SUPPLEMENTAL MATERIAL

Ineffective silencing of transposable elements on the avian W Chromosome

Vera M. Warmuth[1*], Matthias H. Weissensteiner[1,2], Jochen B. W. Wolf[1]

[1] Division of Evolutionary Biology, Faculty of Biology, LMU Munich, Planegg-Martinsried, Germany

[2] Department of Biology, Penn State University, University Park, PA, USA

Contents

Supplemental Tables

- Supplemental Table 1: Supplemental_Table_S1.xlsx
- Supplemental Table 2: Supplemental_Table_S2.xlsx
- Supplemental Table 3: Supplemental_Table_S3.xlsx
- Supplemental Table 4: Supplemental_Table_S4.xlsx
- Supplemental Table 5: Supplemental_Table_S5.xlsx
- Supplemental Table 6: Supplemental_Table_S6.xlsx
- Supplemental Table 7: Supplemental_Table_S7.xlsx
- Supplemental Table 8: Supplemental_Table_S8.xlsx

Supplemental Text

Section 1. Taxonomic considerations

The two taxa considered here - carrion crows (*Corvus (corone) corone*) and hooded crows (*Corvus (corone) cornix*) are regarded as separate species by some (Parkin, Collinson, Helbig, Knox, & Sangster, 2003). Recent genome-wide evidence, however, rather supports a single species with two segregating colour morphs (Knijff, 2014). To confirm the absence of a taxon-specific effect on TE expression, we tested for differential expression between taxa, accounting for differences due to sex and tissue. Only six TEs were found to be differentially expressed (see results), consistent with previous work reporting few gene expression differences except for melanogenesis genes expressed feather follicles (Poelstra, Vijay, Hoeppner, & Wolf, 2015). For the purpose of this study, we therefore consider carrion and hooded crows as members of the same species.

Section 2. Quantification of TE-derived transcripts

Quantification of TE-derived transcripts using short read data suffers from potential ambiguities introduced by multiple mapping. In recent years, a number of approaches have been proposed to minimise this problem, including analysis of TEs at higher taxonomic levels, use of iterative methods to assign multi-mapping reads (Jin et al., 2015; Yang et al., 2019), and/or focus on uniquely mapping reads only (Chuong et al., 2013). In their recent simulation-based study, Teissandier and colleagues (2019) provide a set of best practices for the alignment and quantification of TE transcripts. First, they show that a focus on uniquely mapping reads biases analyses towards older elements with higher (unique) mappability. This is consistent with previous work where analysis of uniquely mapping reads performed worse than an expectation-maximisation algorithm (Yang et al., 2019). Second, the authors establish one of two approaches as best practice for the analysis of TE transcription, i) fractional counting of (all) multi-mapped reads, or ii) randomly reporting and counting of only one valid alignment. Using both these approaches yielded true positive rates of 96% and 93% for paired-end (PE) and single-end (SE) libraries, respectively (Teissandier et al., 2019). A difference between SE and PE libraries was also noted for mapping percentage, with a gain of 2 and 6% in human and mouse, respectively, and an up to 30% gain for some LINE elements in both species when using PE libraries.

Taking the simulated outcome of Teissandier et al. as a guideline, we might expect 2-5% of the mapped (single-end) reads to be false positives. This figure might be higher, if we consider differences in read length (50bp in this study versus 100bp simulated by Teissandier et al. 2019). To assess the extent to which the observed pattern of female bias is an artefact of the random approach to mapping and read counting, we repeated all analyses using the unique approach. Both approaches reveal an overall pattern of female-biased TE transcription, similar numbers of TEs with a sex-biased expression, and a dominance of W-linked TEs among female-biased DETEs. We are therefore confident that the pattern of female-biased TE expression reported here is genuine.

Supplemental Methods

TE content vs. TE transcript abundance: effect of sex-related differences in ploidy

The genome-wide abundance of TEs on the sex chromosomes is naturally expected to differ between males and females owing to differences in ploidy. In crows, dosage compensation is incomplete (Wolf & Bryk, 2011). Therefore, TEs transcribed from the Z Chromosome should be more abundant in males (ZZ) than in females (ZW). The reverse is true for the W Chromosome, which is limited to female birds and is additionally extremely repeat-rich. If variation in TE abundance is closely correlated with differential expression, the presence of a TE-rich W chromosome in females, but not males, could influence the pattern of female-biased TE expression we observe in our data (Shao et al., 2019).

