Chromatin changes induced by a *Plasmodium falciparum* infection in *Anopheles gambiae*

José L. Ruiz[1,2], Rakiswendé S. Yerbanga[3], Thierry Lefèvre[3,4], Jean B. Ouedraogo[3], Victor G. Corces[5] and Elena Gómez-Díaz[1,2]

1 Estación Biológica de Doñana (EBD), Consejo Superior de Investigaciones Científicas, 41092 Seville, Spain

2 Instituto de Parasitología y Biomedicina "López-Neyra" (IPBLN), Consejo Superior de Investigaciones Científicas Granada, 18016, Spain

3 Institut de Recherche en Sciences de la Santé (IRSS), 01 BP 171 Bobo Dioulasso, Burkina Faso

4 MIVEGEC, IRD, CNRS, University of Montpellier, Montpellier, France

5 Department of Biology, Emory University, 1510 Clifton Road NE, Atlanta, GA 30322, USA.

Running title: Malaria-induced chromatin changes in mosquitoes

Key words: human malaria, mosquitoes; histone modifications; gene regulation, ChIP-seq, RNA-seq, transcriptome, epigenome.

Correspondence

Elena Gómez-Díaz

Institute of Parasitology and Biomedicine López-Neyra (IPBLN) , CSIC

Avda. del Conocimiento, 17

618016 Armilla, Granada, Spain

Phone number +34 954466700

Fax number +34954621125

Email: elena.gomez@csic.es

ABSTRACT

Infection by the human malaria parasite leads to important changes in phenotypic traits related to vector competence. However, we still lack a clear understanding of the underlying mechanisms and in particular, of the epigenetic basis for these changes. We have examined genome-wide distribution maps of H3K27ac, H3K9ac, H3K9me3 and H3K4me3 by ChIP-seq and the transcriptome by RNA-seq, of midguts from *Anopheles gambiae* mosquitoes infected by the human malaria parasite *Plasmodium falciparum*. We report 15,916 regions containing differential histone modification enrichment, of which 8,339 locate at promoters and/or intersect with genes. The functional annotation of these regions allowed us to identify infection responsive genes showing differential enrichment in various histone modifications such as CLIP proteases, anti-microbial peptides encoding genes and genes related to melanization responses and the complement system. Further, the motif analysis of differential histone enriched regions predicts binding sites that might be involved in the cis-regulation of this regions such as Deaf1, Pangolin and Dorsal transcription factors (TFs). Some of these TFs are known to regulate immunity gene expression in *Drosophila* and are involved in the Notch and JAK/STAT signaling pathways. The analysis of malaria infection-induced chromatin changes in mosquitoes is important not only to identify regulatory elements and genes underlying mosquito responses to a malaria infection but also for possible applications to the genetic manipulation of mosquitoes and extension to other mosquito-borne systems.

INTRODUCTION

Mosquitoes are the most medically important arthropod vectors of disease. Among mosquito-borne diseases malaria, caused by protozoan parasites of the genus *Plasmodium*, is the deadliest. The human malaria parasite *P. falciparum* is the most prevalent agent in Africa and it is responsible for at least 200 million acute cases worldwide and around half a million deaths each year [1]. It is transmitted by *Anopheles* mosquitoes, being *A. gambiae* the main vector. At a molecular level, *P. falciparum* infection induces drastic and rapid changes in gene expression in mosquito tissues that relate to functions involved in immunity, development, physiology, and reproduction [2]. The immune response during infection is the best characterized of these pathways, and occurs in the midgut of infected mosquitoes. This involves among others the activation of genes involved in epithelial nitration responses, melanization and the complement-like system [3-5]. Furthermore, infection in mosquitoes impacts life-history traits such as vector competence, i.e. the ability of the mosquito to acquire, maintain, and transmit the parasite, and prime subsequent infections [6-8]. There is substantial variability in the responses of the mosquito to a *P. falciparum* infection that depends on the genetic and environmental contexts [9, 10]. However, the mechanisms that regulate phenotypic responses to infection in mosquitoes, and that might mediate memory of the malaria infection-phenotype, are little understood.

Chromatin-associated processes participate in the regulation of gene expression during development as well as in the differentiated tissues of the adult organism. These processes are sensitive to environmental stimuli, they are more or less transitory and can be potentially inherited, allowing living organisms and individual cells to continuously integrate internal and external inputs and to mediate responses through gene regulation [11, 12]. Amongst these, post-translational modifications of histones (hPTMs) impact the structure and/or function of chromatin, with different histone marks yielding distinct functional consequences [13]. For example, in *Drosophila*, as in other

organisms, H3K27ac, H3K9ac and H3K4me3 are linked to gene activation and localize at promoters, whereas H3K9me3 is associated with silencing and occupies broader regions [14]. Various combinations of active and repressive histone modifications define chromatin states that are linked to gene function [15-17]. These modifications can remain, leaving a record of gene activity, i.e. epigenetic memory, that affect or prime the transcriptional response later in life [12, 18-24]. In mosquitoes, despite their relevance for human health, there is very little knowledge on chromatin regulation and its link to mosquito immunity, physiology and behavior [25]. In a previous study, we characterized genome-wide occupancy patterns of various histone modifications and established a link between hPTMs and gene expression profiles in midguts and salivary glands of the human malaria vector *A. gambiae* [26]. In addition, recent reports reveal the role of various transcription factors such as REL2, lola and Deaf1, in *A. gambiae* immune defenses [27, 28]. These findings have set the stage for additional studies aimed at understanding how the chromatin landscape is altered in *P. falciparum* infected mosquito tissues and which are the molecular players involved in these malaria-induced responses.

Available data on the phenotypic and transcriptional responses of mosquitoes to infection by *Plasmodium* is built on the use of mosquito-parasite combinations that in most cases differ from those found in nature. In these studies, infections often take place under standard laboratory conditions, using laboratory-adapted parasite clones and commercially available laboratory mosquito strains. Such experiments are useful to distinguish the contribution of parasite and mosquito genetic factors and also the influence of various environmental variables on the infection output, but they may not reflect the complexity of interactions that take place in natural conditions [29]. In this context, field-based studies are critical as they offer the advantage of allowing a more realistic picture of the molecular interactions in the context of natural transmission.

In this study we aim to provide a comprehensive understanding of the chromatin and the transcriptional responses induced by a *P. falciparum* infection in *A. gambiae* in the conditions found in a malaria endemic area in Africa. For this purpose, we compared genome-wide maps of

histone modification profiles in infected and control mosquito tissues to identify chromatin state transitions associated to infection, and examined the expression pattern of genes that annotate to differential histone modified regions. The motif enrichment analysis on these differential regulatory regions allows us to predict the binding sites of several transcription factors, some of which have been shown to be involved in mosquito immune responses by previous studies.

RESULTS

Chromatin states in malaria infected mosquitoes

A. gambiae were fed with blood obtained from malaria-infected human volunteers in Burkina Faso. Midguts from infected and uninfected (blood fed control) female mosquitoes were dissected. On the pooled samples we carried out ChIP-seq using antibodies to several histone modifications: H3K9ac, H3K27ac, H3K4me3 and H3K9me3. The identified histone modification enriched regions, MACS2 peaks, are listed in Table S2. We find similar numbers of peaks independently of the infection status and a high correlation in ChIP-seq profiles (Figure S1A), indicating that histone modification profiles are comparable between conditions. We annotated ChIP-seq peaks to genomic features using HOMER software (Figure 1A). Distance analysis on ChIP-seq peaks shows that significantly enriched upstream histone modifications regions are mostly located in a window of 2 Kb from the translation start codon (ATG) of *A. gambiae* genes, being scarce at distances greater than 2 Kb (Figure 1B). Based on these findings, we annotated peaks to the promoters of nearby genes when located less than 2 Kb upstream, and used this annotation for all subsequent analysis. Results show that most ChIP-seq peaks correspond to H3K9me3 marked regions (37,343) followed by H3K27ac (35,217), H3K4me3 (19,945) and H3K9ac (6,131) (Table S2).

ChromHMM segmented the genome into 4 distinct chromatin states based on relative enrichment levels of H3K9ac, H3K27ac, H3K4me3 and H3K9me3 at each condition, that we named as follows: hPTM depleted, i.e. low levels of enrichment for all hPTMs, active (H3K9ac/H3K27ac and H3K4me3), repressed (H3K9me3), and bivalent (H3K4me3/H3K9me3) (Figure 1C, Figure

S1B). As expected, most of the genome is in a depleted state, whereas gene bodies and promoters display enrichment in active state 7 (Figure 1D, Figure S1C). Following this categorization, we investigated the association between various combinations of active and repressive histone modifications at promoters and gene bodies and gene expression regulation at each condition. As expected, active chromatin is associated to expressed genes, whereas chromatin marked with H3K9me3 associates to non-expressed genes. Genes showing a bivalent enrichment pattern (H3K4me3 and H3K9me3) are generally expressed at low levels (Figure 1E, Figure S2A). This is corroborated by a positive and statistically significant correlation between levels of enrichment in active histone modifications (H3K9ac, H3K27ac, H3K4me3) and mRNA levels of ChIP-seq peaks annotated genes, whereas the correlation is negative for H3K9me3 (Figure 1E, Figure S2A-B). Following what has been described for *Drosophila* [30], we classified genes according to the gene structure, i.e. length of the coding region and length of intronic segments. We observed that expressed genes with long introns and relatively short coding regions show broad H3K27ac and H3K9ac domains. In contrast, expressed genes with more uniformly distributed coding regions show a more localized H3K27ac/H3K9ac enrichment through the gene body and higher enrichment in H3K4me3 (Figure 1F, Figure S2B).

Differential histone modifications enriched regions associated to a P. falciparum infection

We find a considerable overlap in histone modifications profiles between infected and non-infected mosquitoes, with more than 30,000 common ChIP-seq peaks. However, a portion of the peaks appears to be condition-specific (8,234 in infected and 14,138 in the non-infected) (Figure 2A). Based on this observation, we used the diffReps software to investigate more in deep localized changes in hPTMs enrichment that might occur in response to a *P. falciparum* infection. In total for all four histone modifications, we identify 15,916 regions containing significantly different levels of ChIP-seq signal (P-value < 10E-4). The number of diffReps regions is similar for H3K9ac (2,396) and H3K4me3 (2,837), whereas H3K27ac and H3K9me3 display a larger number of differential regions

(4,810 and 5,873, respectively) (Table S3). Differential regions involving active histone modifications are primarily distributed nearby genes, upstream and downstream, or in the intron. But they also occupy distal intergenic sites, particularly in the case of H3K9me3 (Figure 2B-C). To these differential regions we applied a series of filtering thresholds (see Methods) to get a high confidence set that we classified according to chromatin states transitions (ChromHMM) (Figure 2D). In the majority of cases, the diffReps change involves an enrichment or depletion in a certain histone modification mark without a chromatin state change, but we also report chromatin state shifts between conditions: regions that are active upon infection or regions that are marked with active chromatin marks in control mosquito tissues and change to depleted in the infected. There is also a considerable proportion of regions that switch between the depleted and H3K9me3 enriched states (Figure 2D).

The gene ontology analysis shows that diffReps-annotated genes appear significantly enriched in GO terms related to development, transcription regulation and metabolism (Table S3). In addition to this analysis, we also looked for coincidences between the diffReps-annotated genes and genes that have been reported to be involved in immune responses to infection [31, 32]. Amongst diffReps-annotated genes, we find 133 immune genes (26 considering the high confidence set of diffReps peaks). Genes differing between conditions in their histone modification profiles encode proteins involved in apoptosis (IAP3 and IAP7), Clip-domain serine proteases and serine protease inhibitors (CLIPC, CLIPE, SRPN10 and SRPN4), C-type lectins (CTLs, Figure 2F), antimicrobial peptides (DEF1, Figure 2G), Scavenger Receptor (SRCR domain) with Lysyl Oxidase domain (SCRAL1), and components of Toll, NF, and peptidoglycan recognition protein LC/immune deficiency (PGRP-LC/IMD) signaling pathways, among others (Table S3, Figure S3). Most genes involved in immune responses are enriched in the same chromatin states in both conditions but display a change in the relative abundance of active H3K4me3/K9ac/K27ac, and/or the repressive

H3K9me3 (Figure 2E, Table S3). There are 9 genes within the high-confidence set of diffReps-annotated genes that display differential chromatin state between conditions like the long caspase CASPL2 encoding gene and fibrinogen-related protein encoding gene (Table S3, Figure S3B-C).

Complex relationship between chromatin and gene expression changes

In order to investigate the functional significance of the differentially enriched regions, we selected high confidence differential hPTMs regions that overlap gene bodies or locate up to 2 Kb upstream of genes, 1,208 genes for all four hPTMs, and applied to the whole promoter-gene region a soft clustering approach using Mfuzz. This analysis allowed us to group diffReps-annotated genes based on unique profiles of hPTM enrichment (Figure S4) and to examine the correlation between changes in the histone modification patterns and the expression status. We find that genes/promoters differentially enriched in a condition in active histone marks (H3K27ac, H3K9ac and H3K4me3) tend to be more highly expressed in that condition, whereas promoters that are marked with repressive H3K9me3 or bivalent H3K4me3/H3K9me3 tend to be less expressed (Figure 3A, Figure S5A, Table S4). However, when comparing the ratio of histone enrichment versus the ratio of gene expression values between infected and control mosquitoes, the coefficient of correlation is low and non-significant, meaning that the infection condition influences to a different extent chromatin and gene expression patterns. This is clearly shown when examining the ratio of enrichment of various hPTMs in the infected relative to the control, for infected expressed genes (left panel) or control expressed genes (right panel) (Figure 3B). The ratio is in the same direction, above or below 1 for infected and control Log2 values respectively, only in the case of the H3K9me3 (Figure 3B). Nevertheless, this integrative analysis identified 278 genes in which the differential active or repressive histone modification enrichment coincides with a shift in the gene expression pattern between the infected and control condition (Table S4, Figure 3C, Figure S3B).

In addition to the analysis on the diffReps-annotated genes, we also examined differentially expressed genes and look for differential chromatin marks between conditions. The DESeq2 analysis on the RNA-seq data reveals 713 significantly different expressed transcripts (P-value < 0.05, 184 up-regulated genes and 529 down-regulated genes in the infected vs. the control condition) (Table S5). We find 105 differential expressed genes that contain a diffReps region annotated to the promoter or the gene body. Of those, there are 72 in which the sense of the change (active or repressive histone marks) agrees with the functional prediction (up or down-regulation) (Table S5). Same as above, the switch in the expression status is generally linked to changes in the H3K9me3 enrichment levels (Figure S6A). Examples of diffReps/DESeq2 genes are IAP7 (AGAP007293) and Argonaute 4 (AGAP011717). IAP7 is upregulated in infection and H3K9me3 is depleted in this condition compared to the control (Figure 4AX), while AGAP011717 is expressed at higher levels in the non-infected and this is associated to a gain in active histone modification marks, mainly H3K27ac (Figure 4B).

