

 Open access • Posted Content • DOI:10.1101/2021.08.17.456618

## Epigenetic mechanisms of partial dosage compensation in an avian, female heterogametic system — [Source link](#)

Ana Catalán, Justin Merondun, Ulrich Knief, Jochen B. W. Wolf

**Institutions:** Uppsala University

**Published on:** 19 Aug 2021 - bioRxiv (Cold Spring Harbor Laboratory)

**Topics:** Heterogametic sex, Dosage compensation, Z chromosome, Gene dosage and Autosome

Related papers:

- [Dosage compensation of the sex chromosomes](#)
- [Dosage Compensation Systems](#)
- [Dosage Compensation throughout the Schistosoma mansoni Lifecycle: Specific Chromatin Landscape of the Z Chromosome](#)
- [Untangling the Contributions of Sex-Specific Gene Regulation and X-Chromosome Dosage to Sex-Biased Gene Expression in Caenorhabditis elegans.](#)
- [Genomic imprinting mediates dosage compensation in a young plant XY system.](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/epigenetic-mechanisms-of-partial-dosage-compensation-in-an-42co3ua1ox>

## **Epigenetic mechanisms of partial dosage compensation in an avian, female heterogametic system**

Ana Catalán\*<sup>1,2</sup>, Justin Merondun<sup>2</sup>, Ulrich Knief<sup>2</sup> and Jochen B.W. Wolf<sup>1,2</sup>.

<sup>1</sup> Department of Evolutionary Biology, Evolutionary Biology Centre (EBC), Uppsala University, Norbyvägen 14-18 75236, Uppsala, Sweden.

<sup>2</sup> Division of Evolutionary Biology, LMU Munich, Grosshaderner Straße 2, Planegg-Martinsried 82152, Germany.

\*Corresponding author: [ana.catalan@gmail.com](mailto:ana.catalan@gmail.com)

## Abstract

The evolution of genetic sex determination is often accompanied by degradation of one of the proto sex chromosomes. Male heterogametic systems have evolved convergent, epigenetic mechanisms restoring the resulting imbalance in gene dosage between diploid autosomes (AA) and the hemizygous sex chromosome (X). Female heterogametic systems (AA<sub>f</sub> ZW<sub>f</sub>, AA<sub>m</sub> ZZ<sub>m</sub>) tend to only show partial dosage compensation ( $0.5 < Z_f:AA_f < 1$ ) and dosage balance ( $0.5 < Z_f:ZZ_m < 1$ ). The underlying mechanism remains largely elusive. Here, we quantified gene expression for a total of 15 male and female Eurasian crows (*Corvus (corone) spp.*) raised under common garden conditions. In addition, we characterized aspects of the regulatory landscape quantifying genome-wide ATAC-seq and 5mC methylation profiles. Partial dosage compensation was explained by female upregulation of Z-linked genes accompanied by increased chromatin accessibility on the female Z chromosome. 5mC methylation was strongly reduced in open chromatin-regions and GC islands and showed chromosome-, but no sex-specific variation. With the exception of the pseudo-autosomal region (PAR), female upregulation of gene expression was evenly spread across the Z chromosome without evidence for regional epigenetic regulation, as has for example been suggested for the male hypermethylated region (MHM) in chicken. Our results support the hypothesis that partial dosage compensation in female heterogametic systems is subject to chromosome-wide, epigenetic control mediated by differential chromatin accessibility between the sexes.

## Introduction

The evolution of chromosomal sex-determination appears to be accompanied by degradation of one of the proto sex chromosomes resulting in male (m) or female (f) heterogameity (Bachtrog, 2013; Graves, 2006). In the former case, the set of heteromorphic sex chromosomes resides in males (m: XY; f: XX), in the latter case in females (f: ZW; m: ZZ). In both systems, the heterogametic sex is left with only half of the original copy number of orthologues genes on the shared sex chromosome (X or Z, respectively). In the absence of compensatory mechanisms, this change in gene dosage cuts ancestral transcript levels in half (Gelbart and Kuroda, 2009; Graves, 2016; Schield et al., 2019) and induces an imbalance in gene regulatory networks with genes from diploid autosomes ( $X_m:AA_m = Z_f:AA_f = 0.5$ ) (Bachtrog, 2013; Charlesworth, 1996; Mank, 2009). In the homogametic sex, the sex-chromosomal to autosomal gene dosage ( $XX_f:AA_f$ ,  $ZZ_m:AA_m$ ) remains, in principle, unaffected (Vicoso and Bachtrog, 2011).

Chromosome-level haploinsufficiency, and even dosage change in single genes, will alter stoichiometric ratios of interacting proteins, generally with strong deleterious fitness effects (Birchler et al., 2005; Rice and McLysaght, 2017). A single-copy sex chromosome thus constitutes a serious peril for the hemizygous sex and is expected to prompt an evolutionary response (Charlesworth, 1996; Gu and Walters, 2017; Ohno, 1967). Indeed, male heterogametic systems as divergent as fruit flies and mammals have independently evolved molecular mechanisms that function to restore expression balance (Livernois et al., 2012; Samata and Akhtar, 2018). While the degree of compensation and the underlying mechanisms differ across taxa (Gu and Walters, 2017; Mank, 2009), epigenetic regulation and changes in chromatin structure are common themes (Duncan et al., 2018; Klemm et al., 2019; Park and Kuroda, 2001). For example, in mammals one of the X chromosomes is randomly targeted for inactivation in the homogametic sex by nucleosomal condensation impeding transcription (Graves, 2014). The active X chromosome is then hyper-transcribed to equalize expression levels to the (diploid) autosomal dose (Nguyen and Disteché, 2006). In *Drosophila*, in contrast, hyperactivation is limited to heterogametic males where relaxation of the chromatin structure

on the X restores expression levels to the default double-dosage (Samata and Akhtar, 2018). In *Caenorhabditis elegans* a hitherto unknown mechanism upregulates expression levels in both sexes. While this restores dosage in heterogametic males, expression overshoots in females and requires a secondary mechanism to return gene expression on the X to parity with the autosomes (McDonel et al., 2006). X-specificity is achieved by DNA motifs guiding the dosage compensation complex (Ercan et al., 2007; Jans et al., 2009). A common feature of the abovementioned XY systems is that dosage compensation appears to be initiated at specific sites along the X chromosome (Ercan et al., 2007; Drosophila: Gelbart and Kuroda, 2009; mice: Lin et al., 2007).

Global compensatory mechanisms restoring hemizygous to autosomal balance are not ubiquitous. In heterogametic ZW systems, females tend to only show partial compensation, a discrepancy that may follow from basic evolutionary principles (Naurin et al., 2010). This lack of full dosage compensation is observed across a diverse set of female heterogametic taxa including lepidopterans (Catalán et al., 2018; Gu et al., 2017; Zha et al., 2009), snakes (Schield et al., 2019; Vicoso et al., 2013a) and birds (Itoh et al., 2010, 2007) – the focal taxon of this study for which most information is available. The basal avian group of ratites, where sex chromosomes still recombine and degradation has been trapped at an arrested stage, still show Z:A parity across the largely homomorphic sex chromosomes (Adolfsson and Ellegren, 2013; Vicoso et al., 2013b). Other species in the avian tree of life differ in the degree of morphological and genomic divergence of the Z-chromosome and the size of the diploid pseudo-autosomal region (Rutkowska et al., 2012; Zhou et al., 2014). Yet, all species with heteromorphic sex chromosomes investigated to date show only partial compensation for dosage differences.

Following (Gu et al., 2017), we differentiate between dosage balance and dosage compensation. The former refers to parity in expression between male and female genes located on the shared sex chromosome ( $X_m:XX_{ff}$  or  $Z_f:ZZ_m = 1$ ). The latter refers to expression parity between sex-chromosomes of either sex and the ancestral, non-degenerated proto sex chromosome. In the absence of ancestral information, autosomal levels are generally used as a proxy ( $X_m:AA_m = XX_f:AA_{ff} = 1$  or  $Z_f:AA_f = ZZ_m:AA_m = 1$ ). This proxy seems justified under the conditions of equal expression for autosomal genes between the sexes ( $AA_f:AA_m = 1$ ) and lower intra-autosomal variance relative to variance between autosomes and the sex chromosome. The heterogametic bird species investigated so far tend to show both partial dosage balance ( $0.5 < Z_f:ZZ_m < 1$ ) and partial dosage compensation ( $0.5 < Z_f:AA_f < 1$ ). Depending on the species and tissue, f:m expression ratios of Z-linked genes range from 0.67 – 0.81 (Ellegren et al., 2007; Itoh et al., 2010, 2007; Uebbing et al., 2013; Wolf and Bryk, 2011). Z:A expression ratios similarly range from 0.5 – 0.7 in females, but reach equity in homogametic males (Ellegren et al., 2007; Uebbing et al., 2013; Wolf and Bryk, 2011).

This range between 0.5 (pure dosage effect) and 1 (full compensation) requires a mechanism conveying partial compensation and balance. The nature of this mechanism, however, remains elusive. In some species, a specialized region of greater dosage compensation near the male hypermethylated locus (MHM) would suggest localized epigenetic mechanisms on the Z chromosome analogous to XY systems (Arnold et al., 2008; Melamed and Arnold, 2007; Sun et al., 2019). Revision of the MHM in chicken using chromosome-wide DNA-methylation tiling arrays, however, did not confirm a clear relationship between sex-specific methylation and regional dosage balance (Nätt et al., 2014). Moreover, in the majority of investigated species patterns of f:m gene expression appear to be homogenous along the Z-chromosome and do not call for a regional epigenetic mechanism (Itoh et al., 2010; Sun et al., 2019; Wolf and Bryk, 2011). Thus, it has been hypothesized that partial dosage compensation

in birds is most likely regulated on a gene-by-gene basis (Mank and Ellegren, 2009), possibly in combination with a chromosome-wide epigenetic mechanism (this study).

The genomic tools to quantify gene expression, assay epigenetic profiles and quantify chromatin accessibility are readily available (Jordan et al., 2019). Yet, comprehensive studies investigating epigenetic mechanisms and chromatin structure in birds are rare (Fishman et al., 2019; Xu and Millar, 2020) and in the context of dosage compensation near-absent (Nätt et al., 2014; Sun et al., 2019). In this study, we quantified gene expression in two somatic tissues, spleen and liver, for a total of 15 male and female Eurasian crows (*Corvus (corone) spp.*) raised under common garden conditions. In addition, we characterized the genome-wide chromatin accessibility profile by means of ATAC-seq and generated methylation data to assay chromosomal 5mC profiles similarly known to affect gene expression (Lee et al., 2020; Lindner et al., 2021). Linking these three data types allowed us to gain insight into possible epigenetic mechanisms underlying partial dosage compensation.

## Results

### Chromosome-wide patterns of gene expression, chromatin accessibility and 5mC methylation

The interrogation of open chromatin states and 5mC methylation levels are expected to covary with a set of genomic features such as chromosome length, gene length and GC content, known to differ widely among avian chromosomes (O'Connor et al., 2019). Therefore, categorical comparisons between the long Z chromosome (75 Mb) and all autosomes including microchromosomes (4.88–154.33 Mb; RefSeq GCF\_000738735.5) are only meaningful if these factors are statistically controlled for. Accordingly, subsequent analyses focusing on differences between autosomes and the Z chromosome included these features as constitutive covariates where appropriate. Our measures of interest are also expected to vary in different parts of the chromosome (e.g. genic, intergenic). We therefore additionally considered different scales of integration: entire chromosomes (chromosome-wide), actively expressed genes including +/- 20kb of putative regulatory regions (genes) and flanking regulatory regions without the gene body (regulatory). Statistical results for all measures are summarized across these levels in a **Supplementary Model** file. **Supplementary Table A** summarizes all f:m and Z:A ratios ( $AA_f:AA_m$ ,  $Z_f:ZZ_m$ ,  $Z_f:AA_f$ ,  $ZZ_m:AA_m$ ), both as raw estimates and parameter estimates with confidence intervals controlling for covariates. For chromosome comparisons, the Z chromosome was considered as a whole including the pseudo-autosomal region (PAR). Effects of the PAR were then isolated in a second step.

### Gene expression

RNA-seq data from liver and spleen confirmed partial dosage compensation and dosage balance, as previously demonstrated for brain tissue in the same species (Wolf and Bryk, 2011). In both liver and spleen tissue, expression levels were statistically undistinguishable between autosomal genes of both sexes and genes located on the male Z chromosome ( $AA_m \cong ZZ_m \cong AA_f \cong 1$ ) (**Figure AA, Supplementary Model A**). Expression of Z-linked genes in females was significantly reduced relative to males (parameter estimate  $Z_f:ZZ_m = 0.63/0.62$  for liver/spleen respectively; same nomenclature hereafter) and relative to genes on female autosomes ( $Z_f:AA_f = 0.68/0.68$ ). In both cases, confidence intervals of expression ratios were between 0.5 and 1 confirming the hypothesis of partial dosage compensation through female upregulation of Z-linked genes.

## Chromatin structure

We hypothesized that upregulation on the female's Z chromosome might be under epigenetic control. We therefore performed an ATAC-seq experiment which provides information about nucleosome spacing and chromatin accessibility along the genome (Buenrostro et al., 2013). The density of ATAC-seq signals is expected to scale positively with the accessible proportion of DNA. Accordingly, peak density was taken to reflect global accessibility of a given chromosomal region. The height of individual peaks was used as a measure reflecting local accessibility of putative regulatory regions (Chen et al., 2016). We present the results of both measures in turn.

### Peak density

Analogous to gene expression, chromosome-wide autosomal peak density was statistically undistinguishable between the sexes. On the Z-chromosome, however, peak density was significantly reduced in females relative to males ( $Z_f:ZZ_m = 0.66/0.62$ ). Variation among autosomes was high and raw Z:A ratios were broad, yet overall with significantly lower values in females ( $Z_f:AA_f = 0.66/0.60$ ,  $ZZ_m:AA_m = 0.95/1.02$ , **Supplementary Model B1, Supplementary Figure SE**). To establish the link between gene expression patterns and chromatin accessibility more closely, we next restricted our analysis to peaks located in and around genes. Setting a threshold of 20 kb up- and downstream of the gene model, we scored a median of 6–7 linked peaks per expressed gene (**Supplementary Figure SF**). Gene-centered peak density closely recapitulated the chromosome-wide pattern (**Figure AB, Supplementary Model B2, Supplementary Figure G**). Peak density on the single Z copy in females reached about  $\frac{3}{4}$  of the density of autosomes ( $Z_f:AA_f = 0.77/0.73$ ,  $ZZ_m:AA_m = 0.93/0.97$ ) and of the male Z chromosome ( $Z_f:ZZ_m = 0.78/0.61$ ) suggesting a non-linear relationship between the number of accessible chromosome regions and ploidy. The pattern also remained stable when excluding gene bodies considering only putative regulatory regions up- and downstream of the expressed gene (**Supplementary Model B3**). In summary, peak density closely resembled the partial compensation pattern of gene expression across all scales of integration.

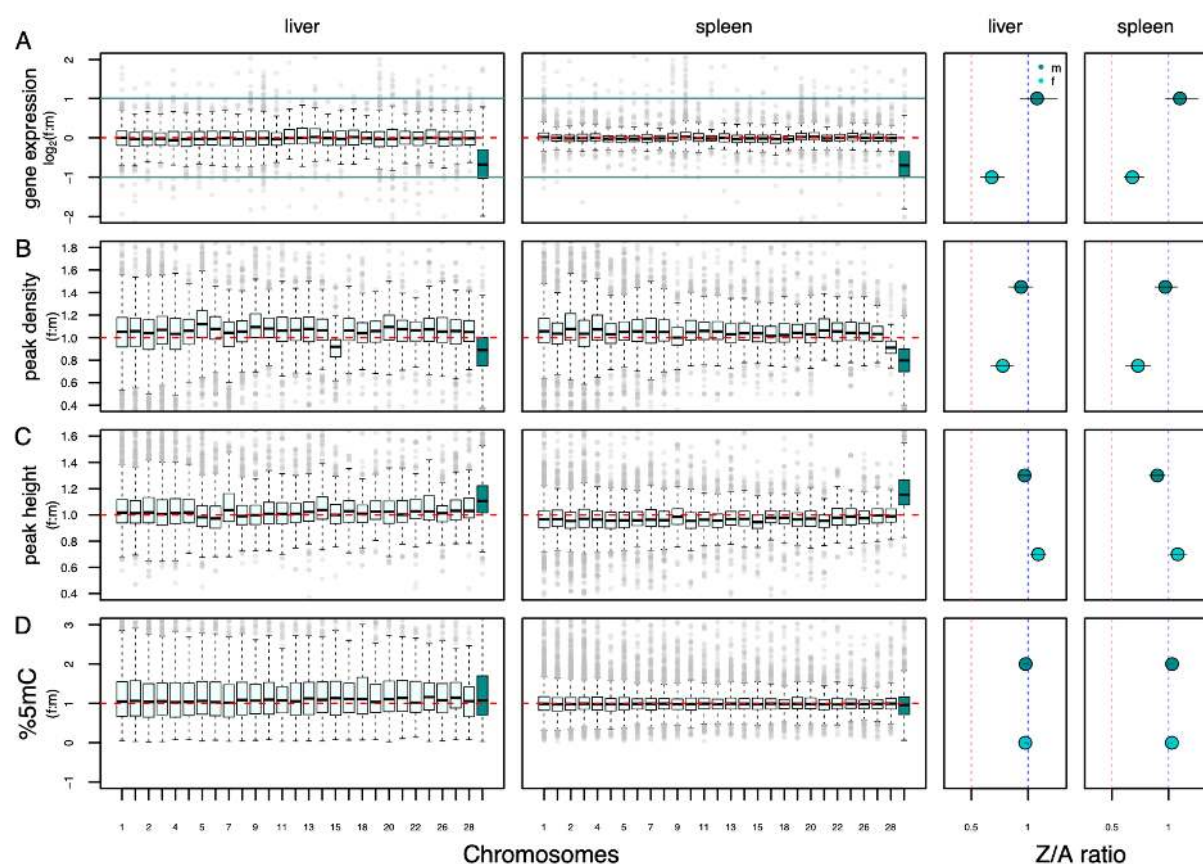
### Peak height

The average height of ATAC-seq peaks identified across autosomes did not differ between the sexes and was undistinguishable from levels on the male Z chromosome. Peak height in females, however, was significantly elevated on the Z when compared to males ( $Z_f:ZZ_m = 1.10/1.16$ ) and when compared to the autosomes ( $Z_f:AA_f = 1.12/1.06$ , **Supplementary Figure SE, Supplementary Model C1**). These differences in peak height were stable when considering only peaks associated with actively expressed genes (**Figure AC, Supplementary Model C2**) and peaks residing in putative regulatory regions up- and downstream of the expressed genes (**Supplementary Model C3, Supplementary Figure G**). This effect was not merely due to few outlier regions with extreme values of peak height, but was consistent across the peak height distribution. In all quantiles, peak height exceeded autosomal values (**Supplementary Figure SH**). The effect was also stable across varying intervals around genes including 1–15kb up- and downstream of the gene model (**Supplementary Figure SI**). Moreover, higher accessibility in females was also observed at the level of ATAC-seq read coverage excluding possible artifacts of the peak calling algorithm. As expected, f:m coverage ratio was centered around 1 for autosomes ( $AA_f:AA_m \approx 1.03/1.03$ ), but was elevated on the female Z beyond the expectation from chromosomal copy number of 0.5 ( $Z_f:ZZ_m = 0.65/0.65$ , **Supplementary Model C4**). Overall, this observation suggests that open chromatin is most accessible on the hemizygous, female Z chromosome.



