Divergent DNA methylation signatures underlying X chromosome regulation in marsupials and eutherians

Devika Singh[1], Dan Sun[1], Andrew G. King[2], David E. Alquezar-Planas[2], Rebecca N. Johnson[2,3], David Alvarez-Ponce[4],*, and Soojin V. Yi[1],*

[1]School of Biological Sciences, Georgia Institute of Technology, Atlanta, Georgia, USA;

[2]Australian Museum Research Institute, Australian Museum, Sydney, New South Wales, Australia;

[3]National Museum of Natural History, Smithsonian Institution, Washington, DC, USA;

[4]Department of Biology, University of Nevada Reno, Reno, Nevada, USA.

*Corresponding Authors:

David Alvarez-Ponce

Department of Biology

University of Nevada, Reno, 1664 N. Virginia Street, Reno, NV 89557

dap@unr.edu

775-682-5735

Soojin V. Yi

School of Biological Sciences

Georgia Institute of Technology, 950 Atlantic Drive, Atlanta, GA 30332.

soojin.yi@biology.gatech.edu

404-385-6084

Abstract

The phenomenon of X chromosome inactivation (XCI) mediated by sex-specific differences in DNA methylation is well characterized in eutherian mammals. Although XCI is shared between eutherians and marsupials, the role of marsupial DNA methylation in this process remains contested. Here we examine genome-wide signatures of DNA methylation from methylation maps across fives tissues from a male and female koala (*Phascolarctos cinereus*) and present the first whole genome, multi-tissue marsupial "methylome atlas." Using these novel data, we elucidate divergent versus common features of marsupial and eutherian DNA methylation and XCI. First, tissue-specific differential DNA methylation in marsupials primarily occurs in gene bodies. Second, the marsupial X chromosome is significantly globally less methylated (hypomethylated) in females compared to males. We show that this pattern is also observed in eutherian X chromosomes. Third, on average, promoter DNA methylation shows little difference between male and female koala X chromosomes, a pattern distinct from that of eutherians. Fourth, the sex-specific DNA methylation landscape upstream of *Rsx*, the primary *lnc*RNA associated with marsupial X chromosome inactivation, is consistent with the epigenetic regulation of female- (and presumably inactive X chromosome-) specific expression. Finally, we utilize the prominent female X chromosome hypomethylation and classify 98 previously unplaced scaffolds as X-linked, contributing an additional 14.6 Mb (21.5 %) to genomic data annotated as the koala X chromosome. Our work demonstrates evolutionarily divergent pathways leading to the functionally conserved pattern of X chromosome inactivation in two deep branches of mammals.

Keywords: X chromosome inactivation, marsupial, DNA methylation, dosage compensation, RSX, koala, whole genome bisulfite sequencing (WGBS)

Introduction

X chromosome inactivation (XCI) is an iconic example of sex chromosome regulation in which one of the two X chromosomes in females is silenced, a mechanism thought to adjust the expression levels of X-linked genes (Lyon 1961). Although XCI has been observed in both eutherian and marsupial mammals (Shevchenko, et al. 2013), there are several notable differences in XCI between the two lineages. First, in eutherians, the transcription of *Xist* RNA exclusively from the inactive X chromosome is essential for the X chromosome inactivation process (Brown, et al. 1992; Heard, et al. 1997; Plath, et al. 2002). However, the *Xist* locus is not present in marsupials (Duret, et al. 2006; Ng, et al. 2007). Instead, another *lnc*RNA gene, *Rsx*, is associated with the marsupial X chromosome inactivation (Grant, et al. 2012). Second, marsupials demonstrate imprinted XCI by selectively silencing the paternal X chromosome (Sharman 1971; Wang, et al. 2014). In contrast, XCI in eutherians occurs randomly between the maternally and paternally derived X chromosomes, although paternal XCI has been observed during early development in rodents (Huynh and Lee 2003; Okamoto, et al. 2004). Third, in eutherians, XCI involves the exclusion of active histone marks and the recruitment of repressive histone marks on the inactive X chromosome (Heard 2005). In comparison, the inactive marsupial X chromosome, while depleted of the active histone marks, shows variable enrichment patterns of repressive histone marks (Koina, et al. 2009; Wang, et al. 2014). Specifically, out of five repressive marks examined in two marsupial studies, H3K9me3, H4K20me3, and HP1α were enriched (Koina, et al. 2009) while H3K27me3 and H3K9me2 (Rens, et al. 2010) were not enriched on the inactive X chromosome. By contrast, all these marks are enriched on the inactive X in eutherians. These differences indicate complex evolutionary pathways leading to the parallel patterns of XCI in the two mammalian lineages and that novel insights into the mechanism of XCI can be gained from comparative studies.

Given that the patterns of repressive histone marks are diverged between eutherian and marsupial X chromosome inactivation, it is of interest to examine DNA methylation marks. In eutherians, active and inactive X chromosomes clearly exhibit differential DNA methylation (Hellman and Chess 2007; Keown, et al. 2017; Sun, et al. 2019). Specifically, the inactive X chromosome shows increased DNA methylation of promoters and decreased DNA methylation across gene bodies and intergenic regions compared to the active X chromosome. Because promoters account for a relatively small portion of all genomic regions, collectively these DNA

65 methylation patterns result in a global reduction of methylation (hypomethylation) of the inactive
66 X chromosome in eutherians (Sun, et al. 2019). It remains unclear whether marsupial X
67 chromosomes undergo similar differential DNA methylation. Interestingly, marsupial genomes
68 harbor an additional copy of *DNMT1*, the gene encoding the DNA methyltransferase responsible
69 for DNA methylation maintenance (Leonhardt, et al. 1992; Goll and Bestor 2005), through gene
70 duplication (Alvarez-Ponce, et al. 2018), which could lead to functional divergence between the
71 mammalian lineages. Several previous studies found little difference in DNA methylation between
72 active and inactive marsupial X chromosomes (Piper, et al. 1993; Loebel and Johnston 1996;
73 Wang, et al. 2014). These studies generally focused on promoter CpG islands. DNA methylation
74 differences between the active and inactive X chromosomes, however, is a chromosome-wide
75 phenomenon where the patterns differ depending on the nature of genomic regions (e.g. promoter,
76 gene body, or intergenic regions) (Hellman and Chess 2007; Keown, et al. 2017; Sun, et al. 2019).
77 Therefore, a comprehensive and unbiased investigation is necessary to understand the role of DNA
78 methylation on X chromosome inactivation in marsupials.

79 The modern koala (*Phascolarctos cinereus*) is an Australian marsupial well-known for its
80 unique biology and life history. This marsupial is the sole surviving member of the family
81 Phascolarctidae following the loss of the Late Pleistocene giant koala (*P. stirtoni*) (Price 2008;
82 Black, et al. 2014). The recently completed koala genome assembly (Johnson, et al. 2018) offers
83 an opportunity to investigate the relationship between DNA methylation epigenetic modifications
84 and regulation of gene expression on the sex chromosome in a marsupial in detail. A previous
85 study characterized DNA methylation across the opossum X chromosome using reduced
86 representation bisulfite sequencing (Waters, et al. 2018), a method that enriches for
87 hypomethylated regions and generally over-represents promoters and CpG islands (Sun, et al.
88 2015). To extend the survey of DNA methylation to another marsupial species across all genomic
89 regions independent of their CpG contents, here we generated unbiased, nucleotide-resolution
90 genomic DNA methylation maps of diverse tissues from a male and a female koala using whole
91 genome bisulfite sequencing.

92 Our data provide the first multi-tissue, whole genome methylome resource of any
93 marsupial with information on tissue-specific variation of DNA methylation. Utilizing these data,
94 here we show previously unknown patterns of chromosome-wide DNA methylation differences
95 between male and female X chromosomes, indicating distinctive impacts of DNA methylation on

X chromosome inactivation in marsupials. We further classify previously undetected X-linked regions from a key marsupial species using characteristic features of X chromosome DNA methylation. Our findings provide new insights into the evolutionary pathways leading to functionally convergent yet mechanistically divergent pathways of X chromosome inactivation and regulation of gene expression in eutherian and marsupial mammals.

Results

Genome-wide patterns of DNA methylation in the modern koala

To analyze global DNA methylation (5mC) patterns, we generated whole genome bisulfite sequencing (WGBS) data across five tissues (brain, lung, kidney, skeletal muscle, and pancreas) from a male ("Ben," Australian Museum registration M.45022) and female koala ("Pacific Chocolate," Australian Museum registration M.47723). Of the total 16,761,785 CpG sites in the currently assembled koala scaffolds, 88-94% of CpGs had at least 3× coverage (Supplementary Table 1). The mean depth of coverage fell between 9.9× and 14.6×. The total GC content across all 1907 scaffolds covering 3.2 Gb was 39.05%. Autosome-linked scaffolds displayed similar GC content ranging between 38.58 and 39.59%. Scaffolds linked to chromosome X had higher GC contents compared to autosomal scaffolds (Supplementary Table 2), as previously observed (Kasai, et al. 2018). Fractional methylation values of all CpGs were used for a hierarchical clustering analysis between tissues (Fig. 1A). We observed a clear clustering of samples by tissue with the pancreas samples exhibiting the most unique methylation signature while the kidney and lung samples share the most similar methylation profiles.

Differential DNA methylation between tissues

We identified shared and tissue specific differentially methylated regions (DMRs) using BSmooth (Hansen, et al. 2012). Tissue specific DMRs were defined as regions that were differentially methylated in a particular tissue compared to all other tissues in a pairwise analysis while shared DMRs were those found in more than one tissue. Fig. 1B shows the total number of shared and tissue specific DMRs. Consistent with the results of the clustering analysis, the pancreas samples contained the greatest number of tissue specific DMRs (N=1,135) followed by the skeletal muscle, brain, kidney and lung (Fig. 1B). Interestingly, across all tissues we found that the majority (50-53%) of tissue specific DMRs fell in gene bodies (Fig. 1C, Supplementary Fig. 1, and

Supplementary Table 3), a significant excess compared to length and GC matched control regions (fold enrichment (FE) = 1.25~1.44, $p < 0.0001$ based on 10,000 bootstraps; Fig. 1C, Supplementary Fig. 1 and Supplementary Table 3). Even though we observed numerous DMRs in intergenic regions, they were significantly depleted in intergenic regions compared to the control regions ($p < 0.05$ based on 10,000 bootstraps; Fig. 1C, Supplementary Fig. 1, and Supplementary Table 3).

Genes containing tissue specific DMRs were enriched in specific biological functions (Table 1). Brain specific DMRs were linked to genes associated with neural developmental processes such as neurogenesis and central nervous system development. DMRs in the skeletal muscle samples were found in genes associated with actin and cytoskeleton organization. Lung and kidney DMRs were merged due to the fact that their methylation profiles were highly similar and consequently both tissues had very few unique DMRs (n=22 and n=119 respectively) and were linked to genes associated with several terms involving embryonic processes including organ development and morphogenesis. Pancreas specific DMRs overlapped the largest number of genes, which showed enrichment for several metabolic processes.

To explore the consequences of tissue specific DMRs on gene expression levels, we integrated our methylome data with a previously generated RNA-seq koala transcriptome reference (Hobbs, et al. 2014). Of the 12 tissues surveyed in that study, three tissues (kidney, brain, and lung) overlapped with the current methylome dataset. From all methylation and expression linked samples, we found significant negative correlations between promoter DNA methylation and gene expression across the entire genome, similar to the patterns observed in eutherian mammals (Table 2, Supplementary Fig. 2). The relationship between gene body DNA methylation and expression was complex. Specifically, both extremely hypomethylated and hypermethylated gene bodies showed high gene expression, again consistent with studies in eutherian mammals and other taxa (Table 2, Supplementary Fig. 2, and also (Lister, et al. 2009; Zemach, et al. 2010; Jjingo, et al. 2012; Spainhour, et al. 2019)). To directly compare the DNA methylation and expression datasets, we considered differentially methylated genes (DMGs) containing DMRs between the brain and kidney samples (n = 1944 genes from n = 4,615 DMRs). We then identified differentially expressed genes (DEGs) between brain and kidney RNA-seq samples. Currently, available RNA-seq data from koala do not include sufficient biological replicates. We overcame this limitation by simulating replicates within each RNA-seq data set (NOISeq, (Tarazona, et al. 2015)) and

identified 600 putative DEGs (probability of differential expression > 95% according to the NOISeq software). We found that DMGs were significantly more likely to be differentially expressed than genes that did not contain DMRs, exhibiting a 1.54-fold enrichment of observed to expected DMGs that were also DEGs, ($\chi^2 = 33.07$, $p < 0.0001$). In addition, differential expression between tissues was significantly correlated with differential DNA methylation between tissues in promoters (Supplementary Fig. 3A). In agreement with the observation from the whole genome, both relative hypomethylation and relative hypermethylation of gene bodies were associated with increased expression (Supplementary Fig. 3B). The methylation patterns of two representative genes containing brain specific DMRs are shown in Fig. 1D and 1E.