Assuming equal regulation of TE expression between sexes and chromosomes, genomic TE abundance should predict sex-specific TE transcript abundance, i.e. autosomal TEs should have equal transcript abundance in both sexes, Z-linked TEs should have higher transcript abundance in males, and W-linked TEs should only be transcribed in females. Significant deviations from such ploidy-based expectations would indicate sex- and/or chromosome specific regulation. To explore this question, we fit a linear model to log-transformed count data, using sex, tissue, and nucleic acid type (RNA or DNA) as categorical explanatory variables for TE expression (formula: log(counts) ~ sex + tissue + type + sex * type). In this model, the statistical interaction between sex and nucleic acid type reveals whether, and to what extent, RNA read counts (*i.e.* transcription) differ between the sexes above and beyond differences in DNA read counts (*i.e.* genomic abundance). To avoid ambiguities due to mis-mapping reads, we focus here only on uniquely mapping reads (both DNA-seq and RNA-seq).

Read counts of autosomal (no ploidy differences) DETEs were similar between the sexes on the DNA level (**Supplemental Fig. S4**, top panel), but significantly higher in females than in males on the RNA level in all three tissues (**Supplemental Fig. S4**, lower panels). The $type_{RNA}$*sex_{male} interaction was negative and highly significant (p< 3.3e-07), corresponding to significantly lower transcription of autosomal DETEs in males compared to females, regardless of tissue type. This suggests less effective silencing of autosomal DETEs in females.

DNA-Seq read counts of Z-based DETEs were higher in males (ZZ) than in females (ZW), in line with expectations from differences in ploidy (**Supplemental Fig. S4**, top panel). RNA-seq read counts were higher in males than in females in the two somatic tissues. Overall, the interaction between sex and type ($type_{RNA}$*sex_{male}) was not significant (p=0.7). However, TE expression in gonads showed the reverse of what would be expected based on ploidy, with female expression exceeding that of males (**Supplemental Fig. S4**, lower panels). Elevated transcription of Z-chromosomal DETEs in female tissue is unexpected, even if we assume partial dosage compensation (Wolf & Bryk, 2011).

On the W-chromosome, both DNA- and RNA-seq read counts were higher in females than in males. In addition, a handful of male DNA-seq reads mapped to the W Chromosome, despite considering only uniquely mapping reads. This likely reflects the presence of highly similar TE copies elsewhere in the genome (**Supplemental Fig. S4**, top panel). RNA-seq reads were high across all female tissues, but near absent in males (**Supplemental Fig. S4**, lower panels). The $type_{RNA}$*sex_{male} interaction was highly significant (p<6.5e-11), indicating a more pronounced difference between males and females on the RNA than on the DNA level. These analyses, together with the results from the differential expression analysis confirm a major contribution of W-chromosomal TEs to female-biased TEs

