

UC Santa Barbara

UC Santa Barbara Previously Published Works

Title

Whole-genome DNA methylomes of tree shrew brains reveal conserved and divergent roles of DNA methylation on sex chromosome regulation.

Permalink

<https://escholarship.org/uc/item/8ck2v1mn>

Journal

Journal of Biology, 22(1)

Authors

Son, Dongmin

Kong, Yifan

Tan, Yulian

et al.

Publication Date

2024-11-28

DOI

10.1186/s12915-024-02071-0

Peer reviewed

RESEARCH

Open Access



Whole-genome DNA methylomes of tree shrew brains reveal conserved and divergent roles of DNA methylation on sex chromosome regulation

Dongmin R. Son^{1,2†}, Yifan Kong^{3,4,5,6†}, Yulian Tan^{3,4,5}, Ting Hu^{3,4,5,6}, Lei Shi^{3,4,5*} and Soojin V. Yi^{1,2*}

Abstract

Background The tree shrew (*Tupaia belangeri*) is a promising emerging model organism in biomedical studies, notably due to their evolutionary proximity to primates. To enhance our understanding of how DNA methylation is implicated in regulation of gene expression and the X chromosome inactivation (XCI) in tree shrew brains, here we present their first genome-wide, single-base-resolution methylomes integrated with transcriptomes from prefrontal cortices.

Results Genome-wide relationships between DNA methylation and gene expression are consistent with those in other mammals. Interestingly, we observed a clear and significant global reduction (hypomethylation) of DNA methylation across the entire female X chromosome compared to male X. Female hypomethylation does not directly contribute to the gene silencing of the inactivated X chromosome nor does it significantly drive sex-specific gene expression in tree shrews. However, we identified a putative regulatory region in the 5' end of the X-inactive-specific transcript (*Xist*) gene, whose pattern of differential DNA methylation strongly relate to its sex-differential expression in tree shrews. Furthermore, differential methylation of this region is conserved across different species. We also provide evidence suggesting that the observed difference between human and tree shrew X-linked promoter methylation is associated with the difference in genomic CpG contents.

Conclusions Our study offers novel information on genomic DNA methylation of tree shrews as well as insights into the evolution of sex chromosome regulation in mammals. Specifically, we show conserved role of DNA methylation in regulation of *Xist* expression and propose genomic CpG contents as a factor in driving sex-differential DNA methylation of X-linked promoters.

Keywords Tree shrew (*Tupaia belangeri*), DNA methylation, Gene expression, X chromosome inactivation, *Xist*, CpG contents

[†]Dongmin R. Son and Yifan Kong contributed equally to this work.

*Correspondence:

Lei Shi
shilei@mail.kiz.ac.cn
Soojin V. Yi
soojinyi@ucsb.edu

Full list of author information is available at the end of the article



Background

The tree shrew (*Tupaia belangeri*) is a small mammal widely found in Southeast Asia and China. The tree shrew offers several advantages to be used as a model species for biomedical studies. For example, the tree shrew has a small body size and short life span, making it easy to rear in laboratories for experimental studies [1]. On the other hand, the tree shrew has a high brain to body ratio, and exhibit several conditions that can model human disorders [2]. Several studies have demonstrated greater genetic similarities between tree shrew and primates than between rodents and primates, especially in genes associated with neuropsychiatric disorders and infectious diseases [3, 4]. Indeed, the tree shrew is the closest group of mammals to primates, providing a useful model system especially in neuroscience [3].

While genetic and transcriptomic studies of tree shrews are becoming available in the literature, epigenetic mechanisms such as DNA methylation of tree shrews have been little explored so far. Given the important role of DNA methylation in regulatory processes such as regulation of gene expression, neuropsychiatric diseases, and the X chromosome inactivation (XCI), such data will advance our understanding of regulatory evolution and enhance the utility of tree shrew as a model species. Here, we generated and analyzed whole genome DNA methylation maps (methylomes) of tree shrew prefrontal cortex from three males and three females. Integrating them with transcriptomic data, we demonstrate genome-wide influence of DNA methylation on gene expression, including the presence of CG and CH methylation which are both associated with gene expression.

Moreover, we examined the role of DNA methylation in the regulation of X chromosome inactivation (XCI). Notably, we show that global differential DNA methylation between the male and female X chromosome is not a driver of XCI in tree shrews. On the other hand, we newly annotated the X-inactive-specific transcript (*Xist*) gene and identified an evolutionary conserved locus of regulation in the 5' end of the gene. Comparing patterns of DNA methylation and X chromosome regulation across different species, we show that the evolutionary patterns of X chromosome DNA methylation is closely associated with the difference in genomic CpG contents. Additionally, we identified putatively Y-linked genomic segments and their hypomethylation. These novel findings illuminate conserved and divergent patterns of genomic DNA methylation and regulation of X chromosome in mammals.

Results

Genomic DNA methylation in tree shrew

We generated the first whole-genome bisulfite sequencing (WGBS) data from the prefrontal cortex (herein referred to as 'PFC') of 6 Chinese tree shrews (3 males and 3 females) to produce DNA methylomes at nucleotide resolution (Additional File 1: Table S1). Previous studies of human and mouse brain methylomes have identified CG and non-CG (CH where H is A, T, C) DNA methylation [5, 6]. Tree shrew PFCs were also highly methylated at CG sites while CH methylations were observed at lower levels compared to CG methylation (Fig. 1A). We observed significantly lower levels of CG methylation in promoter regions (defined as 2 kb upstream of the TSS) than in gene bodies and a sharp drop in CG methylation near transcriptional start sites (TSS) (Fig. 1B, Additional File 1: Fig. S1). Gene body CG methylation levels were higher than nearby intergenic regions. DNA methylation at non-CG sites remained relatively consistent in all genomic context with a slight drop near TSS (Additional File 1: Fig. S2). These patterns were similar across all chromosomes except for the chromosomes 13 and 26, where chromosome 13 showed a significant drop near the end of genes and chromosome 26 exhibited strikingly lower methylation levels across all genomic contexts (Additional File 1: Fig. S2), even though the read coverages of these chromosomes were similarly high to those of other chromosomes (8.7x ~ 13.5x for chromosome 13 and 11.1x ~ 15.2x for chromosome 26).

Integrating methylomes with RNA-seq data from the same samples (Additional File 2: Table S2), all 22,741 protein-coding genes were ranked into 0–20 (lowest to highest) according to their expression levels (Fig. 1C). While genes in all ranks showed hypomethylation near the TSS, highly expressed genes had the lowest methylation levels near TSSs while 0 or lowly expressed genes had relatively higher methylation levels (Fig. 1C, Additional File 1: Fig. S1B), resulting in a significant negative correlation between gene expression and promoter methylation (Additional File 1: Fig. S3, Table 1). In addition, protein-coding genes showed a more pronounced methylation drop near TSS compared to lncRNA genes (Additional File 1: Fig. S2).

For gene bodies, the relationship was slightly negative but not as monotonic as in promoters, with smaller methylation differences between lowly and highly expressed genes (Fig. 1C, Additional File 1: Fig. S1B). Although we identified a negative correlation that was significant for protein-coding genes (Table 1), the overall correlation was bell-shaped, with both lowly and highly expressed genes showing relatively lower

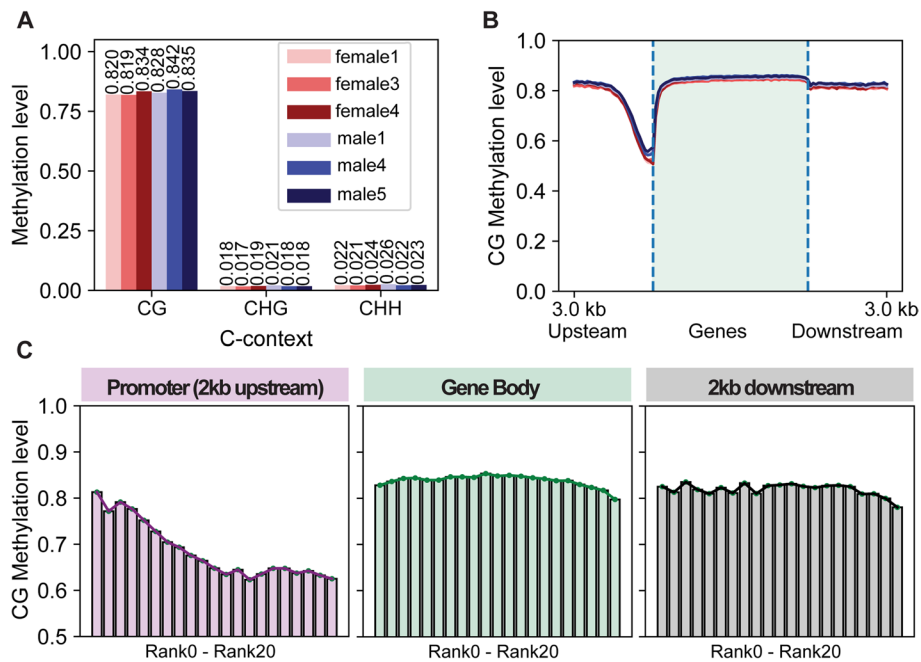


Fig. 1 DNA methylomes of tree shrew and the correlation between DNA methylation and gene expression. Chromosomes 13 and 26 (1001 genes) were excluded due to their unique patterns compared to other chromosomes. **A** Global (weighted) DNA methylation levels of CG, CHG, and CHH in each sample exhibit high levels of CpG methylation and low levels of CH methylation. **B** DNA methylation across gene bodies of 22,741 protein-coding genes in the autosomes and the X chromosome demonstrate decreases of DNA methylation near the transcription start sites. **C** DNA methylation of promoter, gene body, and intergenic regions in 20 groups of genes with different expression levels, ranging from rank 0 to rank 20 (higher expressed genes from left to right). A negative correlation is observed in promoters, while a bell-shaped correlation is observed in gene bodies