Global hypomethylation of female X chromosome in koalas

The koala genome project used cross-species chromosome painting data to identify 24 putative X chromosome scaffolds and 406 putative autosomal scaffolds consisting of 68 Mb and 2.9 Gb of sequences respectively (Johnson, et al. 2018). We utilized this classification to examine sex specific DNA methylation on the X chromosome and autosomes. As expected from 2:1 ratio of X chromosomes in females compared to males, the median depth of coverage of CpGs on the putative X scaffolds were significantly higher (approximately 2-fold higher) in female samples compared to male samples for all five tissues ($p < 2.2 \times 10^{-16}$, Mann-Whitney *U* test, Supplementary Fig. 4A). In addition, the percent of total reads mapped to the putative X scaffolds out of all the mapped reads showed a distinct bimodal distribution whereby the male samples cluster close to 1.3% and the female samples cluster near 2.4% (Supplementary Fig. 4B). In contrast, male and female samples were indistinguishable with respect to read mapping to putative autosomes (Supplementary Fig. 4D).

Next, we characterized sex-based DNA methylation distributions across the autosomes and X chromosomes. We found that the global DNA methylation level of the female X chromosome was lower than that of the male X chromosome in all koala tissues examined (Fig. 2A, B and Supplementary Fig. 5, $p < 2.2 \times 10^{-16}$, Mann-Whitney *U* test). In contrast, autosomes were similarly methylated between females and males (Fig. 2A, B). A comparison to autosomal DNA methylation indicated that the X chromosome exhibits reduced DNA methylation in females. Consequently, we use the term 'female hypomethylation' (as opposed to male hypermethylation) (Fig. 2C) consistently in this work.

To gain better insights into divergent patterns between male and female X chromosomes, we examined DNA methylation in different functional regions (promoters, exons, introns, and intergenic regions). We found that, on average, the female samples were hypomethylated in all functional regions across the X chromosome (Fig. 2D). This difference was most pronounced in gene bodies and intergenic regions, while promoters showed the least sex-based difference. As expected, the autosomal scaffolds did not display a significant variation between female and male methylation levels in any functional region (Fig. 2E). Fig. 2F depicts the DNA methylation difference between male and female X chromosomes in humans. In contrast to the pattern observed in koalas, promoters on the human female X chromosomes were hypermethylated compared to those on the male X chromosome, congruent with previous studies (Hellman and Chess 2007; Keown, et al. 2017; Sun, et al. 2019).

Promoter DNA methylation is not a universal driver of sex-specific expression in koalas

To investigate the implications of the observed sex-specific DNA methylation, we once again utilized the published RNA-seq koala transcriptome reference (Hobbs, et al. 2014). Of the total RNA-seq dataset, only one tissue, kidney, had expression data from both a male and female koala that could be directly compared to our methylome dataset. Consequently, the other tissues were not considered for further analysis. We calculated the log-transformed fold change of female and male gene expression values using NOISeq, capitalizing on its ability to simulate technical replicates. This analysis suggested that of the 209 X-linked genes, 36 (17.2%) exhibit female overexpression while 11 (5.3%) show male overexpression (probability of sex-based differential expression > 95% based on NOISeq, Supplementary Fig. 6A). Although, on average, autosomal genes also exhibited slight female-bias of expression (Supplementary Fig. 6B, C), the increase is more substantial in X chromosome linked genes (mean chromosome X female to male $\log_2$ fold change = 0.50, autosome female to male expression $\log_2$ fold change = 0.24). Using the matched WGBS and RNA-seq data from the kidney samples, we calculated the female and male fractional methylation difference in X chromosome linked promoters and gene bodies and correlated them with gene expression difference (N = 209 gene bodies and N = 206 promoters, excluding 3 promoters with CpGs coverage < 3). In promoters, we found no significant relationship between female and male DNA methylation difference and female-biased gene expression (Supplementary Fig. 7A). In fact, the proportions of significantly female-over expressed genes (probability of sex-

based differential expression > 95% based on NOISeq) were similar between female-hypermethylated promoters and female-hypomethylated promoters (Supplementary Fig. 7B). Interestingly, female and male DNA methylation difference in gene bodies showed an overall negative correlation with gene expression (Spearman's rank correlation coefficient, $\rho = -0.14$, $p = 0.04$, Supplementary Fig. 7C). However, a deeper analysis of the relative DNA methylation and expression levels revealed that both extreme hypomethylation and hypermethylation are associated with increased expression (Supplementary Fig. 7C) attesting to the complexity and heterogeneity of the relationship between gene body DNA methylation and gene expression (Zemach, et al. 2010; Jjingo, et al. 2012; Spainhour, et al. 2019).

The *Rsx* region displays a pattern suggesting methylation driven control of X chromosome regulation in koalas

Above we have demonstrated that male and female differential methylation was the most notable in gene bodies and intergenic regions on the X chromosome. Nevertheless, sex-specific methylation of a key regulator of the X inactivation (XCI) in other marsupials, the long non-coding RNA gene *Rsx*, has been associated with its differential expression (Grant, et al. 2012; Wang, et al. 2014). To determine if differential DNA methylation is similarly associated with differential regulation of *Rsx* in koalas, we isolated a 42 Kb region to examine on the koala X chromosome based on the sequence homology with the partial assembly of the *Rsx* gene from the gray short-tailed opossum, *Monodelphis domestica* (Grant, et al. 2012). We then identified a 29.8 Kb candidate *Rsx* sequence within the isolated region from the alignment to the full koala *Rsx* sequence assembled using PacBio long read sequencing by Johnson et al. (Johnson, et al. 2018). The total examined region spanned 266 CpGs, of which 165 were covered by more than three reads in all 10 samples from the WGBS dataset. The candidate *Rsx* region harbored 93 CpGs over this read depth threshold. Because the WGBS data allows us to explore the entire genomic region around the candidate gene, we used these CpGs to examine the DNA methylation difference between male and female koala samples. Interestingly, we observed a female hypomethylated region containing two CpG islands upstream of the candidate *Rsx* covering 101 CpGs exhibiting a 36% reduction in mean female DNA methylation compared to mean male DNA methylation (mean sex difference: -0.36 ± 0.14, Fig. 3). The mean sex-based DNA methylation difference within this

region was in the top 23% of the distribution of differences across all X-linked scaffolds of the combined tissues.

Having identified an upstream female hypomethylated hotspot, we tested whether *Rsx* expression was exclusive to females, consistent with it being the X chromosome inactivation center in koalas (Hobbs, et al. 2014; Johnson, et al. 2018). We used DESeq2 (Love, et al. 2014) to determine differential gene expression on curated RNA-seq data from 8 female and 7 male koala samples obtained from 12 tissues (Hobbs, et al. 2014). As expected, the expression of the *lnc*RNA annotated within the candidate *Rsx* gene region was significantly greater in females (mean normalized read count = 6987.1) than in males (mean normalized read count = 16, $p < 0.05$ from DeSeq2 using the Wald test). When considering a subset of tissue expression data with matched male and female samples (spleen, kidney, and lung), the results remained robust across different tools to measure differential gene expression (Table 3).

Identification of novel candidate X-linked scaffolds by sex specific methylation patterns

We have demonstrated above that X-linked scaffolds exhibit several distinctive WGBS patterns in koalas. Namely, putative X scaffolds show different depths of coverage per CpG, distinctive clustering based on the proportion of mapped reads, and unique methylation distributions between females and males (Supplementary Fig. 4). We thus quantified DNA methylation differences between females and males to determine if additional candidate X scaffolds existed within the 6.7% of the koala assembly that remained unclassified. We identified an additional 98 scaffolds which showed a clear shift towards female hypomethylation (mean female-male 5mC for all candidate X scaffolds was -0.25 ± 0.12) (Supplementary Table 4). These candidate scaffolds contributed 14.6 Mb (21.5%) to the total genomic region annotated as the koala X chromosome. All candidate scaffolds followed the expected bimodal distribution seen in the putative X scaffolds when we analyzed the percent of total reads mapped to the candidate X scaffolds out of all the mapped reads. The male samples clustered around 0.3% and the female samples clustered around 0.5% (Supplementary Fig. 4C). This clustering pattern could be attributed to the 2:1 ratio of X chromosomes in females compared to males that was not observed in putative autosomes.

Discussion

In this study, we investigated nucleotide-resolution genomic DNA methylation maps of five tissues of a male and a female koala. The overall DNA methylation levels of koala tissues are on par with the patterns observed in other eutherian mammals (Schultz, et al. 2015; Mendizabal, et al. 2016; Keown, et al. 2017), and show a clear pattern of heavy genome-wide DNA methylation punctuated by hypomethylation of CpG islands (e.g., Fig. 1D,E). We identified tissue-specific differential DNA methylated regions (DMRs), which were consistently and significantly enriched in gene bodies. Previous studies of human tissues also observed that DMRs tended to occur in gene bodies (e.g. (Mendizabal and Yi 2015; Peters, et al. 2015; Schultz, et al. 2015)). Gene body methylation is an ancestral form of DNA methylation in animal genomes (e.g., (Zemach, et al. 2010; Yi 2012)), yet its role in regulation of gene expression is less well understood than promoter DNA methylation (Zemach, et al. 2010; Jones 2012). Integrating relative DNA methylation differences and relative gene expression differences, we observe that tissue-specific hypo- and hyper methylation of gene bodies both contribute to increased gene expression (Supplementary Fig. 3B) in koalas, a characteristic previously observed in eutherian mammals and in other taxa (Lister, et al. 2009; Zemach, et al. 2010; Jjingo, et al. 2012; Mendizabal, et al. 2016). Collectively, these observations suggest that gene body methylation may be an important component of regulation of gene expression in koala.

The patterns of DNA methylation across the marsupial X chromosome show significant differences between sexes. Specifically, the female X chromosomes are globally hypomethylated compared to both the male X chromosome and the autosomes of both sexes (Fig. 2). Most of the literature of X chromosome inactivation emphasizes that DNA methylation is increased in promoters of the inactive X chromosomes compared to those of the active X chromosome. However, we show that the hypomethylation of the female X chromosome observed in koala is in fact consistent with DNA methylation data from eutherian mammals. We summarized the current knowledge from eutherian mammals and our proposed model for marsupials in Fig. 4. Concisely, in eutherians, while DNA methylation of promoter regions and CpG islands increase on the inactive X chromosome, gene bodies and intergenic regions of the inactive X chromosome are relatively depleted of DNA methylation compared to the active X chromosome (Fig. 4). Hellmann and Chess (Hellman and Chess 2007) demonstrated the difference between promoter and gene body DNA methylation of active and inactive X chromosomes in humans. An array based DNA

312 methylation study of humans (Cotton, et al. 2015) and whole genome bisulfite sequencing analyses
313 of mouse (Keown, et al. 2017) and humans (Sun, et al. 2019) showed that the hypomethylation of
314 inactive X chromosome is pervasive in gene bodies and intergenic regions.

315 In marsupials, Rens et al. (Rens, et al. 2010) observed hypomethylation of the inactive X
316 chromosome through immuno-staining using antibodies against methyl-cytosines in two species,
317 common brushtail possum (*Trichosurus vulpecula*) and long-nosed potoroo (*Potorous tridactylus*).
318 Waters et al. (Waters, et al. 2018) have also observed female hypomethylation of the X
319 chromosome in gray-tailed opossum (*Monodelphis domestica*) using reduced representation
320 bisulfite sequencing (RRBS). In contrast to Keown et al. (Keown, et al. 2017), Waters et al.
321 (Waters, et al. 2018) did not observe substantial hypomethylation of mouse X chromosomes. They
322 thus concluded that the reduction of methylation in gene bodies was specific to marsupials (Waters,
323 et al. 2018). The discrepancy between Keown et al. (Keown, et al. 2017) and Waters et al. (Waters,
324 et al. 2018) could be due to the inherent bias of the RRBS method, used in the latter study, which
325 disproportionately samples regions with high GC content (Sun, et al. 2015). High GC-content
326 regions tend to be hypomethylated (Elango, et al. 2008; Cotton, et al. 2015) and show less variation
327 of DNA methylation between males and females (Cotton, et al. 2015), which we also demonstrate
328 below (Supplementary Fig. 7).

329 It was previously suggested that differential DNA methylation does not play substantial
330 roles in marsupial X chromosome gene regulation (Piper, et al. 1993; Loebel and Johnston 1996;
331 Wang, et al. 2014) due to the lack of DNA methylation difference in promoters and CpG islands
332 (also in (Waters, et al. 2018)). Here we show that promoters are, on average, clearly
333 hypomethylated in female tissues compared to male tissues in koalas, although the degree of
334 methylation difference is small compared to other genomic regions. This modest sex-based
335 difference may at least partly be attributed to the fact that promoters tend to be GC-rich
336 (Supplementary Fig. 7). Due to the negative correlation between GC-content and DNA
337 methylation, these lowly methylated regions are inherently less variable and would subsequently
338 show subtle differences between sexes ((Cotton, et al. 2015) and Supplementary Fig. 7).
339 Summarizing these findings, we observe that the hypermethylation of promoters on the inactive X
340 chromosome appears to be restricted to eutherian mammals, while chromosome-wide
341 hypomethylation of female X chromosome (Fig. 4), consistent with the hypomethylation of the
342 inactive X chromosome, is conserved between eutherian and marsupial mammals. How

chromosome-wide DNA hypomethylation is linked to chromosome-wide gene silencing is currently unknown, and likely to involve other epigenetic marks. Analyses of DMNT expression, including that of a newly duplicated *DNMT1* (Alvarez-Ponce, et al. 2018), did not indicate significant differential expression between sexes (Results not shown). However, additional analysis of data with broad and balanced sampling of both sexes is necessary to rigorously test the impact of the marsupial *DNMT1* duplication on sex specific gene expression.

