Supplemental Materials

Supplemental Figure Legends

Supplemental Fig.1: Generation and verification of H3K9R mutant genotype. Related to Figure 1. A) Progeny lacking endogenous histone genes were generated by crossing parents heterozygous for the *HisC* deletion and identified by GFP expression using the Gal4-UAS system. We analyzed *HisC* deletion mutants containing a transgenic array of 12 HWT or 12 H3K9R histone clusters integrated at the same VK33 *attP* site on chromosome 3. B) Anti-histone H3 western blot of total cell protein extracted from whole third instar larvae. C) Third instar larval polytene chromosome preparation stained with DAPI and anti-FLASH antibodies to mark the histone locus body (HLB), a nuclear body that recruits factors involved in replication-dependent histone mRNA biosynthesis. Both HWT and K9R transgenes form an HLB (arrowheads) at the expected location on chromosome 3L. D) PCR verification of transgene insertion into the VK33 *attP* transgene landing site. yw contains no VK33 landing site, and VK33 is the empty landing site. Arrows indicate primer locations used for PCR. E) Southern blot verification of histone cluster copy number in various HWT and K9R transgenes. We eliminated a XhoI site from the transgenic *H2A* gene arrays that is present in the H2A gene of the endogenous histone locus. XhoI digestion collapses the endogenous histone gene array to 5kb fragment but leaves the transgenic array intact. A 500bp sequence from *H2A* was used as a probe.

Supplemental Fig.2: FAIRE-seq signal correlates between replicates and occurs at characteristic regions. Related to Figure 2. A) FAIRE sequencing was performed in biological triplicate for HWT and K9R 3rd instar imaginal wing disc samples with three K9R and two HWT replicates meeting quality control standards (see Supplemental Experimental Methods). Normalized counts at each peak are shown in a heat scatter plot with red regions containing the highest number of peaks. R values shown are Pearson correlations. B) Venn diagrams showing peak overlap between replicates. C,D) Correlation analysis (C) and peak overlap (D) of HWT replicates and previously published FAIRE data in wing discs (McKay & Lieb 2013). E) Average HWT FAIRE signal at 5bp bins surrounding transcription start sites (TSS). Signal expressed as average reads per million (RPM). F) Percentage of total aligned reads in HWT and K9R replicates that uniquely mapped or could be mapped to multiple locations.

Supplemental Fig.3: K9R mutants lack DNA copy number changes and have increased FAIRE signal at constitutive heterochromatin. Related to Figure 2. A) Genomic DNA sequencing of HWT and K9R wing discs. Ratio of K9R to HWT read depth at merged FAIRE peak set (Supplemental Supplemental Fig.2) for chromosome 2 and 3. B) Ratio of K9R to HWT FAIRE-seq (top) and genomic DNA-seq (bottom) signal at the largely heterochromatic chromosome 4. C) FAIRE reads were aligned to the dm3 reference genome which contains scaffolds of repetitive heterochromatic regions separate from the largely euchromatic scaffolds. FAIRE signal expressed as Reads per Million (RPM) was used to calculate K9R/HWT ratio for each FAIRE peak. Boxplot shows average ratio for each dm3 scaffold. D) Comparison of HWT and K9R FAIRE signal at regions of K27me3, K9ac, K9me2, and K9me3 enrichment. Histone PTM enrichment was determined by peak calling from modENCODE 3rd instar larval ChIP-seq data (see Supplemental Experimental Procedures, Celniker et al., 2009). A merged FAIRE peak

data set from the two genotypes was used to separate FAIRE peaks that overlap a particular histone PTM peak from FAIRE peaks that did not overlap. Boxplots show the average K9R/HWT ratio of FAIRE signal for the two peak categories.

Supplemental Fig.4: HP1a relocalizes along chromosome arms in K9R mutants. Related to Figure 3. A) Heatscatter plots of HP1a ChIP signal within 1 kb windows for two HWT and three K9R biological replicates. Signal is expressed as reads per million (RPM), and R values indicate Pearson correlation. B) Comparison of HWT HP1a ChIP-seq samples to modENCODE HP1a ChIP-seq samples from 3rd instar larvae. HP1a signal at 1kb windows was normalized to total number of aligned reads and expressed as RPM. R value indicates Pearson correlation. C) Ratio of K9R to HWT normalized HP1a signal plotted versus genomic location in megabases (Mb). Centering of HP1a fold change near zero on chromosome arms supports the *D. virilis* chromatin spike-in normalization procedure. D) Metagene analysis of HP1a signal across all genes on chromosome arms (top) and in pericentromeres (bottom). Schematic indicates 1000 bp upstream (left line) and 500 bp downstream (left light grey box) of the transcription start site (bent arrow) and 500 bp upstream (right light grey box) and 1000 bp downstream (right line) of the transcription termination site (right edge of light grey box). The rest of the coding regions (dark grey box) were scaled into an equal number of windows. E) HP1a ChIP-seq signal from wing discs at peaks called by MACS2. Blue line indicates loess regression line.