## 5mC methylation

DNA methylation constitutes another candidate regulatory mechanism underlying partial dosage compensation. To assess its role, we generated whole-genome bisulfite sequencing data of one female and one male individual for both liver and spleen. First, we investigated the relationship between 5mC methylation with gene expression and chromatin accessibility. Analogous to the chromatin features described above, we assayed 5mC methylation at different scales: chromosome-wide, gene-centered (gene body +/- 20 kb flanking region), putatively regulatory regions in CpG islands and overlapping open chromatin regions of genes (ATAC-seq peaks), both including and excluding the gene body (**Supplementary Figure SJ**). As expected, methylation levels were high chromosome-wide (proportion of global methylation: 0.56/0.58) and around genes (0.63/0.63), but much reduced (and bimodal) in CpG islands (0.05/0.06) and in open chromatin regions including (0.02/0.07) or excluding the gene body (0.02/0.03) (**Supplementary Figure SK**). The Z chromosome as a whole exhibited slight hypermethylation (0.60/0.60) relative to macrochromosomes of similar size (chromosomes 1–21; 0.56/0.57), with minor hypermethylation observed in spleen across microchromosomes (chromosomes 22–28; 0.59/0.62). Important in the context of dosage compensation and dosage balance, hypermethylation on the Z was near-identical between the sexes (**Supplementary Models D1**). When restricting the analyses to genic regions, there was still no difference between the two sexes, but hypermethylation was no longer visible (**Supplementary Models D2**). In CpG islands, the Z chromosome was strongly hypermethylated in liver indicating either an excess of hypermethylated islands on chromosome Z or a reduction of hypermethylated islands on the autosomes (**Supplementary Figure SK**). Male to female methylation levels remained at parity (**Supplementary Model D3**). In open chromatin regions, there was no difference between the sexes, only among chromosomes. Open-chromatin regions in Z-linked genes appeared to be hypermethylated in liver and hypomethylated in spleen for both sexes (**Supplementary Figure SL, Supplementary Models D4 & D5**). Overall, these results indicate a tissue- and chromosome-specific 5mC methylation landscape in crows. They provided no evidence for a possible role of 5mC methylation contributing to dosage compensation or dosage balance. Yet, the fact that 5mC methylation levels were correlated with open-chromatin features across sexes and chromosomes (**Supplementary Figure SM**) allows for a possible indirect contribution.



**Figure A. Gene-centered patterns of dosage effects for gene expression, chromatin accessibility and 5mC methylation.** The two leftmost columns show f:m ratios by chromosome for liver and spleen with the Z chromosome highlighted in color. Horizontal red dashed lines indicate an equal f:m ratio. Boxes encompass interquartile range across genes, whiskers extend to 1.5 x the interquartile range. The two rightmost columns show Z-chromosome to autosomal (Z:A) ratios separately for females ( $Z_f:AA_{ff}$ , light shading) and males ( $ZZ_m:AA_m$ , dark shading). Here, red vertical lines mark a ratio of  $Z/A=0.5$  and blue vertical lines mark a ratio of  $Z/A = 1$ . Bars in dotplots represent 95% confidence intervals of parameter estimates of statistical models controlling for co-variates (**Supplementary Models**) drawn from 1,000 bootstraps. **(A)**  $\log_2(f:m)$  gene expression. **(B)** Female to male (f:m) ATAC-seq peak density. **(C)** Female to male (f:m) ATAC-seq peak height. **(D)** Female to male (f:m) methylation percentages.

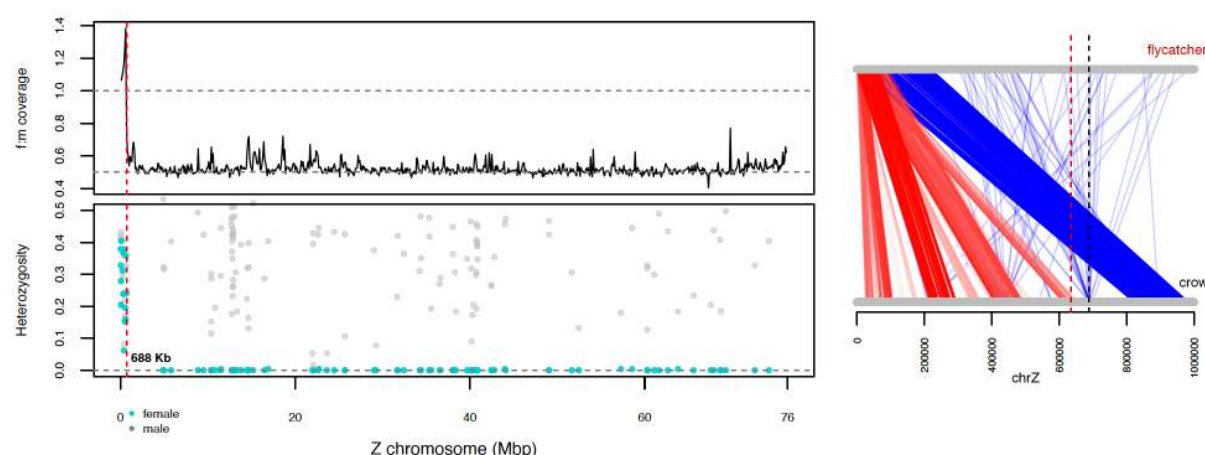
## Patterns of gene expression, chromatin accessibility and 5mC methylation along the Z chromosome

### Identification of the pseudo-autosomal region (PAR)

Next, we examined the distribution of dosage effects along the Z chromosome. In a first step, we identified the pseudo-autosomal region of the crow genome for use as a “diploid control”. Combining information from f:m differences in sequencing coverage, heterozygosity levels along the Z chromosome and orthology with other avian species, we defined a PAR region of ~688 kb from the proximal end of the Z chromosome (**Figure C, Supplementary Figure SN**). In the PAR, we identified 36 annotated genes, 14 of which were expressed in liver and 16 in spleen (for a .gff file see **Supplementary Table SB**). We then calculated f:m ratios inside (PAR) and outside the PAR (nonPAR) for gene expression, peak density and peak height.



Analogous to the chromosome comparisons above, statistical results are summarized in the **Supplementary Model** file for genes and putative regulatory regions. Note that due to the low sample sizes of genes on the PAR, standard errors were large and power was low. **Supplementary Table C** summarizes all PAR:nonPAR ratios between females and males both as raw estimates and parameter estimates with confidence intervals controlling for covariates.



**Figure B.** Identification of the pseudo-autosomal region (PAR) in the European crow. Left upper panel: f:m genome coverage along the Z chromosome. Left bottom panel: heterozygosity levels across the Z chromosome in females (light blue) and males (gray). A drop in female coverage and heterozygosity at 688 kb defines the end of the PAR. Right panel: 1 Mb of Z chromosome alignment between the flycatcher and the European crow. Lines between the Z represent orthologous regions found between the two species. Red: protein coding genes. Blue: non-coding regions. The vertical dashed line marks the end of the PAR region in crows. Vertical dashed red line marks the PAR as identified in the flycatcher (Smeds et al., 2014).

### Gene expression

As expected by gene dose, all genes expressed in the PAR had equal expression levels in females and males in both tissues ( $PAR_f:PAR_m=1.01/1.01$ , **Supplementary Figure SO**). Genes located in the remaining part of the Z chromosome (nonPAR) showed the expected pattern of reduced expression in females ( $nonPAR_f:PAR_f=0.55/0.78$ , **Supplementary Model E1**). For the non-PAR part of the Z chromosome we further categorized genes into fully compensated ( $-0.25 < Z_f:ZZ_m < 1$ ), partially compensated ( $-0.25 > Z_f:ZZ_m > -0.75$ ) and uncompensated ( $-0.75 > Z_f:ZZ_m > -1$ ). The number of genes falling into these categories was similar in liver and spleen. For each tissue, 147/159 genes (28.1/29.2%) were fully compensated, 99/122 genes (19.9/18.3%) were partially compensated and the majority of 277/350 genes (52.9/52.5 %) showed no evidence of compensation according to this definition. The number of genes expressed in both tissues was not enriched for compensation status (Fisher's exact test  $P$ -value: 0.43) indicating no preference for compensation of housekeeping genes.

Subsequent sliding window analysis along the Z chromosome with a window size of 688 kb (PAR regions equivalent) or 1 Mb and a step size of 344 kb or 100 kb, respectively, revealed no clustering of compensated genes (**Figure CA**). Not a single window was enriched for compensated genes (Fisher's exact tests with multiple testing correction, **Supplementary Table SD**). The same result was obtained when the enrichment analysis was done using windows containing 15 genes (mimicking the number of expressed genes on the PAR; **Supplementary Table SD**). A lack of regional concentration of compensated genes was further corroborated by a lack of autocorrelation in  $Z_f:ZZ_m$  values (**Supplementary Figure SP**).

### Peak density

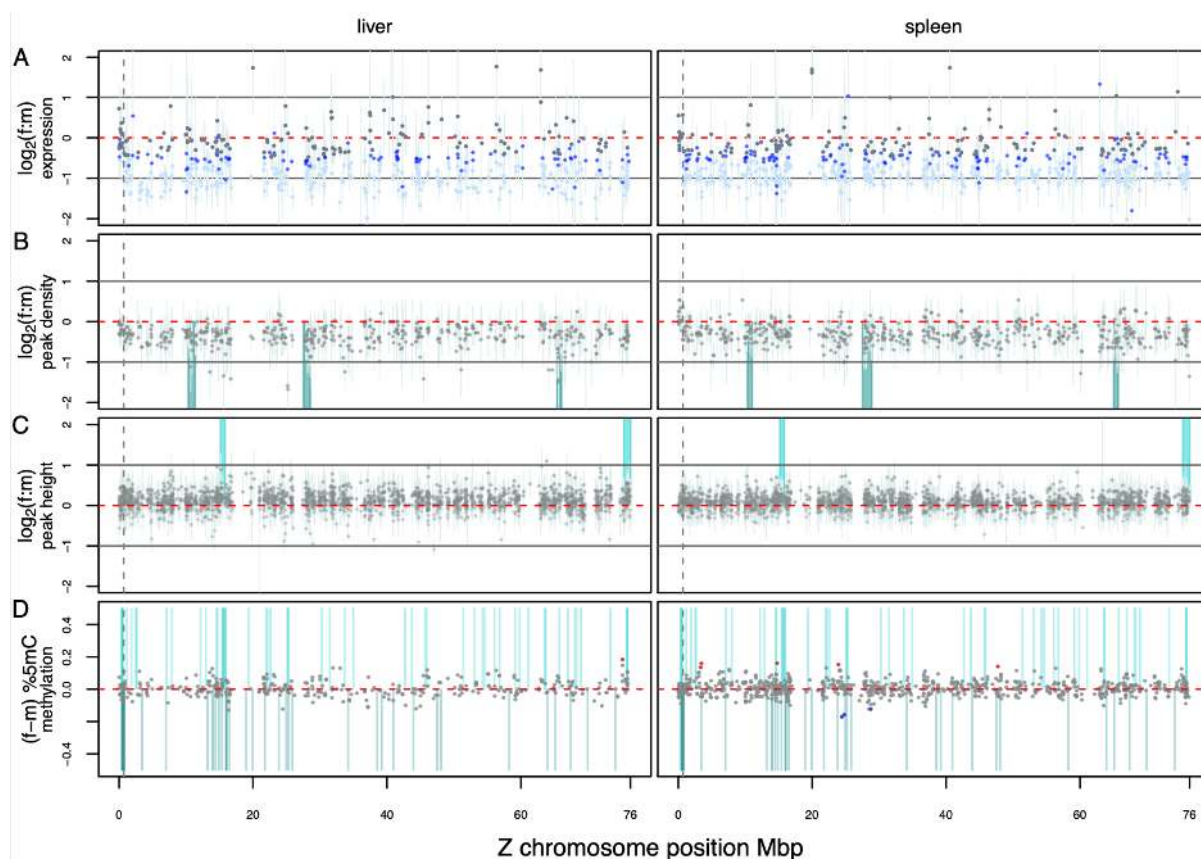
Under the hypothesis that gene expression patterns are governed by chromatin accessibility, we did not expect regionally clustered variation in ATAC-seq peak density and height on the Z chromosome either. As for gene expression, peak density in the PAR was comparable between sexes and autosomal values of the nonPAR region in males. In females, peak density in the nonPAR was lower than in the PAR across all levels of integration (**Supplementary Models E2-E3**). A sliding window analysis (window size of 688 kb and 344 kb shift) identified no region on the Z enriched for a higher density of peaks in females. In males, three non-overlapping regions on the Z were identified with higher peak density in both organs (dark blue areas in **Figure CC**). In spleen, ten additional regions were identified (**Figure CB**, **Supplementary Tables SE**). Overall, these results are consistent with homogeneously reduced peak density outside of the PAR region in females (**Figure CB**).

### Peak height

The height of peaks associated with genes also did not differ between sexes in the PAR region and was close to a ratio of one (**Supplementary Model E4**, **Supplementary Figure SO**). For the remaining part of the Z chromosome, peak height was significantly elevated in females relative to the syntenic nonPAR region in males. This pattern was stable across tissues and across genes or regulatory regions alone (**Supplementary Models E4-E5**). Of all 1761/1770 peaks shared between females and males, 90/78 peaks were female-biased and 16/17 were male-biased in liver and spleen, respectively, although after multiple testing significance was lost (**Supplementary Table SE**). Using a sliding window approach (window size 688 kb, shift 344 kb), we uncovered two regions with significantly higher peaks for females in both liver and spleen (light blue areas in **Figure CC**) and an additional 15 non-overlapping regions present in spleen only (**Supplementary Table SF**), suggesting tissue specific regulation of chromatin accessibility. Overall, these analyses confirm increased peak height in the single copy region of the Z chromosome in females relative to males without much evidence of local clustering (**Figure CC**).

### 5mC methylation – no evidence for a male hypermethylated region

In chicken, two male hypermethylated regions (MHM) have been identified on the Z chromosome, in which genes appear dosage compensated as a result of gene expression arrest through male methylation (Sun et al., 2019; Wright et al., 2015). In the crow data, we found no significant regional enrichment of male methylation. At base-pair resolution, we observed several localized regions of DNA methylation that differed between the sexes along chromosome Z, including the PAR. We identified high autocorrelation within these regions among many of the 5mC positions for both sexes, suggesting sexually differentiated methylated blocks interspersed along the Z chromosome (**Supplementary Figure SQ**). Yet, there was no consistent pattern of exclusively male nor female-biased blocks. We then replicated the above analyses for methylation of genes and CpG islands (see feature dataset **Supplementary Figure SR**). Regions of sexual differentiation differed compared to the analysis at base pair resolution, but likewise identified no male (nor female) hypermethylated region.



**Figure C.** Mean  $\log_2(f.m)$  along the Z for (A) gene expression (B), peak density (C), peak height (D) and methylation. A–C: For each estimate, 95% confidence intervals were calculated as the median of bootstrapped female and male values (10,000 times) of gene-centered estimates. Shaded vertical areas represent regions shared between organs with significant enrichment of compensated genes in males (dark blue, values < 0) or females (light blue, values > 0). A: Gray dots represent dosage compensated genes, royal blue genes with partial compensation and light blue genes with no compensation.

## Discussion

In this study, we quantified chromosome and sex-specific variation in gene expression (RNA-seq), and used two methods to interrogate the underlying epigenetic variation including chromatin features (ATAC-seq peak density and height) and levels of 5mC methylation (WGBS). We interpret our findings in light of dosage compensation and dosage balance.

### Gene expression: $Z_f < (ZZ_m = AA_{f=m})$

Gene expression data provided evidence for incomplete dosage compensation ( $Z_f:AA_f < 1$ ) and dosage balance ( $Z_f:ZZ_m < 1$ ) on a background of autosomal parity between the sexes ( $AA_f = AA_m$ ). The expression pattern of  $Z_f < (ZZ_m = AA_{f=m})$  has been observed in many ZW chromosome systems such as birds, snakes and *Heliconius* butterflies, irrespective of the age of the non-degraded sex-chromosome (Catalán et al., 2018; Gu et al., 2017). Deviations from this pattern have been suggested for moths (Smith et al., 2014) and livebearer fish (Darolti et al., 2019). Our statistical analyses controlling for confounding genomic features show that gene expression on the single-copy part of the female Z chromosome was significantly increased over the expectation of gene dose alone. This was observed relative to the expression of genes

residing on the Z in females and males ( $0.5 < Z_f:ZZ_m < 1$ ), relative to overall expression of autosomal genes when compared to the Z ( $0.5 < Z_f:AA_f < 1$ ), and relative to genes on the dual-copy, pseudo-autosomal region of the female Z chromosome ( $0.5 < \text{nonPAR}_f:\text{PAR}_f < 1$ ). Using autosomal levels of gene expression as an approximation for ancestral, diploid expression levels, this result argues for upregulation of genes located on the hemizygous Z chromosome in females during the course of sex chromosome evolution. This conclusion is consistent with previous point estimates from other avian species exceeding  $Z_f:AA_f$  expression ratios of 0.5 for the majority of genes (Ellegren et al., 2007; Itoh et al., 2010; Julien et al., 2012; Uebbing et al., 2013; Wolf and Bryk, 2011), but see (Chen et al., 2020).