Interestingly, despite the fact that gene regulation by sex-specific promoter DNA methylation is not a global phenomenon in marsupials, the *lnc*RNA *Rsx,* the major player in XCI initiation in marsupials (Grant, et al. 2012), presents an exception. Similar to the mechanism of activation of the *Xist* RNA in eutherian mammals, *Rsx* is exclusively transcribed from and coats the inactive X chromosome initiating the XCI process through which the inactive X chromosome adopts a specific chromatin configuration within cells (Galupa and Heard 2018). Observed signatures unique to female koala samples suggest that *Rsx* expression is regulated by promoter DNA methylation, while gene body methylation of *Rsx* is similar between males and females (Fig. 3). Wang et al. (Wang, et al. 2014) showed that the *Rsx* promoter in opossum is also regulated by differential DNA methylation of CpG islands. Therefore, regulation of the key initiator of XCI via DNA methylation is another parallel feature between eutherians and marsupials, although utilizing distinctive components. In addition, a recent study (Sprague, et al. 2019) found that *Xist* and *Rsx* harbor non-linear sequence similarity. Consequently, their shared functionality may be partially due to characteristics of tandem repeat regions.

Therefore, we show that the overarching hypomethylation of female X chromosome is a conserved feature of X chromosome regulation in eutherian and marsupial mammals. However, X chromosome promoter methylation and the subsequent effect on the regulation of gene expression between males and females appear to be divergent between these two lineages. Regulation of the *Rsx*, on the other hand, is supported by DNA methylation, which mirrors the regulation of the eutherian *Xist* locus. Together, these conclusions illuminate the intricate evolutionary pathways that have diverged and converged to influence X chromosome regulation, XCI, and dosage compensation in eutherian and marsupial mammals.

Materials and Methods

Whole Genome Bisulfite Sequencing Data Generation

All genomic DNA was extracted using a Bioline Isolate II Genomic DNA Extraction Kit (Cat#. BIO-52067) following the recommended protocol with an additional DNAse free RNaseA (100mg/ml) (Qiagen cat. #19101) treatment before column purification. 20mg tissue samples from brain, kidney, lung, skeletal muscle, and pancreas from a female koala, "Pacific Chocolate" (Australian Museum registration M.45022), and a male koala, "Ben" (Australian Museum registration M.47723), were supplied to the Ramaciotti Centre for Genomics for methylome sequencing. The bisulfite conversion was carried out by using the EX DNA Methylation-Lightning Kit (Zymo cat. #D5030) and the WGBS libraries were constructed using the TruSeq DNA methylation kit (Illumina cat.# EGMK81213). The libraries were sequenced on a NovaSeq6000 S2 (Illumina) using the 2 × 100bp PE option. Information on samples, coverage, and the numbers of reads are provided in Supplementary Table 1.

WGBS Data Processing

Paired-end raw fastq files were processed for quality control and adapter trimming using Trim Galore! (version 0.3.7) with a Phred score threshold of 20. Trimmed reads were filtered for phi-X174 (NC_001422.1) to remove the spike-in control before being mapped to the koala reference assembly (phaCin_unsw_v4.1; (Johnson, et al. 2018)) using Bismark with the Bowtie 2 setting and deduplicated. The current koala genome reference assembly was generated from the same female koala ("Pacific Chocolate," Australian Museum registration M.45022) that we used to obtain the female WGBS data. Bisulfite conversion rates were estimated for each WGBS sample following the methodology of methPipe's bsrate (Song, et al. 2013) (Supplementary Table 1). Strand-specific methylation calls were combined, and all samples were filtered to remove CpGs covered by fewer than three reads. See Supplementary Table 1 for an overview of the dataset.

Analyses of tissue differentially methylated regions

A hierarchical clustering tree was drawn using the fractional methylation profiles of all ten samples representing five tissues using *hclust* from R's stats package. The distance matrix was calculated using Euclidean distances and the agglomeration method used was Ward's method. The data for the final tree was visualized using R's dendextend package (Galili 2015).

Before identifying differentially methylated regions (DMRs), all Bismark generated CpG reports were filtered to remove scaffolds that were less than 2 Mb in length. Following this step, we retained 3.03×10^9 positions for downstream analyses, covering 94.8% of the genome. DMRs were called using BSmooth (Hansen, et al. 2012) in 10 pairwise comparisons of all five tissues. Detected DMRs with a minimum fractional methylation difference of 0.3 (30%) between each pair of compared tissues that also contained at least 5 CpG sites were retained. Shared DMRs were defined as those that overlapped at least 50% in multiple tissues. The distribution of common and tissue specific DMRs located among different genomic functional regions (promoters, gene bodies, and intergenic regions) were identified based on koala gene annotations from Ensembl (Phascolarctos_cinereus.phaCin_unsw_v4.1.97 release). Promoters were defined as regions located 1000 bp upstream of the identified transcription start site (TSS). To test the enrichment of tissue specific DMRs in the previously specified genomic regions, we generated 10,000 length and GC content matched control regions for all unique DMRs, and calculated the p-values from the ratio of the number of simulations with values at least as extreme as the observed ones over the total number of simulations.

The genes containing tissue specific DMRs were extracted to make a set of differentially methylated genes unique to each analyzed tissue. As mentioned above, the lung and kidney samples shared the most similar methylation profiles and consequently had few tissue specific DMRs (Fig. 1C). Because of this feature, the corresponding gene sets were combined for these two tissues. Functional annotation and GO term enrichment analysis was performed utilizing the ToppGene Suite (Chen, et al. 2009).

Quantifying sex differences in DNA methylation across chromosomes and functional regions

Johnson et al. (Johnson, et al. 2018) used cross-species chromosome painting data and linked 406 scaffolds spanning 2.9 Gb of sequence data to autosomal scaffolds from chromosomes 1-7 and 24 scaffolds covering 68 Mb of sequence data to chromosome X, leaving 6.7% of the genome unclassified. To explore the same relative amount of genomic space, we randomly sampled a subset of the autosomal scaffolds that were length matched with the X chromosome scaffolds, which we called the "matched autosome" dataset. Separately, all matched autosome and chromosome X scaffolds were divided into 10-kb bins and the mean fractional methylation levels

of the corresponding CpGs were calculated for each of the 5 tissue samples. The difference in mean fractional methylation at each 10-kb bin was computed between male and female samples for all tissues. For our comparative analysis with human sex-specific DNA methylation, we downloaded publicly available WGBS fractional methylation reports from a male brain (Epigenome ID: E071) and a female brain (Epigenome ID: E053) generated by the Roadmap Epigenomics Consortium (Kundaje, et al. 2015) and used CpGs with minimum 3× coverage. Due to its similarity in size to the human X chromosome, we used data from human chromosome 8 as our representative autosome in the comparative analysis.

For the analysis of sex-based fractional DNA methylation across different genomic functional regions, we used the previously defined koala gene annotations from Ensembl (Phascolarctos_cinereus.phaCin_unsw_v4.1.97 release) and human known gene annotations from Ensembl (hg19 release). Individual annotated promoters, exons, introns, and intergenic regions were evenly divided into 20 bins and the mean fractional methylation of CpGs within each bin was computed.

Identification of candidate X-linked scaffolds

To isolate candidate scaffolds potentially located on the X chromosome from the 1,477 unclassified scaffolds covering 6.7% of the koala genome, we binned the remnant scaffolds into 10-kb windows and calculated the mean fractional methylation of the associated CpGs. We then determined the average female and male methylation differences across the bins and plotted the density of the differences for all five tissues. SVY and DS proceeded to independently select scaffolds that exhibited a statistically significant shift towards female hypomethylation from zero based on an initial visual examination. The scaffolds that showed significant female hypomethylation in all five tissues and were selected by both SVY and DS were then further analyzed (n = 98 covering 14.6 Mb of sequence with mean female-male 5mC = -0.25 ± 0.12). As a baseline for an external validation, the percent of reads mapping to the putative X-linked and autosome-linked scaffolds over the total number of mapped reads was computed for the male and female sample in all tissues. The expectation was that the female samples would show and increase in reads mapped to the X-linked scaffolds compared to the male samples due to the 2:1 ratio of X chromosomes in females to males. This sex-biased separation is not expected across length matched autosome-linked scaffolds as autosome counts are balanced in both sexes. Once this

baseline analysis was confirmed, we calculated the percent of reads mapping to the candidate X-linked scaffolds over the total number of mapped reads from all 10 samples.

Annotation of the *Rsx* Region

To annotate the genomic region around *Rsx*, we downloaded the published genome sequence fasta files for the partial assembly of *Rsx* from the gray short-tailed opossum (Grant, et al. 2012) and the complete PacBio assembly of the koala *Rsx* (Johnson, et al. 2018; Sprague, et al. 2019). We then used BLASTN 2.2.29 (Zhang, et al. 2000) to align both sequences to the koala reference genome (phaCin_unsw_v4.1) and obtained genomic coordinates information. As expected, both sequences aligned to a scaffold previously linked to the X chromosome and overlapped with one another. The entire assembled koala *Rsx* sequence aligned with 100% identity and no gaps.

RNA-seq Data Processing

All RNA-seq expression data used in this analysis were obtained from the previously published koala transcriptome reference (Hobbs, et al. 2014). This dataset included eight tissues from a female koala, spleen (SRR1203868), liver (SRR1205138), uterus (SRR1205176), kidney (SRR1205998), lung (SRR1205218), heart (SRR1205223), brain (SRR1205222), adrenal glands (SRR1205224), and seven tissues from a male koala, salivary gland (SRR1207973), kidney (SRR1207974), testes (SRR1207975), bone marrow (SRR1106690), lymph node (SRR1106707), liver (SRR1121764), and spleen (SRR1122141). To process the acquired data, we followed the protocol outlined by Pertea et al. (Pertea, et al. 2016). Briefly, raw reads were filtered for low quality and trimmed of adapter sequences by Trim Galore! (version 0.3.7) before alignment to the reference koala genome (phaCin_unsw_v4.1) using HISAT2 (Kim, et al. 2015). We used the koala GTF annotation from Ensembl (Phascolarctos_cinereus.phaCin_unsw_v4.1.97.gtf.gz release) to assemble mapped reads into transcripts using StringTie 2.0 (Pertea, et al. 2016) with the -e-b--A <gene_abund.tab> flags. We chose to use StringTie's functionality for an additional *de novo* transcript assembly to quantify transcripts not currently annotated in the koala reference, specifically those associated with the *lnc*RNA *Rsx*. First, a new GTF annotation was generated including novel transcripts using the --merge flag and then the previously generated mapped reads were reassembled into transcripts guided by the novel GTF file.

Analysis of sex-based differential gene expression

Gene expression for the previously annotated and novel transcripts was measured for all fifteen samples available. Using the prepDE.py script from Stringtie 2.0 (Pertea, et al. 2016), we generated a raw expression count matrix containing the data from all male and female samples. Subsequently, we inputted the count matrix into DeSeq2 1.22.2 (Love, et al. 2014), which uses the median of ratios method to normalize raw expression data for sequencing depths and RNA composition to perform differential gene expression analysis. Considering all eight female and seven male samples, we first sought to specifically test the male and female differential expression of the annotated *Rsx* region from the novel transcript analysis. Only one 30.4 kb transcript, a novel *lnc*RNA, overlapped with the annotated *Rsx* region (overlap > 90% of transcript) and was used to evaluate gene expression. From DeSeq2's output, we collected data on the normalized male and female expression counts, the log transformed fold change ($\log_2$[female expression/male expression]), and the p-value associated with the sex-based differential expression. We also performed a permutation of this analysis by only considering the tissues that had both a male and female sample, namely the kidney, lung, and spleen.

NOISeq 2.26.1 (Tarazona, et al. 2015) was used for differential expression analysis due to its ability to simulating technical replicates within given RNA-seq data sets when no replicates are available. Using the raw count matrix from annotated transcripts, we generated sex-specific differentially methylated genes for one dataset considering only the male and female kidney data and another dataset considering the three tissues with male and female samples (kidney, lung, and spleen). We compared the gene expression outputs for the three tissue datasets generated by both DeSeq2 and NOISeq and found highly consistent results when considering the female over male expression change values for all genes (Spearman's rank correlation coefficient, $\rho = 0.997$, $p < 2.2 \times 10^{-16}$).

Availability of data and materials

The raw and processed methylation datasets generated in this study have been deposited in the NCBI Gene Expression Omnibus (Edgar, et al. 2002).

Acknowledgements

This work was supported by a grant from the National Science Foundation (MCB 1818288) and a Pilot Grant from the Smooth Muscle Plasticity COBRE of the University of Nevada, Reno (funded by the National Institutes of Health grant 5P30GM110767-04) to DAP, grants by the National Science Foundation (MCB 1615664) and the National Institute of Health (R01MH103517) to SVY. In addition, DS was partially supported by the NIH Training Grant in Computational Biology and Biomedical Genomics (T32 GM105490).

References

Alvarez-Ponce D, Torres-Sanchez M, Feyertag F, Kulkarni A, Nappi T. 2018. Molecular evolution of DNMT1 in vertebrates: Duplications in marsupials followed by positive selection. PLOS ONE 13.