Table 1 Correlation analysis of mean promoter and gene body DNA methylation levels and ranked gene expression of genes of an autosome (chromosome 8) and the X chromosome of the tree shrew. (A) for protein-coding genes and (B) for lncRNA genes. Spearman’s rank correlation coefficient (ρ) and P -value for each chromosome in males and females are shown

(A) Protein-coding genes	Chromosome	Number of genes analyzed	Genomic Region	Spearman’s ρ (P -value) (male)	Spearman’s ρ (P -value) (female)
	Chr 8	597	Promoter	-0.31 (7.24×10^{-15})	-0.34 (4.95×10^{-17})
			Gene body	-0.29 (5.49×10^{-13})	-0.34 (2.31×10^{-17})
	Chr X	823	Promoter	-0.34 (5.58×10^{-23})	-0.31 (6.34×10^{-20})
			Gene body	-0.31 (1.42×10^{-18})	-0.26 (2.83×10^{-13})
(B) lncRNA genes	Chromosome	Number of genes analyzed	Genomic Region	Spearman’s ρ (P -value) (male)	Spearman’s ρ (P -value) (female)
	Chr 8	1759	Promoter	-0.06 (1.76×10^{-2})	-0.04 (7.96×10^{-2})
			Gene body	-0.07 (1.73×10^{-3})	-0.10 (1.53×10^{-5})
	Chr X	1058	Promoter	-0.09 (1.14×10^{-3})	-0.11 (2.60×10^{-4})
			Gene body	-0.04 (1.79×10^{-1})	-0.04 (2.19×10^{-1})

methylation compared to genes with median expression levels (Fig. 1C, Additional File 1: Fig. S3). While earlier studies in mammals linked high gene body

methylation to high expression [6–8], the results in tree shrews align with findings in humans, suggesting that the relationship is more complex than previously

thought—non-monotonic and bell-shaped [9]—and showing only slight differences in gene body methylation, as seen in a mouse study [10].

Differential DNA methylation of female and male tree shrew X chromosomes

We compared DNA methylation of autosomes and the X chromosome in the tree shrews PFC. We selected chromosome 8 as a representative autosome for comparison with the X chromosome, since it has a similar size and GC-content to the X chromosome. The mean coverage depth at CpG sites across the X chromosome was approximately twice as high in females as in males, as expected from the 2:1 ratio of the X chromosome in females compared to in males (Additional File 1: Fig. S4). We identified a region on the X chromosome where reads were mapped only in males, both in the methylomes and in the transcriptomes (Additional File 1: Fig. S4). We inferred that this included a part of the Y-linked regions that were incorrectly assembled into the X chromosome in the reference assembly (ChrX: 4,542,400–6,144,400) and subsequently removed it from further analysis (Methods).

We observed that the X chromosomes of females displayed significantly lower levels of DNA methylation in gene bodies and promoters compared to those of males ($P=6.13 \times 10^{-110}$ and 1.02×10^{-195} for promoters and gene bodies, respectively, Mann–Whitney *U* test, Fig. 2, Table 2). Comparisons to the autosomes demonstrated that this pattern was due to the reduced DNA methylation in females or “female hypomethylation” (Fig. 2A and B, Additional File 1: Fig. S3). This pattern was consistently observed across the entire X chromosome and was pervasive across different functional regions and also in non-CpG methylation (Fig. 2C, Additional File 1: Fig. S6). Furthermore, most of the differentially methylated regions (DMRs) between females and males (Additional File 2: Table S3) were located on the X chromosome, with a total length of 890,424 bp, compared to the total

length of 155,177 bp across all autosomes (Additional File 1: Fig. S7). The observation of female X hypomethylation was not due to a bias from different read depths between females and males, as we observed the same patterns when we re-analyzed equal depth data by randomly sampling half of the sequencing reads on female X chromosomes (Additional File 1: Fig. S8). The observed pattern resembled that in marsupial koalas and differed from that in humans, where promoters were typically hypermethylated in females [11].

Promoter methylation difference between females and males associate with CpG counts

We found that the degree of female hypomethylation, compared to males, in tree shrew X-linked promoters was highly depended on the density of CpGs within the promoter regions. First, female hypomethylated promoters had fewer numbers of CpGs compared to female hypermethylated promoters (Mann–Whitney *U* test, $P=6.70 \times 10^{-39}$, Fig. 2E, note that these counts reflect CpG density as the total number of nucleotides are the same for all 2 kb-sized promoters). Second, the difference between female and male promoter methylation decreased as the number of CpGs with promoters increased (Fig. 2F).

We propose that the dependency of promoter hypomethylation on CpG counts can explain the observed difference between species. We compared the distribution of CpG contents in human, using the metric CpG observed/expected ratio (CpG O/E) [12]. We found that while human X chromosome and autosomes included a large number of high CpG O/E promoters, koala X promoters mostly consist of low CpG O/E promoters (Fig. 2G, Additional File 1: Fig. S9). Tree shrew promoters resembled the pattern observed in koalas, where most X-linked promoters had low CpG O/E (Fig. 2G, Additional File 1: Fig. S9), which we found to be associated with female hypomethylation (Fig. 2F). In contrast, CpG

(See figure on next page.)

Fig. 2 Global patterns of female X hypomethylation in the tree shrew PFC and its association with CpG counts. **A** Mean fractional DNA methylation levels of all CpGs in males and females demonstrates that the female X chromosome is globally hypomethylated compared to chromosome 8 and the male X chromosome. **B** Distributions of DNA methylation level differences of CpG sites between females and males in autosomes and the X chromosomes show that the X chromosome is generally hypomethylated. **C** The distribution of expression levels, promoter DNA methylation levels, and gene body DNA methylation levels of genes (1857 genes including both protein-coding genes and lncRNA genes) across the X chromosome in females (red) and males (blue). **D** The differences between females and males for expression, promoter methylation and gene body DNA methylation of genes across the X chromosome. The yellow dot in the gene expression plot represents the *Xist* gene, which is upregulated in females. **E** Genes with female hypomethylated (mean 5mC male–female > 0.05, 1154 genes) promoters tend to have fewer CpGs compared to those with female hypermethylated (mean 5mC male–female < -0.05, 155 genes) promoters. **F** Comparisons of DNA methylation (Y-axis on the left) levels and their differences between males and females (Y-axis on the right) according to the numbers of CpGs in promoters (X-axis). Female promoters are clearly hypomethylated compared to male promoters when CpG counts are low. As CpG counts increases, both promoters are generally lowly methylated. Promoters with large CpG counts (> 80) are on average female hypermethylated. **G** A comparison of CpG O/E ratios in the promoter regions among humans, koalas, and tree shrews, showing a similarity between koalas and tree shrews compared to humans, specifically in the X chromosome promoters

