Divergent DNA methylation signatures of 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
dap@unr.edu
775-682-5735

Soojin V. Yi
soojin.yi@biology.gatech.edu
404-385-6084

Abstract

X chromosome inactivation (XCI) mediated by differential DNA methylation between sexes is well characterized in eutherian mammals. Although XCI is shared between eutherians and marsupials, the role of DNA methylation in marsupial XCI 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. First, tissue-specific differential DNA methylation in marsupials primarily occurs in gene bodies. Second, females show significant global reduction (hypomethylation) of X chromosome DNA methylation compared to males. We show that this pattern is also observed in eutherians. 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 XCI, 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 functionally conserved patterns of XCI in two deep branches of mammals.

Keywords: X chromosome inactivation, marsupial, DNA methylation, 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 (1). Although XCI occurs in both eutherian and marsupial mammals (2), there are several notable differences in XCI between the two lineages. First, in eutherians, the transcription of *Xist* RNA from the inactive X chromosome is essential for XCI (3-5). However, the *Xist* locus is not present in marsupials (6, 7). Instead, another *lnc*RNA gene, *Rsx*, drives marsupial XCI (8). Second, marsupials exhibit 'imprinted' XCI by selectively silencing the paternal X chromosome (9, 10). In contrast, XCI in eutherians occurs randomly between the maternally and paternally derived X chromosomes, although paternal XCI has been observed during early rodent development (11, 12). Third, while eutherian XCI involves the exclusion of active histone marks and the recruitment of repressive histone marks on the inactive X chromosome (13), marsupial X chromosomes do not show a consistent pattern (10, 14). These differences suggest that evolutionary pathways leading to XCI likely differ between eutherians and marsupials, and that novel insights into the mechanism of XCI can be gained from comparative studies.

The role of DNA methylation on marsupial XCI has been particularly controversial. Immunofluorescent labeling studies observed relative hypomethylation of the inactivate X chromosome in marsupials (15, 16). Other studies found little difference in DNA methylation between active and inactive marsupial X chromosomes (10, 17, 18). Recently, Waters et al. (19) analyzed reduced representation bisulfite sequencing (RRBS) data of a male and female opossum (*Monodelphis domestica*) and proposed that female X chromosomes in marsupials, but not in eutherians, exhibit hypomethylation near the transcription start sites (TSSs). Notably, these previous DNA methylation studies either examined a small number of CpGs or employed methodologies that over-represent promoters and CpG islands (in case of RRBS, (20)). Since patterns of DNA methylation vary greatly among distinctive genomic regions with different functional consequences, it is necessary to extend our knowledge to unbiased, whole-genome assays of DNA methylation.

We used the whole-genome bisulfite sequencing, which captures the DNA methylation state of every nucleotide in the genome, of the modern koala (*Phascolarctos cinereus*) to understand the role of DNA methylation on XCI in marsupials and to resolve previous

controversies. We compared DNA methylation maps of five diverse tissues from a male and a female koala, providing the first multi-tissue, whole genome methylome resource of any marsupial with information on tissue-specific variation of DNA methylation. Utilizing these novel resources, we show distinctive impacts of DNA methylation on tissue-specific gene expression in marsupials, as well as on XCI in eutherians and marsupials. We further classify previously undetected X-linked regions from this 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 XCI and regulation of gene expression in two deep branches of mammals.

Results

Genome-wide patterns of DNA methylation in the modern koala

We generated whole genome bisulfite sequencing (WGBS) data from 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). The mean depth of coverage fell between 9.9× and 14.6× (Supplementary Table 1). Autosome-linked scaffolds displayed 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), similar to (21). We performed a hierarchical clustering analysis and observed a clear clustering of samples by tissue (Fig. 1A) 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 (22). 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 multiple tissues (Fig. 1B). Consistent with the results of the clustering analysis, the pancreas samples contained the most tissue-specific DMRs followed by the skeletal muscle, brain, kidney and lung (Fig. 1B). 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

97 10,000 bootstraps; Fig. 1C, Supplementary Fig. 1 and Supplementary Table 3). On the other hand,
98 DMRs were significantly depleted in intergenic regions compared to the control regions ($p < 0.05$
99 based on 10,000 bootstraps; Fig. 1C, Supplementary Fig. 1, and Supplementary Table 3). Genes
100 containing tissue-specific DMRs were enriched in specific biological functions (Table 1). For
101 example, brain specific DMRs were linked to genes associated with neural developmental
102 processes such as neurogenesis and central nervous system development.

103 We integrated our methylome data with a previously generated RNA-seq koala
104 transcriptomes (23). Of the 12 tissues surveyed in that study, three tissues (kidney, brain, and lung)
105 overlapped with the tissues examined here. Promoter DNA methylation and gene expression were
106 significantly negatively correlated across the entire genome (Table 2, Supplementary Fig. 2). In
107 contrast, the relationship between gene body DNA methylation and expression was complex; both
108 extremely hypomethylated and hypermethylated gene bodies showed high gene expression (Table
109 2, Supplementary Fig. 2). We considered differentially methylated genes (DMGs) containing
110 DMRs between the brain and kidney samples (n = 1944 genes from n = 4,615 DMRs) compared
111 with differentially expressed genes (DEGs) between brain and kidney RNA-seq samples.
112 Currently, available RNA-seq data from koalas do not include sufficient biological replicates. We
113 overcame this limitation by simulating replicates within each RNA-seq data set (NOISeq, (24))
114 and identified 600 putative DEGs (probability of differential expression > 95% according to the
115 NOISeq). We found that DMGs were significantly more likely to be differentially expressed than
116 genes that did not contain DMRs, exhibiting a 1.54-fold enrichment ($\chi^2 = 33.07$, $p < 0.0001$). In
117 addition, differential expression between tissues was significantly correlated with differential
118 promoter DNA methylation between tissues (Supplementary Fig. 3A). Both relative
119 hypomethylation and relative hypermethylation of gene bodies were associated with increased
120 expression (Supplementary Fig. 3B). In Fig. 1D and 1E, we show DNA methylation patterns of
121 two representative genes containing brain specific DMRs.
122

123 **Global hypomethylation of female X chromosome in koalas**

124 The koala genome project used cross-species chromosome painting data to identify 24 putative X
125 chromosome scaffolds and 406 putative autosomal scaffolds (25). As expected from 2:1 ratio of
126 X chromosomes in females compared to males, the median depth of coverage of CpGs on the
127 putative X scaffolds were consistently higher (approximately 2-fold) in female samples compared

128 to male samples ($p < 2.2 \times 10^{-16}$, Mann-Whitney *U* test, Supplementary Fig. 4A). Additionally, the
129 proportion of reads mapped to the putative X scaffolds showed a distinct bimodal distribution
130 whereby the male samples cluster close to 1.3% and the female samples cluster near 2.4%
131 (Supplementary Fig. 4B). In contrast, male and female samples were indistinguishable with respect
132 to read mapping to putative autosomes (Supplementary Fig. 4D).
133 The global DNA methylation level of the female X chromosome was lower than that of the
134 male X chromosome in all koala tissues examined (Fig. 2A, B and Supplementary Fig. 5, $p < 2.2$
135 $\times 10^{-16}$, Mann-Whitney *U* test). A comparison to autosomal DNA methylation indicated that the X
136 chromosome exhibits reduced DNA methylation in females. Consequently, we use the term
137 'female hypomethylation' (as opposed to male hypermethylation) (Fig. 2C) consistently in this
138 work. As a comparative basis of the DNA methylation patterns observed in eutherian mammals,
139 we performed parallel analyses on human DNA methylation data (methods). Human X
140 chromosomes were also globally hypomethylated in females compared to males (Fig. 2A, C).
141 Female hypomethylation was observed in all functional regions across the X chromosome (Fig.
142 2D), but most pronounced in gene bodies and intergenic regions. Promoters showed the least
143 difference between males and females. As expected, the autosomal scaffolds did not display a
144 significant variation between female and male methylation levels in any functional region (Fig.
145 2E). Fig. 2F depicts the DNA methylation difference between male and female X chromosomes
146 in humans. In contrast to the pattern observed in koalas, promoters on the human female X
147 chromosomes were hypermethylated compared to those on the male X chromosome, congruent
148 with previous studies (26-28).
149
150 **Promoter DNA methylation is not a universal driver of sex-specific expression in koalas**
151 To investigate the implications of the observed sex-specific DNA methylation, we once again
152 utilized the published RNA-seq koala transcriptomes (23). Of the total RNA-seq dataset, only one
153 tissue (kidney) had expression data from both a male and female koala that could be directly
154 compared to our methylome dataset. Consequently, the other tissues were not considered for
155 further analysis. We calculated the log-transformed fold change of female and male gene
156 expression values using NOISeq, capitalizing on its ability to simulate technical replicates. Of the
157 209 X-linked genes, 36 (17.2%) exhibit female overexpression while 11 (5.3%) show male
158 overexpression (probability of differential expression > 95% based on NOISeq, Supplementary

159 Fig. 6A). Although, on average, autosomal genes also exhibited slight female-bias of expression
160 (Supplementary Fig. 6B, C), the increase was more substantial in the X chromosome (mean
161 chromosome X female to male $\log_2$ fold change = 0.50, autosome female to male expression $\log_2$
162 fold change = 0.24). We calculated the female and male fractional methylation difference in X
163 chromosome-linked promoters and gene bodies correlated with gene expression difference (N =
164 209 gene bodies and N = 206 promoters, excluding 3 promoters with CpGs coverage < 3). In
165 promoters, no significant relationship was observed (Supplementary Fig. 7A). In fact, the
166 proportions of significantly female-over expressed genes were similar between female-
167 hypermethylated promoters and female-hypomethylated promoters (Supplementary Fig. 7B).
168 Interestingly, female and male DNA methylation difference in gene bodies showed an overall
169 negative correlation with gene expression (Spearman's rank correlation coefficient, $\rho = -0.14$, $p =$
170 0.04, Supplementary Fig. 7C). A deeper analysis of the relative DNA methylation and expression
171 levels revealed that both extreme hypomethylation and hypermethylation are associated with
172 increased expression (Supplementary Fig. 7C) attesting to the complexity and heterogeneity of the
173 relationship between gene body DNA methylation and gene expression.
174