Black KH, Price GJ, Archer M, Hand SJ. 2014. Bearing up well? Understanding the past, present and future of Australia's koalas. Gondwana Research 25:1186-1201.

Brown CJ, Hendrich BD, Rupert JL, Lafreniere RG, Xing Y, Lawrence J, Willard HF. 1992. 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 71:527-542.

Chen J, Bardes EE, Aronow BJ, Jegga AG. 2009. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. Nucleic Acids Res 37:W305-311.

Cotton AM, Price EM, Jones MJ, Balaton BP, Kobor MS, Brown CJ. 2015. Landscape of DNA methylation on the X chromosome reflects CpG density, functional chromatin state and X-chromosome inactivation. Hum Mol Genet 24:1528-1539.

Duret L, Chureau C, Samain S, Weissenbach J, Avner P. 2006. The Xist RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. Science 312:1653-1655.

Edgar R, Domrachev M, Lash AE. 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 30:207-210.

Elango N, Kim S-H, Program NCS, Vigoda E, Yi SV. 2008. Mutations of different molecular origins exhibit contrasting patterns of regional substitution rate variation. PLoS Computational Biology 4:e1000015.

Galili T. 2015. dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. Bioinformatics (Oxford, England) 31:3718-3720.

Galupa R, Heard E. 2018. X-Chromosome Inactivation: A Crossroads Between Chromosome Architecture and Gene Regulation. Annual Review of Genetics 52:535-566.

Goll MG, Bestor TH. 2005. Eukaryotic Cytosine Methyltransferases. Annual Review of Biochemistry 74:481-514.

Grant J, Mahadevaiah SK, Khil P, Sangrithi MN, Royo H, Duckworth J, McCarrey JR, VandeBerg JL, Renfree MB, Taylor W, et al. 2012. Rsx is a metatherian RNA with Xist-like properties in X-chromosome inactivation. Nature 487:254.

Hansen KD, Langmead B, Irizarry RA. 2012. BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. Genome biology 13:R83.

570 Heard E. 2005. Delving into the diversity of facultative heterochromatin: the epigenetics of the
571 inactive X chromosome. Curr Opin Genet Dev 15:482-489.
572 Heard E, Clerc P, Avner P. 1997. X-chromosome inactivation in mammals. Annu Rev Genet
573 31:571-610.
574 Hellman A, Chess A. 2007. Gene body-specific methylation on the active X chromosome. Science
575 315:1141-1143.
576 Hobbs M, Pavasovic A, King AG, Prentis PJ, Eldridge MDB, Chen Z, Colgan DJ, Polkinghorne
577 A, Wilkins MR, Flanagan C, et al. 2014. A transcriptome resource for the koala
578 (Phascolarctos cinereus): insights into koala retrovirus transcription and sequence
579 diversity. BMC Genomics 15:786.
580 Huynh KD, Lee JT. 2003. Inheritance of a pre-inactivated paternal X chromosome in early mouse
581 embryos. Nature 426:857-862.
582 Jjingo D, Conley AB, Yi SV, Lunyak VV, Jordan IK. 2012. On the presence and role of human
583 gene-body DNA methylation. Oncotarget 3:462.
584 Johnson RN, O'Meally D, Chen Z, Etherington GJ, Ho SYW, Nash WJ, Grueber CE, Cheng Y,
585 Whittington CM, Dennison S, et al. 2018. Adaptation and conservation insights from the
586 koala genome. Nature Genetics 50:1102-1111.
587 Jones PA. 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat
588 Rev Genet 13:484-492.
589 Kasai F, O'Brien PCM, Pereira JC, Ferguson-Smith MA. 2018. Marsupial chromosome DNA
590 content and genome size assessed from flow karyotypes: invariable low autosomal GC
591 content. Royal Society open science 5:171539-171539.
592 Keown CL, Berletch JB, Castanon R, Nery JR, Disteche CM, Ecker JR, Mukamel EA. 2017.
593 Allele-specific non-CG DNA methylation marks domains of active chromatin in female
594 mouse brain. Proceedings of the National Academy of Sciences 114:E2882.
595 Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory
596 requirements. Nature methods 12:357-360.
597 Koina E, Chaumeil J, Greaves IK, Tremethick DJ, Graves JA. 2009. Specific patterns of histone
598 marks accompany X chromosome inactivation in a marsupial. Chromosome Res 17:115-
599 126.
600 Krueger F, Andrews SR. 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq
601 applications. Bioinformatics (Oxford, England) 27:1571-1572.
602 Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang
603 Z, Wang J, Ziller MJ, et al. 2015. Integrative analysis of 111 reference human epigenomes.
604 Nature 518:317-330.
605 Leonhardt H, Page AW, Weier H-U, Bestor TH. 1992. A targeting sequence directs DNA
606 methyltransferase to sites of DNA replication in mammalian nuclei. Cell:865-873.
607 Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z,
608 Ngo Q-M. 2009. Human DNA methylomes at base resolution show widespread
609 epigenomic differences. Nature 462:315-322.
610 Loebel D, Johnston PG. 1996. Methylation analysis of a marsupial X-linked CpG island by
611 bisulfite genomic sequencing. Genome research 6:114-123.
612 Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-
613 seq data with DESeq2. Genome biology 15:550.
614 Lyon MF. 1961. Gene Action in the X-chromosome of the Mouse (Mus musculus L.). Nature
615 190:372-373.

Mendizabal I, Shi L, Keller TE, Konopka G, Preuss TM, Hsieh TF, Hu E, Zhang Z, Su B, Yi SV. 2016. Comparative Methylome Analyses Identify Epigenetic Regulatory Loci of Human Brain Evolution. Mol Biol Evol 33:2947-2959.

Mendizabal I, Yi SV. 2015. Whole-genome bisulfite sequencing maps from multiple human tissues reveal novel CpG islands associated with tissue-specific regulation. Human Molecular Genetics 25:69-82.

Ng K, Pullirsch D, Leeb M, Wutz A. 2007. Xist and the order of silencing. EMBO reports 8:34-39.

Ogawa Y, Kakumoto K, Yoshida T, Kuwako K-i, Miyazaki T, Yamaguchi J, Konno A, Hata J, Uchiyama Y, Hirai H, et al. 2018. Elavl3 is essential for the maintenance of Purkinje neuron axons. Scientific Reports 8:2722.

Okamoto I, Otte AP, Allis CD, Reinberg D, Heard E. 2004. Epigenetic dynamics of imprinted X inactivation during early mouse development. Science 303:644-649.

Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. 2016. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nature Protocols 11:1650-1667.

Peters TJ, Buckley MJ, Statham AL, Pidsley R, Samaras K, V Lord R, Clark SJ, Molloy PL. 2015. De novo identification of differentially methylated regions in the human genome. Epigenetics & Chromatin 8:6.

Piper AA, Bennett AM, Noyce L, Swanton MK, Cooper DW. 1993. Isolation of a clone partially encoding hill kangaroo X-linked hypoxanthine phosphoribosyltransferase: Sex differences in methylation in the body of the gene. Somatic Cell and Molecular Genetics 19:141-159.

Plath K, Mlynarczyk-Evans S, Nusinow DA, Panning B. 2002. Xist RNA and the mechanism of X chromosome inactivation. Annu Rev Genet 36:233-278.

Price GJ. 2008. Is the modern koala (Phascolarctos cinereus) a derived dwarf of a Pleistocene giant? Implications for testing megafauna extinction hypotheses. Quaternary Science Reviews 27:2516-2521.

Rens W, Wallduck MS, Lovell FL, Ferguson-Smith MA, Ferguson-Smith AC. 2010. Epigenetic modifications on X chromosomes in marsupial and monotreme mammals and implications for evolution of dosage compensation. Proceedings of the National Academy of Sciences 107:17657.

Schultz MD, He Y, Whitaker JW, Hariharan M, Mukamel EA, Leung D, Rajagopal N, Nery JR, Urich MA, Chen H, et al. 2015. Human body epigenome maps reveal noncanonical DNA methylation variation. Nature 523:212-216.

Sharman GB. 1971. Late DNA Replication in the Paternally Derived X Chromosome of Female Kangaroos. Nature 230:231-232.

Shevchenko AI, Zakharova IS, Zakian SM. 2013. The evolutionary pathway of x chromosome inactivation in mammals. Acta naturae 5:40-53.

Song Q, Decato B, Hong EE, Zhou M, Fang F, Qu J, Garvin T, Kessler M, Zhou J, Smith AD. 2013. A Reference Methylome Database and Analysis Pipeline to Facilitate Integrative and Comparative Epigenomics. PLOS ONE 8:e81148.

Spainhour JCG, Lim HS, Yi SV, Qiu P. 2019. Correlation Patterns Between DNA Methylation and Gene Expression in The Cancer Genome Atlas. Cancer Informatics 18:1176935119828776.

Sprague D, Waters SA, Kirk JM, Wang JR, Samollow PB, Waters PD, Calabrese JM. 2019. Nonlinear sequence similarity between the Xist and Rsx long noncoding RNAs suggests shared functions of tandem repeat domains. RNA 25:1004-1019.

Sun D, Maney DL, Layman TS, Chatterjee P, Yi SV. 2019. Regional epigenetic differentiation of the Z Chromosome between sexes in a female heterogametic system. Genome research 29:1673-1684.

Sun Z, Cunningham J, Slager S, Kocher J-P. 2015. Base resolution methylome profiling: considerations in platform selection, data preprocessing and analysis. Epigenomics 7:813-828.

Tarazona S, Furio-Tari P, Turra D, Pietro AD, Nueda MJ, Ferrer A, Conesa A. 2015. Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. Nucleic Acids Res 43:e140.

Wang X, Douglas KC, Vandeberg JL, Clark AG, Samollow PB. 2014. Chromosome-wide profiling of X-chromosome inactivation and epigenetic states in fetal brain and placenta of the opossum, Monodelphis domestica. Genome research 24:70-83.

Waters SA, Livernois AM, Patel H, O'Meally D, Craig JM, Marshall Graves JA, Suter CM, Waters PD. 2018. Landscape of DNA Methylation on the Marsupial X. Molecular Biology and Evolution 35:431-439.

Yi SV. 2012. Birds do it, bees do it, worms and ciliates do it too: DNA methylation from unexpected corners of the tree of life. Genome biology 13:174.

Zemach A, McDaniel IE, Silva P, Zilberman D. 2010. Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science 328:916-919.

Zhang Z, Schwartz S, Wagner L, Miller W. 2000. A greedy algorithm for aligning DNA sequences. J Comput Biol 7:203-214.

690 **Figure Legends**

691 **Figure 1. Overview of DNA methylation patterns across the koala genome.** (A) Hierarchical
692 clustering of all 10 Whole Genome Bisulfite (WGBS) CpG reports representing five tissues using
693 Euclidean distances and Ward's agglomeration method. (B) Matrix reporting tissue-specific and
694 shared differentially methylation regions (DMRs) based on the pairwise comparison of DMRs
695 between tissues. (C) Considering data from brain, the total count of tissue specific DMRs falling
696 within one of three annotated genomic regions: promoters, gene bodies, and intergenic regions.
697 The enrichment of DMRs in each functional region is shown through a comparison with length
698 and GC matched control regions (***$p < 0.0001$, n.s. Not significant, from 10,000 bootstraps).
699 Error bars depict standard deviation. (D-E) Examples of two brain-specific DMRs (highlighted in
700 grey) with the corresponding CpG fractional methylation values reported for all WGBS samples
701 including the two brain (red) and eight remaining tissues (blue). Line smoothing performed using
702 local regression (LOESS). (D) *ELAVL3* is a neural specific RNA-binding protein highly expressed
703 in adult mammalian brains and linked to the maintenance of Purkinje neuron axons (Ogawa, et al.
704 2018). A 1.84 kb region containing three CpG Islands (CGI) and overlapping the first exon of
705 *ELAVL3* was highly methylated (hypermethylated) in the brain samples compared to all other
706 tissues, and this gene was up-regulated in the brain compared to the kidneys (log-transformed brain
707 to kidney ratio = 15.4, probability of differential expression > 98% from NOISeq). (D) *ALDOC*
708 encodes a catalytic enzyme responsible for the conversion of fructose-1,6-bisphosphate to
709 glyceraldehyde 3-phosphate and dihydroxyacetone phosphate. A 945 bp, brain-specific lowly
710 methylated (hypomethylated) DMR overlaps the promoter and part of *ALDOC*'s gene body, and
711 this gene was up-regulated in koala brain samples compared to the kidney samples (log-
712 transformed brain to kidney ratio = 5.45, probability of differential expression > 96% from
713 NOISeq).
714
715 **Figure 2. Global patterns of female and male DNA methylation (5mC) across eutherian and**
716 **marsupial X chromosomes.** (A) Comparison of chromosome-wide sex-based DNA methylation
717 differences in human brain and koala brain samples. The distribution of differences in female and
718 male mean CpG fractional methylation is shown for autosomes and X chromosomes in human and
719 koala data (B) The combined distribution of female and male differences of mean CpG fractional
720 methylation from koala brain, lung, kidney, skeletal muscle and pancreas samples. For A-B, the

