Title
Epigenetic Factors Regulating Toxoplasma gondii Recrudescence Fate

Permalink
https://escholarship.org/uc/item/4pz745qs

Author
Ciampossin, Loic

Publication Date
2023

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Epigenetic Factors Regulating *Toxoplasma gondii* Recrudescence Fate

A Thesis submitted in partial satisfaction
of the requirements for the degree of

Master of Science

in

Biomedical Sciences

by

Locic Clement Ciampossin

September 2023

Thesis Committee:
  Dr. Karine Le Roch, Chairperson
  Dr. Emma Wilson
  Dr. Meera Nair

The Thesis of Loic Clement Ciampossin is approved:

Committee Chairperson

University of California, Riverside

Acknowledgment

I would like to thank Dr. Karine Le Roch, my colleagues, and our collaborators for being welcomed and given the opportunity to train and work on a number of compelling projects to further our understanding of the molecular underpinnings of parasite biology.

ABSTRACT OF THE THESIS

Epigenetic Factors Regulating *Toxooplasma gondii* Recrudescence Fate

by

Loic Clement Ciampossin

Master of Science, Graduate Program in Biomedical Sciences
University of California, Riverside, September 2023
Dr. Karine Le Roch, Chairperson

Toxoplasma gondii is a prolific parasite known to infect a wide variety of warm-blooded animals, including humans. The parasite can interconvert between a dormant form of the parasite (bradyzoites), and an acute form of the parasite (tachyzoites), which manifests as Toxoplasmosis, in a process called recrudescence. This process is tightly controlled through a series of changes in gene expression in response to environmental pressure. Histone modifications are effective post-translational modifications, influencing the parasite's gene expression through different stages of its life cycle. Despite knowing that recrudescence is, in part, epigenetically regulated, little is known about how environmental factors, such as the type of host cell, affect this process. This study uses histone modifications to observe changes in the epigenomic profile of the parasite as they develop within two different host cells, Astrocytes (ASTs) and Human Foreskin Fibroblasts (HFFs). Using Chromatin Immunoprecipitation followed by deep Sequencing (ChIP–seq) against three different histone marks, we revealed a higher abundance of histone H3K4me3 activating marks within ASTs. This observation was supported by an increased binding affinity of H3K4me3 in ASTs and a decline in the number of affinity binding sites in HFF as recrudescence progressed. Of the annotated genes that were identified to have a high affinity with H3K4me3, several had potential roles in invasion, chromatin remodeling, and cell cycle progression.

Table of Contents

Chapter 1

List of Figures

Introduction

Materials and Methods

Results

Appendix

List of Tables

Results

Appendix

.

Chapter 1

Chapter 1.1 Introduction

Toxoplasmosis is a parasitic disease caused by *Toxoplasma gondii*, a prolific obligate endoparasite that belongs to the phylum Apicomplexa. Toxoplasmosis is considered to be a leading cause of death attributed to foodborne illness in the United States. Nearly one-third of the world population, with an estimated 60 million people in the United States, are chronically infected with the parasite [1]. While the infection rate is high, most of the infected individuals are either asymptomatic or have mild symptoms that may include fever and general pain. Macular degeneration and encephalitis can occur in severe cases. In immunocompetent people, the immune system generally quickly contains the infection, but the parasite is rarely eradicated from the host and is latent throughout its lifetime [2, 3]. In immunodeficient patients, *T. gondii* can however cause severe neurologic damage and even death. Symptoms can also be serious to the fetus if the parasite is acquired during pregnancy.

As described above, *T. gondii* belongs to the Apicomplexa, a large group of parasitic organisms that possess a unique set of organelles, including a non-photosynthetic plastid called apicoplast and an apical complex structure critical for host cell invasion. *T. gondii* has a complex life cycle, infecting all warm-blooded animals, but the parasite can only undergo sexual reproduction within its definitive hosts (felids) (Fig. 1) [4].

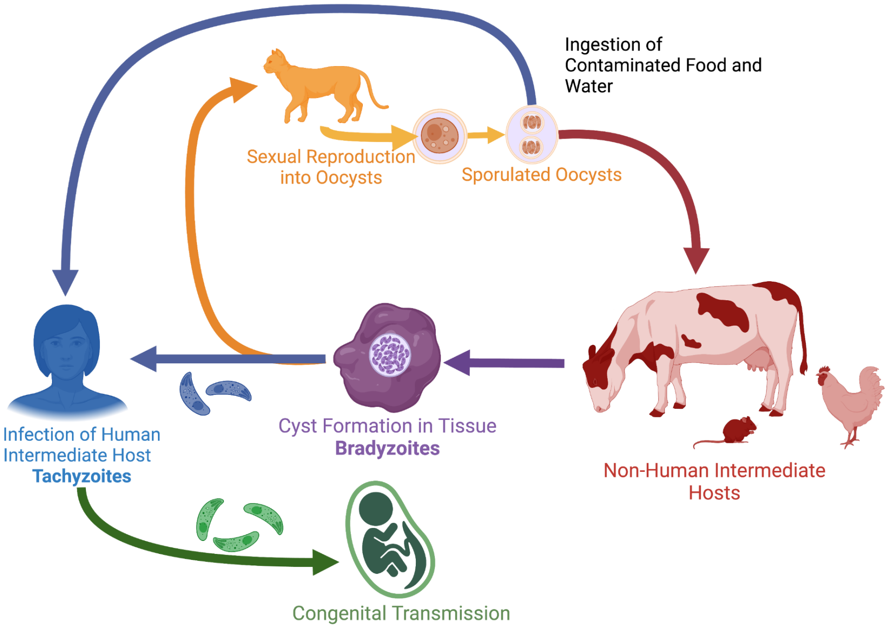


Figure 1. The life cycle of *T. gondii*. Diagram depicting the sexual stage (oocysts) within its definitive host (cats) and asexual stages (bradyzoites and tachyzoites) within intermediate hosts (Humans and Non-Humans).

Environmental conditions and dietary habits can significantly impact the infection. Ingestion of raw or undercooked meat, especially sheep, is generally associated with high transmission rates. There are three developmental *T. gondii*'s life-cycle stages that can infect cells through ingesting contaminated food or water (Fig.1). **(i)** the tachyzoite, an asexual and aggressive form of rapid multiplication found in acute infections; **(ii)** the bradyzoite, an asexual form of slow multiplication of the chronic infection from the tissue cysts; and finally, **(iii)** the sporozoite, which is released from the mature sporulated oocysts excreted in feline feces. Infections are also possible through congenital transmission from the mother through the placenta, blood transfusion or organ transplantation (Fig.1) [4].

The inability to completely clear infection of *T. gondii* is closely related to an immune stress cascade that leads **tachyzoites** to differentiate into its second asexual form, the **bradyzoite** [5]. Bradyzoites are morphologically, physiologically, and behaviorally distinct from tachyzoites [6, 7]. Many of these distinctions facilitate immune escape and even drug resistance for the parasite. The most distinguishable of these changes is that bradyzoites do not undergo lytic replication but instead slow their metabolism and limit the replication rate within a protective cyst. The cyst wall is a modified parasitophorous vacuole (PV) that develops into a several nanometer thick and heavily glycosylated, non-permeable barrier. These morphological changes to the cyst wall are accompanied by modifications to the antigens and receptors on the surface of the cyst, potentially de-escalating an immune response and denying the uptake of stage-specific drugs [5-7].

Bradyzoites are problematic throughout *T. gondii's* pathogenesis as they can remain in their dormant state for extended periods until they undergo **bradyzoite to tachyzoite (BtT)** recrudescence and return to their acute pathogenic form. Recrudescence occurs opportunistically when pressure from the immune system declines, either from age or immunodeficiency, such as co-infection with the HIV virus, which causes long-term immune compromisation [4]. Re-invasion begins with an excystation event, wherein the protective cyst of the parasite ruptures and releases bradyzoites into the local tissue. Once released (day 0), the parasites differentiate and propagate over the next few days as the pathogenesis progresses [8, 9]. Within the brain and other neurological tissue, serial excystation and subsequent cyst formation events contribute to the formation of progressive tissue lesions, clinically manifesting as encephalitis and other lesion-related complications.

BtT recrudescence was believed to be the sole recrudescence pattern in *T. gondii*. However, preliminary work has suggested that two distinct recrudescence pathways could exist [9]. The proposed second pathway is a **bradyzoite to bradyzoite (BtB)** recrudescence. This phenomenon proposes that bradyzoites can undergo an excystation event, but a small portion of the population **can differentiate into tachyzoites or will quickly re-differentiate back into bradyzoites.**

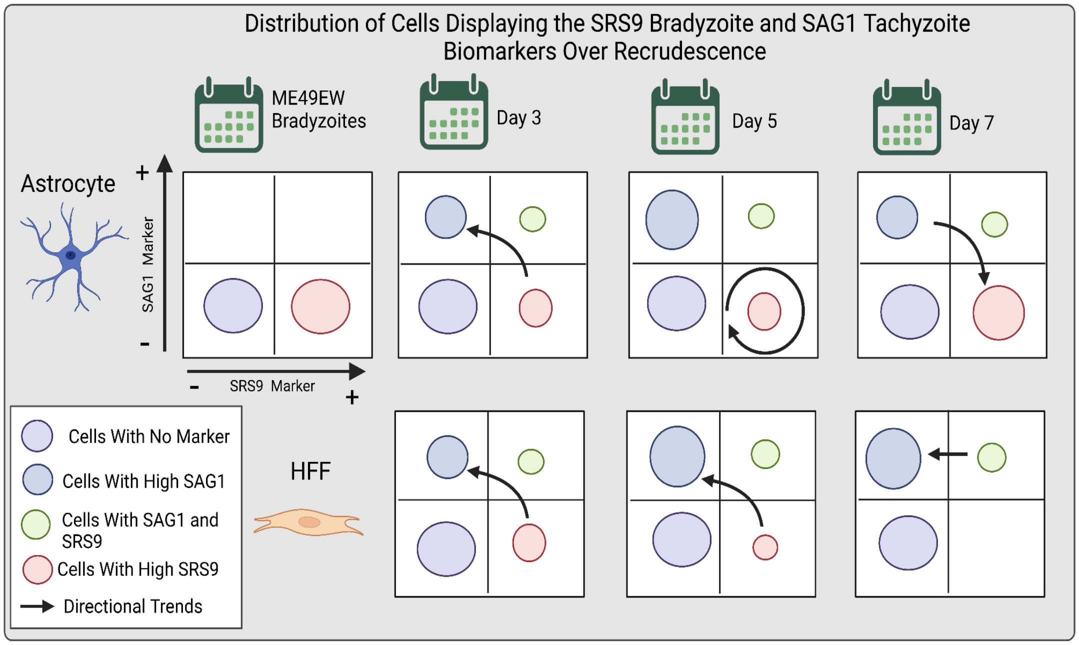


Figure 2. The proposed BtT and BtB recrudescence pathways in two different host cell lines. Simplification of a fluorescence-activated cell sorting (FACS) experiment using SRS9 and SAG1 as markers of bradyzoite and tachyzoite, respectively, during infection at day 0, 3, 5, and 7. Arrows are used to represent the trend of signal migration within the sample visually. The experiment was performed in parallel with two different cell lines: AST = Astrocytes and HFF = Human Foreskin Fibroblasts. The outputs are color coordinated: Purple = cells with neither marker, blue = cells with high SAG1, green = cells with both SAG1 and SRS9 and red = cells with high SRS9.

These divergent recrudescence models were recently identified using complementary cellular, molecular and transcriptomic experiments. Those experiments included fluorescence-activated cell sorting (FACS), microscopy and single-cell sequencing along with well-known tachyzoite/bradyzoite markers of differentiation, SAG-1 and SRS9, respectively. To start these experiments, the authors of preliminary data by Amber L. et al., used a clone of T. gondii ME49, ME49EW, and cultured it in two distinct host cells; the **Human Foreskin Fibroblast (HFFs)** and the more biologically relevant **neonatal mouse (C57B/6) Astrocyte (AST)** cell lines (Fig. 3).

Using these two models, the team was able to demonstrate different rates of recrudescence pathways under distinct environmental conditions [9]. In this set of experiments, the team of researchers infected their two cell lines with ME49EW and followed the parasite morphologies with a set of distinct markers at four time points that included days 0, 3, 5, and 7 post-inoculation. Post-excystation (Day 0) corresponds to a starting bradyzoite infection with little time given for the cells to begin the differentiation process. At day 3, the differentiation rate of the parasites into tachyzoites from bradyzoites was maximized in both cell cultures. At day 5, in the AST cells (modeling BtB recrudescence), the number of parasites within the bradyzoite-like population plateaued, maintaining a steady turnover rate. In contrast, the fibroblast infecting bradyzoite population (modeling BtT recrudescence) continued to decline as they differentiated into tachyzoites. At day 7, the number of bradyzoites within the HFF and AST cell lines continued to diverge. Within the AST, a re-differentiation from tachyzoites back into bradyzoites was observed alongside the general maintenance of a small bradyzoite population, resulting in the growth of bradyzoite-like parasites. Meanwhile, in HFF, the bradyzoite population was greatly depleted.