Motif analysis of differential histone enriched regions identifies transcription factor binding sites involved in transcriptional responses to infection

On the set of ChIP-seq high confidence diffReps regions that coincide with MACS2 peaks of significant enrichment, 2,018 peaks for all four histone modification marks (Table S3), we conducted DNA-binding motif enrichment analysis. For histone modification peaks such an approach is useful because it allows the discovery of unanticipated sequence signals associated with specific histone marks, like transcription factor binding sites. This analysis revealed multiple novel motifs that are significantly enriched in the sequences of H3K9ac, H3K27ac, H3K4me3 and H3K9me3 differentially enriched peaks. Table 1 lists the novel motifs identified by HOMER on the set of diffReps peaks for each histone modification mark and their similarities with known TF binding sites previously described in *Drosophila*. We find that the binding sites predicted on ChIP-seq peaks show similarities with transcription factors such as Deaf1, pangolin (pan) and dorsal (dl), linked to

immunity gene expression regulation in *Drosophila*. We also find binding sites for several transcription factors associated to developmental functions like Caudal (cad), and reproduction like Eip74EF (Table 1).

As a validation of our strategy to predict regulatory sites and TF binding, we report the overlap between FAIRE-seq peaks determined in *A. gambiae* hemocytes in a previous study, with our histone modifications ChIP-seq peaks in malaria infected tissues. Even if the experimental conditions and cell types are not comparable, we find 9,136 MACS2 ChIP-seq peaks and 2,690 diffReps regions (10% and 17% of the total, respectively) that intersect with FAIRE-seq peaks.

These results collectively suggest that some of the transcription factors reported in our study are probably involved in chromatin remodeling processes and the regulation of transcription of the genes that are elicited in mosquitoes in response to a malaria infection.

DISCUSSION

Human malaria is a mosquito-borne disease responsible for around half-million deaths per year, being *A. gambiae* the main disease vector in Africa [1]. A malaria infection alters the phenotype and vector competence of mosquitoes with consequences for transmission and malaria epidemiology. However, the molecular players that regulate malaria infection-triggered responses are still poorly known. A considerable amount of work exists on the genomic basis of mosquito resistance to infection (reviewed by [33]), but there is a dearth of epigenomic studies in mosquitoes on the relationship between chromatin and gene expression regulation. In a previous study, we characterised for the first time genome-wide profiles of various histone modifications in *A. gambiae* and compared these patterns with chromatin maps published for *Drosophila* [26]. Here, we go a step further and performed an integrative analysis of ChIP-seq and RNA-seq data in the context of a malaria infection and in the natural conditions of transmission of the disease.

In the study of global patterns of histone modification marks in infected mosquitoes, we report various chromatin states that link to functional gene expression: H3K9ac/H3K4me3/H3K27ac histone acetylation marks associated to the promoters of active genes and repressive H3K9me3 associated to silent genes. In agreement with previous studies in *Drosophila*, our results in mosquitoes show that gene structure is related to differences in the distribution of H3K27ac and H3K4me3 enrichment in the gene body, being the most common difference in the chromatin state of expressed genes that contain long introns relative to exons. We also report bivalent chromatin domains, that have both repressive H3K9me3 and activating H3K4me3 histone marks in the same region. This pattern is typical of transcription factor genes that are expressed at low levels, and have been also found associated to genes involved in development as well as gene imprinting [34, 35]. Despite the majority of peaks annotate to promoters or gene bodies, we report a considerable number of ChIP-seq peaks (~40%) that locate more than 2 Kb apart of the nearest gene. It has been described that H3K27ac associates with active enhancers in *Drosophila* and other model organisms [36, 37]. In mosquitoes, we identified a number of distal and intragenic H3K27ac-enriched regions that could be marking enhancer-like sites. Future ChIP studies on enhancer-specific marks combined with chromatin accessibility mapping are required to confirm the presence of enhancer-like elements in mosquitoes.

Once we showed a relationship between chromatin composition and function, our main purpose was to identify malaria-induced chromatin changes in mosquitoes. We could identify 15,916 histone modified regions (2,564 considering the high confidence set of peaks) that appear differentially methylated or acetylated upon infection. Various chromatin states: active, repressive and bivalent chromatin, were identified in the set of differential histone marked regions. We also observed that differential regions generally display enrichments or depletions of individual marks at specific gene segments, but maintaining the same chromatin state. Importantly, there are 107 promoter or intragenic regions and 133 more distal sites (>2 Kb), that annotate to genes that are implicated in different pathways of mosquito innate immunity either as activators or inhibitors of various responses (see [38, 39] for a review). These include apoptosis, Clip-domain serine proteases and serine protease

inhibitors, antimicrobial peptides, enzymes that catalyze generation and detoxification of reactive oxygen species, and components of Toll, NF and IMD immune signaling pathways.

The integration of differentially histone modified regions with expression levels of the annotated genes resulted in the identification of 278 malaria responsive genes. These genes showed local differences in the enrichment of specific active/repressive histone marks that correlate with gene expression changes in the same direction. However, this is not the general rule and most of the differential regions by ChIP-seq do not display noticeable differences in gene expression, and the other way around for differentially expressed genes. It might be that there is a threshold enrichment level that is necessary to activate transcription. It is also unexpected that comparing infected with control, only a few of these malaria responsive genes correspond to immune factors. A possible explanation is that the majority of the immune factors that have been described belong to the early innate immunity response that takes place between 2-24h post-infection, at the ookinete stage, where the most part of the parasite recognition and killing occur [40, 41]. Contrary, our samples correspond to the oocyst stage, approximately six and seven days after an infective blood feeding, and the immune factors playing a role at this stage still remain poorly characterized [42].

Cis-regulatory elements are implicated in the control gene expression because they contain specific DNA sequences that are binding sites for transcription factors and other chromatin remodelers and appear often enriched in certain histone modification marks. These elements are well mapped in *Drosophila* [43], but very few have been identified in mosquitoes [27, 44]. The analysis of differential ChIP-seq peaks located at promoters or gene bodies identified significant enrichment in binding sites that match consensus sequences of TFs previously described in *Drosophila* like transcription factor deformed epidermal autoregulatory factor-1 (DEAF-1) [45]. Indeed, DEAF-1 is an important regulator of *Drosophila* immunity that could act as enhancer by inducing genes encoding antimicrobial peptides [46]. Another example that we report in this study is dorsal (dl), a TF that functions downstream of the Toll pathway [45]. Finally, as a validation of our motif analysis, we

find that a portion of the differential histone modified ChIP-peaks matches FAIRE-seq peaks described by a previous study on hemocytes [27]. In that study, authors reported that FAIRE sequences were enriched in binding sites for DEAF-1, which is one of the top motifs reported in the present study. New approaches to profiling chromatin accessibility such as ATAC-seq will be useful to further characterize cis-regulatory sequences and TF binding “in vivo”.

In summary, this study charts genomic landscapes of various active and repressive histone modifications in infected mosquitoes and integrates these profiles with RNA-seq data to quantify gene expression. This way, we have identified malaria responsive genes that display changes in the abundance of specific histone modification marks. However, the relationship between chromatin and gene expression changes at differential regions is complex, and only a subset of genes shows correlated patterns that agree with the predicted function. Further research to identify regulatory sequences associated to these changes and the transcription factors that associate to them could provide new molecules and targets for vector control.

METHODS

Mosquito rearing and experimental infections

Three- to five-day-old female *A. gambiae* mosquitoes were sourced from an outbred colony established in 2008 and repeatedly replenished with F1 from wild-caught mosquito females collected in Kou Valley (11°23'14"N, 4°24'42"W), 30 km from Bobo-Dioulasso, south-western Burkina Faso (West Africa). Mosquitoes were maintained under standard insectary conditions (27 ± 2°C, 70 ± 5% relative humidity, 12:12 LD). Two independent experimental infections, biological replicates, were carried out by membrane blood feeding in the laboratory as described previously [47-50]. Briefly, females were fed through membranes on gametocyte-infected blood from malaria patients. To eliminate transmission-blocking factors, venous blood was collected and the serum was replaced by a non-immune AB serum to avoid transmission of human blocking factors. Dissection of mosquito midguts and salivary glands was performed in situ on adult females at 7 days post-blood

meal. To determine infection levels, mosquito guts were dissected 7 days after blood feeding and were stained with 2% Mercurochrome before microscopic examination. Tissues were maintained in ice-cold Schneider's insect culture medium (Sigma-Aldrich) and fresh tissues were immediately processed for chromatin and RNA analyses.

RNA isolation, library preparation and sequencing

We prepared RNA-seq libraries from RNA isolated from two biological replicates of midgut control and infected samples. Total RNA was extracted from fresh mosquito tissues (~25 midguts and ~ 50 salivary glands) using the mirVana™ RNA Isolation Kit (Ambion®) according to the manufacturer protocol and used for mRNA library preparation. RNA concentration was quantified using a Qubit® 2.0 Fluorometer, and RNA integrity was determined with an Agilent 2100 Bioanalyzer. Illumina libraries were prepared and sequenced at the HudsonAlpha Institute for Biotechnology, using an Illumina HiSeq2000 sequencer, standard directional RNA-seq library construction, 50 bp paired end reads with ribosomal reduction (RiboMinus™ Eukaryote Kit, Ambion®).

RNA-seq analysis

We mapped RNA paired directional reads to *A. gambiae* PEST strain genome version 4.3 publicly available at Vector Base (https://www.vectorbase.org/) using TopHat v2.0.13 [51]. We aligned reads using the option of library type set as first-strand for directional RNA-seq. We used SAMtools v.1.6 (http://samtools.sourceforge.net) for SAM and BAM file manipulation and conversion. We performed quality control analysis using QualiMap v2.2.1 [52]. Statistics of the RNA-seq analysis for each condition and replicate are shown in Table S1.

Quantification and differential gene expression analysis was conducted using HTSeq/DESeq2 packages [53, 54]. To count reads, HTSeq configuration parameters were set for a strand-specific assay to separate between sense and antisense transcripts. We used the matrix of raw

read counts as input for DESeq2 R package, which performs library normalization and uses negative binomial generalized linear models to identify differentially expressed genes. The design included condition as main factor and infection as covariable to control for inter-experiment variability. In this analysis, genes were considered significantly differentially expressed if the P-value was below 0.05.

Sets of differentially expressed genes between conditions, infected and control, were annotated based on the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) using DAVID [55, 56].

Chromatin Immunoprecipitation

Chromatin immunoprecipitation was performed in mosquito as previously described by us [26]. Antibodies to histone modifications used in this study were anti-H3K9ac (Millipore #07-352), anti-H3K4me3 (Abcam ab8580), anti-H3K27ac (Abcam ab4729), and anti-H3K9me3 (Abcam ab8898). ChIP-seq libraries were prepared following the procedure in Bowman et al. [57], and using the HiFi Kapa Sybr library preparation kit (KapaBiosystems). To obtain the quantity needed to perform ChIP-seq, the samples were pooled, resulting in one biological replicate for infected mosquito midguts and one biological replicate of uninfected tissues. ChIP-seq libraries were sequenced at the HudsonAlpha Institute for Biotechnology using an Illumina HiSeq2000 sequencer.

ChIP-seq data analysis

We performed quality control of Illumina reads using QualiMap v2.2.1 [52] (Table S1). Correlation analysis was performed deepTools2 (v2.5.0) [58]. We reported Spearman *rho* correlation coefficients between each pair of histone modification datasets.

We mapped reads for various histone modifications and the input to *A. gambiae* PEST strain genome version 4.3 (https://www.vectorbase.org/) using Bowtie v2.2.9 [59] with default parameters except for --no-mixed. Reads were trimmed 5 bases from each read 5’ end (--trim5 5). Mapped

reads were then sorted and deduplicated using SAMtools v.1.6. We applied a quality threshold of 10 in MAPQ score. All libraries were downsampled (SAMtools) to the same number of reads (9M) for further downstream analysis. To calculate the enrichment, we used BEDtools software suite (v.2.25.0) [60] to get the number of reads overlapping with regions of interest. Resulting read counts were normalized (RPKM), input corrected and log2 transformed using R. We conducted peak calling using MACS2 (v.2.1.1) [61] "callpeak" module with -t and -c options and default parameters, except for -g 2.73e8 --keep-dup all -B --SPMR -q 0.01 –nomodel. We further assessed statistical significance of MACS2 peaks using the RECAP software [62] to recalibrate peak calling P-values. Over 99% of MACS2 peaks remained significant according to a recalibrated P-value threshold of 0.05 (Table S2). For visualization in IGV [63, 64], tracks of input corrected ChIP-seq signal were computed using MACS2 "bdgcmp" module (-m ppois) on each pair of fragment pileup and control lambda bedGraph files from peak calling analysis (https://github.com/tao-liu/MACS/wiki/Build-Signal-Track).

To quantitative compare histone modification profiles between infected and control mosquito tissues, we used diffReps software [65]. This method uses a sliding window approach to identify regions that show significant changes in ChIP-seq signal, without constraining regions to compare by peak calling. ChIP-seq data for infected and control and the corresponding inputs were provided with -tr, -co, --btr and --bco options. Due to the lack of biological replicates, the statistical test used was G-test (-me gt). Rest of parameters were set as default except for --window 1000, as recommended for the scanning of histone mark peaks. We performed annotation of MACS2 peaks and diffReps regions to genomic features (TSSs, exons, introns, and intergenic regions) using the annotatePeaks.pl module in HOMER (v.3.12) [66]. Based on the density distribution of the distances from upstream MACS2 peaks and diffReps regions to the nearest ATG site, we considered 2 Kb upstream from the translation start site ATG as the putative promoter region.

We used the chromatin-state segmentation software ChromHMM [67] to compute genome-wide chromatin state predictions in each condition based on relative enrichment levels of histone

modification marks. For the binarization of the genome we used default parameters except for -b 200. We chose a 4 states model assuming chromatin states with high levels of enrichment of each histone modification. The chromatin states we found were: depleted (low levels of enrichment for all histone modifications, high levels of enrichment for H3K9me3, high levels of enrichment for H3K9me3 and H3K4me3 (bivalent state) and high levels of enrichment for H3K9ac, H3K27ac and H3K4me3. According to the ChromHMM segmentation, most of the genome is on a depleted state. We assigned predicted chromatin states to different features, such as MACS2 peaks and diffReps regions using the intersect tool from BEDtools and we required a minimum overlap of the region of 51% (-f 0.51).

To obtain a high confidence set of diffReps regions we applied a filtering based on multiple thresholds. We filtered out those diffReps regions located in depleted chromatin states (ChromHMM segmentation) at each corresponding condition and displaying FDR > 0.05. We also divided regions in three quantile groups (cut2 function in the Hmisc R package) according to their mean values in log2 Fold Change and in average normalized counts and fold enrichment vs input at each corresponding condition. Regions that fall in the lower quantiles were discarded (Table S3).