721 difference of female and male mean fractional methylation was calculated for 10 kb bins across
722 each autosome- or X-linked scaffold (*** indicated p < 2.2x10-16, Mann-Whitney U test). (C)
723 Mean female and male fractional DNA methylation across autosomes and X chromosomes from
724 human brain and koala brain samples. For D-F, mean fractional methylation levels across X-linked
725 or autosome-linked gene regions in male and female whole genome bisulfite (WGBS) samples
726 across genomic functional regions (promoters, exons, introns, and intergenic regions). Functional
727 regions were divided into 20 equally spaced bins and the mean fractional methylation of the
728 corresponding CpGs were calculated for all male (purple) and female (orange) samples. Line
729 smoothing was performed using local regression (LOESS). (D) Average fractional methylation of
730 CpGs in 100-bp sliding windows using a 10 bp step size in a 5 Kb region upstream and downstream
731 of all chromosome X linked gene's transcription start sites (TSSs) across all five koala tissues.
732
733 **Figure 3. Annotation of genomic of DNA methylation (5mC) around *Rsx*.** The genomic
734 coordinates are shown above the figure in grey isolating a 57 kb window containing the annotated
735 *Rsx* region (blue gene model below). The top panel identifies CpG islands (CGI), the middle panel
736 reports the absolute male (purple) and female (orange) fractional methylation at each CpG, and the
737 bottom panel show the female and male fractional methylation difference. Highlighted in grey
738 across all panels is the identifies female hypomethylated hotspot containing 101 CpGs and two
739 CGIs upstream of the Rsx transcription start site.
740
741 **Figure 4. Model of DNA methylation (5mC) patterns across genomic functional regions for**
742 **eutherian mammals and marsupials.** In summary, eutherian mammals exhibit increased DNA
743 methylation of promoter regions and CpG islands coupled with a relative depletion of DNA
744 methylation across gene bodies and intergenic regions on the inactive X chromosome (X_I)
745 compared to the active X chromosome (X_A) in females. Marsupial mammals share similar DNA
746 methylation depletion in gene bodies and intergenic regions of the inactive X chromosome;
747 however, they diverge from eutherian mammals in their promoter methylation patterns. Marsupial
748 promoters are modestly hypomethylated in the female X chromosomes (X_A and X_I) compared to
749 the male X chromosome (X_M).

Tables

Table 1. Functional annotation of enriched biological processes associated with gene sets containing tissue specific differentially methylated regions (DMRs).

Tissue	GO biological process term	Accession ID	p-value	q-value
Brain	central nervous system development	GO:0007417	5.79×10^{-13}	2.71×10^{-09}
Brain	generation of neurons	GO:0048699	1.47×10^{-12}	3.45×10^{-09}
Brain	head development	GO:0060322	2.84×10^{-12}	3.78×10^{-09}
Brain	Neurogenesis	GO:0022008	3.22×10^{-12}	3.78×10^{-09}
Brain	brain development	GO:0007420	4.98×10^{-12}	4.67×10^{-09}
Pancreas	response to endoplasmic reticulum stress	GO:0034976	2.14×10^{-11}	1.51×10^{-07}
Pancreas	oxoacid metabolic process	GO:0043436	1.58×10^{-10}	4.04×10^{-07}
Pancreas	organic acid metabolic process	GO:0006082	1.72×10^{-10}	4.04×10^{-07}
Pancreas	response to endogenous stimulus	GO:0009719	6.34×10^{-09}	1.12×10^{-05}
Pancreas	carboxylic acid metabolic process	GO:0019752	2.29×10^{-08}	3.24×10^{-05}
Skeletal Muscle	actin filament-based process	GO:0030029	1.26×10^{-09}	5.78×10^{-06}
Skeletal Muscle	actin cytoskeleton organization	GO:0030036	5.04×10^{-09}	1.16×10^{-05}
Skeletal Muscle	cytoskeleton organization	GO:0007010	1.50×10^{-08}	2.30×10^{-05}
Skeletal Muscle	cellular carbohydrate metabolic process	GO:0044262	2.19×10^{-08}	2.52×10^{-05}
Skeletal Muscle	embryonic morphogenesis	GO:0048598	2.37×10^{-07}	2.17×10^{-04}
Lung and Kidney	embryonic skeletal system morphogenesis	GO:0048704	1.50×10^{-10}	3.52×10^{-07}
Lung and Kidney	embryonic organ morphogenesis	GO:0048562	1.61×10^{-10}	3.52×10^{-07}
Lung and Kidney	embryonic skeletal system development	GO:0048706	5.24×10^{-10}	7.65×10^{-07}
Lung and Kidney	embryonic organ development	GO:0048568	1.13×10^{-08}	1.23×10^{-05}
Lung and Kidney	pattern specification process	GO:0007389	3.31×10^{-08}	2.90×10^{-05}

Gene ontology (GO) terms are presented for the top five most significantly enriched results of each tissue after correcting for multiple testing (FDR < 0.05). As the numbers of tissue specific DMRs for lung (n=22) and kidney (n=119) samples were so few, the corresponding gene sets were combined for this analysis.

Table 2. Correlation analysis of mean promoter and gene body DNA methylation and ranked gene expression.

Tissue	Genomic Region	Gene Count	Rho (p-value)
Brain	Promoter	5,396	-0.08 ($p = 2.28\times10^{-9}$)
	Gene body	5,443	-0.16 ($p < 2.2\times10^{-16}$)
Kidney	Promoter	9,268	-0.12 ($p < 2.2\times10^{-16}$)
	Gene body	9,379	-0.12 ($p < 2.2\times10^{-16}$)
Lung	Promoter	9,192	-0.13 ($p < 2.2\times10^{-16}$)
	Gene body	9,265	-0.19 ($p < 2.2\times10^{-16}$)

Spearman's rank correlation coefficients (ρ) and associated significances are reported for all tissues with both whole genome bisulfite sequencing (WGBS) data and RNA-seq expression data.

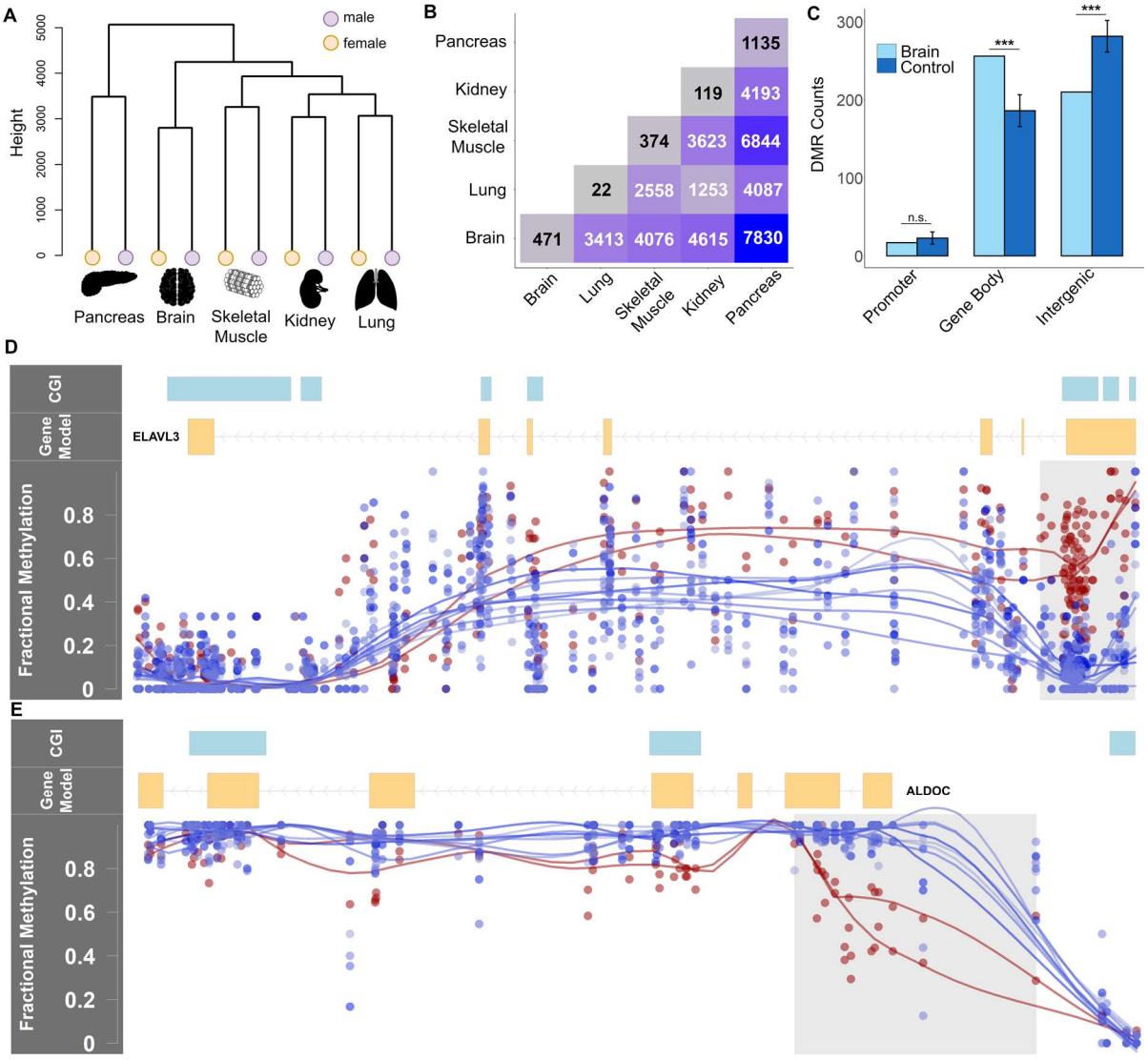
Table 3. Sex-based differential expression of the *lnc*RNA *Rsx* utilizing different data subsets and expression quantification tools.

Expression Dataset	Tool	Female Count	Male Count	Significance
All Data (n = 15)	DeSeq2	6987.1	16	p-value = 0.05
Matched Data (n = 6)	DeSeq2	6837.6	0	p-value = 2.04×10^{-30}
Matched Data (n = 6)	NOISeq	7872.4	0.67	Probability = 99.99%
Kidney Data (n = 2)	NOISeq	4074.4	0.68	Probability = 99.99%

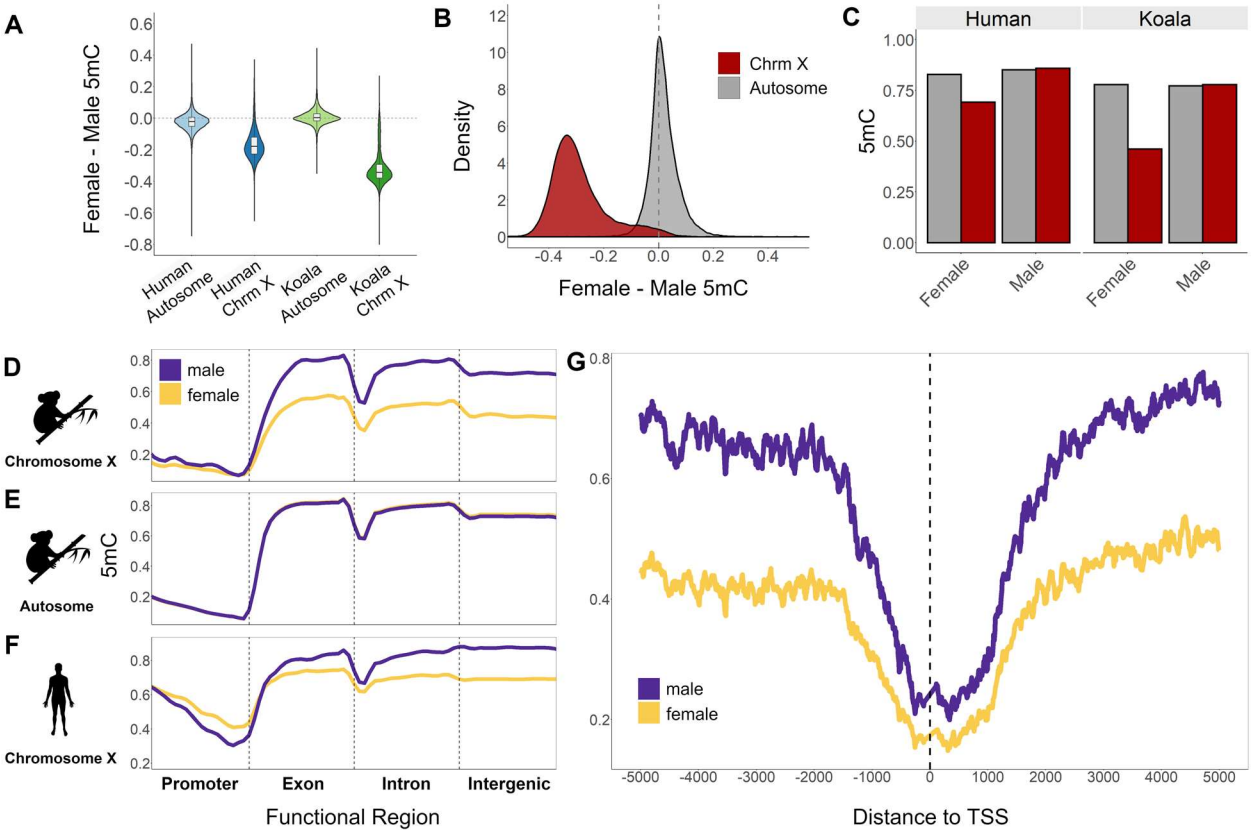
Normalized expression count values and significance of sex-based differential expression is shown for three data subsets using two expression quantification tools. All data refers to the dataset considering all 15 RNA-seq samples (7 male and 8 female). Matched data includes the tissues with both male and female RNA-seq samples (brain, kidney, and lung), and the kidney data is reported independently. DeSeq2 reports significance as an associated p-value from the Wald test while NOISeq reports a probability of differential expression threshold.