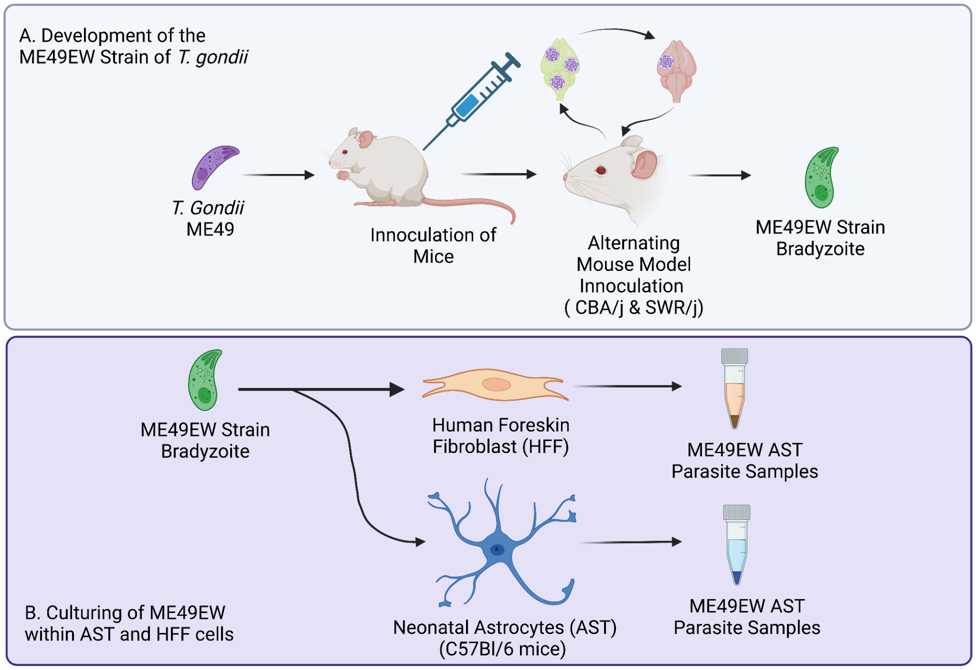


Figure 3. Summary of the generation of the ME49EW strain from the ME49 strain of *T. gondii*. (A) Cysts of *T. gondii* ME49 strain were alternatively inoculated into CBA/j (vulnerable) and SWR/j (resistant) mice. The ME49EW clone forms bradyzoite cysts with high yields. (B) A summary for culturing ME49EW parasites into Human Foreskin Fibroblasts (HFFs) and Astrocytes (AST) to obtain parasite samples.

While single-cell sequencing experiments validate the presence of these two distinct recrudescence pathways, **the molecular mechanisms regulating these two contrasting developmental processes need to be determined**.

Of the many different mechanisms regulating gene expression in an organism, **epigenetics** provide an adaptive layer to adjust the change in gene expression to a dynamic environment. Epigenetic mechanisms induce a reversible alteration of gene expression in an organism without changing the DNA sequence. They can include but are not limited to non-coding RNAs, chromatin structure, DNA methylation, and histone modifications [10]. Histones are protein complexes formed by the association of 2 sets of 4 core histone proteins, H2A, H2B, H3, and H4. This complex interacts with DNA, to form a nucleosome. Nucleosomes have several secondary flexible external structures called histone tails, which can be modified to alter the interaction between the histones and the DNA. **Post-translational modifications (PTMs)**, such as methylation, acetylation, or ubiquitination to the histone tails, are well-known epigenetic mechanisms in all eukaryotic organisms and have also been demonstrated to regulate gene expression in *T. gondii* [11, 12]. Thus, changes in epigenetics features may explain the changes in the recrudescence models that have been observed between AST and HFF cell lines.

Histone modifications are functionally categorized by one of three effects on the genome structure and gene expression: activation, repression, or poised modification. An activating histone modification includes the conversion of heterochromatin into euchromatin by weakening the association of DNA around the histone. This makes the DNA more accessible to transcription machinery or even directly recruits certain transcription factors. H3K4me3 is an activating histone modification composed of a tri-methylation of the 4th Lysine on the H3 histone tail and is primarily situated near promoter regions of actively transcribed genes [10]. In contrast, Histone H3 Lysine 9 tri-methylation (H3K9me3) mark typically converts euchromatin into heterochromatin by constricting DNA around the nucleosome, making the DNA inaccessible. This prevents certain transcription factors from binding to the DNA and thus represses transcription [10]. In *T. gondii*, H3K9me3 is primarily found near the centromeres [13, 14]. The poised modifications that include Histone H3 Lysine 14 acetylation (H3K14ac) can incorporate activating and repressing histone modifications and are the least understood category of histone modification. It is believed to be most relevant for genes involved in the cell cycle, cell identity, and somatic development [12]. In *T. gondii*, H3K14ac may be a marker of a poised chromatin state, but its role remains poorly studied [14, 15].

To analyze epigenetics patterns at a genome-wide level, scientists have developed a technique called **Chromatin immunoprecipitation followed by deep sequencing (ChIP–seq)**. ChIP–seq is a powerful method to map DNA-binding proteins, including histone modifications, in a genome-wide manner at single base-pair resolution. Genome alignment of the sequenced reads can provide detailed profiling of histone modifications and nucleosome positions, enabling a greater understanding of epigenetic mechanisms controlling gene regulation during parasite development and their adaptation into their host cells.

Overall, we found two major differences between the host manifested at the genome level. The **Transcription Start Site (TSS)** coverage and the binding affinity of H3K4me3 were higher in the AST, with a corresponding depletion in the number of affinity binding sites for HFF as recrudescence progressed. At the individual gene level, we found differential binding of H3K4me3 on several genes between the host cells. A number of these genes held significance to the development of the parasite, including invasion, chromatin remodeling, metabolism and cell cycle regulation.

Chapter 1.2 Materials and Methods

ME49EW Strain

T. gondii ME49 strain, a strain commonly used to study pathogenesis in mouse models, was selected for this study. ME49 is a type II strain, presenting the *T. gondii* Surface Antigen 2 (SAG2), associated with symptomatic toxoplasmosis in humans and is known to infect over 50% of immunocompromised patients [9]. Using this strain, our collaborators inoculated two mice populations, CBA/j, a naive mouse model used to study pathogenesis, and SWR/j mice, a model which showed *T. gondii* resistance to strong infection and cyst formation. After being maintained in these culture systems for more than 20 years, ME49 developed into a high cyst yield strain that can grow under standard conditions [9]. Two partitions of the parasite clones were then selected to be cultured in two cell lines; the immortalized Human Foreskin Fibroblast cell line (HFFs), and the C57Bl/6 mice neonatal Astrocytes (AST). Parasites were then extracted at day 0, 3, 5 and 7 post-inoculation to follow changes in their epigenetic features.

ChIP-seq

For our ChiP-seq experiments, *T. gondii* parasites (approximately 22 million per sample) were crosslinked with paraformaldehyde (1%), and resuspended in 1ml of a nuclear extraction buffer (10 mM HEPES, 10 mM KCl, 0.1 mM EDTA, 0.1 mM EGTA, 1 mM DTT, 0.5 mM AEBSF, 1X Roche protease inhibitor tablet, and 1X Roche phosphatase inhibitor tablet) and incubated on ice for 30-minutes. Following the incubation, 0.1mM Igepal CA-630 was added to each sample and homogenized by passing the sample through a 26 gauge needle. After centrifugation at 5,000 rpm, the pellet was resuspended in shearing buffer (0.1% SDS, 1 mM EDTA, 10 mM Tris–HCl (pH 7.5), 1X Roche protease inhibitor tablet, and 1X Roche phosphatase inhibitor tablet) and transferred to a 130 µl Covaris sonication tube. Each sample was subjected to mechanical shearing by a Covaris S220 Ultrasonicator for 7 minutes and 20 seconds (duty cycle: 5%, intensity peak power: 140, cycles per burst: 200, bath temperature: 6°C). ChIP dilution buffer (30 mM Tris–HCl (pH 8), 3 mM EDTA, 0.1% SDS, 30 mM NaCl, 1.8% TritonX-100, 1X protease inhibitor tablet, 1X phosphatase inhibitor tablet) was then added before centrifugation for 10 minutes at 13,000 rpm at 4°C, and the supernatant containing the sheared chromatin was collected.

Each sample was pre-cleared by adding 13 µl of washed protein A agarose/salmon sperm DNA beads for 1 hour at 4°C with agitation. The samples were then transferred to new tubes after pelleting out the pre-clearing beads using centrifugation at 1,000 rpm. For each sample, approximately 10% by volume was kept as an input control. After, the appropriate antibody (H3K4me3, H3K9me3, H3K14ac, or IgG) was added to the samples and incubated overnight at 4°C with rotation. A biological replicate was included for each modification, and several samples were used for inputs, samples that did not undergo immunoprecipitation with an antibody and IgG-negative controls.

Protein A agarose/salmon sperm DNA beads (25 µl/ sample) were washed with an unsupplemented ChIP dilution buffer before being blocked with a 1mg/ml BSA solution at 4°C for 1 hour. After blocking, the beads were washed three times with the unsupplemented ChIP dilution buffer. Washed and blocked beads were added to each sample and incubated for 1hr at 4°C with rotation. Samples were removed from beads by centrifugation, and the bead/antibody/protein complex underwent a series of washes, each with a 15-minute incubation followed by centrifugation at 1,000 rpm. The washes were as follows: twice with low-salt immune complex wash buffer (1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris–HCl (pH 8), 150 mM NaCl), twice with high-salt immune complex wash buffer (1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris–HCl (pH 8), 500 mM NaCl), twice with LiCl wash buffer, and twice with TE buffer. The antibody-protein complexes were then eluted twice from the beads with an elution buffer (1% SDS, 0.1 M sodium bicarbonate). A 5M NaCl solution was added to each sample and incubated at 45°C overnight to reverse crosslinking.

The following day, 15 µl of 20 mg/ml RNase A was added to each sample and incubated at 37°C for 30 minutes. Following incubation, 10 µl of 0.5 M EDTA, 20 µl of 1 M Tris–HCl (pH 7.5), and 2 µl of 20 mg/mL proteinase K were added and incubated for 2 hr at 45°C. DNA recovery was performed via two extractions with phenol/chloroform/isoamyl alcohol (25:24:1) and a final chloroform extraction. The samples were then precipitated overnight at a temperature of -20°C.

After precipitation, samples were subjected to a 13,000 rpm centrifugation for 30 minutes at 4°C to pellet the DNA. After a wash with 80% ethanol, the pellet was air-dried and resuspended in 50 µl of nuclease-free water. DNA was purified using Agencourt AMPure XP beads (Beckman Coulter, CA, USA). Libraries were then prepared from this DNA using a KAPA library preparation kit (KK8230 & KK8500) according to the manufacturer’s instructions and sequenced on an Illumina NovaSeq 6000 machine.

Post-Sequencing Processing

Reads were processed as detailed below. We used three critical outputs: BAM files, Peak files, and TDF or WIG files.

FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) [15] was first used to analyze raw read quality. Any adapter sequences and low-quality reads (Phred < 25) were removed using Trimmomatic (http://www.usadellab.org/cms/?page=trimmomatic) [16]. The trimmed reads were mapped against the ME49 (ToxoDB.v62) genome using Bowtie2 (version 2.3.4.1), (https://bowtie-bio.sourceforge.net/bowtie2/index.shtml) [17]. Through Samtools (https://software.broadinstitute.org/software/igv/) [18], only properly paired reads with mapping quality 40 or higher were retained. Reads were then deduplicated by PicardTools (https://broadinstitute.github.io/picard/) [19]. During each of these steps, mapping statistics for each factor were collected and stored in a combined file. A script was used so that genome-wide read counts per nucleotide were normalized by dividing millions of mapped reads for each sample (for all samples, including input) and subtracting input read counts from the ChIP and IgG counts. Bedtools (https://bedtools.readthedocs.io/en/latest/) [20] is a toolset that converts the processed BAM files from binary to a human-readable format as a BED file. The BED files were converted into TDFs to be used with the IGV program and visualize the data on genome tracks. The BAM files were further processed using MACS2 (https://pypi.org/project/MACS2/) [21] to generate peak set data for each experimental condition. The BAM and Peak outputs were processed using DiffBind (https://bioconductor.org/packages/release/bioc/html/DiffBind.html) [22] to establish pearson correlations and visually represent them through correlation heat maps. The Bam Files were then processed with an in-house script to generate a broad epigenetic profile, scaled on the y-axis by read depth and by kilobases of nucleotide (Kb) on the x-axis. Next, we used DeepTools to observe the courage of the histone modifiers across set genomic intervals and form a data “bin”, specifically, 3 bins of reads from 1 kb upstream of the TSS, the TSS itself, and 2kb downstream of the TSS [23]. The read data from the bins were then layered into a heat map to show read density relative to the TSS, and a line plot was generated to better visualize the distribution of H3K4me3. DiffBind was run again, but with a different protocol to analyze the binding affinity of H3K4me3 in the parasites through each experimental condition, visualizing the binding affinity strength on the y-axis and the number of affinity sizes as the x axis of a box plot. Finally, DiffBind was used again with the peak set information from MACS2, to run a differential analysis on peaks and read depth via DESeq2 (https://bioconductor.org/packages/release/bioc/html/DESeq2.html) [24] to formulate a lists of genes found to be differentially covered by histone modifications between two or more samples. Several restrictions were placed on the gene list outputs to ensure meaningful results: a log2 fold change of absolute value greater than 1, an FDR of 0.05 or lower, and a combined concentration of 6 or greater. Then the lists were manually reviewed using IGV tracks with false positives removed from the final list.