Supplemental Fig.5: Protein-coding genes have similar expression levels in HWT and K9R genotypes. Related to Figure 4. A) Comparison of RNA-seq data from three biological replicates of HWT and K9R wing discs. Scatterplots show FPKM (Fragments Per Kilobase of transcript per Million mapped reads) calculated by Cufflinks at each gene. R value represents Pearson correlation. B) Comparison of an average of HWT RNA-seq samples to wing disc RNA-seq from McKay & Lieb 2013. R value represents Pearson correlation. C) K9R/HWT ratio of expression for 46 genes located in heterochromatin on chromosome 2 that were identified using data from Corradini et al. 2007. Only genes with transcripts included in the dm6 reference transcriptome were included in the analysis. Red indicates the two genes, *CG30440* and *chitinase 3 (Cht3),* with significantly different expression between K9R and HWT samples as determine by DESeq2 (p value<0.05). *Concertina (cta), light (lt), rolled (rl)* are also indicated. D) Scatterplots of K9R/HWT expression ratios for transcripts that do not overlap (top) or overlap (bottom) H3K9ac enriched regions. Only 4% of transcripts from genes containing H3K9ac are differentially expressed (210/5260) compared to 14% of transcripts that are not associated with H3K9ac (950/6563). Moreover, H3K9ac is not associated with changes in FAIRE signal (see Supplemental Fig.3D). Red dots indicate statistical significance as determined by DESeq2 ($p<0.05$).

Supplemental Fig.6: Changes in gene expression correlate with changes in FAIRE and HP1a signal. Related to Figure 4. A) Transcripts with significantly different expression between HWT and K9R as called by DESeq2 ($p<0.05$) were separated into one of nine chromatin states based on classification in Kharchenko et al. 2010. The percentage of transcripts that fall into each category is compared to the average of 25 iterations of randomly selected transcripts. Error bars indicate standard deviation. B) K9R/HWT ratio of RNA-seq signal for transcripts identified in Cufflinks transcriptome assembly but not in reference transcriptome. Red indicates statistical significance called by DESeq2 ($p<0.05$). C) Shown in red is the percent of

differentially expressed transcripts that overlap or are within 1 kb, 5 kb, or 10 kb of a statistically different FAIRE peak between HWT and K9R as determined by edgeR ($p<0.01$). 25 random selections of transcripts were put through the same pipeline and the average percent of transcripts within the various distances are shown in grey. Error bars indicate standard deviation. D) Similar analysis to C was performed with HP1a ChIP-seq peaks. E) K9R/HWT ratio of RNA-seq signal for genes on chromosome four. Red dots indicate statistical significance called by DESeq2 ($p<0.05$). F) Pie charts showing fraction of transcripts that were annotated as protein-coding (black) or non-coding (grey) or were unannotated in reference transcriptome but identified in transcriptome assembly (green). Left pie chart shows all differentially expressed transcripts increased in K9R relative to HWT, and right pie chart shows all transcripts in the newly assembled transcriptome.

Supplemental Fig.7: Transposons are activated and mobilized in K9R mutants. Related to Figures 5 and 6. A) Genome browser shot of sequencing reads mapping to the 42AB piRNA cluster from anti-HP1a immunoprecipitated (IP) chromatin samples compared to input samples. Top rows show uniquely mapping reads (MAPQ>10) while bottom rows show multiple mapping reads randomly distributed to one of a maximum of 10 possible assignments. HP1a ChIP-seq reads are enriched in the HWT sample (blue) but not the K9R sample (red). Signal represented as reads per million (RPM). B,C) Transposon FAIRE-seq and RNA-seq data from Figure 5A and B with various *gypsy* transposon families labeled. D,E) Whole genome sequencing data from wing disc (D) or larvae (E) was analyzed for transposon insertion or depletion events using TIDAL (Rahman et al. 2015). To examine the effect of genome coverage on detection we randomly subsampled reads from each genotype. Scatterplot shows the number of insertions (solid line) or depletions (dashed line) at various fold coverages (x axis) in K9R (red) and HWT (blue) samples. Venn diagrams indicate the number of insertions or depletions unique to or shared by each genotype. See also Supplemental Experimental Procedures.