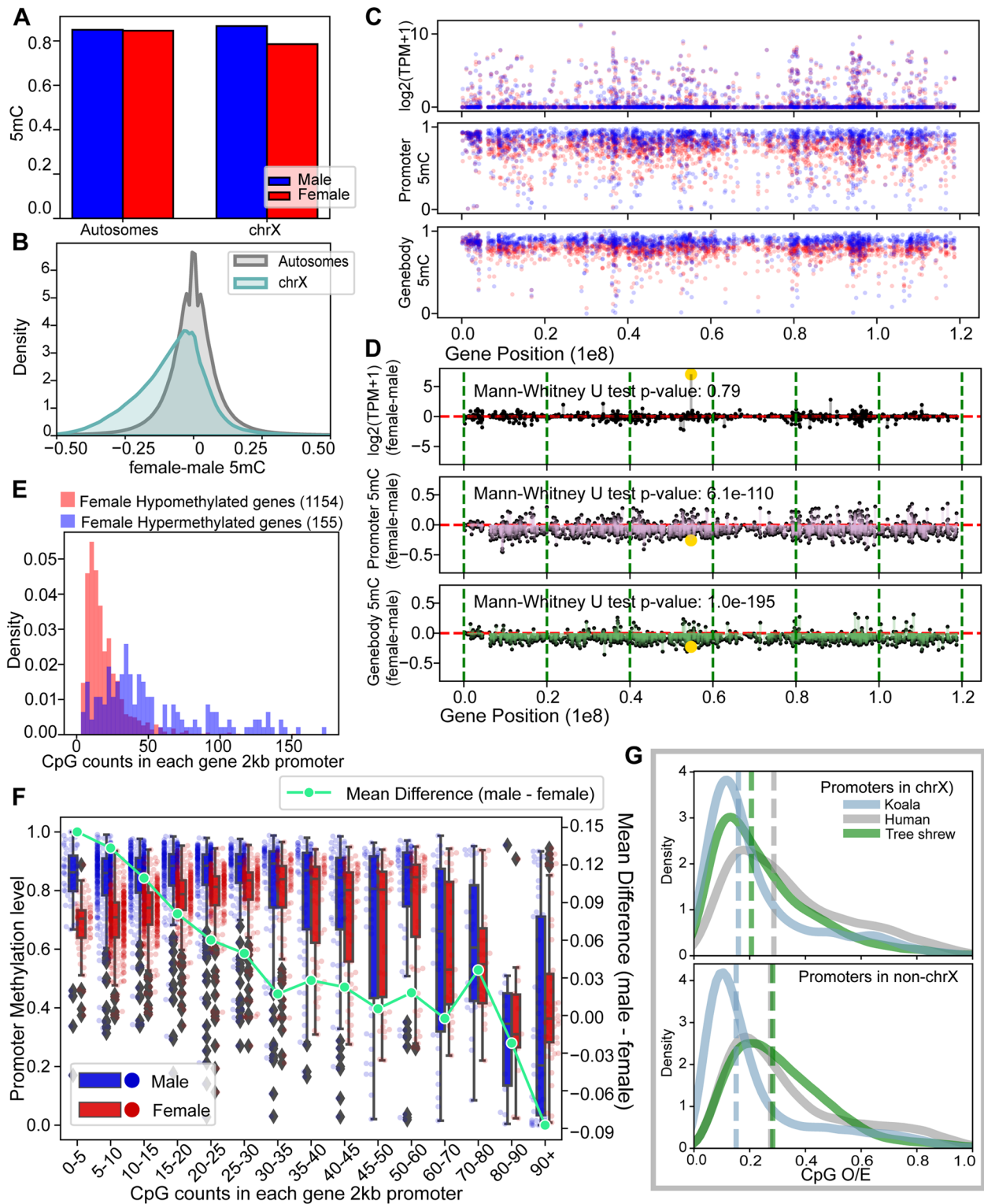


Fig. 2 (See legend on previous page.)

Table 2 Differences in DNA methylation levels and expression levels between female and male tree shrews, along with Mann–Whitney *U* test results and its *P*-values

		Number of genes compared	Mann–Whitney <i>U</i> test <i>P</i> -value
Chr 8	Promoter 5mC	2324	0.12
	Gene body 5mC	2245	0.02
	Expression	2341	0.86
Chr X	Promoter 5mC	1790	6.13×10^{-110}
	Gene body 5mC	1667	1.02×10^{-195}
	Expression	1856	0.79

O/E from autosomes were similar between humans and tree shrews compared to that in koalas (Additional File 1: Fig. S9).

Hypomethylation of the tree shrew female X does not drive sex-specific expression

If the primary functional outcome of X chromosome DNA hypomethylation is an upregulation in gene expression, as supported by the negative correlation observed between DNA methylation levels and gene expression (Table 1), then we would expect to see an overall increase in gene expression across female X-linked genes. However, this was not the case and we observed no global difference of gene expression between the male and female X chromosomes, as expected under the functional XCI (Fig. 2D, Table 2). Similarly, chromosome 8, where there was no discernible DNA methylation difference between males and females, showed no global difference of gene expression between males and females (Additional File 1: Fig. S5, Table 2). Therefore, we concluded that female hypomethylation of the tree shrew X chromosome did not lead to upregulation of genes.

Out of the total 32,302 autosomal genes including both protein-coding and non-coding RNA genes, 479 genes were significantly differentially expressed between males and females (adjusted *P*-value < 0.1 based on DESeq2, Fig. 3A, Additional File 2: Table S4). In comparison, 14 out of 783 X-linked genes were

significantly differentially expressed (Fig. 3C, Additional File 2: Table S5). The proportion of sex-specific genes was not statistically different between the autosomes and X chromosomes (Chi-square test, *P* = 0.48).

Across the entire set of autosomes, there was a slight excess of male-biased genes, characterized by a greater number of male upregulated genes compared to female upregulated genes. This trend remained consistent when different *P*-value thresholds were employed (Additional File 1: Fig. S10). On the other hand, there was an excess of female biased genes on the X chromosome (Additional File 1: Fig. S10). The average log2 fold-change of female to male expression was –0.097 for autosomes and –0.024 for chromosome X. However, it is worth noting that this difference was not significant (Mann–Whitney *U* test, *P* = 0.13, Fig. 3B) and the average log2 fold-change values showed variability depending on the stringency of filtering steps in the analysis (Additional File 1: Fig. S10).

We examined whether female chromosome X hypomethylation contribute to the sex-specific expression. Among the genes on the X chromosome, 443 gene had available information regarding log2 fold-change expression and the gene body DNA methylation level, while 439 genes had available information regarding log2 fold-change expression and the promoter DNA methylation level. Including these genes, we conducted an analysis of the relationship between gene expression and DNA methylation levels (Fig. 3D and E). We found that DNA methylation difference between females and males on the X chromosome did not exhibit a significant correlation with the corresponding gene expression difference (Spearman’s rank correlation test: $\rho = -0.004$, *P* = 0.94 for gene bodies; $\rho = 0.08$, *P* = 0.12 for promoters). Likewise, there was no significant trend between DNA methylation difference and gene expression difference when only significant sex-specific genes were analyzed (Spearman’s rank correlation test: $\rho = -0.24$, *P* = 0.46 for gene bodies; $\rho = -0.18$, *P* = 0.57 for promoters) (Fig. 3D and E). These observations suggest that female X chromosome hypomethylation does not contribute to sex-specific gene expression.

(See figure on next page.)

Fig. 3 Patterns of differential expression between males and females in relation to differential DNA methylation. Sex-specific expressed genes and their correlation with DNA methylation levels. **A** The MA plot illustrating differentially expressed genes across autosomes (left) and the X chromosome (right). Ashr-shrunken log fold-change values are used for the visualization. Blue dots represent male upregulated genes and red dots represents female upregulated genes. **B** The density distribution graph displays the log-transformed female-to-male expression ratio for genes. There was no significant difference between the X chromosome and autosomes (Mann–Whitney *U* test *P*-value = 0.13). **C** The distribution of female (red) and male (blue) upregulated genes identified by DESeq2 across the X chromosome. The *Xist* gene is marked with a yellow box. **D, E** The Y-axis represents the difference in mean DNA methylation levels between females and males is across gene bodies (**D**) and promoters (**E**). The X-axis represents the log2 fold-change of female-to-male expression difference. Spearman’s rank correlation coefficient and *P*-value are reported

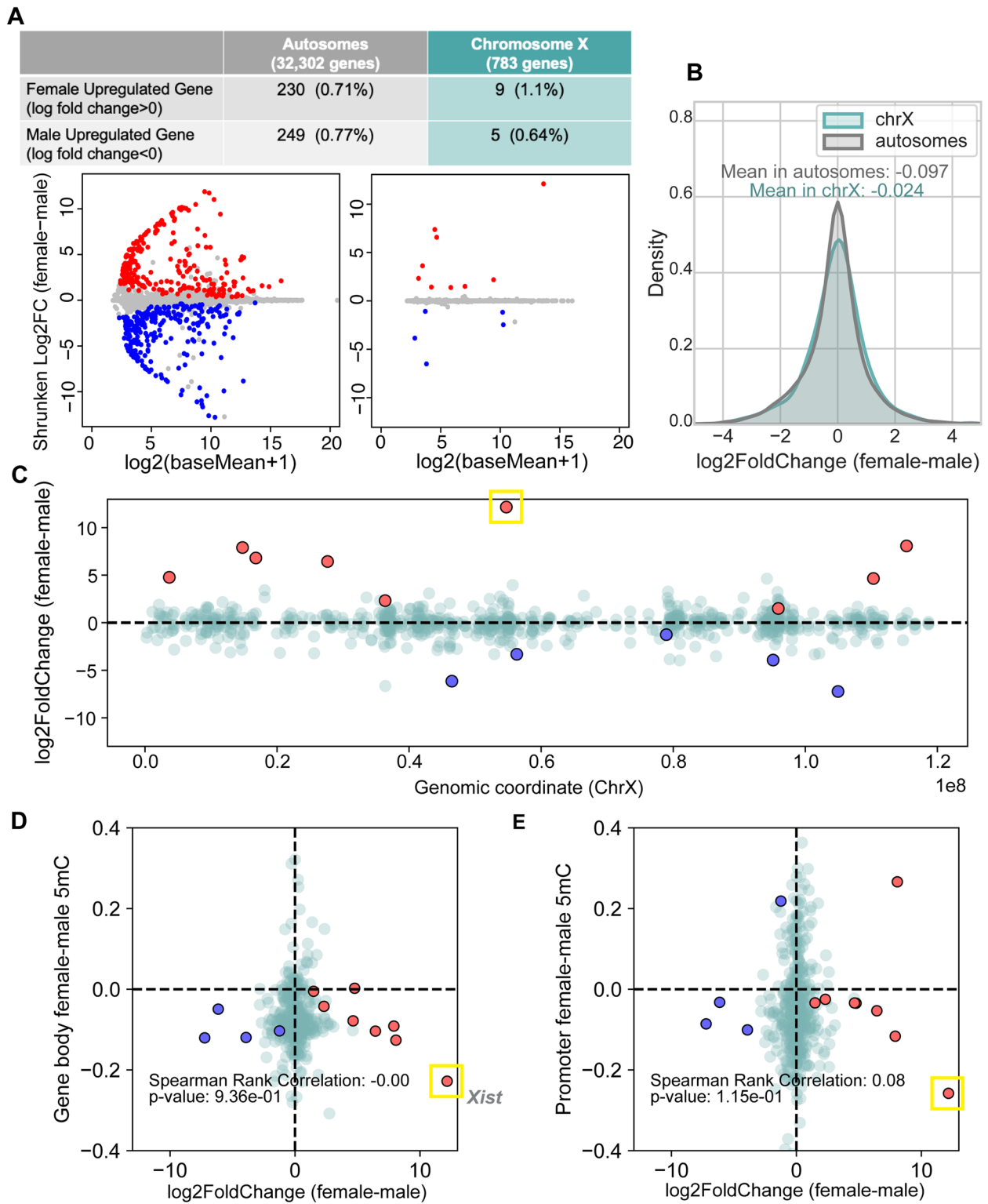


Fig. 3 (See legend on previous page.)

