Transcriptional states of retroelement-inserted regions and specific KRAB zinc finger protein association are correlated with DNA methylation of retroelements in human male germ cells

Kei Fukuda1*, Yoshinori Makino2, Satoru Kaneko3, Chikako Shimura1, Yuki Okada2, Kenji Ichiyanagi4, Yoichi Shinkai1*

1 Cellular Memory Laboratory, RIKEN Cluster for Pioneering Research, Wako 351-0198, Japan.
2 Laboratory of Pathology and Development, Institute for Quantitative Biosciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo, Tokyo 113-0032, Japan.
3 Department of Obstetrics and Gynecology, Ichikawa General Hospital, Tokyo Dental College, 5-11-13, Sugano, Ichikawa, Chiba, 272-0824, Japan
4 Laboratory of Genome and Epigenome Dynamics, Department of Animal Sciences, Graduate School of Bioagricultural Sciences, Nagoya University, Nagoya 464-8601, Japan

*Correspondence: kei.fukuda@riken.jp (K.F.) and yshinkai@riken.jp (Y.S.)

Running title: Retroelement DNA methylation in human germ cells

KEYWORDS: Retroelements, DNA methylation, human male germ cells
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 (KRAB-ZFPs) are associated with DNA methylation of retroelements in human primordial germ cells (hPGCs). 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.
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, Burns, & Boeke, 2012; Maksakova et al., 2013; Ostertag, Goodier, Zhang, & Kazazian, 2003). As retrotransposons cause genome instability, because insertional mutagenesis, or transcriptional perturbation and are often deleterious to host species (O’Donnell & 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 & 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, Helleboid, & Trono, 2017; Jacobs et al., 2014; Wolf et al., 2015). KRAB-ZFPs repress retroelements by recruiting KAP1/TRIM28 (Sripathy, Stevens, & Schultz, 2006) and other repressive epigenetic modifiers (Schultz, Ayyanathan, Negorev, Maul, & Rauscher, 2002; Schultz, Friedman, & Rauscher, 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 (Gkountela 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 (Gkountela 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 (S. 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, Ichyanagi, Fukuda, Glinka, & Sasaki, 2017). Epigenetic alterations in retroelements and dysfunction of retroelement silencing pathways in male germ cells are associated with male sterility linked to azoosperma (Aravin, Sachidanandam, Girard, Fejes-Toth, & Hannon, 2007; Bourc'his & Bestor, 2004; Carmell et al., 2007; Heyn et al., 2012; Urdinguio 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 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 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-ZNPs 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 (Fig. 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, 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 SV A 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 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, Chang, & Feschotte, 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 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: 2,054; 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 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, Cui, Yuan, Qi, & Lin, 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)
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), 2,008 regions were identified as DMRs (donor 1 < donor 2: 332, donor 2 < donor 1: 1,676). Of the 1,676 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 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 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) (Leitao 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 (FSH), luteinizing hormone (LH), sex hormone-binding globulin (SHBG) 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 towards 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-ZNPs 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 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 2H, I). 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-quality 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, Okuda, Yusa, & Shinkai, 2018; N. 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 & Shinkai, 2020; N. 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
In eukaryotes, gene bodies are the most conserved targets of DNA methylation. Gene body DNA methylation levels are often correlated with transcriptional levels (Teissandier & 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, Miura, Ito, & Lorincz, 2020). Furthermore, antisense RNAs embedded within protein-coding genes are selectively silenced by H3K36 methyltransferase SET2 in Saccharomyces cerevisiae (Venkatesh, Li, Gogol, & Workman, 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, Bubb, & Quinn, 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 & 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.

**FIGUER LEGENDS**

**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 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 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 Fig. 1C-E.

Figure 2 Identification of KRAB-ZFPs associated with DNA demethylation resistance in 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. SVA_A copies were classified by DNA methylation levels in 19W 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 sub-type. (E) Position of ZNF257 and ZNF28 binding motifs along SVA consensus sequences. VNTR1 and VNTR2 is composed of multiple copy number of tandem repeats, and the copy number of these VNTR 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 sub-type. (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.