### Epigenetic dosage effects: from pattern to mechanism?

Epigenetic variation is an important determinant of variation in gene expression (Ando et al., 2019). Sex- and chromosome-specific regulation of epigenetic variation thus constitutes a plausible mechanism to equalize differences in gene dose (Duncan et al., 2018; Graves, 2014; Park and Kuroda, 2001; Samata and Akhtar, 2018). We here quantified broad-scale patterns of chromatin structure and 5mC methylation serving as proxies for epigenetic mechanisms with direct and indirect effects on transcription (Klemm et al., 2019; Lindner et al., 2021). This approach is based on the rationale that these two epigenetic methods are, in essence, functionally related to gene expression. Several lines of evidence support this hypothesis. ATAC-seq peak height is known to translate into higher chromatin accessibility (Corces et al., 2018), and other studies have seen a correlation between the number of chromatin accessible sites and the level of gene expression (Marlétaz et al., 2018; Werner et al., 2018). A regulatory role of open chromatin is not necessarily restricted to the close vicinity of genes, as regulatory elements can be found megabases away from the target gene (Hug et al., 2017). Local variation in 5mC methylation is of undoubted importance for gene expression (Veselovska et al., 2015), and has previously been discussed in the context of dosage compensation (Richard et al., 2017). Also, for several of the chromatin accessibility metrics and methylation levels considered here, we observe an expected relationship with broad-scale patterns of gene expression. Clearly, in the absence of a detailed regulatory map in crows inferring process from (broad-scale) pattern is a leap of faith, and results need to be interpreted with caution. Many of the open chromatin and methylated regions will not bear any direct functional relationship to gene expression as has been shown by other studies (Klemm et al., 2019). Still, we believe that the overall patterns described in the study do contain meaningful information. To mitigate potential biases and reduce potential background noise concealing functionally active genomic regions, we quantified patterns along a gradient of increased functional resolution ranging from the coarse category of chromosomes, to genes and fine-scaled overlap of ATAC-seq peaks with 5mC methylation data.

An additional challenge of interpreting patterns of epigenetic variation across sexes and chromosomes is the influence of confounding variables. Unsurprisingly, variation in both the ATAC-seq and 5mC methylation data was in part explained by chromosomal features such as chromosome length, gene length or GC content. All of these variables are closely related to recombination rate and are known to influence many aspects of genome biology in birds (Nam et al., 2010; O'Connor et al., 2019). For example, in the neighborhood of genes, nucleosome spacing (peak density) and accessibility of nucleosome free regions (peak height) were both influenced by chromosome length and gene length. In short, open chromatin was sparser, but more accessible the longer the genes and chromosomes. The degree of 5mC methylation also depended on chromosome length and GC content with hypomethylated genes residing on longer chromosomes. Clearly, these confounding variables need to be statistically accounted for to obtain sex-specific parameter estimates isolating dose effects between the Z chromosome and the highly variable autosomes. The use of flexible, mixed-effects linear models serves this



purpose and has been advocated in the context of dosage compensation before, though has rarely been put into practice (Gu and Walters, 2017). Discrepancies between our model-based estimates and Z:A ratios derived from raw data in this study underline the importance of this aspect (**Supplementary Table A, C**).

#### Density of open chromatin regions: $Z_f < (ZZ_m = AA_{f=m})$

Patterns of open chromatin density closely mirrored patterns of gene expression. Chromatin accessible regions were spaced at equal rate across male and female autosomes and the male Z chromosome. Density of open chromatin peaks was only reduced on the female Z. Strikingly, density was not reduced to half, but to a ratio in between 0.5 and 1, mimicking the pattern of partial dosage compensation for gene expression. This pattern was stable across scales of resolution: genome-wide, gene-centered and regulatory. This may partly be explained by the fact that the majority of all the identified peaks (55%) resided near expressed genes ( $\pm 20\text{kb}$ ) diminishing the influence of intergenic regions. Nonetheless, near-identical f:m and Z:A ratios and confidence intervals across scales of integration speak for a rather homogenous, chromosome-wide effect. Under the justified assumption that homologous chromosomes share a substantial part of open chromatin regions, this observation does not necessarily require an independent epigenetic mechanism of the paternally inherited, single-copy Z chromosome of females.  $Z_f/ZZ_m$  ratios between 0.5 and 1 may simply reflect the proportion of open chromatin peaks constitutively shared between homologous chromosomes. Differentiating between this scenario and active depletion of nucleosome density on the female Z would require quantification of ATAC-seq peaks for each of the male homologues separately.

#### Accessibility of open chromatin regions: $Z_f > (ZZ_m = AA_{f=m})$

In contrast to gene expression and peak density, peak height was elevated on the female Z chromosome relative to males and autosomes of both sexes. Again, estimates in f:m and Z:A ratios were near-identical across scales of integration pointing towards a chromosome-wide mechanism. We therefore hypothesize that a majority of genes achieve partial dosage compensation and balance through a sex-specific, chromosome-wide mechanism increasing chromatin accessibility throughout the female Z chromosome. A similar chromatin Z-specific remodeling mechanism leading to a permissive chromatin landscape has been suggested in other, non-avian female heterogametic systems (Picard et al., 2019). In principle, such a mechanism is compatible with a gene-by-gene mechanism of compensation, as has previously been suggested for birds (Mank, 2009; Mank and Ellegren, 2009; Warnefors et al., 2017). Gene-to-gene variation in the degree of compensation can still be achieved by variation in the distribution of open chromatin regions and other epigenetic marks. It is, however, incompatible with different levels of compensation strength corresponding to evolutionary strata on the Z chromosome (Wright et al., 2015; Xu et al., 2019b).

#### Epigenetic patterns along the Z

Our hypothesis of a chromosome-wide epigenetic mechanism was supported by our analysis of epigenetic patterns along the Z chromosome. To address the question of heterogeneity along the Z chromosome, we performed window-based analyses. First, we isolated the homogametic PAR region, which proved to be comparable in size and gene content to other passerines (Xu et al., 2019a; Zhou et al., 2014). Parameter estimates of expression levels, peak density and peak height on the female PAR were statistically comparable to genes on the male PAR. Outside of the female PAR, we identified a reduction of gene expression and peak density, but an increase in peak height. Due to the small size of the PAR, confidence intervals were large, but estimates reflected the overall patterns observed between the hemizygous Z chromosome and the autosomes. Approximately 47% of the expressed genes that were classified to be



compensated were randomly distributed along the Z. For peak density, the sliding window analysis uncovered three shared regions in liver and spleen that showed significantly lower peak density when compared to the PAR in females. These regions could contain male specific regulation centers supporting higher transcription of genes beneficial to males. For peak height, two regions on the female Z were significantly elevated compared to the PAR in both tissues, pointing towards highly localized regions of female-bias, which could function as additional coordination centers for gene regulation. Overall, however, these findings are consistent with chromosome-wide regulation of open chromatin and provide no evidence for a localized ‘organization center’ of dosage compensation.

### 5mC methylation: no global nor regional dosage effect

A lack of regional regulation of Z-linked gene expression was likewise evident when considering 5mC methylation levels. As expected, methylation differed strongly among levels of integration. Gene bodies were most strongly methylated, followed by CpG islands and regions of accessible chromatin. Chromatin accessibility is expected to show a strong negative relationship with DNA methylation (Taudt et al., 2016), as DNA methylation hinders the anchoring of the transcription machinery and tightens up the chromatin structure (Duncan et al., 2018; Moore et al., 2013). Despite these expected genome-wide correlations, patterns of DNA methylation levels were neither sex-specific, nor did they show any clear relationship to the pattern of partial dosage compensation observed for gene expression and ATAC-seq data. Whole-genome methylation data rather portrayed a conserved methylomic landscape on the Z chromosome between sexes. This is consistent with previous reports in chicken and the white-throated sparrow (Sun et al., 2019). Yet, we found no evidence for male-specific hypermethylation on the Z chromosome, as suggested for chicken (Arnold et al., 2008; Melamed and Arnold, 2007; Sun et al., 2019) and other species in the Galloanserae clade (Wright et al., 2015). Instead, we identified regions of increased differences in methylation consisting of hypo- and hypermethylated sites in both sexes. The lack of a MHM region in the European crow further supports a growing body of evidence that male-specific hypermethylated regions are far from ubiquitous across Aves (Itoh et al., 2010; Uebbing et al., 2013), and these regions are also absent in whole-genome bisulfite data in the white-throated sparrow (Sun et al., 2019). The lack of a MHM region further indicates that in the European crow, the DNA methylation landscape on the Z chromosome does not contribute to regional sex-specific gene expression to achieve dosage compensation. This is consistent with mechanisms detected in zebra finch, where methylation is autonomous from dosage compensating mechanisms in brain tissue (Diddens et al., 2021). Instead, we hypothesize that the female’s Z has evolved a chromatin environment that increases permissiveness along the entire chromosome and that gene-specific control of expression is fine-tuned by specific chromatin accessible regions.

### Conclusions

This study provides evidence for partial dosage compensation and dosage balance in crows as a result of female upregulation of Z-linked gene expression. It further sheds light on the potential underlying epigenetic mechanisms suggesting a chromosome-wide effect of increased chromatin accessibility on the hemizygous Z chromosome. 5mC methylation appeared to play no role, nor did regional ‘centers’ of regulation on the Z chromosome, such as the male hypermethylated region, find any support. The data also provided no hint at an influence of discrete genomic regions differing in the age of recombination cessation (evolutionary strata (Xu et al., 2019a). Dedicated, comparative dosage compensation studies across the avian tree will be needed to see whether signals of gene expression and epigenetic patterns still reflect the ghost of past evolutionary strata. In summary, this work highlights the role of chromatin dynamics in the process of dosage compensation and constitutes a first step towards a functional understanding of partial dosage compensation mechanism in birds.

## Material and methods

### Taxonomic considerations

The two taxa considered here – carrion crows (*Corvus (corone) corone*) and hooded crows (*Corvus (corone) cornix*) – have been considered as separate species by some (Linnaeus, 1758; Parkin et al., 2003). Recent genome-wide evidence, however, rather supports treatment as a single species with two segregating colour morphs (Knief et al., 2019; Knijff, 2014; Poelstra et al., 2014; Vijay et al., 2016). Relevant for the context of this study, previous work has also found near-identical gene expression profiles (Warmuth et al., 2021) with marked differences only for genes expressed in melanocytes underlying plumage divergence (Poelstra et al., 2015; Wu et al., 2019). For the purpose of this study, we therefore consider carrion and hooded crows as members of the same species.

### Data set and sample collection

In May 2014, crow hatchlings of an approximate age of 21 days were obtained directly from the nest (Weissensteiner et al., 2015). Hooded crows (*C. (corone) cornix*) were sampled in the area around Uppsala, Sweden (59°52'N, 17°38'E), and carrion crows (*C. (corone) corone*) in the area around Konstanz, Germany (47°45'N, 9°10'E). A single individual was selected from each nest to avoid any confounding effects of relatedness. After transfer of carrion crows to Sweden by airplane, all crows were hand-raised indoors at Tovetorp field station, Sweden (58°56'55"N, 17°8'49"E). When starting to feed by themselves, they were released to large roofed outdoor enclosures (6.5 x 4.8 x 3.5 m), specifically constructed for the purpose. All crows were maintained under common garden conditions in groups of a maximum of six individuals separated by sub-species and sex. In October 2016, at an age of approximately 2.5 years, individuals were euthanized by cervical dislocation. Tissues were dissected from eight females (3 *corone*, 5 *cornix*) and seven males (4 *corone*, 3 *cornix*). Tissue for RNA extraction was conserved in RNAlater (ThermoFisherScientific) and stored at -80°C. Tissue designated for ATAC-seq was flash frozen at -80°C immediately after dissection. Metadata on all individuals can be retrieved in **Supplementary Table SJ**.

*Regierungspräsidium Freiburg* granted permission for the sampling of wild carrion crows in Germany (Aktenzeichen 55- 8852.15/05). Import into Sweden was registered with the *Veterinärämkt Konstanz* (Bescheinigungsnummer INTRA.DE.2014.0047502) and its Swedish counterpart *Jordbruksverket* (Diarienummer 6.6.18-3037/14). Sampling permission in Sweden was granted by *Naturvårdsverket* (Dnr: NV-03432-14) and *Jordbruksverket* (Diarienummer 27-14). Animal husbandry and experimentation was authorized by *Jordbruksverket* (Diarienummer 5.2.18-3065/13, Diarienummer 27-14) and ethically approved under the Directive 2010/63/EU on the Protection of Animals used for Scientific Purposes by the *European Research Council* (ERCStG-336536).

### RNA-seq

#### Data generation and processing

RNA-seq reads were generated for liver and spleen. Frozen tissue was disrupted with a TissueRuptor (Qiagen, Hilden, Germany), and total RNA was extracted with the RNeasy Plus Universal Kit (Qiagen, Hilden, Germany). RNA quantity and quality were determined with an Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA). RNA-seq libraries were prepared from 500 ng total RNA using the TruSeq stranded mRNA library preparation kit (Illumina Inc, Cat# 20020594/5) including polyA selection. The library preparation was

performed according to the manufacturers' protocol (#1000000040498). Fifty base pair single-end reads were generated on a HiSeq2500, v4 sequencer (Illumina). Sequencing was performed by the SNP&SEQ Technology Platform in Uppsala, Sweden.

Base pair quality was assessed with FastQC v0.11.5 (Andrews, 2010) and bases with a Phred quality score < 20 were removed. All reads with a minimum length of 20 bp were kept for mapping. Reads were mapped to the crow genome version 5.6 (NCBI: ASM73873v4) with STAR (Dobin et al., 2013). An annotation lift-over was done from crow genome version 2.5 (Poelstra et al., 2014) to the version 5.6. We used StringTie (Pertea et al., 2015) to predict novel isoforms, that were merged into the annotation lift-over file if not overlapping with already annotated features. Raw read counts per annotated transcript were calculated with HTSeq (v0.9.1) (Anders et al., 2015). Fragments Per Kilobase per Million mapped reads (FPKM) were calculated and normalized with Cufflinks (Trapnell et al., 2013). For empirical summary statistics of central tendency, we filtered out genes if the mean FPKM of the focal gene across all females and across all males was <1. Thus, genes with FPKM > 1 were considered expressed. For all model-based inference, "actively expressed genes" were identified by calculating zFPKM-values for each sample and tissue using the zFPKM R package (v1.10.0) and we considered genes with zFPKM > -3 as expressed (Hart et al., 2013).

#### Quantification of dosage compensation and dosage balance

For each tissue, f:m expression ratios were calculated per chromosome (A1A1<sub>f</sub>:A1A1<sub>m</sub>, A2A2<sub>f</sub>:A2A2<sub>m</sub>, ... Z<sub>f</sub>:ZZ<sub>m</sub>) as  $\log_2(\text{median}(\text{Gene}_i\text{FPKM}_{\text{female}})/\text{median}(\text{Gene}_i\text{FPKM}_{\text{male}}))$  for each gene. Sex-specific Z:A ratios (Z<sub>f</sub>:AA<sub>f</sub>, ZZ<sub>m</sub>:AA<sub>m</sub>) were calculated as  $\text{median}(\log_2(\text{FPKM of all Z-linked genes})) / \text{median}(\log_2(\text{FPKM of all A}_i\text{-linked genes}))$ . The degree of dosage compensation was categorized by partitioning  $\log_2(\text{f:m})$  expression of each gene into three categories: (1) compensated:  $\log_2(\text{f:m}) > -0.25$  and  $\log_2(\text{f:m}) < 0.25$ , (2) partially compensated:  $-0.25 > \log_2(\text{f:m}) > -0.5$  and  $0.25 < \log_2(\text{f:m}) < 0.5$  and (3) uncompensated:  $\log_2(\text{f:m}) < -0.75$  and  $\log_2(\text{f:m}) > 0.75$ . All  $\log_2(\text{f:m}) > 2$  were classified as female or male biased genes.

#### ATAC-seq

##### Data generation and processing

We followed an ATAC-seq protocol specifically developed for frozen tissue (Corces et al., 2017). In brief, ~ 10 mg of cryopreserved liver and spleen tissue samples were ground using a TissueRuptor (Qiagen, Hilden, Germany) in liquid nitrogen and subsequently homogenized in HB Buffer (see buffer recipes in Corces et al. 2017). Cells were counted with a hemocytometer and 50,000 cells were used for transposition. Transposition was carried out for 60 min in a 50µl reaction with 2.5 µl (100nM final) of Tn5 transposase (Illumina Nextera) per sample. Libraries were amplified using the primers published in (Buenrostro et al. 2013). The number of necessary library PCR amplification cycles was determined by qRT-PCR using SYBR Green 10x in DMSO (Sigma-Aldrich) and NEBNext Master Mix 2x (New England BioLabs). DNA concentration of the resulting library was determined with a Qubit 2.0 Fluorometer (Life Technologies, Grand Island, NY) and its quality was assessed using an Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA). Libraries that passed the quality controls (bioanalyzer fragment distribution following nucleosomal patterning) were selected for sequencing. Two technical replicates were prepared per sample and sequenced (50 bp paired-end) on an Illumina HiSeq1500 sequencer in an inhouse sequencing facility at LMU Munich and on an Illumina NovaSeq S1 flowcell by the SNP&SEQ Technology Platform in Uppsala.

Sequencing data quality was assessed with FastQC (v0.11.5), and base pairs with a sequencing Phred quality score < 20 were removed. Nextera adapters were trimmed, and reads with a minimum length of 36 bp were selected for further mapping using Cutadapt (v1.9.1)

(Martin, 2011). Selected reads were mapped to the crow genome version 5.6. (NCBI accession number: GCA\_000738735.4) using Bowtie2 (v2.3.2) (Langmead et al., 2009), with the following parameters: bowtie2 -very-sensitive -X 2000, allowing for fragments of up to 2kb to align. Only properly paired and uniquely mapped reads with a mapping quality score > Q20 were filtered out using samtools (v1.4) (Li et al., 2009). PCR duplicate reads were removed with Picard tools (v2.0.1).

### Quality control

Following guidelines of ATAC-seq quality control assessment, we included the following indicators to judge data quality: (1) *Insert size distribution* following a nucleosomal spacing periodicity (**Supplementary Figure SA**) (2) *Spearman correlation coefficients* between technical replicates (**Supplementary Figure SB**) (3) *The number of usable reads* (4) *The fraction of reads in peaks* (FRiP) as calculated with deepTools -plotEnrichment (**Supplementary Figure SC**) and (5) *ATAC-seq coverage enrichment at transcription start sites* (TSS), (**Supplementary Figure SD**). More details on quality control are provided in the **Supplement**. Technical replicates that survived the quality check were merged by biological replicate for all subsequent analyses. A summary of the usable reads can be found in **Supplementary Table SG**.