Differential methylation of the *Xist* associated with its sex-specific expression

The *Xist* gene is known for its key role in the initiation and maintenance of XCI in mammals [13, 14]. However, how *Xist* gene is regulated in the tree shrew remained unknown prior to our study. In fact, the current tree shrew genome database did not include a specifically annotated *Xist* gene. Here, we first identified the putative *Xist* gene region in the tree shrew genome. Briefly, we conducted a BLASTN search using the human *Xist* gene sequence as a query against the tree shrew reference genome to obtain a potential genomic coordinate

of *Xist* gene. We then identified novel transcripts from tree shrew RNA-seq samples using StringTie’s functionality for de novo transcript assembly (See Method). We identified a single gene from these novel transcripts. We found that longer novel transcripts were produced from female samples compared to male samples (Fig. 4A). Furthermore, this gene exhibited extremely high expression levels in our female data (Fig. 3C). The new annotation puts the tree shrew *Xist* on chr X: 54,721,977–54,759,147.

The *Xist* displays the most pronounced differential expression between the male and female tree shrew PFC (Fig. 2D, Fig. 3C, indicated by a yellow box), consistent

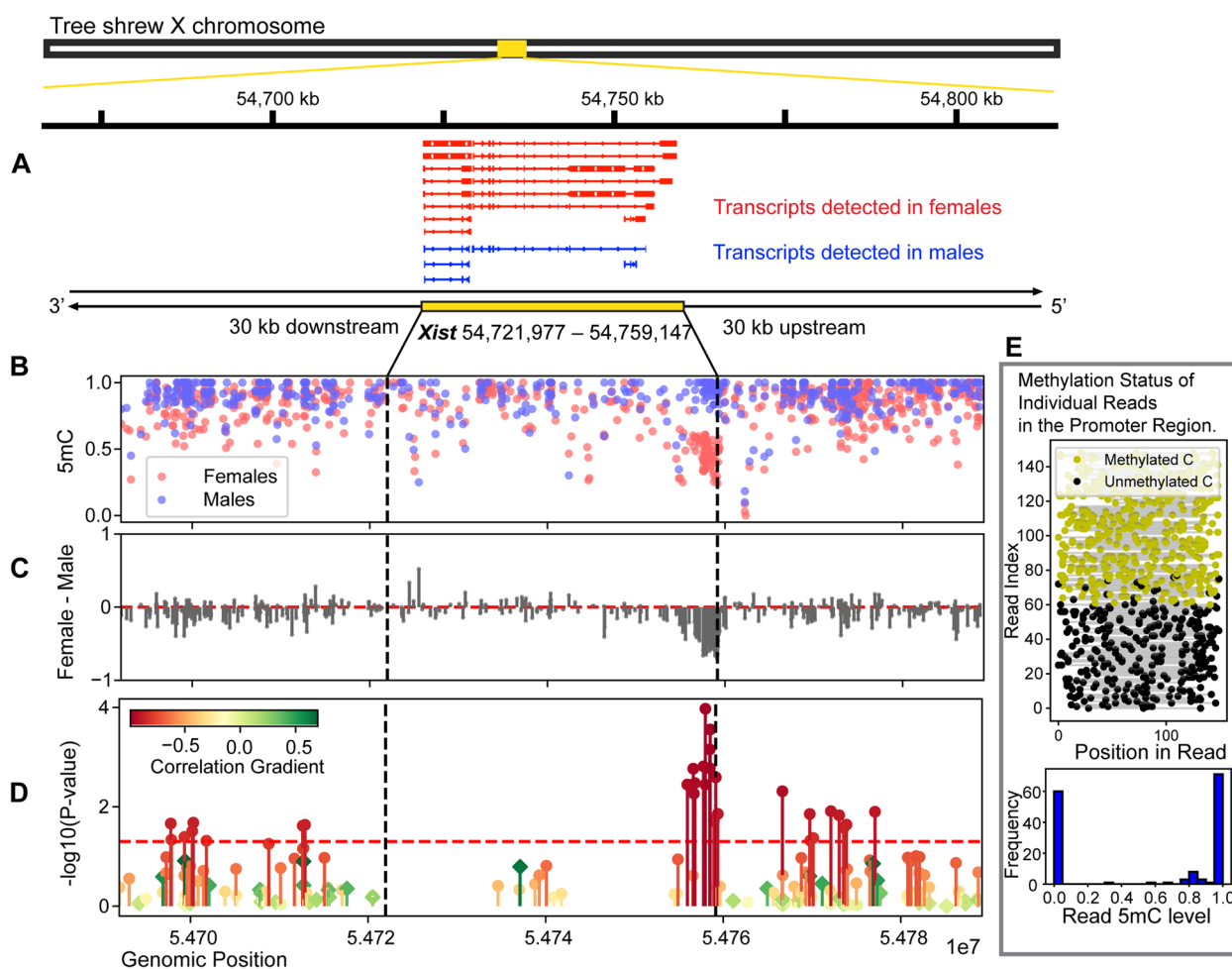


Fig. 4 DNA methylation difference between females and males for the newly annotated tree shrew *Xist* gene. **A** Longer transcripts are detected in female samples compared to male samples in the *Xist* gene region. **B** Fractional methylation levels of CpG sites around *Xist* are indicated and **C** the differences of DNA methylation levels of each CpG site between females and males are calculated. A CpG island near the 5' end of the gene displays marked female hypomethylation. **D** The methylation levels at each CpG site are correlated with gene expression levels using six samples. The Y-axis values represent the significance of the Pearson correlation, with a red horizontal line indicating the threshold for a P-value of 0.05. The color gradients indicate their corresponding Pearson correlation coefficient, with red for a negative correlation and green for a positive correlation. Circular points represent female-hypomethylated CpGs, while diamond-shaped points indicate male-hypomethylated CpGs. **E** The methylation states of individual reads from female1 mapped to the *Xist* promoter (positioned between chrX: 54,755,000–54,761,000) indicate allele-specific methylation, with the reads being either 0% or 100% methylated

with its key role in X-chromosome inactivation (XCI) in eutherian mammals. While *Xist* is a rapidly evolving non-coding RNA, specific regions are known to exhibit a high degree of conservation among eutherian mammals. Notably, the 5' end of *Xist* is known to be relatively well conserved compared to other regions [8, 15, 16]. Specifically, the 5' region of *Xist*, including a CpG island, is hypomethylated on the inactive X chromosome and highly methylated on the active X chromosome to maintain *Xist* repression in human and mouse [16]. DNA methylation in this region is known to play an active role in controlling *Xist* transcription in mouse [17–19] and in human somatic cells [20]. We therefore investigated differential DNA methylation between females and male tree shrews within *Xist* region (across the 37.2 kb gene body and 30 kb upstream and downstream regions) and identified a CpG island near the 5' end that exhibited a striking pattern of female hypomethylation (Fig. 4B, C). We examined the relationships between DNA methylation of CpGs in this region and the *Xist* gene expression levels across the six samples (Fig. 4D) and found several differentially methylated CpG positions within this female hypomethylated CpG island exhibiting significant correlations (13 CpGs located near the 5' end of *Xist* gene with Pearson correlation, $P < 0.05$). Furthermore, analysis of the methylation state of individual sequence reads in this region suggests that it is allele-specific, with one allele being hypermethylated and the other hypomethylated (Fig. 4E). Even though we do not have information on the active and inactive X, based on the prior knowledge in other mammals [16, 21], these hypomethylated reads are likely to arise from the inactive X chromosome. These results indicate that the tree shrew *Xist* gene is regulated by allele-specific methylation of CpGs that are female-hypomethylated and inversely correlating with expression. Therefore, regulation of *Xist* by hypomethylation of its 5' end CpG island in the inactive X chromosome is conserved across mouse, human, and tree shrew. We processed the public WGBS data of these species and found it can be consistently detected using WGBS data (Additional File 1: Fig. S11).

Identification of Y-linked contigs and the lower methylation level of the Y chromosome

Leveraging the availability of both female (XX) and male (XY) samples, we explored the level of DNA methylation in the Y-chromosome, a topic sparsely explored so far. The reference tree shrew genome lacks the Y chromosome assembly. To detect Y-linked segments, we developed a method to compare CpG read-depths between female and male samples. Scanning the contigs currently unassembled using this method, we detected potentially Y-linked segments as those where female samples showed

a lack of read counts compared to male samples. We demonstrate our methodology by computing $1 - (\text{mean read-depth in females}) / (\text{mean read-depth in males})$ in Fig. 5. For autosomal-linked contigs, since mean read-depths should be similar between males and females, this metric should be near zero. For X-linked contigs, the mean read-depth in females should be greater than that in males, and this metric should be less than zero. For Y-linked contigs, this metric should be closer to 1. Indeed, we observed three distinct peaks as expected: the left peak likely represents X-linked contigs, the middle peak corresponds to autosome-linked contigs, and the right peak indicates Y-linked contigs (Fig. 5A).

