Received: 5 July 2020 | Revised: 25 November 2020 | Accepted: 4 January 2021

DOI: 10.1111/mec.15793

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Genome-wide variation in DNA methylation linked to developmental stage and chromosomal suppression of recombination in white-throated sparrows

Dan Sun[1,2] | Thomas S. Layman[1] | Hyeonsoo Jeong[1] | Paramita Chatterjee[1] | Kathleen Grogan[3] | Jennifer R. Merritt[3] | Donna L. Maney[3] | Soojin V. Yi[1]

[1]School of Biological Sciences, Institute for Bioengineering and Bioscience, Georgia Institute of Technology, Atlanta, GA, USA

[2]Department of Epigenetics and Molecular Carcinogenesis, The University of Texas MD Anderson Cancer Center, Smithville, TX, USA

[3]Department of Psychology, Emory University, Atlanta, GA, USA

Correspondence
Soojin V. Yi, School of Biological Sciences, Institute for Bioengineering and Bioscience, Georgia Institute of Technology, Atlanta, GA 30332, USA.

Present address
Kathleen Grogan, Departments of Anthropology & Biology, University of Cincinnati, Cincinnati, OH, USA

Funding information
National Institutes of Health, Grant/Award Number: R01MH082833 and R01MH103517; National Science Foundation, Grant/Award Number: IOS-1656247 and MCB-1615664

Abstract

Much of our knowledge on regulatory impacts of DNA methylation has come from laboratory-bred model organisms, which may not exhibit the full extent of variation found in wild populations. Here, we investigated naturally-occurring variation in DNA methylation in a wild avian species, the white-throated sparrow (*Zonotrichia albicollis*). This species offers exceptional opportunities for studying the link between genetic differentiation and phenotypic traits because of a nonrecombining chromosome pair linked to both plumage and behavioural phenotypes. Using novel single-nucleotide resolution methylation maps and gene expression data, we show that DNA methylation and the expression of DNA methyltransferases are significantly higher in adults than in nestlings. Genes for which DNA methylation varied between nestlings and adults were implicated in development and cell differentiation and were located throughout the genome. In contrast, differential methylation between plumage morphs was concentrated in the nonrecombining chromosome pair. Interestingly, a large number of CpGs on the nonrecombining chromosome, localized to transposable elements, have undergone dramatic loss of DNA methylation since the split of the ZAL2 and ZAL2^{m} chromosomes. Changes in methylation predicted changes in gene expression for both chromosomes. In summary, we demonstrate changes in genome-wide DNA methylation that are associated with development and with specific functional categories of genes in white-throated sparrows. Moreover, we observe substantial DNA methylation reprogramming associated with the suppression of recombination, with implications for genome integrity and gene expression divergence. These results offer an unprecedented view of ongoing epigenetic reprogramming in a wild population.

KEYWORDS
DNA methylation, epigenetic reprogramming, gene expression, recombination, transposable elements

1 | INTRODUCTION

DNA methylation is a key epigenetic mark that adds another layer of information to the genomic DNA (Lister & Ecker, 2009). The best-known effect of DNA methylation, which has been observed most often in mammalian studies, is the repression of transcription resulting from methylation at CpG sites in *cis*-regulatory regions (Schübeler, 2015). The role of DNA methylation in other genomic regions is less well understood, although there are examples linking DNA methylation of gene bodies and intergenic regions to regulation of gene expression (Aran et al., 2013; Jjingo et al., 2012; Mendizabal et al., 2017). DNA methylation in nongenic regions has been implicated in many regulatory processes, including maintenance of genome stability and silencing transposable elements (TEs) (Burns, 2017; Deniz et al., 2019; Jones, 2012; Robertson & Jones, 2000). DNA methylation is also known to associate with development and aging (Bell et al., 2019; Horvath, 2013; Lister et al., 2013; Price et al., 2019; Sun & Yi, 2015).

Most of our knowledge about DNA methylation comes from studies of humans and laboratory mice. As a result of recent methodological advances in genomics and epigenomics, more data from natural populations are now being generated, often using reduced representation bisulphite sequencing (Lea et al., 2016; Thompson et al., 2017; Wogan et al., 2020). Nevertheless, detailed information on variation in DNA methylation and its associations with gene expression remains sparse, particularly for natural populations. In this study, we provide rare insight into how DNA methylation varies in a wild species of songbird. We used deep whole-genome bisulphite sequencing (WGBS) to generate single-nucleotide-resolution maps of DNA methylation in a wild passerine, the white-throated sparrow (*Zonotrichia albicollis*). This species is an exceptional vertebrate system for understanding links between the evolution of genomes and complex behavioural phenotypes (Maney et al., 2020; Merritt et al., 2020; Sun et al., 2018; Tuttle et al., 2016). Two naturally occurring plumage morphs, white-striped (WS) and tan-striped (TS), are completely linked to a supergene that segregates with an aggressive phenotype in both sexes. Birds of the WS morph, which are heterozygous for a rearranged second chromosome (ZAL2^m), are on average more aggressive than birds of the TS morph (Horton, Hudson, et al., 2014; Kopachena & Falls, 1993), which are homozygous for the standard arrangement (ZAL2) (Thomas et al., 2008; Thorneycroft, 1966, 1975). In addition, WS birds invest less in parenting behaviour than do TS birds (Horton, Hudson, et al., 2014; Horton et al., 2014; Kopachena & Falls, 1993; Maney, 2008; Maney et al., 2015; Tuttle, 2003; Zinzow-Kramer et al., 2015). Thus, the supergene is associated with a complex phenotype involving both aggression and parenting.

This unique chromosomal polymorphism is maintained in the population through disassortative mating; that is, nearly all mating pairs consist of one TS and one WS individual (Thorneycroft, 1966, 1975; Tuttle et al., 2016). As a consequence, the ZAL2^m chromosome is nearly always in a state of heterozygosity, experiencing little recombination. Cessation of recombination causes several genetic changes, including reduction of gene expression, accumulation of transposable elements, and ultimately, genetic degeneration of the nonrecombining region (Charlesworth & Charlesworth, 2000; Yi & Charlesworth, 2000). The ZAL2 and ZAL2^m chromosomes are in an early stage of genetic differentiation, having accumulated approximately 1% nucleotide divergence (Sun et al., 2018; Thomas et al., 2008; Tuttle et al., 2016). Based on mutation rates estimated in closely related species, the rearrangement is thought to have occurred approximately 2–3 million years ago (Huynh et al., 2011; Tuttle et al., 2016). The ZAL2^m chromosome has yet to exhibit signs of substantial genetic degeneration (i.e., only a handful of genes have become pseudogenes [Sun et al., 2018]). Despite this modest genetic divergence, genes on the nonrecombining ZAL2^m chromosome exhibit reduced expression, and ZAL2 appears to have evolved incipient dosage compensation, indicating rapid regulatory evolution preceding large-scale genetic differences between the ZAL2 and ZAL2^m chromosomes (Sun et al., 2018). Our novel whole-genome DNA methylation maps offer a unique opportunity to investigate how DNA methylation changes in the early stage of chromosomal differentiation following the cessation of recombination.

We investigated patterns of DNA methylation in 12 samples of brain tissue from female white-throated sparrows of both morphs. These samples were taken from seven adults and five nestlings, thus spanning two developmental time points. In this study we had three main objectives, as follows. First, we investigated variation in DNA methylation between the two developmental time points. Second, we examined DNA methylation divergence between the ZAL2 and ZAL2^m chromosomes. Finally, we looked for predictive relationships between differential DNA methylation and regulation of gene expression. Our novel and comprehensive data on nucleotide-resolution whole-genome methylation maps reveal substantial variation in DNA methylation between developmental stages and plumage morphs. By integrating this data set with novel gene expression data collected from the same individuals, as well as an open chromatin map of a WS bird, we demonstrate that variation in DNA methylation between nestlings and adults is widespread across the genome and correlated with variation in the expression of developmental genes. Furthermore, by identifying allele-specific methylation and its potential evolutionary origins, we provide a rare glimpse into epigenetic reprogramming of a chromosome following a recent cessation of recombination.

2 | MATERIALS AND METHODS

2.1 | Sample collection

We collected 12 birds (seven breeding adults [four WS and three TS] and five nestlings [three WS and two TS]) for our analysis (Tables S1–S3). A workflow of our data set and analysis is presented in Figure S1. To remove the effect of sex in a cost-effective manner, only female birds were analysed in the current study. Adults were collected using mist nets at our field site near Argyle, Maine, USA, as previously described (Horton, Hudson, et al., 2014; Zinzow-Kramer et al.,

2015). Nestlings were collected from nests at day seven post-hatch (Grogan et al., 2019). Only one nestling per nest was used for the analysis. The hypothalamus, a brain region containing neuroendocrine cell groups that respond to social cues in this species (Maney et al., 2008), was microdissected from each brain as previously described (Zinzow-Kramer et al., 2015). For ATAC-seq, the hypothalamus was microdissected from a nonbreeding WS male adult bird (sample ID 17031) collected at our field site in Atlanta, Georgia, USA during fall migration (Merritt et al., 2018).

2.2 | Whole genome sequencing, SNP calling, identification of fixed differences, and kinship analysis

Whole genome sequencing libraries from the same 12 birds used for WGBS analysis were generated from DNA extracted from the white-throated sparrow livers using a Qiagen DNeasy Blood and Tissue DNA kit. For each sample, 500 ng–1 μg of DNA was extracted and sheared on a Covaris ultrasonicator to 200–600 base pairs (bp) at the Emory Integrated Genomics Core. The DNA fragment ends were repaired, and A-overhangs were added before Nextera barcode adaptors were ligated to the DNA fragments overnight. Finally, the libraries were PCR-amplified to increase concentration and enrich for adaptor-ligated DNA fragments. WGS libraries were sequenced using Illumina HiSeq X Ten with 150 x 2 bp paired-end reads at Macrogen Clinical Laboratory. Read coverage and processing statistics are presented in Table S1.

To identify SNPs occupying CpG sites, we first removed adaptor sequences and low-quality bases from the sequencing reads using the parameters "-q 30 -O 1 -m 50 --trim-n --pair-filter any" using cutadapt 1.18 (Martin, 2011). Trimmed reads were then aligned to the NCBI white-throated sparrow reference genome (version: *Zonotrichia_albicollis-1.0.1*), which is from a TS individual, using Bowtie2 v2.3.4.2 (Langmead & Salzberg, 2012) with the --very-sensitive-local option, and the alignment rate was ~95% per sample. Technical duplicates (PCR and optical duplicates) were then discarded by Picard Tools 2.19.0 (https://broadinstitute.github.io/picard/). SNP calling was conducted on clean and aligned reads using GATK 4.0 (Van der Auwera et al., 2013; DePristo et al., 2011; McKenna et al., 2010). Specifically, SNPs were called using Haplotypecaller with the -ERC GVCF option, and joint genotyping of all samples was performed with the GenotypeGVCF. Finally, SNPs with MAF <0.05, meanDP < 5 and meanDP > 80 were discarded using VCFtools 0.1.15 (Danecek et al., 2011).