We performed GO terms overrepresentation tests for set of genes of interest annotated to diffReps regions using PANTHER Overrepresentation Test [68, 69]. We chose Fisher's Exact with FDR multiple tests correction and applied a threshold of FDR < 0.05. Sets of differentially expressed genes between conditions, infected and control, were annotated based on the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) using DAVID. We used ChromHMM segmentation and plotEnrichment function in chromstaR R package [70] to assess enrichment of predicted chromatin states in certain features of interest, e.g subset of immune factors encoding genes. Average profile plots and heatmaps representing histone modifications enrichments (RPKM normalized and input-corrected) centered on gene coordinates were built using ngs.plot (v.2.61) [71].

Integration of RNA and ChIP-seq data

To connect patterns of histone modifications with the regulation of gene expression we ordered genes annotated to MACS2 peaks by mRNA levels and showed histone modifications enrichment levels at those gene bodies and promoter regions. Correlation between histone modifications enrichment levels and mRNA levels was assessed using Spearman's rank correlation test (cor.test R function). Additionally, in order to connect differential enrichment of histone modifications with regulation of gene expression, we filtered those genes containing high-confidence diffReps regions in promoters and gene bodies and performed a soft clustering approach using the Mfuzz R package [72] over the ratio of histone modifications between conditions (ratio of enrichment at infected to non-infected samples). Using a standard m fuzzy c-means parameter of 1.7, a total of 30 clusters were created. Clusters with certain patterns of histone modifications were isolated creating unique groups (Figure S4). Only elements with membership value higher than 51% within each particular cluster were considered. Next, we used the clustering order based on the ratio of histone modifications enrichment to show mRNA levels of the corresponding genes (ratio of mRNA levels at infected to non-infected samples). To check the validity of our results and to further assess the functional output and transcriptional shift associated to different chromatin states: we focused on patterns showing maximum enrichment of certain histone modification at both infected and non-infected, we categorized genes into high, medium or low expressed groups at each condition by dividing the mRNA levels in three quantile groups according to their means (cut2 function) and we filtered out low-expressed genes. Based on the soft-clustering of each region and each gene mRNA level we then isolated those cases where differential histone enrichment profiles and functional output coincide with the expected patterns. We also performed a soft clustering analysis with the same parameters as above computing histone modifications enrichment ratios at promoters and gene bodies of significant differentially expressed genes according to DESeq2 (P-value < 0.05), with similar results (Figure S7-8, Table S6).

Heatmaps showing histone modifications enrichment and mRNA levels were built using iheatmapr R package [73]. Bar and violin plots were produced using ggplot2 R package [74]. For comparative and visualization purposes, histone enrichment signals and mRNA levels were log2 transformed. When computing histone modifications enrichments and ratios, a pseudocount (0.1) was added to get finite values when input correcting (dividing by 0 in ratios) or converting the signal to log2 scale. When categorizing values in quantile groups according to their means, High, Medium and Low groups, we used cut2 function in Hmisc R package.

Motif analysis

On the set of MACS2 peaks that intersect with high confidence diffReps regions and annotate to genes (promoters and gene bodies) we conducted *de novo* motif analysis using HOMER software (v.3.12) [66]. For this analysis, we considered the center of the ChIP-seq peak region and slopped 100 bp in each direction. We limited the number of background sequences to double the number of ChIP-seq target sequences for each histone modification mark. Only motifs enriched in more than 5% of the targets sequences and below a threshold P-value of 10E-15 were considered, and results corresponding to low complexity motifs and offsets or degenerate versions of highly enriched motifs were avoided. The purpose of this analysis was to identify enrichment in particular sequence footprints associated with changes in histone modification occupancy that are associated to a *P. falciparum* infection.

REFERENCES

1. WHO, *World malaria report*. 2017: Geneva.
2. Vlachou, D., et al., *Functional genomic analysis of midgut epithelial responses in Anopheles during Plasmodium invasion.* Curr Biol, 2005. **15**(13): p. 1185-95.
3. Blandin, S. and E.A. Levashina, *Mosquito immune responses against malaria parasites.* Curr Opin Immunol, 2004. **16**(1): p. 16-20.

4. Akman-Anderson, L., M. Olivier, and S. Luckhart, *Induction of nitric oxide synthase and activation of signaling proteins in Anopheles mosquitoes by the malaria pigment, hemozoin.* Infect Immun, 2007. **75**(8): p. 4012-9.

5. Blandin, S.A., E. Marois, and E.A. Levashina, *Antimalarial responses in Anopheles gambiae: from a complement-like protein to a complement-like pathway.* Cell Host Microbe, 2008. **3**(6): p. 364-74.

6. Bartholomay, L.C. and K. Michel, *Mosquito immunobiology: the intersection of vector health and vector competence.* Annual review of entomology, 2018. **63**: p. 145-167.

7. Pollitt, L.C., et al., *Existing infection facilitates establishment and density of malaria parasites in their mosquito vector.* PLoS pathogens, 2015. **11**(7): p. e1005003.

8. Simões, M.L. and G. Dimopoulos, *A mosquito mediator of parasite-induced immune priming.* Trends in parasitology, 2015. **31**(9): p. 402-404.

9. Lambrechts, L., et al., *Environmental influence on the genetic basis of mosquito resistance to malaria parasites.* Proc Biol Sci, 2006. **273**(1593): p. 1501-6.

10. Lefevre, T., et al., *Non-genetic determinants of mosquito competence for malaria parasites.* PLoS Pathog, 2013. **9**(6): p. e1003365.

11. Feil, R. and M.F. Fraga, *Epigenetics and the environment: emerging patterns and implications.* Nat Rev Genet, 2012. **13**(2): p. 97-109.

12. Gut, P. and E. Verdin, *The nexus of chromatin regulation and intermediary metabolism.* Nature, 2013. **502**(7472): p. 489-98.

13. Kouzarides, T., *Chromatin modifications and their function.* Cell, 2007. **128**(4): p. 693-705.

14. Yin, H., et al., *A high-resolution whole-genome map of key chromatin modifications in the adult Drosophila melanogaster.* PLoS Genet, 2011. **7**(12): p. e1002380.

15. Merkling, S.H., et al., *The epigenetic regulator G9a mediates tolerance to RNA virus infection in Drosophila.* PLoS Pathog, 2015. **11**(4): p. e1004692.

16. Anreiter, I., J.M. Kramer, and M.B. Sokolowski, *Epigenetic mechanisms modulate differences in Drosophila foraging behavior.* Proc Natl Acad Sci U S A, 2017. **114**(47): p. 12518-12523.

17. Dobson, A.J., et al., *Nutritional Programming of Lifespan by FOXO Inhibition on Sugar-Rich Diets.* Cell Rep, 2017. **18**(2): p. 299-306.

18. Bantignies, F., et al., *Inheritance of Polycomb-dependent chromosomal interactions in Drosophila.* Genes Dev, 2003. **17**(19): p. 2406-20.

19. Margueron, R. and D. Reinberg, *Chromatin structure and the inheritance of epigenetic information.* Nat Rev Genet, 2010. **11**(4): p. 285-96.

20. Campos, E.I., J.M. Stafford, and D. Reinberg, *Epigenetic inheritance: histone bookmarks across generations.* Trends Cell Biol, 2014. **24**(11): p. 664-74.

21. Stroud, H., et al., *Early-life gene expression in neurons modulates lasting epigenetic states.* Cell, 2017. **171**(5): p. 1151-1164. e16.

22. Coleman, R.T. and G. Struhl, *Causal role for inheritance of H3K27me3 in maintaining the OFF state of a Drosophila HOX gene.* Science, 2017. **356**(6333): p. eaai8236.

23. Osborne, A.J. and P.K. Dearden, *A 'phenotypic hangover': the predictive adaptive response and multigenerational effects of altered nutrition on the transcriptome of Drosophila melanogaster.* Environmental epigenetics, 2017. **3**(4): p. dvx019.

24. Ciabrelli, F., et al., *Stable Polycomb-dependent transgenerational inheritance of chromatin states in Drosophila.* Nature genetics, 2017. **49**(6): p. 876.

25. Gomez-Diaz, E., et al., *Epigenetics of host-pathogen interactions: the road ahead and the road behind.* PLoS Pathog, 2012. **8**(11): p. e1003007.

26. Gómez-Díaz, E., et al., *Insights into the epigenomic landscape of the human malaria vector Anopheles gambiae.* Frontiers in genetics, 2014. **5**: p. 277.

27. Pérez-Zamorano, B., et al., *Identification of cis-regulatory sequences reveals potential participation of lola and Deaf1 transcription factors in Anopheles gambiae innate immune response.* PloS one, 2017. **12**(10): p. e0186435.

28. Zakovic, S. and E.A. Levashina, *NF-κB-Like Signaling Pathway REL2 in Immune Defenses of the Malaria Vector Anopheles gambiae.* Frontiers in cellular and infection microbiology, 2017. **7**: p. 258.

29. Tripet, F., F. Aboagye-Antwi, and H. Hurd, *Ecological immunology of mosquito–malaria interactions.* Trends in parasitology, 2008. **24**(5): p. 219-227.

30. Kharchenko, P.V., et al., *Comprehensive analysis of the chromatin landscape in Drosophila melanogaster.* Nature, 2011. **471**(7339): p. 480.

31. Brucker, R.M., et al., *Insect Innate Immunity Database (IIID): an annotation tool for identifying immune genes in insect genomes.* PLoS One, 2012. **7**(9): p. e45125.

32. Waterhouse, R.M., et al., *Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes.* Science, 2007. **316**(5832): p. 1738-1743.

33. Rinker, D.C., R.J. Pitts, and L.J. Zwiebel, *Disease vectors in the era of next generation sequencing.* Genome Biol, 2016. **17**(1): p. 95.

34. Bernstein, B.E., et al., *A bivalent chromatin structure marks key developmental genes in embryonic stem cells.* Cell, 2006. **125**(2): p. 315-26.

35. Voigt, P., W.W. Tee, and D. Reinberg, *A double take on bivalent promoters.* Genes Dev, 2013. **27**(12): p. 1318-38.

36. Creyghton, M.P., et al., *Histone H3K27ac separates active from poised enhancers and predicts developmental state.* Proc Natl Acad Sci U S A, 2010. **107**(50): p. 21931-6.

37. Pradeepa, M.M., *Causal role of histone acetylations in enhancer function.* Transcription, 2017. **8**(1): p. 40-47.

38. Lemaitre, B. and J. Hoffmann, *The host defense of Drosophila melanogaster.* Annu Rev Immunol, 2007. **25**: p. 697-743.

39. Marois, E., *The multifaceted mosquito anti-Plasmodium response.* Curr Opin Microbiol, 2011. **14**(4): p. 429-35.

40. Habtewold, T., Z. Groom, and G.K. Christophides, *Immune resistance and tolerance strategies in malaria vector and non-vector mosquitoes.* Parasites & vectors, 2017. **10**(1): p. 186.

41. Baton, L.A. and L.C. Ranford-Cartwright, *Ookinete destruction within the mosquito midgut lumen explains Anopheles albimanus refractoriness to Plasmodium falciparum (3D7A) oocyst infection.* International journal for parasitology, 2012. **42**(3): p. 249-258.

42. Smith, R.C. and C. Barillas-Mury, *Plasmodium Oocysts: Overlooked Targets of Mosquito Immunity.* Trends Parasitol, 2016. **32**(12): p. 979-990.

43. Negre, N., et al., *A cis-regulatory map of the Drosophila genome.* Nature, 2011. **471**(7339): p. 527-31.

44. Behura, S.K., et al., *High-throughput cis-regulatory element discovery in the vector mosquito Aedes aegypti.* BMC Genomics, 2016. **17**: p. 341.

45. Govind, S., *Innate immunity in Drosophila: Pathogens and pathways.* Insect Sci, 2008. **15**(1): p. 29-43.

46. Reed, D.E., et al., *DEAF-1 regulates immunity gene expression in Drosophila.* Proc Natl Acad Sci U S A, 2008. **105**(24): p. 8351-6.

47. Alout, H., et al., *Insecticide exposure impacts vector–parasite interactions in insecticide-resistant malaria vectors.* Proc. R. Soc. B, 2014. **281**(1786): p. 20140389.

48. Vantaux, A., et al., *A heavy legacy: offspring of malaria-infected mosquitoes show reduced disease resistance.* Malaria journal, 2014. **13**(1): p. 442.

49. Da, D.F., et al., *Experimental study of the relationship between Plasmodium gametocyte density and infection success in mosquitoes; implications for the evaluation of malaria transmission-reducing interventions.* Experimental parasitology, 2015. **149**: p. 74-83.

50. Gendrin, M., et al., *Antibiotics in ingested human blood affect the mosquito microbiota and capacity to transmit malaria.* Nature communications, 2015. **6**: p. 5921.

51. Kim, D., et al., *TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions.* Genome Biol, 2013. **14**(4): p. R36.

52. Okonechnikov, K., A. Conesa, and F. Garcia-Alcalde, *Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data.* Bioinformatics, 2016. **32**(2): p. 292-4.

53. Anders, S., P.T. Pyl, and W. Huber, *HTSeq--a Python framework to work with high-throughput sequencing data.* Bioinformatics, 2015. **31**(2): p. 166-9.

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

55. Jiao, X., et al., *DAVID-WS: a stateful web service to facilitate gene/protein list analysis.* Bioinformatics, 2012. **28**(13): p. 1805-1806.

56. Huang, D.W., B.T. Sherman, and R.A. Lempicki, *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.* Nature protocols, 2008. **4**(1): p. 44.

57. Bowman, S.K., et al., *Multiplexed Illumina sequencing libraries from picogram quantities of DNA.* BMC genomics, 2013. **14**(1): p. 466.

58. Ramirez, F., et al., *deepTools2: a next generation web server for deep-sequencing data analysis.* Nucleic Acids Res, 2016. **44**(W1): p. W160-5.

59. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2.* Nat Methods, 2012. **9**(4): p. 357-9.

60. Quinlan, A.R. and I.M. Hall, *BEDTools: a flexible suite of utilities for comparing genomic features.* Bioinformatics, 2010. **26**(6): p. 841-2.

61. Zhang, Y., et al., *Model-based analysis of ChIP-Seq (MACS).* Genome Biol, 2008. **9**(9): p. R137.

62. Chitpin, J.G., A. Awdeh, and T.J. Perkins, *RECAP reveals the true statistical significance of ChIP-seq peak calls.* bioRxiv, 2018: p. 260687.

63. Robinson, J.T., et al., *Integrative genomics viewer.* Nat Biotechnol, 2011. **29**(1): p. 24-6.

64. Thorvaldsdottir, H., J.T. Robinson, and J.P. Mesirov, *Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration.* Brief Bioinform, 2013. **14**(2): p. 178-92.