Supplemental Materials and Methods

Stocks (with origin)
yw (Bloomington Stock Center #6598)
Δ*HisC, UAS-2xEYFP / CyO* (gift from Alf Herzig)
Δ*HisC, twi-GAL4 / CyO* (gift from Alf Herzig)
yw^{122}*; ΔHisC, FRT40A / CyO twi-GAL4, UAS-GFP* (gift from Alf Herzig)
yw^{122}*; Ubi-GFPS65Tnls, FRT40A / CyO* (BSC# 5629)
w,ovo ; ScO,CyO (ovo=10185-3-8) (gift from Stephen Helfand)

Antibodies (species, source) (concentration)
α-H3K9me2 (mouse, Abcam 1220) (1:500)
α-H3K9me3 (rabbit, Active Motif 39161) (1:1000)
α-HP1-immunofluorescence (mouse, DSHB C1A9) (1:1500)
α-HP1-ChIP (rabbit, Covance PRB-291C)
α-H3 (rabbit, Abcam 1791) (1:30,000)
α-Tubulin (mouse, Sigma T6074) (1:30,000)
α-FLASH (rabbit, (Yang et al. 2009)) (1:1000)

Molecular confirmation of 12xK9R transgenes
PCR analysis of insertion into VK33 was monitored with the following primers:

attPF 5'-CCTTCACGTTTTCCCAGGT-3'
attPR 5'- CGACTGACGGTCGTAAGCAC-3'
attBF 5'- GGAACTAGGCTAGCATAACTTCGTA-3'
attBR 5'- AGTGTGTCGCTGTCGAGATG-3'

Amplification of intact landing site only occurs with attPF/attPR primers. If transgene is inserted into landing site, amplification occurs only with attPF/attBR or attBF/attPR primer sets. Southern blots were performed as in McKay et al. 2015.

Immunofluorescence
Salivary gland polytene chromosome spreads for Supplemental Fig. 1 were performed as previously described (Cai et al. 2010). Salivary glands from different genotypes were prepared on the same slide to control for differences in squashing. FLASH localizes to histone genes and was used to distinguish *yw* polytene chromosomes containing endogenous histone genes at the HisC locus located on chromosome 2L from polytene chromosomes containing ectopic transgenic histone genes on chromosome 3L. anti-K9me2 staining was used to distinguish K9R from HWT. Images shown are single confocal images taken at a constant gain on a Leica TCS SP5 AOBS UV/spectral confocal laser-scanning system mounted on an inverted DM IRE2 microscope.

FAIRE
Reads were aligned to the reference genome (dm6 Release 6.04) using Bowtie2 default parameters (Langmead and Salzberg 2012). FAIRE was performed in biological triplicate for both K9R and HWT genotypes with one HWT replicate failing to meet quality standards (only

37% of reads aligned to the reference genome and aligned reads had low correlation of FAIRE signal at peaks (R=0.8)). Since the remaining replicates were highly correlated to each other (R>0.94) with 67-90% of reads aligned, reads from each replicate of a particular genotype were pooled for subsequent analyses. Peaks were called on individual replicates using MACS2 with a shift size of 110bp and a cutoff of 0.01 (Zhang et al. 2008). A union peak set was generated by concatenating individual sample peaks and merging peaks with BEDTools (Quinlan and Hall 2010). Peak overlaps between replicates were calculated with BEDTools intersect. Coverage at each peak in the union set was normalized to sequencing depth. Average Counts per Million values are calculated by averaging the normalized HWT and K9R FAIRE signal at a particular peak. Genomic input DNA from the FAIRE procedure was sequenced on an Illumina MiSeq. The alignment pipeline described above was used to examine copy number at the union peak set.