Figures

Figure 1.

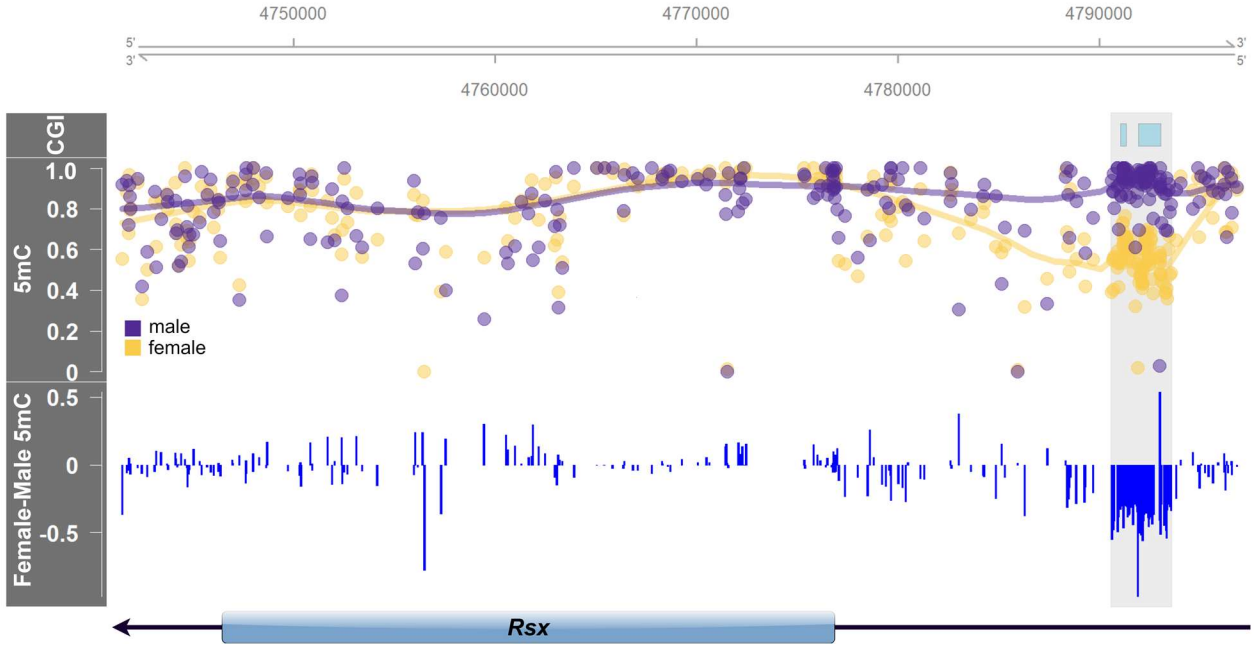


821 **Figure 2.**
822



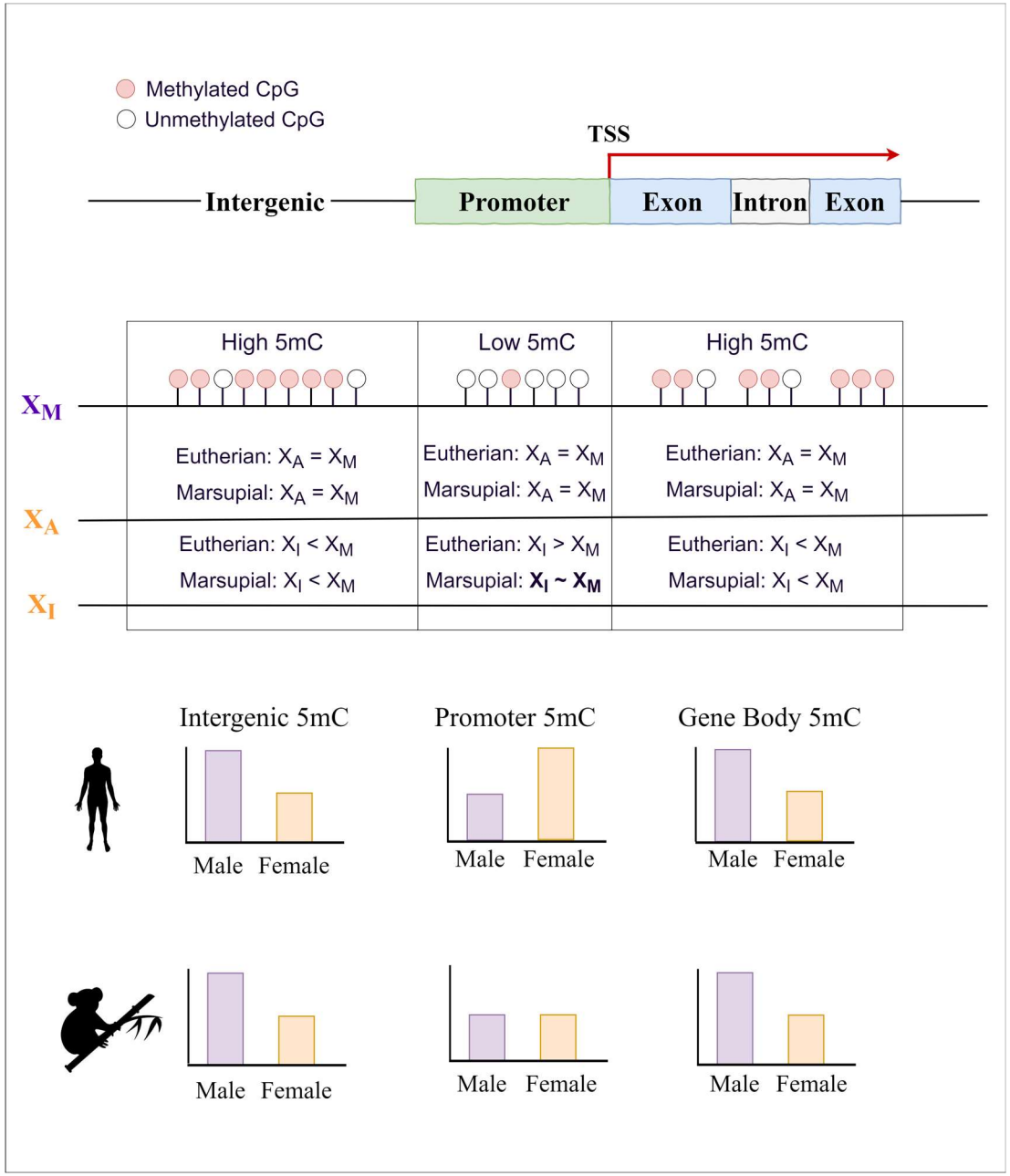
823
824
825
826

827 **Figure 3.**



828
829

Figure 4.


Methylated CpG
Unmethylated CpG
TSS
Intergenic
Promoter
Exon
Intron
Exon
High 5mC
Low 5mC
High 5mC
X_M
X_A
X_I
Eutherian: $X_A = X_M$
Marsupial: $X_A = X_M$
Eutherian: $X_I < X_M$
Marsupial: $X_I < X_M$
Eutherian: $X_A = X_M$
Marsupial: $X_A = X_M$
Eutherian: $X_I > X_M$
Marsupial: $X_I \sim X_M$
Eutherian: $X_A = X_M$
Marsupial: $X_A = X_M$
Eutherian: $X_I < X_M$
Marsupial: $X_I < X_M$
Intergenic 5mC
Promoter 5mC
Gene Body 5mC
Male Female
Male Female
Male Female
Male Female
Male Female
Male Female

Supplementary Tables

Supplementary Table 1. Overview of whole genome bisulfite sequencing (WGBS) data for all 10 koala samples.

Sample	Australian Museum registration	Name	Tissue	Mapped Reads	De-duplicated Reads	Total CpGs	Coverage (%)	Mean Depth	% Reads Cov > 3×	Bisulfite Conversion Rate
WGM145_01_S1	M.45022	Pacific Chocolate	Brain	602055711	228244836	16761785	97.5	13.99	93.4	98.2
WGM145_04_S2	M.47723	Ben	Brain	677593548	233893124	16761785	97.6	14.61	93.8	98.0
WGM145_06_S3	M.45022	Pacific Chocolate	Kidney	559911229	213274030	16761785	97.2	13.62	92.5	98.7
WGM145_08_S4	M.47723	Ben	Kidney	589121434	202094771	16761785	97.2	12.93	92.3	98.7
WGM145_09_S5	M.45022	Pacific Chocolate	Lung	637878987	210002505	16761785	97.5	13.8	93.2	98.6
WGM145_12_S6	M.47723	Ben	Lung	663204935	198555030	16761785	97.3	12.54	92.4	98.6
WGM145_14_S7	M.45022	Pacific Chocolate	Skeletal Muscle	592423486	168450992	16761785	96.7	10.93	90.0	98.7
WGM145_16_S8	M.47723	Ben	Skeletal Muscle	605979530	168220022	16761785	96.8	11.12	90.4	98.6
WGM145_19_S9	M.45022	Pacific Chocolate	Pancreas	563288200	159866998	16761785	96.2	9.857	88.0	98.6
WGM145_20_S10	M.47723	Ben	Pancreas	598508860	166573663	16761785	96.6	10.56	89.4	98.7

Supplementary Table 2. GC-content for all scaffolds linked to annotated autosomes (chromosomes 1-7) and the X chromosome.

Chromosome	Size (Mb)	Total GC Content (%)
1	730.67	38.72
2	595.75	38.58
3	480.35	39.35
4	412.55	39.38
5	299.29	39.56
6	248.35	38.66
7	250.28	39.20
X	67.91	42.02

Supplementary Table 3. Enrichment and significance of all tissue specific DMRs compared to length and GC matched control regions.

Tissue	Genomic Region	DMR Counts (%)	Enrichment	p-value
Pancreas	Promoter	52 (4.5%)	1.13	0.74
	Gene body	612 (52.5 %)	1.44	**< 0.0001**
	Intergenic	502 (43.0 %)	-1.40	**< 0.0001**
Brain	Promoter	17 (3.5%)	-1.35	0.25
	Gene body	256 (53.0 %)	1.38	**< 0.0001**
	Intergenic	210 (43.5%)	-1.34	**< 0.0001**
Skeletal Muscle	Promoter	19 (4.9 %)	-1.23	0.32
	Gene body	202 (51.8 %)	1.27	**< 0.0001**
	Intergenic	169 (43.3 %)	-1.26	**0.001**
Kidney	Promoter	5 (4.0 %)	-1.34	0.40
	Gene body	62 (50.0 %)	1.25	**< 0.0001**
	Intergenic	57 (46.0%)	-1.21	**0.04**
Lung	Promoter	1 (4.3 %)	-1.61	0.51
	Gene body	12 (52.2%)	1.37	**< 0.0001**
	Intergenic	10 (43.5%)	-1.31	0.14

Reported are the total counts of tissue specific differentially methylated regions (DMRs) falling within one of three annotated genomic regions: promoters, gene bodies, and intergenic regions. The enrichment of DMRs in each functional region is shown through a fold change comparison with a control dataset generated from 10,000 bootstraps using length and GC matched control regions. All significant p-values ($p < 0.05$) are highlighted in bold.