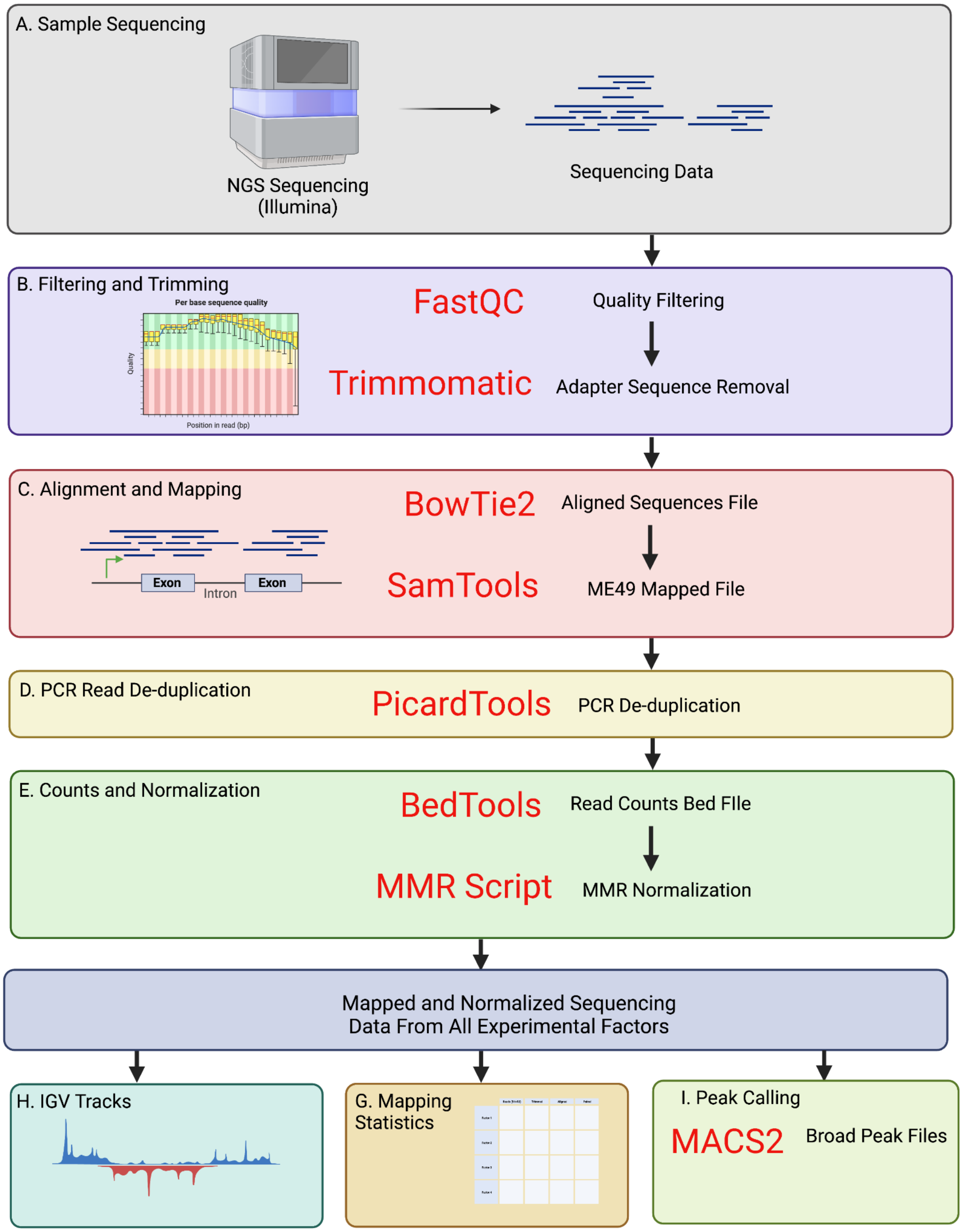


Figure 4. Visualization of our Post-Sequencing Processing Pipeline.

Chapter 1.3 Results

Profiling Histone Modifications Using ChIP-seq Experiment

To better understand the epigenetic mechanisms regulating parasite development in distinct environmental conditions and to also better understand the molecular insights into the bradyzoite to tachyzoite (BtT) or bradyzoite-to-bradyzoite (BtB) phenomenon, we used ChIP-seq experiment (Fig. 5). The goal of our experiment was to examine whether parasites infecting AST and HFF cells display differentially histone markers that can explain *T. gondii* recrudescence from its native bradyzoites at day 0 through days 3, 5, and 7 post excystation (Fig. 5). We first decided to profile three histone marks: H3K4me3, H3K14ac, and H3K9me3. H3K4me3 is a histone mark associated with transcriptional activation (euchromatin) [13, 25]. H3K9me3 is a histone modification associated with repression (heterochromatin) [12, 13, 25]. Finally, H3K14ac is an understudied histone mark but is potentially associated with poised transcription [14, 26]. We, therefore, performed ChIP-seq experiments as described in the method section in duplicate using ME49EW samples collected at day 0, 3, 5, and 7 in infected AST and HFF cells. We also used two controls, including the IgG negative control and the input. Please note that not all libraries have not yet been sequenced, and some time points are still missing for our analysis.

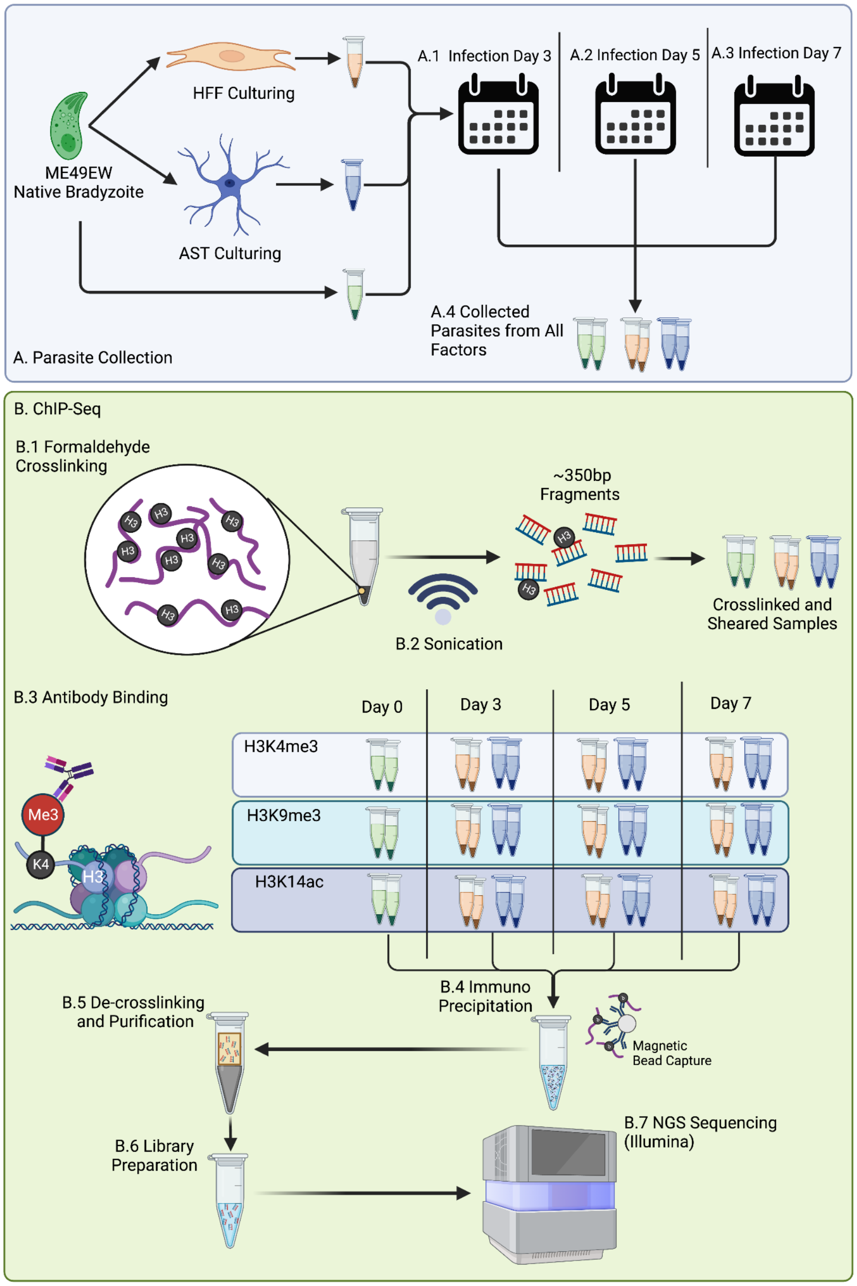


Figure 5. Experimental design for this project. (A) ME49EW native bradyzoites were collected as day 0 and cultured in AST and HFF cells. Then, parasites were collected at days 3, 5, and 7 post-excystation. (B) Summary of ChIP-seq experiments on the native bradyzoites (day 0) and the three recrudescence time points (days 3, 5, and 7) for both AST and HFF cells.

Post-Processing Read Counts

After library preparation and sequencing for each sample, reads were processed to remove low-quality reads. High-quality reads were then aligned to the ME49 genome (ToxoDB.v62) and filtered to remove PCR duplicates. The mean of all reads totaled 13,048,430, with H3K4me3 samples averaging 24,740,529.9 reads, and H3K9me3 at 12,190,750 reads. Reads obtained for H3K14ac were significantly lower with only 3,066,840 reads (Supplemental Table 1, see Appendix).

The Correlation Coefficients of Experimental Factors

After mapping our reads to the *T. gondii* genome, we then aimed to identify the significant peaks of each sample using the MACS2 peak caller [21]. In brief, MACS2 uses fragment pileups from properly-aligned paired-end data, which are then normalized by Million Mapped Reads (MMR). MACS2 calls peaks via a bias model comparing the signal of the sample to the background noise at every base pair using a Poisson test [27]. Next, DiffBind, a program for differential binding analysis of ChIP-seq peak data [22] identified sites that are differentially enriched between two or more samples and clustered the data by Pearson correlation.

For the histone H3K4me3 marks, the correlations of called peaks revealed that all the samples are closely correlated, especially between replicates and host cells (Fig. 6A). Samples collected in AST and the HFF cells formed two distinct clusters, indicating that the host cell may impact the epigenetic profile of *T. gondii.* Samples collected in AST at days 5 and 7 had a higher correlation (0.99) than samples collected at days 5 and 3 (0.88), suggesting a different epigenetic pattern in early time points.

We, unfortunately, cannot yet confirm this trend for parasites extracted in HFF cells as samples collected in day 3 have yet to be sequenced. For the histone H3K9me3 marks, samples displayed a moderate to high correlation coefficient (Fig. 6B). For these samples, a minimal number of peaks were identified surrounding the centromeres, and a weak correlation could be due to noise identified by MACS2 peak caller. Data must be further validated pending sequencing of our second replicates at days 5 and 7 in parasites extracted from AST cells. Thus, the final clustering will be adjusted once we have sequenced all biological replicates.

For the histone H3K14ac marks, DiffBind analysis detected only one consensus peak between the biological replicates. This result, the absence of clustering, and the low correlation obtained between replicates (Fig. 6C), indicate that the H3K14ac antibody may have a low affinity for parasite H3K14ac marks. These samples were therefore removed from our analysis.