Supplemental Figures

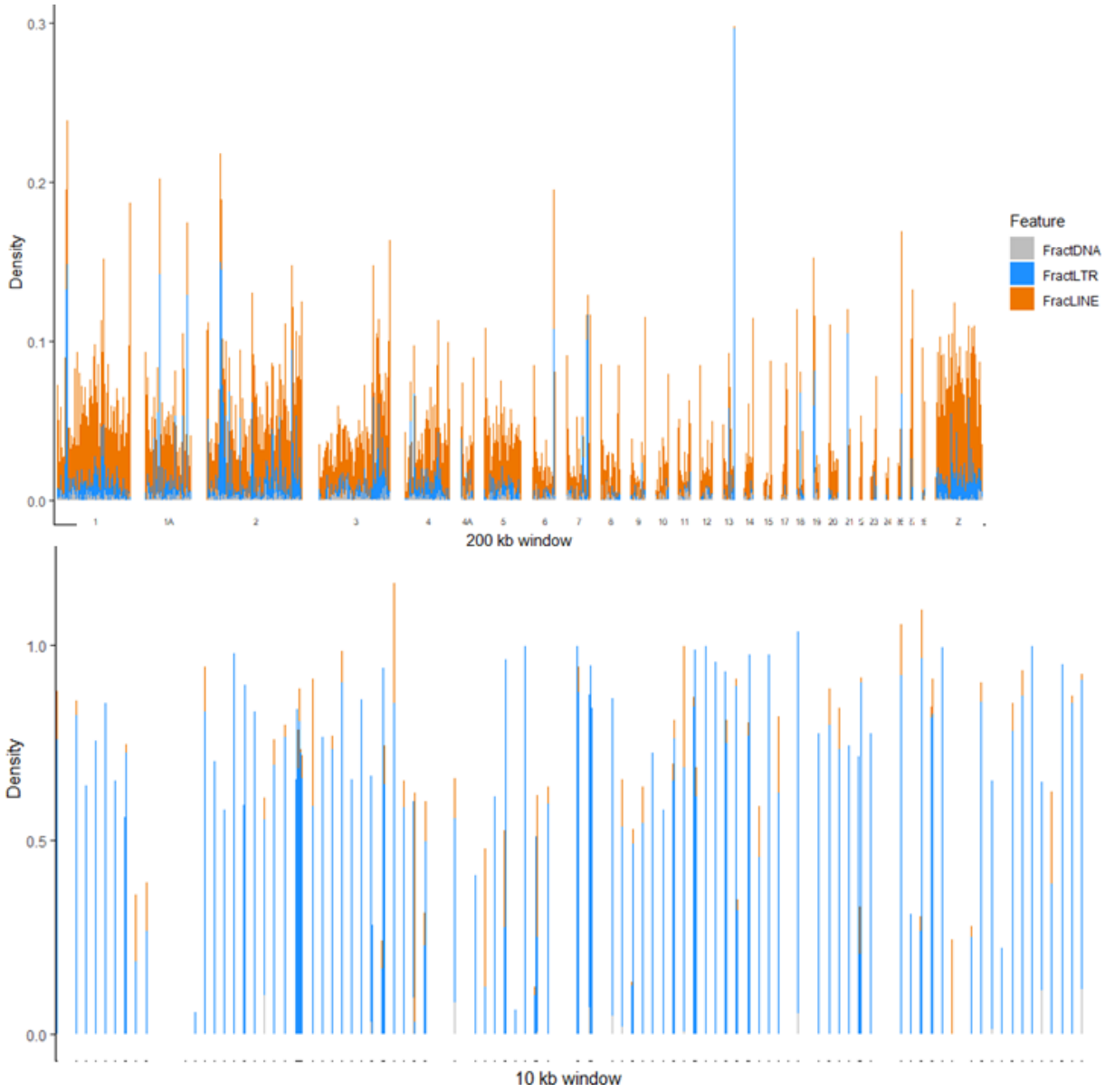

Figure S1. Density of the three most dominant TE classes (DNA transposons, LINE, and LTR retrotransposons) along the chromosomes of the hooded crow reference genome (top panel) and the 101 W-linked contigs (bottom panel). Density is expressed as the fraction of bases per 200 kb window that are covered by TEs. Autosomal scaffolds < 0.45 Mb in length are not included.