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

177 Sex-specific DNA methylation of a key regulator of XCI in other marsupials, the *lnc*RNA gene
178 *Rsx*, has been associated with its differential expression (8, 10). Based on the sequence homology
179 with the *Rsx* gene from the gray short-tailed opossum (*Monodelphis domestica*) (8), we identified
180 a 29.8 Kb candidate *Rsx* sequence (Methods), using PacBio long read sequencing by Johnson et
181 al. (25). Interestingly, we observed a female hypomethylated region containing two CpG islands
182 upstream of the candidate *Rsx* covering 101 CpGs exhibiting a 36% reduction in mean female
183 DNA methylation compared to mean male DNA methylation (mean sex difference: -0.36 ± 0.14,
184 Fig. 3). The mean DNA methylation difference between sexes within this region was in the top
185 23% of the distribution of differences across all X-linked scaffolds of the combined tissues. The
186 expression of the *lnc*RNA annotated within the candidate *Rsx* gene region was significantly greater
187 in females (mean normalized read count = 6987.1) than in males (mean normalized read count =
188 16, $p < 0.05$ from DeSeq2 using the Wald test). When considering a subset of tissue expression

189 data with matched male and female samples (spleen, kidney, and lung), the results remained robust
190 across different tools to measure differential gene expression (Table 3).
191
192 **Identification of novel candidate X-linked scaffolds by sex-specific methylation patterns**
193 We have demonstrated above that X-linked scaffolds exhibit several distinctive WGBS patterns in
194 koalas, including different depths of coverage per CpG, distinctive clustering based on the
195 proportion of mapped reads, and unique methylation distributions between females and males
196 (Supplementary Fig. 4). We thus quantified DNA methylation differences between females and
197 males to determine if additional candidate X scaffolds existed within the 6.7% of the koala
198 assembly that remained unclassified. We identified an additional 98 scaffolds that showed a clear
199 shift towards female hypomethylation (mean female-male 5mC for all candidate X scaffolds was
200 -0.25 ± 0.12) (Supplementary Tables 4,5). These candidate scaffolds contributed 14.6 Mb (21.5%)
201 to the total genomic region annotated as the koala X chromosome. All candidate scaffolds followed
202 the expected bimodal distribution seen in the putative X scaffolds; male samples clustered around
203 0.3% and the female samples clustered around 0.5% (Supplementary Fig. 4C). This clustering
204 pattern could be attributed to the 2:1 ratio of X chromosomes in females compared to males that
205 was not observed in putative autosomes.
206
207 **Discussion**
208 In this study, we analyzed novel nucleotide-resolution genomic DNA methylation maps of five
209 tissues of a male and a female koala. The overall DNA methylation levels of koala tissues are on
210 par with those in other mammals (27, 29, 30), and show heavy genome-wide DNA methylation
211 punctuated by hypomethylation of CpG islands (e.g., Fig. 1D,E). Tissue-specific differential DNA
212 methylated regions (DMRs) were significantly enriched in gene bodies. Gene body methylation is
213 an ancestral form of DNA methylation in animal genomes (e.g., (31, 32)), yet its role in regulation
214 of gene expression is less well understood than promoter DNA methylation owing in part to the
215 heterogeneity of gene body regions with respect to sequence length, composition and functionality
216 of introns and exons, and CpG density compared to promoter regions (31, 33). For example, DNA
217 methylation levels of the first exons/introns of genes are negatively correlated with gene
218 expression (34-36), and tend to be different from downstream genic regions (35). On the other
219 hand, high levels of cumulative gene body DNA methylation is positively correlated with gene

220 expression and may reduce spurious transcription of intragenic RNA (37, 38). These studies
221 illustrate the potential for both positive and negative regulatory control of gene expression by gene
222 body DNA methylation.

223 We show that female X chromosomes are globally hypomethylated compared to both the
224 male X chromosome and the autosomes of both sexes (Fig. 2). Even though it may appear
225 counterintuitive at the first glance, the hypomethylation of female X chromosome appears as a
226 common feature of eutherian and marsupial mammals. Hellmann and Chess (26) showed that the
227 inactive X chromosomes of humans had reduced gene body DNA methylation. Recent whole
228 genome bisulfite sequencing analyses of mouse (27) and humans (28) showed that the
229 hypomethylation of inactive X chromosome is pervasive in gene bodies and intergenic regions
230 (Fig. 4). Waters et al. (19) performed reduced representation bisulfite sequencing (RRBS) of
231 several species including mouse and opossum, and concluded that the reduction of methylation in
232 gene bodies was specific to marsupials, but not in eutherians (19). The reason why they did not
233 observe methylation difference in mouse might be due to the inherent bias of RRBS, which
234 disproportionately samples regions with high GC content (20). High GC-content regions tend to
235 be hypomethylated (39, 40) and show less variation of DNA methylation. For example, the degree
236 of methylation difference between sexes in koala promoters is small compared to other genomic
237 regions, at least partly due to high GC content of promoters (Supplementary Fig. 8). How
238 chromosome-wide DNA hypomethylation is linked to chromosome-wide gene silencing is
239 currently unknown. Interestingly, marsupial genomes harbor an additional copy of *DNMT1* (41),
240 which could lead to functional divergence between the mammalian lineages. Analyses of DMNT
241 expression however did not indicate significant differential expression of *DNMT*s between sexes
242 (probability of differential expression using NOISeq < 95%). Additional analysis with broad and
243 balanced sampling of both sexes is necessary to rigorously test the impact of the marsupial *DNMT1*
244 duplication on gene expression.

245 Interestingly, DNA methylation signatures of *Rsx,* the major player in XCI initiation in
246 marsupials (8), suggest that koala *Rsx* expression is regulated by promoter CpG island DNA
247 methylation (Fig. 3). Wang et al. (10) also showed differential DNA methylation of *Rsx* promoter
248 in opossum. Therefore, regulation of the key initiator of XCI via DNA methylation is another
249 parallel feature between eutherians and marsupials Notably, a recent study (42) found that *Xist* and

250 *Rsx* harbor non-linear sequence similarity. Consequently, their shared functionality may be
251 partially due to characteristics of tandem repeat regions.
252 In summary, we show that gene body DNA methylation is an important contributor to
253 differential expression between tissues in koala. We also show that the global hypomethylation of
254 female X chromosome is a conserved feature of X chromosome regulation in eutherian and
255 marsupial mammals. However, X chromosome promoter methylation and the subsequent effect
256 on the regulation of gene expression appear to be divergent between these two lineages.
257 Regulation of the *Rsx*, on the other hand, is supported by DNA methylation, which mirrors the
258 regulation of the eutherian *Xist* locus. Together, these conclusions illuminate the intricate
259 evolutionary pathways that have diverged and converged to influence X chromosome regulation,
260 XCI, and dosage compensation in eutherian and marsupial mammals.
261

262 # Methods

263 **Whole genome bisulfite sequencing and processing**

264 All genomic DNA was extracted using a Bioline Isolate II Genomic DNA Extraction Kit (Cat#.
265 BIO-52067) following the recommended protocol with an additional DNAse free RNaseA
266 (100mg/ml) (Qiagen cat. #19101) treatment before column purification. 20mg tissue samples from
267 brain, kidney, lung, skeletal muscle, and pancreas from a female koala, "Pacific Chocolate"
268 (Australian Museum registration M.45022), and a male koala, "Ben" (Australian Museum
269 registration M.47723), were supplied to the Ramaciotti Centre for Genomics for methylome
270 sequencing. The bisulfite conversion was carried out by using the EX DNA Methylation-Lightning
271 Kit (Zymo cat. #D5030) and the WGBS libraries were constructed using the TruSeq DNA
272 methylation kit (Illumina cat.# EGMK81213). The libraries were sequenced on a NovaSeq6000
273 S2 (Illumina) using the 2 × 100bp PE option. Information on samples, coverage, and read counts
274 are provided in Supplementary Table 1. Processing of the WGBS data followed previous studies
275 (30). Bisulfite conversion rates were estimated for each WGBS sample following the methodology
276 of methPipe's bsrate (43) (Supplementary Table 1). Strand-specific methylation calls were
277 combined, and all samples were filtered to remove CpGs covered by fewer than three reads.
278