65. Shen, L., et al., *diffReps: detecting differential chromatin modification sites from ChIP-seq data with biological replicates.* PloS one, 2013. **8**(6): p. e65598.

66. Heinz, S., et al., *Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities.* Mol Cell, 2010. **38**(4): p. 576-89.

67. Ernst, J. and M. Kellis, *ChromHMM: automating chromatin-state discovery and characterization.* Nat Methods, 2012. **9**(3): p. 215-6.

68. Thomas, P.D., et al., *PANTHER: a library of protein families and subfamilies indexed by function.* Genome Res, 2003. **13**(9): p. 2129-41.

69. Mi, H., et al., *PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium.* Nucleic Acids Res, 2010. **38**(Database issue): p. D204-10.

70. Taudt, A., et al., *chromstaR: Tracking combinatorial chromatin state dynamics in space and time.* bioRxiv, 2016: p. 038612.

71. Shen, L., et al., *ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases.* BMC Genomics, 2014. **15**: p. 284.

72. Kumar, L. and E.F. M, *Mfuzz: a software package for soft clustering of microarray data.* Bioinformation, 2007. **2**(1): p. 5-7.

73. Schep, A.N. and S.K. Kummerfeld, *iheatmapr: interactive complex heatmaps in R.* J Open Source Software, 2017. **2**(16): p. 359.

74. Wickham, H., *ggplot2: elegant graphics for data analysis*. 2016: Springer.

TABLE 1. List of consensus motifs corresponding to known transcription factors significantly enriched in the set of differential histone modified regions that are associated to a *P. falciparum* infection

Histone	Motif	P-value	% Targets / % Background	TF BS predicted (score)		
H3K9ac	CGTTCCCYWTTT	1E-23	18.40 / 1.06	**kni**/dmmpmm/fly (0.64)	http://flybase.org/reports/FBgn0001320	Development, gene expression
	AACKATTT	1E-23	18.40 / 1.02	**br-Z2**/dmmpmm/fly (0.727)	http://flybase.org/reports/FBgn0283451	Development, gene expression, reproduction, response to stimulus
	GTGCGTRA	1E-23	18.40 / 1.20	**z**/dmmpmm/fly (0.71)	http://flybase.org/reports/FBgn0000629.htm	Gene expression, nucleus
	GCGATAGA	1E-21	20.80 / 1.68	**Dref**/dmmpmm/fly (0.75)	http://flybase.org/reports/FBgn0015664.htm	Development, gene expression, metabolism
	MCCGAGCN	1E-20	16.00 / 0.64	**Deaf1**/dmmpmm/fly (0.70)	http://beta.flybase.org/reports/FBgn0013799.htm	Inmune system, development, gene expressio
	TAATCCRY	1E-19	16.00 / 0.93	**Gsc**/dmmpmm/fly (0.88)	http://flybase.org/reports/FBgn0010323	Gene expression, other molecular function
	CATCCTGGCG	1E-17	24.00 / 3.41	**brk**/dmmpmm/fly (0.679)	http://flybase.org/reports/FBgn0024250	Development, gene expression, signalling, response to stimulus
	TCCTTCGA	1E-15	17.60 / 2.23	**ttk**/dmmpmm/fly (0.71)	http://flybase.org/reports/FBgn0003870	Development, reproduction, response to stimulus, cell organisation-biogenesis
	CTTGTTCTTC	1E-15	13.60 / 0.58	**pan**/dmmpmm/fly (0.63)	http://flybase.org/reports/FBgn0085432	Signalling, gene expression, Inmune system
	TTCGTAATAC	1E-15	13.60 / 1.46	**gt**/dmmpmm/fly (0.66)	http://flybase.org/reports/FBgn0001150	Development, gene expression, transcription factor
	GCTTGCTY	1E-14	16.80 / 2.34	**grh**/dmmpmm/fly (0.72)	http://flybase.org/reports/FBgn0259211	Development, biogenesis
	AGCTTTAA	1E-12	15.70 / 2.29	**zen**/dmmpmm/fly (0.70)	http://flybase.org/reports/FBgn0004053	Development, gene expression, transcription factor
H3K27ac	GGCTCGTC	1E-29	11.74 / 0.43	**h**/dmmpmm/fly(0.676)	http://flybase.org/reports/FBgn0001168.htm	Development, gene expression, DNA bindin
	ATTTTTCCCC	1E-21	9.39 / 0.65	**dl**/dmmpmm/fly (0.69)	http://flybase.org/reports/FBgn0260632.htm	Inmune system, development, gene expressio
	AGAACAGTAA	1E-21	9.39 / 0.40	**ara**/dmmpmm/fly (0.67)	http://flybase.org/reports/FBgn0015904	Development, gene expression
	GTGCAGCTCG	1E-21	9.39 / 0.65	**sna**/dmmpmm/fly (0.65)	http://flybase.org/reports/FBgn0003448	Development, gene expression, DNA bindin
	CAAAAACGCAAC	1E-21	9.39 / 0.67	**hb**/dmmpmm/fly (0.744)	http://flybase.org/reports/FBgn0001180	Development, gene expression, transcription factor
	CTATGTTT	1E-16	9.86 / 0.84	**pan**/dmmpmm/fly (0.80)	http://flybase.org/reports/FBgn0085432	Signalling, gene expression, Inmune system
	TCGATCGTCG	1E-15	7.51 / 0.00	**Dref**/dmmpmm/fly (0.60)	http://flybase.org/reports/FBgn0015664.html	Development, gene expression, DNA binding, transcription factor, cell cycle, DNA metabolism
	ATWTCTGC	1E-13	8.45 / 1.00	**grh**/dmmpmm/fly (0.714)	http://flybase.org/reports/FBgn0259211	Development, biogenesis
	CTTCADTGCGGA	1E-13	6.57 / 0.66	**slbo**/dmmpmm/fly (0.661)	http://flybase.org/reports/FBgn0005638	Development, gene expression, reproduction
	CACGAAGT	1E-12	12.68 / 2.32	**Eip74EF**/dmmpmm/fly (0.772)	http://flybase.org/reports/FBgn0000567	Development, reproduction
H3K4me3	ARTTTTGTGT	1E-30	5.16 / 0.27	**br-Z3**/dmmpmm/fly (0.70)	http://flybase.org/reports/FBgn0283451	Development, gene expression, reproduction, response to stimuli
	TTTGATTCGTAA	1E-30	5.16 / 0.28	**ara**/dmmpmm/fly (0.67)	http://flybase.org/reports/FBgn0015904	Development, gene expression
	CTTCTTGCCCGA	1E-23	5.34 / 0.49	**Eip74EF**/dmmpmm/fly (0.62)	http://flybase.org/reports/FBgn0000567	Development, gene expression, reproduction
	CGAARAAGAR	1E-18	5.34 / 0.56	**pan**/dmmpmm/fly (0.70)	http://flybase.org/reports/FBgn0085432	Development, gene expression, response to stimulus, signalling, immune system
	TCCTCGTCGTTG	1E-16	7.55 / 1.60	**Aef1**/dmmpmm/fly (0.56)	http://flybase.org/reports/FBgn0005694.htm	Other molecular function
H3K9me3	AGCGCCTGGT	1E-26	8.82 / 0.61	**brk**/dmmpmm/fly (0.688)	http://flybase.org/reports/FBgn0024250	Development, gene expression, response to stimulus, signalling
	YCTGTGACCG	1E-24	8.46 / 0.59	**Hth**/dmmpmm/fly (0.72)	http://flybase.org/reports/FBgn0001235	Development, gene expression, DNA binding, transport/localisation, transcription facto
	TCYKGWAKCKGA	1E-24	8.46 / 0.58	**Pnt**/dmmpmm/fly (0.60)	http://flybase.org/reports/FBgn0003118	Development, gene expression, response to stimulus, signalling, immune system, cell cycle/proliferation, DNA binding,
	ACTCCAGATA	1E-21	7.72 / 0.57	**z**/dmmpmm/fly (0.61)	http://flybase.org/reports/FBgn0000629.html	Development, gene expression, protein metabolism, cell organisation/biogenesis, enzym
	CTAYTTAT	1E-21	7.72 / 0.18	**bin**/dmmpmm/fly (0.829)	http://flybase.org/reports/FBgn0045759	Development, gene expression, transcription factor
	ATCTCGGG	1E-21	7.72 / 0.53	**Deaf1**/dmmpmm/fly (0.68)	http://beta.flybase.org/reports/FBgn0013799.html	Development, gene expression, immune system, response to stimulus, transcription facto
	CATTCGAC	1E-21	9.56 / 0.81	**sd**/dmmpmm/fly (0.70)	http://flybase.org/reports/FBgn0003345	Development, gene expression, other molecular function, DNA binding, cell cycle/proliferatio
	RCGAAATTTTTG	1E-20	7.35 / 0.22	**cad**/dmmpmm/fly (0.76)	http://flybase.org/reports/FBgn0000251	Development, gene expression, response to stimulus, transcription factor, DNA binding, immune system
	ACTTTTCACTTT	1E-18	8.46 / 0.91	**tll**/dmmpmm/fly (0.77)	http://flybase.org/reports/FBgn0003720	Development, gene expression, cell cycle/proliferation, transcription factor, DNA binding
	TTAGACGA	1E-20	7.35 / 0.40	**exd**/dmmpmm/fly (0.67)	http://flybase.org/reports/FBgn0000611	Development, gene expression, other molecular function, DNA binding, transcription facto

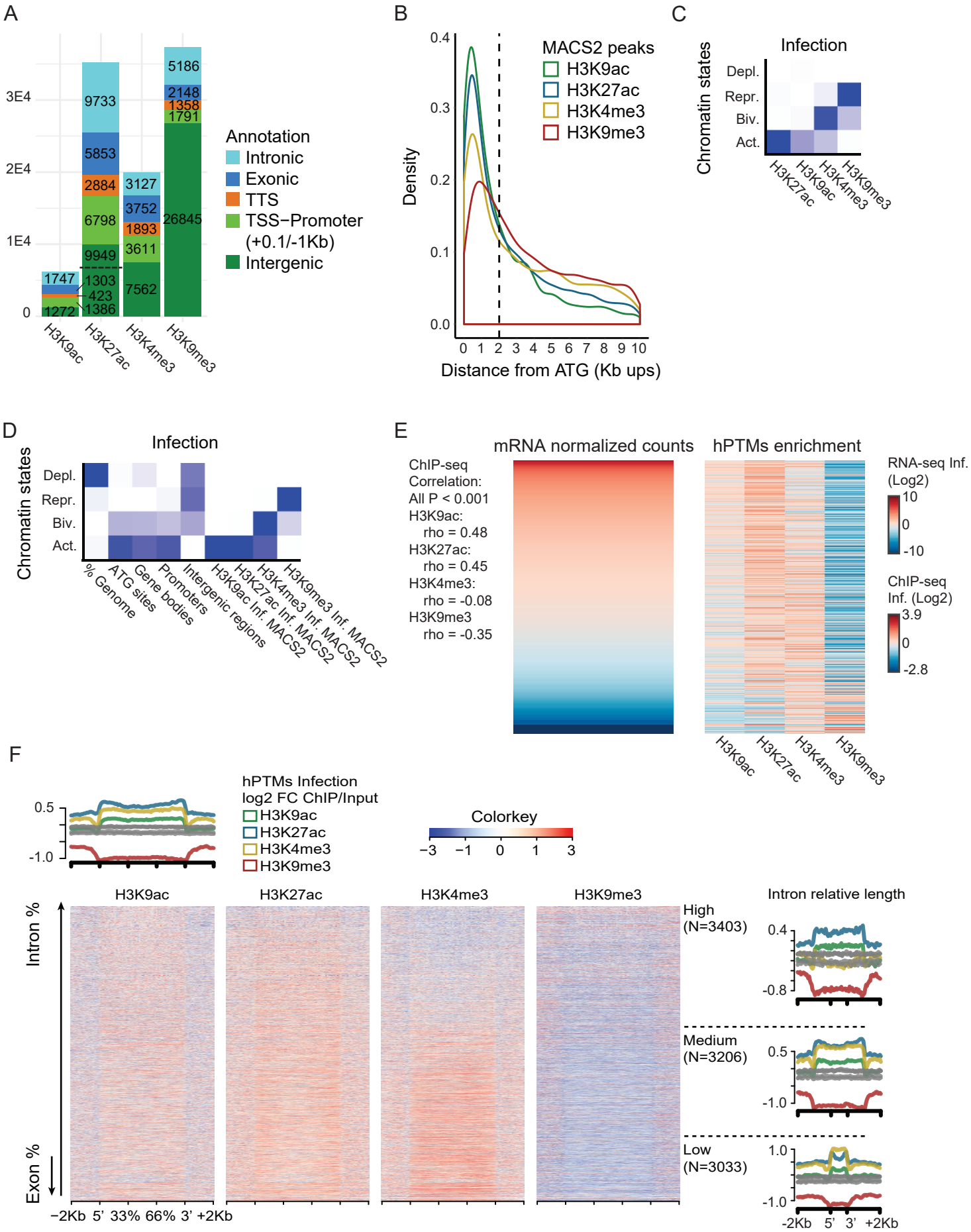

A
3E4
2E4
1E4
0
9733
5853
2884
6798
9949
1747
1303
423
1272
1386
3127
3752
1893
3611
7562
5186
2148
1358
1791
26845
Annotation
Intronic
Exonic
TTS
TSS−Promoter (+0.1/-1Kb)
Intergenic
H3K9ac
H3K27ac
H3K4me3
H3K9me3
B
Density
0.4
0.3
0.2
0.1
0.0
0 1 2 3 4 5 6 7 8 9 10
Distance from ATG (Kb ups)
MACS2 peaks
H3K9ac
H3K27ac
H3K4me3
H3K9me3
C
Infection
Chromatin states
Depl.
Repr.
Biv.
Act.
H3K27ac
H3K9ac
H3K4me3
H3K9me3
D
Infection
Chromatin states
Depl.
Repr.
Biv.
Act.
% Genome
ATG sites
Gene bodies
Promoters
Intergenic regions
H3K9ac Inf. MACS2
H3K27ac Inf. MACS2
H3K4me3 Inf. MACS2
H3K9me3 Inf. MACS2
E
mRNA normalized counts
hPTMs enrichment
ChIP-seq
Correlation:
All P < 0.001
H3K9ac:
rho = 0.48
H3K27ac:
rho = 0.45
H3K4me3:
rho = -0.08
H3K9me3
rho = -0.35
RNA-seq Inf. (Log2)
10
0
-10
ChIP-seq Inf. (Log2)
3.9
0
-2.8
H3K9ac
H3K27ac
H3K4me3
H3K9me3
F
0.5
-1.0
hPTMs Infection
log2 FC ChIP/Input
H3K9ac
H3K27ac
H3K4me3
H3K9me3
Colorkey
−3 −1 0 1 3
H3K9ac
H3K27ac
H3K4me3
H3K9me3
Intron %
Exon %
−2Kb 5' 33% 66% 3' +2Kb
Intron relative length
High (N=3403)
0.4
-0.8
Medium (N=3206)
0.5
-1.0
Low (N=3033)
1.0
-1.0
-2Kb 5' 3' +2Kb

