30**:1014–1028. DOI: <https://doi.org/10.1093/humrep/dev053>, PMID: 25753583
- Velazquez Camacho O**, Galan C, Swist-Rosowska K, Ching R, Gamalinda M, Karabiber F, De La Rosa-Velazquez I, Engst B, Koschorz B, Shukeir N, Onishi-Seebacher M, van de Nobelen S, Jenuwein T. 2017. Major satellite repeat RNA stabilize heterochromatin retention of Suv39h enzymes by RNA-nucleosome association and RNA:DNA hybrid formation. *eLife* **6**:e25293. DOI: <https://doi.org/10.7554/eLife.25293>, PMID: 28760199
- Venkatesh S**, Li H, Gogol MM, Workman JL. 2016. Selective suppression of antisense transcription by Set2-mediated H3K36 methylation. *Nature Communications* **7**:13610. DOI: <https://doi.org/10.1038/ncomms13610>, PMID: 27892455
- von Meyenn F**, Berrens RV, Andrews S, Santos F, Collier AJ, Krueger F, Osorno R, Dean W, Rugg-Gunn PJ, Reik W. 2016. Comparative Principles of DNA Methylation Reprogramming during Human and Mouse In Vitro Primordial Germ Cell Specification. *Developmental Cell* **39**:104–115. DOI: <https://doi.org/10.1016/j.devcel.2016.09.015>, PMID: 27728778
- Watanabe T**, Cui X, Yuan Z, Qi H, Lin H. 2018. MIWI2 targets RNAs transcribed from piRNA-dependent regions to drive DNA methylation in mouse prospermatogonia. *The EMBO Journal* **37**:18. DOI: <https://doi.org/10.15252/embj.201695329>, PMID: 30108053
- Wolf G**, Yang P, Fuchtbauer AC, Fuchtbauer E-M, Silva AM, Park C, Wu W, Nielsen AL, Pedersen FS, Macfarlan TS. 2015. The KRAB zinc finger protein ZFP809 is required to initiate epigenetic silencing of endogenous retroviruses. *Genes & Development* **29**:538–554. DOI: <https://doi.org/10.1101/gad.252767.114>, PMID: 25737282
- Xiang X**, Tao Y, DiRusso J, Hsu F-M, Zhang J, Xue Z, Pontis J, Trono D, Liu W, Clark AT. 2022. Human reproduction is regulated by retrotransposons derived from ancient Hominidae-specific viral infections. *Nature Communications* **13**:463. DOI: <https://doi.org/10.1038/s41467-022-28105-1>, PMID: 35075135
- Zoch A**, Auchynnikava T, Berrens RV, Kabayama Y, Schöpp T, Heep M, Vasiliauskaitė L, Pérez-Rico YA, Cook AG, Shkumatava A, Rappsilber J, Allshire RC, O'Carroll D. 2020. SPOCD1 is an essential executor of piRNA-directed de novo DNA methylation. *Nature* **584**:635–639. DOI: <https://doi.org/10.1038/s41586-020-2557-5>, PMID: 32674113