Epigenetic Landscape of *T. gondii* Infecting Different Host Cells

To evaluate the global epigenetic changes to the parasite's genome in different conditions, the normalized Per Million (CPM) of mapped Chip_seq reads was processed by an in-house script to visualize the distribution of the histone modification sites by set genomic intervals (bins). The histone H3K4me3 marks were distributed across the entire genome, and profiles are relatively similar between the experimental conditions (Figure 7 and 9). This is expected as H3K4me3 is a prominent marker for actively transcribed genes. For the histone H3K9me3 marks, we confirmed their association with the centromeres of all *T. gondii* chromosomes (Figs. 8 and 9A). No H3K9me3 centromeric signals were detected in chromosome VIIb. However, two recent publications have highlighted the miss assembly of the *T. gondii* genome with VIIa and VIIb forming a singular large chromosome [28, 29]. DiffBind analysis confirmed that no significant differences could be detected between cell lines and time points, validating its main role in centromere function. However, a closer inspection of our profiles revealed a moderate signal on AST day 5 and HFF day 7. Data will be needed to be further confirmed with an analysis of additional biological replicates. Considering the results obtained above, we focused our analysis using ChiP-seq profiles obtained against the histone H3K4me3 marks.

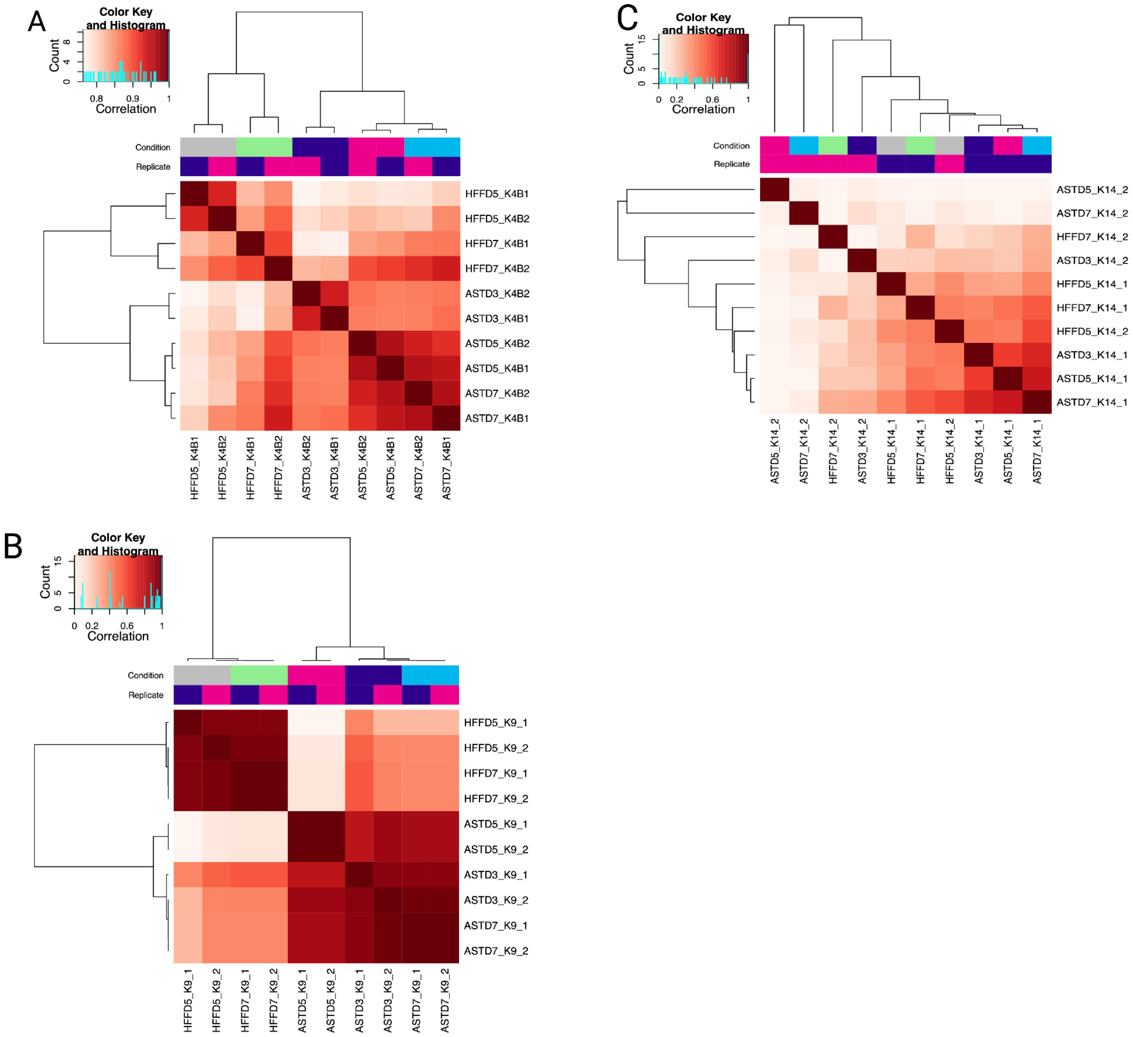


Figure 6. Pearson correlations between the differentially expressed peaks of all samples. The correlations between the called peaks of days 3, 5, and 7 in AST and HFF cell lines are represented as heatmaps for H3K4me3 (A), H3K9me3 (B), and H3K14ac (C) modifications. The samples are annotated as follows: AST = Astrocytes, HFF = Human Foreskin Fibroblasts, D3 = day 3, D5 = day 5 and D7 = day 7, histone modification K4 = H3K4me3, K9 = H3K9me3, K14 = H3K14ac.

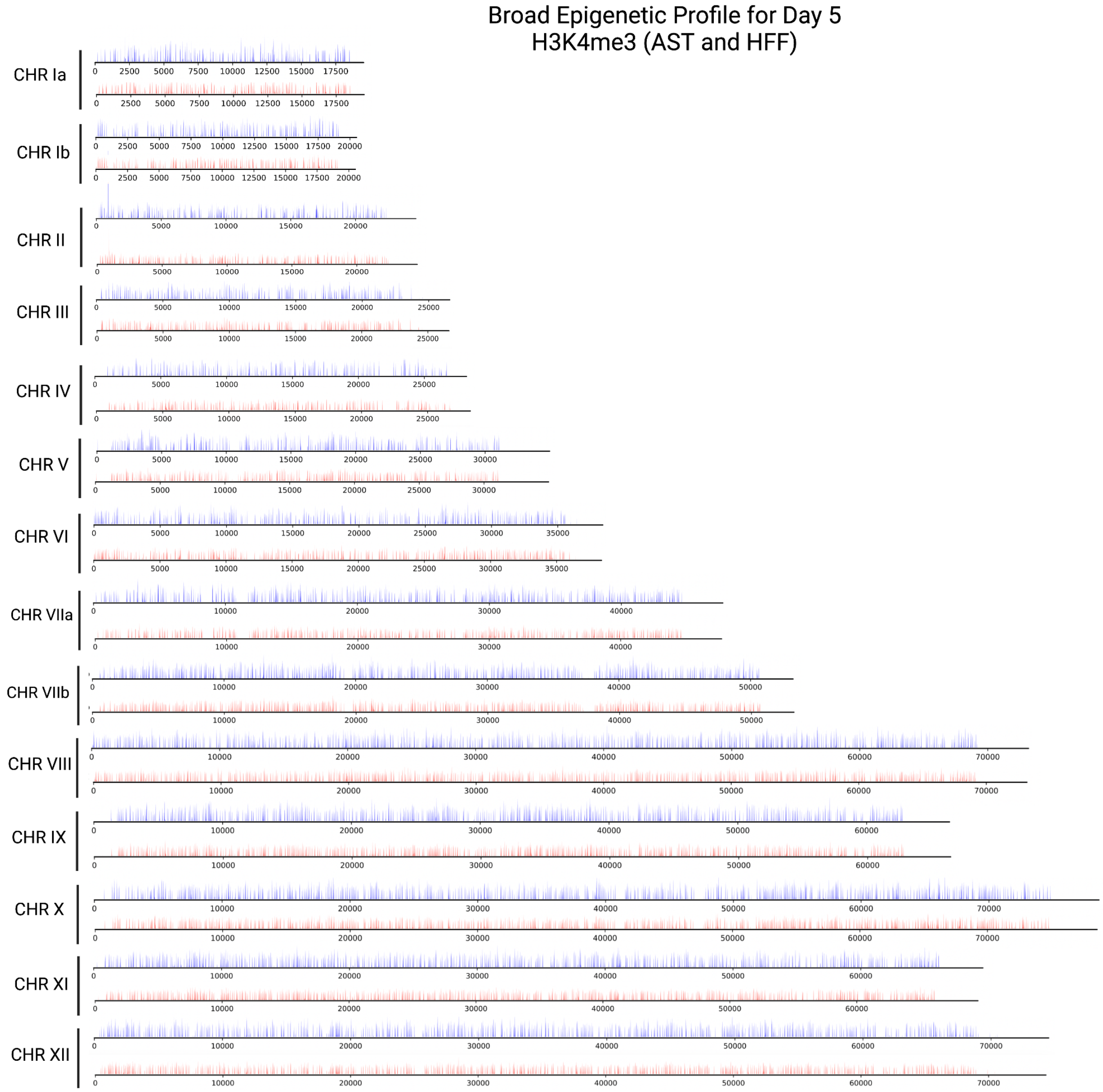


Figure 7. ChIP-seq analysis showing genome-wide distribution of H3K4me3 in different cell lines. H3K4me3 distribution across all *T. gondii* chromosomes for AST and HFF cell lines at day 5 of recrudescence. The x-axis represents chromosome size in kilobases. The samples are coordinated by color: day 5 AST = blue and day 5 HFF = red.

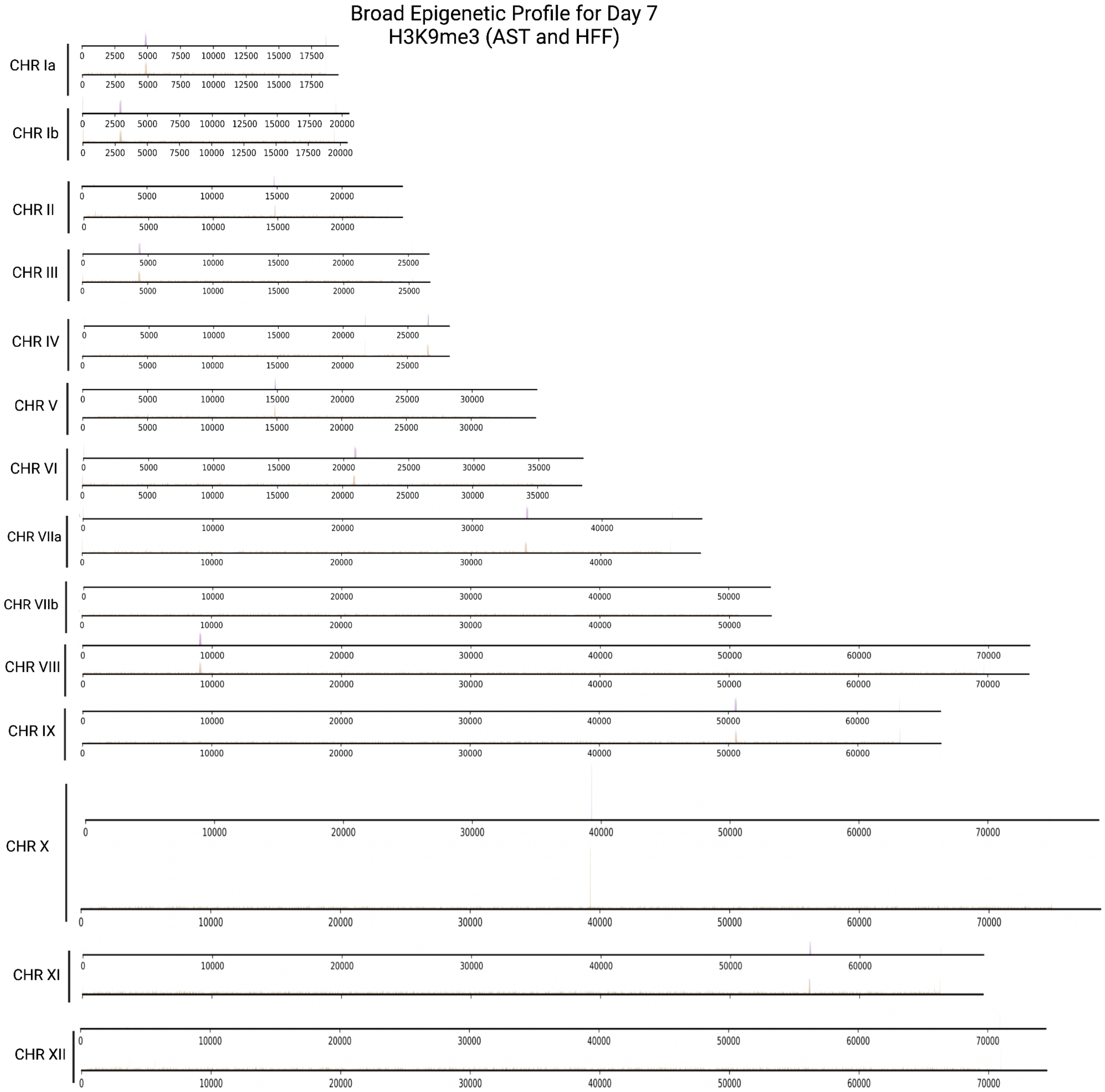


Figure 8. ChIP-seq analysis showing the genome-wide distribution of H3K9me3 in different cell lines. H3K9me3 distribution across all *T. gondii* chromosomes for AST cell lines at day 7 of recrudescence. The x-axis represents chromosome size in kilobases. The samples are coordinated by color: day 7 AST = purple and day 5 HFF = gold.