**Figure 3 Identification of 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 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 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.
Figure 4 DNA methylation dynamics of retroelements during human spermatogenesis

(A) Scatter plots showing DNA methylation levels of each retroelement copy in male hPGCs at 19 weeks of gestation and sperm. WGBS data from two sperm donors (Hammoud et al. 2014) were used for this analysis. Donor1 and donor2 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 SVA (D). “Other” indicates groups except for Common High, Common Low and High and Low, such as Low methylated in donor1 and Mediumly methylated in donor2. 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 donor1, and sperm donor2. 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.

Figure 5 Transcription associated regulation of DNA methylation of SVA during spermatogenesis

(A) Phylogenetic analysis of SVA_B (left) and D (right) copies low-methylated in male 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 & low and common high).

Figure 6 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 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. (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-seq for analyzing SVA methylation. (C) Bar plots showing the 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 donor1 and sperm donor2 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 DMRs. (G) Representative view of DMRs overlapping with SVA. Black and green boxes represent SVA and DMR, respectively.

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 hPGCs tended to be maintained during spermatogenesis. The mode of DNA methylation changes in 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.

SUPPLEMENTARY FIGURE AND FILE LEGENDS

Figure 1-figure supplement 1 Mappability of WGBS reads on transposable elements

(A) Violin plots of WGBS read mappability in each retroelement types. Less than 10% of L1HS-derived reads were uniquely mapped in the majority of L1HS copies. (B) Percentage of copies measured for DNA methylation (>= 10 CpG with >= 5 reads).

Figure 2-figure supplement 1 Correlation between ZNF257/ZNF28 binding motifs and DNA methylation levels in SVA_A copies.

(A) Enrichment of SVA_A in ZNF257 and ZNF28 peaks. We randomly extracted genomic regions whose length and number matched the SVA copy and calculated the overlap with the ZNF257/28 peaks. This procedure was repeated 1,000 times, and the average value was used as the background. P-value was
calculated by Prop.test in R. (B) Violin plots of read mappability in SVA_A copies. Read mappability was significantly higher in “high” group of SVA_A copy than other groups. P-value was calculated by Tukey’s test. (C) Enrichment of ZNF257 and ZNF28 on SVA_A copies classified by DNA methylation levels in male hPGCs at 19 weeks of gestation. Only SVA_A copies with 50~70% mappability were used in this figure. (D) Position of ZNF28/257 binding motifs along VNTR region of SVA_A consensus sequence. TR represents a unit of VNTR. (E) Scatter plot of DNA methylation levels and VNTR1 copy number of SVA_A copies. Pearson’s r and P-value were calculated by cor.test in R. (F) Scatter plot of DNA methylation levels and the number of ZNF257 (left) or ZNF28 (right) binding motifs in SVA_A copies. (G) Correlation of DNA methylation levels in VNTR of SVA_A and the presence of ZNF257/28 peaks. Black box represents the presence of ZNF257/28 peaks. (H) Representative view of correlation of DNA methylation in SVA_A and ZNF257/28 binding.

Figure 2-figure supplement 2 Expression of KRAB-ZFP genes in somatic cells and hPGCs.

(A) Heatmap of expression levels of KRAB-ZFP genes in hPGCs and the neighboring somatic cells. Y-axis represents the data from each cell which categorized by cell-type and weeks of gestation. ACTB and GAPDH are house-keeping genes and KIT is germ-cell marker. (B) Violinplot showing expression levels (log2(FPKM)) of KRAB-ZFP genes in hPGCs (N=233) and the neighboring somatic cells (N=86). P-value was calculated by T-test.
Figure 3-figure supplement 1 Enrichment of ZNF649/93 on L1PA4.

(A) Enrichment of ZNF649 and ZNF93 on L1PA4 classified by DNA methylation levels in male hPGCs at 19 weeks of gestation. (B) Violin plot showing the mappability of L1PA4 copies classified by DNA methylation levels in male hPGCs at 19 weeks of gestation. P-value were calculated by Tukey’s test.

Figure 3-figure supplement 2 ZNF850 association is correlated with DNA demethylation resistance of LTR12 family in hPGCs.