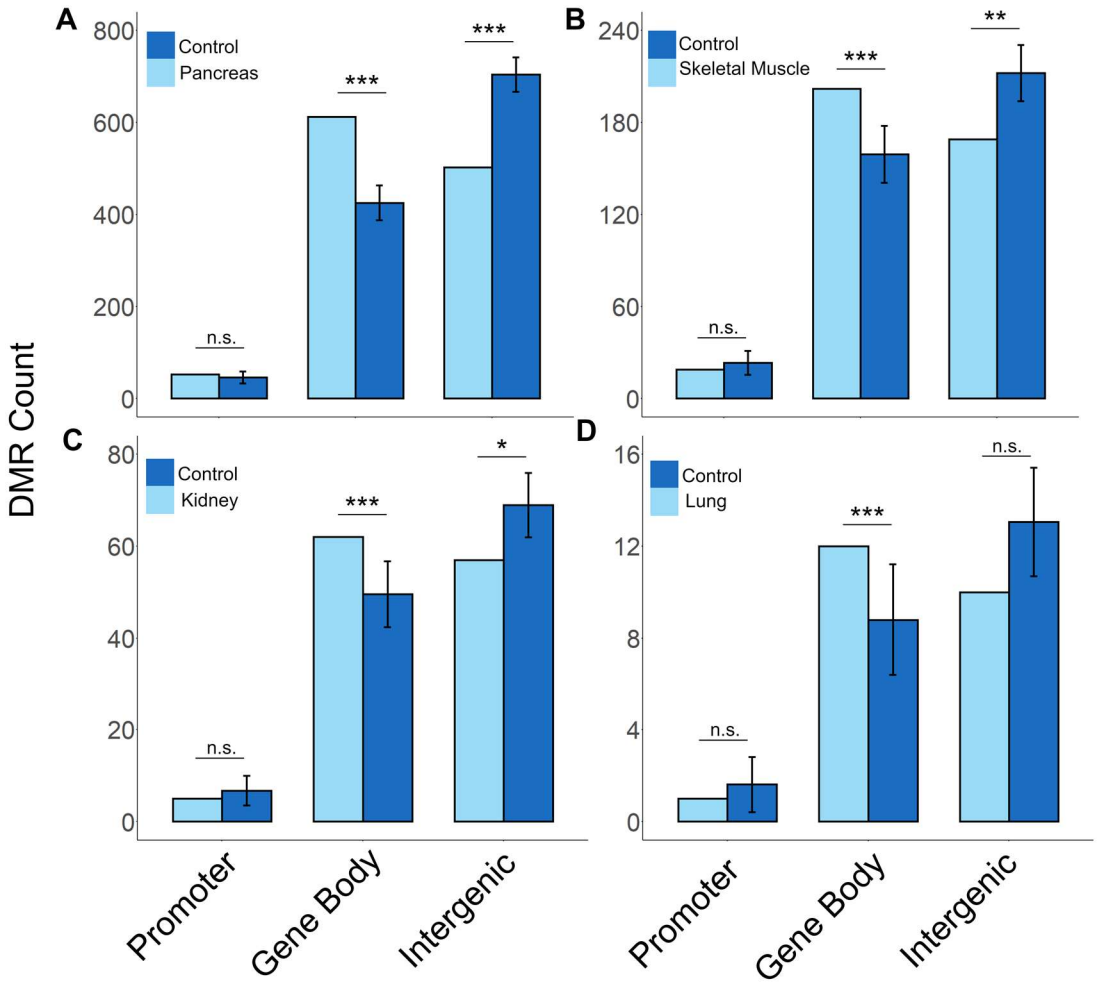
Supplementary Table 4. Mean and median sex-based DNA methylation difference calculated for all candidate X-scaffolds (n=98) by tissue.

Tissue	Median Male-Female 5mC	Mean Male-Female 5mC
Pancreas	-0.2282 ± 0.15	-0.2066 ± 0.12
Brain	-0.3273 ± 0.14	-0.2911 ± 0.12
Skeletal Muscle	-0.2748 ± 0.15	-0.2431 ± 0.12
Kidney	-0.2850 ± 0.15	-0.2567 ± 0.12
Lung	-0.2706 ± 0.14	-0.2437 ± 0.11
Combined Tissues	-0.2773 ± 0.15	-0.2484 ± 0.12

The female and male mean fractional DNA methylation (methylated reads/total reads per CpG) was calculated for all CpGs within 10 kb bins across candidate scaffold.

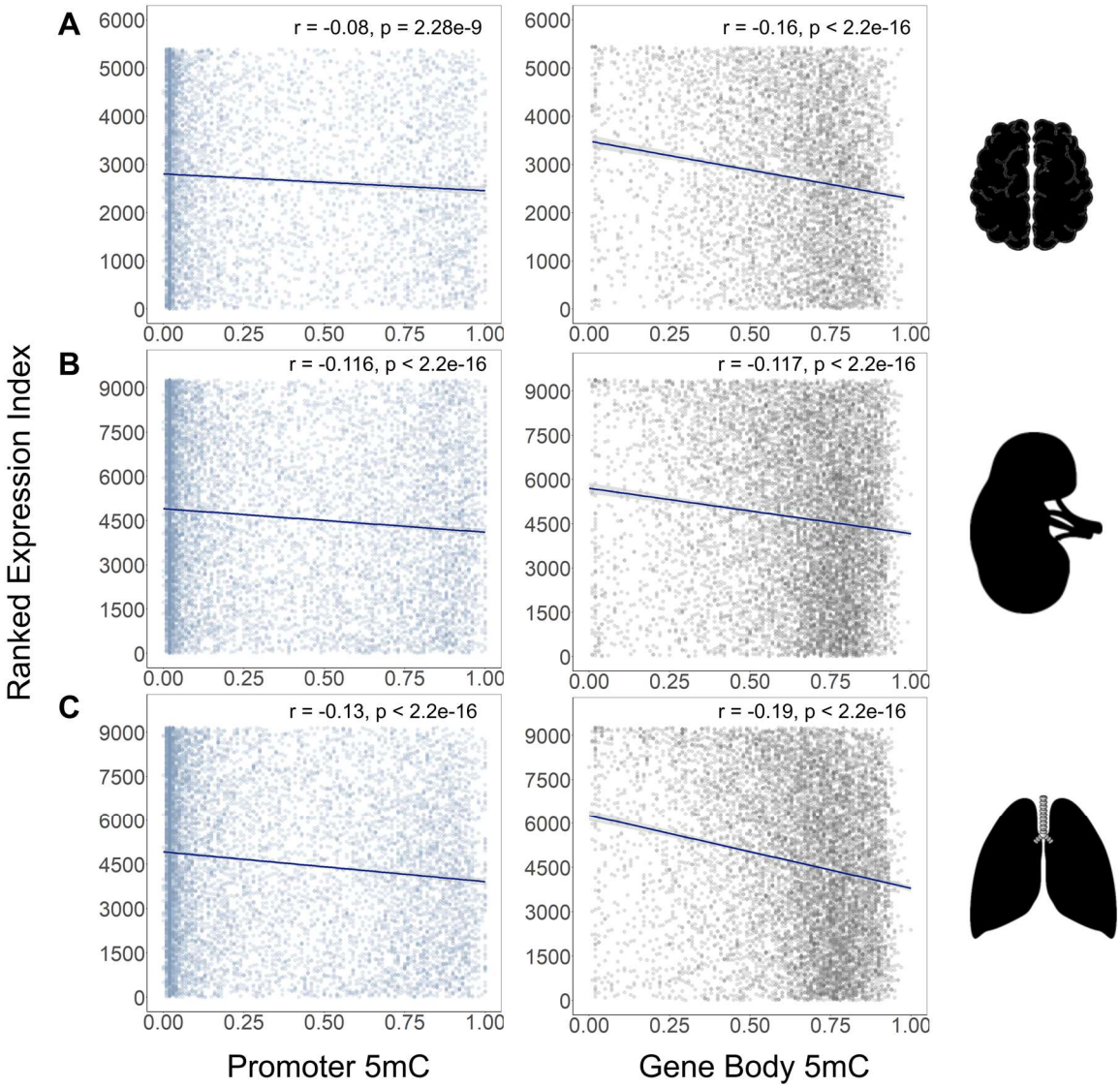
Supplementary Figures

Supplementary Figure 1. Enrichment of tissue specific differentially methylated regions (DMRs) falling within genomic functional regions.



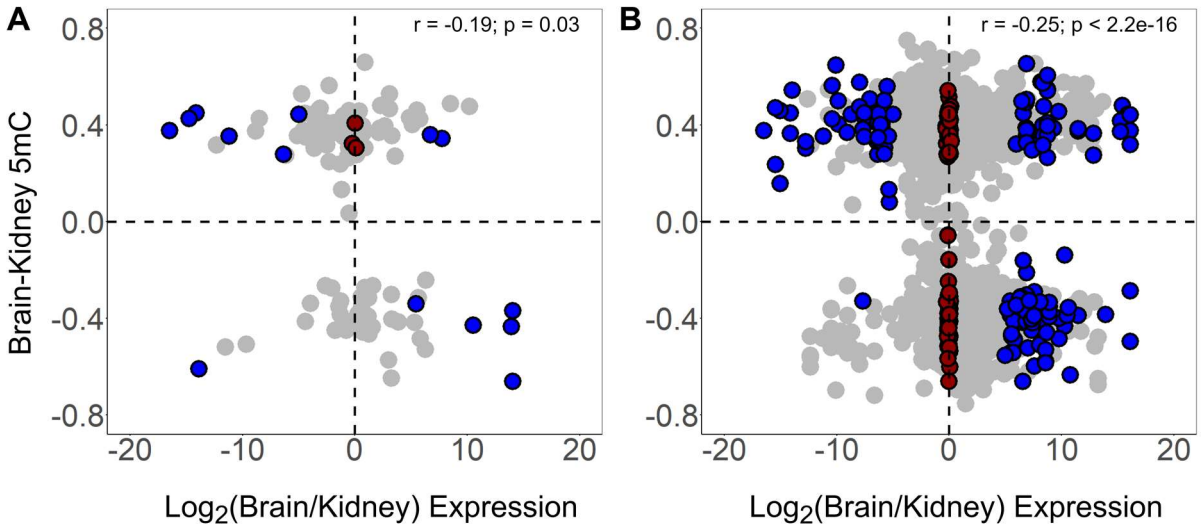
Data shown for (A) pancreas, (B) skeletal muscle, (C) kidney, and (D) lung. The enrichment of DMRs in each functional region (promoter, gene body, and intergenic regions) is shown through a comparison with length and GC matched control regions (*** indicates $p < 0.0001$, ** indicates $p < 0.001$, * indicates $p < 0.05$, and non-significance is shown by n.s. based on 10,000 bootstraps). Error bars indicate standard deviation.

Supplementary Figure 2. Correlation of gene expression and DNA methylation (5mC) in CpGs across promoters and gene bodies.



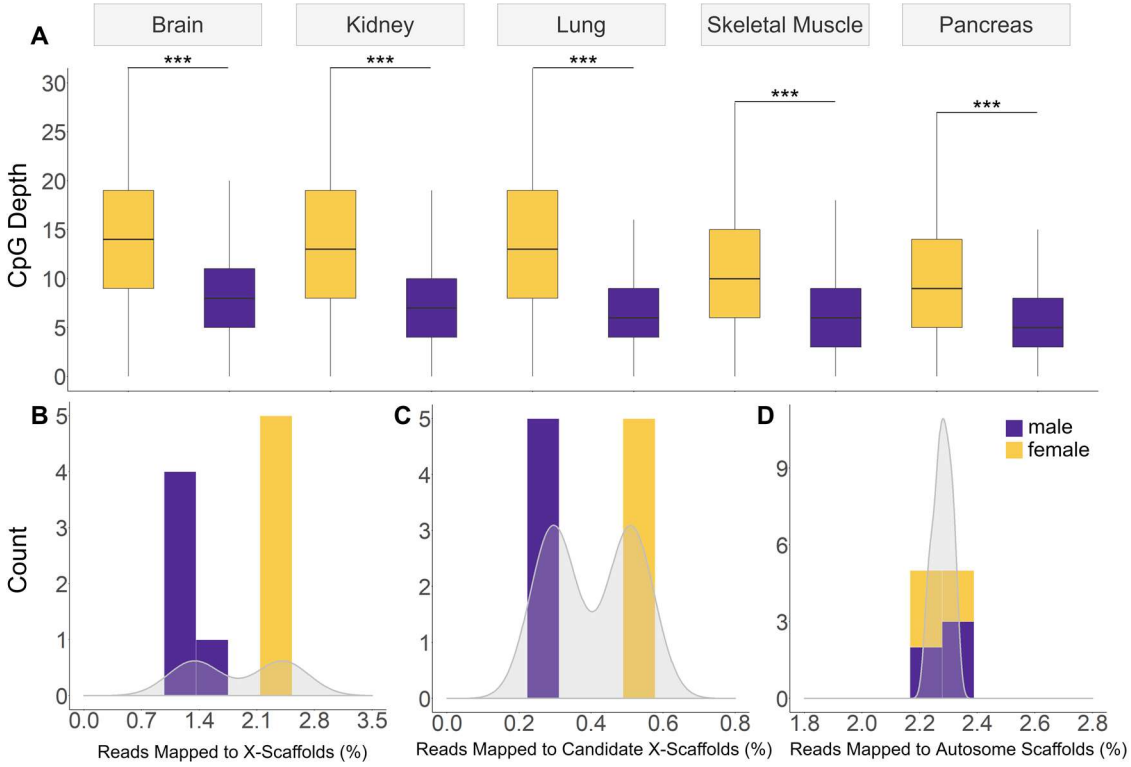
Three tissues with both whole genome bisulfite sequencing (WGBS) DNA methylation data and RNA-seq gene expression data are shown, (A) brain (n = 5,396 promoters and n = 5,443 gene bodies), (B) kidney (n = 9,268 promoters and n = 9,379 gene bodies), and (C) lung (n = 9,192 promoters and n = 9,265 gene bodies). For A-C, TPM expression values were ranked from low to high for each gene and correlated with mean fractional DNA methylation (methylated reads/total reads per CpG site). Spearman's rank correlation coefficients and the associated p-values are reported.

Supplementary Figure 3. Correlation of tissue dependent DNA methylation (5mC) and gene expression from brain and kidney samples.