Figure 1

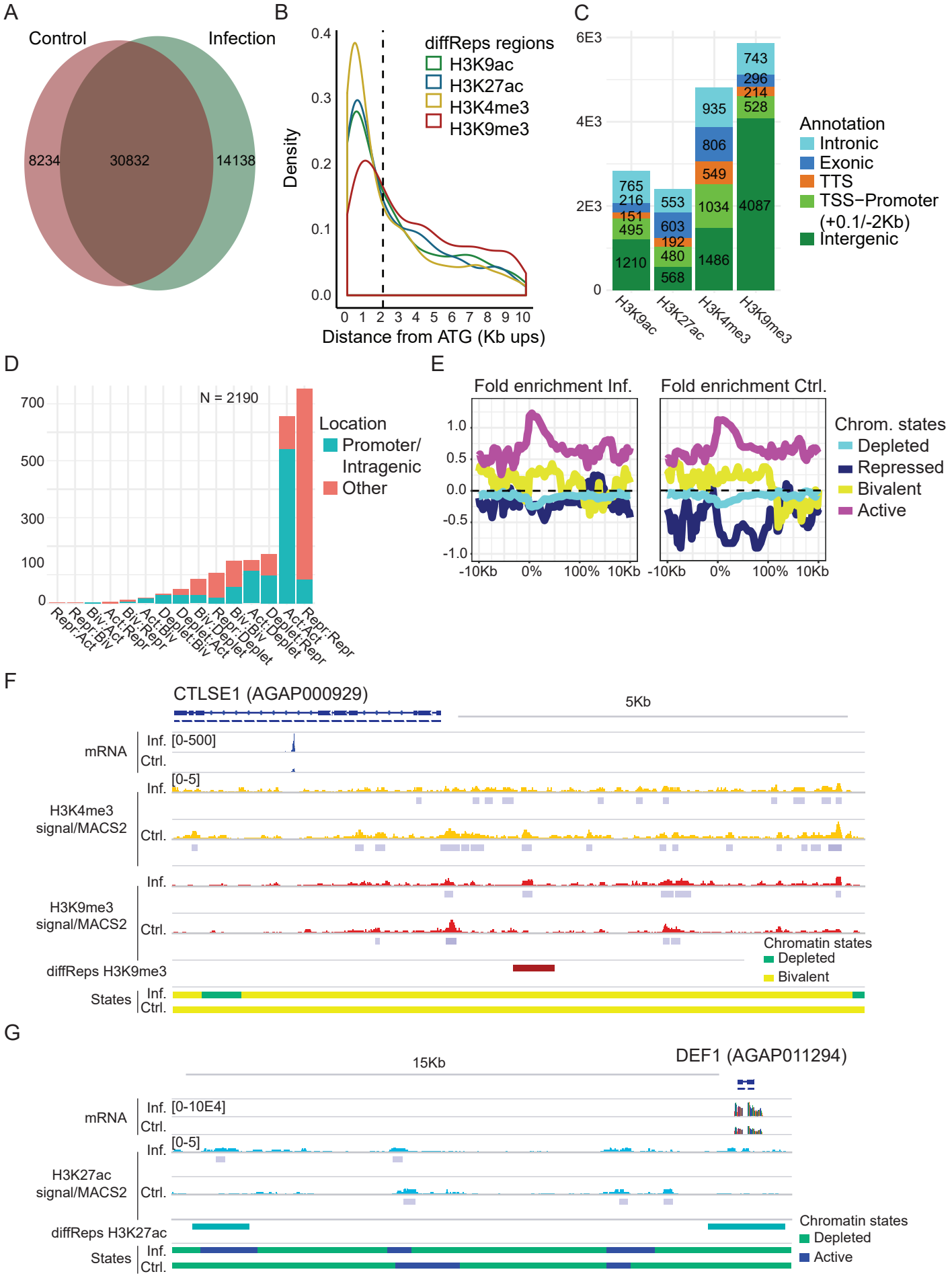

A
Control
Infection
8234
30832
14138
B
diffReps regions
H3K9ac
H3K27ac
H3K4me3
H3K9me3
Density
Distance from ATG (Kb ups)
C
Annotation
Intronic
Exonic
TTS
TSS−Promoter (+0.1/-2Kb)
Intergenic
H3K9ac
H3K27ac
H3K4me3
H3K9me3
D
N = 2190
Location
Promoter/ Intragenic
Other
E
Fold enrichment Inf.
Fold enrichment Ctrl.
Chrom. states
Depleted
Repressed
Bivalent
Active
F
CTLSE1 (AGAP000929)
5Kb
mRNA
Inf.
Ctrl.
[0-500]
[0-5]
H3K4me3 signal/MACS2
H3K9me3 signal/MACS2
diffReps H3K9me3
States
Chromatin states
Depleted
Bivalent
G
DEF1 (AGAP011294)
15Kb
[0-10E4]
H3K27ac signal/MACS2
diffReps H3K27ac
Chromatin states
Depleted
Active

Figure 2

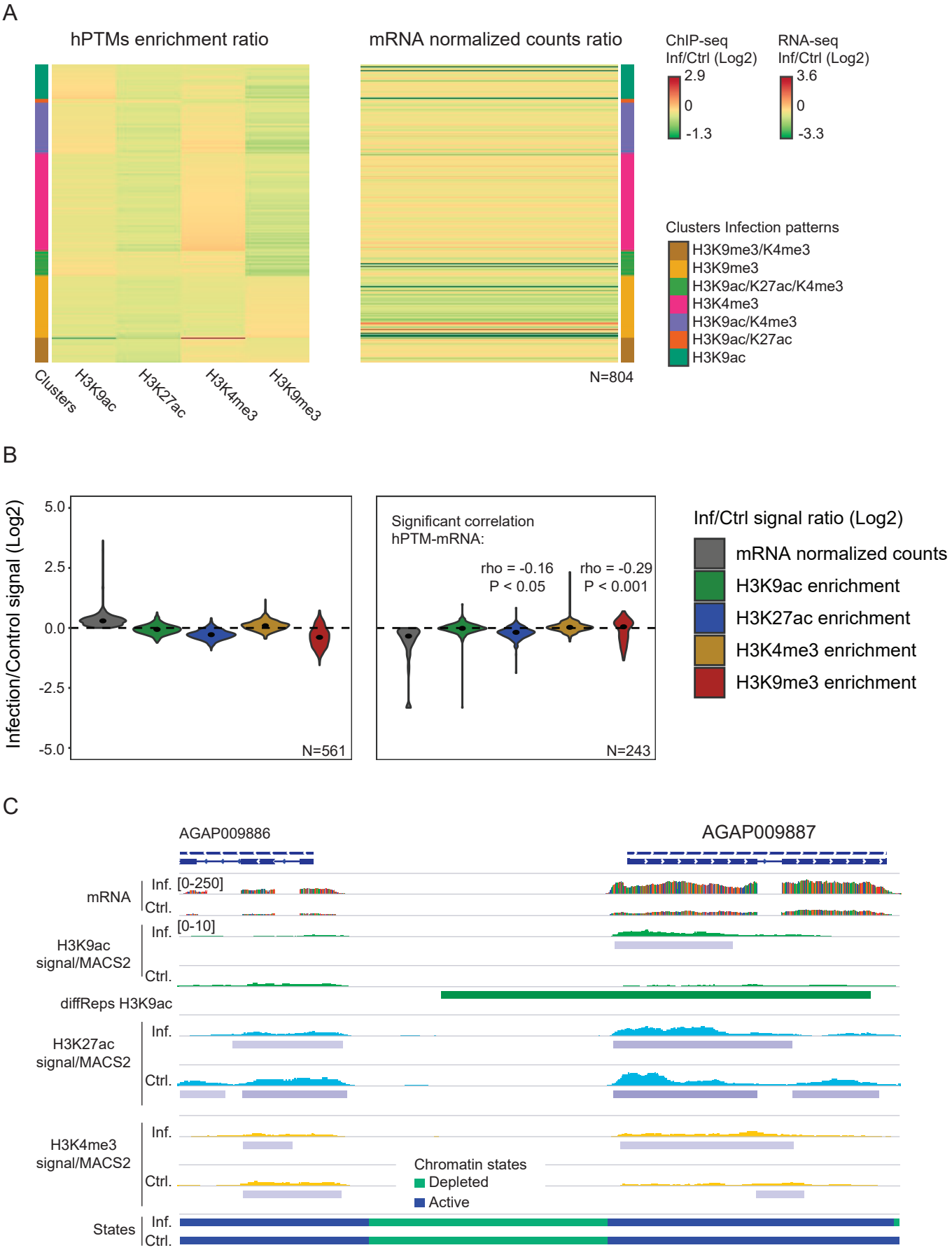
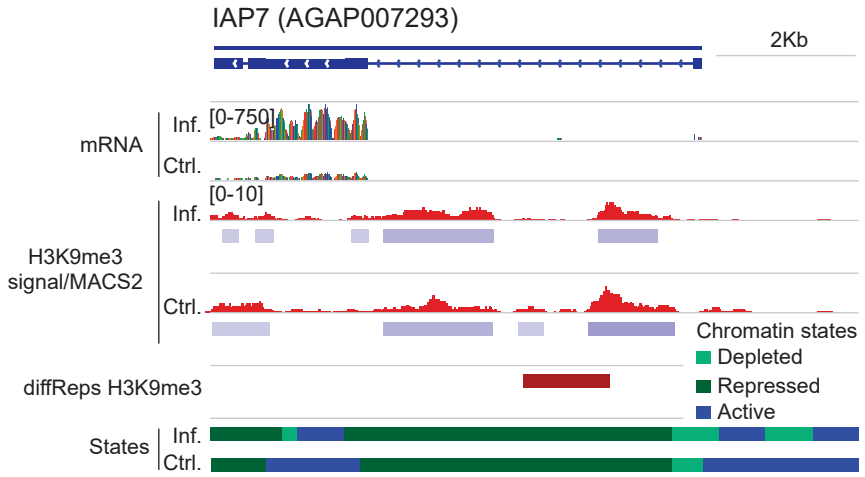

A
hPTMs enrichment ratio
mRNA normalized counts ratio
ChIP-seq Inf/Ctrl (Log2)
2.9
0
-1.3
RNA-seq Inf/Ctrl (Log2)
3.6
0
-3.3
Clusters Infection patterns
H3K9me3/K4me3
H3K9me3
H3K9ac/K27ac/K4me3
H3K4me3
H3K9ac/K4me3
H3K9ac/K27ac
H3K9ac
Clusters
H3K9ac
H3K27ac
H3K4me3
H3K9me3
N=804
B
Infection/Control signal (Log2)
5.0
2.5
0.0
-2.5
-5.0
N=561
Significant correlation hPTM-mRNA:
rho = -0.16
P < 0.05
rho = -0.29
P < 0.001
N=243
Inf/Ctrl signal ratio (Log2)
mRNA normalized counts
H3K9ac enrichment
H3K27ac enrichment
H3K4me3 enrichment
H3K9me3 enrichment
C
AGAP009886
AGAP009887
mRNA
Inf.
Ctrl.
[0-250]
H3K9ac signal/MACS2
[0-10]
diffReps H3K9ac
H3K27ac signal/MACS2
H3K4me3 signal/MACS2
Chromatin states
Depleted
Active
States

Figure 3

A


IAP7 (AGAP007293)
2Kb
mRNA
Inf.
[0-750]
Ctrl.
H3K9me3
signal/MACS2
Inf.
[0-10]
Ctrl.
Chromatin states
Depleted
Repressed
Active
diffReps H3K9me3
States
Inf.
Ctrl.

B

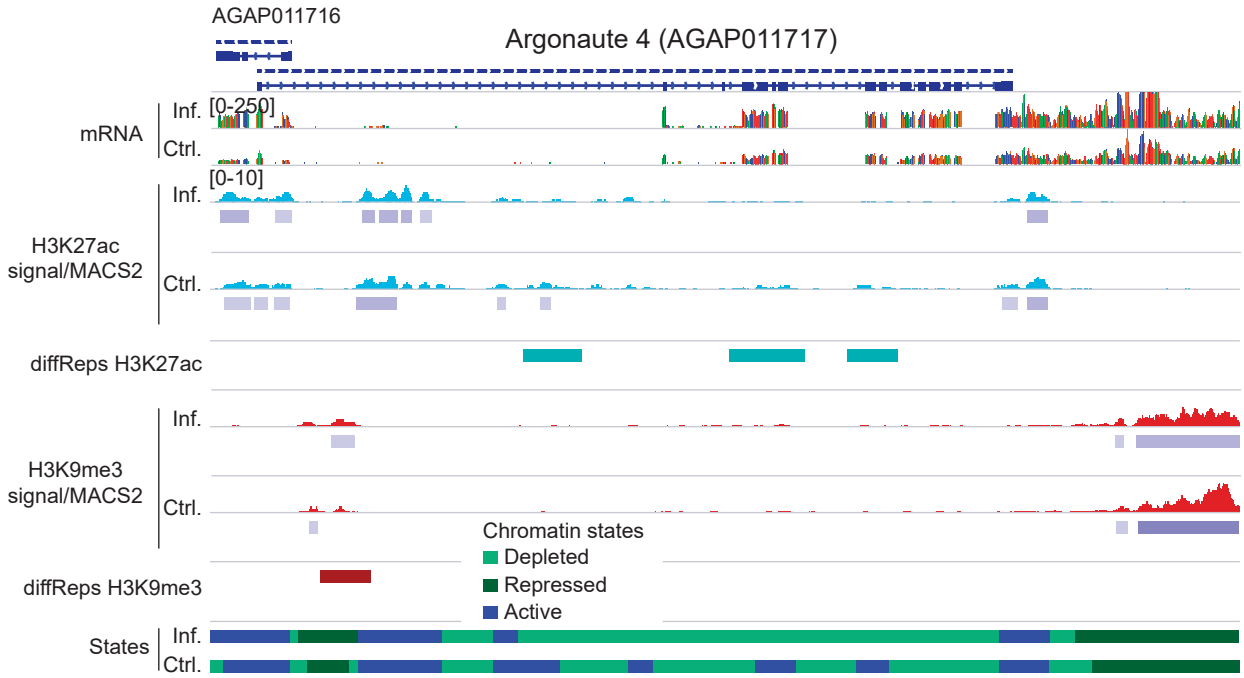

AGAP011716
Argonaute 4 (AGAP011717)
mRNA
Inf.
[0-250]
Ctrl.
H3K27ac
signal/MACS2
Inf.
[0-10]
Ctrl.
diffReps H3K27ac
H3K9me3
signal/MACS2
Inf.
Ctrl.
Chromatin states
Depleted
Repressed
Active
diffReps H3K9me3
States
Inf.
Ctrl.