To minimize potential mapping bias towards the reference genome (ZAL2/ZAL2) caused by differences between ZAL2 and ZAL2^m, with the final set of SNPs, we identified putatively fixed differences between ZAL2 and ZAL2^m using the same procedure as described by Sun et al. (2018). For further alignment, we used a genome with putatively fixed differences masked by *N*'s in the reference (*N*-masked genome), as in other studies (e.g., Wu et al., 2020). To check whether the individuals in our analyses were related, we performed a kinship analysis using KING (Manichaikul et al., 2010). The kinship coefficients between the 12 individuals in this study were all practically zero (the maximum kinship coefficient was 0.00277; Figure S2), indicating that they were not closely related.

2.3 | Whole genome bisulphite sequencing

WGBS libraries were prepared using the following custom protocol. First, genomic DNA was extracted from the hypothalamus samples using a Qiagen DNeasy Blood and Tissue DNA kit. For each sample, 100 ng–1 μg of DNA was pooled with 1%–5% lambda phage DNA to test for bisulphite conversion efficiency. The DNA samples were then sheared on a Covaris ultrasonicator to 200–600 bp. The DNA fragment ends were repaired, and A-overhangs were added before bisulphite compatible adaptors were ligated to the DNA fragments overnight. Then, the DNA fragments were bisulphite-converted and PCR-amplified to increase concentration and enrich for adaptor-ligated DNA fragments. WGBS libraries were then sequenced using Illumina HiSeq X Ten at Macrogen Clinical Laboratory. On average, 431 and 134 million 150 bp × 2 reads were generated for WS and TS birds, respectively (Table S2).

2.4 | Preprocessing of whole genome bisulphite sequencing data

WGBS reads were trimmed as described above. The trimmed reads were aligned to the *N*-masked reference genome with parameters "--bowtie2-X 1000" using BISMARK v0.20.0 (Krueger & Andrews, 2011). The average mapping efficiency of samples was ~70% for all samples (Table S2). Duplicated reads and nonbisulphite-converted reads were discarded by *deduplicate_bismark* (parameter: -p) and *filter_non_conversion* (parameter: percentage_cutoff 20), respectively. Finally, *bismark_methylation_extractor* was run to extract CpG methylation calls. To obtain bisulphite conversion rates, raw reads were aligned to the phage lambda genome using Bismark (same parameters). Because lambda DNA is not methylated and therefore should be completely bisulphite-converted, the percentage of methylated cytosines of lambda DNA was taken as the nonconversion rate. Bisulphite conversion rates were above 99.8% in all samples (Table S2).

To call allele-specific methylation values, SNPSPLIT 0.3.4 (Krueger & Andrews, 2016) was run with parameters "--bisulphite --paired" using fixed differences between ZAL2 and ZAL2^m. Then, bismark_methylation_extractor was run for allele-separated reads. For WS birds, consistent with the genotype (ZAL2/ZAL2^m), the percentage of reads assigned to each chromosome was ~4%–4.5% (Table S2); for TS birds, the percentage of reads assigned to ZAL2 was ~8%–9% but to ZAL2^m 0%–0.01% (Table S2), which was consistent with the genotype (ZAL2 / ZAL2). After this procedure, the median sequencing depths were at least nine reads per sample and four per allele (Table S2). Only CpG sites with at least five reads aligned were

retained for further analysis (e.g., Mendizabal et al., 2019). Finally, because cytosine polymorphisms could hamper accurate calling of methylation, we excluded any CpGs in the reference genome that were polymorphic within the sequenced samples. After these quality control steps, a total of 3,880,473 CpGs were used to test for differential methylation between morphs and developmental stages, and 317,499 CpGs were used to test for differential methylation between alleles.

2.5 | Analysis of differential DNA methylation

Differentially methylated CpGs between developmental stages and morphs (or alleles) were detected by the dispersion shrinkage for sequencing data (DSS) package, version 2.30.1 (Wu et al., 2015) using Wald tests under the default setting. This method explicitly accounts for the characteristics of next-generation sequencing data and allows us to identify sites that are affected by different covariates. The null distributions of *p*-values for stage and allele comparisons largely followed uniform distributions (Figure S3). We then applied FDR to the *p*-values, and CpGs with FDR-corrected *p*-values <.05 and absolute values of differences in methylation greater than 10% were defined as DMCs, similar to the criteria in literature (Herb et al., 2012; Jeong et al., 2020; Yuen et al., 2010). BEDTOOLS version 2.28.0 (Quinlan & Hall, 2010) was run to assign DMCs to different gene features. If a DMC was within multiple gene features, we prioritized the assignment in the following order based on a feature's likelihood to influence gene expression (as per (Yu et al., 2015): upstream (10 Kb upstream of transcription start sites [TSS]), exons, introns, downstream (10 Kb downstream of TTS [transcription termination sites]), transposable elements [TEs] and intergenic regions.

Some authors identify differentially methylated regions (DMRs) rather than DMCs. While DMRs are useful to find genomic regions of interest, defining DMRs using different criteria can yield inconsistent results (as discussed by Lea et al., 2017; Roessler et al., 2016). Our comparisons of results obtained by identifying DMRs vs. DMCs showed that while the overall results were highly similar, some aspects of data were less well presented in DMR analyses (Figure S4). Therefore, we decided to focus and present DMC analyses in the current paper.

2.6 | Principal component analysis of DNA methylation data and chromosomal distribution of morph- and stage-DMCs

We stored DNA methylation data generated from all samples as a methylrawDB object using methylkit 1.9.4 (Akalin et al., 2012). The object was then converted into a percent methylation matrix, with only CpG sites with more than five reads in all samples retained. PCA analysis of all CpGs or CpGs inside and outside the ZAL2/ZAL2 m inversions was performed using the PCASamples function in methylkit (parameter: obj.return = T). The returned prcomp result was used to plot sample clusters with the autoplot function in ggfortify 0.4.5 (Tang et al., 2016).

To visualize enrichment/depletion of morph- and stage-DMCs along chromosomes, we calculated the fold enrichment with a 95% confidence interval by comparing the observed number of DMCs with the expected number of DMCs per chromosome. The null chromosomal distribution of expected numbers of DMCs was generated by 100 random selections of an equal total number of CpGs from the genome.

2.7 | ATAC-seq library preparation, sequencing, data preprocessing, and peak calling

For one sample (hypothalamus of a WS male, ID 17031), 10,000–200,000 cells were homogenized in EMEM (Eagle's Minimum Essential Medium) and phosphate-buffered saline. The cells were pelleted in a centrifuge and resuspended in a lysis buffer made of nonionic detergent (made in-house from Tris, NaCl, $MgCl_2$, and IGEPAL CA-630). After cell lysis, nuclei were isolated by centrifugation and added to a tagmentation reaction mix (Illumina Nextera DNA Library Prep Kit, Cat#: FC-121-1030). During tagmentation, the sequencing adapters were inserted into accessible chromatin regions by Tn5 transposase. Adapter-tagmented fragments were purified using Invitrogen Agencourt AMPure XP beads (Cat#: A63880) at 1.4×, 1× and 0.6× to remove fragments across a range of sizes. Purified samples were bar-coded (Illumina Nextera Index Kit, cat#: FC-121-1011), and amplified (Fisher KAPA HiFi HotStart Kit, Cat#: NC0295239). The ATAC-seq libraries were then sequenced using a MiSeq sequencer (Illumina; Reagent Kit v3) with 150 cycles (75 bp paired-end reads) in the Molecular Evolution Core at Georgia Tech.

We aligned the trimmed ATAC-seq reads (trimming was performed as above) to the *N*-masked reference genome using BOWTIE2 version 2.3.4.2 (parameters: -X 2000 --no-mixed --no-discordant) (Langmead & Salzberg, 2012), which allowed a maximal insert size of 2 Kb between paired reads, and discarded unmapped or discordant alignments. The mapping efficiency for this sample was 82.13%. The aligned reads were then deduplicated using markdup of SAMtools 1.7 (Li et al., 2009). As a result, we obtained 23 million clean mapped reads. To identify ZAL2 and $ZAL2^m$-specific ATAC-seq peaks, we followed the strategy proposed by (Jung et al., 2019). Specifically, we first called peaks in the overall sample using MACS2 version 2.1.1.20160309 (Zhang et al., 2008) with "-g 1.1e+9 -f BAMPE -p 0.05 -B --SPMR –nomodel" options. We next assigned reads to ZAL2 and $ZAL2^m$ using SNPsplit 0.3.4 (Krueger & Andrews, 2016) with parameters "--paired" using fixed differences between ZAL2 and $ZAL2^m$. The number of ZAL2 and $ZAL2^m$ reads mapped to the ATAC-seq peaks were counted using BEDTOOLS version 2.28.0 (Quinlan & Hall, 2010), and the differences in allelic read counts were tested by a two-tailed binomial test. Peaks with FDR-corrected $p < 0.05$ were denoted as allele-specific.

2.8 | RNA-seq library preparation, sequencing, data processing, and analysis of differential expression

RNA extraction and library preparation of the female samples were performed as previously described (Zinzow-Kramer et al., 2015). The libraries were then sequenced on the HiSeq 4000 at 150 bp paired-end reads to ~40 million reads per sample (Table S3). RNA-seq raw reads were trimmed as above and then aligned to the *N*-masked genome by HISAT2 2.1.0 (Kim et al., 2015). Secondary alignments were filtered by SAMTOOLS 1.7 (Li et al., 2009) to ensure that only primary alignments were retained. SNPSPLIT 0.3.4 (Krueger & Andrews, 2016) was run to assign reads to ZAL2 or ZAL2^m for the WS samples (Table S3). Expression levels (raw read counts) were then quantified by STRINGTIE version 1.3.4d (Pertea et al., 2015).

To identify genes that were differentially expressed between the morphs and developmental stages, we calculated size factors, normalized libraries with these factors, and then identified differential expression with "design = ~ stage +morph" (stage as the adjusted covariate) and "design = ~ morph +stage" (morph as the adjusted covariate), respectively, using the DESEQ2 1.22.2 package (Love et al., 2014) in R 3.5 (R Core Team, 2019). To identify genes that were differentially expressed between ZAL2 and ZAL2^m, we normalized libraries with the size factors generated in the morph comparison step and identified differential expression with "design = ~ stage +allele" (stage as the adjusted covariate). The null distributions of *p*-values for expression comparisons between developmental stages or between alleles followed approximately uniform distributions (Figure S3), and FDR-adjusted *p*-values < .05 were denoted as differentially expressed genes. Differences in the numbers of DMCs between promoters (defined as 1.5-kb upstream of TSSs) of adult- and nestling-biased genes were assessed by binomial tests.

2.9 | Transposable element annotation and expression/methylation analysis

We adopted both de novo and homology-based approaches to annotate repetitive sequences in the reference genome. First, de novo discovery of TEs was performed by RepeatModeler 1.0.9 (Smit & Hubley, 2008–2015). The generated library was merged with the avian Repbase library (20181026 version), which was used to annotate TEs in the reference genome using RepeatMasker 4.0.9 (parameters: -xsmall -s -nolow -norna -nocut) (Smit et al., 2013–2015).

To distinguish expression of genes from that of TEs and to estimate TE expression per subfamily, we used the TEcount program of TETOOLKIT 2.0.3 (Jin et al., 2015) We entered aligned RNA-seq reads, gene annotation, and TE annotation (at least 10 kb from any genes). DESeq2 was run on the generated count tables for library normalization across samples. For TE subfamilies with baseMean >3, the allelic difference in expression was tested using a Mann-Whitney U test. For TE subfamilies with more than 10 CpGs, allelic differences in DNA methylation between ZAL2 and ZAL2^m were detected using a Mann-Whitney U test while either including or excluding the ZAL2^m extremely hypomethylated DMCs.