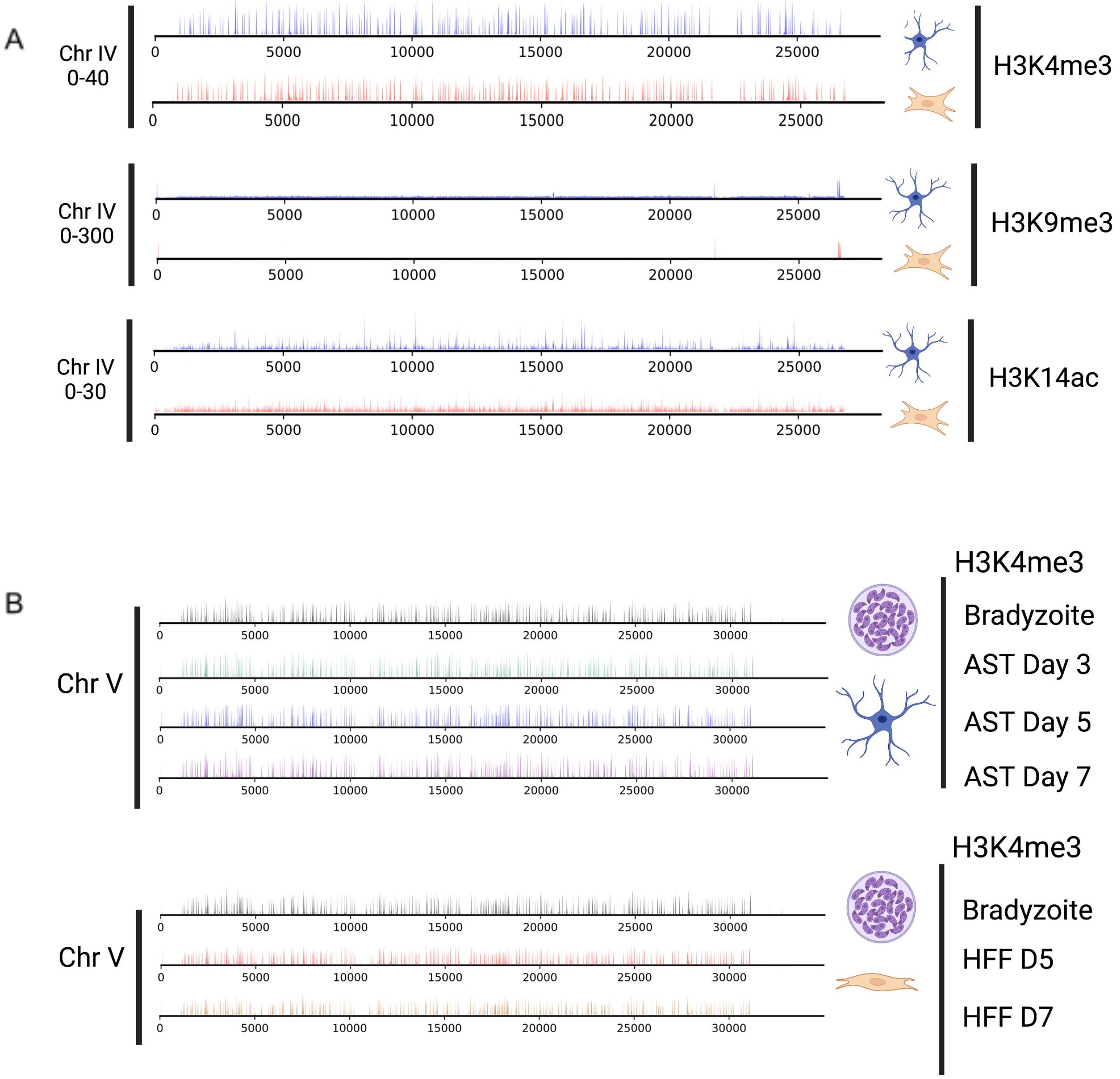


Figure 9. ChIP-seq analysis showing chromosome distribution of H3K4me3, H3K9me3 and H3K14ac in different cell lines. (A) Distribution of H3K4me3, H3K9me3, and H3K14ac on chromosome IV for the AST and HFF cell lines at day 5. The figure is coordinated by color: day 5 AST = blue and day 5 HFF = red. (B) Distribution of H3K4me3 marks on chromosome V for the AST and HFF samples at the different time points. The figure is annotated with colors for host cell and time points: native bradyzoite = black, day 3 AST = teal, day 5 AST = blue, day 7 AST = purple, day 5 HFF = red and day 7 HFF = gold.

Differential Analysis of Histone H3K4me3 Marks Between Host Cells and Time

As described above, H3K4me3 is a chromatin modification known to mark the transcription start sites of active genes. To identify differences in the distribution of H3K4me3 marks in and around the TSS of genes, we binned our normalized reads into 3 parts, 1kb upstream of the TSS, the TSS itself, and 2kb downstream of the TSS for every gene in the genome. These bins were then layered to visualize the relative abundance of the histone mark around the TSS and plotted as heatmaps.

In parasites cultured in ASTs, we observed a low occupancy of H3K4me3 marks around the TSS of the native bradyzoites. This result indicates a certain dormancy of the bradyzoite stage. In parasites extracted at days 3, 5, and 7, an increased level of H3K4me3 marks is detected across all chromosomes around the TSS of most genes (Figure 10A). The line plots further supported this data (Fig. 10B). We also detected that the parasites have a nearly 5-fold greater abundance of H3K4me3 at days 3, 5, and 7 compared to the bradyzoites (Fig. 10B), demonstrating differentiation into tachyzoite-like parasites which necessitates parasite growth and active gene expression.

The heatmap of the H3K4me3 marks for parasite infecting HFF cells followed a similar pattern, where H3K4me3 marks were weaker in the TSS of the bradyzoites compared to recrudescing parasites at days 3, 5 and 7 (Fig. 11A). This was also supported by a line plot (Fig. 11B) which also confirmed that parasites cultured in HFFs at days 3, 5 and 7 have an approximately 3-fold higher abundance of H3K4me3 marks compared to the bradyzoite sample.

To compare profiles between parasites cultured in ASTs and HFFs, we compared the line plots obtained with samples from both host cells. Our results revealed that the H3K4me3 marks surrounding the TSS were 1.5x higher in parasite cultures in ASTs compared to the one cultured in HFFs (Fig. 12), indicating more active parasite replications in ASTs.

To further confirm the difference in binding affinity for H3K4me3 marks between parasites cultured in AST and HFF, we performed an additional binding affinity analysis using DiffBind. Results shown in figure 13 confirm that H3K4me3 marks within the *T. gondii* genome have a higher affinity in AST compared to HFF (Fig. 13). This was true for parasites extracted on day 5 (Fig. 13A) and day 7 (Fig. 13B). The pattern was detailed in the figure, representing affinity sites that favored the HFF (+) or AST samples (-), respectively. The width of the boxes in the (+) section clearly showed fewer affinity sites than in the (-) AST section. All together our results suggested that parasites cultured in HFF cells are most likely stressed indicating a model that is most likely not biologically relevant.

To investigate at the gene level changes in histone marks across time points and conditions, we used the normalized peakset data and DiffBind to extract specific gene loci. A Python script trawled through the ME49 genome assembly and extracted the gene’s loci.

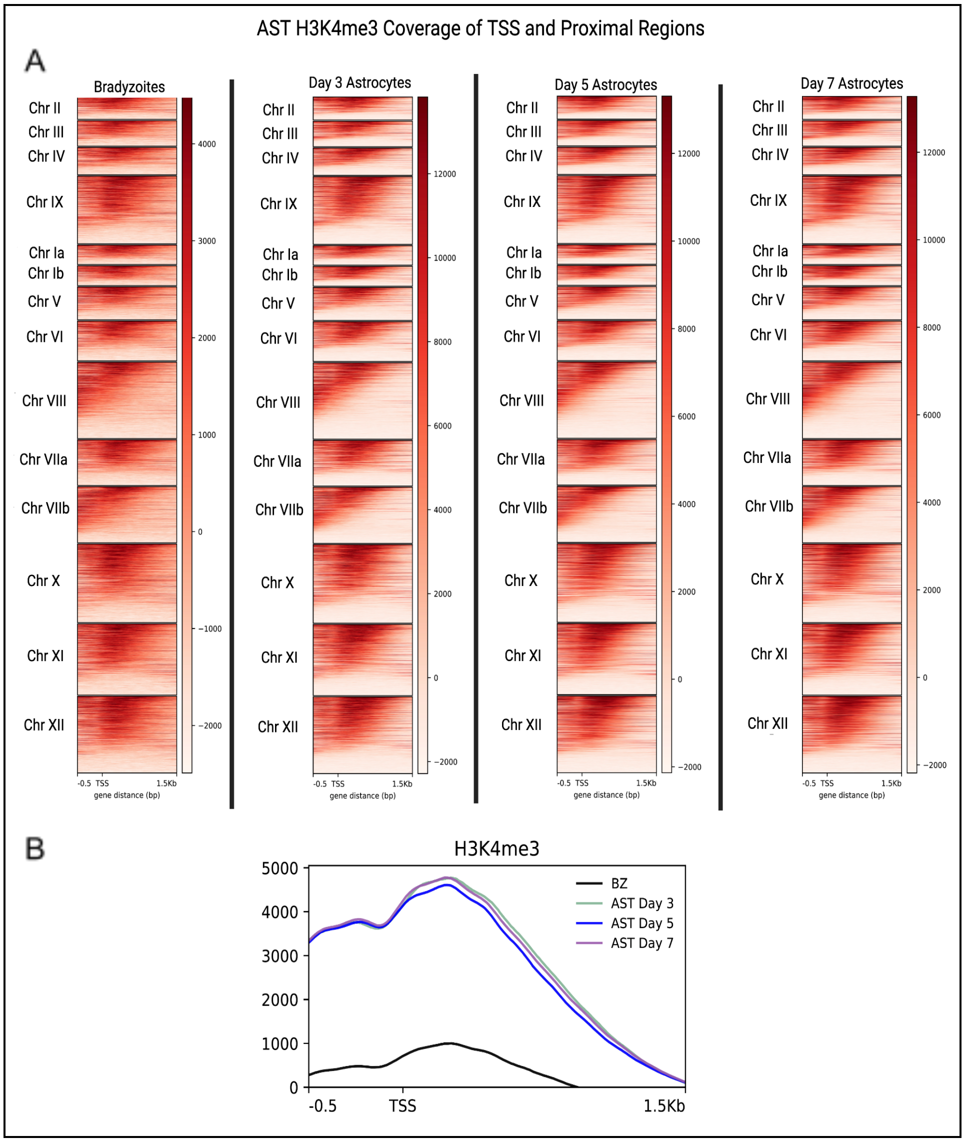


Figure 10. Distribution of H3K4m3 in native bradyzoites and parasites cultured in the AST cell line. (A) Heatmaps showing the distribution of H3K4me3 marks from 0.5kb upstream to 1.5kb downstream of the TSS for each chromosome. (B) Graph showing the global distribution of H3K4me3 marks for bradyzoites (BZ) and AST days 3, 5, and 7 samples.