Figure 4

FIGURE LEGENDS

FIGURE 1. Histone modifications profiles and gene expression regulation in *P. falciparum* infected mosquitoes

A. Annotation of MACS2 ChIP-seq peaks for each histone modification to genomic features: TSS-Promoters, TTSs, Intergenic, Intron and Exon regions.

B. Density plot showing the position of MACS2 peaks for each histone modification with respect to the ATG protein initiation codon.

C. Heatmap of emission parameters from ChromHMM analysis using a four chromatin-states model based on histone modifications enrichment patterns in the infected condition. The predicted states are: Deplet. (depleted, low levels of all hPTMs), Repr. (repressive, H3K9me3 enrichment), Biv (bivalent, H3K4me3/H3K9me3 enrichment) and Act. (active, H3K27ac/H3K9ac/H3K4me3 enrichment). Darker blue indicates higher enrichment of a particular histone modification.

D. Heatmap showing the overlap of various genomic features, including MACS2 peaks located in promoters (2-2Kb from the ATG) or gene bodies in the infection condition, with the predicted chromatin states. Darker blue in the first column indicates higher percentage of the genome overlapped by a given state. For other columns it indicates the likelihood of finding a particular chromatin state in each genomic feature compared to what it would be expected by chance.

E. Heatmaps showing mRNA levels (left) and histone modifications enrichment profiles (right) of genes displaying a MACS2 peak in the promoter or the gene body. Data corresponds to the infected condition. Genes are ordered by mRNA levels. ChIP-seq enrichment at the promoters and the gene bodies is normalized (RPKM) and input-corrected. Data is log2-scaled and mean-centered. Spearman rank correlation coefficient

(rho) and corresponding P-value are shown for the association between for each histone modification enrichment levels and mRNA levels.

F. Heatmaps showing histone modifications enrichment profiles at high and medium expressed genes in the infection condition. Genes are ordered by the percentage of the bodies containing introns and exons. Average profile plots show density of normalized (RPKM) and input corrected ChIP-seq reads for each histone modification at high and medium expressed genes (top) and at those genes classified by the percentage of gene bodies containing introns (right).

FIGURE 2. Changes in histone modifications enrichment in response to infection

A. Venn diagram showing the overlap between MACS2 peaks identified in infected and control conditions.

B. Density plot showing the position of differential diffReps regions for each histone modification with respect to the ATG initiation codon.

C. Annotation of diffReps regions for each histone modification to genomic features: TSS-Promoters, TTSs, Intergenic, Intron and Exon regions. The diffReps regions located -2Kb/+0.1Kb from the ATG are annotated to TSS-Promoter regions.

D. Barplot showing the number and location of high confidence diffReps regions and the chromatin states transitions between conditions associated to the peak region.

E. Profile plots showing predicted chromatin states in infection (left) and control (right) conditions at genes encoding for immune-related factors [1]. The graphs represent chromatin states fold enrichment (log(observed/expected)) with respect to the scaled gene bodies +- 10Kb.

F-G. Histone enrichment profiles in the regions containing the CTLSE1 (AGAP000929) and DEF1 (AGAP011294) encoding genes. Tracks show normalized/input-corrected

ChIP-seq signals and RNA-seq mapped reads counts for each condition. The location of diffReps regions, MACS2 peaks and predicted chromatin states for each condition are included. All tracks are shown at equal scale.

FIGURE 3. Association between histone modifications differential enrichment and changes in gene expression.

A. Heatmaps showing clusters of genes (-2Kb) grouped by unique histone modification profiles identified in the soft clustering analysis (left) and corresponding changes in mRNA levels (right). ChIP-seq enrichment at the promoters and gene bodies is normalized (RPKM) and input-corrected. The signal corresponds to the s ratio of ChIP-seq and mRNA levels in the infected versus the control condition. Data is log2-scaled and mean-centered.

B. Ratio of gene expression and histone modifications enrichment between infected and control conditions for Mfuzz gene clusters more highly expressed in infected (left) and control (right) conditions. Data is the log2-scaled ration between the infected and the control as in (A). Spearman rank correlation coefficient (rho) and corresponding P-value are shown for significant correlations between histone enrichment modifications and mRNA levels.

C. Histone enrichment profiles in the region containing the AGAP009887 coding gene. Tracks show normalized/input-corrected ChIP-seq signals and RNA-seq mapped reads counts for each condition. The location of diffReps regions, MACS2 peaks and predicted chromatin states for each condition are included. All tracks are shown at equal scale.

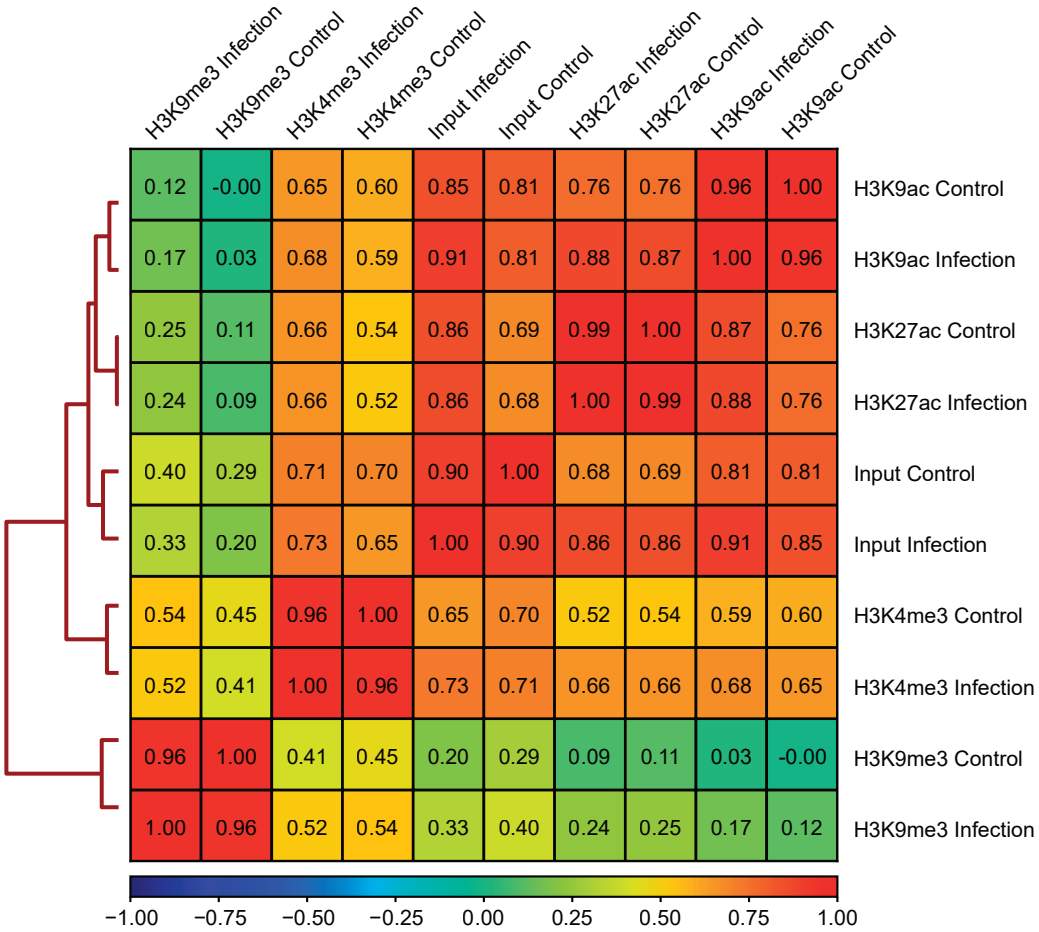
FIGURE 4. Significant differential gene expression and association with histone modifications differential enrichment.

A-B. Histone enrichment profiles in the regions containing the IAP7 (AGAP007293) and Argonaute 4 (AGAP011717) encoding genes. Tracks show normalized/input-corrected ChIP-seq signals and RNA-seq mapped reads counts for each condition. The location of diffReps regions, MACS2 peaks and predicted chromatin states for each condition are included. All tracks are shown at equal scale.

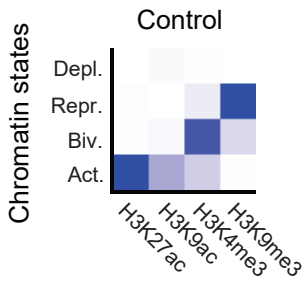
REFERENCES

1. Brucker, R.M., et al., Insect Innate Immunity Database (IIID): an annotation tool for identifying immune genes in insect genomes. PLoS One, 2012. 7(9): p. e45125.

A



B



C

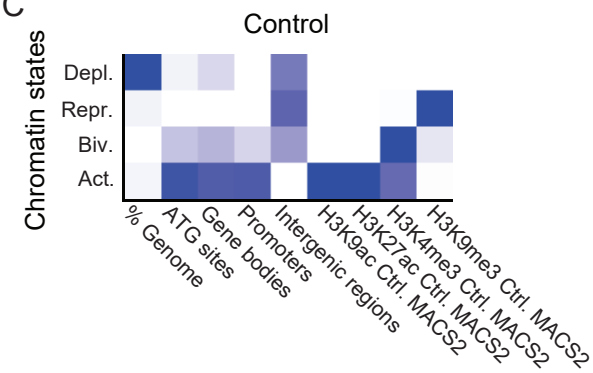


Figure S1

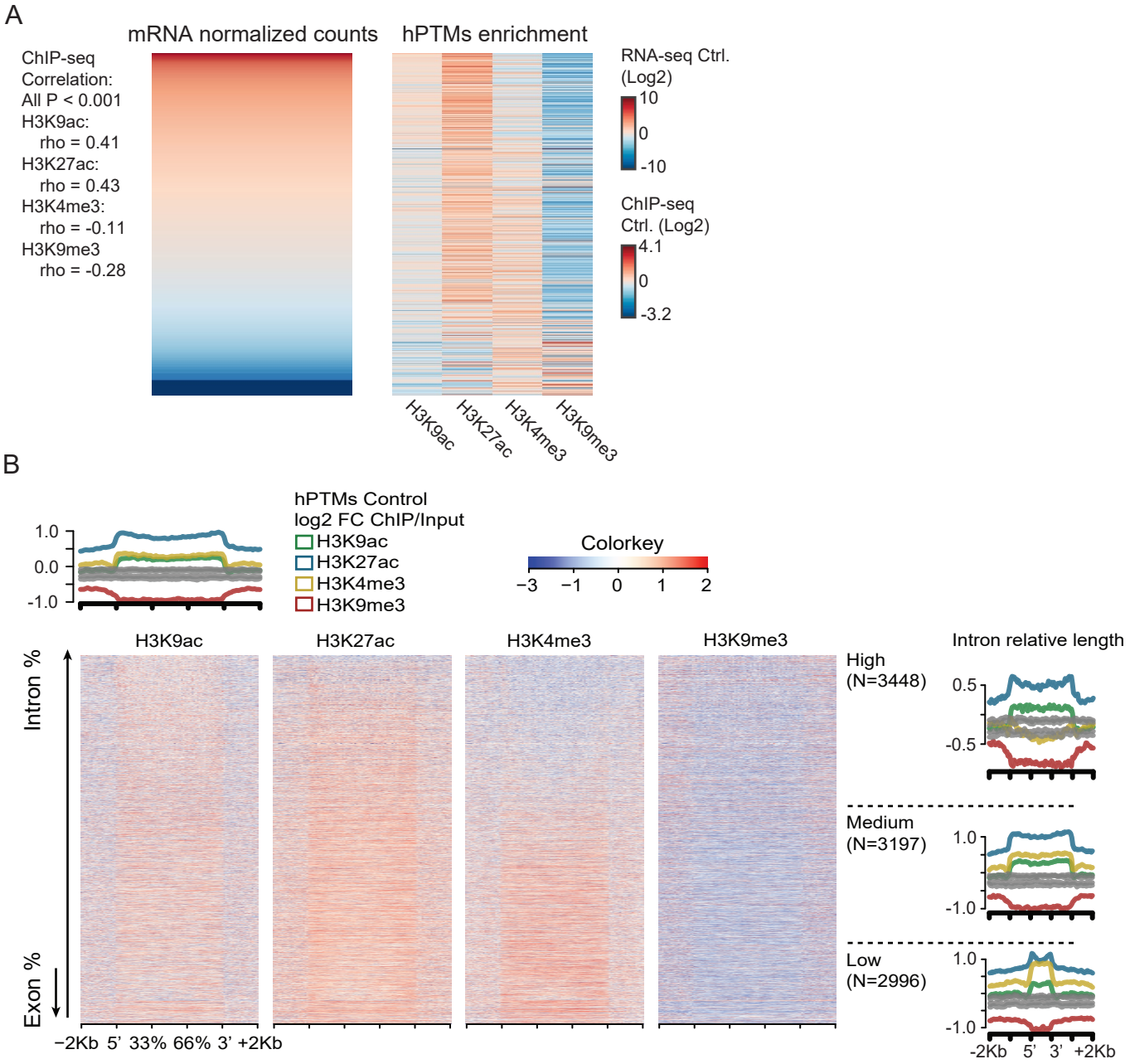

A
mRNA normalized counts
hPTMs enrichment
ChIP-seq
Correlation:
All P < 0.001
H3K9ac:
rho = 0.41
H3K27ac:
rho = 0.43
H3K4me3:
rho = -0.11
H3K9me3
rho = -0.28
RNA-seq Ctrl.
(Log2)
10
0
-10
ChIP-seq
Ctrl. (Log2)
4.1
0
-3.2
H3K9ac
H3K27ac
H3K4me3
H3K9me3
B
hPTMs Control
log2 FC ChIP/Input
H3K9ac
H3K27ac
H3K4me3
H3K9me3
Colorkey
−3 −1 0 1 2
1.0
0.0
-1.0
Intron %
Exon %
−2Kb 5' 33% 66% 3' +2Kb
Intron relative length
High
(N=3448)
0.5
-0.5
Medium
(N=3197)
1.0
-1.0
Low
(N=2996)
1.0
-1.0
-2Kb 5' 3' +2Kb