ChIP

Chromatin from 3rd instar larvae was prepared with a protocol modified from Soruco et al. 2013. Three biological replicates for both the HWT and K9R genotypes were prepared concurrently with a sample of *Drosophila virilis* larvae. For each replicate, 150 larvae were Dounce homogenized in PBS containing 1x of Halt protease inhibitor cocktail (Thermo Scientific). The homogenized samples were filtered through Miracloth and brought to a final volume of 40 mL with PBS+protease inhibitor. Formaldehyde crosslinking and subsequent washes were performed as in Soruco et al. 2013. Chromatin was sonicated to 100-500bp fragments using a Branson sonifier 450. 6x RIPA buffer was added to a final concentration of 1x (140mM NaCl,10mM Hepes pH 8.0,1mM EDTA pH 8.0, 1% Triton X-100, 0.1% sodium deoxycholate, 1mM DTT, 0.5mM PMSF, 1x protease inhibitor cocktail). 10 µg of chromatin was used for each immunoprecipitation, and 0.25 µg of *D. virilis* chromatin was added to each sample. Each replicate was precleared with protein A/G magnetic beads while HP1a antibody (Covance) was incubated with beads according to manufacturer's instructions (EMD Millipore). For each genotype, 0.33 µg of chromatin was taken from each replicate and pooled to use as an input. Antibody/bead mixture was then incubated with chromatin extract overnight at 4C (7 µL of antibody used for each immunoprecipitation). Subsequently, beads were washed 2x in 1x RIPA buffer for 5 minutes, 1x in 1xRIPA+1M NaCl for 5 minutes, 1x in 1x RIPA+0.5M NaCl for 10 minutes, 1x in TEL buffer for 10 minutes (0.25M LiCl, 1% NP-40, 1% sodium deoxycholate, 1mM EDTA pH 8.0, 10mM Tris-HCl pH 8.0), and 2x in TE buffer for 5 minutes. Immunocomplexes were eluted in 300 µL of elution buffer (1%SDS, 250mM NaCl in TE) at 65C for 15 minutes with intermittent vortexing. Input samples were brought to a final volume of 300 µL with elution buffer. Next, all samples were treated with 100 µg/mL of RNaseA for 30 minutes at 37C and then 200 µg/mL of proteinase K for 2 hours at 55C. Crosslinks were reversed overnight at 65C and DNA was then ethanol precipitated. Libraries were prepped with 20ng of DNA using ThruPLEX DNA-seq Kit (Rubicon Genomics) and sequenced on an Illumina HiSeq2500 at the UNC High-Throughput Sequencing Core. Sequencing data was analyzed similarly to FAIRE datasets with the exception that ChIP signal was calculated within 1kb windows instead of peaks due to broad redistribution of HP1a that precluded peak calling along chromosome arms.

Transposon mobilization

Analysis of transposon mobilization in sequencing datasets from Supplemental Fig. 7 was performed with TIDAL (Rahman et al. 2015). We also used TEMP to analyze paired-end reads

from wing discs but could not analyze the single-end reads from whole larvae with this program (Zhuang et al. 2014). Comparable to TIDAL analysis in wing discs, the number of insertions or deletions was similar in the two genotypes; however, TEMP called 4 times the number of insertions events. For insertions events, 678 events were HWT specific, 607 events were shared, and 544 events were K9R specific. For depletion events, 41 were HWT specific, 111 were shared, and 39 were K9R specific. We note that the interpretation of these data is complicated as the overlap in called transposon insertion or depletion events across programs is minimal (Song et al. 2014; Rahman et al. 2015). Additionally, we assayed mobilization in somatic tissues that will be mosaic for individual mobilization events that likely represent a very small fraction of the total cells in any given animal, limiting our detection ability.

Supplemental References

Celniker SE, Dillon LAL, Gerstein MB, Gunsalus KC, Henikoff S, Karpen GH, Kellis M, Lai EC, Lieb JD, MacAlpine DM, et al. 2009. FEATURE Unlocking the secrets of the genome. *Nature* **459**: 927–930.

Quinlan AR, Hall IM. 2010. BEDTools : a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842.

Song J, Liu J, Schnakenberg SL, Ha H, Xing J, Chen KC. 2014. Variation in piRNA and Transposable Element Content in Strains of Drosophila melanogaster. *Genome Biol Evol* **6**: 2786–2798.

Yang X, Burch BD, Yan Y, Marzluff WF, Dominski Z. 2009. FLASH , a Proapoptotic Protein Involved in Activation of Caspase-8 , Is Essential for 3 0 End Processing of Histone Pre-mRNAs. *MOLCEL* **36**: 267–278.

Zhuang J, Wang J, Theurkauf W, Weng Z. 2014. TEMP: A computational method for analyzing transposable element polymorphism in populations. *Nucleic Acids Res* **42**: 6826–6838.