We applied a filter to identify 81 potential Y-linked contigs with aforementioned “ $1 - (\text{mean read-depth in females}) / (\text{mean read-depth in males})$ ” values greater than 0.95, 938 autosome-linked contigs with values between -0.5 and 0.5 , and 563 X-linked contigs with values smaller than -0.5 (Additional File 2: Table S6). The read-depth of each cytosine across these contigs was graphically represented to show a lack of female counts mapped (Fig. 5). Subsequently, we calculated and compared DNA methylation levels of these contigs against those of other chromosomes. Notably, the Y-linked contigs in the tree shrew PFC exhibited markedly lower methylation levels compared to both autosomes and the X chromosome (Fig. 5B, C).

Furthermore, contig18904 (23,877 bp in size) is identified as Y-linked using this methodology, wherein the sex-determining region Y (*Sry*) genes of humans and rhesus monkeys align in the tree shrew genome (Additional File 1: Fig. S12), exhibiting 79% identity for human-tree shrew alignment and 82% identity for rhesus monkey-tree shrew alignment based on BLASTN (see the “Methods” section). Interestingly, no expressed *Sry* transcripts are detected in our tree shrew prefrontal cortex transcriptome data of male samples. *Sry* gene expression is known to be tissue-specific and under the control of DNA methylation in mice and humans. [22–24]. Our study reveals the putative *Sry* locus in the tree shrew genome (Additional File 1: Fig. S12) and indicates its extremely low DNA methylation and low expression in PFCs.

Discussion

This study presents the first genome-wide analysis of DNA methylation of the emerging model species, the tree shrew. In human and mouse, DNA methylation of *cis*-regulatory regions is known to dampen the expression of associated genes [25]. DNA methylation of gene bodies is also known to contribute to regulation of gene expression, although the directionality is not as straightforward [9, 25, 26]. We observed significant and strong negative correlations between promoter DNA methylation and

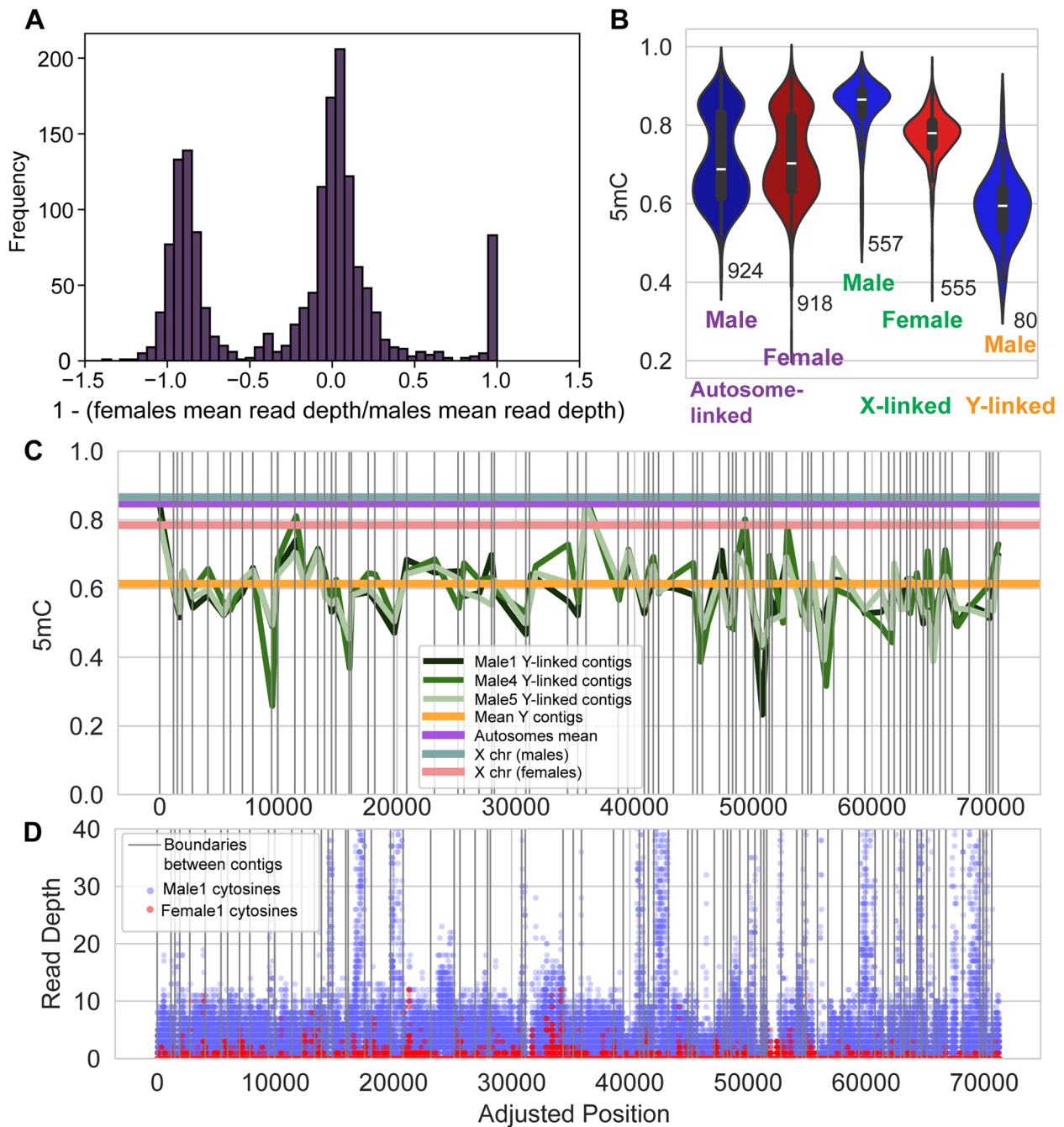


Fig. 5 DNA methylation in the Y-linked contigs of the tree shrew PFC. **A** The distribution of the metric $[1 - (\text{mean read-depth in females}) / (\text{mean read-depth in males})]$ for 1616 contigs. Three peaks are observed, which potentially correspond to from X-linked, autosome-linked, and Y-linked contigs. **B** Mean methylation levels for the identified contigs were analyzed. Notably, the putatively X-linked contigs exhibited female X hypomethylation. The number of contigs with available mean DNA methylation levels in each set is indicated. **C** The methylation levels of these contigs, compared with autosomes and the X chromosome. The Y-linked contigs exhibit markedly lower methylation levels than other chromosomes. **D** The read-depth of cytosines in the CG context are visualized for male 1 and female 1 sample across each Y-linked contig. The contigs are sorted and connected create an adjusted position

gene expression across the tree shrew genome, which is consistent with the aforementioned model as well as with previous studies [8–10, 27]. Gene body DNA methylation

was also associated with expression, although its effect was less pronounced compared to promoter methylation. We observed similar levels of overall gene expression between the male and female tree shrew X chromosomes,

as expected under functional XCI. Interestingly, female tree shrew X chromosome exhibited substantially lower level of DNA methylation compared to the male X chromosome (Fig. 2, Table 2). From the comparisons to autosomes and the male X chromosome, we demonstrate that this was due to the reduction of DNA methylation in the female X chromosome in tree shrews. Examining literature that have specifically addressed the global DNA methylation difference between the male and female X chromosome using chromosome-wide methods, Hellmann and Chess [28] demonstrated that gene bodies were hypomethylated in the human female X chromosome, and Keown et al. [29] showed that intergenic regions were hypomethylated in mouse female X chromosome. Sun et al. [30] and Singh et al. [11] analyzed WGBS data and showed that human female X chromosome was hypomethylated compared to the male X chromosome except for the promoter regions. Here, we show that in tree shrews, female X chromosome is globally hypomethylated compared to male X chromosome and autosomes, including both gene bodies and promoter regions, which is similar to the pattern observed in a marsupial, koalas [11].

While we observed that while the majority of the promoters (86.6%, Fig. 2E) was hypomethylated in female X, some promoters were hypermethylated in the female X chromosome of tree shrew. Intriguingly, these promoters were those harboring high density of CpGs (Fig. 2E). CpG-rich promoters are known to be associated with highly and broadly expressed genes in diverse vertebrate species [12]. Our findings of the close association between female promoter hypomethylation and CpG density provide insights into the observed difference between species where similar hypomethylation was observed in the promoters of female tree shrews and koalas, in contrast to the hypermethylation detected in human female X promoters. We show that CpG O/E ratio within X chromosome promoters is relatively similar between koalas and tree shrews, while the CpG O/E ratio between humans and tree shrews is similar across other chromosomes (Additional File 1: Fig. S9). In addition, a recent in-depth analysis of DNA methylation difference between the female and male human X chromosome demonstrated that CpG-rich promoters tend to be hypermethylated in female X, suggesting that a similar principle could be applied to variation within the human genome [31].

It is noteworthy that we found little difference in gene expression but strong differential methylation between the male and female X chromosomes, especially considering the pervasive genome-wide negative correlation between promoter methylation and gene expression. If the negative correlation between DNA methylation level

and expression level is directly involved in gene silencing in XCI, the female X chromosome would likely exhibit hypermethylation to silence the genes on the inactivated X chromosome. The result implies that the genome-wide pattern of negative correlation does not play a direct role in XCI in tree shrews. Future studies are needed to investigate other regulatory mechanisms that regulate XCI in this species.