279 **Analyses of tissue differentially methylated regions**

280 A hierarchical clustering tree was drawn using the fractional methylation profiles of all ten samples
281 using hclust from R's stats package. The distance matrix was calculated using Euclidean distances
282 and the agglomeration method used was Ward's method. The data for the final tree was visualized
283 using R's dendextend package (44). Bismark generated CpG reports were filtered to remove
284 scaffolds that were less than 2 Mb in length, retaining 3.03×10^9 (94.8%) of the genome. DMRs
285 were called using BSmooth (22), with additional filtering of a minimum fractional methylation
286 difference of 0.3 (30%) and at least 5 CpG sites. DMRs were considered shared between tissues if
287 they overlapped by at least 50%. Using koala gene annotations from Ensembl
288 (Phascolarctos_cinereus.phaCin_unsw_v4.1.97 release), promoters were defined as regions
289 located 1000 bp upstream of the identified transcription start site (TSS). To test the enrichment of
290 tissue-specific DMRs in functional regions, we generated 10,000 length and GC content matched
291 controls for all unique DMRs. The lung and kidney samples shared the most similar methylation
292 profiles and consequently had few tissue-specific DMRs (Fig. 1C). Because of this feature, the
293 corresponding gene sets were combined for these two tissues. Functional annotation and GO term
294 enrichment analysis was performed utilizing the ToppGene Suite (45).
295
296 **Differential DNA methylation between sexes**
297 Johnson et al. (25) used cross-species chromosome painting data and linked 406 scaffolds spanning
298 2.9 Gb of sequence data to autosomal scaffolds from chromosomes 1-7 and 24 scaffolds covering
299 68 Mb of sequence data to chromosome X, leaving 6.7% of the genome unclassified. To explore
300 the same relative amount of genomic space, we randomly sampled a subset of the autosomal
301 scaffolds that were length matched with the X chromosome scaffolds, which we called the
302 "matched autosome" dataset. These scaffolds were divided into 10-kb bins and the difference in
303 mean fractional methylation at each 10-kb bin was compared between male and female samples
304 for all tissues. For the analysis of human data, we used WGBS fractional methylation reports from
305 a male brain (Epigenome ID: E071) and a female brain (Epigenome ID: E053) and the human
306 known gene annotations from Ensembl (hg19 release). Due to its similarity in size to the human
307 X chromosome, we used data from human chromosome 8 as our representative autosome in the
308 comparative analysis.
309
310 **Identification of candidate X-linked scaffolds**

311 To isolate candidate X-linked scaffolds from the 1,477 unclassified koala scaffolds, we binned the
312 unclassified scaffolds into 10-kb windows and calculated the mean fractional methylation of the
313 associated CpGs. We then determined the average female and male methylation differences across
314 the bins and plotted the density of the differences for all five tissues. SVY and DS proceeded to
315 independently select scaffolds that exhibited a statistically significant shift towards female
316 hypomethylation from zero. The scaffolds that showed significant female hypomethylation in all
317 five tissues and were selected by both SVY and DS were isolated (n = 98 covering 14.6 Mb of
318 sequence with mean female-male 5mC = -0.25 ± 0.12). As an external validation, the percent of
319 reads mapping to the putative X-linked and autosome-linked scaffolds over the total number of
320 mapped reads was computed for the male and female sample in all tissues.
321
322 **Annotation of the koala *Rsx***
323 To annotate the genomic region around *Rsx*, we downloaded the published genome sequence fasta
324 files for the partial assembly of *Rsx* from the gray short-tailed opossum (8) and the complete
325 PacBio assembly of the koala *Rsx* (25, 42). We used BLASTN 2.2.29 (46) to align both sequences
326 to the koala reference genome (phaCin_unsw_v4.1) and obtained genomic coordinates. The entire
327 assembled koala *Rsx* sequence aligned with 100% identity and no gaps. Only one 30.4 kb
328 transcript, a novel *lnc*RNA, overlapped with the annotated *Rsx* region (overlap > 90% of transcript)
329 and was used to evaluate gene expression.
330
331 **Analysis of differential gene expression**
332 All RNA-seq expression data used in this analysis were obtained from the previously published
333 koala transcriptomes from eight females and seven males (23). To process the acquired data, we
334 followed the protocol outlined by Pertea et al. (47), using the koala GTF annotation from Ensembl
335 (Phascolarctos_cinereus.phaCin_unsw_v4.1.97.gtf.gz release) to assemble mapped reads into
336 transcripts using StringTie 2.0 (47) with the -e-b--A <gene_abund.tab> flags. We used StringTie's
337 functionality for *de novo* transcript assembly to identify candidate *Rsx* transcripts. An updated
338 GTF annotation was generated including novel transcripts using the --merge flag and then the
339 previously generated mapped reads were reassembled into transcripts guided by this GTF file.
340 DeSeq2 1.22.2 (48) was used to perform differential gene expression analysis between males and

341 females. NOISeq 2.26.1 (24) was used for differential expression analysis due to its ability to
342 simulate technical replicates within given RNA-seq data sets when no replicates are available.
343

344 Data accessibility

345 The raw and processed methylation datasets generated in this study have been deposited and
346 accessible through GEO Series accession number GSE149600 at
347 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE149600.
348

349 Competing interests

350 The authors declare that they have no competing interests.
351

352 Authors' contributions

353 D.S., D.A.P. and S.V.Y. formed the research design; D.A.P., A.G.K., D.E.A.P, and R.N.J.
354 generated the data; D.S., D.S. and S.V.Y. performed the analysis and wrote the initial draft ; all
355 authors contributed to and approved the final manuscript.
356

357 Funding statement

358 This work was supported by a grant from the National Science Foundation (MCB 1818288) and a
359 Pilot Grant from the Smooth Muscle Plasticity COBRE of the University of Nevada, Reno (funded
360 by the National Institutes of Health grant 5P30GM110767-04) to DAP, grants by the National
361 Science Foundation (MCB 1615664) and the National Institute of Health (R01MH103517) to
362 SVY. DS was partially supported by the NIH Training Grant in Computational Biology and
363 Biomedical Genomics (T32 GM105490).
364

365 References

366 1. Lyon MF. Gene Action in the X-chromosome of the Mouse (Mus musculus L.). Nature.
367 1961;190(4773):372-3.
368 2. Shevchenko AI, Zakharova IS, Zakian SM. The evolutionary pathway of x chromosome
369 inactivation in mammals. Acta Naturae. 2013;5(2):40-53.
370 3. Brown CJ, Hendrich BD, Rupert JL, Lafreniere RG, Xing Y, Lawrence J, et al. The human
371 XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is
372 highly localized within the nucleus. Cell. 1992;71(3):527-42.

373 4. Heard E, Clerc P, Avner P. X-chromosome inactivation in mammals. Annu Rev Genet.
374 1997;31:571-610.
375 5. Plath K, Mlynarczyk-Evans S, Nusinow DA, Panning B. Xist RNA and the mechanism of
376 X chromosome inactivation. Annu Rev Genet. 2002;36:233-78.
377 6. Duret L, Chureau C, Samain S, Weissenbach J, Avner P. The Xist RNA gene evolved in
378 eutherians by pseudogenization of a protein-coding gene. Science. 2006;312(5780):1653-5.
379 7. Ng K, Pullirsch D, Leeb M, Wutz A. Xist and the order of silencing. EMBO Rep.
380 2007;8(1):34-9.
381 8. Grant J, Mahadevaiah SK, Khil P, Sangrithi MN, Royo H, Duckworth J, et al. Rsx is a
382 metatherian RNA with Xist-like properties in X-chromosome inactivation. Nature. 2012;487:254.
383 9. Sharman GB. Late DNA Replication in the Paternally Derived X Chromosome of Female
384 Kangaroos. Nature. 1971;230(5291):231-2.
385 10. Wang X, Douglas KC, Vandeberg JL, Clark AG, Samollow PB. Chromosome-wide
386 profiling of X-chromosome inactivation and epigenetic states in fetal brain and placenta of the
387 opossum, Monodelphis domestica. Genome Res. 2014;24(1):70-83.
388 11. Huynh KD, Lee JT. Inheritance of a pre-inactivated paternal X chromosome in early mouse
389 embryos. Nature. 2003;426(6968):857-62.
390 12. Okamoto I, Otte AP, Allis CD, Reinberg D, Heard E. Epigenetic dynamics of imprinted X
391 inactivation during early mouse development. Science. 2004;303(5658):644-9.
392 13. Heard E. Delving into the diversity of facultative heterochromatin: the epigenetics of the
393 inactive X chromosome. Current opinion in genetics & development. 2005;15(5):482-9.
394 14. Koina E, Chaumeil J, Greaves IK, Tremethick DJ, Graves JA. Specific patterns of histone
395 marks accompany X chromosome inactivation in a marsupial. Chromosome research : an
396 international journal on the molecular, supramolecular and evolutionary aspects of chromosome
397 biology. 2009;17(1):115-26.
398 15. Ingles ED, Deakin JE. Global DNA Methylation patterns on marsupial and devil facial
399 tumour chromosomes. Mol Cytogenet. 2015;8:74-.
400 16. Rens W, Wallduck MS, Lovell FL, Ferguson-Smith MA, Ferguson-Smith AC. Epigenetic
401 modifications on X chromosomes in marsupial and monotreme mammals and implications for
402 evolution of dosage compensation. Proceedings of the National Academy of Sciences.
403 2010;107(41):17657.
404 17. Piper AA, Bennett AM, Noyce L, Swanton MK, Cooper DW. Isolation of a clone partially
405 encoding hill kangaroo X-linked hypoxanthine phosphoribosyltransferase: Sex differences in
406 methylation in the body of the gene. Somatic Cell and Molecular Genetics. 1993;19(2):141-59.
407 18. Loebel D, Johnston PG. Methylation analysis of a marsupial X-linked CpG island by
408 bisulfite genomic sequencing. Genome Res. 1996;6(2):114-23.
409 19. Waters SA, Livernois AM, Patel H, O'Meally D, Craig JM, Marshall Graves JA, et al.
410 Landscape of DNA Methylation on the Marsupial X. Molecular Biology and Evolution.
411 2018;35(2):431-9.
412 20. Sun Z, Cunningham J, Slager S, Kocher J-P. Base resolution methylome profiling:
413 considerations in platform selection, data preprocessing and analysis. Epigenomics.
414 2015;7(5):813-28.
415 21. Kasai F, O'Brien PCM, Pereira JC, Ferguson-Smith MA. Marsupial chromosome DNA
416 content and genome size assessed from flow karyotypes: invariable low autosomal GC content. R
417 Soc Open Sci. 2018;5(8):171539-.