### ATAC-seq peak calling

Differences in read coverage among samples can influence the sensitivity and specificity of peak identification. To avoid calling higher peak density in samples with more read depth, we down-sampled all the libraries after mapping and filtering to the read number of the smallest library. This resulted in ~60 million workable reads in liver and ~90 million in spleen for all samples (**Supplementary Table SH**). Even though we had a third more reads for spleen, we did not detect more peaks in this tissue, corroborating the fact that we reached a saturation peak-detection point. Down-sampling was done with samtools view -s and coverage equality across libraries was checked by calculating the coverage for windows of 10 bp using deepTools -bamCoverage -bs 10 -p max (v2.27.1) (Ramírez et al., 2016). After library normalization, peaks were called using MACS2 (v2.1.0) macs callpeak -t -n -g 1e9 -f BAMPE (Zhang et al., 2008).

### ATAC-seq analysis

We quantified two metrics to describe chromatin accessibility for each individual: ATAC-seq peak height and peak density. Peak height was defined as the fold-change value estimated with MACS2 (Zhang et al., 2008). Peak density was defined as the number of ATAC-seq peaks present in a given genomic interval divided by the length of the interval. Chromosome-wide peak density was calculated as the number of peaks per chromosome divided by its length. Gene-centered peak density was accordingly normalized to the length of the focal expressed gene. Empirical chromosome-wide estimates of female to male ratios (f:m) for peak height and density were calculated as the median(peak height or density)<sub>female</sub>/median(peak height or density)<sub>male</sub>. Z-chromosome to autosomes ratios (Z:A) for each sex were calculated as median(Z chromosome peak height or density)/median(autosomes peak height or density). For peak-centered peak heights f:m were calculated as f:m = median(Gene<sub>i</sub> peak height<sub>female</sub>)/median(Gene<sub>i</sub> peak height<sub>male</sub>). For gene-centered peak density estimates, f:m ratios were calculated as f:m = Gene<sub>i</sub> peak density<sub>female</sub>/ Gene<sub>i</sub> peak density<sub>male</sub>. Gene-centered Z:A ratios were calculated as described for chromosome-wide.



## 5mC methylation

### Data generation and processing

We generated whole-genome bisulfite sequencing (WGBS) libraries for one female and one male individual for both liver and spleen tissue. Post-bisulfite converted libraries were generated using the Illumina TruSeq Methylation Kit (EGMK91324) following the manufacturer's protocol at the SNP&SEQ Technology Platform in Uppsala, Sweden. Raw reads were trimmed using Trim\_Galore! v0.6.6 (Krueger, 2012), clipping the first 9 bases from the 5' ends, and the last base of the 3' end, following standard post-bisulfite converted WGBS library protocols and as indicated by methylation bias plots (**Supplementary Figure SS**). Base and tile quality was assessed with FastQC v0.11.9 after trimming (Andrews, 2010). Cleaned reads were then aligned to the reference genome version 5.6 (NCBI: ASM73873v4 with default settings, followed by read de-duplication and extraction of CpG methylation state using Bismark v0.23.0 (Krueger and Andrews, 2011). We retained only CpGs on assigned chromosomes (1 – 28 and Z) and removed any positions overlapping transition mutations (C - T, or G - A SNPs), identified from population resequencing data generated from a previous study (discussed below; (Poelstra et al., 2014). We then further subset and excluded any sites overlapping repeats identified with RepeatMasker v4.1.1 (Smit et al., 2013), using the chicken repeat library and not including short simple repeats (-nolow). We then summarized coverage by sample (**Supplementary Figure SS**). We also created a CpG island track for assembly version 5.6 because of our interest in regulatory regions, using makeCGI (Wu et al., 2010), requiring stringent thresholds of a length of at least 250bp, a GC content > 60%, and an observed to expected CG ratio > 75%, resulting in 18,526 islands (**Supplementary Table SI**).

### Population resequencing reanalysis

We reanalyzed population resequencing data for 15 male hooded (*C. (corone) cornix*) and 15 male carrion (*C. (corone) corone*) crows from an existing study (Poelstra et al., 2014) to identify transition SNPs which could confound WGBS methylation calls (*e.g.* a TG genotype at a CG site will incorrectly be called as non-methylated, it is therefore best practices to exclude C – T and A – G sites (Barrow and Byun, 2014). In short, resequencing data was trimmed with BBtools v38.18 (Bushnell, 2021) and aligned to the crow reference genome using BWA v0.7.17 (Li and Durbin, 2009). Duplicates were removed with GATK v4.1.4.1 (Auwera et al., 2013) and variants were called with Freebayes v1.3.2 (Garrison and Marth, 2012) without population priors. SNPs were filtered with bcftools (Li, 2011), excluding INDELs and sites exhibiting the strongest allelic balance (AB), depth (DP), strand bias (SRP, SAP) and read placement (EPP) biases, identified from the top and bottom 2.5% of sites for each metric (only the bottom 2.5% were removed for SRP, SAP, and EPP because only low scores indicate bias). We furthermore removed sites with a quality (QUAL) below 20, a genotype quality (GQ) below 20, and required at least 75% of individuals to share a site with at least a genotype depth of 3 reads. We then retained only C – T or A – G SNP sites using bash.

### Genome-wide data analyses

We calculated the proportion of methylated to non-methylated sites considering individual 5mC calls only if covered by at least 5 reads. Methylation data was further only considered if a feature of interest (see below) contained at least three 5mC calls, and was scorable for both male and female tissue-specific libraries. DNA methylation analysis was divided into five distinct feature-based analyses to determine the relative impacts of putatively increasing regulatory importance (**Supplementary Figure SJ**). First, we considered chromosome-wide levels, with the entirety of each chromosome as a feature. Second, we analyzed positions within genes, including each gene and its 20 kb up- and downstream sequence as a feature. We then examined levels within CpG islands, with each island as a distinct feature. Our fourth and fifth analyses analyzed DNA methylation within open chromatin that was shared between males and females



(requiring peaks to overlap by at least 75% between sexes to be considered). As peaks differed between liver and spleen, a separate coordinate file was created for each tissue, which was then overlapped with the tissue-specific DNA methylation data. Specifically, the fourth analysis analyzed DNA methylation in open chromatin that either overlapped or was within 20 kb upstream or downstream of a gene. The fifth analysis only included the open chromatin upstream and downstream of a gene, not the peaks overlapping the gene itself.

#### Identification of a male hypermethylated region

Two male hypermethylated (MHM) regions have been identified in *Gallus gallus* on the Z chromosome, although only one region has been found in other avian clades (Sun et al., 2019). Here, we explored the possibility of an MHM region in the European crow by examining base-pair resolution DNA methylation divergence and genomic autocorrelation between sexes and across tissues. We started our search by blasting the known MHM sequences (Sun et al., 2019) against the hooded crow reference, but detected no hits on the Z chromosome. We hypothesized that if the European crow harbored an MHM, it would exhibit male-specific hypermethylation and strong genomic autocorrelation within that region. To test this, we used the same filtered 5mC calls as above, but instead kept each individual position as a feature, and retained only sites shared between both sexes. We then visualized divergence (both f-m and  $\log_2(f/m)$ ) along the Z using ggplot2 (Wickham, 2016). To detect regions of high autocorrelation in each sex, we sampled 50 consecutive methylation positions and calculated Spearman correlations of 5mC levels between the first and second half of sites. We then proceeded to examine the remainder of the Z chromosome in tiled 50 base-pair windows. We classified blocks of differentiation as regions exhibiting methylation differences greater than 25% between the sexes and Spearman correlations greater than 0.3, merging any regions within 10 kb. We repeated this process except with a larger 250 base-pair block, and arrived at the same results.

#### Identification of the pseudo autosomal region (PAR)

We took three approaches to identify the PAR in the European crow. (1) Genomic regions on the Z belonging to the PAR are expected to have f:m<sub>coverage</sub> of approximately 1 for whole-genome shotgun sequencing data, genomic regions outside the PAR of approximately 0.5. To estimate sex-specific coverage along the Z chromosome we used 150bp paired-end Illumina sequencing data from one female and one male crow with NCBI bioproject\_id: PRJNA192205 (Poelstra et al., 2014). Reads were mapped with BWA mem (v0.7.17) (Li and Durbin, 2009) to version 5.6 of the crow genome (NCBI: ASM73873v4). Only concordant paired reads with mapping quality >20 were kept using samtools view -u -h -q 20 -f 0x2 (Li et al., 2009). Read duplicates were removed with Picard tools and repeats were masked and filtered using RepeatMasker and the chicken library (v4.0.7) (Smit et al., 2015). Coverage was calculated with bamcoverage deepTools (v2.5.1) (Ramírez et al., 2016) in 50kb windows for the Z chromosome and chromosome 1 and 2 which are similar in size and GC content: bamCoverage --binSize=50000 --normalizeUsingRPKM -p 2. Differences in chromosomal coverage were assessed as follows: the coverage value of each 50 kb window along chromosomes 1,2 and Z was divided by the median coverage of these autosomes. This was done separately for the female and male sample. The normalized coverage was used to calculate  $\log_2(f:m_{\text{coverage}})$ . (2) In addition to the coverage-based approach, we assessed synteny with the PAR of collared flycatcher (*Ficedula albicollis*; (Smeds et al., 2014). We aligned the Z of the flycatcher genome version 1.5 (Kawakami et al., 2014) using the genome aligner lastZ (v1.04.00) (Harris, 2007) and selected only those contigs with > 90% identity. Anchors to homologous regions between the species were plotted in R (R Core Team, 2015) (3) In heterogametic females, we expect to find heterozygous sites in the PAR as opposed to regions outside the PAR, where no heterozygosity should be found. We used GoldenGate genotype data from the Z chromosome (N = 129 SNPs) collected on 522 hooded and carrion crows and

their hybrids (data described in Knief et al. (2019)). For each SNP, we calculated the observed heterozygosity in females (N = 277 individuals) and males (N = 245 individuals). We transferred SNP positions to crow genome version 5.6 and observed 15 heterozygous SNPs in females on the Z chromosome between 27,730 and 687,785 bp.

## Statistical analysis

### Linear models

We fitted linear mixed-effects models with a Gaussian error structure using the lme4 (v1.1-27.1) and lmerTest (3.1-3) R packages (Bates et al., 2015; Kuznetsova et al., 2017). We transformed the data to match normality (dependent variables: FPKM and ATAC peak height were log<sub>2</sub>-transformed, ATAC peak number was square-root-transformed and percent 5mC was arcsine-square-root-transformed). It has been shown that Gaussian models are robust and that violating the normality assumption has only minor effects on parameter estimates and P-values (Knief and Forstmeier, 2021). For each dependent variable, we fitted whether the gene/ATAC peak/methylation window was located on an autosome or the Z chromosome (factor with two levels) and the sex of the individual (factor with two levels) as interacting fixed effects. We controlled for gene length, chromosome length and GC content where appropriate and Z-transformed these covariates. Individual ID and gene ID nested in chromosome ID were modeled as random effects. For the methylation data set, we had only one female and one male sequenced. Because of that, we did not fit individual ID as a random effect. Results should therefore be interpreted with caution, as we are unable to separate sex from individual effects. Exact model specifications are provided in **Supplementary Models**. We used the effects (v4.2-0) R package (Fox, John, Sanford Weisberg, 2011) to obtain the four parameter estimates (autosomes in females, autosomes in males, Z in females, Z in males). We calculated ratios on these parameter estimates and bootstrapped 95% confidence intervals using the bootMer() function in lme4 with 1,000 replicates. These ratios are reported in **Supplementary Table A**, model estimates can be found in **Supplementary Models**. Fixed effect estimates were considered significant when  $P < 0.05$ .

For chromosome-wide peak density, we additionally fitted mixed-effects models with a quasipoisson or negative binomial error structure (log-link), using peak number per chromosome as our dependent variables and the same fixed and random effects as above, except that we included chromosome-length as an offset term (Zuur et al., 2009). For this, we used the glmmTMB (v1.1.2) (Brooks et al., 2017) package in R. Exact model specifications are provided in **Supplementary Models** and ratios calculated on parameter estimates are reported in **Supplementary Table A**.

In a second set of models, we restricted the data to the Z chromosome and separated the PAR from the rest of the Z (nonPAR), such that we estimated four parameters (nonPAR in females, nonPAR in males, PAR in females, PAR in males). We controlled for gene length (Z-transformed) by fitting it as a covariate and included individual ID and gene ID as random effects. We used the effects (v4.2-0) R package (Fox & Weisberg 2019) to obtain the four parameter estimates, calculated ratios on these parameter estimates and bootstrapped 95% confidence intervals using the bootMer() function in lme4 with 1,000 replicates. These ratios are reported in **Supplementary Table C**, model estimates can be found in **Supplementary Models**.

### Identification of female to male ratios along the Z-chromosome

To identify putative clusters of dosage compensated genes on the Z chromosome and to identify regions on the Z enriched or depleted for peak height and peak density, we performed a sliding

window analysis. We used a window size of 688 kb (the size of the PAR) and 1 Mb with a sliding step of 344 kb (half PAR size) and 100 kb respectively. Gene expression and ATAC-seq peak metrics on the PAR were used as reference to identify deviating regions along the Z chromosome. A Fisher's exact test was performed for each window to detect enrichment or depletion of dosage compensated genes along the Z (categories as defined above), with a Bonferroni correction for multiple testing. To detect regions on the Z enriched or depleted or enriched in number of peaks, we additionally used a Wilcoxon test as alternative = "less" and alternative="greater". To further survey the Z for potential clustered centers of regulation, we conducted an autocorrelation analysis using the `acf()` function from the R package `forecast` v5.8 (Rob J. Hyndman and Yeasmin Khandakar, 2019). In order to identify peaks with significantly different heights between females and males on the Z, we first identified consensus peaks present in either of the sexes. We selected peaks present in 80% of both female and male samples. Peaks overlapping by more than 50% between individuals were merged into a single peak to avoid peak redundancy across samples. For these consensus-peak coordinates we used extracted peak heights as `peak_fold_change` computed in MACS2. Only peaks were considered, where at least three individuals had non-zero values. Differences in the peak height distribution between females and males for a given window was tested in R using a Wilcoxon test.

## Acknowledgements

We express our gratitude to Sven Jakobsson for providing the infrastructure for animal husbandry at Tovetorp research station. We would also like to thank Christen Bossu, Jelmer Poelstra and Matthias Weissensteiner for their contribution in obtaining samples. Kristaps Solokovskis, Thomas Giegold, Nils Andbjør, Tamara Volkmer, Barbara Martinschitsch and Luisa Sontheimer provided invaluable support in raising and maintaining the captive crow population. Martin Wikelski, Inge Müller and additional staff from the Max-Planck-Institute for Ornithology in Radolfzell facilitated sampling in Germany and transport to Sweden. We are further grateful to Gabriele Kumpfmüller with whom the ATAC-seq protocol was established. The Swedish sequencing facility is part of the National Genomics Infrastructure (NGI) Sweden and Science for Life Laboratory. We further thank Stefan Krebs at the Gene Center Munich for discussions and sequencing the ATAC-seq libraries.

## Funding

Funding was provided to J. B. W. W. by the European Research Council (ERCStG-336536 FuncSpecGen.), the Swedish Research Council (Vetenskapsrådet; 621-2013-4510), the Knut and Alice Wallenberg Foundation (Knut och Alice Wallenbergs Stiftelse.), Tovetorp fieldstation through Stockholm University (Stockholms Universitet) and LMU Munich.

## References

- Adolfsson S, Ellegren H. 2013. Lack of Dosage Compensation Accompanies the Arrested Stage of Sex Chromosome Evolution in Ostriches. *Mol Biol Evol* **30**:806–810. doi:10.1093/molbev/mst009
- Anders S, Pyl PT, Huber W. 2015. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**:166–169. doi:10.1093/bioinformatics/btu638
- Ando M, Saito Y, Xu G, Bui NQ, Medetgul-Ernar K, Pu M, Fisch K, Ren S, Sakai A,

- Fukusumi T, Liu C, Haft S, Pang J, Mark A, Gaykalova DA, Guo T, Favorov A V., Yegnasubramanian S, Fertig EJ, Ha P, Tamayo P, Yamasoba T, Ideker T, Messer K, Califano JA. 2019. Chromatin dysregulation and DNA methylation at transcription start sites associated with transcriptional repression in cancers. *Nat Commun* **10**:2188. doi:10.1038/s41467-019-09937-w
- Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>. doi:<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
- Arnold AP, Itoh Y, Melamed E. 2008. A bird's-eye view of sex chromosome dosage compensation. *Annu Rev Genomics Hum Genet* **9**:109–127. doi:10.1146/annurev.genom.9.081307.164220
- Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, Banks E, Garimella K V., Altshuler D, Gabriel S, DePristo MA. 2013. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Curr Protoc Bioinforma* **43**. doi:10.1002/0471250953.bi1110s43
- Bachtrog D. 2013. Y-chromosome evolution: Emerging insights into processes of Y-chromosome degeneration. *Nat Rev Genet* **14**:113–124. doi:10.1038/nrg3366
- Barrow TM, Byun H-M. 2014. Single nucleotide polymorphisms on DNA methylation microarrays: precautions against confounding. *Epigenomics* **6**:577–579. doi:10.2217/epi.14.55
- Bates D, Mächler M, Bolker BM, Walker SC. 2015. Fitting linear mixed-effects models using lme4. *J Stat Softw* **67**. doi:10.18637/jss.v067.i01
- Birchler JA, Riddle NC, Auger DL, Veitia RA. 2005. Dosage balance in gene regulation: biological implications. *Trends Genet* **21**:219–226.
- Brooks EM, Kristensen K, van Benthem K, Bolker MB. 2017. glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *R J*. doi:10.32614/RJ-2017-066
- Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**:1213–8. doi:10.1038/nmeth.2688
- Bushnell B. 2021. BBTools: BMAP short read aligner, and other bioinformatic tools.
- Catalán A, Macias-Munoz A, Briscoe AD. 2018. Evolution of Sex-Biased Gene Expression and Dosage Compensation in the Eye and Brain of Heliconius Butterflies. *Mol Biol Evol* **35**:2120–2134. doi:10.1093/molbev/msy111
- Charlesworth B. 1996. The evolution of chromosomal sex determination and dosage compensation. *Curr Biol* **6**:149–162. doi:10.1016/S0960-9822(02)00448-7
- Chen J, Wang M, He X, Yang JR, Chen X. 2020. The evolution of sex chromosome dosage compensation in animals. *J Genet Genomics* **47**:681–693. doi:10.1016/j.jgg.2020.10.005
- Chen X, Shen Y, Draper W, Buenrostro JD, Litzenburger U, Cho SW, Satpathy AT, Carter AC, Ghosh RP, East-Seletsky A, Doudna JA, Greenleaf WJ, Liphardt JT, Chang HY. 2016. ATAC-seq reveals the accessible genome by transposase-mediated imaging and sequencing. *Nat Methods* **13**:1013–1020. doi:10.1038/nmeth.4031
- Corces MR, Granja JM, Shams S, Louie BH, Seoane JA, Zhou W, Silva TC, Groeneveld C, Wong CK, Cho SW, Satpathy AT, Mumbach MR, Hoadley KA, Robertson AG, Sheffield NC, Felau I, Castro MAA, Berman BP, Staudt LM, Zenklusen JC, Laird PW, Curtis C, Greenleaf WJ, Chang HY. 2018. The chromatin accessibility landscape of primary human cancers. *Science (80- )* **362**. doi:10.1126/science.aav1898
- Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-armstrong NA, Vesuna S, Satpathy AT, Rubin AJ, Montine KS, Wu B, Kathiria A, Cho SW, Mumbach MR, Carter