2.10 | Cross-species comparison of DNA methylation

We used the brain methylation data from the great tit compiled by Sun et al. (2019) to infer evolutionary changes in DNA methylation of allele-DMCs. To do so, we aligned the white-throated sparrow reference genome to the great tit reference genome (Parus_major1.1) and to the zebra finch (Taeniopygia_guttata-3.2.4) reference genome using minimap2-2.16 (parameters: --secondary=no -c) (Li, 2018). Only alignments for which we were confident, defined by the highest mapping score (MAPQ = 60), were retained. ZAL2/ZAL2^m CpGs were identified if they could be aligned to the zebra finch chromosome homologous to ZAL2/ZAL2^m. The paftools liftover program of minimap2 was then run to find dinucleotides in the great tit genome that were orthologous to CpG sites in the sparrow genome and fractional methylation levels were calculated for each position.

3 | RESULTS

3.1 | Contrasting effects of developmental stage and morph on genome-wide DNA methylation maps

We examined patterns of DNA methylation and gene expression in samples of hypothalamus, a brain region containing cell groups critical to many social and developmental traits. Our experimental cohort included birds of both morphs in two developmental stages: adults aged more than one year (4 WS, 3 TS) and nestlings at post-hatch day seven (3 WS, 2 TS). We investigated the effects of morph and developmental stage (also referred to as "stage" in this manuscript) on genome-wide DNA methylation maps. These white-throated sparrows were not related, based on the kinship coefficient analysis (Manichaikul et al., 2010). Specifically, using all SNPs we detected in this study, the maximum kinship coefficient was 0.00277 (Section 2).

Whole genome bisulphite sequencing maps of the samples were generated (see Section 2). Bisulphite conversion rates, determined from spiked-in unmethylated lambda phage DNA, were all 99.8%. We generated on average 431 million reads of 150 bps for WS birds, and 134 million reads for TS birds. Greater coverage for the heterozygous WS birds was necessary to recover sufficient reads for both ZAL2 and ZAL2^m chromosomes. After removing duplicates, we mapped the reads to an *N*-masked reference genome to avoid mapping bias due to the polymorphisms between ZAL2/ZAL2^m chromosomes. Following these procedures, WS birds and TS birds had on average 33.4x and 13x coverage, respectively, per CpG (Table S2).

To first gain an understanding of genome-wide variation in DNA methylation, we performed a principal component analysis (PCA) of all preprocessed CpGs (Materials and Methods) prior to separating the ZAL2 and ZAL2^m alleles (Figure 1a). The first principal

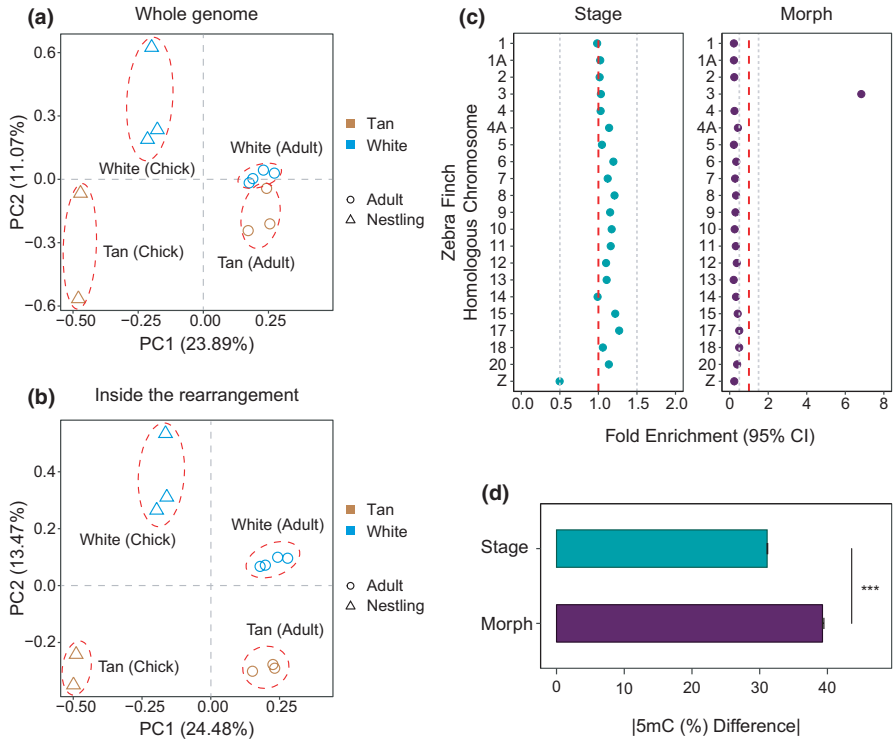

FIGURE 1 The effects of developmental stage and plumage morph on DNA methylation patterns in white-throated sparrows. We show PCA of WGBS samples for (a) all CpGs and (b) CpGs within the rearranged portion on ZAL2/ZAL2^{m}. (c) Fold enrichment with 95% confidence interval (95% CI) for the chromosomal distribution of DMCs (using homologous chromosomes in zebra finch for designation; ZAL2/ZAL2^{m} is equivalent to zebra finch chromosome 3). The fold enrichment and confidence intervals were calculated by comparing the real distribution of DMCs with the null distribution generated by 100 random selections of the same number of CpGs. The red dashed lines indicate no depletion/enrichment (enrichment score = 1) of DMCs on a chromosome, and the grey dashed lines depict boundaries for moderate depletion (0.5) or enrichment (1.5) of DMCs. Only chromosomes larger than 10 Mb are shown. (d) Mean absolute differences (+ 1.96 × standard errors) in fractional DNA methylation (5 mC [%]) for stage-DMCs (286,434) and morph-DMCs (4507). Effect sizes were smaller for stage-DMCs than for morph-DMCs. ***p < .001; Mann–Whitney U test. Standard error bars are shown. For (a–d) only CpGs with at least five reads mapped were used for analysis (Section 2)

component (PC1), which distinguished adults from nestlings, accounted for the largest amount of variation in DNA methylation (~20%). The second principal component (PC2) separated the TS and WS morphs and explained ~11% of the variation among samples. To examine the variation in DNA methylation associated with the chromosomal rearrangement, we then performed the same analyses using only the CpGs within the rearranged portion of the ZAL2/ZAL2^{m} chromosome (Figure 1b), which produced a clearer separation of the morphs. Consequently, developmental stage and morph were determined to be the top two factors explaining variation in DNA methylation in our data.

Using the same whole-genome CpG data, we identified significantly differentially methylated CpGs (herein referred to as "DMCs") between adults and nestlings, as well as between WS and TS birds, using a method designed specifically for the WGBS analysis (Wu et al., 2015). In addition to correcting for multiple testing using the FDR method (FDR-adjusted p < .05), we restricted the value of the absolute methylation difference to be equal to or >10%. Following these procedures, we identified 286,434 DMCs between adults and nestlings (referred to as "stage-DMCs"), and 4,507 DMCs between TS and WS birds (referred to as "morph-DMCs").

Stage-DMCs and morph-DMCs were distinct from each other with respect to both the chromosomal distribution and the effect sizes (Figure 1c, d). In terms of the chromosomal distribution, stage-DMCs were distributed across the genome, but depleted from the Z chromosome. In comparison, morph-DMCs were largely restricted to the ZAL2/ZAL2^{m} chromosomes (Figure 1c), indicating that nearly all differences in DNA methylation between the morphs were due to CpGs on the nonrecombining chromosomal pair. Consistent with this conclusion, PCA analysis of a different chromosome, ZAL1, clearly separated developmental stages but not the morphs (Figure S5). The effect sizes, measured as absolute differences in DNA methylation between the two morphs, were on average substantially greater than for the stage-DMCs (Figure 1d).

3.2 | Global hypermethylation of CpGs in adults relative to nestlings

Interestingly, most stage-DMCs (97.7% of all stage-DMCs) were more highly methylated (hyper-methylated) in adults than in nestlings (Figure 2a). We examined the expression levels of DNA