While the global pattern of differential DNA methylation was not associated with differential gene expression of male and female X chromosome, we find evidence supporting the role of DNA methylation in the regulation of *Xist*, the key regulator of XCI in other mammals. We annotated *Xist* from the tree shrew genome and show that the female and male X chromosomes generate distinctive transcripts. The transcripts from the male X chromosomes tended to be short and only a few copies were identified, which may potentially arise from unstable RNAs known to continue being expressed on the active X chromosome while stable RNAs are accumulating on the inactive X chromosome [32, 33]. The *Xist* was highly expressed in the female X chromosome and was strongly differentially methylated between the female and male X chromosomes. We further identified a regulatory region (CpG island) in the 5' region of the *Xist* gene. The hypomethylation of this CpG island was tightly correlated with the expression of *Xist* and exhibited a pattern of allele-specific DNA methylation, consistent with the exclusive expression of *Xist* from the inactive X [16, 21]. Moreover, hypomethylation of that specific CpG island was observed in other mammalian species. These observations indicate that DNA methylation is a key mechanism of regulation of *Xist*, the initiator of XCI, in diverse mammals.

We also explored DNA methylation of the Y-linked segments, utilizing the abundance of reads thanks to the next-generation sequencing approach. Epigenetics of the mammalian Y chromosome is currently little understood in large part due to the difficulties associated with the sequencing and assembly of the Y chromosome. Nevertheless, we show that contigs that could be best explained by the Y-linkage exhibit markedly reduced levels of DNA methylation compared to the X chromosome and autosomes (Fig. 5). Even though we have not and cannot attempt to identify specific genomic regions from these data, it is notable that DNA methylation levels of the same putative Y-contigs from the three males are highly correlated (Fig. 5), indicating that we are observing reproducible patterns of DNA methylation from the tree shrew Y-chromosome. The observation that the putative Y-linked segments exhibit reduced methylation supports the idea that DNA methylation tends to be reduced in less transcriptionally active regions [34]. The

comprehensive DNA methylome and gene expression data from tree shrews provide new insights into the evolution of genome, methylome and their interactions on the regulation of X chromosomes.

Conclusions

Our first DNA methylomes from the tree shrew prefrontal cortex yield information on conserved and divergent patterns of DNA methylation and gene expression in mammals. In addition to confirming the association between DNA methylation and gene expression at genomic scale, we identified pervasive female X hypomethylation in tree shrew that was not associated with an increase of expression. This pattern appears as a conserved feature of mammalian X chromosomes. We annotated the *Xist* gene and demonstrated that the allele-specific hypomethylation of a 5' CpG island of this gene is a key regulatory feature of mammalian X chromosomes. We also annotated the *Sry* gene and show that the Y-linked contigs are hypomethylated compared to other chromosomes, advancing our understanding of this understudied chromosome. On the other hand, sex-specific DNA methylation of X-linked promoters displayed divergent patterns between species, which we hypothesize is associated with the divergence of genomic CpG contents. These findings provide novel understanding of the genomic DNA methylation system in tree shrews and offers insights into the evolution of X chromosome regulation by DNA methylation in mammals.

Methods

Tissue samples

The adult Chinese tree shrews used in this study were obtained from the Laboratory Animal Center of Kunming Institute of Zoology and handled in accordance with the guidelines approved by the Animal Care and Use Committee of the Kunming Institute of Zoology, Chinese Academy of Sciences. The anesthesia and euthanasia procedures were performed following a protocol adhering to the guidelines of the American Veterinary Medical Association (AVMA) Guidelines for the Euthanasia Animals. Briefly, each tree shrew was initially anesthetized by being placed in a small sealed box with a small amount of isoflurane near the nostril. It was then followed by a deep anesthesia and euthanasia with an intramuscular injection of 300 μ L ketamine (0.05 g/mL) and 300 μ L pentobarbital (0.7%) in sequence according to the weight of adult tree shrew (~200 g). Once the tree shrew did not respond to their toes being pinched, it indicated that the tree shrew had been in deep anesthesia. After manually confirming that the tree shrew was in deep anesthesia and reached apnea and cardiac arrest status,

the euthanasia step was completed. The tree shrews were removed from the cloth bag for dissection procedures. Following sacrifice of both male and female tree shrews (Sample information, Additional File 1: Table S1), their brains were dissected and cortex tissues were rapidly frozen using liquid nitrogen for long-term storage at -80°C . All protocols of this study were approved by the internal review board of Kunming Institute of Zoology, Chinese Academy of Sciences (Ethics Approval number: IACUC NO. IACUC-RE-2022-11-013).

RNA sequencing

Total RNA was extracted from PFC tissues using TRIzol Reagent, and the integrity of RNA was assessed using Agilent 2100 bioanalyzer. The library preparation began with total RNA as the initial template. Specifically, mRNA with PolyA tails was enriched from total RNA using Oligo(dT) magnetic beads. The resulting mRNA was then randomly fragmented by divalent cations in fragmentation buffer. The first strand cDNA was synthesized in the M-MuLV reverse transcriptase system using the fragmented mRNA. The second strand cDNA was subsequently synthesized in the DNA polymerase I system using dNTPs. The obtained double-stranded cDNA was purified, end repaired, and then poly-A tails and sequencing adaptors were ligated. AMPure XP beads were used to screen for cDNA fragments within the size range of 370–420 bp followed by PCR amplification and product purification. Library quality assessment was performed on the Agilent Bioanalyzer 2100 system, while cluster generation took place on the cBot Cluster Generation System. Finally, library preparations were sequenced on an Illumina Novaseq platform generating paired-end reads of 150 bp each.

Whole-genome bisulfite sequencing

Genomic DNA was extracted from PFC tissues by phenol–chloroform extraction and ethanol precipitation. The extracted genomic DNA was quantified using Qubit fluorometer, and the 200 ng of gDNA containing 1% unmethylated Lambda DNA was then randomly fragmented into 300 bp small fragments using the Covaris LE220R ultrasonic fragmentation instrument. These small fragments were then subjected to terminal repair and adenylation before being fitted with methylated adapters. The bisulfite treatment step was performed using the EZ DNA Methylation-Gold kit (Zymo Research) following the manufacturer's instructions. The resulting single-stranded DNA was PCR amplified and the PCR products were purified. Similarly, libraries obtained were quantified using Qubit fluorometer and their size distribution was analyzed by Agilent BioAnalyzer (Agilent). Paired-end sequencing

was performed using an Illumina NovaSeq6000 according to Illumina-provided protocols. Finally, standardized WGBS data analysis pipeline was employed for analyzing the resulting data.

Processing whole-genome bisulfite sequencing data

We first performed quality and adapter trimming using TrimGalore v0.6.7 (Babraham Institute) with paired-end mode and default parameters. Subsequently, reads were mapped to the tree shrew reference genome (TS_3.0) from the tree shrewDB [35, 36], using Bismark v0.24.0 [37]. Following deduplication using Bismark, we obtained coverage for over 96% CpG sites, with a read-depth between 12 and 18X (Additional File 1: Table S1). The female4 and male4 samples exhibited relatively lower mapping efficiency. Additionally, we mapped the reads to the lambda phage genome (NC_001416.1) to estimate the bisulfite conversion rate in each sample, resulting in values 99.1–99.3% (Additional File 1: Table S1).

The data, comprising the counts of methylated and unmethylated cytosines in each C-context at individual cytosines, were generated as cytosine report files using Bismark methylation extractor with `-bedGraph -cytosine_report -CX_context` options. These output cytosine reports were used as input files for ViewBS v0.1.11 [38]. CpG sites that are detected as a positional difference of 1 and are located on different DNA strands are merged into a single CpG site, considering the symmetrical nature of CpG methylation. Subsequently, we calculated the fractional methylation level by taking the ratio of methylated cytosine reads to the total read count for each cytosine with at least five reads, covering approximately 74–81% of the total 34,833,278 CpG sites in the tree shrew genome (Additional File 1: Fig. S13). These outputs were used in downstream analyses to calculate the mean methylation level of specific genomic regions. We also compiled an additional list of highly covered CpG sites, each with at least 10 reads, to demonstrate that there was no bias in our key findings due to different CpG coverage cutoffs (Additional File 1: Fig. S13).

Processing RNA-seq data

We mapped the RNA-seq reads to the tree shrew reference genome (TS_3.0) using HISAT2 v2.2.1 [39] with the `-dta` option (Additional File 2: Table S2). Here, the reference annotation information [36] were embedded in the genome index using HISAT2 `-ss` and `-exon` options. Transcripts were then assembled for each sample using StringTie v2.2.1 [40], and we generated an updated GTF annotation included novel transcripts using the `-merge` flag. This process was guided by the reference GTF annotation using the StringTie `-G` flag. We obtained transcript and gene abundance information using the `-eB` and

`-A` options. These output files were used to define gene regions and estimate expression levels.

Among the 224,473 transcripts identified by StringTie, 183,669 transcripts were guided by the reference genome, while the rest 40,804 transcripts were detected as novel transcripts. In the downstream analysis, we excluded 18,623 transcripts generated from StringTie that lacked strand information for promoter methylation level calculation.

Quantifying methylation levels

We employed ViewBS v0.1.11 [38] to estimate the global weighted methylation levels [41] and to analyze methylation landscape near gene and promoter regions. Cytosines in all C-context from all 30 chromosome and X chromosome were included in the analysis. We designated the longest transcript of each gene as a gene body in the process. For further analysis, we estimated the mean methylation level [41] in the gene body region for each individual gene, including those with more than three CpG sites with fractional methylation values covered by at least five reads. Promoter regions are defined as the 2 kb upstream region of each gene body's start site, and mean methylation levels are calculated in the same way.