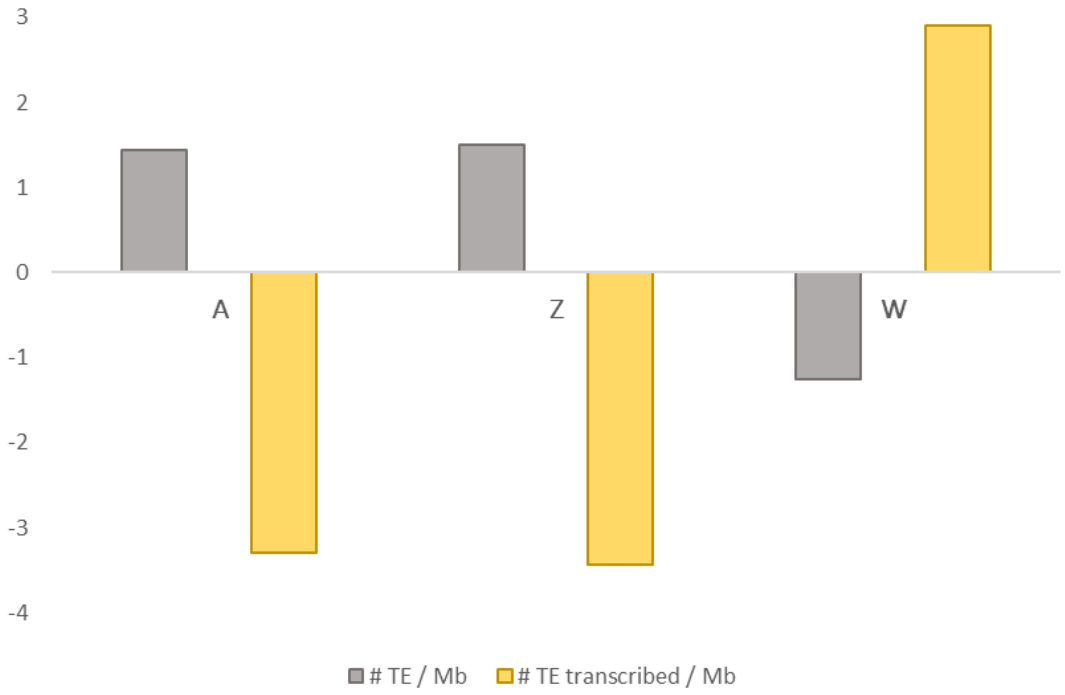

Figure S2. χ^2 residuals illustrating the difference between the observed and expected values for TE abundance (grey bars) and TE transcription (yellow bars) for the three chromosome types. χ^2 residuals follow a normal distribution and values exceeding +/- 2.58 can be taken as evidence for a statistically significant deviation from the expectation of a given class at Type I error probability of 0.01. TEs derived from autosomes and the Z Chromosome are significantly less often expressed than would be expected by their genomic abundance. On the W Chromosome, the pattern is reversed with TEs expression far exceeding the expectation.