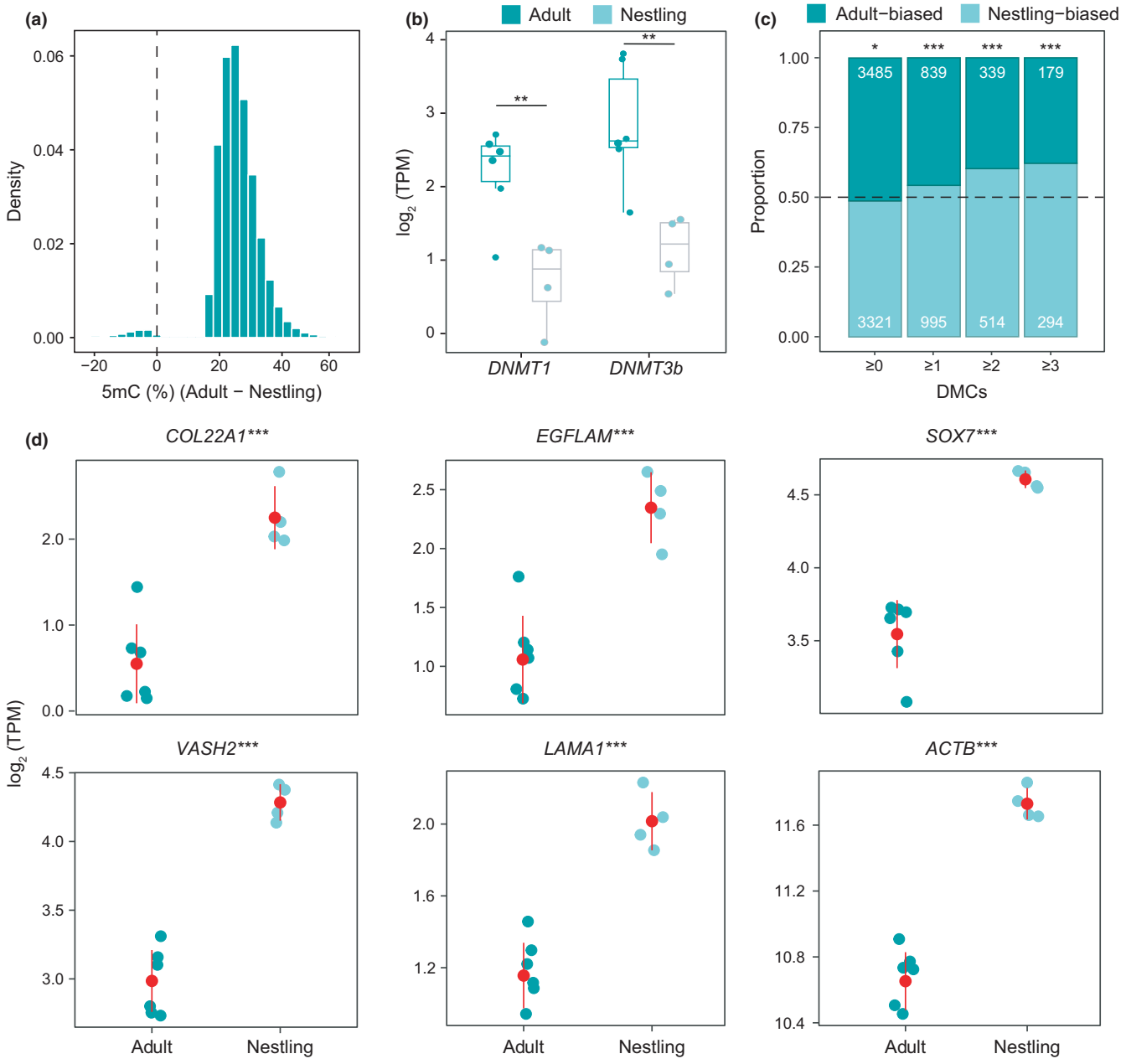

FIGURE 2 Hypermethylation in adults relative to nestlings. (a) The density distribution of differences in methylation between adults and nestlings shows that most stage-DMCs are hypermethylated in adults, compared with nestlings. (b) Both *DNMT1* and *DNMT3b* were more highly expressed in adults than in nestlings (tested by DESeq2, ***$p < .001$), consistent with the observed hypermethylation in adults. (c) The proportions of adult-biased and nestling-biased genes with zero, more than one, two, or three stage-DMCs in their promoters. The numbers of DE genes that are biased in each developmental stage are marked. The differences in the number of DE genes between adults and nestlings were tested by a binomial test (***$p < .001$). (d) Adults in general have lower gene expression levels than nestlings for stage-DE genes (tested by DESeq2, ***$p < .001$) associated with developmental processes (GO:0032502) and which have at least three stage-DMCs in the promoters. Shown here are some examples. Each dot represents a sample with both WGBS and RNA-seq data. Mean +/- standard deviations are depicted as red lines

methyltransferases (DMNTs) in RNA-seq data of the same individuals. Consistent with the observed genome-wide hypermethylation of samples from adults, DNA methyltransferases DNMT1 and DNMT3b had significantly higher expression in adults than in nestlings (Figure 2b; note that DNMT3a is not annotated in the reference genome due to the poor assembly quality around that region). Genes harbouring stage-DMCs in promoters (defined as 1.5-kb upstream of TSSs) were significantly enriched for gene ontology (GO) terms related to development and cell differentiation (Figure S6).

We identified a total of 6806 genes that were differentially expressed between nestlings and adults using FDR-adjusted $p < .05$, demonstrating that gene expression profiles change dramatically between the two developmental stages. Among these

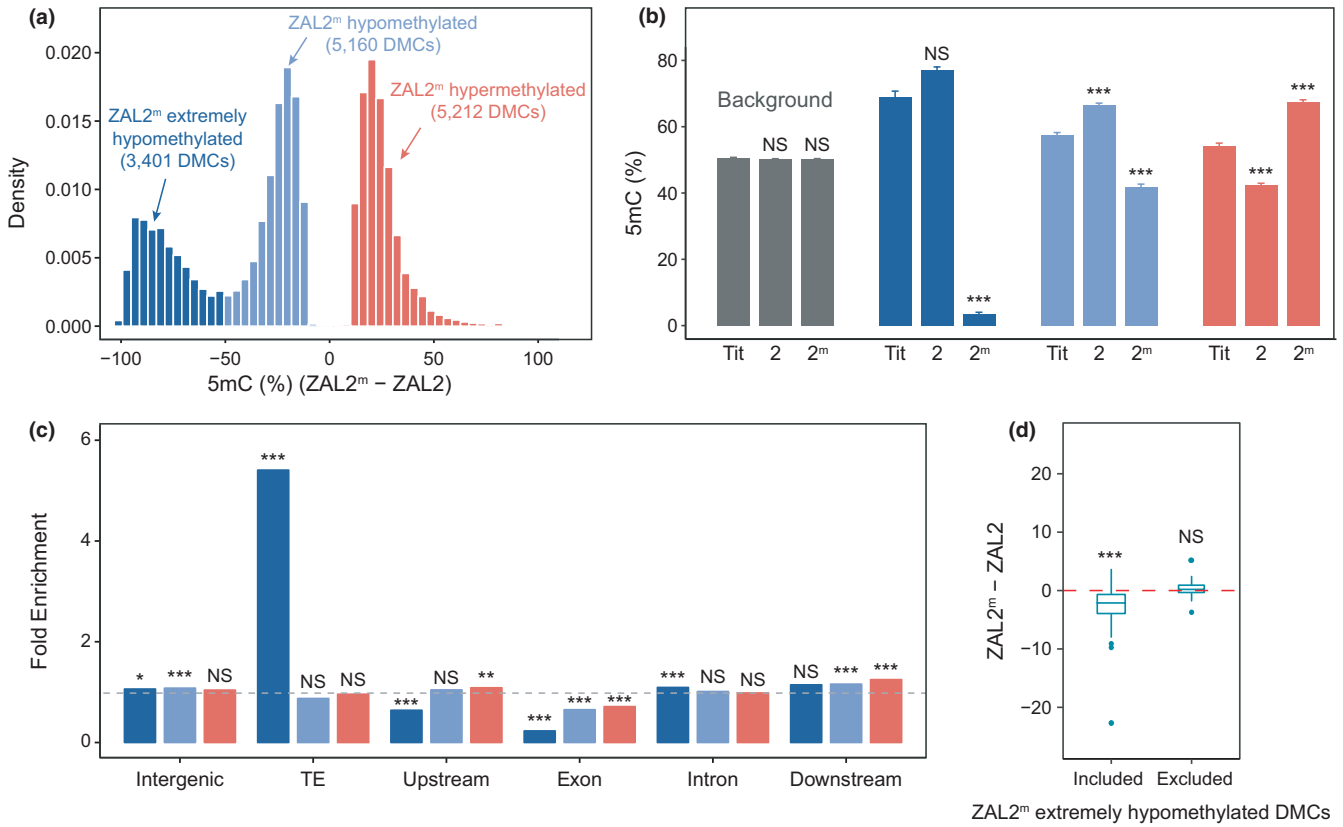

FIGURE 3 Characterization of the three classes of allele-DMCs. (a) The effect sizes of the differences in DNA methylation between ZAL2 and ZAL2^m alleles (allele-DMCs) fall into three distinct groups. (b) Changes in DNA methylation levels relative to the ancestral methylation levels inferred by comparison with an outgroup species, great tit. ***p < .001, Mann Whitney U test. (c) Fold enrichment of allele-DMCs within different genomic regions relative to the background (all CpGs on ZAL2/ZAL2^m). The dashed line corresponds to a fold enrichment of 1 (no enrichment or depletion). Intergenic regions were defined as regions that were at least 10 Kb away from any genes, and upstream/downstream distal regions were defined as 10 Kb upstream/downstream of the transcription start site (TSS)/transcription end site (TES). For B-C, all ZAL2/ZAL2^m-linked CpGs were used as the control, and enrichment or depletion was assessed by a two proportion Z-test. (d) Differences in TE methylation (5mC%) between ZAL2^m and ZAL2 after including or excluding the ZAL2^m extremely hypomethylated DMCs. Only TE subfamilies with more than 10 CpGs were used for analysis. For (b–d), NS, not significant; *p < .05; **p < .01; ***p < .001, Mann–Whitney U test

genes, about equal numbers were expressed more highly in adults vs. more highly in nestlings (3485 adult-biased genes vs. 3321 nestling-biased genes). In contrast, genes harbouring stage-DMCs in promoters tended to be more highly expressed in nestlings, and this trend increased as the number of stage-DMCs in each promoter increased (Figure 2c). In addition, for the GO term "developmental process", there were 74 genes harbouring DMCs in their promoters. Among these, 23 genes were downregulated in adults (31.08%). In contrast, out of all 17,365 genes, only 3,321 genes (19.12%) were downregulated in adults. Therefore, there was a significant enrichment of promoter-hypermethylated developmental genes that were downregulated in adults (19.12%, p = .014 by proportion test). These observations suggest that hypermethylation of promoter CpGs might contribute to the downregulation of early developmental genes in adults. Examples of several developmental genes harbouring promoter DMCs and exhibiting reduced expression in adults compared to nestlings are shown in Figure 2d.

3.3 | Differential methylation of the ZAL2 and ZAL2^m chromosomes is driven by substantial hypomethylation of CpGs on the nonrecombining ZAL2^m

Because the effects of morph on DNA methylation were nearly exclusive to the ZAL2/ZAL2^m chromosomes, we next investigated the DNA methylation patterns of these two chromosomes more deeply. To do so, we used WGBS data from WS individuals and separated the ZAL2 and ZAL2^m alleles (see Section 2). We then used DSS (Wu et al., 2015) to detect CpGs that were differentially methylated between ZAL2 and ZAL2^m (referred to as "allele-DMCs"). We identified 13,773 allele-DMCs using the same criteria we used in the genome-wide analysis (FDR-adjusted p < 0.05, absolute methylation difference >10%).

To examine the degree and direction of differences in DNA methylation between ZAL2 and ZAL2^m, we first plotted the sizes of these differences (level of ZAL2^m methylation − ZAL2 level of

methylation) in a histogram (Figure 3a). Their distribution revealed that allele-DMCs tended to be less methylated (also referred to as"hypomethylated") on ZAL2 ^{m} than on ZAL2 (Figure 3a). We identified three distinct groups of allele-DMCs. Approximately 75% of allele-DMCs showed a difference in DNA methylation between −0.5 to 0.5, i.e., a <50% difference in DNA methylation between ZAL2 and ZAL2^{m} (light blue and red, Figure 3a). These allele-DMCs were equally likely to be more methylated on either ZAL2 or ZAL2^{m} (depicted as "ZAL2^{m} < ZAL2" and "ZAL2^{m} > ZAL2" in Figure 3a, respectively). Interestingly, the remaining 25% of allele-DMCs showed extreme differential DNA methylation, with ZAL2^{m} alleles exhibiting markedly lower DNA methylation than their ZAL2 counterparts (depicted as "ZAL2^{m} ≪ ZAL2", dark blue, in Figure 3a). We will refer to these three categories of allele-DMCs as "ZAL2 ^{m} hypomethylated" (ZAL2^{m} < ZAL2, light blue in Figure 3a), "ZAL2^{m} hypermethylated" (ZAL2^{m} > ZAL2, red in Figure 3a), and "ZAL2^{m} extremely hypomethylated" (ZAL2^{m} ≪ ZAL2, dark blue in Figure 3a) in the remainder of the study.

To understand the evolutionary changes in DNA methylation leading to the three distinctive categories of allele-DMCs, we compared levels of DNA methylation in these three categories of CpGs, as well as those that did not exhibit differential DNA methylation, with corresponding levels of DNA methylation in a passerine outgroup, the great tit (Laine et al., 2016). This comparison revealed that CpGs that were not differentially methylated between the ZAL2 and ZAL2^{m} chromosomes showed similar methylation levels in the white-throated sparrow and great tit, suggesting that they have maintained similar levels of DNA methylation through evolutionary time (Figure 3b, grey columns). In comparison, ZAL2^{m} extremely hypomethylated (ZAL2^{m} ≪ ZAL2) DMCs bore a clear signature of hypomethylation on the ZAL2^{m} since the split from the great tit (Figure 3b). This pattern contrasts clearly with that of other allele-DMCs, which exhibited signs of both increased and decreased DNA methylation compared with great tit (Figure 3b, light blue and red). Together these observations indicate that although both ZAL2 and ZAL2^{m} have undergone changes in DNA methylation since the divergence from the great tit, a number of CpGs on ZAL2^{m} have experienced a strong reduction in DNA methylation since the split of the ZAL2 and ZAL2^{m} chromosomes.