Identification of sex-specific expressed genes and methylated regions

Following the processing of RNA-seq data, we obtained read count information for each sample. For the sex-specific gene analysis, we employed DESeq2 v1.38.3 [42]. We restricted our analysis to genes with at least 5 counts in 3 or more samples among total six samples (Additional File 1: Fig. S10). We identified sex-specific genes with an adjusted *P*-value less than 0.1 (Additional File 2: Table S4, Table S5). To detect differentially methylated regions (DMRs) between females and males, we used DSS v2.48.0 [43], specifying a minimum region length of 50 bp and a minimum of 3 CpG sites (default options). We also identified the closest gene to each DMR, generating a potential list of sex-specific expressed genes that may be regulated by sex-specific DMRs (Additional File 2: Table S3).

Sex-specific read-depths on the X chromosome

Comparing the sex-specific read-depths on the X chromosome, the average read-depths in CpG sites were roughly twice in females compared to males for the X-linked sites, corresponding to the number of the X chromosome in females and males (Additional File 1: Fig. S4, Fig. S8). We discovered that certain X-linked regions exhibited similar coverages between the female and male samples, which are potentially originated from the PAR (pseudoautosomal region) of the X chromosome.

Intriguingly, we also observed one region on the X chromosome (genomic coordinate 4,542,400–6144400) where female sample showed a complete lack of read counts (Additional File 1: Fig. S4). We hypothesized that region may contain portion of the Y chromosome that was wrongly annotated. This region also harbored several potential Y chromosome genes that exhibited 0 expression in female samples but significant expression in male samples (Additional File 1: Fig. S4). Based on these findings, we have concluded that this region is wrongly included in X chromosome reference from the Y chromosome. To address this issue, we have excluded both the CpG sites and genes located within this region. A total of 74 genes were identified within this region and subsequently removed from the X chromosome analysis. We identified an additional region with this characteristic on chromosome 26 (genomic coordinate < 3,132,930) and excluded 85 genes in the region in sex-specific gene analysis.

Identification of the *Xist* gene in the tree shrew genome

The *Xist* gene was not annotated in the reference annotation data. We initiated a process to locate the gene in the tree shrew X chromosome. Our approach involved performing BLASTN v2.13.0+ [44] analysis using the human *Xist* gene (NR_001564.2) against the tree shrew reference genome. This allowed us to identify a potential range of the genomic coordinate of the *Xist* gene. Subsequently, we employed StringTie's functionality for de novo transcript assembly [40]. We investigated the transcripts identified by StringTie and reference annotation in the potential range of the *Xist* gene. We found these transcripts exhibited female-specific expression patterns. Based on these findings, we defined these specific transcripts and their associated gene as the *Xist* gene located within the region of chrX 54,721,977–54,759,147 in the tree shrew genome. In examining differential DNA methylation between females and males within *Xist* region, we included all CpG sites detected at least twice in females or males and its fractional methylation levels were averaged in females and males each. Bismark BAM outputs, after the deduplication step, are used to investigate read-level DNA methylation states in the promoter region.

Comparative studies among tree shrew, human, and mouse

To compare the sex-specific pattern near the *Xist* gene region within tree shrew, human, and mouse, we downloaded the public data for two male and two female samples of mouse fetal brains from the Gene Expression Omnibus database with accession# GSE157553 [45, 46] and the data for one male and one female

human sample of prefrontal cortex with accession# GSE37202 [47, 48]. The processed files, along with the CpG coverage files for mice and the fractional methylation value files for humans, are formatted uniformly. Utilizing CpG sites with a read depth of 3 or greater, DNA methylation levels were calculated across 2-kb-sized windows overlapping by 1 kb around the *Xist* gene region for all species.

To compare the CpG density in the genomes of human, koala, and tree shrew, we calculated the CpG observed/expected ratio across 1000-bp-sized windows (with 500 bp overlaps) in the genome of each species. Each window is annotated as either a promoter (2000 bp upstream of TSS) or gene body, depending on whether the window center falls within the promoter or gene body region. We utilized the human genome T2T-chm13v2.0 genome and gene annotation [49] and the koala phaCin_unsw_v4.1 genome and RefSeq annotation for the analysis [50]. To identify X-linked regions, we referenced the list of X-linked scaffolds provided in [11].

Detection of Y-linked contigs and DNA methylation

We employed a bioinformatic method to identify putative Y-linked regions in our data. As the current tree shrew genome assembly does not provide a Y chromosome, we utilized the currently unassembled contigs in the tree shrew reference genome. We calculated the read-depth of cytosines in the CG context for each of these contigs and then averaged the values separately for male and female samples. Subsequently, for contigs containing more than 40 cytosines, we computed $[1 - (\text{mean read-depth in females}) / (\text{mean read-depth in males})]$. Using this calculated metric distribution (Fig. 5), we generated a set of potentially 938 autosome-linked contigs (with values between -0.50 and 0.50), 563 X-linked contigs (with values smaller than -0.5), and 81 Y-linked contigs (with values greater than 0.95). To validate our approach, we calculated the mean methylation levels for each set of contigs by averaging the cytosines covered by at least five reads in individual contigs, along with their corresponding read depths and GC contents (Table S6). BLASTN v2.13.0+ [44] is utilized to identify the sex-determining region Y (*Sry*)-related region in the tree shrew genome, employing the human *Sry* gene (NIH Gene ID 6736) and the rhesus monkey *Sry* gene (NIH Gene ID 574155).

Abbreviations

XCI	X chromosome inactivation
<i>Xist</i>	X-inactive-specific transcript
WGBS	Whole-genome bisulfite sequencing
PFC	Prefrontal cortex
TSS	Transcription start site
DMR	Differentially methylated region
<i>Sry</i>	Sex-determining region Y
PAR	Pseudoautosomal region

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12915-024-02071-0>.

Additional File 1: Table S1. Overview of samples from 3 male and 3 female tree shrews and the processing of whole-genome bisulfite sequencing (WGBS) data. Fig. S1. Overview of DNA methylation data. Fig. S2. DNA methylation patterns near gene region. Fig. S3. Relationships between DNA methylation and gene expression for promoters and gene bodies for the X chromosome and a representative autosome (Chromosome 8) in male and female tree shrews. Fig. S4. Sex-specific read-depth on the X chromosome. Fig. S5. Comparisons of promoter and gene body DNA methylation in females and males for chromosome 8. Fig. S6. Global patterns of female X hypomethylation in tree shrews. Fig. S7. Differentially methylated regions (DMRs) between female and male tree shrews in the prefrontal cortex. Fig. S8. The observed hypomethylation of female X chromosomes is not due to biases arising from differing read depths between female (XY) and male (XX) samples. Fig. S9. A comparison of CpG O/E ratios among humans, koalas, and tree shrews. Fig. S10. Overview of differentially expressed genes between the male and female tree shrews. Fig. S11. A conserved DNA methylation pattern near the *Xist* gene. Fig. S12. The putative *Sry* gene region. Fig. S13. Information on CpG site coverage in our data and the reproduction of results using highly covered CpG sites.

Additional File 2: Table S2. Summary of the RNA-seq dataset and the number of expressed genes detected on individual chromosomes at various raw count thresholds. Table S3. List of DMRs. A list of differentially methylated regions (DMRs) between female and male tree shrews, detected by DSS. The associated gene and the distance between the TSS and the DMR are provided. Table S4. List of DEGs (Autosome). A list of sex-specific differentially expressed genes (DEGs) between female and male tree shrews on autosomes, generated from DESeq2 outputs. Table S5. List of DEGs (ChrX). A list of sex-specific differentially expressed genes (DEGs) between female and male tree shrews on the X chromosome, generated from DESeq2 outputs. Table S6. List of contigs and their associated chromosomal information. Read depth for females and males and DNA methylation levels are calculated, and GC content is included.

Acknowledgements

We thank Yuhua Ma of experimental animal center at Kunming Institute of Zoology for taking care of tree shrew. We thank the members of Shi lab for critical reading the manuscript. We thank the California NanoSystems Institute and the Life Science Computing Group at UCSB for computational support and members of the Yi lab for comments and discussions on the previous versions of the manuscript.

Authors' contributions

LS and YK conceived the methylome and transcriptome sequencing. YK, YT and TH performed tree shrew brain dissection, RNA and DNA extraction, and generated the sequencing libraries. DRS performed the processing and computational analysis of genomic data, generated figures, and prepared the manuscript. SVY directed the analysis of the data, participated in developing analysis methods, interpreting the results, and preparing the manuscript. LS participated in preparing the manuscript and interpreting the results. All authors read and approved the final manuscript.

Funding

LS is supported by the Pioneer Hundred Talents Program of the Chinese Academy of Sciences and the Yunnan Revitalization Talent Support Program Young Talent Project. This study was supported by the grants from Natural Science Foundation of China (32170630), National Key Research and Development Program of China (2021YFF0702700), Science and Technology Major Project of Science and Technology of Department of Yunnan Province (202102AA100057), Yunnan Applied Basic Research Projects (202201AS070043 and 202401AS070072), and Spring City Project from Kunming Science and Technology Bureau (2022SCP007) to L.S. SVY is partially supported by National Science Foundation (EF-2021635) and National Institutes of Health (HG011641).

Data availability

Sequence data that support the findings of this study have been deposited to the GEO accession GSE270314 and GSE270313.

Declarations

Ethics approval and consent to participate

All protocols of this study were approved by the internal review board of Kunming Institute of Zoology, Chinese Academy of Sciences (Ethics Approval number: IACUC NO. IACUC-RE-2022-11-013).