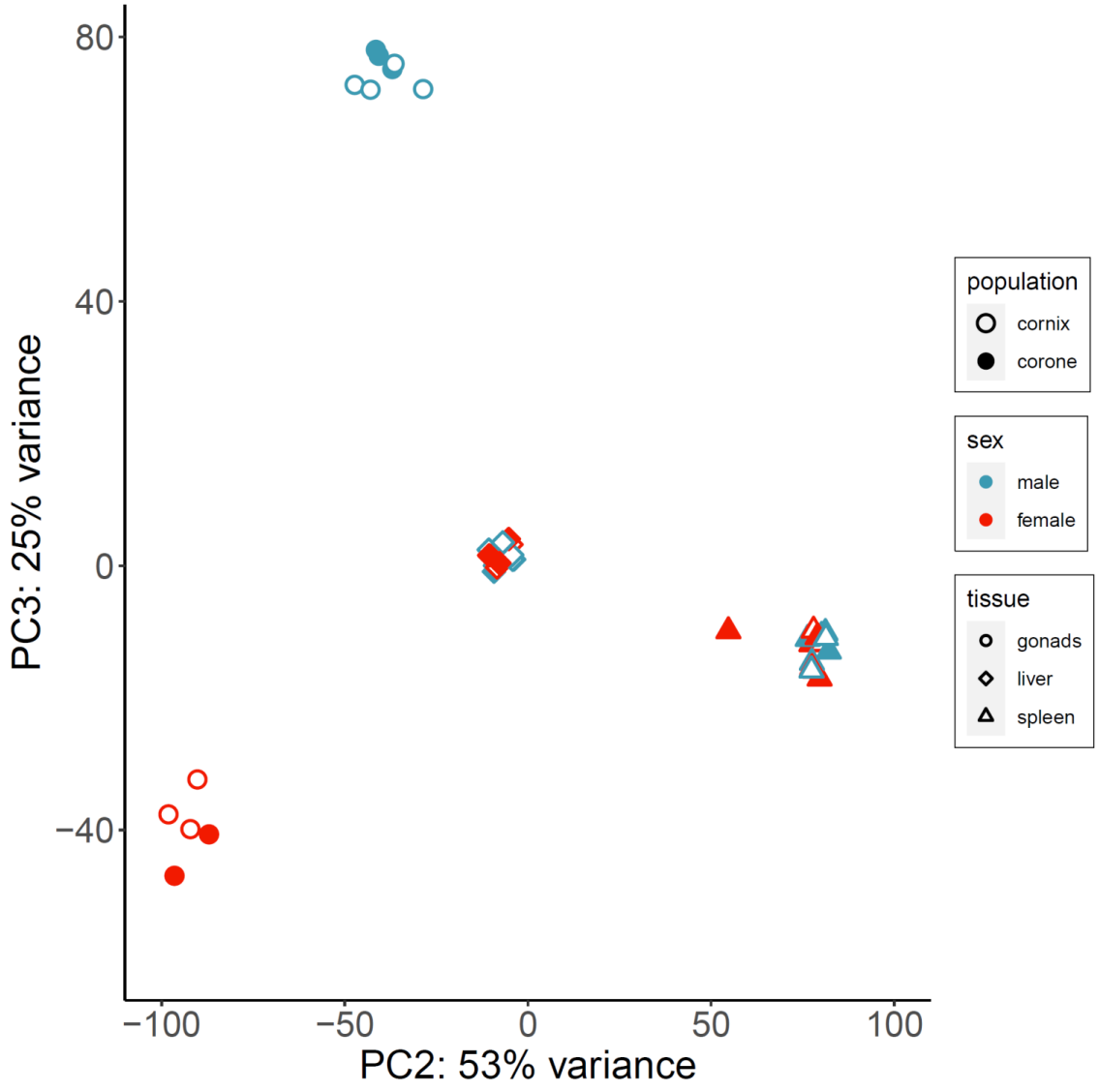

Figure S3. Principal components analysis of RNA-Seq data. PC2 and PC3 of normalized and vst-transformed read counts for coding genes in the hooded crow genome (including W-linked contigs).