418 22. Hansen KD, Langmead B, Irizarry RA. BSmooth: from whole genome bisulfite sequencing
419 reads to differentially methylated regions. Genome Biol. 2012;13(10):R83.
420 23. Hobbs M, Pavasovic A, King AG, Prentis PJ, Eldridge MDB, Chen Z, et al. A
421 transcriptome resource for the koala (Phascolarctos cinereus): insights into koala retrovirus
422 transcription and sequence diversity. BMC Genomics. 2014;15(1):786.
423 24. Tarazona S, Furio-Tari P, Turra D, Pietro AD, Nueda MJ, Ferrer A, et al. Data quality
424 aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. Nucleic acids
425 research. 2015;43(21):e140.
426 25. Johnson RN, O'Meally D, Chen Z, Etherington GJ, Ho SYW, Nash WJ, et al. Adaptation
427 and conservation insights from the koala genome. Nature Genetics. 2018;50(8):1102-11.
428 26. Hellman A, Chess A. Gene body-specific methylation on the active X chromosome.
429 Science. 2007;315:1141-3.
430 27. Keown CL, Berletch JB, Castanon R, Nery JR, Disteche CM, Ecker JR, et al. Allele-
431 specific non-CG DNA methylation marks domains of active chromatin in female mouse brain.
432 Proceedings of the National Academy of Sciences. 2017;114(14):E2882.
433 28. Sun D, Maney DL, Layman TS, Chatterjee P, Yi SV. Regional epigenetic differentiation
434 of the Z Chromosome between sexes in a female heterogametic system. Genome Res.
435 2019;29(10):1673-84.
436 29. Schultz MD, He Y, Whitaker JW, Hariharan M, Mukamel EA, Leung D, et al. Human body
437 epigenome maps reveal noncanonical DNA methylation variation. Nature. 2015;523(7559):212-
438 6.
439 30. Mendizabal I, Shi L, Keller TE, Konopka G, Preuss TM, Hsieh TF, et al. Comparative
440 Methylome Analyses Identify Epigenetic Regulatory Loci of Human Brain Evolution. Mol Biol
441 Evol. 2016;33(11):2947-59.
442 31. Zemach A, McDaniel IE, Silva P, Zilberman D. Genome-wide evolutionary analysis of
443 eukaryotic DNA methylation. Science. 2010;328(5980):916-9.
444 32. Yi SV. Birds do it, bees do it, worms and ciliates do it too: DNA methylation from
445 unexpected corners of the tree of life. Genome Biol. 2012;13(10):174.
446 33. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond.
447 Nature reviews Genetics. 2012;13(7):484-92.
448 34. Chuang T-J, Chen F-C, Chen Y-Z. Position-dependent correlations between DNA
449 methylation and the evolutionary rates of mammalian coding exons. Proceedings of the National
450 Academy of Sciences. 2012;109(39):15841.
451 35. Brenet F, Moh M, Funk P, Feierstein E, Viale AJ, Socci ND, et al. DNA methylation of
452 the first exon is tightly linked to transcriptional silencing. PloS one. 2011;6(1):e14524.
453 36. Anastasiadi D, Esteve-Codina A, Piferrer F. Consistent inverse correlation between DNA
454 methylation of the first intron and gene expression across tissues and species. Epigenetics &
455 Chromatin. 2018;11(1):37.
456 37. Neri F, Rapelli S, Krepelova A, Incarnato D, Parlato C, Basile G, et al. Intragenic DNA
457 methylation prevents spurious transcription initiation. Nature. 2017;543(7643):72-7.
458 38. Huh I, Zeng J, Park T, Yi SV. DNA methylation and transcriptional noise. Epigenetics &
459 chromatin. 2013;6(1):9-.
460 39. Cotton AM, Price EM, Jones MJ, Balaton BP, Kobor MS, Brown CJ. Landscape of DNA
461 methylation on the X chromosome reflects CpG density, functional chromatin state and X-
462 chromosome inactivation. Hum Mol Genet. 2015;24(6):1528-39.

463 40. Elango N, Kim S-H, Program NCS, Vigoda E, Yi SV. Mutations of different molecular
464 origins exhibit contrasting patterns of regional substitution rate variation. PLoS Computational
465 Biology. 2008;4(2):e1000015.
466 41. Alvarez-Ponce D, Torres-Sanchez M, Feyertag F, Kulkarni A, Nappi T. Molecular
467 evolution of DNMT1 in vertebrates: Duplications in marsupials followed by positive selection.
468 PloS one. 2018;13(4).
469 42. Sprague D, Waters SA, Kirk JM, Wang JR, Samollow PB, Waters PD, et al. Nonlinear
470 sequence similarity between the Xist and Rsx long noncoding RNAs suggests shared functions of
471 tandem repeat domains. RNA. 2019;25(8):1004-19.
472 43. Song Q, Decato B, Hong EE, Zhou M, Fang F, Qu J, et al. A Reference Methylome
473 Database and Analysis Pipeline to Facilitate Integrative and Comparative Epigenomics. PLOS
474 ONE. 2013;8(12):e81148.
475 44. Galili T. dendextend: an R package for visualizing, adjusting and comparing trees of
476 hierarchical clustering. Bioinformatics. 2015;31(22):3718-20.
477 45. Chen J, Bardes EE, Aronow BJ, Jegga AG. ToppGene Suite for gene list enrichment
478 analysis and candidate gene prioritization. Nucleic acids research. 2009;37(Web Server
479 issue):W305-11.
480 46. Zhang Z, Schwartz S, Wagner L, Miller W. A greedy algorithm for aligning DNA
481 sequences. Journal of computational biology : a journal of computational molecular cell biology.
482 2000;7(1-2):203-14.
483 47. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis
484 of RNA-seq experiments with HISAT, StringTie and Ballgown. Nature Protocols.
485 2016;11(9):1650-67.
486 48. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for
487 RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.
488 49. Ogawa Y, Kakumoto K, Yoshida T, Kuwako K-i, Miyazaki T, Yamaguchi J, et al. Elavl3
489 is essential for the maintenance of Purkinje neuron axons. Scientific Reports. 2018;8(1):2722.

490
491

492 **Figure Legends**

493 **Figure 1. Overview of DNA methylation patterns across the koala genome.** (A) Hierarchical
494 clustering of DNA methylation of five tissues. (B) Tissue-specific and shared differentially
495 methylation regions (DMRs) between tissues. (C) Enrichments of DMRs in each functional region
496 in comparison to length and GC matched control regions (***$p < 0.0001$, n.s. Not significant, from
497 10,000 bootstraps) from brain samples. Error bars depict standard deviation. (D-E) Examples of
498 two brain-specific DMRs (highlighted in grey) with corresponding CpG fractional methylation for
499 the two brain (red) and eight remaining tissues (blue). Line smoothing performed using local
500 regression (LOESS). (D) *ELAVL3* is a neural specific RNA-binding protein linked to the
501 maintenance of Purkinje neuron axons (49). A 1.84 kb region containing three CpG Islands (CGI)
502 and overlapping the first exon of *ELAVL3* was hypermethylated in the brain compared to all other
503 tissues, and this gene was up-regulated in the brain compared to the kidneys (probability of
504 differential expression > 98% from NOISeq). (D) *ALDOC* encodes a catalytic enzyme responsible
505 for the conversion of fructose-1,6-bisphosphate to glyceraldehyde 3-phosphate and
506 dihydroxyacetone phosphate. A 945 bp, brain-specific hypomethylated DMR overlaps the
507 promoter and part of *ALDOC*'s gene body. This gene was up-regulated in brain samples compared
508 to kidney samples (probability of differential expression > 96% from NOISeq).

509

510 **Figure 2. Global patterns of female and male DNA methylation (5mC) in eutherian and**
511 **marsupial X chromosomes.** (A) Comparison of chromosome-wide DNA methylation between
512 sexes in human and koala brains. (B) Distributions of the DNA methylation difference between
513 female and male koalas in autosomes and the X chromosome. (C) Comparing fractional DNA
514 methylation between males and females demonstrate that the female X chromosomes are
515 hypomethylated in both humans and koalas. (D-F) DNA methylation differences between males
516 and females of (D) the X chromosome and (E) autosomes of koalas, and (F) the X chromosome of
517 humans for all male (purple) and female (orange) samples. Line smoothing was performed using
518 local regression (LOESS). (G) Average fractional methylation of CpGs in 100-bp sliding windows
519 using a 10 bp step size in a 5 Kb region upstream and downstream of all chromosome X linked
520 gene's transcription start sites (TSSs) across koala tissues.

521

522 **Figure 3. Annotation of genomic of DNA methylation (5mC) around *Rsx*.** The top panel
523 identifies CpG islands (CGI), the middle panel reports the absolute male (purple) and female
524 (orange) fractional methylation at each CpG, and the bottom panel shows the female and male
525 fractional methylation difference. Highlighted in grey across all panels is the female
526 hypomethylated region upstream of the *Rsx* TSS.
527
528 **Figure 4. Model of DNA methylation (5mC) patterns across genomic functional regions for**
529 **eutherian mammals and marsupials.** Eutherian mammals exhibit increased DNA methylation
530 of promoter regions and CpG islands coupled with a relative depletion of DNA methylation across
531 gene bodies and intergenic regions on the inactive X chromosome (X_I) compared to the active X
532 chromosome (X_A) in females. Marsupial mammals share similar DNA methylation depletion in
533 gene bodies and intergenic regions of the inactive X chromosome; however, they diverge from
534 eutherian mammals in their promoter methylation patterns. Marsupial promoters are modestly
535 hypomethylated in the female X chromosomes (X_A and X_I) compared to the male X chromosome
536 (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.