- AC, Kasowski M, Orloff LA, Risca VI, Kundaje A, Khavari PA, Montine TJ, Greenleaf WJ, Chang HY. 2017. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues **14**. doi:10.1038/nmeth.4396
- Darolti I, Wright AE, Sandkam BA, Morris J, Bloch NI, Farré M, Fuller RC, Bourne GR, Larkin DM, Breden F, Mank JE. 2019. Extreme heterogeneity in sex chromosome differentiation and dosage compensation in livebearers. *Proc Natl Acad Sci U S A* **116**:19031–19036. doi:10.1073/pnas.1905298116
- Diddens J, Coussement L, Frankl-Vilches C, Majumdar G, Steyaert S, Ter Haar SM, Galle J, De Meester E, De Keulenaer S, Van Crielinge W, Cornil CA, Balthazart J, Van Der Linden A, De Meyer T, Vanden Berghe W. 2021. DNA Methylation Regulates Transcription Factor-Specific Neurodevelopmental but Not Sexually Dimorphic Gene Expression Dynamics in Zebra Finch Telencephalon. *Front Cell Dev Biol* **9**:583555. doi:10.3389/fcell.2021.583555
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**:15–21. doi:10.1093/bioinformatics/bts635
- Duncan CG, Grimm SA, Morgan DL, Bushel PR, Bennett BD, Roberts JD, Tyson FL, Merrick BA, Wade PA. 2018. Dosage compensation and DNA methylation landscape of the X chromosome in mouse liver. *Sci Rep* **8**:1–17. doi:10.1038/s41598-018-28356-3
- Ellegren H, Hultin-Rosenberg L, Brunstrom B, Dencker L, Kultima K, Scholz B. 2007. Faced with inequality: chicken do not have a general dosage compensation of sex-linked genes. *Bmc Biol* **5**. doi:40
- Ercan S, Giresi PG, Whittle CM, Zhang X, Green RD, Lieb JD. 2007. X chromosome repression by localization of the *C. elegans* dosage compensation machinery to sites of transcription initiation. *Nat Genet* **39**:403–408. doi:10.1038/ng1983
- Fishman V, Battulin N, Nuriddinov M, Maslova A, Zlotina A, Strunov A, Chervyakova D, Korablev A, Serov O, Krasikova A. 2019. 3D organization of chicken genome demonstrates evolutionary conservation of topologically associated domains and highlights unique architecture of erythrocytes' chromatin. *Nucleic Acids Res* **47**:648–665. doi:10.1093/nar/gky1103
- Fox, John, Sanford Weisberg and JF. 2011. An R companion to applied regression.
- Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. *arXiv:12073907 [q-bio]*.
- Gelbart ME, Kuroda MI. 2009. *Drosophila* dosage compensation: a complex voyage to the X chromosome. *Development* **136**:1399–1410. doi:10.1242/dev.029645
- Graves JAM. 2016. Evolution of vertebrate sex chromosomes and dosage compensation. *Nat Rev Genet*. doi:10.1038/nrg.2015.2
- Graves JAM. 2014. The epigenetic sole of sex and dosage compensation. *Nat Genet* **46**:215–217. doi:10.1038/ng.2903
- Graves JAM. 2006. Sex chromosome specialization and degeneration in mammals. *Cell* **124**:901–914. doi:10.1016/j.cell.2006.02.024
- Gu L, Walters JR. 2017. Evolution of sex chromosome dosage compensation in animals: A beautiful theory, undermined by facts and bedeviled by details. *Genome Biol Evol* **9**:2461–2476. doi:10.1093/gbe/evx154
- Gu L, Walters JR, Knipple DC. 2017. Conserved patterns of sex chromosome dosage compensation in the Lepidoptera (WZ/WW): Insights from a moth neo-Z chromosome. *Genome Biol Evol* **9**:802–816. doi:10.1093/gbe/evx039
- Harris R. 2007. Improved pairwise alignment of genomic DNA. Pennsylvania State University, State College, PA.
- Hart T, Komori HK, LaMere S, Podshivalova K, Salomon DR. 2013. Finding the active genes in deep RNA-seq gene expression studies. *BMC Genomics* **14**. doi:10.1186/1471-2164-



14-778

- Hug CB, Grimaldi AG, Kruse K, Vaquerizas JM. 2017. Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. *Cell* **169**:216-228.e19. doi:10.1016/j.cell.2017.03.024
- Itoh Y, Melamed E, Yang X, Kampf K, Wang S, Yehya N, Van Nas A, Replogle K, Band MR, Clayton DF, Schadt EE, Lusis AJ, Arnold AP. 2007. Dosage compensation is less effective in birds than in mammals. *J Biol* **6**:1–5. doi:10.1186/jbiol53
- Itoh Y, Replogle K, Kim YH, Wade J, Clayton DF, Arnold AP. 2010. Sex bias and dosage compensation in the zebra finch versus chicken genomes: General and specialized patterns among birds. *Genome Res* **20**:512–518. doi:10.1101/gr.102343.109
- Jans J, Gladden JM, Ralston EJ, Pickle CS, Michel AH, Pferdehirt RR, Eisen MB, Meyer BJ. 2009. A condensin-like dosage compensation complex acts at a distance to control expression throughout the genome. *Genes Dev* **23**:602–618. doi:10.1101/gad.1751109
- Jordan W, Rieder LE, Larschan E. 2019. Diverse Genome Topologies Characterize Dosage Compensation across Species. *Trends Genet* **35**:308–315. doi:10.1016/j.tig.2019.02.001
- Julien P, Brawand D, Soumillon M, Necsulea A, Liechti A, Schütz F, Daish T, Grützner F, Kaessmann H. 2012. Mechanisms and evolutionary patterns of mammalian and avian dosage compensation. *PLoS Biol* **10**:1–20. doi:10.1371/journal.pbio.1001328
- Kawakami T, Smeds L, Backström N, Husby A, Qvarnström A, Mugal CF, Olason P, Ellegren H. 2014. A high-density linkage map enables a second-generation collared flycatcher genome assembly and reveals the patterns of avian recombination rate variation and chromosomal evolution. *Mol Ecol* **23**:4035–4058. doi:10.1111/mec.12810
- Klemm SL, Shipony Z, Greenleaf WJ. 2019. Chromatin accessibility and the regulatory epigenome. *Nat Rev Genet* **20**:207–220. doi:10.1038/s41576-018-0089-8
- Knief U, Bossu CM, Saino N, Hansson B, Poelstra J, Vijay N, Weissensteiner M, Wolf JBW. 2019. Epistatic mutations under divergent selection govern phenotypic variation in the crow hybrid zone. *Nat Ecol Evol* **3**:570–576. doi:10.1038/s41559-019-0847-9
- Knief U, Forstmeier W. 2021. Violating the normality assumption may be the lesser of two evils. *Behav Res Methods*. doi:10.3758/s13428-021-01587-5
- Knijff P de. 2014. How carrion and hooded crows defeat Linnaeus’s curse. *Science (80- )* **344**:1345–1346. doi:10.1126/science.1255744
- Krueger F. 2012. Trim Galore! : A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files. Available at [https://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/).
- Krueger F, Andrews SR. 2011. Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**:1571–1572. doi:10.1093/bioinformatics/btr167
- Kuznetsova A, Brockhoff PB, Christensen RHB. 2017. lmerTest Package: Tests in Linear Mixed Effects Models . *J Stat Softw* **82**. doi:10.18637/jss.v082.i13
- Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**:1–10. doi:gb-2009-10-3-r25 [pii]n10.1186/gb-2009-10-3-r25
- Lee HJ, Hou Y, Chen Y, Dailey ZZ, Riddihough A, Jang HS, Wang T, Johnson SL. 2020. Regenerating zebrafish fin epigenome is characterized by stable lineage-specific DNA methylation and dynamic chromatin accessibility. *Genome Biol* **21**:52. doi:10.1186/s13059-020-1948-0
- Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**:2987–2993. doi:10.1093/bioinformatics/btr509
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**:1754–1760. doi:10.1093/bioinformatics/btp324

- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**:2078–2079. doi:10.1093/bioinformatics/btp352
- Lin H, Gupta V, Vermilyea MD, Falciani F, Lee JT, O'Neill LP, Turner BM. 2007. Dosage compensation in the mouse balances up-regulation and silencing of X-linked genes. *PLoS Biol* **5**:2809–2820. doi:10.1371/journal.pbio.0050326
- Lindner M, Verhagen I, Viitaniemi HM, Laine VN, Visser ME, Husby A, van Oers K. 2021. Temporal changes in DNA methylation and RNA expression in a small song bird: within- and between-tissue comparisons. *BMC Genomics* **22**:36. doi:10.1186/s12864-020-07329-9
- Linnaeus C. 1758. *Systema Naturae per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis*.
- Livornois A, Graves J, Waters P. 2012. The origin and evolution of vertebrate sex chromosomes and dosage compensation. *Heredity (Edinb)* **108**:50–58. doi:10.1038/nrg.2015.2
- Mank JE. 2009. The W, X, Y and Z of sex-chromosome dosage compensation. *Trends Genet* **25**:226–233. doi:10.1016/j.tig.2009.03.005
- Mank JE, Ellegren H. 2009. All dosage compensation is local: Gene-by-gene regulation of sex-biased expression on the chicken Z chromosome. *Heredity (Edinb)* **102**:312–320. doi:10.1038/hdy.2008.116
- Marlétaz F, Firbas PN, Maeso I, Tena JJ, Bogdanovic O, Perry M, Wyatt CDR, de la Calle-Mustienes E, Bertrand S, Burguera D, Acemel RD, van Heeringen SJ, Naranjo S, Herrera-Ubeda C, Skvortsova K, Jimenez-Gancedo S, Aldea D, Marquez Y, Buono L, Kozmikova I, Permany J, Louis A, Albuixech-Crespo B, Le Petillon Y, Leon A, Subirana L, Balwierz PJ, Duckett PE, Farahani E, Aury JM, Mangenot S, Wincker P, Albalat R, Benito-Gutiérrez E, Cañestro C, Castro F, D'Aniello S, Ferrier DEK, Huang S, Laudet V, Marais GAB, Pontarotti P, Schubert M, Seitz H, Somorjai I, Takahashi T, Mirabeau O, Xu A, Yu JK, Carninci P, Martinez-Morales JR, Crollius HR, Kozmik Z, Weirauch MT, Garcia-Fernández J, Lister R, Lenhard B, Holland PWH, Escriva H, Gómez-Skarmeta JL, Irimia M. 2018. Amphioxus functional genomics and the origins of vertebrate gene regulation. *Nature* **564**:64–70. doi:10.1038/s41586-018-0734-6
- Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. <https://journal.embnnet.org/index.php/embnnetjournal/article/view/200/479>.
- McDonel P, Jans J, Peterson BK, Meyer BJ. 2006. Clustered DNA motifs mark X chromosomes for repression by a dosage compensation complex. *Nature* **444**:614–618. doi:10.1038/nature05338
- Melamed E, Arnold AP. 2007. Regional differences in dosage compensation on the chicken Z chromosome. *Genome Biol* **8**. doi:R202
- Moore LD, Le T, Fan G. 2013. DNA methylation and its basic function. *Neuropsychopharmacology* **38**:23–38. doi:10.1038/npp.2012.112
- Nam K, Mugal C, Nabholz B, Schielzeth H, Wolf JBW, Backstrom N, Kunstner A, Balakrishnan CN, Heger A, Ponting CP, Clayton DF, Ellegren H. 2010. Molecular evolution of genes in avian genomes. *Genome Biol* **11**:R68.
- Nätt D, Agnvall B, Jensen P. 2014. Large Sex Differences in Chicken Behavior and Brain Gene Expression Coincide with Few Differences in Promoter DNA-Methylation. *PLoS One* **9**:e96376. doi:10.1371/journal.pone.0096376
- Naurin S, Hansson B, Bensch S, Hasselquist D. 2010. Why does dosage compensation differ between XY and ZW taxa? *Trends Genet* **26**:15–20. doi:10.1016/j.tig.2009.11.006
- Nguyen DK, Disteché CM. 2006. Dosage compensation of the active X chromosome in mammals **38**:47–53. doi:10.1038/ng1705
- O'Connor RE, Kiazim L, Skinner B, Fonseka G, Joseph S, Jennings R, Larkin DM, Griffin

- DK. 2019. Patterns of microchromosome organization remain highly conserved throughout avian evolution. *Chromosoma* **128**:21–29. doi:10.1007/s00412-018-0685-6
- Ohno S. 1967. Sex chromosomes and sex-linked genes. Berlin, New York: Springer.
- Park Y, Kuroda MI. 2001. Epigenetic Aspects of X-Chromosome Dosage Compensation. *Science* (80- ) **293**:1083–1085. doi:10.1126/science.1063073
- Parkin DT, Collinson M, Helbig AJ, Knox AG, Sangster G. 2003. The taxonomic status of Carrion and Hooded Crows. *Br Birds* **96**:274–290.
- Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* **33**:290–295. doi:10.1038/nbt.3122
- Picard MAL, Vicoso B, Roquis D, Bulla I, Augusto RC, Arancibia N, Grunau C, Boissier J, Cosseau C, Mank J. 2019. Dosage Compensation throughout the Schistosoma mansoni Lifecycle: Specific Chromatin Landscape of the Z Chromosome. *Genome Biol Evol* **11**:1909–1922. doi:10.1093/gbe/evz133
- Poelstra JW, Vijay N, Bossu CM, Lantz H, Ryll B, Müller I, Baglione V, Unneberg P, Wikelski M, Grabherr MG, Wolf JBW. 2014. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science* (80- ) **344**:1410–4. doi:10.1126/science.1253226
- Poelstra JW, Vijay N, Hoepfner MP, Wolf JBW. 2015. Transcriptomics of colour patterning and coloration shifts in crows. *Mol Ecol* **24**:4617–4628. doi:10.1111/mec.13353
- R Core Team. 2015. R: A language and environment for statistical computing. <http://www.r-project.org>.
- Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**:W160–W165. doi:10.1093/nar/gkw257
- Rice AM, McLysaght A. 2017. Dosage-sensitive genes in evolution and disease. *BMC Biol* **15**:1–10. doi:10.1186/s12915-017-0418-y
- Richard G, Legeai F, Prunier-Leterme N, Bretaudeau A, Tagu D, Jaquière J, Le Trionnaire G. 2017. Dosage compensation and sex-specific epigenetic landscape of the X chromosome in the pea aphid. *Epigenetics and Chromatin* **10**:1–17. doi:10.1186/s13072-017-0137-1
- Rob J. Hyndman, Yeasmin Khandakar. 2019. Automatic Time Series Forecasting: The forecast Package for R. *J Stat Softw* **27**:22.
- Rutkowska J, Lagisz M, Nakagawa S. 2012. The long and the short of avian W chromosomes: no evidence for gradual W shortening. *Biol Lett* **8**:636–638. doi:10.1098/rsbl.2012.0083
- Samata M, Akhtar A. 2018. Dosage Compensation of the X Chromosome: A Complex Epigenetic Assignment Involving Chromatin Regulators and Long Noncoding RNAs, Annual Review of Biochemistry. doi:10.1146/annurev-biochem-062917-011816
- Schild DR, Card DC, Hales NR, Perry BW, Pasquesi GM, Blackmon H, Adams RH, Corbin AB, Smith CF, Ramesh B, Demuth JP, Betrán E, Tollis M, Meik JM, Mackessy SP, Castoe TA. 2019. The origins and evolution of chromosomes, dosage compensation, and mechanisms underlying venom regulation in snakes. *Genome Res* **29**:590–601. doi:10.1101/gr.240952.118
- Smeds L, Kawakami T, Burri R, Bolivar P, Husby A, Qvarnström A, Uebbing S, Ellegren H. 2014. Genomic identification and characterization of the pseudoautosomal region in highly differentiated avian sex chromosomes. *Nat Commun* **5**:1–7. doi:10.1038/ncomms6448
- Smit A, Hubley R, Green P. 2015. RepeatMasker. <http://www.repeatmasker.org>.
- Smit A, Hubley R, Green P. 2013. RepeatMasker.
- Smith G, Chen Y-RR, Blissard GW, Briscoe AD. 2014. Complete dosage compensation and sex-biased gene expression in the moth Manduca sexta. *Genome Biol Evol* **6**:526–537. doi:10.1093/gbe/evu035