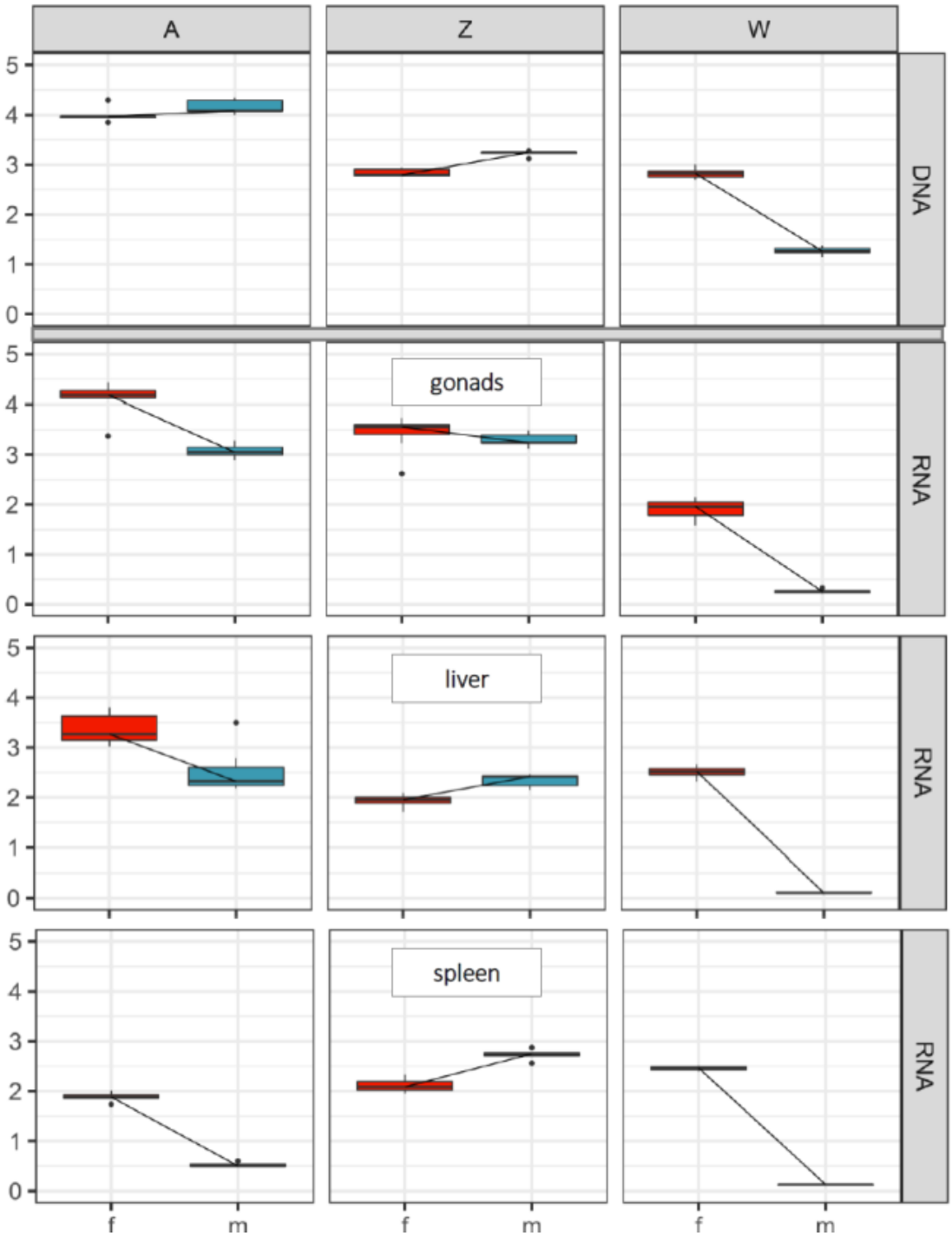

Figure S4. Genomic abundance (DNA-Seq, top panel) and transcript abundance (RNA-Seq, lower panels) of sex-biased DETEs in males and females across three tissues. The data represent normalized read counts averaged across all DETEs in females (red) and males (blue). Note that sex-biased DETEs include those with male-biased expression.