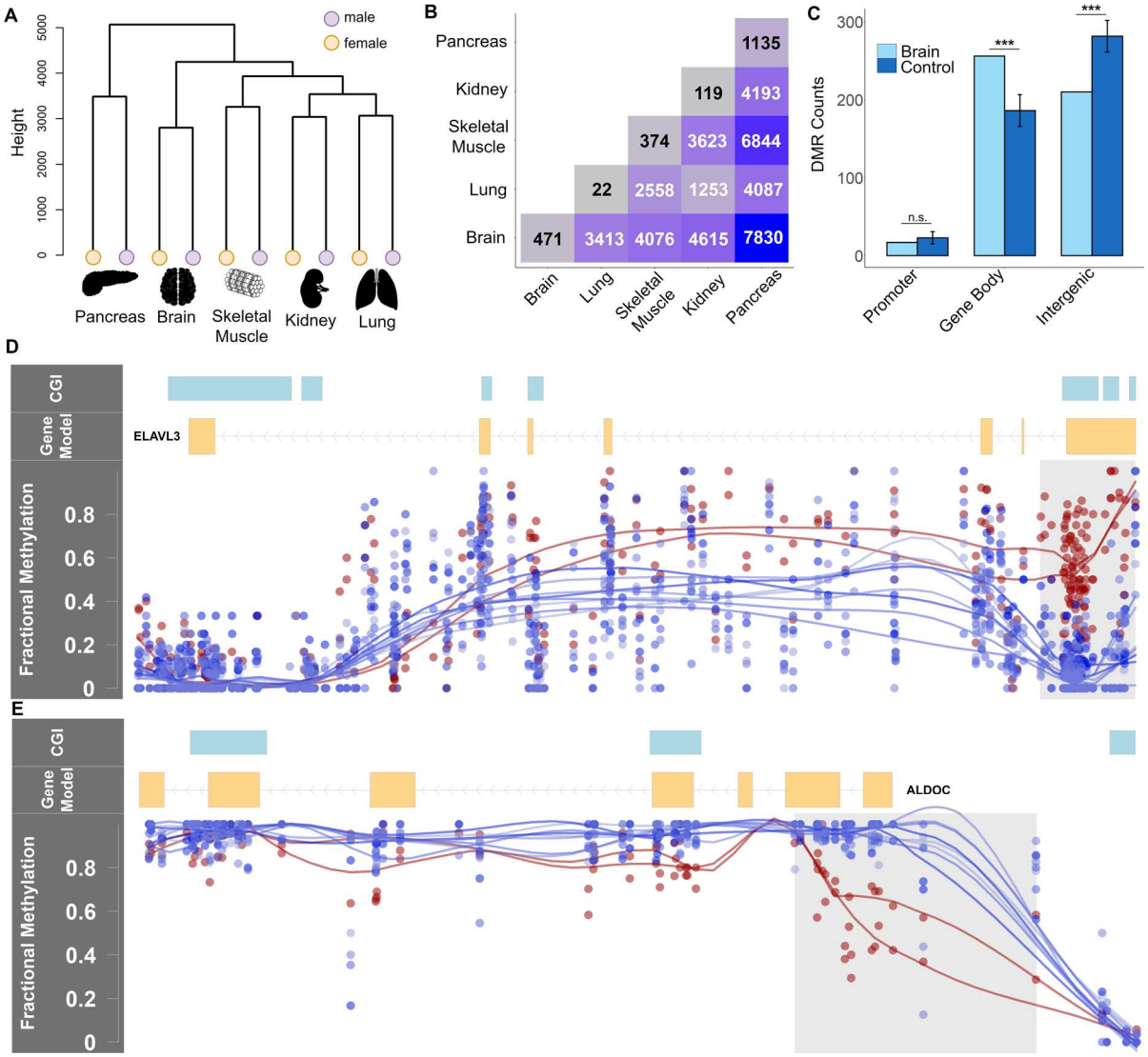


Figure 2.

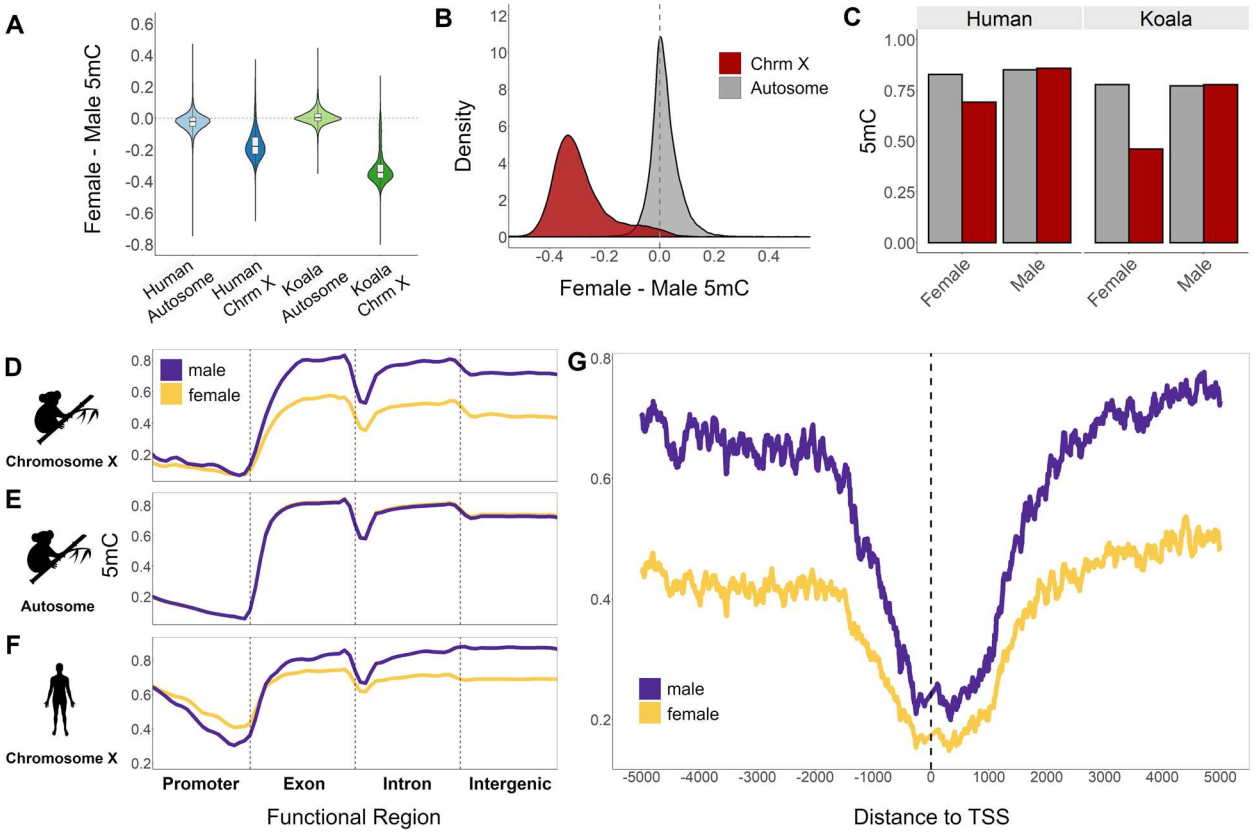

A
Female - Male 5mC
Human Autosome
Human Chrm X
Koala Autosome
Koala Chrm X
B
Density
Chrm X
Autosome
Female - Male 5mC
C
Human
Koala
5mC
Female
Male
Female
Male
D
male
female
Chromosome X
E
5mC
Autosome
F
Chromosome X
Promoter
Exon
Intron
Intergenic
Functional Region
G
male
female
Distance to TSS

Figure 3.