We then tested whether allele-DMCs are enriched in specific functional regions. While the occurrence of other allele-DMCs was similar to all CpGs, ZAL2^{m} extremely hypomethylated allele-DMCs were five-fold enriched in TEs ($p < 2.2 \times 10^{-16}$ using a two proportion *Z*-test, Figure 3c). They were also slightly enriched in intronic regions, while slightly (yet significantly) depleted in regions upstream of transcription start sites (TSSs) where CpG islands are typically located (e.g., Mendizabal & Yi, 2016), Figure 3c). Currently, TEs in white-throated sparrow are poorly annotated. We used a de novo annotation (see Section 2) and identified subfamilies of TEs (we could not identify individual TEs with confidence due to low mappability). At the subfamily level, we observed higher expression of TEs on ZAL2^{m} than ZAL2 (Figure S7), which suggested potentially higher TE activity on ZAL2^{m}. We also observed that TEs were more hypomethylated on ZAL2^{m} than ZAL2, and that this pattern was driven by ZAL2^{m} extremely hypomethylated DMCs (Figure 3d). Given that these effects were estimated at the TE subfamily level, more data are necessary to show a direct link between methylation of TEs and their insertion activity in the ZAL2^{m} chromosome.

3.4 | Potential regulatory consequences of ZAL2 and ZAL2^{m}-specific DNA methylation

One of the best-known impacts of differential DNA methylation, when it occurs in promoters, is silencing of gene expression (Schübeler, 2015). Therefore, we first examined the expression levels of genes harbouring allele-DMCs in their promoters (defined as 1.5-kb upstream of TSSs). We found 325 genes with at least one allele-DMC in the promoter. For those genes, the divergence of gene expression was negatively correlated with the divergence of DNA methylation in the promoter (Figure 4a). This relationship was consistent with the aforementioned idea that promoter methylation dampens gene expression, although the degree of correlation was relatively weak (but significant). As we restricted our gene sets to those including more and more allele-DMCs in their promoters, the correlation coefficients increased (Figure 4b–c). These observations indicate that divergence in DNA methylation of promoters can explain some of the divergence in gene expression between ZAL2 and ZAL2^{m}.

Recent studies have demonstrated that the regulatory impacts of differential DNA methylation may extend far beyond promoters; differential DNA methylation is often predictive of differential expression of distant genes (Heyn et al., 2016; Hon et al., 2013; Murrell et al., 2004; Stadler et al., 2011) and differential accessibility of long-range chromatin (Guo et al., 2017; Lin et al., 2017; Liu et al., 2018; Lorincz et al., 2004). We thus investigated the relationship between allele-DMCs and chromatin accessibility. We generated a map of accessible chromatin regions using ATAC-seq on DNA isolated from the hypothalamus of a white-throated sparrow of the WS morph (ZAL2/ZAL2^{m}). We assigned open chromatin peaks to either ZAL2 or ZAL2^{m} (see Section 2) and examined the overlap of each peak with allele-DMCs. In the absence of enrichment or depletion of allele-DMCs in these peaks, the number of ATAC-seq peaks that overlap allele-DMCs should be proportional simply to the number of CpGs, regardless of their allele-specific methylation status (Table S4). In contrast to this prediction, we found both statistically significant enrichment and depletion of allele-DMCs within ATAC-seq peaks (Figure 4d, Table S4). Only one ZAL2^{m} extremely hypomethylated allele-DMC was located within ATAC-seq peaks on each chromosome, which represents a significant (p = .037, two proportion *Z*-test) and marginally significant (p = .058) depletion of allele-DMCs from this category on ZAL2 and ZAL2^{m}, respectively (Figure 4d). In contrast, other categories of allele-DMCs were enriched in allele-specific ATAC-seq peaks. Specifically, ZAL2^{m} hypomethylated allele-DMCs were enriched in ATAC-seq peaks specific to the ZAL2^{m} chromosome, but not

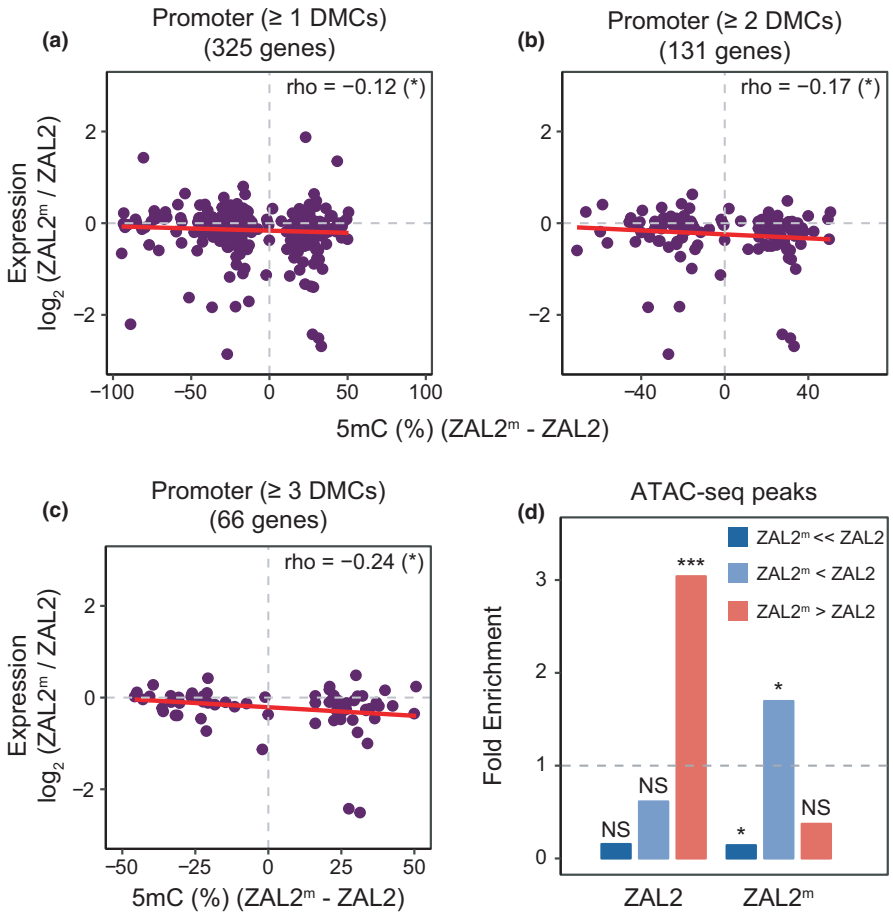

FIGURE 4 The potential role of allelic differences in DNA methylation in differential gene regulation. (a–c) Relationships between allelic differences in DNA methylation and allelic differences in gene expression for genes harbouring more than one, two, and three allele-DMCs in their promoters. Allelic differences in DNA methylation across DMCs in a region were averaged. The strength and direction of association were measured by Spearman's rank correlation coefficient, and the relationship was fit with a linear regression line (in red). (d) Fold enrichment of allele-DMCs occurring within ZAL2 or ZAL2^m-specific ATAC-seq peaks (37 and 23 peaks, respectively). The dashed line corresponds to a fold enrichment of 1 (no enrichment or depletion relative to the background of all ZAL2/ZAL2^m CpGs). NS, not significant; *p < .05; ***p < .001. Two proportion Z-test (also in Table S4)

in those specific to ZAL2 (Figure 4d). ZAL2^m hypermethylated allele-DMCs, on the other hand, were enriched in ZAL2 peaks but not in ZAL2^m peaks (Figure 4d). These observations suggest that differential DNA methylation between alleles correlates with their differential accessibility.

4 | DISCUSSION

Naturally occurring morphological and behavioural polymorphisms in white-throated sparrows offer a tremendous opportunity for studying the links between chromosomal differentiation and phenotypic traits (Maney et al., 2020; Merritt et al., 2020; Sun et al., 2018; Tuttle et al., 2016). In this work, we present new and extensive epigenomic and transcriptomic data from this non-model organism, broadening our perspective on development and chromosomal evolution. We showed that developmental stages and plumage morphs are associated with distinct patterns of genome-wide DNA methylation in this species. The comparison between nestlings and adults revealed significant differences in DNA methylation that are widespread across the genome, except for the Z chromosome (Figure 1c). As previous studies of DNA methylation across development/aging have typically excluded sex chromosomes (e.g., Day et al., 2013; Kim et al., 2018), we have yet to understand why stage-DMCs are underrepresented on the Z chromosome. Future studies that include sex chromosomes would reveal whether our observation is specific to white-throated sparrows or extends to other taxa.

Interestingly, stage-DMCs were predominantly hypermethylated in adults (Figure 2a), consistent with the significantly higher expression of DNMTs in adults (Figure 2b). Previous studies have demonstrated widespread hypermethylation in the brains of humans and mice (Li et al., 2014; Sun & Yi, 2015). As far as we are aware, however, changes in DNA methylation associated with aging have not been demonstrated outside of mammalian systems. Our observation of pronounced hypermethylation in adult brains, compared to nestling brains, in this avian species suggests that it may represent a shared molecular mechanism between mammals and birds. Previous studies in mammals have shown that DNA methylation regulates downstream pathways of neuronal and glial cellular differentiation (Fan et al., 2005; Murao et al., 2016; Wu et al., 2010), and that differential DNA methylation between cell types is critical for the differentiation of gene expression between them (Mendizabal et al., 2019). Results of GO analysis and differential gene expression suggest that hypermethylation of promoter CpGs in adult brains might contribute to the downregulation of early developmental genes (Figure S6).

In contrast, morph differences in DNA methylation were nearly exclusive to the ZAL2/ZAL2^m chromosomes. We identified nearly 14,000 CpGs that were differentially methylated, at a relatively stringent cutoff of FDR-corrected p < .05. Utilizing these CpGs and outgroup data, we observed both hyper- and hypomethylation of

the nonrecombining ZAL2^m chromosome as well as its counterpart, ZAL2, since their divergence. As DNA methylation varies strongly with underlying genetic variation in mammals and plants (Eichten et al., 2013; Keller et al., 2016; McClay et al., 2015; Yi, 2017) some of the observed epigenetic divergence could have been due to divergence of linked positions. Most of the CpGs that were differentially methylated between the ZAL2 and ZAL2^m chromosomes were equally likely to be more methylated on either ZAL2 or ZAL2^m. However, we discovered a group of CpGs that show extremely reduced DNA methylation on the ZAL2^m chromosome (referred to as "ZAL2^m extremely hypomethylated" in the Section 3). This group accounted for a quarter of all allele-differentially methylated CpGs (Figure 3). Cross-species comparisons solidified that these CpGs underwent massive hypomethylation on the nonrecombining ZAL2^m chromosome (Figure 3b).