(A) Scatter plots showing the fraction of low-methylated or highly methylated LTR12 copies which overlaps of KRAB-ZFP peaks. (B) Fraction of LTR12 copies in each group overlapping with ZNF850 peaks. P-value was calculated by prop.test in R. (C) Fraction of LTR12C copies in each group containing ZNF850 binding motif. P-value was calculated by prop.test in R. (D) The position of ZNF850 binding motif along LTR12C consensus sequence. (E) Enrichment of ZNF850 ChIP-seq reads within and around LTR12C copies. (F) Comparison of sequences at ZNF850 binding motifs between highly and low-methylated LTR12C copies. (G) Representative view of correlation between ZNF850 peak and DNA methylation of LTR12C in hPGCs.

Figure 6-figure supplement 1 Inter-individual epigenetic variation in SVAs was correlated with
testosterone levels in bloods.

(A-C) Violin plots showing DNA methylation levels of SVAs in adult skeletal muscle (A), adult CD-4 positive helper T-cell (B) and sperm from healthy and oligozoospermic men (C). In Figure 6-figure supplement 1C, NC and OAT represent normal control and oligozoospermic men, respectively. (D) Boxplots showing various physiological conditions between donors with high methylated “High and Low” SVAs (N=4) and donors with low methylated ones (N=7) in Figure 6-figure supplement 1C. P-value was calculated by Wilcoxon signed-rank test.

Supplementary File 1. List of Full-length L1/SVA/LTR12

Supplementary File 2. Statistical analysis of DNA methylation levels of repeat elements in human PGCs (Related to Fig. 1C-E)

SOURCE DATA LEGENDS

Figure 1-source data 1 : Raw data of graphs in Figure 1
Figure 2-source data 1 : Raw data of graphs in Figure 2
Figure 3-source data 1 : Raw data of graphs in Figure 3
Figure 4-source data 1 : Raw data of graphs in Figure 4
Figure 5-source data 1 : Raw data of graphs in Figure 5
Figure 6-source data 1 : Raw data of graphs in Figure 6
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, USA). 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× 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
TTATTGTAATTTTTTGTGGATTTTTTTTTTTAG. SVA_1_Rv
AAAAAACCTCCTACATCCCCAAC SVA_2_Fw TTAATGGTTTTAGGTTGGAGTGTAGTG
SVA_2_Rv CAAAAAAACTCCTCACTTCCAATA. SVA_3_Fw
TTTGGGAGGTGTATTTAATAGTTTATTGAGAAA SVA_3_Rv
TAAACAAAAATCTCTAATTTTTCTAAACAAAAAACC). 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 9 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 & 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 & 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, Bailey, & Noble, 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, Tsuji, & Mituyama, 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 & 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 (MCL) approach. The tree was drawn to scale, with branch lengths measured as the number of substitutions per site. There were 10153 positions in the final dataset.

Evolutionary analyses were conducted using MEGA6 (Tamura, Stecher, Peterson, Filipski, & Kumar, 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, Shao, Liu, & Nestler, 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 (IGV) (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.

COMPETING INTEREST STATEMENT

Not applicable.

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.

REFERENCES


Carmell, M. A., Girard, A., van de Kant, H. J., Bourc’his, D., Bestor, T. H., de Rooij, D. G., &


Grant, C. E., Bailey, T. L., & Noble, W. S. (2011). FIMO: scanning for occurrences of a given


Manakov, S. A., Pezic, D., Marinov, G. K., Pastor, W. A., Sachidanandam, R., & Aravin, A. A.


De novo DNA methylation of endogenous retroviruses is shaped by KRAB-ZFPs/KAP1 and ESET. Development, 140(3), 519-529. doi:10.1242/dev.087585


A. Fraction of low methylated SVA_A copy overlapping with KRAB-ZNP peaks (%)

B. [Graph showing distribution of SVA_A copy numbers]

C. [Graph showing RPKM values for SVA_A]

D. [Graph showing RPKM values for SVA_A and SVA_B]

E. [Graph showing RPKM values for SVA_A, SVA_B, SVA_C, SVA_D, SVA_F]

F. [Graph showing ZNF257 and ZNF28 motifs]

G. [Graph showing copy number of VNTR1]

H. [Graph showing copy number of VNTR1 for SVA_A, SVA_B, SVA_C, SVA_D, SVA_F]

I. [Graph showing ZNF257 and ZNF28 motif numbers]

J. [Graph showing P-values for copy number of VNTR1 and ZNF257/ZNF28 motif numbers]
Fukuda et al., Figure 2-figure supplement 1