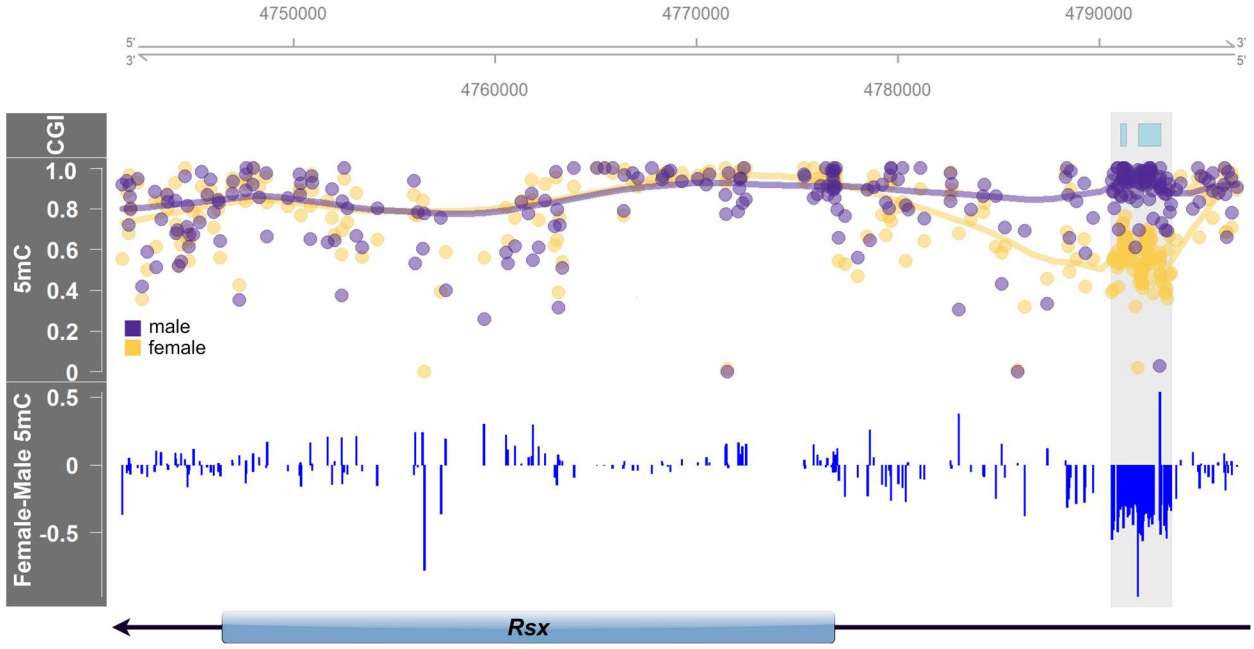
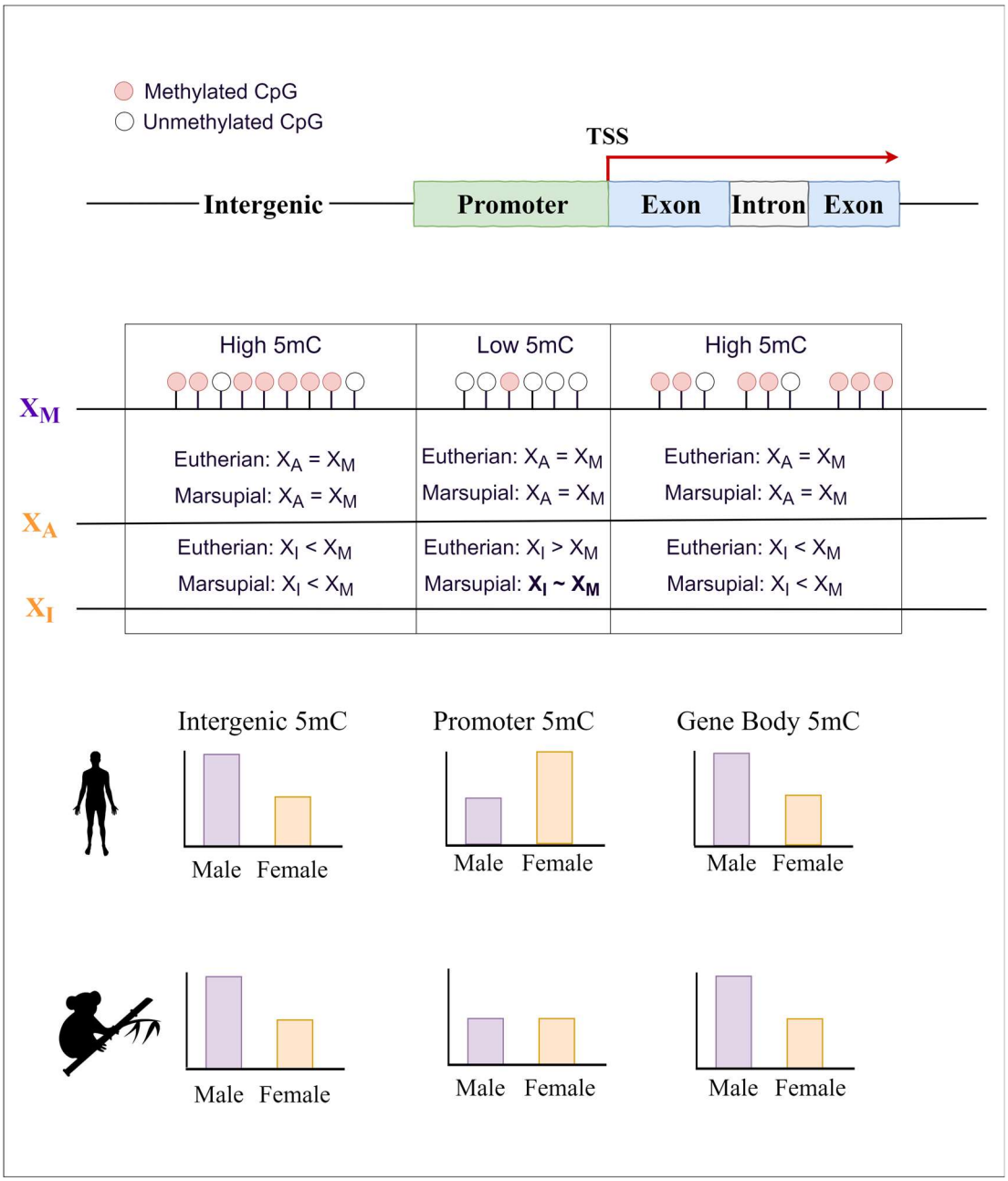

4750000
4770000
4790000
4760000
4780000
5'
3'
3'
5'
CGI
5mC
1.0
0.8
0.6
0.4
0.2
0
male
female
Female-Male 5mC
0.5
0
-0.5
Rsx

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: $X_A = X_M$
Marsupial: $X_A = X_M$
Eutherian: $X_A = X_M$
Marsupial: $X_A = X_M$
Eutherian: $X_I < X_M$
Marsupial: $X_I < X_M$
Eutherian: $X_I > X_M$
Marsupial: $X_I \sim X_M$
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.

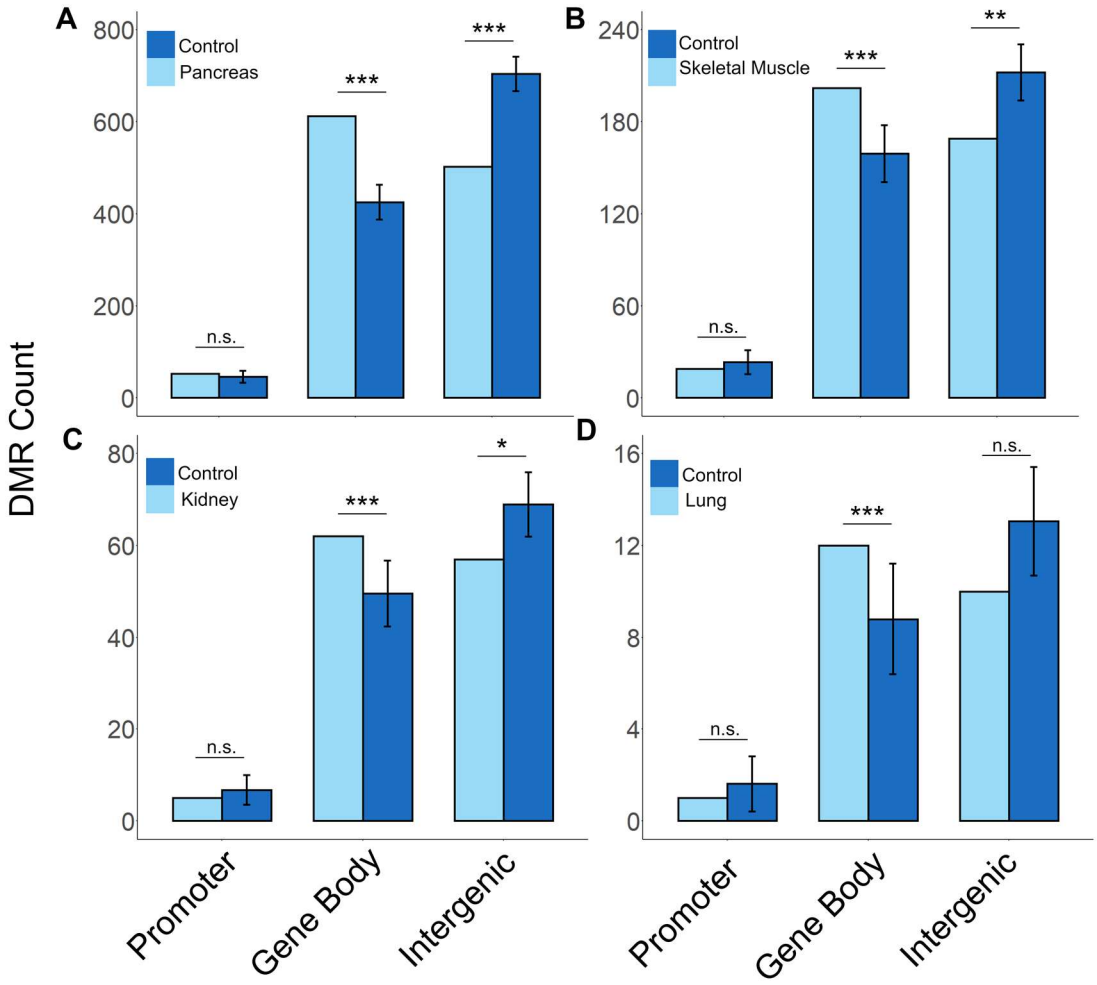
Supplementary Table 4. Mean and median sex-based DNA methylation difference calculated for all candidate X-scaffolds (n=98) by tissue.

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.