Consent for publication

Authors consent for publication.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Ecology, Evolution and Marine Biology, University of California Santa Barbara, Santa Barbara, USA. ²Department of Molecular, Cellular and Developmental Biology, Neuroscience Research Institute, University of California Santa Barbara, Santa Barbara, USA. ³Key Laboratory of Genetic Evolution & Animal Models, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650201, China. ⁴Key Laboratory of Animal Models and Human Disease Mechanisms of Yunnan Province, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, Yunnan, China. ⁵National Resource Center for Non-Human Primates, Kunming Primate Research Center, and National Research Facility for Phenotypic & Genetic Analysis of Model Animals (Primate Facility), Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan 650107, People's Republic of China. ⁶Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming, Yunnan 650201, People's Republic of China.

Received: 27 June 2024 Accepted: 13 November 2024

Published online: 28 November 2024

References

1. Yao Y-G. Creating animal models, why not use the Chinese tree shrew (*Tupaia belangeri chinensis*)? *Zool Res.* 2017;38:118–26.
2. Li H, Xiang B-L, Li X, Li C, Li Y, Miao Y, et al. Cognitive deficits and Alzheimer's disease-like pathologies in the aged Chinese tree shrew. *Mol Neurobiol.* 2024;61:1892–906.
3. Fan Y, Huang Z-Y, Cao C-C, Chen C-S, Chen Y-X, Fan D-D, et al. Genome of the Chinese tree shrew. *Nat Commun.* 2013;4:1426.
4. Yamashita A, Fuchs E, Taira M, Yamamoto T, Hayashi M. Somatostatin-immunoreactive senile plaque-like structures in the frontal cortex and nucleus accumbens of aged tree shrews and Japanese macaques. *J Med Primatol.* 2012;41:147–57.
5. Jeong H, Mendizabal I, Berto S, Chatterjee P, Layman T, Usui N, et al. Evolution of DNA methylation in the human brain. *Nat Commun.* 2021;12:2021.
6. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature.* 2009;462:315–22.
7. Laurent L, Wong E, Li G, Huynh T, Tsigiris A, Ong CT, et al. Dynamic changes in the human methylome during differentiation. *Genome Res.* 2010;20:320–31.
8. Rauch TA, Wu X, Zhong X, Riggs AD, Pfeifer GP. A human B cell methylome at 100–base pair resolution. *Proc Natl Acad Sci.* 2009;106:671–8.
9. Jingo D, Conley AB, Yi SV, Lunyak VV, Jordan IK. On the presence and role of human gene-body DNA methylation. *Oncotarget.* 2012;3:462–74.
10. Al Adhami H, Bardet AF, Dumas M, Cleroux E, Guibert S, Fauque P, et al. A comparative methylome analysis reveals conservation and divergence of DNA methylation patterns and functions in vertebrates. *BMC Biol.* 2022;20:70.

11. Singh D, Sun D, King AG, Alquezar-Planas DE, Johnson RN, Alvarez-Ponce D, et al. Koala methylomes reveal divergent and conserved DNA methylation signatures of X chromosome regulation. *Proc R Soc B*. 2021;288:20202244.
12. Elango N, Yi SV. DNA Methylation and structural and functional bimodality of vertebrate promoters. *Mol Biol Evol*. 2008;25:1602–8.
13. Penny GD, Kay GF, Sheardown SA, Rastan S, Brockdorff N. Requirement for Xist in X chromosome inactivation. *Nature*. 1996;379:131–7.
14. Pontier DB, Gribnau J. Xist regulation and function eXplored. *Hum Genet*. 2011;130:223–36.
15. Brown CJ, Hendrich BD, Rupert JL, Lafrenière RG, Xing Y, Lawrence J, et al. The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell*. 1992;71:527–42.
16. Hendrich BD, Brown CJ, Willard HF. Evolutionary conservation of possible functional domains of the human and murine XIST genes. *Hum Mol Genet*. 1993;2:663–72.
17. Beard C, Li E, Jaenisch R. Loss of methylation activates Xist in somatic but not in embryonic cells. *Genes Dev*. 1995;9:2325–34.
18. Norris DP, Patel D, Kay GF, Penny GD, Brockdorff N, Sheardown SA, et al. Evidence that random and imprinted Xist expression is controlled by preemptive methylation. *Cell*. 1994;77:41–51.
19. Panning B, Jaenisch R. DNA hypomethylation can activate Xist expression and silence X-linked genes. *Genes Dev*. 1996;10:1991–2002.
20. Tinker AV, Brown CJ. Induction of XIST expression from the human active X chromosome in mouse/human somatic cell hybrids by DNA demethylation. *Nucleic Acids Res*. 1998;26:2935–40.
21. Brown CJ, Ballabio A, Rupert JL, Lafreniere RG, Grompe M, Tonlorenzi R, et al. A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature*. 1991;349:38–44.
22. Gimelli G, Giorda R, Beri S, Gimelli S, Zuffardi O. A 46, X, inv(Y) young woman with gonadal dysgenesis and gonadoblastoma: cytogenetics, molecular, and methylation studies. *Am J Med Genet A*. 2006;140:40–5.
23. Larney C, Bailey TL, Koopman P. Switching on sex: transcriptional regulation of the testis-determining gene Sry. *Development*. 2014;141:2195–205.
24. Nishino K, Hattori N, Tanaka S, Shiota K. DNA methylation-mediated control of Sry gene expression in mouse gonadal development. *J Biol Chem*. 2004;279:22306–13.
25. Schübeler D. Function and information content of DNA methylation. *Nature*. 2015;517:321–6.
26. Huh I, Zeng J, Park T, Yi SV. DNA methylation and transcriptional noise. *Epigenetics Chromatin*. 2013;6:9.
27. Mendizabal I, Berto S, Usui N, Toriumi K, Chatterjee P, Douglas C, et al. Cell type-specific epigenetic links to schizophrenia risk in the brain. *Genome Biol*. 2019;20:135.
28. Hellman A, Chess A. Gene body-specific methylation on the active X chromosome. *Science*. 2007;315:1141–3.
29. Keown CL, Berletch JB, Castanon R, Nery JR, Distechi CM, Ecker JR, et al. Allele-specific non-CG DNA methylation marks domains of active chromatin in female mouse brain. *Proc Natl Acad Sci U S A*. 2017;114:E2882–90.
30. Sun D, Maney DL, Layman TS, Chatterjee P, Yi SV. Regional epigenetic differentiation of the Z Chromosome between sexes in a female heterogametic system. *Genome Res*. 2019;29:1673–84.
31. Morgan R, Loh E, Singh D, Mendizabal I, Yi SV. DNA methylation differences between the female and male X chromosomes in human brain. *bioRxiv*. 2024;2024.04.16.589778.
32. Panning B, Dausman J, Jaenisch R. X chromosome inactivation is mediated by Xist RNA stabilization. *Cell*. 1997;90:907–16.
33. Sheardown SA, Duthie SM, Johnston CM, Newall AE, Formstone EJ, Arkell RM, et al. Stabilization of Xist RNA mediates initiation of X chromosome inactivation. *Cell*. 1997;91:99–107.
34. Makova KD, Pickett BD, Harris RS, Hartley GA, Cechova M, Pal K, et al. The complete sequence and comparative analysis of ape sex chromosomes. *Nature*. 2024;630:401–11.
35. Fan Y, Yu D, Yao Y-G. Tree shrew database (TreeshrewDB): a genomic knowledge base for the Chinese tree shrew. *Sci Rep*. 2014;4:7145.
36. Ye M-S, Zhang J-Y, Yu D-D, Xu M, Xu L, Lv L-B, et al. Comprehensive annotation of the Chinese tree shrew genome by large-scale RNA sequencing and long-read isoform sequencing. *Zool Res*. 2021;42:692–709.
37. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*. 2011;27:1571–2.
38. Huang X, Zhang S, Li K, Thimmapuram J, Xie S, Wren J. ViewBS: a powerful toolkit for visualization of high-throughput bisulfite sequencing data. *Bioinformatics*. 2018;34:708–9.
39. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*. 2019;37:907–15.
40. Peratea M, Peratea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*. 2015;33:290–5.
41. Schultz MD, Schmitz RJ, Ecker JR. 'Leveling' the playing field for analyses of single-base resolution DNA methylomes. *Trends Genet*. 2012;28:583–5.
42. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550.
43. Wu H, Xu T, Feng H, Chen L, Li B, Yao B, et al. Detection of differentially methylated regions from whole-genome bisulfite sequencing data without replicates. *Nucleic Acids Res*. 2015;43:e141.
44. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009;10:421.
45. Islam M, Strawn M, Behura SK. Fetal origin of sex-bias brain aging. *FASEB J*. 2022;36:e22463.
46. Islam M, Strawn M, Behura SK. Fetal origin of sex-bias brain aging. <http://identifiers.org/geo/GSE157553> (2022).
47. Zeng J, Konopka G, Hunt BG, Preuss TM, Geschwind D, Yi SV. Divergent whole-genome methylation maps of human and chimpanzee brains reveal epigenetic basis of human regulatory evolution. *Am J Hum Genet*. 2012;91:455–65.
48. Zeng J, Yi SV. Evolutionary significance of DNA methylation in human and chimpanzee brains. <http://identifiers.org/geo/GSE37202> (2012).
49. Rhie A, Nurk S, Cechova M, Hoyt SJ, Taylor DJ, Altemose N, et al. The complete sequence of a human Y chromosome. *Nature*. 2023;621:344–54.
50. Johnson RN, O'Meally D, Chen Z, Etherington GJ, Ho SYW, Nash WJ, et al. Adaptation and conservation insights from the koala genome. *Nat Genet*. 2018;50:1102–11.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.