A. Number of SVA_A located in KRAB-ZNF binding peaks.

B. Mappability (50%-70%).

C. DNA methylation levels in 19W hPGCs.

D. VNTR1 copy number.

E. ZNF257 binding motif.

F. ZNF28 binding motif.

G. DNA methylation levels in 19W hPGCs.

H. Highly methylated SVA_A in hPGCs.

I. Hypo-methylated SVA_A in hPGCs.
Fukuda et al., Figure 2-figure supplement 2

A

B

Actb

Kit

Znf28

Znf257

Znf91

P = 4.9x10^-11

P < 2.2e-16

P = 2.4x10^-7

P = 8.3x10^-5

P = 3.1x10^-12

P = 2.7x10^-4

N. S.
A. Fraction of L1 copies overlapping with KRAB-ZNF peaks.

B. ZNF93 and ZNF649 with L1HS, L1PA2, L1PA3, L1PA4, L1PA5, L1PA6, and L1PA7.

C. Fraction of ZNF93 and ZNF649 in low, medium, and high categories.

D. ZNF93 and ZNF649 motifs with P-values.

E. Fraction of ZNF93 and ZNF649 in low, medium, and high categories.

F. ZNF649 motif in L1 with ORF1, ORF2, and ORF3.

G. L1HS and L1PA2, L1PA3, and L1PA4 with highly methylated L1PA4 in hPGCs.

H. ZNF649 binding motif with WGBS, 19W hPGCs rep1, 19W hPGCs rep2, ZNF649 ChIP-seq, and L1PA4.

Fukuda et al., Figure 3
A

Fukuda et al., Figure 3-figure supplement 1

B
A. Fraction of low methylated SVA_A copy overlapping with KRAB-ZNP peaks (%)

B. Fraction of low methylated SVA_A copy overlapping with KRAB-ZNP peaks (%)

C. Fraction of LTR12C copies having ZNF850 motif

D. -log10(qValue) along LTR12C (bp)

E. ZNF850 enrichment around LTR12C

F. Motif and Consensus

G. Highly methylated LTR12C in hPGCs and Hypo-methylated LTR12C in hPGCs
DNA methylation in 19W hPGCs (%)

<table>
<thead>
<tr>
<th>Gene</th>
<th>Sperm donor 1</th>
<th>Sperm donor 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1PA2</td>
<td>389 copies</td>
<td></td>
</tr>
<tr>
<td>SVA_D</td>
<td>641 copies</td>
<td></td>
</tr>
<tr>
<td>LTR12C</td>
<td>2,031 copies</td>
<td></td>
</tr>
</tbody>
</table>

DNA methylation in sperm (%)

- Common High
- High and Low
- Common Low
- Other

Fukuda et al., Figure 4
A. SVA_B~F

Low-methylated in 19W PGCs

Sense Anti-sense Non-genic Fraction

B. RNA expression of genes with SVA copies by antisense orientation

Log2(RPM)

C. Enrichment of RNA-seq reads from undifferentiated spermatogonia

Low-methylated in 19W PGCs

D. Fukuda et al., Figure 5
Fukuda et al., Figure 6-figure supplement 1

A. Skelatal muscle CD4-positive helper T cells

DNA methylation levels (%)

Common High  High and Low  Common Low

B. CD4-positive helper T cells

DNA methylation levels (%)

Common High  High and Low  Common Low

C. Common High  High and Low  Common Low

DNA methylation levels (%)

D. Box plots for various parameters:

- Age (y.o.)
- Total Testosterone (nmol/l)
- Free Testosterone (nmol/l)
- FSH (IU/l)
- SHBG (nmol/l)
- LH (IU/l)
- Prostate (mIU/l)

Group comparisons with p-values:
- Age: 0.7
- Total Testosterone: 0.024
- Free Testosterone: 0.012
- FSH: 1
- SHBG: 0.85
- LH: 0.79
L1 (3,333 copies)

With ZNF649 binding motif

LTR12 (2,286 copies)

With ZNF850 binding motif

SVA (1,249 copies)

With ZNF257/28 binding motifs

Located in high transcriptionally active regions with antisense direction

Located in low transcriptionally active regions

Located in transcriptionally inactive regions or transcriptionally active regions with sense direction

Located in transcriptionally active regions with antisense direction