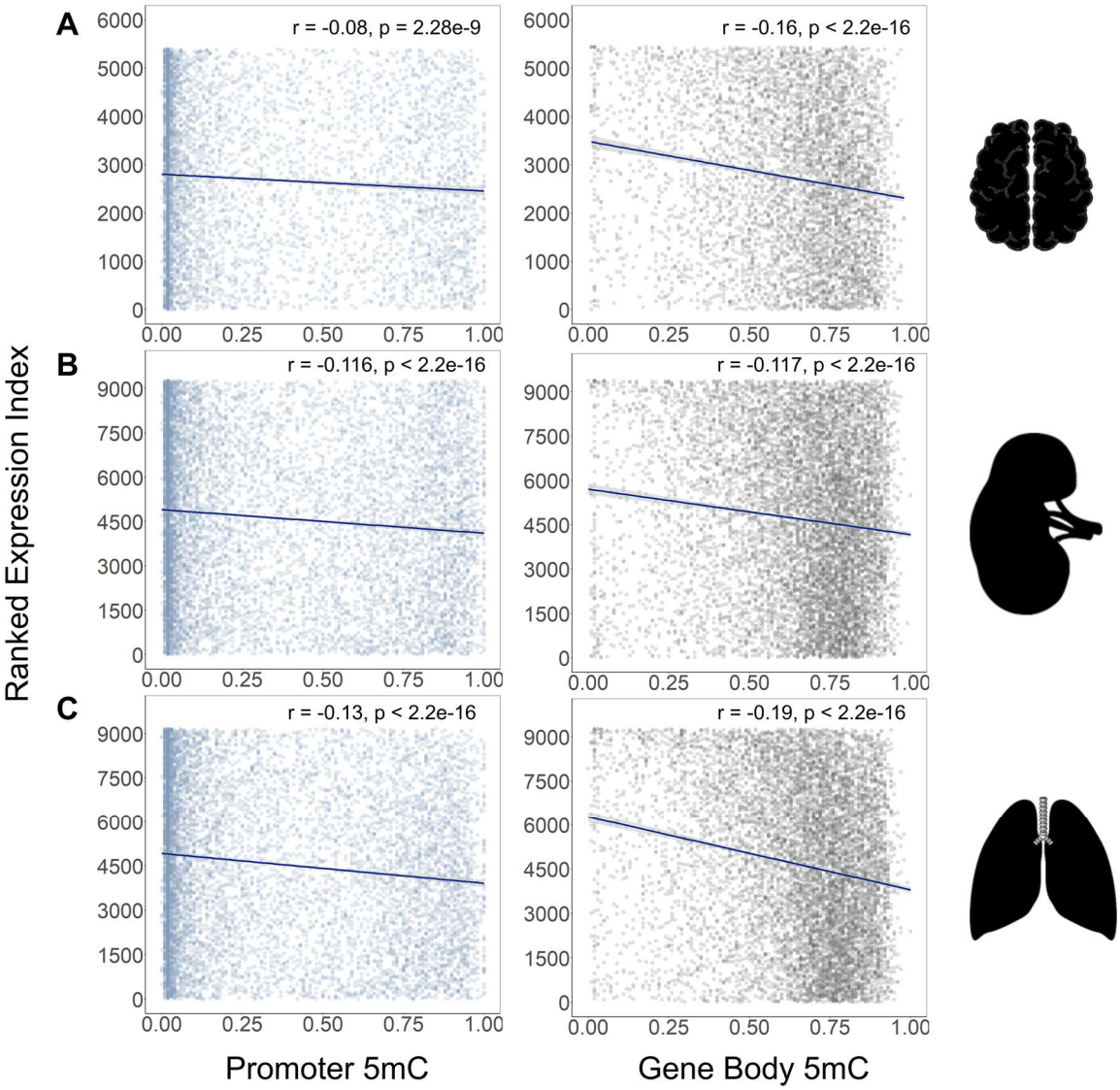
Supplementary Figures

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



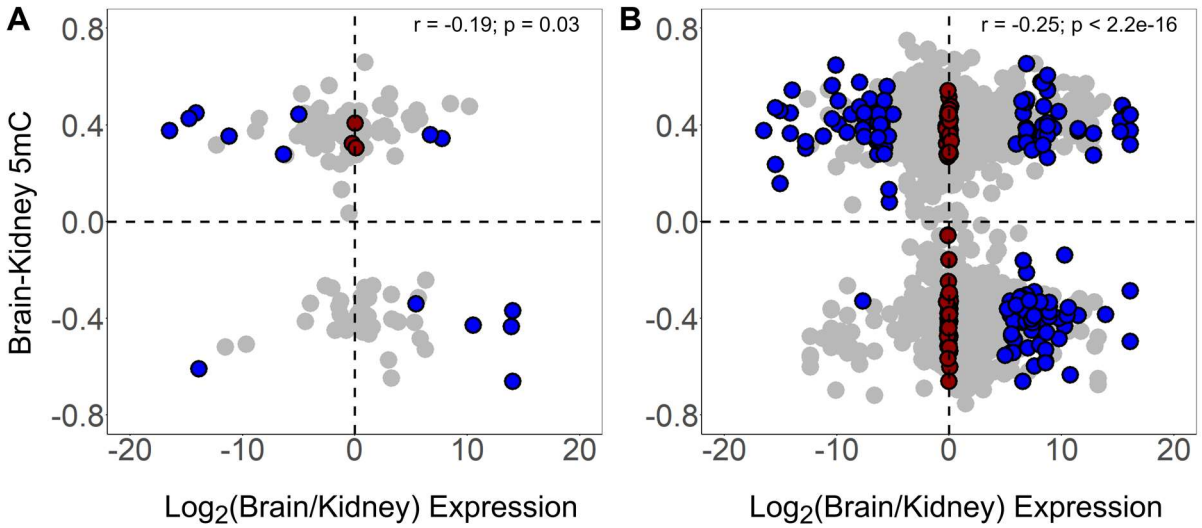
Data shown for (A) pancreas, (B) skeletal muscle, (C) kidney, and (D) lung. The enrichment of DMRs in each functional region (promoter, gene body, and intergenic regions) is shown through a comparison with length and GC matched control regions (*** indicates $p < 0.0001$, ** indicates $p < 0.001$, * 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.



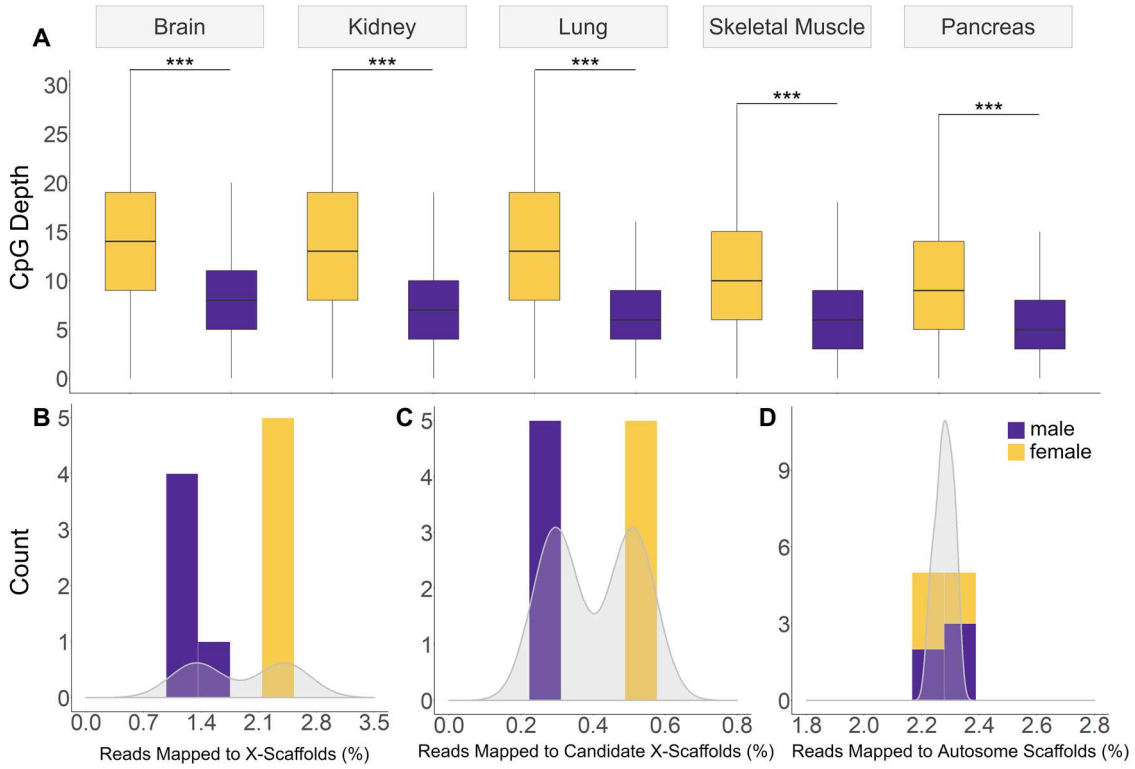
Three tissues with both whole genome bisulfite sequencing (WGBS) DNA methylation data and RNA-seq gene expression data are shown, (A) brain (n = 5,396 promoters and n = 5,443 gene bodies), (B) kidney (n = 9,268 promoters and n = 9,379 gene bodies), and (C) lung (n = 9,192 promoters and n = 9,265 gene 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.