Figure S2

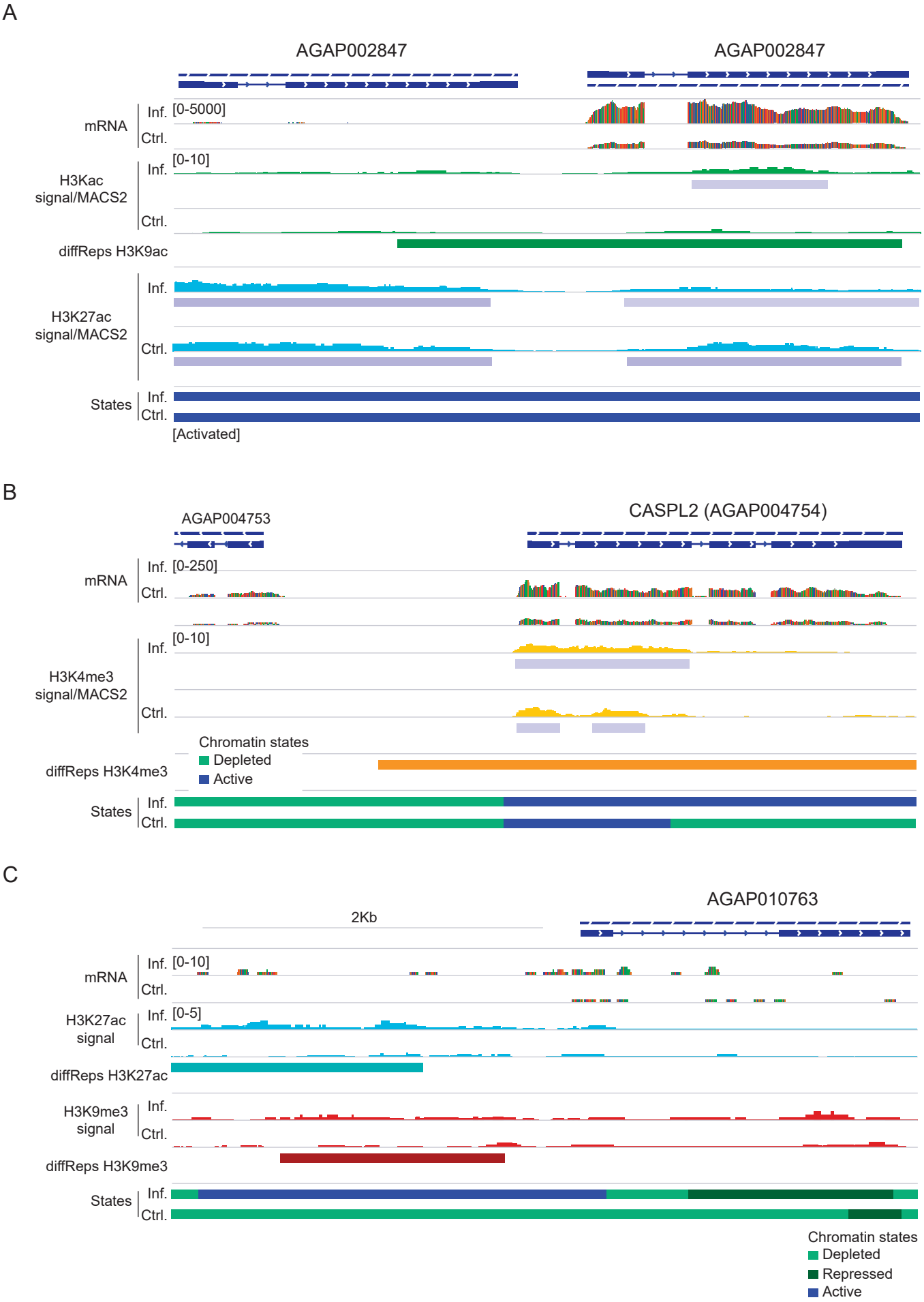
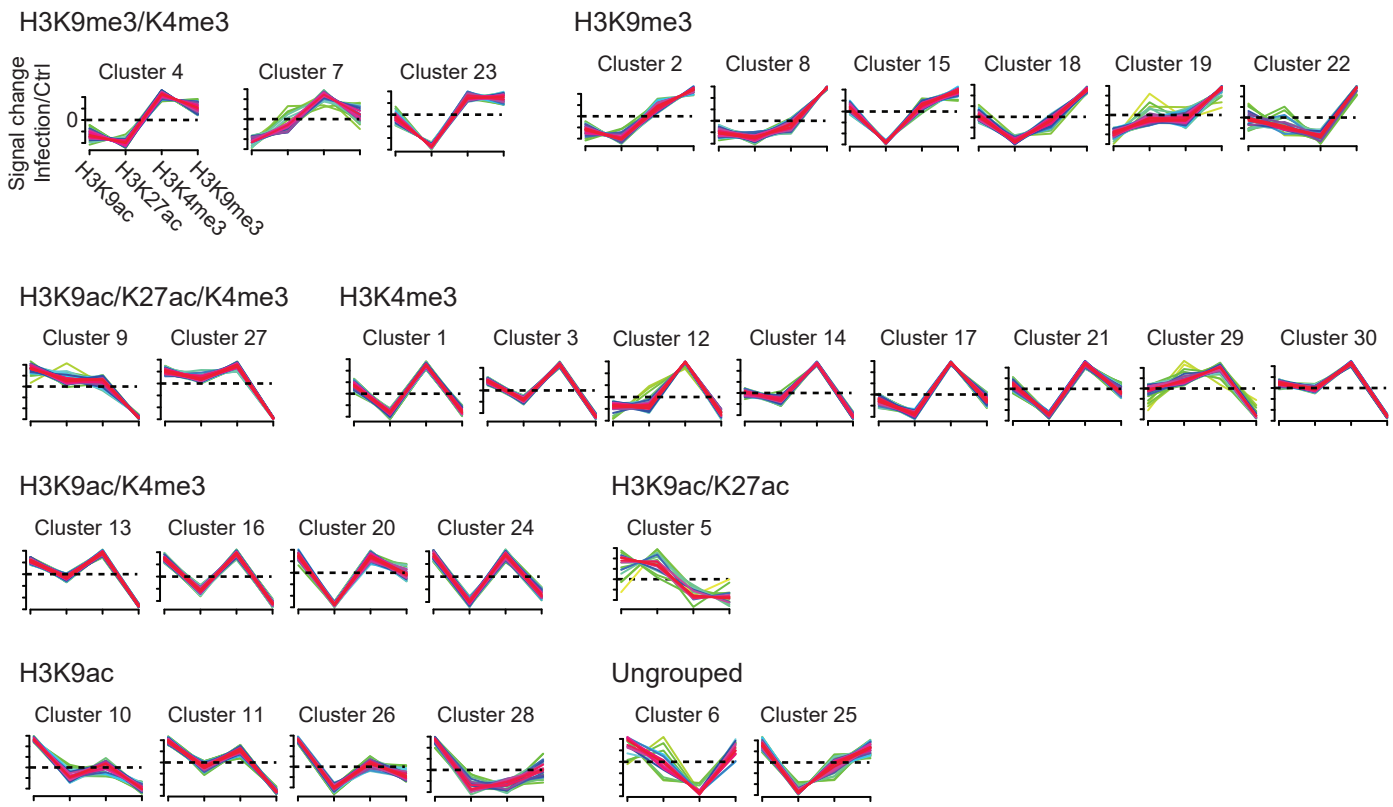

A
AGAP002847
AGAP002847
mRNA
Inf.
[0-5000]
Ctrl.
H3Kac signal/MACS2
Inf.
[0-10]
Ctrl.
diffReps H3K9ac
H3K27ac signal/MACS2
Inf.
Ctrl.
States
Inf.
Ctrl.
[Activated]
B
AGAP004753
CASPL2 (AGAP004754)
mRNA
Inf.
[0-250]
Ctrl.
H3K4me3 signal/MACS2
Inf.
[0-10]
Ctrl.
Chromatin states
Depleted
Active
diffReps H3K4me3
States
Inf.
Ctrl.
C
AGAP010763
2Kb
mRNA
Inf.
[0-10]
Ctrl.
H3K27ac signal
Inf.
[0-5]
Ctrl.
diffReps H3K27ac
H3K9me3 signal
Inf.
Ctrl.
diffReps H3K9me3
States
Inf.
Ctrl.
Chromatin states
Depleted
Repressed
Active

Figure S3

A

Mfuzz analysis Infection patterns



Mfuzz analysis Control patterns

H3K9me3/K4me3	H3K9me3	H3K4me3
Cluster 5	Cluster 3,9,10,11,12,13,14,16,27,29,30	Cluster 6,22

H3K27ac	H3K9ac/K27ac	H3K9ac
Cluster 15,17,18,20,21,23,24,25,26,28	Cluster 2,4,8	Cluster 7,19

Figure S4

A

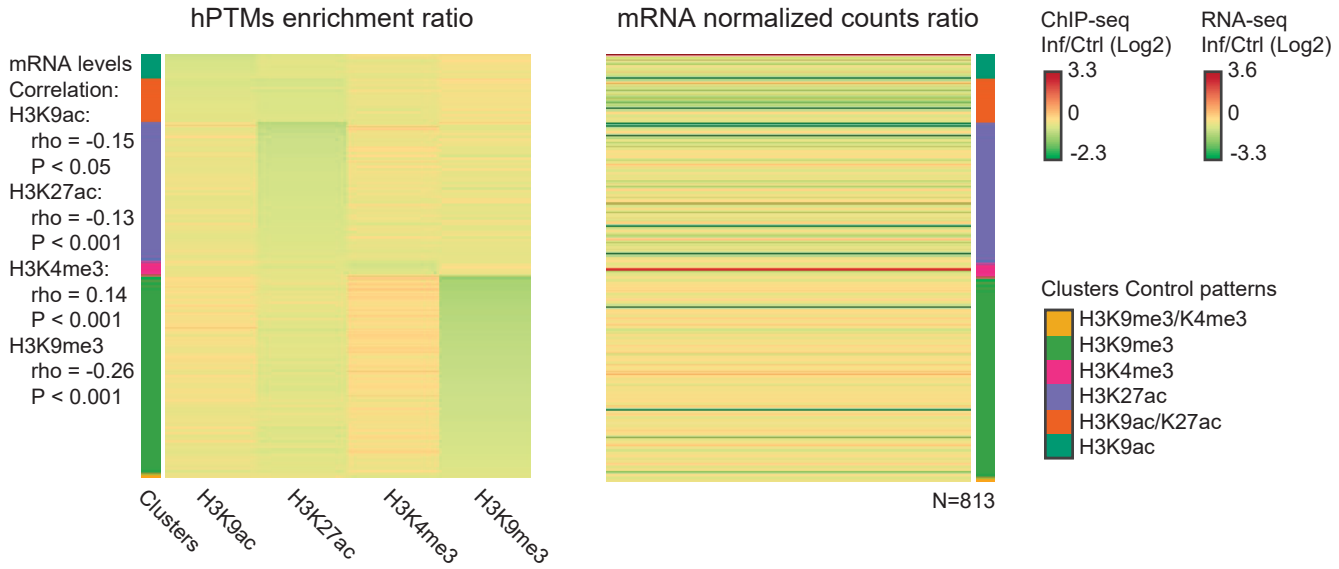

hPTMs enrichment ratio
mRNA normalized counts ratio
mRNA levels
Correlation:
H3K9ac:
rho = -0.15
P < 0.05
H3K27ac:
rho = -0.13
P < 0.001
H3K4me3:
rho = 0.14
P < 0.001
H3K9me3
rho = -0.26
P < 0.001
Clusters
H3K9ac
H3K27ac
H3K4me3
H3K9me3
N=813
ChIP-seq
Inf/Ctrl (Log2)
3.3
0
-2.3
RNA-seq
Inf/Ctrl (Log2)
3.6
0
-3.3
Clusters Control patterns
H3K9me3/K4me3
H3K9me3
H3K4me3
H3K27ac
H3K9ac/K27ac
H3K9ac

Figure S5

A

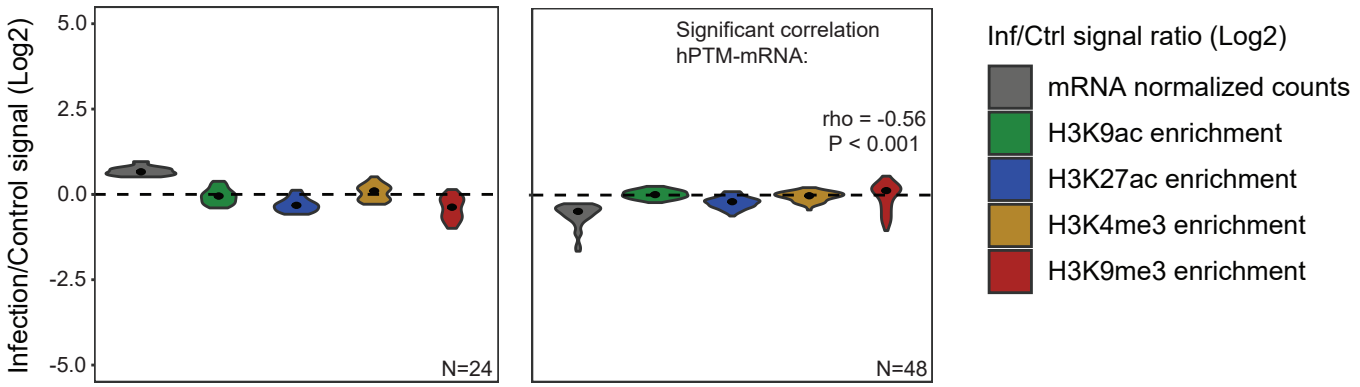

Infection/Control signal (Log2)
5.0
2.5
0.0
-2.5
-5.0
N=24
Significant correlation
hPTM-mRNA:
rho = -0.56
P < 0.001
N=48
Inf/Ctrl signal ratio (Log2)
mRNA normalized counts
H3K9ac enrichment
H3K27ac enrichment
H3K4me3 enrichment
H3K9me3 enrichment