- Sun D, Maney DL, Layman TS, Chatterjee P, Yi S V. 2019. Regional epigenetic differentiation of the Z Chromosome between sexes in a female heterogametic system. *Genome Res* **29**:1673–1684. doi:10.1101/gr.248641.119
- Taudt A, Colomé-Tatché M, Johannes F. 2016. Genetic sources of population epigenomic variation. *Nat Rev Genet* **17**:319–332. doi:10.1038/nrg.2016.45
- Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol* **31**:46–53. doi:10.1038/nbt.2450
- Uebbing S, Künstner A, Mäkinen H, Ellegren H. 2013. Transcriptome sequencing reveals the character of incomplete dosage compensation across multiple tissues in flycatchers. *Genome Biol Evol* **5**:1555–1566. doi:10.1093/gbe/evt114
- Veselovska L, Smallwood SA, Saadeh H, Stewart KR, Krueger F, Maupetit Méhouas S, Arnaud P, Tomizawa S ichi, Andrews S, Kelsey G. 2015. Deep sequencing and de novo assembly of the mouse oocyte transcriptome define the contribution of transcription to the DNA methylation landscape. *Genome Biol* **16**:1–17. doi:10.1186/s13059-015-0769-z
- Vicoso B, Bachtrog D. 2011. Lack of global dosage compensation in *Schistosoma mansoni*, a female-heterogametic parasite. *Genome Biol Evol* **3**:230–235. doi:10.1093/gbe/evr010
- Vicoso B, Emerson JJ, Zektser Y, Mahajan S, Bachtrog D. 2013a. Comparative Sex Chromosome Genomics in Snakes: Differentiation, Evolutionary Strata, and Lack of Global Dosage Compensation. *PLOS Biol* **11**:e1001643. doi:10.1371/journal.pbio.1001643
- Vicoso B, Kaiser VB, Bachtrog D. 2013b. Sex-biased gene expression at homomorphic sex chromosomes in emus and its implication for sex chromosome evolution. *Proc Natl Acad Sci* **110**:6453–6458. doi:10.1073/pnas.1217027110
- Vijay N, Bossu CM, Poelstra JW, Weissensteiner MH, Suh A, Kryukov AP, Wolf JBW. 2016. Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. *Nat Commun* **7**:13195. doi:10.1038/ncomms13195
- Warmuth VM, Weissensteiner MH, Wolf JBW. 2021. Ineffective silencing of transposable elements on the avian W chromosome. *submitted*.
- Warnefors M, Mössinger K, Halbert J, Studer T, VandeBerg JL, Lindgren I, Fallahshahroudi A, Jensen P, Kaessmann H. 2017. Sex-biased microRNA expression in mammals and birds reveals underlying regulatory mechanisms and a role in dosage compensation. *Genome Res* **27**:1961–1973. doi:10.1101/gr.225391.117
- Weissensteiner MH, Poelstra JW, Wolf JBW. 2015. Low-budget ready-to-fly unmanned aerial vehicles: an effective tool for evaluating the nesting status of canopy-breeding bird species. *J Avian Biol* **46**:425–430. doi:10.1111/jav.00619
- Werner MS, Sieriebriennikov B, Prabh N, Loschko T, Lanz C, Sommer RJ. 2018. Young genes have distinct gene structure, epigenetic profiles, and transcriptional regulation. *Genome Res* **28**:1675–1687. doi:10.1101/gr.234872.118
- Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.
- Wolf JBW, Bryk J. 2011. General lack of global dosage compensation in ZZ/ZW systems? Broadening the perspective with RNA-seq. *BMC Genomics* **12**:1–10. doi:10.1186/1471-2164-12-91
- Wright AE, Zimmer F, Harrison PW, Mank JE. 2015. Conservation of Regional Variation in Sex-Specific Sex Chromosome Regulation. *Genetics* **201**:587–598. doi:10.1534/genetics.115.179234
- Wu C-C, Klaesson A, Buskas J, Ranefall P, Mirzazadeh R, Söderberg O, Wolf JBW. 2019. In situ quantification of individual mRNA transcripts in melanocytes discloses gene regulation of relevance to speciation. *J Exp Biol* **222**:jeb194431. doi:10.1242/jeb.194431
- Wu H, Caffo B, Jaffee HA, Irizarry RA, Feinberg AP. 2010. Redefining CpG islands using hidden Markov models. *Biostatistics* **11**:499–514. doi:10.1093/biostatistics/kxq005



- Xu L, Auer G, Peona V, Suh A, Deng Y, Feng S, Zhang G, Blom MPK, Christidis L, Prost S, Irestedt M, Zhou Q. 2019a. Dynamic evolutionary history and gene content of sex chromosomes across diverse songbirds. *Nat Ecol Evol* **3**:834–844. doi:10.1038/s41559-019-0850-1
- Xu L, Wa Sin SY, Grayson P, Edwards S V, Sackton TB. 2019b. Evolutionary Dynamics of Sex Chromosomes of Paleognathous Birds. *Genome Biol Evol* **11**:2376–2390. doi:10.1093/gbe/evz154
- Xu M, Millar SE. 2020. How a Bird Gets Its Feathers: Insights from Chromatin Looping. *Dev Cell* **53**:493–495. doi:10.1016/j.devcel.2020.05.008
- Zha XF, Xia QY, Duan J, Wang CY, He NJ, Xiang ZH. 2009. Dosage analysis of Z chromosome genes using microarray in silkworm, *Bombyx mori*. *Insect Biochem Mol Biol* **39**:315–321. doi:10.1016/j.ibmb.2008.12.003
- Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**:R137. doi:10.1186/gb-2008-9-9-r137
- Zhou QQ, Zhang J, Bachtrog D, An NN, Huang Q, Jarvis ED, Gilbert MTP, Zhang G. 2014. Complex evolutionary trajectories of sex chromosomes across bird taxa. *Science (80- )* **346**:1246338–1246338. doi:10.1126/science.1246338
- Zuur FA, Ieno NE, Walker N, Saveliev AA, Smith MG. 2009. Mixed effect models and extensions in ecology with R. Springer.



## **Supplementary Material**

### **Epigenetic mechanisms of partial dosage compensation in an avian, female heterogametic system**

Ana Catalán<sup>\* 1,2</sup>, Justin Merondun<sup>2</sup>, Ulrich Knief<sup>2</sup> and Jochen B.W. Wolf<sup>1,2</sup>.

*Supplementary Text (ATAC-seq quality control)* *pages 2*

*Supplementary Figures* *pages 3-21*

*Supplementary References* *page 22*

*Supplementary Tables* are provided as external files

## Supplementary Text

### Pre-processing of ATAC-seq data

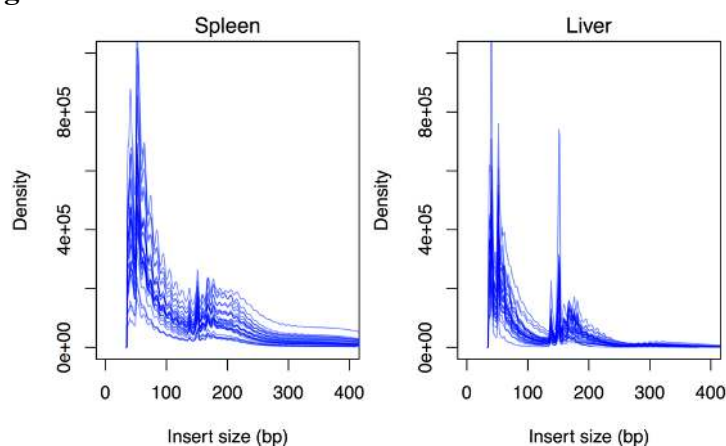
Quality control of ATAC-seq data was performed as described in the main document. Here, we provide details for each of the components. For each library we first examined the insert size distribution which should show a downward laddering pattern reflecting the amount and length of DNA fragments from nucleosome-free region and from regions with associated nucleosomes (Buenrostro et al., 2013). For all libraries, we observed an enrichment of fragments corresponding to the nucleosome-free region that ranged in length from 36 – 130 bp. The second enriched fragment observed, ranged from 150 - 250 bp in both organs, which corresponds to the mononucleosome-bound region (**Figure SA**). The detection of a mononucleosome ensures the boundaries for a successful detection of nucleosome-free regions, where putatively, the regulatory regions lie.

Next, we quantified the correlation between our two technical and biological replicates (8 females, 7 males for liver and spleen). For the technical replicates, Spearman correlation coefficients drawn from ATAC-seq coverage estimates ranged from 0.86 – 0.97 for liver and 0.72 – 0.98 for spleen (**Figure SB**). For biological replicates, Spearman correlation coefficients ranged from 0.87 to 0.95, where correlation was higher within the same sex for both tissues. The number of total usable reads (see methods), varied among samples (**Supplementary Table SB**). Following the ENCODE guidelines for ATAC-seq quality, 50 million fragments are recommended for paired-end experiments. Due to a high correlation between technical replicates, we pooled these for downstream analysis, resulting in ~60 million reads for the smallest biological sample in liver and ~92 million reads in spleen.

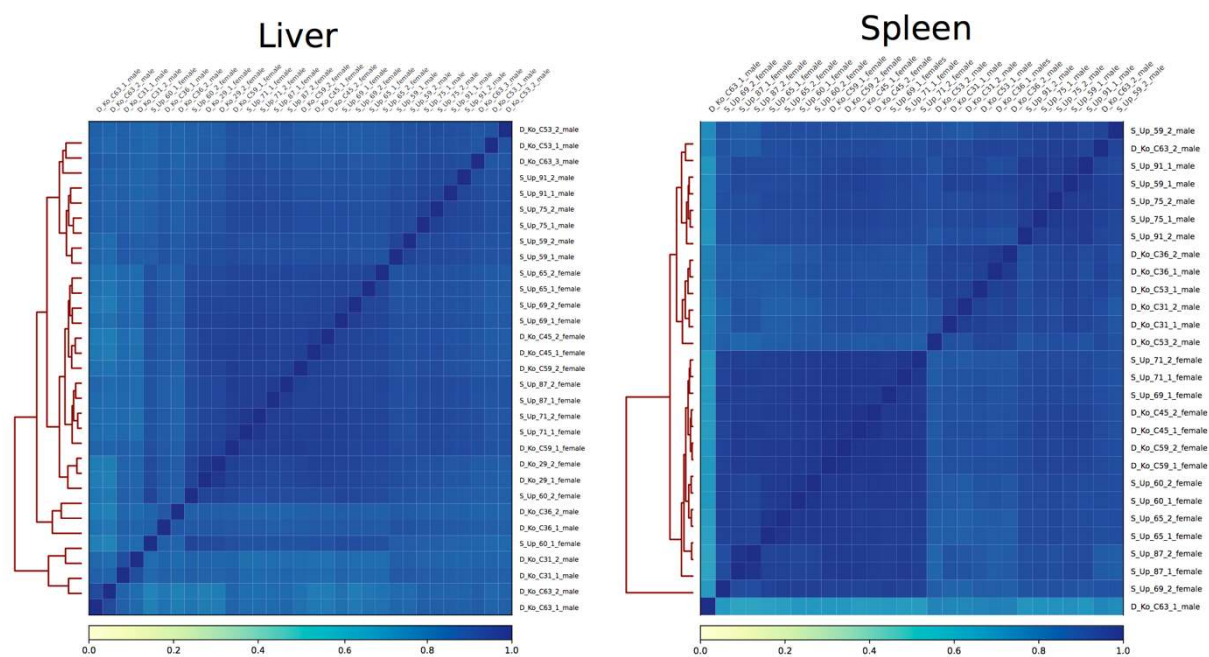
We also revised the FRiP (Fragments of Reads in Peaks) score values for each pooled sample, which ranged from ~12 – 51% (**Figure SC**) and which evaluates the number of reads that were mapped into peaks. According the ENCODE guidelines, FRiP values exceeding 0.2 are acceptable. From our sample assessment, only one sample fell below this threshold which was accordingly removed from all subsequent analyses.

We further checked for an enrichment of mapped reads in putative transcription start sites (TSS) of the expressed genes which we defined as genes with FPKM expression levels exceeding 1. The putative TSS was identified as 2kb upstream and downstream of the first nucleotide of the expressed gene. Wig files were produced for each sample and heatmaps showing coverage enrichment around the TSS were done with Deeptools v.3.5.0 (deepTools - computeMatrix -b 2000 -a 2000) (Ramírez et al., 2016). **Figure SD**, shows two examples, one of liver and one of spleen, of coverage enrichment around the TSS (+/- 2kb) (**Figure SD**).

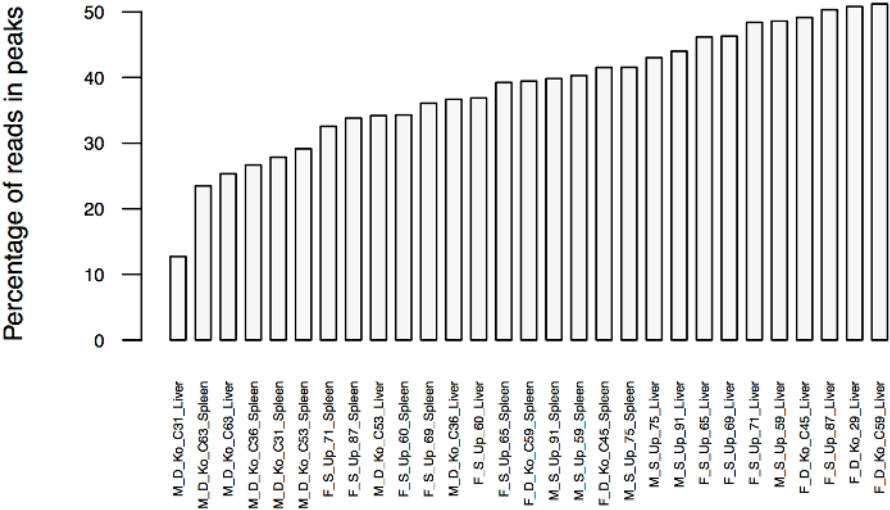
## Supplementary figures



**Figure SA.** ATAC-seq insert size distribution for all biological replicates for spleen (left) and liver (right). The position of the first nucleosome can be observed between ~150-250 bp.

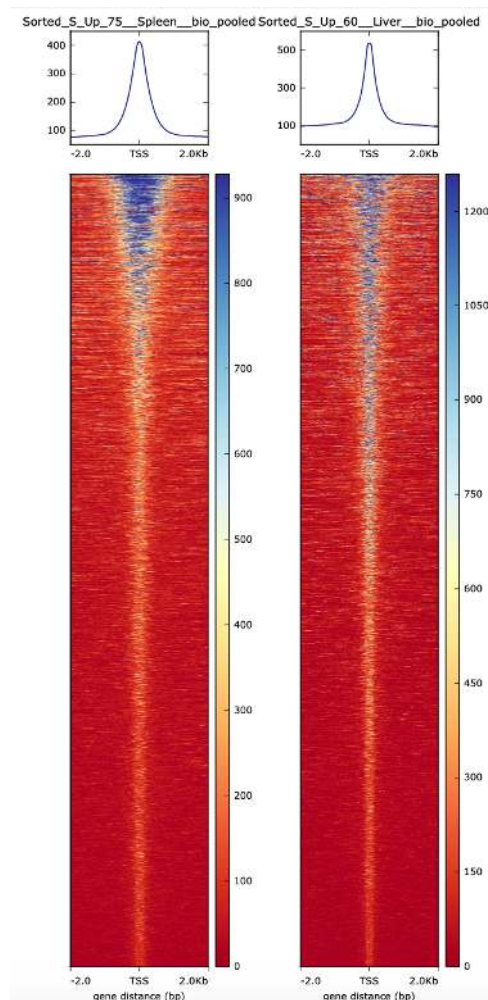


**Figure SB.** Across sample correlation between technical and biological replicates using genome coverage as the correlation parameter in liver and spleen.

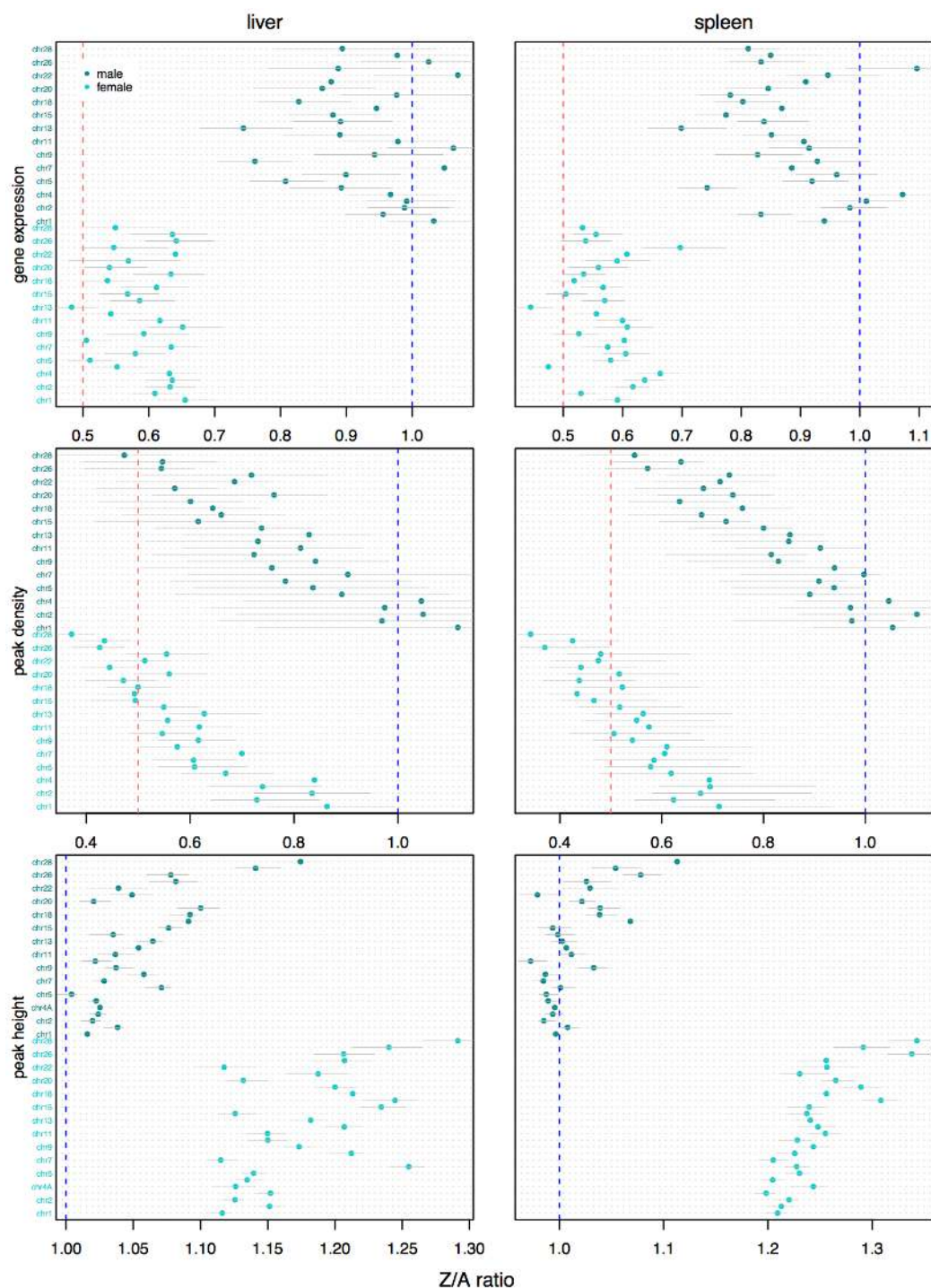


**Figure SC.** Percentage of reads mapped in reads for pooled samples in liver and spleen. (blue). Vertical dashed lines show the median expression for both density distributions.