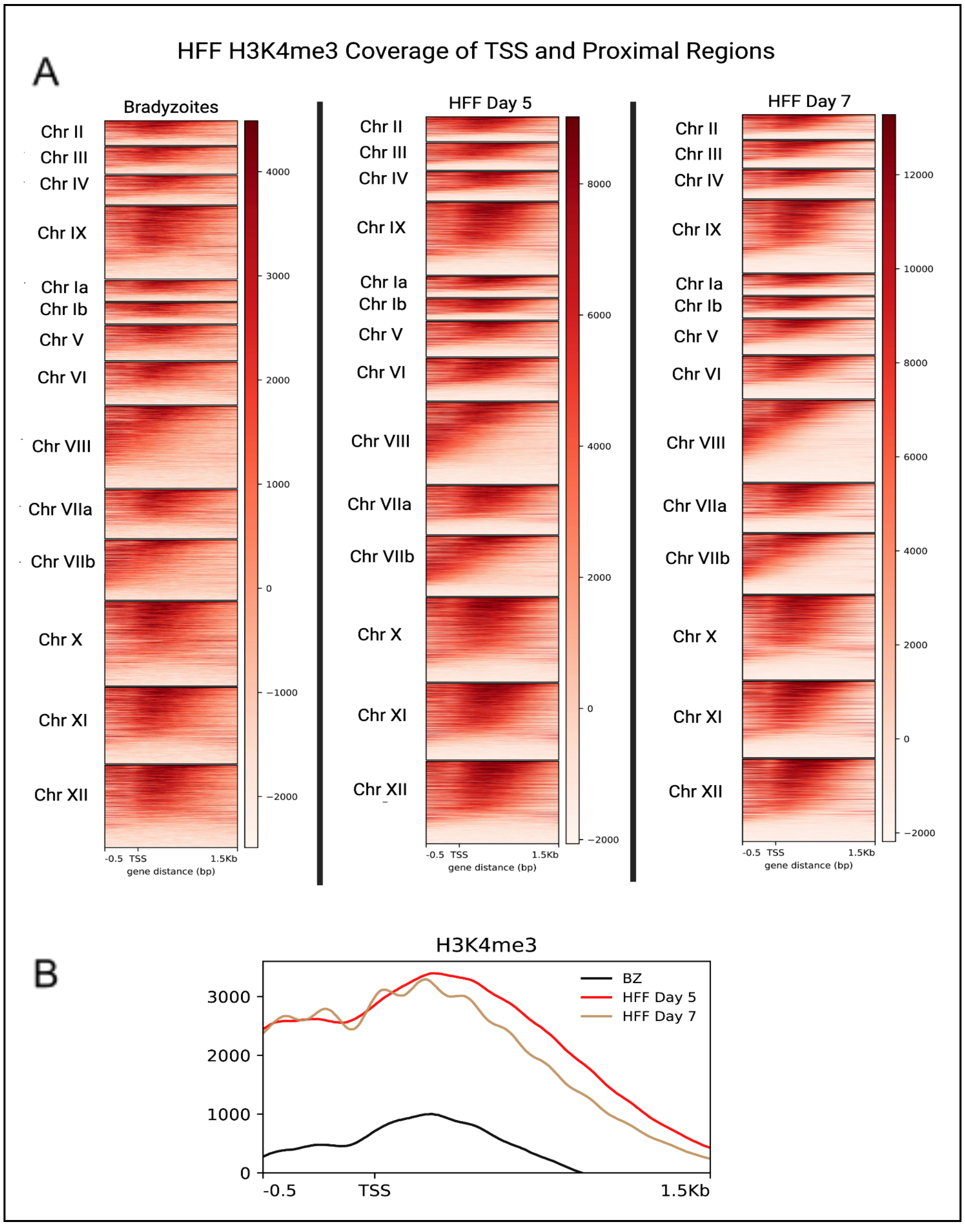


Figure 11. Distribution of H3K4m3 in native bradyzoites and parasites cultured in the HFF cell line. (A) Heatmaps showing the distribution of H3K4me3 from 0.5kb upstream to 1.5kb downstream of the TSS for each chromosome. (B) Graph showing the global distribution of H3K4me3 marks for bradyzoites (BZ) and HFF days 5 and 7 samples.

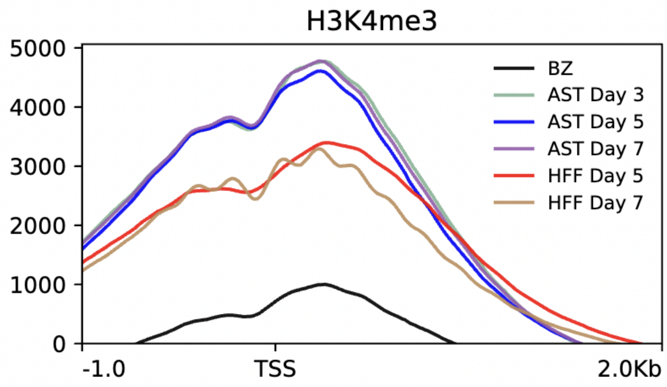


Figure 12. Combined line plot showing the distribution of H3K4me3. Normalized reads were binned from 1kb upstream to 2kb downstream of the TSS for native bradyzoites (BZ), AST days 3, 5, and 7, and HFF days 5 and 7 samples.

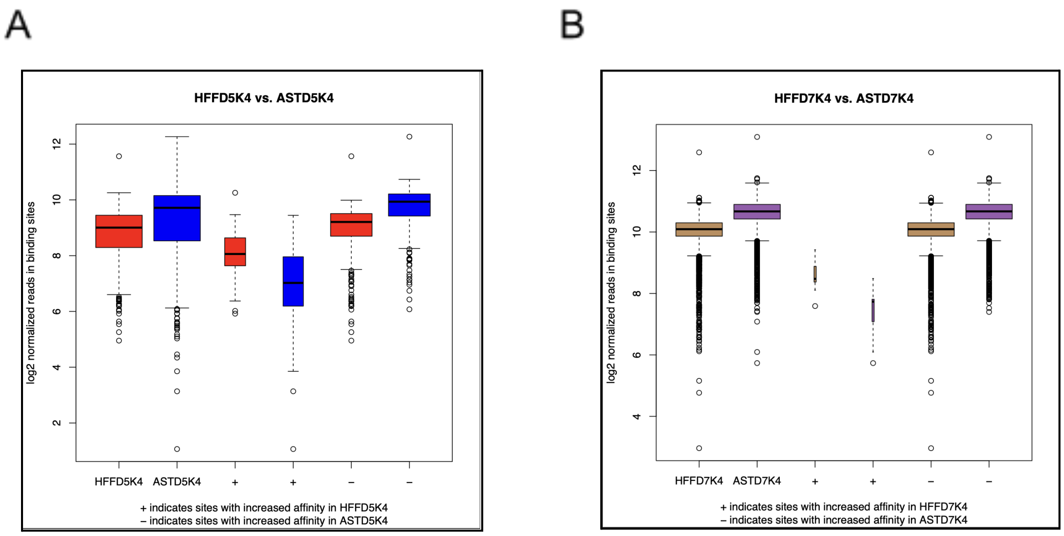


Figure 13. The binding affinity of H3K4me3 in parasites cultured in AST and HFF cells at day 5 (A) and Day 7 (B). DiffBind processed normalized reads, identifying affinity binding sites by their relative read densities over a predetermined genomic interval, making them independent of peak set information. Samples collected in ASTs have a higher binding affinity when compared to parasites cultured in HFFs. DiffBind visualizes the number of affinity binding sites through the width of the boxes in the plot. Day 5 HFFs have fewer affinity binding sites, as seen in the days 5 and 7 (+) columns, while the majority of genes saw a higher mean binding affinity for days 5 and 7 AST as seen in the (-) columns.

Differential H3K4me3 signal in AST and HFF Hosted Parasites at Day 5 and 7

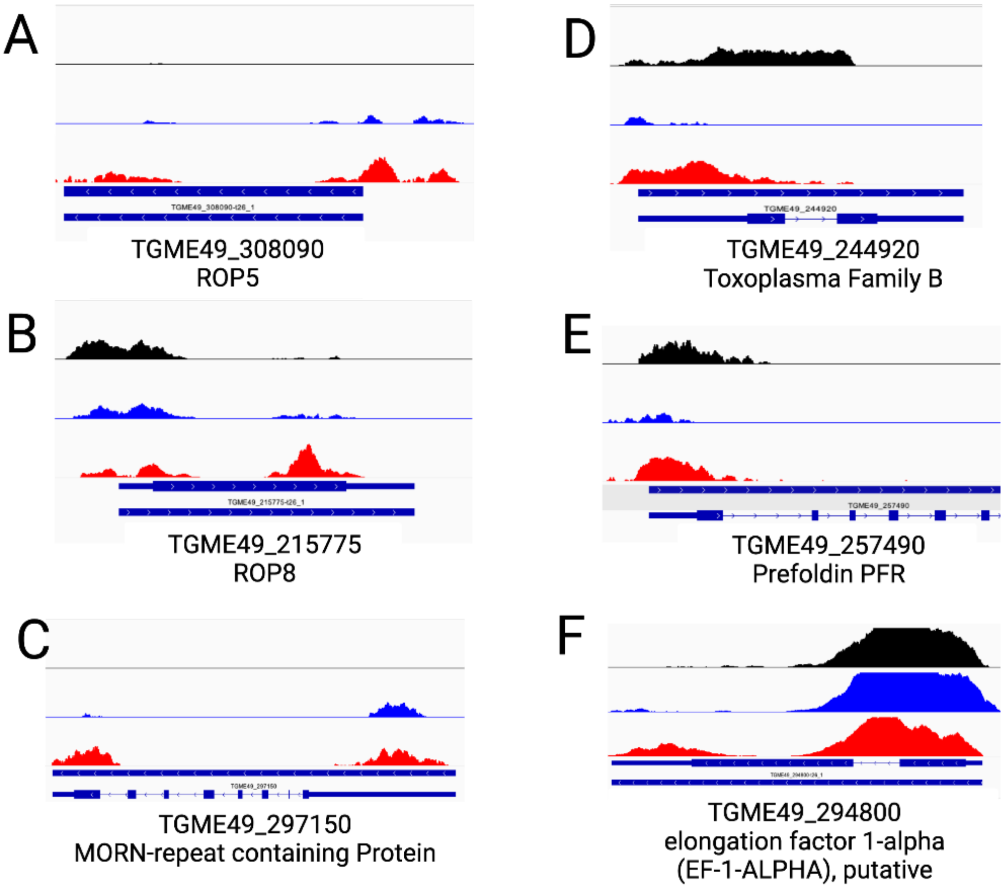
27 peaks were identified exhibiting significant differences in peak level at day 5. Three of these peaks were called in intergenic regions and could potentially regulate novel genes or non-coding genes. 24 were identified in the promoter region of annotated coding genes (Table 1). Of the 24 genes, 12 were described as hypothetical proteins. The 12 hypothetical genes identified in our initial list could potentially guide future experiments to understand their potential impact in parasite development and adaptation.

Of the 12 genes with functional annotation, six had a higher level of H3K4me3 marks in parasites cultured in HFF cells, including two invasion-related proteins (ROP5 and ROP8) [30-32], two RNA transportation and processing proteins (Prefoldin and EF-1-Alpha) [33-35], and one predicted histone-modifying enzyme (MORN-repeat containing protein). These genes were individually visualized using the IGV tracks (Fig. 14). For the six genes that exhibited higher levels of H3K4me3 marks in parasites cultured in ASTs, we found a histone variant (H2A1) [36], a cyclin Dependant kinase (CDK), an invasion gene (SAG1), two surface antigens (SAG1 and SRS17A) [37, 38] and one RNA binding protein. Interestingly, an inspection of the IGV track revealed a peak in one of the introns of the CDK, suggesting a possible splice site variant or activation of novel coding or non-coding gene (Figure 15).

Table 1: List of genes with differential H3K4me3 signal in parasites cultured in AST and HFF cell lines at day 5.

Seqnames	Called	Fold	p.value	FDR	Gene	Short_Name	Description
TGME49_chrXII	HFF	10.31503501	4.29E-13	2.80E-10	TGME49_308090	ROP5	rhoptry protein ROP5
TGME49_chrII	HFF	2.825222054	3.95E-21	1.42E-17	TGME49_297150		MORN repeat-containing protein
TGME49_chrVI	HFF	1.753573647	2.67E-11	1.27E-08	TGME49_244920		*Toxoplasma gondii* family B protein
TGME49_chrIV	HFF	1.559414377	5.30E-06	0.0004550651392	intergenic		
TGME49_chrVIII	HFF	1.538817737	2.61E-06	0.0002676769233	intergenic		
TGME49_chrX	HFF	1.37841341	1.24E-09	3.83E-07	TGME49_215775	ROP8	rhoptry protein ROP8
TGME49_chrX	HFF	1.169106242	1.16E-06	0.0001427030919	TGME49_237450		hypothetical protein
TGME49_chrX	HFF	1.150139488	0.0001322896827	0.004587210455	TGME49_214900		hypothetical protein
TGME49_chrIa	HFF	1.147972092	4.97E-06	0.0004335089918	TGME49_294800		"elongation factor 1-alpha (EF-1-ALPHA), putative"
TGME49_chrVIIb	HFF	1.113448627	3.18E-08	6.40E-06	TGME49_257490		prefoldin subunit superfamily protein
TGME49_chrIX	AST	-0.9996668707	6.61E-09	1.92E-06	TGME49_265320		hypothetical protein
TGME49_chrV	AST	-1.071467718	1.04E-05	0.0007402300876	TGME49_213930		"3' exoribonuclease family, domain 1 domain-containing protein"
TGME49_chrVIIb	AST	-1.159595048	1.04E-10	4.55E-08	TGME49_261250		histone H2A1
TGME49_chrIb	AST	-1.175595921	2.47E-10	9.22E-08	TGME49_321360		clustered-asparagine-rich protein
TGME49_chrIX	AST	-1.214169773	1.82E-08	4.45E-06	TGME49_264630		hypothetical protein
TGME49_chrVIII	AST	-1.306423769	3.24E-10	1.13E-07	TGME49_233460	SAG1	SAG-related sequence SRS29B
TGME49_chrVIIa	AST	-1.38886466	1.29E-12	6.74E-10	TGME49_203300		hypothetical protein
TGME49_chrVIII	AST	-1.574312568	1.95E-13	1.70E-10	TGME49_268985		hypothetical protein
TGME49_chrIX	AST	-1.851795891	5.99E-13	3.48E-10	TGME49_291170		hypothetical protein
TGME49_chrX	AST	-1.869781943	1.36E-10	5.50E-08	intergenic		
TGME49_chrXII	AST	-1.987562723	1.48E-06	0.0001726362902	TGME49_245770		hypothetical protein