Experimental studies in human and mouse cell lines have demonstrated that recombination following double strand breaks can recruit DNA methyltransferases and increase DNA methylation (Cuozzo et al., 2007; Morano et al., 2014; O'Hagan et al., 2008). At the genome scale, methylation-associated SNPs and germline methylation levels are both positively correlated with inferred recombination rates in humans (Sigurdsson et al., 2009; Zeng & Yi, 2014). Hypomethylation of the nonrecombining chromosome in white-throated sparrows, ZAL2^m, fits this broad observation, and supports a potential molecular link between recombination and DNA methylation. Interestingly, extreme hypomethylation of the ZAL2^m chromosome preferentially occurred in TEs (Figure 3c). ATAC-seq profiles of ZAL2^m extremely hypomethylated CpGs indicate that they tend to occur outside of accessible chromatin (Figure 4d). Hypomethylation is known to activate TEs (Rodríguez-Paredes & Esteller, 2011), further increasing TE insertion (Gaudet et al., 2003; Howard et al., 2007). We showed that at the subfamily level, TEs on ZAL2^m exhibit higher expression than those on ZAL2, which is consistent with the effects of hypomethylation on TE activity (Figure 3d). Given that an increase in TE insertion is hypothesized to be one of the first genomic changes during the evolution of nonrecombining chromosomes in *Drosophila* (Zhou et al., 2013), a similar mechanism may be operating in the ZAL2/ZAL2^m system, potentially fueled by the extreme hypomethylation. Additional data on TE transcription and a better-annotated reference genome in this species will be necessary to investigate the relationship between DNA methylation and TE activity on the ZAL2^m chromosome.

Integrating our gene expression data and chromatin accessibility data, we present results consistent with potential regulatory roles of allele-specific DNA methylation. First, when allele-DMCs were present in the promoter, the degree of differential methylation of those promoters was correlated with the degree of differential expression of the genes (Figure 4a–c). Second, the landscape of open chromatin on the ZAL2 and ZAL2^m chromosomes in a WS bird suggested significant associations between allele-specific hypomethylation and allele-specific open chromatin peaks (Figure 4d). The comparison between ATAC-seq peaks and DNA methylation should be taken with caution because of a limitation in our data; the tissue sample used for ATAC-seq was from a nonbreeding (winter) male while the adult WGBS data were from breeding (summer) females. ATAC-seq and WGBS data from the same birds are currently lacking. A recent study of 66 ATAC-seq maps from 20 different tissues of male and female mice (Liu et al., 2019) demonstrated that the majority of accessible regions between tissues overlapped and that the correlation between male and female tissues was extremely high. For example, in samples of cerebellum in mice, the correlation in accessible regions between males and females was 0.96 in (Liu et al., 2019). In the present study, the associations between DNA methylation and chromatin accessibility are consistent with those observed in model organisms (Guo et al., 2017; Liu et al., 2018; Lorincz et al., 2004) and suggest that changes in allele-specific DNA methylation may correlate with the chromatin landscape. Together, these observations indicate possibilities for widespread functional impacts of differential DNA methylation in the genome of this interesting species.

We should consider potential caveats in our study. First, our sample size of 12 is modest, and likely to lack power to detect many true differences in DNA methylation (e.g., Lea et al., 2017). In the future, it will be useful to perform targeted validation studies using more samples. Nevertheless, the *p*-value distribution from our differential DNA methylation analysis suggests that we were able to detect many true positives (Figure S3). In addition, although our data were generated from a specific region of the brain, the hypothalamus, this region is itself heterogeneous, containing a variety of nuclei and cell types. Recent studies have shown that differential DNA methylation between cell types is substantial (e.g., Jeong et al., 2020; Lister et al., 2013; Mendizabal et al., 2019). In addition, the cell composition of this region may change over the course of development, which could have driven some of the stage-related differences in methylation that we observed. However, it is notable that even in cell-type resolved samples, DNA methylation associated with age was observed in the same direction in human and mouse (Lister et al., 2013; Sun & Yi, 2015). Also note that differences in methylation between ZAL2 and ZAL2^m, since methylation was estimated from the same WS samples in these cases, should be free from such a bias. Finally, our study was designed to examine CpGs that are conserved on both the ZAL2 and ZAL2^m chromosomes, so that we could identify differentially methylated CpGs. It should be noted that CpGs that are specific to either chromosome might play important roles, particularly with respect to morph differences in endocrinology and behaviour (Merritt et al., 2020). We intend to study the potential impacts of chromosome-specific CpGs in follow-up studies.

In conclusion, our comprehensive epigenetic study in white-throated sparrows has revealed significant effects of developmental stage and plumage morph on DNA methylation landscapes. We show that effects of developmental stage on DNA methylation are pervasive and probably affect regulation of developmental genes. In contrast, morph differences in DNA methylation are mostly enriched on ZAL2/ZAL2^m, and involve both hyper- and hypomethylation of the recombination-suppressed ZAL2^m as well as its counterpart, ZAL2. On the basis of a comparison with an outgroup, we also discovered a large number of CpGs for which DNA

methylation has been dramatically reduced specifically on the ZAL2^{m} chromosome. We propose that these different varieties of allelic DNA methylation divergence have led to specific functional consequences. Together, our results not only provide a novel data set from a wild avian species, but also raise several hypotheses on which we hope future studies will build to further illuminate the connection between genotype and phenotype and pathways of chromosome evolution.

ACKNOWLEDGEMENTS

This study was funded by an NIH grant (R01MH082833) to DLM and an NSF grant (IOS-1656247) to DLM and SVY, and by NIH (R01MH103517) and NSF (MCB-1615664) grants to SVY. We thank Ben Long for comments on the manuscript.

AUTHOR CONTRIBUTIONS

D.S., D.L.M., and S.V.Y. designed the experiments. P.C., and T.L. performed WGBS. T.L. performed ATAC-seq. K.E.G. did the brain dissections, extracted RNA/DNA and led RNA-seq. J.R.M. provided input for ATAC-seq analyses. D.S., and H.J. performed analyses. D.S., and S.V.Y. drafted the initial manuscript and D.S., D.L.M., and S.V.Y. edited the manuscript. All authors reviewed and approved the submission.

DATA AVAILABILITY STATEMENT

WGBS, WGS, ATAC-seq and RNA-seq data in this study have been deposited in the Sequence Read Archive (SRA) with the project accession number PRJNA643514. Processed DNA methylation data per WS/TS sample can be accessed through FigShare https://doi.org/10.6084/m9.figshare.12597503 and per allele of each WS sample through FigShare https://doi.org/10.6084/m9.figshare.12597554 (Sun et al., 2020). Code for differential methylation and expression analysis can be found at https://github.com/soojinyilab/sparrow_WGBS_paper.

ORCID

Dan Sun https://orcid.org/0000-0003-2147-9308
Kathleen Grogan https://orcid.org/0000-0002-0112-1766
Jennifer R. Merritt https://orcid.org/0000-0002-1189-0454
Donna L. Maney https://orcid.org/0000-0002-1006-2358
Soojin V. Yi https://orcid.org/0000-0003-1497-1871

REFERENCES

Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F. E., Figueroa, M. E., Melnick, A., & Mason, C. E. (2012). methylKit: A comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biology*, *13*(10), R87. https://doi.org/10.1186/gb-2012-13-10-r87

Aran, D., Sabato, S., & Hellman, A. (2013). DNA methylation of distal regulatory sites characterizes dysregulation of cancer genes. *Genome Biology*, *14*(3), R21. https://doi.org/10.1186/gb-2013-14-3-r21

Auwera, G. A., Carneiro, M. O., 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, M. A. (2013). From FastQ data to high confidence variant calls: The Genome Analysis Toolkit best practices pipeline. *Current Protocols in Bioinformatics*, *43*, 11.10.11-11.10.33. https://doi.org/10.1002/0471250953.bi1110s43

Bell, C. G., Lowe, R., Adams, P. D., Baccarelli, A. A., Beck, S., Bell, J. T., Christensen, B. C., Gladyshev, V. N., Heijmans, B. T., Horvath, S., Ideker, T., Issa, J.-P., Kelsey, K. T., Marioni, R. E., Reik, W., Relton, C. L., Schalkwyk, L. C., Teschendorff, A. E., Wagner, W., … Rakyan, V. K. (2019). DNA methylation aging clocks: challenges and recommendations. *Genome Biology*, *20*(1), 249. https://doi.org/10.1186/s13059-019-1824-y

Burns, K. H. (2017). Transposable elements in cancer. *Nature Reviews Cancer*, *17*(7), 415–424. https://doi.org/10.1038/nrc.2017.35

Charlesworth, B., & Charlesworth, D. (2000). The degeneration of Y chromosomes. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *355*(1403), 1563–1572. https://doi.org/10.1098/rstb.2000.0717

Cuozzo, C., Porcellini, A., Angrisano, T., Morano, A., Lee, B., Pardo, A. D., Messina, S., Iuliano, R., Fusco, A., Santillo, M. R., Muller, M. T., Chiariotti, L., Gottesman, M. E., & Avvedimento, E. V. (2007). DNA damage, homology-directed repair, and DNA methylation. *PLoS Genetics*, *3*(7), e110. https://doi.org/10.1371/journal.pgen.0030110

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Day, K., Waite, L. L., Thalacker-Mercer, A., West, A., Bamman, M. M., Brooks, J. D., Myers, R. M., & Absher, D. (2013). Differential DNA methylation with age displays both common and dynamic features across human tissues that are influenced by CpG landscape. *Genome Biology*, *14*(9), R102. https://doi.org/10.1186/gb-2013-14-9-r102

Deniz, O., Frost, J. M., & Branco, M. R. (2019). Regulation of transposable elements by DNA modifications. *Nature Reviews Genetics*, *20*(7), 417–431. https://doi.org/10.1038/s41576-019-0106-6

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*(5), 491–498. https://doi.org/10.1038/ng.806

Eichten, S. R., Briskine, R., Song, J., Li, Q., Swanson-Wagner, R., Hermanson, P. J., Waters, A. J., Starr, E., West, P. T., Tiffin, P., Myers, C. L., Vaughn, M. W., & Springer, N. M. (2013). Epigenetic and genetic influences on DNA methylation variation in maize populations. *The Plant Cell*, *25*(8), 2783–2797

Fan, G., Martinowich, K., Chin, M. H., He, F., Fouse, S. D., Hutnick, L., Hattori, D., Ge, W., Shen, Y., Wu, H., ten Hoeve, J., Shuai, K., & Sun, Y. E. (2005). DNA methylation controls the timing of astrogliogenesis through regulation of JAK-STAT signaling. *Development*, *132*(15), 3345. https://doi.org/10.1242/dev.01912

Gaudet, F., Hodgson, J. G., Eden, A., Jackson-Grusby, L., Dausman, J., Gray, J. W., Leonhardt, H., & Jaenisch, R. (2003). Induction of tumors in mice by genomic hypomethylation. *Science*, *300*(5618), 489. https://doi.org/10.1126/science.1083558

Grogan, K. E., Horton, B. M., Hu, Y., & Maney, D. L. (2019). A chromosomal inversion predicts the expression of sex steroid-related genes in a species with alternative behavioral phenotypes. *Molecular and Cellular Endocrinology*, *495*, 110517. https://doi.org/10.1016/j.mce.2019.110517

Guo, H., Hu, B., Yan, L., Yong, J., Wu, Y., Gao, Y., Guo, F., Hou, Y. U., Fan, X., Dong, J. I., Wang, X., Zhu, X., Yan, J., Wei, Y., Jin, H., Zhang, W., Wen, L. U., Tang, F., & Qiao, J. (2017). DNA methylation and chromatin accessibility profiling of mouse and human fetal germ cells. *Cell Research*, *27*(2), 165–183. https://doi.org/10.1038/cr.2016.128

Herb, B. R., Wolschin, F., Hansen, K. D., Aryee, M. J., Langmead, B., Irizarry, R., Amdam, G. V., & Feinberg, A. P. (2012). Reversible switching between epigenetic states in honeybee behavioral subcastes. *Nature Neuroscience*, *15*, 1371–1373.