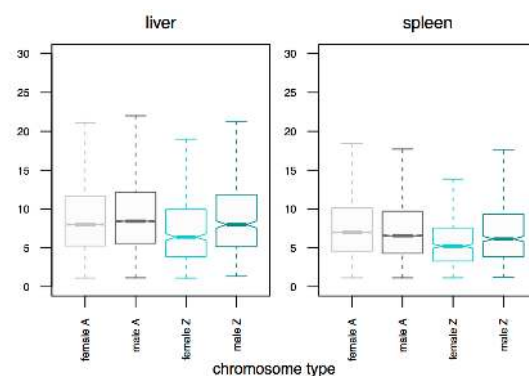




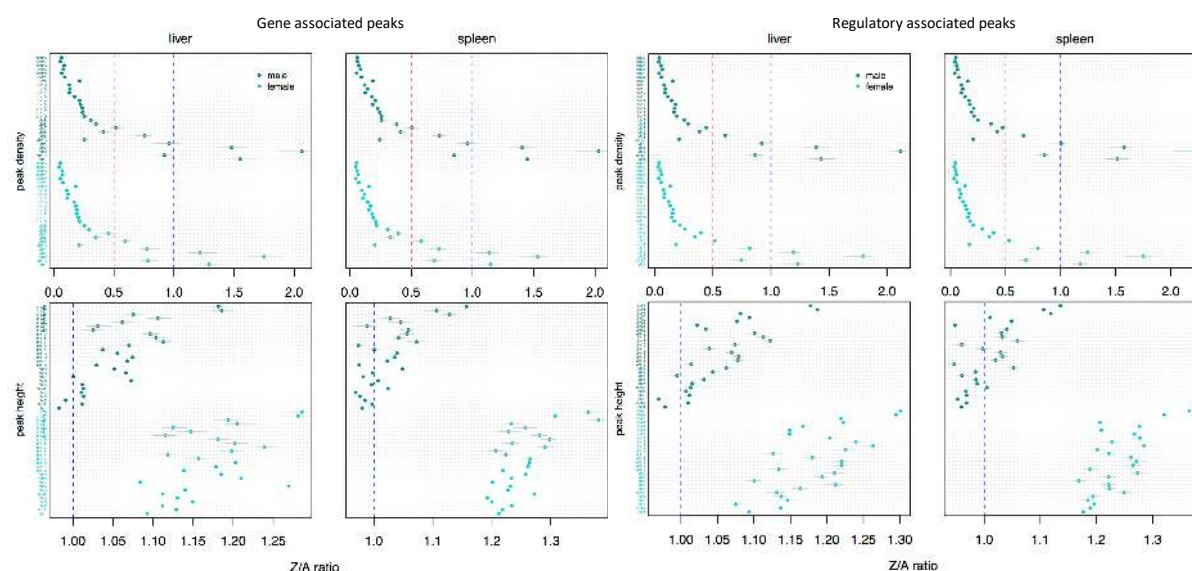
**Figure SD.** Upper panels: summary of coverage density +/- 2kb of the TSS, in spleen (left) and liver (right). Bottom panels: Enrichment of mapped reads presented as heatmaps along the defined TSS, in spleen (left) and liver (liver).



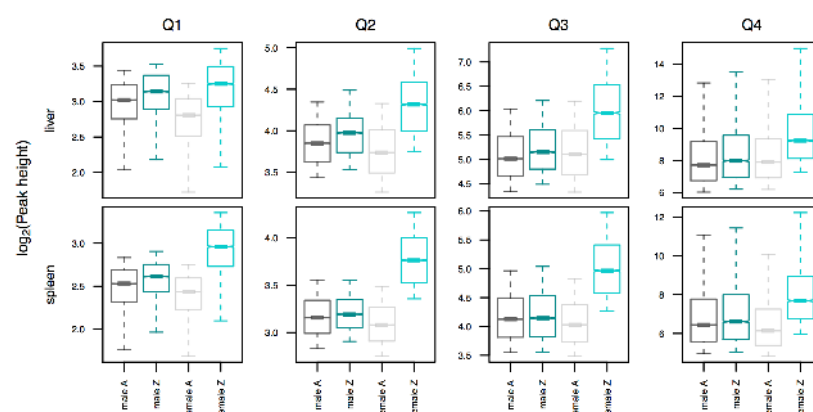
**Figure SE.** Raw Z:A ratio for gene expression, peak density (chromosome-wide) and peak height (chromosome-wide) shown for each autosome in females (light blue) and males (dark blue). 95% confidence intervals drawn from 10,000 bootstraps.



**Figure SF.** Boxplot showing the distribution of the number of associated peaks per expressed gene in the autosomes and the Z chromosome in a region of 20 kb upstream and downstream of the focal gene. Y-lab indicates the number of peaks associated to a gene.

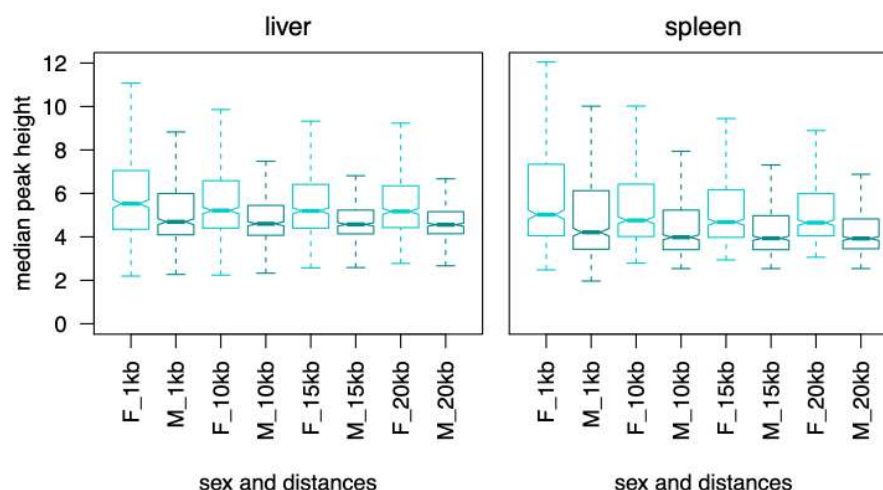


**Figure SG.** Z:A ratio for peak density and peak height shown for each autosome in females (light blue) and males (dark blue). 95% confidence intervals drawn from 10,000 bootstraps. Left panel: shows gene-centered Z:A ratios. Right panel: shows Z:A only in up- and downstream regions of expressed genes excluding the gene body. Note the strong dependence on chromosome length.

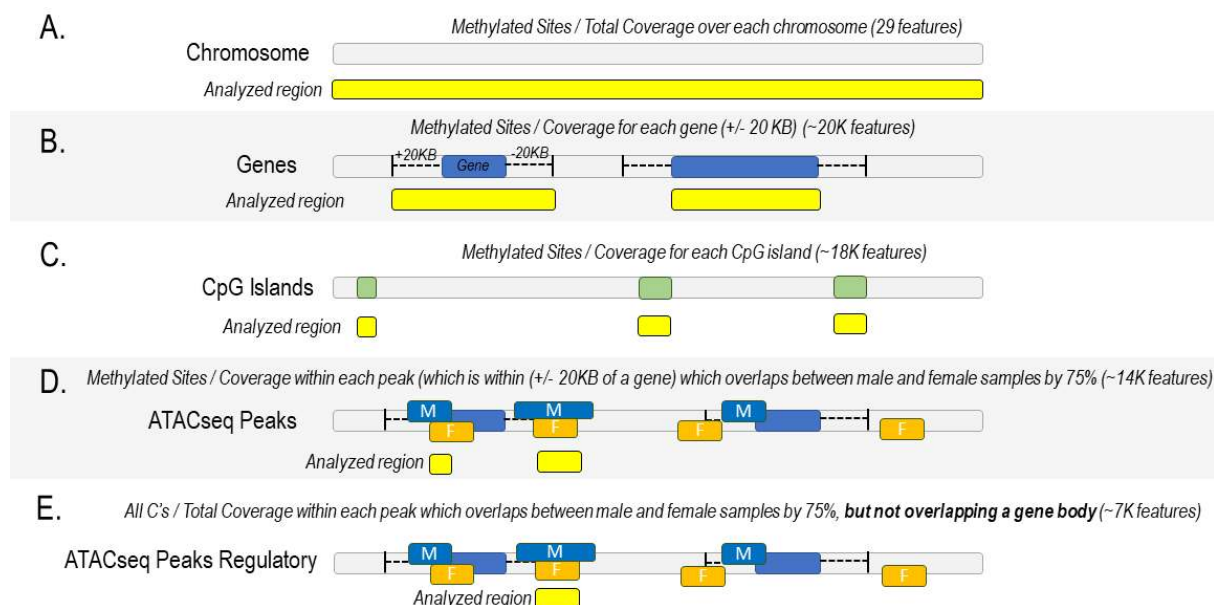


**Figure SH.** Boxplots presenting  $\log_2$  gene-centred peak height distribution broken down into four quantiles (Q1 – Q4). A: autosomes, Z: Z chromosome. Upper panel liver, bottom panel spleen.

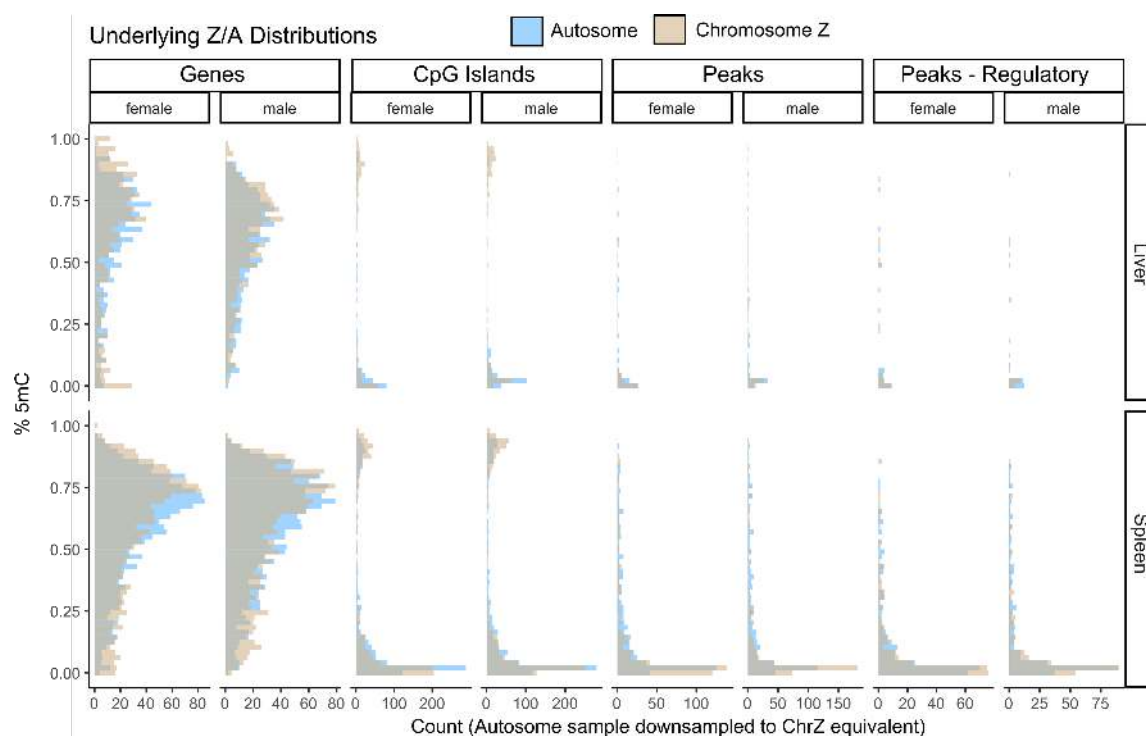




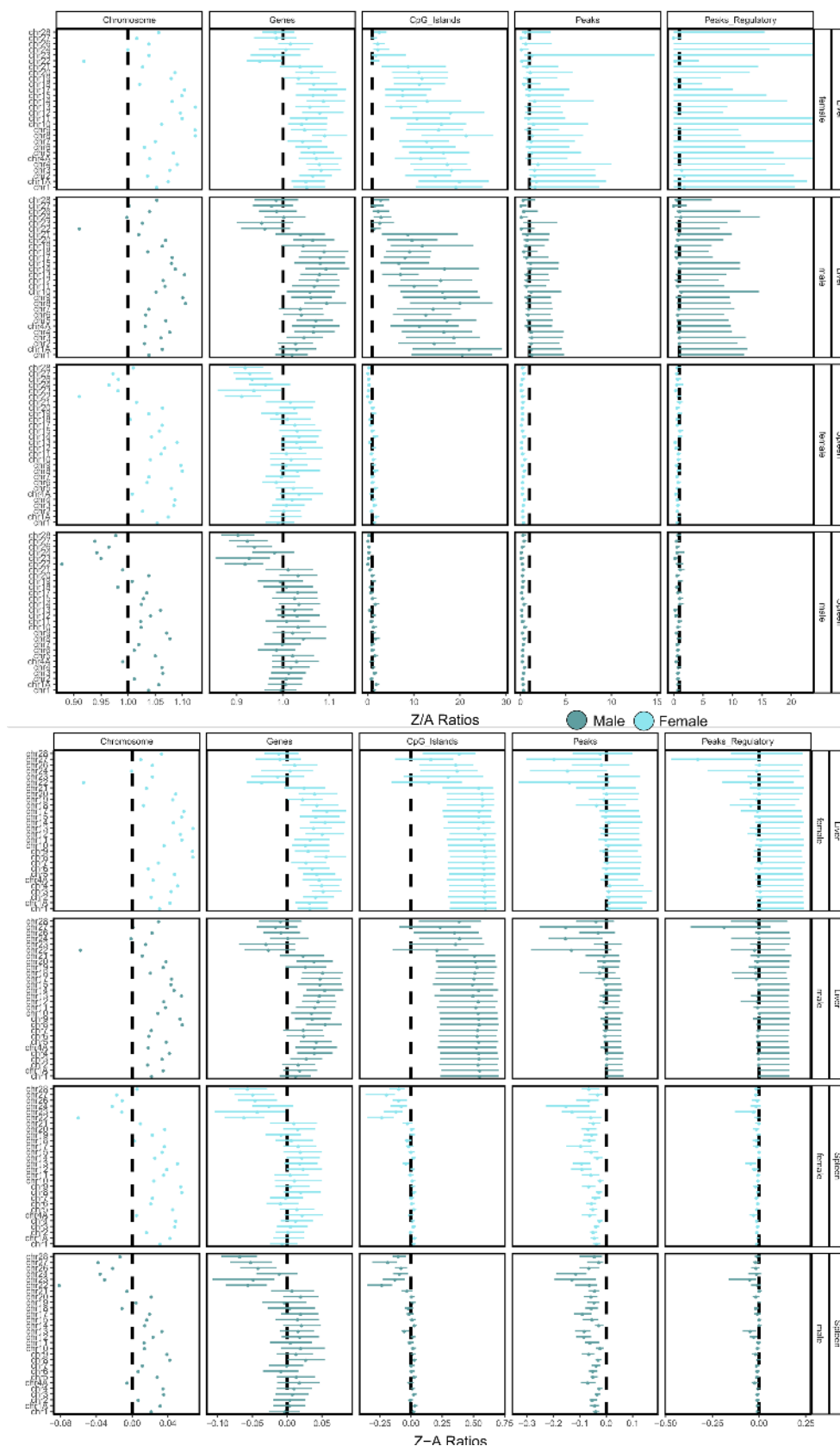
**Figure SI.** Boxplots showing gene-centred peak height distribution including differently sized regions around the gene body (1 kb, 10 kb, 15 kb and 20 kb) in females (F) and males (M) on the Z chromosome.



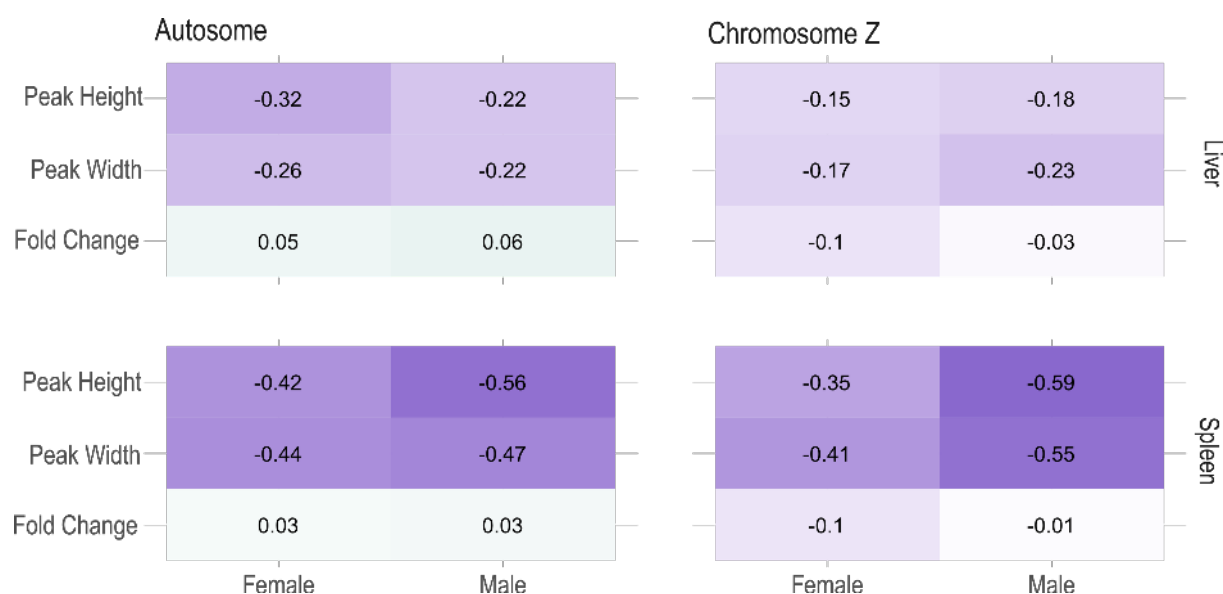
**Figure SJ.** Cartoon depicting the five feature-based analyses used to examine 5mC DNA methylation. DNA methylation levels for each feature were calculated as the cumulative proportion of methylated reads (sum of methylated reads divided by the total number of reads per feature).



**Figure SK.** Frequency distributions of 5mC CpG DNA methylation underlying data in **Figure SL** and in **Supplementary Table A** for four of the feature categories (chromosome excluded because of a single feature unit for Chromosome Z). Autosomal distributions were randomly subsampled to be equivalent to the number of features present in the Z chromosome.