TGME49_chrXII	AST	-2.031067177	5.97E-14	6.25E-11	TGME49_277230		hypothetical protein
TGME49_chrIX	AST	-2.04758607	5.12E-15	8.94E-12	TGME49_290310		hypothetical protein
TGME49_chrIV	AST	-2.064551594	2.42E-13	1.81E-10	TGME49_319360	SRS17A	SAG-related sequence SRS17A
TGME49_chrVI	AST	-2.280199668	5.43E-21	1.42E-17	TGME49_239910		cyclin-dependent kinase
TGME49_chrV	AST	-9.330481312	4.84E-10	1.58E-07	TGME49_287250		hypothetical protein
TGME49_chrXII	AST	-10.82302061	1.87E-14	2.45E-11	TGME49_245980		hypothetical protein



A: TGME49_308090
ROP5

B: TGME49_215775
ROP8

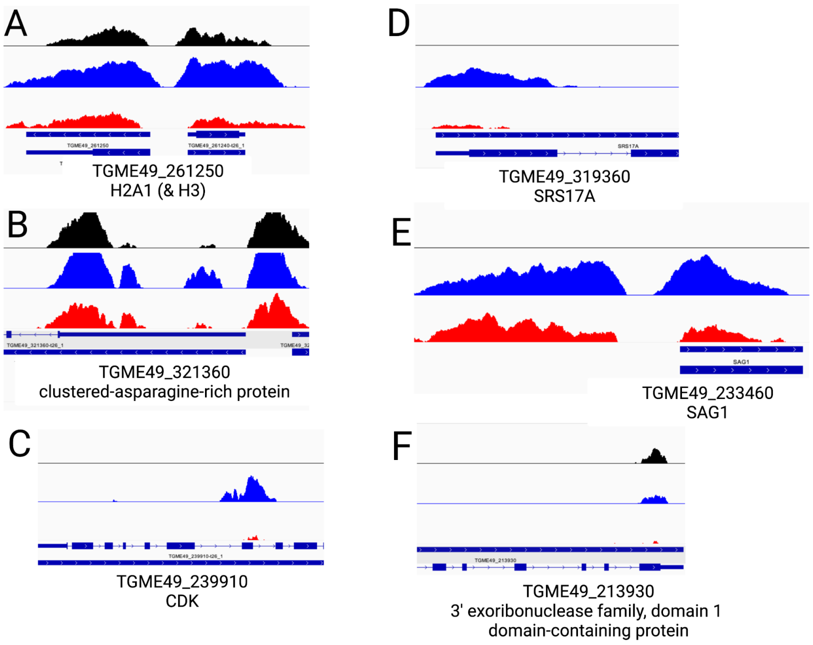
C: TGME49_297150
MORN-repeat containing Protein

D: TGME49_244920
Toxoplasma Family B

E: TGME49_257490
Prefoldin PFR

F: TGME49_294800
elongation factor 1-alpha
(EF-1-ALPHA), putative

Figure 14. Individual IGV tracks highlighting genes with distinctly higher H3K4me3 profiles in HFF parasites at day 5. IGV tracks show genes called between the AST (blue), HFF (red), and native bradyzoite (black). All parts of this figure show genes from the day 5 list with a higher abundance of H3K4me3 on the HFF sample.



TGME49_261250
H2A1 (& H3)

TGME49_321360
clustered-asparagine-rich protein

TGME49_239910
CDK

TGME49_319360
SRS17A

TGME49_233460
SAG1

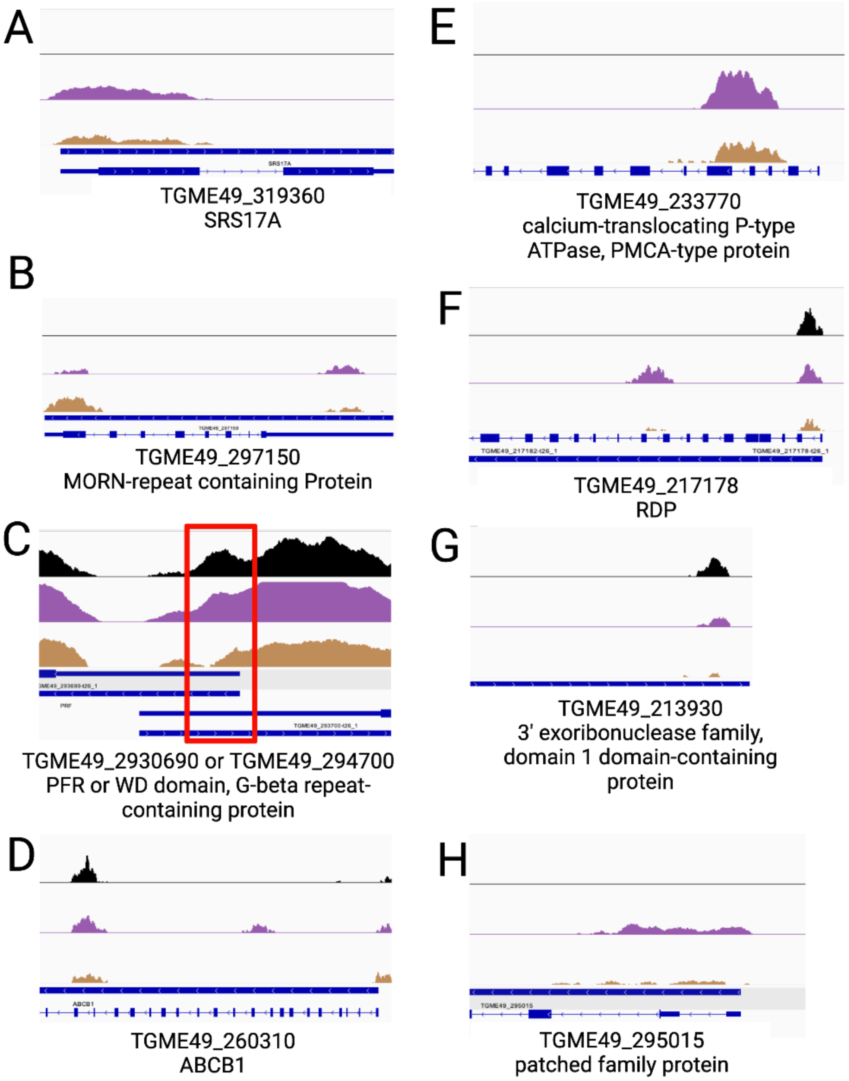
TGME49_213930
3' exoribonuclease family, domain 1
domain-containing protein

Figure 15. Individual IGV tracks highlighting genes with distinctly higher H3K4me3 profiles in AST parasites at day 5. IGV tracks show genes called between the AST (blue), HFF (red), and native bradyzoite (black). All parts of this figure show genes from the day 5 list with a higher abundance of H3K4me3 on the AST sample.

At day 7, 17 peaks were found to exhibit significant differential binding of peaks. Only 1 of these peaks was called in an intergenic region, and potentially relevant in the regulation of other non-coding genes or novel genes. 16 were identified in the promoter region of annotated coding genes (Table 2). Of these 16 genes, 6 were identified as hypothetical proteins. Of the 10 genes with functional annotations, all had a higher H3K4me3 mark in the parasites cultured in the AST cells, including an invasion related protein (Profilin PFR) [39], two active transporter proteins (ABCB1 and calcium-translocating P-type ATPase) [40, 41], two RNA processing proteins (RDP and 3' exoribonuclease family, domain 1 domain-containing protein) [42], and one predicted histone-modifying enzyme (MORN-repeat containing protein). These genes were individually visualized using the IGV tracks (Figure 16).

Table 2. List of genes with differential H3K4me3 signal in parasites cultured in AST and HFF cell lines at day 7.

Seqnames	Called	Fold	p.value	FDR	Gene	Short_ Name	Description
TGME49_chrIX	HFF	2.47924127	7.39E-30	1.95E-26	TGME49_290310		hypothetical protein
TGME49_chrXI	AST	-0.985830672	0.0002146247024	0.000275244292	TGME49_217182		RNA-dependent RNA polymerase RDP
TGME49_chrIa	AST	-1.002867517	2.39E-05	3.53E-05	intergenic		
TGME49_chrIa	AST	-1.007490414	5.87E-15	3.68E-13	TGME49_293690	PRF	profilin PRF
TGME49_chrIV	AST	-1.073593146	7.98E-11	8.63E-10	TGME49_319360	SRS17A	SAG-related sequence SRS17A
TGME49_chrVI	AST	-1.077458624	9.47E-12	1.52E-10	TGME49_239910		cyclin-dependent kinase
TGME49_chrIa	AST	-1.113299596	1.04E-21	5.47E-19	TGME49_293650		hypothetical protein
TGME49_chrXII	AST	-1.148158566	1.66E-15	1.31E-13	TGME49_250910		hypothetical protein
TGME49_chrVI	AST	-1.158658053	7.62E-06	1.25E-05	TGME49_243382		hypothetical protein
TGME49_chrVIIb	AST	-1.168785236	2.73E-13	8.46E-12	TGME49_260310	ABCB1	ATP-binding cassette transporter ABC.B1
TGME49_chrVIII	AST	-1.198069219	4.66E-19	1.02E-16	TGME49_233770		calcium-translocating P-type ATPase, PMCA-type protein
TGME49_chrVIIa	AST	-1.234740167	2.46E-06	4.57E-06	TGME49_205050		hypothetical protein
TGME49_chrXII	AST	-1.247501145	6.97E-16	6.33E-14	TGME49_245980		hypothetical protein
TGME49_chrII	AST	-1.248622749	3.05E-09	1.63E-08	TGME49_297150		MORN repeat-containing protein
TGME49_chrIa	AST	-1.330517484	4.60E-12	8.51E-11	TGME49_295015		patched family protein
TGME49_chrV	AST	-1.786061883	2.33E-16	2.61E-14	TGME49_213930		"3' exoribonuclease family, domain 1 domain-containing protein"



TGME49_319360
SRS17A

TGME49_233770
calcium-translocating P-type ATPase, PMCA-type protein

TGME49_297150
MORN-repeat containing Protein

TGME49_217178
RDP

TGME49_2930690 or TGME49_294700
PFR or WD domain, G-beta repeat-containing protein

TGME49_213930
3' exoribonuclease family, domain 1 domain-containing protein

TGME49_260310
ABCB1

TGME49_295015
patched family protein

Figure 16. Individual IGV tracks highlighting genes with distinctly higher H3K4me3 profiles in AST parasites at day 7. IGV tracks show genes called between the AST (purple), HFF (brown), and native bradyzoite (black). All parts of this figure show genes from the day 7 list with a higher abundance of H3K4me3 on the AST sample.

Chapter 1.4 Discussion and Conclusion

To investigate at the mechanistic level how *T. gondii* could adapt to two different cell lines during its development; we analyzed changes in epigenetic profiles via ChiP-seq experiment against three histone marks throughout parasite infection. Experiments done using Histone H3K4me3 and Histone H3K9me3 marks were highly reproducible. Unfortunately, low specificity of our H3K4ac antibody prevented us from extracting any meaningful information. H3K9me3 signals across the genome confirmed the presence of peaks at the centromeres of all true chromosomes (Fig. 8, see appendix for more) and a few telomeric ends. The absence of the signal in chromosome VIIb is in agreement with the corrected genome assembly of *T. gondii* and the fact that chromosomes VIIa and VIIb form a single large chromosome. Investigation of histone H3K4me3 coverage confirmed strong localization of this active mark near TSS of active genes. There was a significant disparity in the abundance of H3K4me3 at different time points analyzed and between parasites cultured in either AST or HFF cells. We detected approximately 3-fold higher abundance in parasites extracted at days 5 and 7 compared to the bradyzoite sample indicating a significant bradyzoite to tachyzoite transition. We also detected approximately 1.5 times greater signal in the promoters of active genes when parasites were cultured in AST compared to HFF cells. This has most likely a downstream effect on the transcriptional behavior of the parasite and will need to be confirmed using RNA-seq analysis.