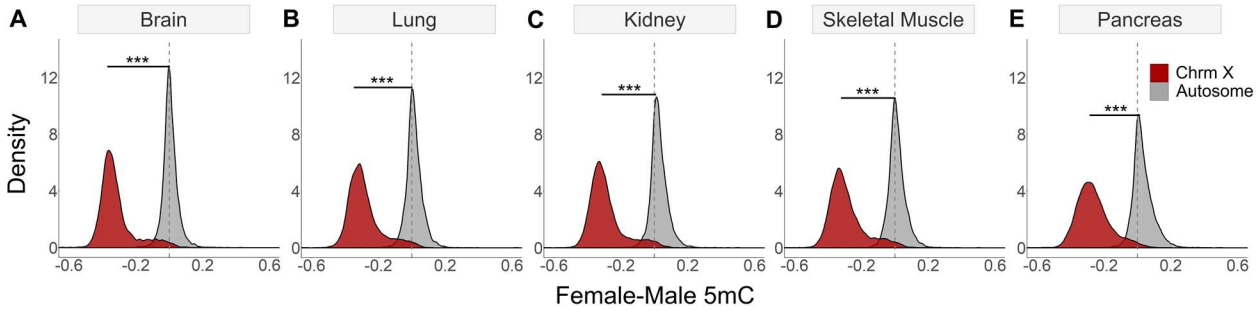
(A) The mean brain and kidney DNA methylation difference calculated for all CpGs across each gene promoter matched with the corresponding log-transformed ratio of brain to kidney expression. (B) The mean brain and kidney DNA methylation difference calculated for all CpGs across each gene body and matched 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.



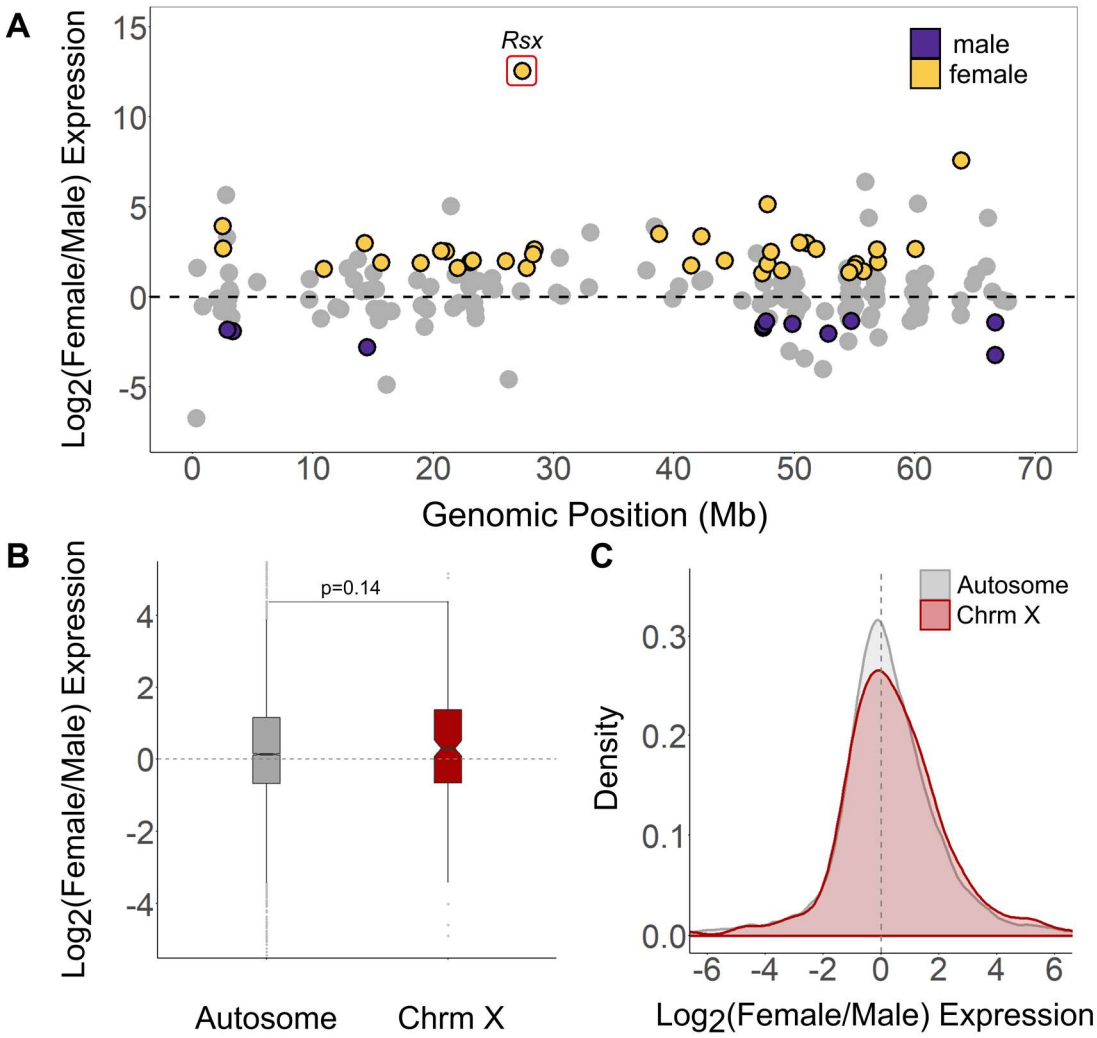
(A) For each of the five tissues, box-and-whisker plots of CpG depths across the X chromosome in male (purple) and female (orange) samples (*** indicates p < 2.2x10-16, Mann-Whitney U test). Histogram and distribution of sex-based read mapping per sample (n=10) to (B) X-linked scaffolds, (C) candidate 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.



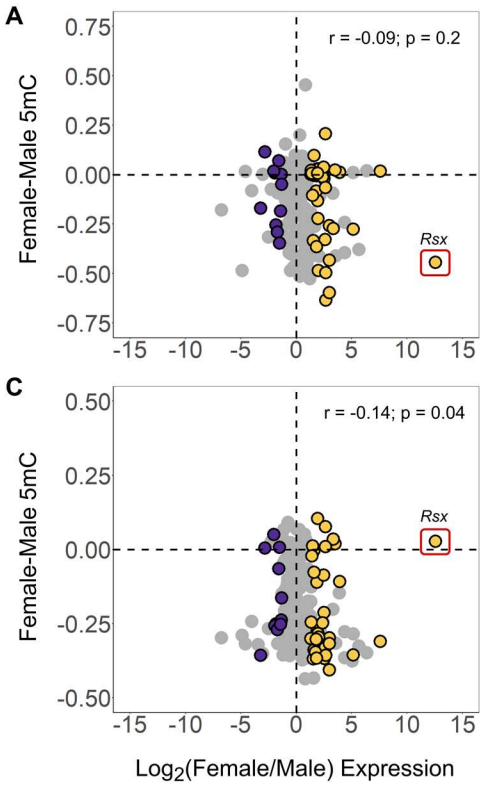
The distribution of female and male mean fractional methylation difference from (A) brain, (B) lung, (C) kidney, (D) skeletal muscle and (E) pancreas samples across autosomes and X chromosomes. For A-E, the female and male mean fractional methylation (methylated reads/total reads per CpG) was calculated 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-test).

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

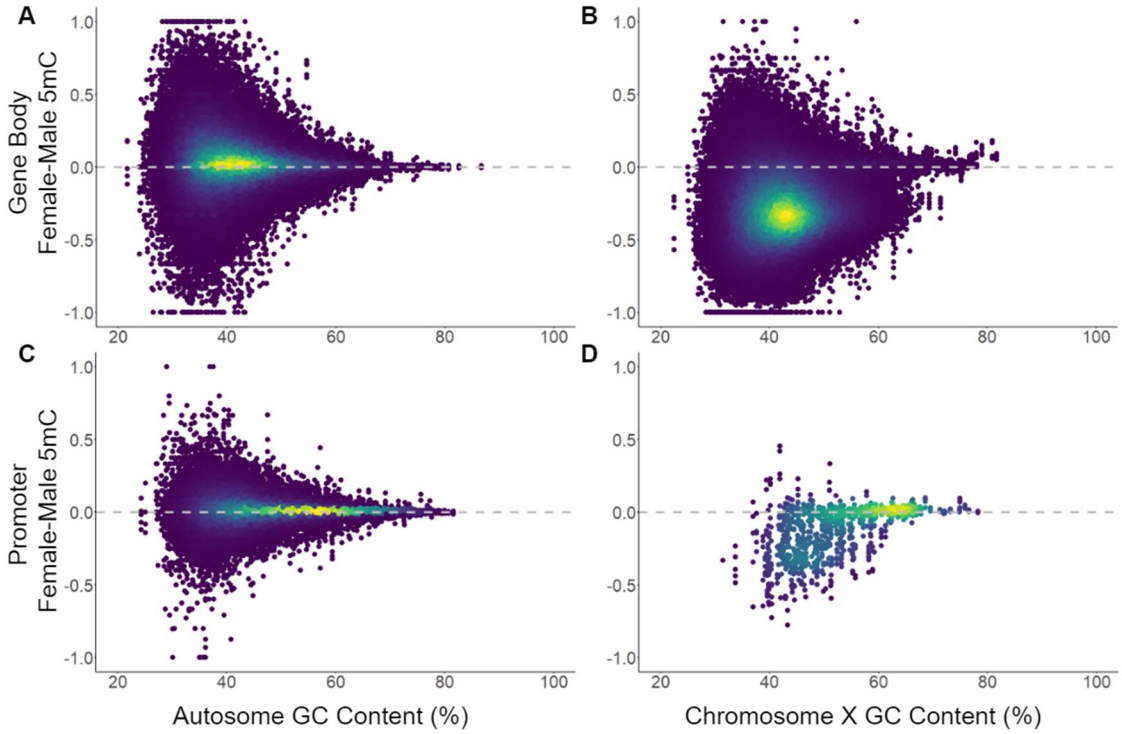
Promoter	Male Over Expressed / Total Genes	Female Over Expressed / Total Genes
Female Hypermethylated	$\frac{5}{32} = 0.16$	$\frac{14}{44} = 0.32$
Female Hypomethylated	$\frac{6}{57} = 0.11$	$\frac{22}{73} = 0.30$

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 8. 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.