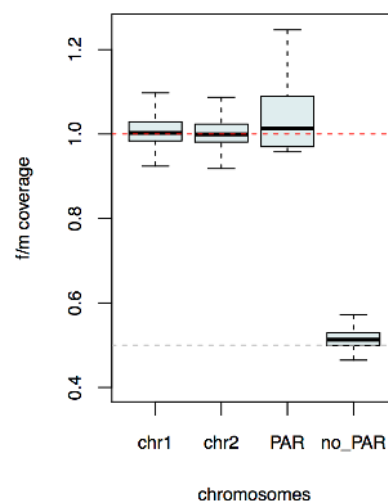


**Figure SL.** 5mC CpG DNA methylation differences between chromosome Z and autosomal chromosomes for both sexes and tissues. Distributions were drawn from 10,000 bootstrap replicate samplings of the median. **A.** Ratio of DNA methylation on chromosome Z relative to the chromosome along the Y axis. Note variable X axes across facets. **B.** Same as panel A, except indicating the absolute difference in methylation (Chromosome Z – Chromosome N).

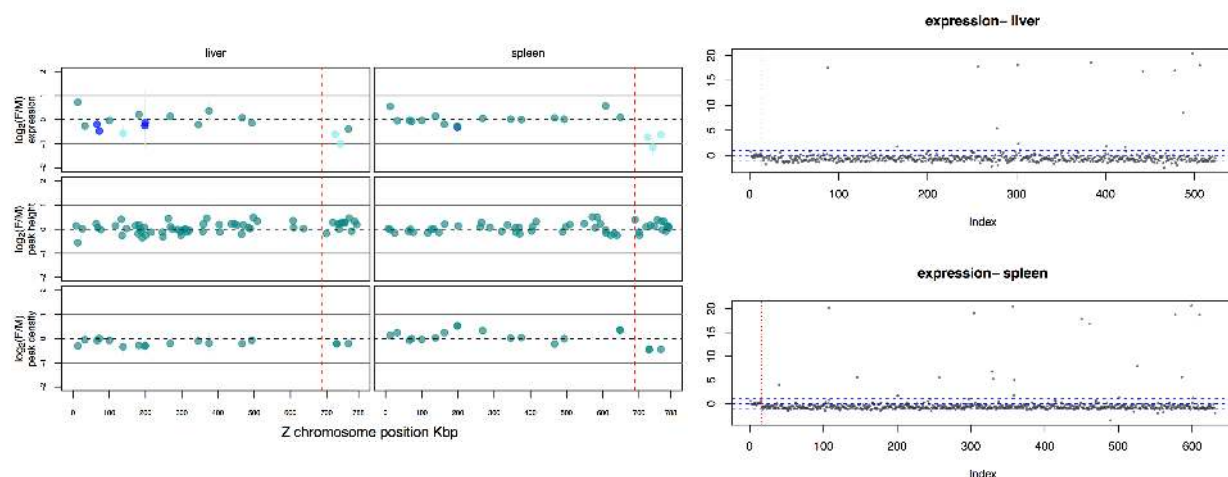


**Figure SM.** Spearman correlations between 5mC CpG methylation level and gene expression (FPKM) and open chromatin (peak height, peak width), calculated within each feature for each sex and tissue individually.

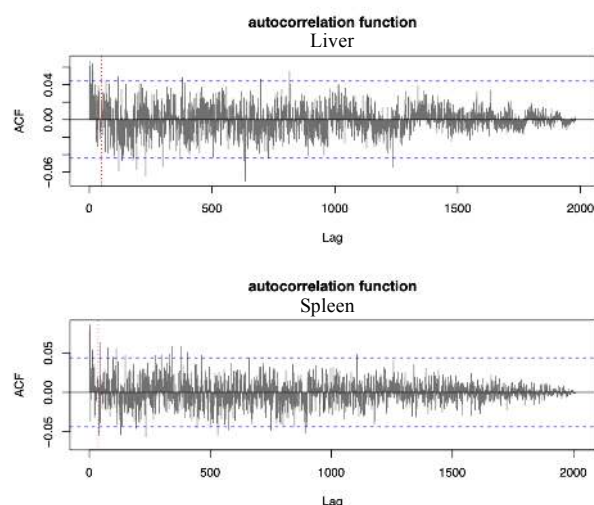




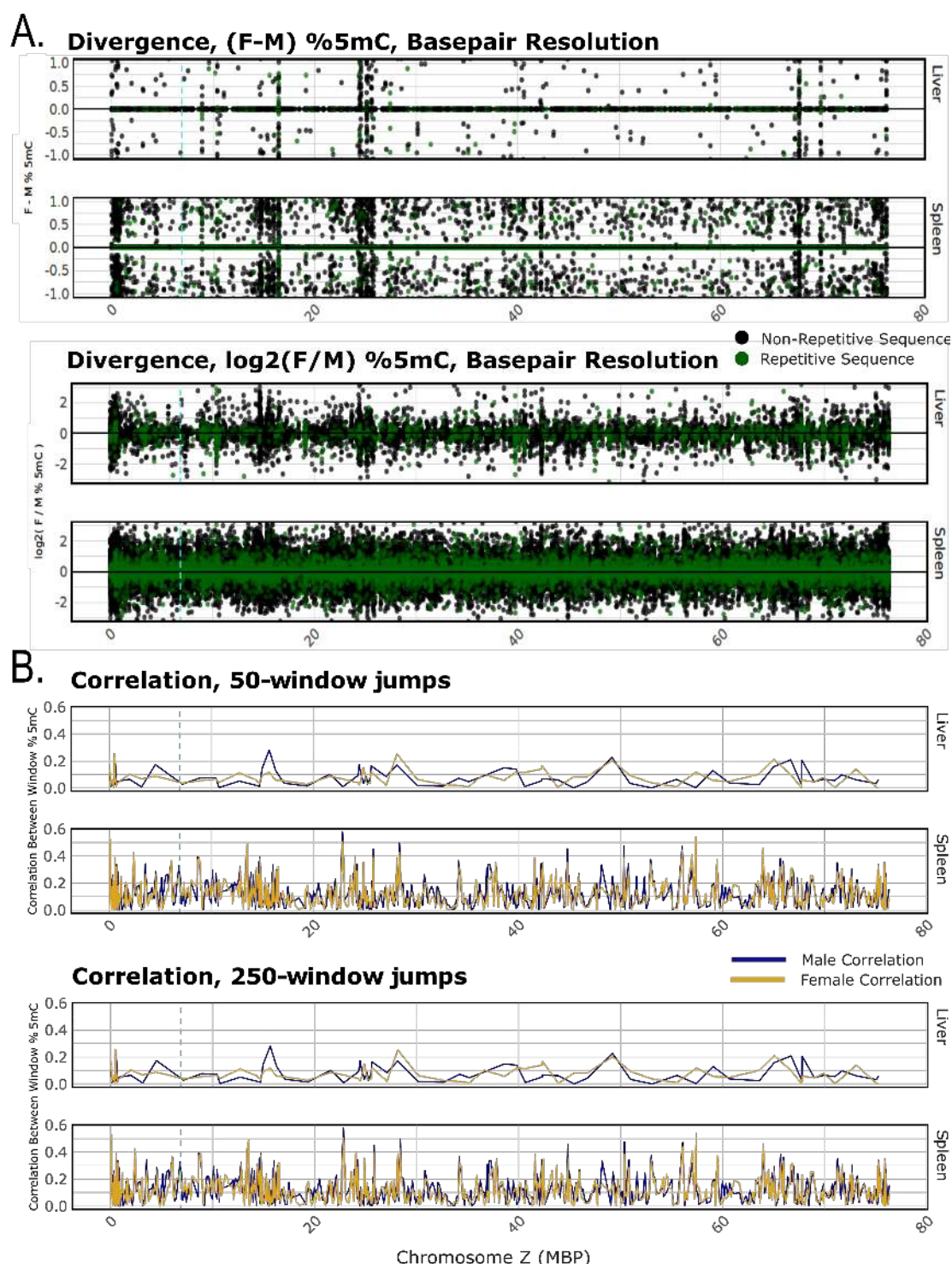
**Figure SN.** Boxplot of female to male read coverage distribution from whole genome shotgun sequencing of one individual of each sex as calculated in 50 kb windows, in chromosome 1 (chr1), chromosome 2 (chr2), the pseudoautosomal region (PAR) and the Z chromosome without the PAR. F/M coverage ratio on the shown autosomes and the PAR lie near a ratio of 1 (red line), whereas the hemizygous part of the Z lies close to a ratio of 0.5 (grey line).



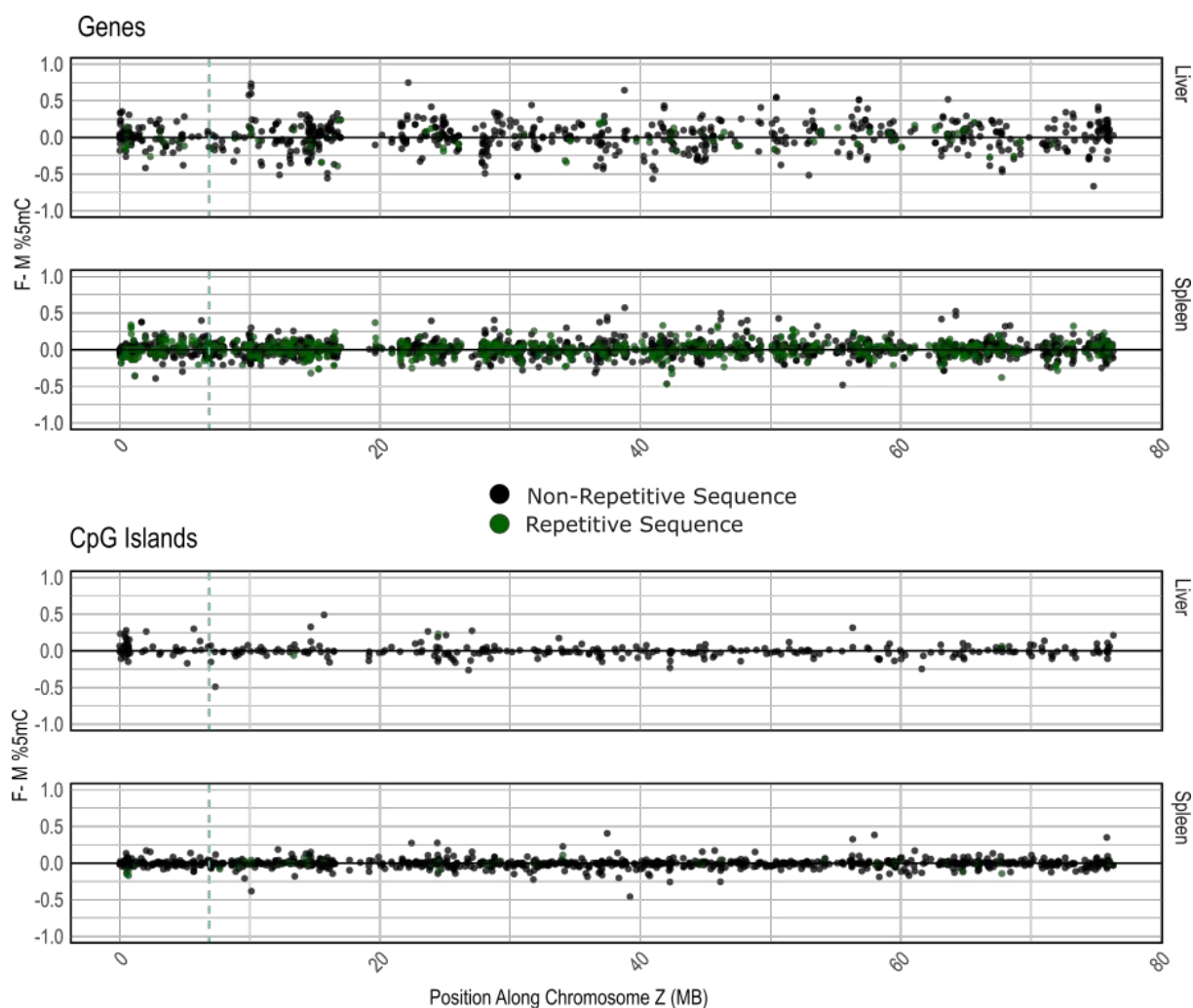
**Figure S0.** Left panel: The PAR region on the Z chromosome is shown in spleen and liver. Upper panel:  $\log_2(F/M)$  gene expression. Middle panel:  $\log_2(F/M)$  peak height. Bottom panel:  $\log_2(F/M)$  peak number. Right panel:  $\log_2(F/M)$  gene expression of all genes expressed in the Z. Red dotted vertical line marks the PAR.



**Figure SP.** Autocorrelation of gene expression present on the Z chromosome. Red vertical line shows the PAR border. Blue lines represent autocorrelation values significantly different from zero. Except for the PAR region, there is no clear trend of clusters that would point at dosage local centers of dosage compensation.

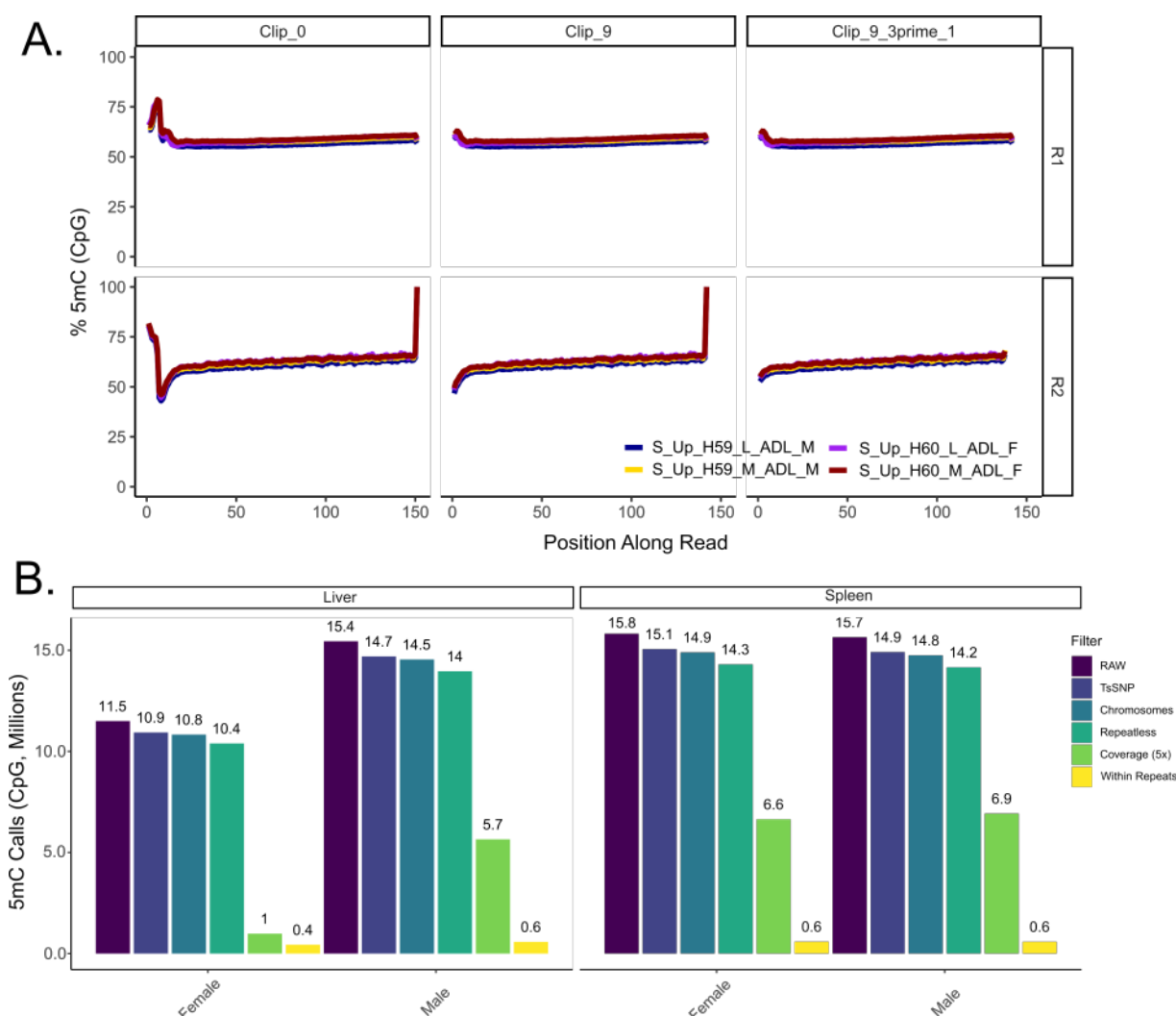


**Figure SQ.** Sex-specific DNA methylation patterning along the Z chromosome at base-pair resolution calculated for both tissues and sexes. **A.** Divergence values shown for the difference in 5mC methylation between females and males (f-m; top) and the ratio between sexes ( $\log_2(f/m)$ ; bottom) along the Z chromosome suggest interspersed blocks of increased divergence, concomitant for both sexes. There was no evidence of a male-only hypermethylated region. **B.** Genomic autocorrelation between  $n$ -sized base-pair blocks. Spearman correlations were calculated between the first  $n/2$  and last  $n/2$  methylation calls in a block. Blocks were analyzed in both 50 and 250 base-pair blocks. We hypothesized that an MHM region in crow would exhibit male-specific hypermethylation, and strong male genomic autocorrelation.



**Figure SR.** Sex-specific DNA methylation patterning along the Z chromosome shown as the difference between female and male methylation levels (f-m). Similar to Figure SQ, except instead of base-pair resolution, we examined DNA methylation divergence over genes and CpG islands.





**Figure SS.** Results of quality control for whole-genome bisulfite-sequencing libraries for male and female hooded crow samples, created for both spleen and liver. Panel **A** indicates methylation levels averaged across all reads in a library along read position, used to identify systematic methylation biases along read lengths, with the far-right panel indicating our final cleaned reads. As our libraries were created with a post-bisulfite conversion kit, we trimmed the first 9 reads from the 5' end, alleviating this known bias. We then iteratively revisited this step by trimming the last base from the 3' end, due to the elevated 5mC levels in the last position on the R2 read. Sample name: Individual: S\_Up\_H59/H60; tissue:liver (L) or spleen (M); age: adult (ADL); sex: male (M) or females (F)). Panel **B** indicates total 5mC positions called for each sample, as well as sites retained after filtering for C-T and G-A SNPs, retaining sites on assigned chromosomes, removing sites within identified repeats, and filtering for 5x coverage.

## References

- Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**:1213–8. doi:10.1038/nmeth.2688
- Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**:W160–W165. doi:10.1093/nar/gkw257