Heyn, H., Vidal, E., Ferreira, H. J., Vizoso, M., Sayols, S., Gomez, A., Moran, S., Boque-Sastre, R., Guil, S., Martinez-Cardus, A., Lin, C. Y., Royo, R., Sanchez-Mut, J. V., Martinez, R., Gut, M., Torrents, D., Orozco, M., Gut, I., Young, R. A., & Esteller, M. (2016). Epigenomic analysis detects aberrant super-enhancer DNA methylation in human cancer. *Genome Biology*, *17*(1), 11. https://doi.org/10.1186/s13059-016-0879-2

Hon, G. C., Rajagopal, N., Shen, Y., McCleary, D. F., Yue, F., Dang, M. D., & Ren, B. (2013). Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nature Genetics*, *45*(10), 1198–1206. https://doi.org/10.1038/ng.2746

Horton, B. M., Hudson, W. H., Ortlund, E. A., Shirk, S., Thomas, J. W., Young, E. R., Zinzow-Kramer, W. M., & Maney, D. L. (2014). Estrogen receptor alpha polymorphism in a species with alternative behavioral phenotypes. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(4), 1443–1448. https://doi.org/10.1073/pnas.1317165111

Horton, B. M., Moore, I. T., & Maney, D. L. (2014). New insights into the hormonal and behavioural correlates of polymorphism in white-throated sparrows, *Zonotrichia albicollis*. *Animal Behaviour*, *93*, 207–219. https://doi.org/10.1016/j.anbehav.2014.04.015

Horvath, S. (2013). DNA methylation age of human tissues and cell types. *Genome Biology*, *14*(10), 3156. https://doi.org/10.1186/gb-2013-14-10-r115

Howard, G., Eiges, R., Gaudet, F., Jaenisch, R., & Eden, A. (2007). Activation and transposition of endogenous retroviral elements in hypomethylation induced tumors in mice. *Oncogene*, *27*, 404. https://doi.org/10.1038/sj.onc.1210631

Huynh, L. Y., Maney, D. L., & Thomas, J. W. (2011). Chromosome-wide linkage disequilibrium caused by an inversion polymorphism in the white-throated sparrow (Zonotrichia albicollis). *Heredity*, *106*(4), 537–546.

Jeong, H., Mendizabal, I., Berto, S., Chatterjee, P., Layman, T., Usui, N., Toriumi, K., Douglas, C.,Singh, D., Huh, I., Preuss, T. M., Konopka, G., Yi, S. V. (2020). Distinctive cell-type and context specific DNA methylation trajectory during human brain evolution. *bioRxiv(2020.07.14.203034)*. https://doi.org/10.1101/2020.07.14.203034

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. https://doi.org/10.1093/bioinformatics/btv422

Jjingo, D., Conley, A. B., Yi, S. V., Lunyak, V. V., & Jordan, I. K. (2012). On the presence and role of human gene-body DNA methylation. *Oncotarget*, *3*(4), 462–474. https://doi.org/10.18632/oncotarget.497

Jones, P. A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Reviews Genetics*, *13*(7), 484–492. https://doi.org/10.1038/nrg3230

Jung, Y. H., Kremsky, I., Gold, H. B., Rowley, M. J., Punyawai, K., Buonanotte, A., Lyu, X., Bixler, B. J., Chan, A. W. S., & Corces, V. G. (2019). Maintenance of CTCF- and transcription factor-mediated interactions from the gametes to the early mouse embryo. *Molecular Cell*, *75*(1), 154–171.e155. https://doi.org/10.1016/j.molcel.2019.04.014

Keller, T. E., Lasky, J. R., & Yi, S. V. (2016). The multivariate association between genomewide DNA methylation and climate across the range of *Arabidopsis thaliana*. *Molecular Ecology*, *25*(8), 1823–1837. https://doi.org/10.1111/mec.13573

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*, *12*(4), 357–360. https://doi.org/10.1038/nmeth.3317

Kim, S., Wyckoff, J., Morris, A. T., Succop, A., Avery, A., Duncan, G. E., & Jazwinski, S. M. (2018). DNA methylation associated with healthy aging of elderly twins. *Geroscience*, *40*(5–6), 469–484. https://doi.org/10.1007/s11357-018-0040-0

Kopachena, J. G., & Falls, J. B. (1993). Aggressive performance as a behavioral correlate of plumage polymorphism in the white-throated sparrow (*Zonotrichia albicollis*). *Behaviour*, *124*(3/4), 249–266.

Krueger, F., & Andrews, S. R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*, *27*(11), 1571–1572.

Krueger, F., & Andrews, S. R. (2016). SNPsplit: allele-specific splitting of alignments between genomes with known SNP genotypes. *F1000Research*, *5*, 1479. https://doi.org/10.12688/f1000research.9037.2

Laine, V. N., Gossmann, T. I., Schachtschneider, K. M., Garroway, C. J., Madsen, O., Verhoeven, K. J. F., de Jager, V., Megens, H.-J., Warren, W. C., Minx, P., Crooijmans, R. P. M. A., Corcoran, P., Sheldon, B. C., Slate, J., Zeng, K., van Oers, K., Visser, M. E., & Groenen, M. A. M. (2016). Evolutionary signals of selection on cognition from the great tit genome and methylome. *Nature Communications*, *7*, 10474. https://doi.org/10.1038/ncomms10474

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–U354. https://doi.org/10.1038/Nmeth.1923

Sun, et al. (2020). WGBS data from white-throated sparrows. https://doi.org/10.6084/m9.figshare.12597503 and https://doi.org/10.6084/m9.figshare.12597554

Lea, A. J., Altmann, J., Alberts, S. C., & Tung, J. (2016). Resource base influences genome-wide DNA methylation levels in wild baboons (*Papio cynocephalus*). *Molecular Ecology*, *25*(8), 1681–1696. https://doi.org/10.1111/mec.13436

Lea, A. J., Vilgalys, T. P., Durst, P. A. P., & Tung, J. (2017). Maximizing ecological and evolutionary insight in bisulfite sequencing data sets. *Nature Ecology & Evolution*, *1*(8), 1074–1083. https://doi.org/10.1038/s41559-017-0229-0

Li, G. E., Zhang, W., Baker, M. S., Laritsky, E., Mattan-Hung, N., Yu, D., Kunde-Ramamoorthy, G., Simerly, R. B., Chen, R., Shen, L., & Waterland, R. A. (2014). Major epigenetic development distinguishing neuronal and non-neuronal cells occurs postnatally in the murine hypothalamus. *Human Molecular Genetics*, *23*(6), 1579–1590. https://doi.org/10.1093/hmg/ddt548

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

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*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lin, X., Su, J., Chen, K., Rodriguez, B., & Li, W. (2017). Sparse conserved under-methylated CpGs are associated with high-order chromatin structure. *Genome Biology*, *18*(1), 163. https://doi.org/10.1186/s13059-017-1296-x

Lister, R., & Ecker, J. R. (2009). Finding the fifth base: Genome-wide sequencing of cytosine methylation. *Genome Research*, *19*(6), 959–966. https://doi.org/10.1101/gr.083451.108

Lister, R., Mukamel, E. A., Nery, J. R., Urich, M., Puddifoot, C. A., Johnson, N. D., Lucero, J., Huang, Y., Dwork, A. J., Schultz, M. D., Yu, M., Tonti-Filippini, J., Heyn, H., Hu, S., Wu, J. C., Rao, A., Esteller, M., He, C., Haghighi, F. G., … Ecker, J. R. (2013). Global epigenomic reconfiguration during mammalian brain development. *Science*, *341*(6146), 1237905

Liu, C., Wang, M., Wei, X., Wu, L., Xu, J., Dai, X. I., Xia, J., Cheng, M., Yuan, Y., Zhang, P., Li, J., Feng, T., Chen, A. O., Zhang, W., Chen, F., Shang, Z., Zhang, X., Peters, B. A., & Liu, L. (2019). An ATAC-seq atlas of chromatin accessibility in mouse tissues. *Scientific Data*, *6*(1), 65. https://doi.org/10.1038/s41597-019-0071-0

Liu, G., Wang, W., Hu, S., Wang, X., & Zhang, Y. (2018). Inherited DNA methylation primes the establishment of accessible chromatin during genome activation. *Genome Research*, *28*(7), 998–1007.

Lorincz, M. C., Dickerson, D. R., Schmitt, M., & Groudine, M. (2004). Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. *Nature Structural and Molecular Biology*, *11*(11), 1068–1075.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550.

Maney, D. L. (2008). Endocrine and genomic architecture of life history trade-offs in an avian model of social behavior. *General and Comparative Endocrinology*, *157*(3), 275–282. https://doi.org/10.1016/j.ygcen.2008.03.023

Maney, D. L., Goode, C. T., Lange, H. S., Sanford, S. E., & Solomon, B. L. (2008). Estradiol modulates neural responses to song in a seasonal songbird. *The Journal of Comparative Neurology*, *511*(2), 173–186. https://doi.org/10.1002/cne.21830

Maney, D. L., Horton, B. M., & Zinzow-Kramer, W. M. (2015). Estrogen receptor alpha as a mediator of life-history trade-offs. *Integrative and Comparative Biology*, *55*(2), 323–331. https://doi.org/10.1093/icb/icv005

Maney, D. L., Merritt, J. R., Prichard, M. R., Horton, B. M., & Yi, S. V. (2020). Inside the supergene of the bird with four sexes. *Hormones and Behavior*, *126*, 104850. https://doi.org/10.1016/j.yhbeh.2020.104850

Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., & Chen, W. M. (2010). Robust relationship inference in genome-wide association studies. *Bioinformatics*, *26*(22), 2867–2873. https://doi.org/10.1093/bioinformatics/btq559

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *Embnet Journal*, *17*(1), 10. https://doi.org/10.14806/ej.17.1.200

McClay, J. L., Shabalin, A. A., Dozmorov, M. G., Adkins, D. E., Kumar, G., Nerella, S., Clark, S. L., Bergen, S. E., Hultman, C. M., Magnusson, P. K. E., Sullivan, P. F., Aberg, K. A., & van den Oord, E. J. C. G. (2015). High density methylation QTL analysis in human blood via next-generation sequencing of the methylated genomic DNA fraction. *Genome Biology*, *16*(1), 291. https://doi.org/10.1186/s13059-015-0842-7

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

Mendizabal, I., Berto, S., Usui, N., Toriumi, K., Chatterjee, P., Douglas, C., Huh, I., Jeong, H., Layman, T., Tamminga, C. A., Preuss, T. M., Konopka, G., & Yi, S. V. (2019). Cell type-specific epigenetic links to schizophrenia risk in the brain. *Genome Biology*, *20*(1), 135. https://doi.org/10.1186/s13059-019-1747-7