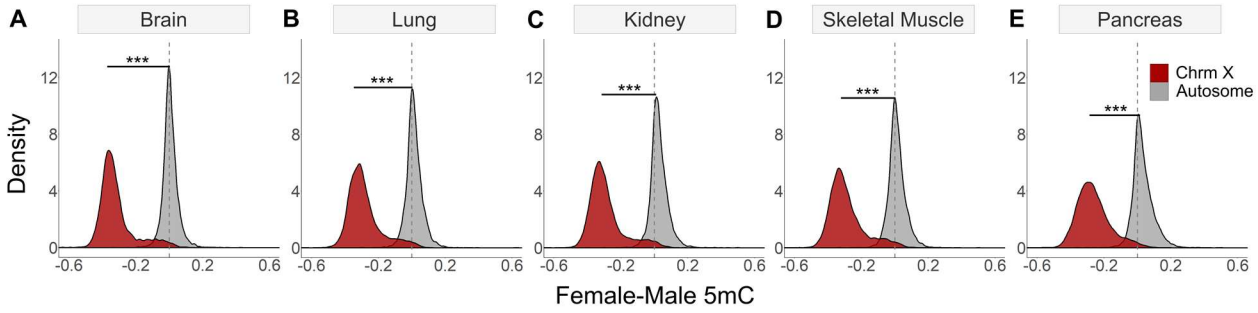
(A) The mean brain and kidney DNA methylation difference calculated for all CpGs across each gene promoter matched with the corresponding log-transformed ratio of brain to kidney expression. (B) The mean brain and kidney DNA methylation difference calculated for all CpGs across each gene body and matched with corresponding log-transformed ratio of brain to kidney expression. For A and B, Spearman's rank correlation coefficient and the associated p-value is reported. Blue dots indicate genes that are significantly differentially express between brain and kidney samples (probability of differential expression > 95% based on NOISeq) and red dots show all genes that are significantly similarly methylated in brain and kidney samples (probability of differential expression < 5% based on NOISeq).

Supplementary Figure 4. Sex-specific CpG depth of coverage and read mapping to autosomes and X chromosomes.



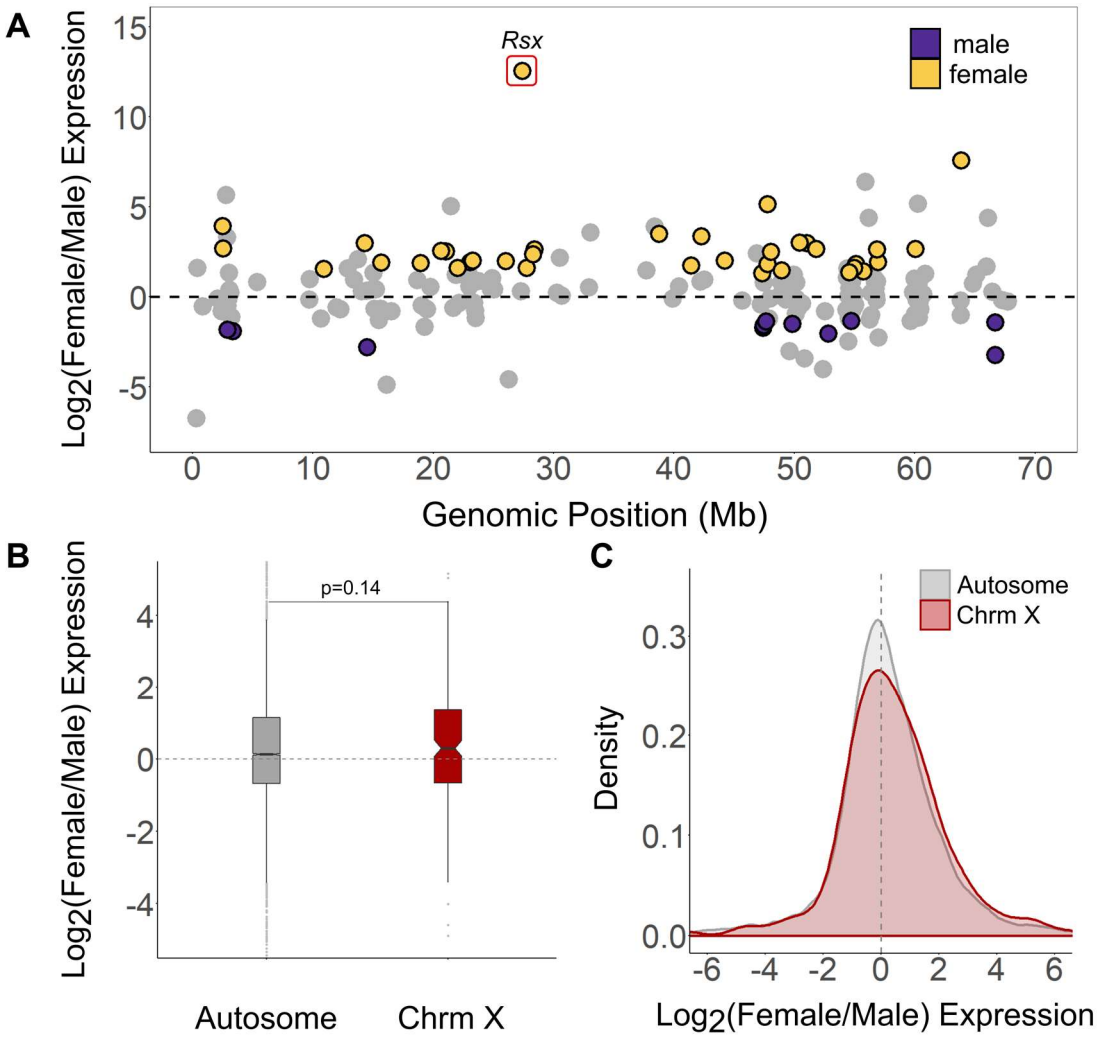
(A) For each of the five tissues, box-and-whisker plots of CpG depths across the X chromosome in male (purple) and female (orange) samples (*** indicates p < 2.2x10-16, Mann-Whitney U test). Histogram and distribution of sex-based read mapping per sample (n=10) to (B) X-linked scaffolds, (C) candidate X-linked scaffolds, and (D) a subset of autosome-linked scaffolds matched in length with all known X-linked scaffolds. For A-C, the percent of reads mapping to the scaffold category of interested over the total number of mapped reads in the genome was calculated for all male (n=5) and female samples (n=5). The known X-linked and candidate X-linked scaffolds show a bimodal distribution with an increase of read mapping to female samples expected from the 2:1 ratio X chromosomes in females to males. This bimodality is not observed in autosomes.

Supplementary Figure 5. The distribution of sex-based CpG fractional DNA methylation (5mC) differences across autosomes and X chromosomes.



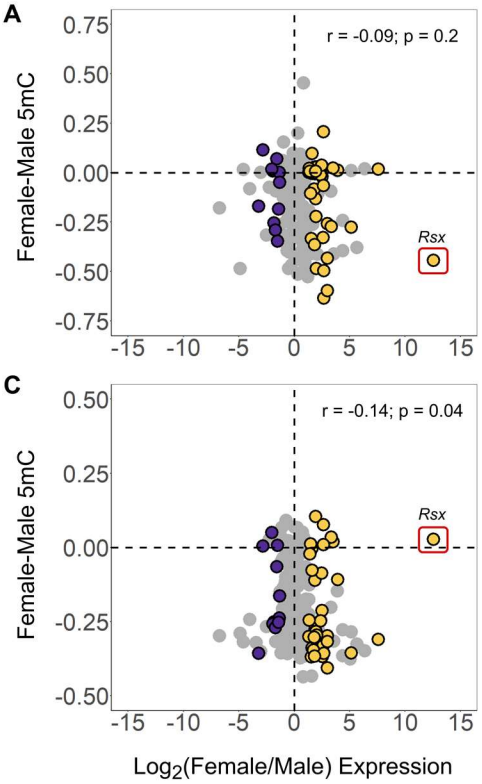
The distribution of female and male mean fractional methylation difference from (A) brain, (B) lung, (C) kidney, (D) skeletal muscle and (E) pancreas samples across autosomes and X chromosomes. For A-E, the female and male mean fractional methylation (methylated reads/total reads per CpG) was calculated for all CpGs within 10 kb bins across each autosome- or X-linked scaffold. All tissues exhibited a significant shift towards female hypomethylation in the X chromosome compared to the autosome (*** indicates $p < 2.2x10-16$, Welch's t-t

Supplementary Figure 6. Female and male gene expression across autosomes and the X chromosome using kidney RNA-seq data.



(A) The relative genomic location of all genes linked to X chromosome scaffolds aligned by scaffold length and the log-transformed female to male expression ratio generated by NOISeq. Orange dots indicate the 36 genes with significant female-biased expression and purple dots indicate the 11 genes with significant male-biased expression (probability of differential expression > 95% based on NOISeq). For all autosome-linked genes (n = 10,414) and chromosome X-linked genes (n= 209), a box-and-whisker plot (B) and density distribution (C) of the log-transformed female to male expression ratio (p = 0.14, Mann-Whitney U test).

Supplementary Figure 7. Correlation analysis of sex-based DNA methylation (5mC) and gene expression across chromosome X using kidney WGBS and RNA-seq data.



B

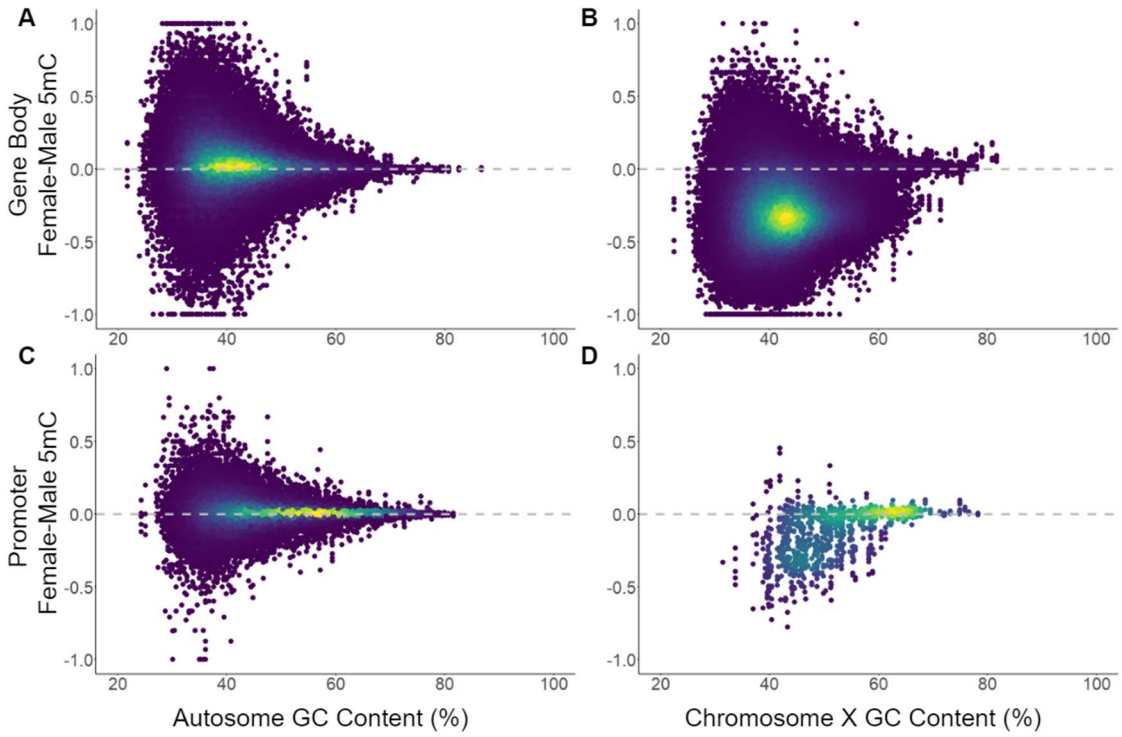
Promoter	Male Over Expressed / Total Genes	Female Over Expressed / Total Genes
Female Hypermethylated	$\frac{5}{32} = 0.16$	$\frac{14}{44} = 0.32$
Female Hypomethylated	$\frac{6}{57} = 0.11$	$\frac{22}{73} = 0.30$

D

Gene Body	Male Over Expressed / Total Genes	Female Over Expressed / Total Genes
Female Hypermethylated	$\frac{3}{11} = 0.27$	$\frac{7}{15} = 0.47$
Female Hypomethylated	$\frac{8}{79} = 0.10$	$\frac{29}{103} = 0.28$

(A) The mean female and male fractional methylation difference calculated for all CpGs across each gene promoter matched with the corresponding log-transformed ratio of female to male expression. (B) The ratio of the number female hypermethylated and female hypomethylated gene promoters that show either significant male of female biased expression over the total number of genes in each category. (C) The mean female and male fractional methylation difference calculated for all CpGs across each gene body and matched with corresponding log-transformed ratio of female to male expression. (D) The ratio of the number female hypermethylated and female hypomethylated gene bodies that show either significant male of female biased expression over the total number of genes in each category. For A and C, Spearman's rank correlation coefficient and the associated p-value is reported. The Rsx gene was excluded from the correlation calculation. Orange dots indicate the 36 genes with significant female-biased expression and purple dots indicate the 11 genes with significant male-biased expression (probability of differential expression > 95% based on NOISeq).

Supplementary Figure 7. Sex-based DNA methylation (5mC) by GC-content across autosomes and the X chromosome.



For A and B, the mean female and male methylation difference calculated from CpGs in 1 Kb bins across (A) autosomes and (B) X chromosomes. For C and D, mean female and male methylation difference calculated from CpGs located in promoter regions (defined as regions 1 kb upstream of known gene TSSs) in (C) autosomes and (B) X chromosomes. For A-D, data from all five tissues (brain, kidney, lung, pancreas, and skeletal muscle) are reported. All plots are coloured by data density where blue represents low density regions and yellow represents high density regions.