Figure S6

A

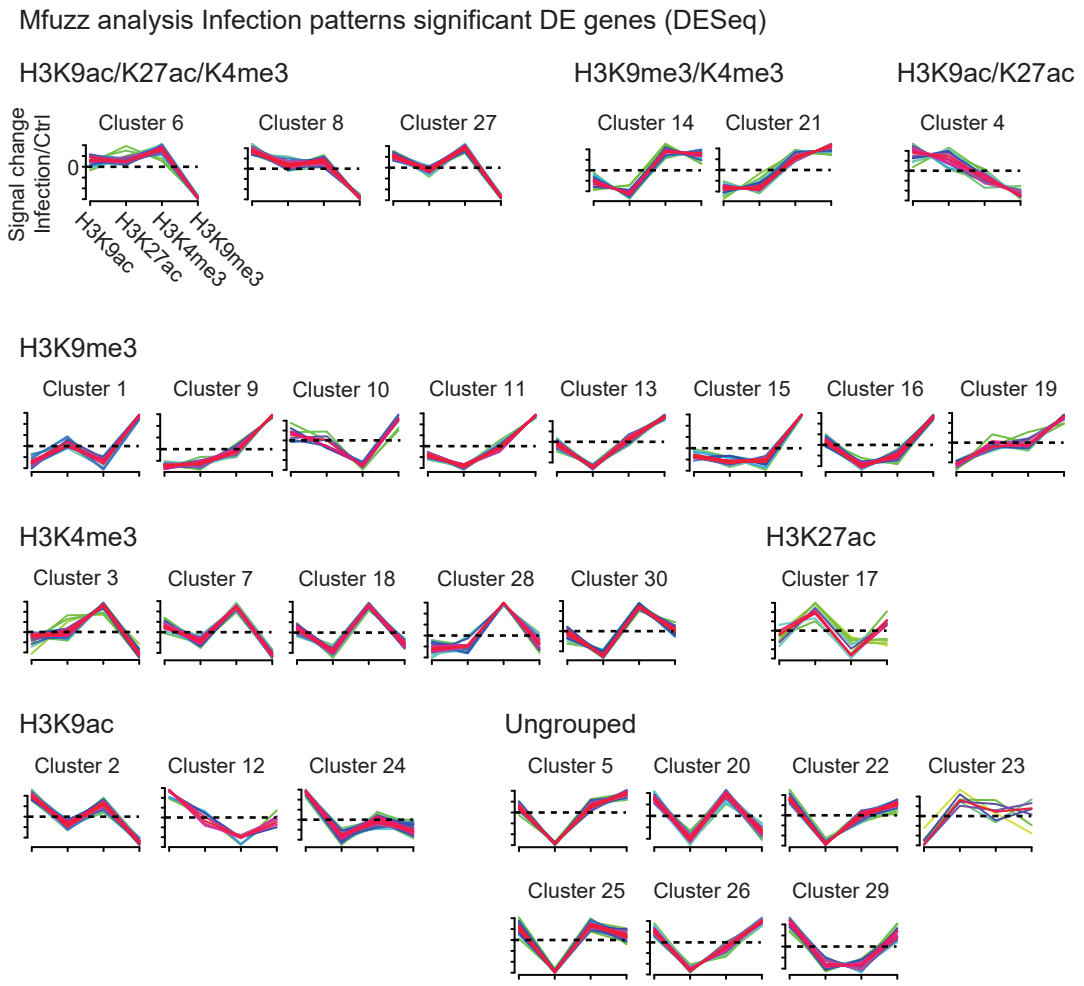
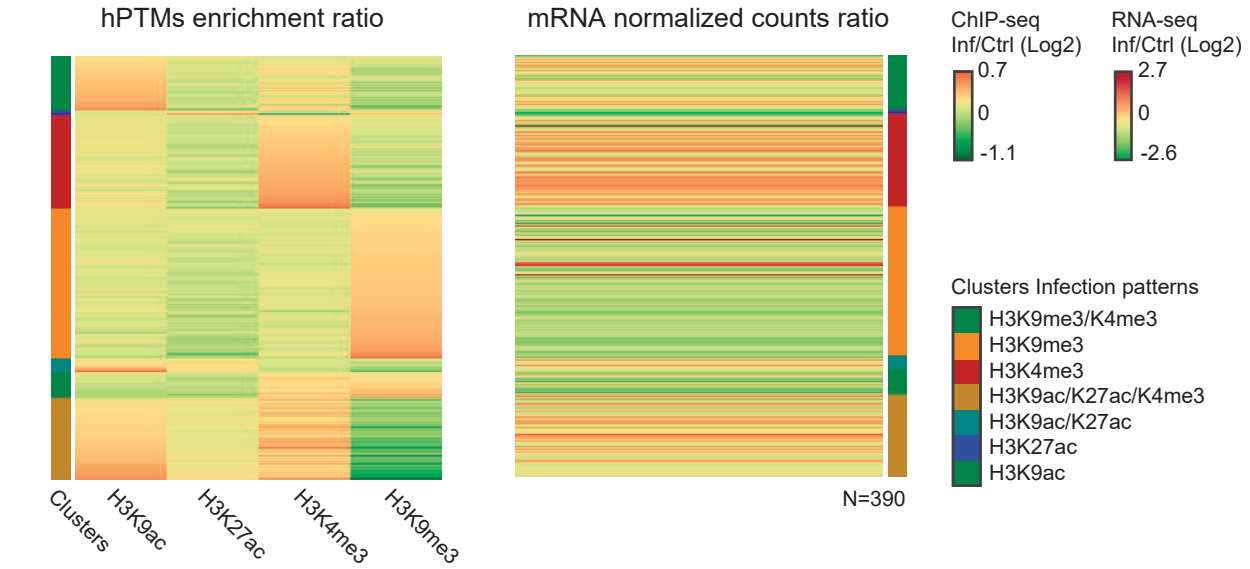

Mfuzz analysis Infection patterns significant DE genes (DESeq)
H3K9ac/K27ac/K4me3
H3K9me3/K4me3
H3K9ac/K27ac
Signal change Infection/Ctrl
0
Cluster 6
Cluster 8
Cluster 27
Cluster 14
Cluster 21
Cluster 4
H3K9ac
H3K27ac
H3K4me3
H3K9me3
H3K9me3
Cluster 1
Cluster 9
Cluster 10
Cluster 11
Cluster 13
Cluster 15
Cluster 16
Cluster 19
H3K4me3
Cluster 3
Cluster 7
Cluster 18
Cluster 28
Cluster 30
H3K27ac
Cluster 17
H3K9ac
Cluster 2
Cluster 12
Cluster 24
Ungrouped
Cluster 5
Cluster 20
Cluster 22
Cluster 23
Cluster 25
Cluster 26
Cluster 29

Figure S7

A


hPTMs enrichment ratio
mRNA normalized counts ratio
ChIP-seq
Inf/Ctrl (Log2)
0.7
0
-1.1
RNA-seq
Inf/Ctrl (Log2)
2.7
0
-2.6
Clusters
H3K9ac
H3K27ac
H3K4me3
H3K9me3
N=390
Clusters Infection patterns
H3K9me3/K4me3
H3K9me3
H3K4me3
H3K9ac/K27ac/K4me3
H3K9ac/K27ac
H3K27ac
H3K9ac

B

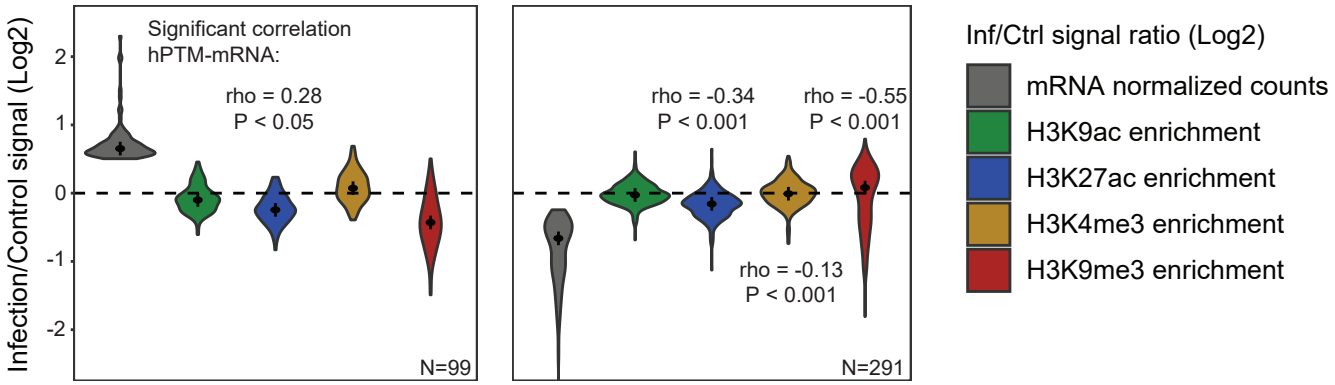

Infection/Control signal (Log2)
2
1
0
-1
-2
Significant correlation
hPTM-mRNA:
rho = 0.28
P < 0.05
N=99
rho = -0.34
P < 0.001
rho = -0.55
P < 0.001
rho = -0.13
P < 0.001
N=291
Inf/Ctrl signal ratio (Log2)
mRNA normalized counts
H3K9ac enrichment
H3K27ac enrichment
H3K4me3 enrichment
H3K9me3 enrichment

Figure S8

SUPPLEMENTAL FIGURES LEGENDS

FIGURE S1

A. Spearman correlation plots between histone modifications and input ChIP-seq libraries. Libraries are subsampled to equal number of reads (9M) and ordered following a hierarchical clustering approach.

B. Heatmap of emission parameters from ChromHMM analysis using a four chromatin-states model based on histone modifications enrichment patterns in the control condition. The predicted states are: Deplet. (depleted, low levels of all hPTMs), Repr. (repressive, H3K9me3 enrichment), Biv (bivalent, H3K4me3/H3K9me3 enrichment) and Act. (active, H3K27ac/H3K9ac/H3K4me3 enrichment). Darker blue indicates higher enrichment of a particular histone modification.

C. Heatmap showing the overlap of various genomic features, including MACS2 peaks located in promoters (-2Kb from the ATG) or gene bodies in the control condition, with the predicted chromatin states. Darker blue in the first column indicates higher percentage of the genome overlapped by a given state. For other columns, it indicates the likelihood of finding a particular chromatin state in each genomic feature compared to what it would be expected by chance.

FIGURE S2

A. Heatmaps showing mRNA levels (left) and histone modifications enrichment profiles (right), of genes displaying a MACS2 peak in the promoter or the gene body. Data corresponds to the control condition. Genes are ordered by mRNA levels. ChIP-seq enrichment at the promoters and gene bodies is normalized (RPKM) and input-corrected. Data is log2-scaled and mean-centered. Spearman rank correlation coefficient

(rho) and corresponding P-value are shown for the association between for each histone modification enrichment levels and mRNA levels.

B. Heatmaps showing histone modifications enrichment profiles at high and medium expressed genes in the control condition. Genes are ordered by the percentage of the bodies containing introns and exons. Average profile plots show density of normalized (RPKM) and input corrected ChIP-seq reads for each histone modification at high and medium expressed genes (top) and at those genes classified by the percentage of gene bodies containing introns (right).

FIGURE S3

A-C. Histone enrichment profiles in the regions containing the Niemann-Pick Type C-2 genes (AGAP002847-AGAP002848) and CASPL2 (AGAP004754) and fibrinogen-related protein (AGAP010763) encoding genes. Tracks show normalized/input-corrected ChIP-seq signals and RNA-seq mapped reads counts for each condition. The location of diffReps regions, MACS2 peaks and predicted chromatin states for each condition are included. All tracks are shown at equal scale.

FIGURE S4

A. Histone modifications enrichment profiles for clusters identified in the Mfuzz soft clustering analysis. Signal is computed as the ratio of normalized (RPKM) input-corrected ChIP-seq reads in infection relative to the control condition at promoters and genes bodies containing high confidence differential diffReps regions. Data is log2-scaled. Yellow/green colored profiles represent regions with low membership and red/purple elements with high membership value to each cluster. Clusters are classified

in groups that share unique histone modifications enrichment patterns. Some clusters showing unspecific profiles remain ungrouped.

FIGURE S5

A. Heatmaps showing clusters of genes (-2Kb) grouped by unique histone modification profiles identified in the soft clustering analysis (left) and the corresponding changes in mRNA levels (right). ChIP-seq enrichment at the promoters and gene bodies is normalized (RPKM) and input-corrected. The signal corresponds to the ratio of ChIP-seq and mRNA levels in the infected versus the control condition. Data is log2-scaled and mean-centered.

FIGURE S6

A. Ratio of gene expression and histone modifications enrichment between infected and control conditions for genes with significant differential gene expression (DESeq2) and containing high confidence diffReps regions in the promoter or the gene body. Genes are divided by highly expressed in infected (left) and non-infected (right) mosquito tissues. ChIP-seq enrichment at promoters and gene bodies is normalized (RPKM) and input-corrected. Data is the log2-scaled ratio of mRNA and histone modifications enrichment levels between the infected and the control conditions. Spearman rank correlation coefficient (rho) and corresponding P-value are shown for significant correlations between histone enrichment modifications and mRNA levels.

FIGURE S7

A. Histone modifications enrichment profiles for clusters identified in the Mfuzz soft clustering analysis. Signal is computed as the ratio of normalized (RPKM) input-

corrected ChIP-seq reads in the infection relative to the control condition for genes showing significant differential levels of expression according to DESeq2 analysis. Data is log2-scaled. Yellow/green colored profiles represent regions with low membership and red/purple elements with high membership value to each cluster. Clusters are classified in groups that share unique histone modifications enrichment patterns. Some clusters showing unspecific profiles remain ungrouped.

Figure S8.

A. Heatmaps showing clusters of genes (-2Kb) grouped by unique histone modification profiles identified in the soft clustering analysis (left) and corresponding changes in mRNA levels (right). ChIP-seq enrichment at promoters and gene bodies is normalized (RPKM) and input-corrected. The signal corresponds to the ratio of ChIP-seq and mRNA levels in the infected versus the control condition. Data is log2-scaled and mean-centered.

B. Ratio of gene expression and histone modifications enrichment between infected and control conditions for DESeq2 gene clusters more highly expressed in infected (left) and control (right) conditions. Data is the log2-scaled ration between the infected and the control as in (A). Spearman rank correlation coefficient (rho) and corresponding P-value are shown for significant correlations between histone enrichment modifications and mRNA levels.

SUPPLEMENTARY TABLES LEGENDS

TABLE S1. Summary of RNA-seq and ChIP-seq alignment statistics.

TABLE S2. Set of MACS2 ChIP-seq peaks for each histone modification in the infected and non-infected mosquito tissues. Information on the annotation to genomic

features, location, Gene Ontology (GO) terms and data from PFAM and KEGG databases, is included.

TABLE S3. Set of differential histone modifications regions between infected and non-infected mosquito tissues. Information on the annotation to genomic features, location, predicted chromatin states, MACS2 peaks overlap, mRNA levels of the nearest gene, GO terms and data from PFAM and KEGG databases, is included. The set of high confidence diffReps regions and the GO Biological Processes overrepresentation analysis are shown in separate sheets.

TABLE S4: Results of the soft clustering analysis on the set of genes with high confidence diffReps regions located in the promoter or the gene body. Genes are clustered based on the histone modifications enrichment profile (ratio of enrichment in infected versus controls). The ratio is normalized (RPKM) and input-corrected. The ChIP-seq signal for each histone modification and mRNA levels of the corresponding genes are included.

TABLE S5: Set of genes showing significant differential gene expression according to DESeq2 analysis. Information on GO terms and PFAM and KEGG database annotations are included. Overlap with high confidence diffReps regions is indicated.

TABLE S6: Results of the soft clustering analysis on the set of significant differentially expressed genes according to DESeq2 analysis. Genes are clustered based on the histone modifications enrichment profile (ratio of enrichment in infected versus control). The ratio is normalized (RPKM) and input-corrected. The ChIP-seq signal for each histone modification and mRNA levels of the corresponding genes are included.

REFERENCES

1. Brucker, R.M., et al., Insect Innate Immunity Database (IIID): an annotation tool for identifying immune genes in insect genomes. PLoS One, 2012. 7(9): p. e45125.