Mendizabal, I., & Yi, S. V. (2016). Whole-genome bisulfite sequencing maps from multiple human tissues reveal novel CpG islands associated with tissue-specific regulation. *Human Molecular Genetics*, *25*(1), 69–82. https://doi.org/10.1093/hmg/ddv449

Mendizabal, I., Zeng, J., Keller, T. E., & Yi, S. V. (2017). Body-hypomethylated human genes harbor extensive intragenic transcriptional activity and are prone to cancer-associated dysregulation. *Nucleic Acids Research*, *45*(8), 4390–4400. https://doi.org/10.1093/nar/gkx020

Merritt, J. R., Davis, M. T., Jalabert, C., Libecap, T. J., Williams, D. R., Soma, K. K., & Maney, D. L. (2018). Rapid effects of estradiol on aggression depend on genotype in a species with an estrogen receptor polymorphism. *Hormones and Behavior*, *98*, 210–218. https://doi.org/10.1016/j.yhbeh.2017.11.014

Merritt, J. R., Grogan, K. E., Zinzow-Kramer, W. M., Sun, D., Ortlund, E. A., Yi, S. V., & Maney, D. L. (2020). A supergene-linked estrogen receptor drives alternative phenotypes in a polymorphic songbird. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(35), 21673. https://doi.org/10.1073/pnas.2011347117

Morano, A., Angrisano, T., Russo, G., Landi, R., Pezone, A., Bartollino, S., Zuchegna, C., Babbio, F., Bonapace, I. M., Allen, B., Muller, M. T., Chiariotti, L., Gottesman, M. E., Porcellini, A., & Avvedimento, E. V. (2014). Targeted DNA methylation by homology-directed repair in mammalian cells. Transcription reshapes methylation on the repaired gene. *Nucleic Acids Research*, *42*(2), 804–821. https://doi.org/10.1093/nar/gkt920

Murao, N., Noguchi, H., & Nakashima, K. (2016). Epigenetic regulation of neural stem cell property from embryo to adult. *Neuroepigenetics*, *5*, 1–10. https://doi.org/10.1016/j.nepig.2016.01.001

Murrell, A., Heeson, S., & Reik, W. (2004). Interaction between differentially methylated regions partitions the imprinted genes Igf2 and H19 into parent-specific chromatin loops. *Nature Genetics*, *36*(8), 889–893. https://doi.org/10.1038/ng1402

O'Hagan, H. M., Mohammad, H. P., & Baylin, S. B. (2008). Double strand breaks can initiate gene silencing and SIRT1-dependent onset of DNA methylation in an exogenous promoter CpG island. *PLOS Genetics*, *4*(8), e1000155. https://doi.org/10.1371/journal.pgen.1000155

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, *33*(3), 290–295. https://doi.org/10.1038/nbt.3122

Price, A. J., Collado-Torres, L., Ivanov, N. A., Xia, W., Burke, E. E., Shin, J. H., Tao, R., Ma, L., Jia, Y., Hyde, T. M., Kleinman, J. E., Weinberger, D. R., & Jaffe, A. E. (2019). Divergent neuronal DNA methylation patterns across human cortical development reveal critical periods and a unique role of CpH methylation. *Genome Biology*, *20*(1), 196. https://doi.org/10.1186/s13059-019-1805-1

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

R Core Team (2019). *R: a language and environment for statistical computing*. R Core Team. Retrieved from https://www.R-project.org/

Robertson, K. D., & Jones, P. A. (2000). DNA methylation: past, present and future directions. *Carcinogenesis*, *21*(3), 461–467. https://doi.org/10.1093/carcin/21.3.461

Rodríguez-Paredes, M., & Esteller, M. (2011). Cancer epigenetics reaches mainstream oncology. *Nature Medicine*, *17*, 330. https://doi.org/10.1038/nm.2305

Roessler, K., Takuno, S., & Gaut, B. S. (2016). CG methylation covaries with differential gene expression between leaf and floral bud tissues of *Brachypodium distachyon*. *PLoS One*, *11*(3), e0150002. https://doi.org/10.1371/journal.pone.0150002

Schübeler, D. (2015). Function and information content of DNA methylation. *Nature*, *517*(7534), 321–326. https://doi.org/10.1038/nature14192

Sigurdsson, M. I., Smith, A. V., Bjornsson, H. T., & Jonsson, J. J. (2009). HapMap methylation-associated SNPs, markers of germline DNA methylation, positively correlate with regional levels of human meiotic recombination. *Genome Research*, *19*(4), 581–589. https://doi.org/10.1101/gr.086181.108

Smit, A. F. A., & Hubley, R. (2008–2015). RepeatModeler Open-1.0.

Smit, A. F. A., Hubley, R., & Green, P. (2013–2015). *RepeatMasker Open-4.0*. Retrieved from http://www.repeatmasker.org

Stadler, M. B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Scholer, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E. J., Gaidatzis, D., Tiwari, V. K., & Schubeler, D. (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature*, *480*(7378), 490–495.

Sun, D., Huh, I., Zinzow-Kramer, W. M., Maney, D. L., & Yi, S. V. (2018). Rapid regulatory evolution of a nonrecombining autosome linked

to divergent behavioral phenotypes. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(11), 2794.

Sun, D., Maney, D. L., Layman, T. S., Chatterjee, P., & Yi, S. V. (2019). Regional epigenetic differentiation of the Z Chromosome between sexes in a female heterogametic system. *Genome Research*, *29*, 1673–1684.

Sun, D., & Yi, S. V. (2015). Impacts of chromatin states and long-range genomic segments on aging and DNA methylation. *PLoS One*, *10*(6), e0128517. https://doi.org/10.1371/journal.pone.0128517

Tang, Y., Horikoshi, M., & Li, W. (2016). ggfortify: unified interface to visualize statistical result of popular R packages. *The R Journal*, *8*(2), 478–489.

Thomas, J. W., Caceres, M., Lowman, J. J., Morehouse, C. B., Short, M. E., Baldwin, E. L., Maney, D. L., & Martin, C. L. (2008). The chromosomal polymorphism linked to variation in social behavior in the white-throated sparrow (*Zonotrichia albicollis*) is a complex rearrangement and suppressor of recombination. *Genetics*, *179*(3), 1455–1468. https://doi.org/10.1534/genetics.108.088229

Thompson, M. J., vonHoldt, B., Horvath, S., & Pellegrini, M. (2017). An epigenetic aging clock for dogs and wolves. *Aging*, *9*(3), 1055–1068. https://doi.org/10.18632/aging.101211

Thorneycroft, H. B. (1966). Chromosomal polymorphism in the white-throated sparrow, *Zonotrichia albicollis* (Gmelin). *Science*, *154*(3756), 1571–1572. https://doi.org/10.1126/science.154.3756.1571

Thorneycroft, H. B. (1975). Cytogenetic study of white-throated sparrow, *Zonotrichia albicollis* (Gmelin). *Evolution*, *29*(4), 611–621.

Tuttle, E. M. (2003). Alternative reproductive strategies in the white-throated sparrow: behavioral and genetic evidence. *Behavioral Ecology*, *14*(3), 425–432. https://doi.org/10.1093/beheco/14.3.425

Tuttle, E. M., Bergland, A. O., Korody, M. L., Brewer, M. S., Newhouse, D. J., Minx, P., Stager, M., Betuel, A., Cheviron, Z. A., Warren, W. C., Gonser, R. A., & Balakrishnan, C. N. (2016). Divergence and functional degradation of a sex chromosome-like supergene. *Current Biology*, *26*(3), 344–350. https://doi.org/10.1016/j.cub.2015.11.069

Wogan, G. O. U., Yuan, M. L., Mahler, D. L., & Wang, I. J. (2020). Genome-wide epigenetic isolation by environment in a widespread Anolis lizard. *Molecular Ecology*, *29*(1), 40–55. https://doi.org/10.1111/mec.15301

Wu, H., Coskun, V., Tao, J., Xie, W., Ge, W., Yoshikawa, K., Li, E., Zhang, Y., & Sun, Y. E. (2010). Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes. *Science*, *329*(5990), 444–448. https://doi.org/10.1126/science.1190485

Wu, H., Xu, T., Feng, H., Chen, L. I., Li, B., Yao, B., Qin, Z., Jin, P., & Conneely, K. N. (2015). Detection of differentially methylated regions from whole-genome bisulfite sequencing data without replicates. *Nucleic Acids Research*, *43*(21), e141–e141. https://doi.org/10.1093/nar/gkv715

Wu, X., Galbraith, D. A., Chatterjee, P., Jeong, H., Grozinger, C. M., & Yi, S. V. (2020). Lineage and Parent-of-origin effects in DNA methylation of Honey Bees (*Apis mellifera*) revealed by reciprocal crosses and whole-genome Bisulfite sequencing. *Genome Biology and Evolution*, *12*(8), 1482–1492. https://doi.org/10.1093/gbe/evaa133.

Yi, S. V. (2017). Insights into epigenome evolution from animal and plant methylomes. *Genome Biology and Evolution*, *9*(11), 3189–3201. https://doi.org/10.1093/gbe/evx203

Yi, S., & Charlesworth, B. (2000). Contrasting patterns of molecular evolution of genes on the new and old sex chromosomes of *Drosophila miranda*. *Molecular Biology and Evolution*, *17*, 703–717.

Yu, G., Wang, L.-G., & He, Q.-Y. (2015). ChIPseeker: An R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*, *31*(14), 2382–2383. https://doi.org/10.1093/bioinformatics/btv145

Yuen, R. K., Peñaherrera, M. S., von Dadelszen, P., McFadden, D. E., & Robinson, W. P. (2010). DNA methylation profiling of human placentas reveals promoter hypomethylation of multiple genes in early-onset preeclampsia. *European Journal of Human Genetics: EJHG*, *18*(9), 1006–1012. https://doi.org/10.1038/ejhg.2010.63

Zeng, J., & Yi, S. V. (2014). Specific modifications of histone tails, but not DNA methylation, mirror the temporal variation of mammalian recombination hotspots. *Genome Biology and Evolution*, *6*(10), 2918–2929. https://doi.org/10.1093/gbe/evu230

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nussbaum, C., Myers, R. M., Brown, M., Li, W., & Liu, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biology*, *9*(9), R137. https://doi.org/10.1186/gb-2008-9-9-r137.

Zhou, Q., Ellison, C. E., Kaiser, V. B., Alekseyenko, A. A., Gorchakov, A. A., & Bachtrog, D. (2013). The epigenome of evolving *Drosophila* neo-sex chromosomes: dosage compensation and heterochromatin formation. *PLoS Biology*, *11*(11), e1001711. https://doi.org/10.1371/journal.pbio.1001711

Zinzow-Kramer, W. M., Horton, B. M., McKee, C. D., Michaud, J. M., Tharp, G. K., Thomas, J. W., Tuttle, E. M., Yi, S., & Maney, D. L. (2015). Genes located in a chromosomal inversion are correlated with territorial song in white-throated sparrows. *Genes Brain and Behavior*, *14*(8), 641–654. https://doi.org/10.1111/gbb.12252

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Sun D, Layman TS, Jeong H, et al. Genome-wide variation in DNA methylation linked to developmental stage and chromosomal suppression of recombination in white-throated sparrows. *Mol Ecol*. 2021;30:3453–3467. https://doi.org/10.1111/mec.15793