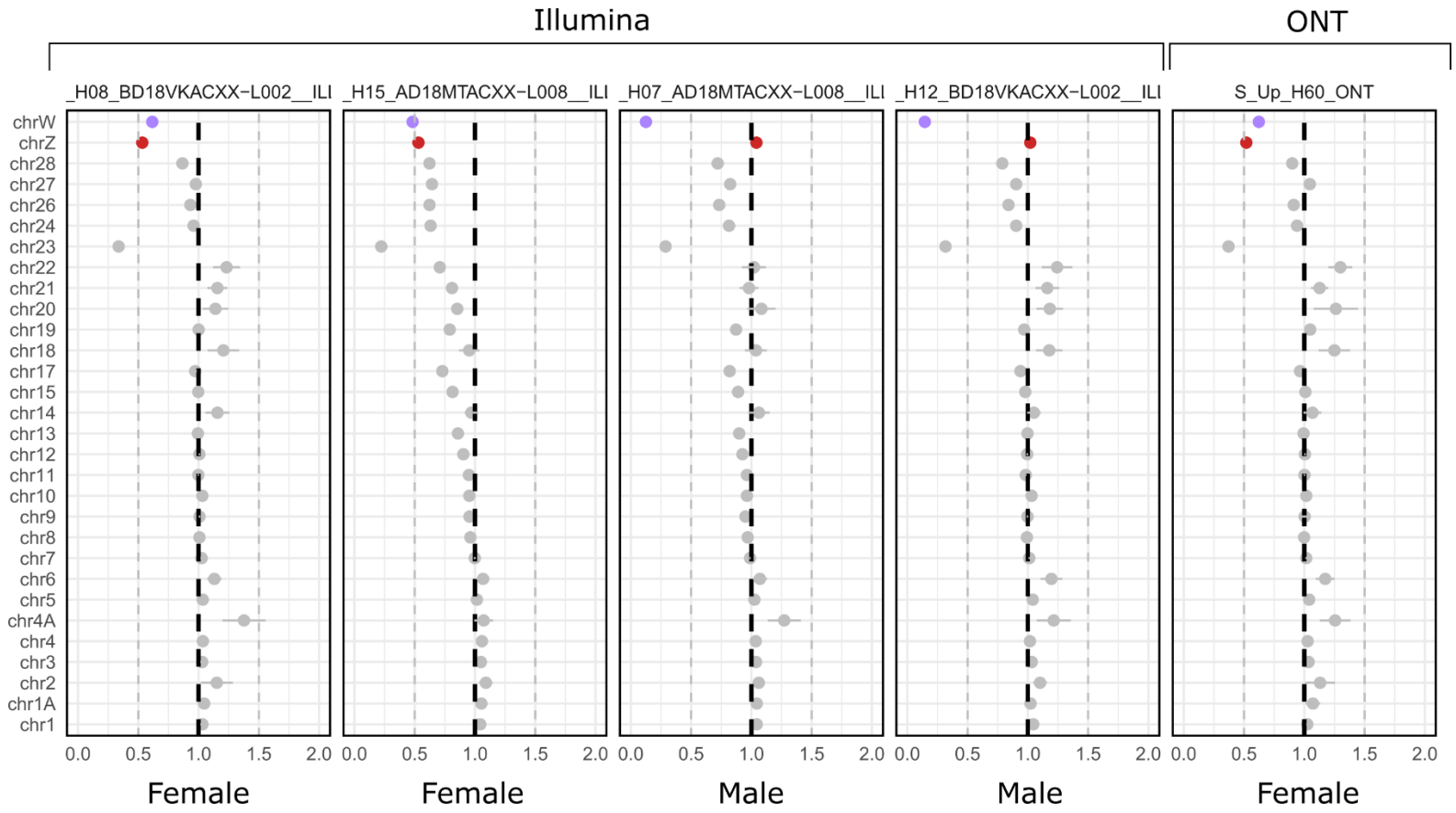

Figure S5. Relative sequencing read depth on sex chromosomes and autosomes. To determine whether the W Chromosome linked sequences identified in this study adhere to theoretical expectations regarding sequencing read depth in female and male individuals, we aligned short-read (first four columns) and long-read (last column, individual that was used for the *de novo* assembly) sequencing data to a combination of the hooded crow reference assembly (including autosomes and Z Chromosome) and the newly identified W-linked contigs. As expected, sequence read depth (relative to chromosome 1) of the W-linked contigs was half of that of Chromosome 1 in female individuals, and approaching zero in male individuals, providing further evidence that the tentatively W Chromosome linked contigs are indeed of W Chromosome origin.

Supplemental References

Chuong, E. B., Rumi, M. A. K., Soares, M. J., & Baker, J. C. (2013). Endogenous retroviruses function as species-specific enhancer elements in the placenta. *Nature Genetics*, *45*(3), 325–329.

Jin, Y., Tam, O. H., Paniagua, E., & Hammell, M. (2015). TEtranscripts: A package for including transposable elements in differential expression analysis of RNA-seq datasets. *Bioinformatics*, *31*(22), 3593–3599.

Knijff, P. de. (2014). How carrion and hooded crows defeat Linnaeus's curse. *Science*, *344*(6190), 1345–1346.

Parkin, D. T., Collinson, M., Helbig, A. J., Knox, A. G., & Sangster, G. (2003). The taxonomic status of Carrion and Hooded Crows. *British Birds*, *96*(6), 274–290.

Poelstra, J. W., Vijay, N., Hoeppner, M. P., & Wolf, J. B. W. (2015). Transcriptomics of colour patterning and coloration shifts in crows. *Molecular Ecology*, *24*(18), 4617–4628.

Shao, X., Lv, N., Liao, J., Long, J., Xue, R., Ai, N., Fan, X. (2019). Copy number variation is highly correlated with differential gene expression: A pan-cancer study. *BMC Medical Genetics*, *20*(1), 175.

Teissandier, A., Servant, N., Barillot, E., & Bourc'his, D. (2019). Tools and best practices for retrotransposon analysis using high-throughput sequencing data. *Mobile DNA*, *10*(1), 52.

Wolf, J. B., & Bryk, J. (2011). General lack of global dosage compensation in ZZ/ZW systems? Broadening the perspective with RNA-seq. *BMC Genomics*, *12*(1), 91.

Yang, W. R., Ardeljan, D., Pacyna, C. N., Payer, L. M., & Burns, K. H. (2019). SQuIRE reveals locus-specific regulation of interspersed repeat expression. *Nucleic Acids Research*, *47*(5), e27–e27.