Overall, this global decrease in H3K4me3 signal in HFFs could suggest that the parasite is struggling to develop properly within this host cell. This is not entirely surprising as the preferred parasite host cells are in the brain and, to a lesser extent, muscular or visceral tissue. It is highly possible that ME49 parasites are not well adapted to HFF cells. Thus, despite their long history as a common lab model for *T. gondii*, the HFFs may impose biologically irrelevant changes on the development and propagation of the parasites.

Using DiffBind, we also extracted a list of genes or regions of the genome which experienced differential binding of H3K4me3 between host cells and parasite development. Overall, parasites cultured in HFF cells had increased signal in invasion genes (ROP5 and ROP8). An increased level of H3K4me3 was also observed in the promoter of histone methyl-transferase (MORN-repeat-containing protein) indicating a potential compensatory mechanism for the challenge of parasite growth in HFF cell line. In AST cells, two high H3K4me3 peaks were identified near the histone H2A1 and a CDK. The H2A1 histone is a variant of the H2A core histone protein and is believed to be the canonical H2A histone critical for cell division and S phase in early bradyzoites but not in mature bradyzoites or tachyzoites [35]. This has exciting implications as it serves as preliminary evidence for possible large-scale chromatin remodeling experienced in parasites cultured in AST cells and the potential of rapid re-differentiation of the parasite back into bradyzoite-like parasites. The CDK encoded by TGME49_239910 may be an important protein regulating cell cycle progression known to be upregulated in bradyzoites. Combined with the identification of a peak in the intronic region, CDK might also be subjected to splice variation that may affect the mechanism regulating the recrudescence phenomenon observed previously.

At day 7, the gene list includes SAG1, a known tachyzoite biomarker, a calcium-translocating P-type ATPase and patched family protein, two genes controlling signaling pathways and cellular proliferation. Higher signal of H3K4me3 marks in the promoter of both of these genes may suggest their role in distinct replication behavior and development observed in parasites cultured in ASTs.

Gene lists made between time points from the same host cell can be found in the appendix section as: Supplemental Table 2 (days 3 and 5 AST), Supplemental Table 3 (days 5 and 7), Supplemental Table 4 (HFF days 5 and 7)

Overall, while epigenetics profiling identifies potential key regulators of parasite development in two distinct host cells, our data will need to be further validated using alternative methods, including transcriptional profiling and genome editing.

Chapter 1.6 References

1. Flegr, Jaroslav et al. “Toxoplasmosis--a global threat. Correlation of latent toxoplasmosis with specific disease burden in a set of 88 countries.” PloS one vol. 9,3 e90203. (2014), doi:10.1371/journal.pone.0090203

2. Tenter, A M et al. “Toxoplasma gondii: from animals to humans.” International journal for parasitology vol. 30,12-13 (2000): 1217-58. doi:10.1016/s0020-7519(00)00124-7

3. Weiss, Louis M, and Jitender P Dubey. “Toxoplasmosis: A history of clinical observations.” International journal for parasitology vol. 39,8 (2009): 895-901. doi:10.1016/j.ijpara.2009.02.004

4. Sanchez, Syrian G, and Sébastien Besteiro. “The pathogenicity and virulence of Toxoplasma gondii.” Virulence vol. 12,1 (2021): 3095-3114. doi:10.1080/21505594.2021.2012346

5. Halonen, Sandra K, and Louis M Weiss. “Toxoplasmosis.” Handbook of clinical neurology vol. 114 (2013): 125-45. doi:10.1016/B978-0-444-53490-3.00008-X

6. Cerutti, Aude et al. “The bradyzoite: A Key Developmental Stage for the Persistence and Pathogenesis of Toxoplasmosis.” Pathogens (Basel, Switzerland) vol. 9,3 234. (2020), doi:10.3390/pathogens9030234

7. Dubey, J P et al. “Structures of Toxoplasma gondii tachyzoites, bradyzoites, and sporozoites and biology and development of tissue cysts.” Clinical microbiology reviews vol. 11,2 (1998): 267-99. doi:10.1128/CMR.11.2.267

8. Silva, Rodrigo C da et al. “Recrudescence of Toxoplasma gondii infection in chronically infected rats (Rattus norvegicus).” Experimental parasitology vol. 125,4 (2010): 409-12. doi:10.1016/j.exppara.2010.04.003

9. An ex vivo model of Toxoplasma recrudescence Amber L. Goerner, Edward A. Vizcarra, David D. Hong, Kristina V. Bergersen, Carmelo A. Alvarez, Michael A. Talavera, Emma H. Wilson, Michael W. White. bioRxiv 2020.05.18.101931; doi: https://doi.org/10.1101/2020.05.18.101931

10. Weinhold, Bob. “Epigenetics: the science of change.” Environmental health perspectives vol. 114,3 (2006): A160-7. doi:10.1289/ehp.114-a160

11. Kim, Kami. "The Epigenome, Cell Cycle, and Development in Toxoplasma." Annual Review of Microbiology, 2018, https://doi.org/10.1146/annurev-micro-090817-062741. Accessed 13 Apr. 2023.

12. Bougdour, A., Braun, L., Cannella, D. and Hakimi, M.-A., Chromatin modifications: implications in the regulation of gene expression in Toxoplasma gondii. Cellular Microbiology, 12: 413-423. (2010). https://doi.org/10.1111/j.1462-5822.2010.01446.x

13. Gissot M, Kelly KA, Ajioka JW, Greally JM, Kim K. Epigenomic modifications predict active promoters and gene structure in Toxoplasma gondii. PLoS Pathog. ;3(6):e77. (2007) doi: 10.1371/journal.ppat.0030077. PMID: 17559302; PMCID: PMC1891328.

14. Nardelli, Sheila C et al. “The histone code of Toxoplasma gondii comprises conserved and unique posttranslational modifications.” mBio vol. 4,6 e00922-13. (2013), doi:10.1128/mBio.00922-13

15. Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

16. Bolger, Anthony M et al. “Trimmomatic: a flexible trimmer for Illumina sequence data.” *Bioinformatics (Oxford, England)* vol. 30,15 (2014): 2114-20. doi:10.1093/bioinformatics/btu170

17. Langmead, Ben, and Steven L Salzberg. “Fast gapped-read alignment with Bowtie 2.” *Nature methods* vol. 9,4 357-9. 4 Mar. 2012, doi:10.1038/nmeth.1923

18. Heng Li and others, The Sequence Alignment/Map format and SAMtools, *Bioinformatics*, Volume 25, Issue 16, August 2009, Pages 2078–2079, https://doi.org/10.1093/bioinformatics/btp352

19. PicatedTools. http://broadinstitute.github.io/picard

20. Aaron R. Quinlan , Ira M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features, *Bioinformatics*, Volume 26, Issue 6, March 2010, Pages 841–842, https://doi.org/10.1093/bioinformatics/btq033

21. Zhang, Y., Liu, T., Meyer, C.A. et al. Model-based Analysis of ChIP-Seq (MACS). Genome Biol 9, R137 (2008). https://doi.org/10.1186/gb-2008-9-9-r137

22. Wu, Dai-Ying et al. “Identifying differential transcription factor binding in ChIP-seq.” *Frontiers in genetics* vol. 6 169. 29 Apr. 2015, doi:10.3389/fgene.2015.00169

23. Ramírez, Fidel, Devon P. Ryan, Björn Grüning, Vivek Bhardwaj, Fabian Kilpert, Andreas S. Richter, Steffen Heyne, Friederike Dündar, and Thomas Manke. deepTools2: A next Generation Web Server for Deep-Sequencing Data Analysis. Nucleic Acids Research (2016). doi:10.1093/nar/gkw257.

24. Love MI, Huber W, Anders S (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." *Genome Biology*, **15**, 550. doi:10.1186/s13059-014-0550-8.

25. Brooks CF, Francia ME, Gissot M, Croken MM, Kim K, Striepen B. Toxoplasma gondii sequesters centromeres to a specific nuclear region throughout the cell cycle. *Proc Natl Acad Sci U S A.* 2011;**108**(9):3767–3772.

26. Karmodiya, K., Krebs, A.R., Oulad-Abdelghani, M. et al. H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells. BMC Genomics 13, 424 (2012). https://doi.org/10.1186/1471-2164-13-424

27. Reske, J.J., Wilson, M.R. & Chandler, R.L. ATAC-seq normalization method can significantly affect differential accessibility analysis and interpretation. Epigenetics & Chromatin 13, 22 (2020). https://doi.org/10.1186/s13072-020-00342-y

28. Xia, Jing et al. "Third-generation sequencing revises the molecular karyotype for Toxoplasma gondii and identifies emerging copy number variants in sexual recombinants." Genome research vol. 31,5 (2021): 834-851. doi:10.1101/gr.262816.120

29. Berná L, Marquez P, Cabrera A, Greif G, Francia ME, Robello C. 2021. Reevaluation of the Toxoplasma gondii and Neospora caninum genomes reveals misassembly, karyotype differences, and chromosomal rearrangements. Genome Res (this issue) 31: 823–833. 10.1101/gr.262832.120

30. Behnke MS, Fentress SJ, Mashayekhi M, Li LX, Taylor GA, Sibley LD (2012) The Polymorphic Pseudokinase ROP5 Controls Virulence in Toxoplasma gondii by Regulating the Active Kinase ROP18. PLoS Pathog 8(11): e1002992. https://doi.org/10.1371/journal.ppat.1002992

31. El Hajj, Hiba et al. "Inverted topology of the Toxoplasma gondii ROP5 rhoptry protein provides new insights into the association of the ROP2 protein family with

the parasitophorous vacuole membrane." *Cellular microbiology* vol. 9,1 (2007): 54-64. doi:10.1111/j.1462-5822.2006.00767.x

32. Foroutan, Masoud et al. "Vaccination with a novel multi-epitope ROP8 DNA vaccine against acute Toxoplasma gondii infection induces strong B and T cell responses in mice." Comparative immunology, microbiology and infectious diseases vol. 69 (2020): 101413. doi:10.1016/j.cimid.2020.101413

33. Liang, Jiaxin et al. "The functions and mechanisms of prefoldin complex and prefoldin-subunits." *Cell & bioscience* vol. 10 87. 20 Jul. 2020, doi:10.1186/s13578-020-00446-8

34. Kumar, Vikash et al. "Prefoldin subunit 6 of Plasmodium falciparum binds merozoite surface protein-1." *FEBS open bio* vol. 12,5 (2022): 1050-1060. doi:10.1002/2211-5463.13022

35. Wang, Shuai et al. "*Toxoplasma gondii* Elongation Factor 1-Alpha (TgEF-1α) Is a Novel Vaccine Candidate Antigen against Toxoplasmosis." *Frontiers in microbiology* vol. 8 168. 13 Feb. 2017, doi:10.3389/fmicb.2017.00168

36. Maria C. Dalmasso, David O. Onyango, Arunasalam Naguleswaran, William J. Sullivan, Sergio O. Angel,Toxoplasma H2A Variants Reveal Novel Insights into Nucleosome Composition and Functions for this Histone Family, Journal of Molecular Biology, Volume 392, Issue 1, 2009, Pages 33-47, ISSN 0022-2836, https://doi.org/10.1016/j.jmb.2009.07.017.

37. Wang, Yanhua, and Hong Yin. "Research progress on surface antigen 1 (SAG1) of Toxoplasma gondii." *Parasites & vectors* vol. 7 180. 13 Apr. 2014, doi:10.1186/1756-3305-7-180

38. Wasmuth, James D et al. "Integrated bioinformatic and targeted deletion analyses of the SRS gene superfamily identify SRS29C as a negative regulator of Toxoplasma virulence." *mBio* vol. 3,6 e00321-12. 13 Nov. 2012, doi:10.1128/mBio.00321-12

39. Plattner, Fabienne et al. "Toxoplasma profilin is essential for host cell invasion and TLR11-dependent induction of an interleukin-12 response." Cell host & microbe vol. 3,2 (2008): 77-87. doi:10.1016/j.chom.2008.01.001

40. Ebert, Caroline et al. "Expression and function of ABC-transporter protein ABCB1 correlates with inhibitory capacity of Ruxolitinib in vitro and in vivo." *Haematologica* vol. 101,3 (2016): e81-5. doi:10.3324/haematol.2015.136754

41. Larrazabal, Camilo et al. “P-Glycoprotein Inhibitors Differently Affect *Toxoplasma gondii*, *Neospora caninum* and *Besnoitia besnoiti* Proliferation in Bovine Primary Endothelial Cells.” *Pathogens (Basel, Switzerland)* vol. 10,4 395. 25 Mar. 2021, doi:10.3390/pathogens10040395

42. Venkataraman, Sangita et al. “RNA Dependent RNA Polymerases: Insights from Structure, Function and Evolution.” *Viruses* vol. 10,2 76. 10 Feb. 2018, doi:10.3390/v10020076

Chapter 1.7 Appendix

Supplemental Material

Supplemental Table 1. ChIP-seq Raw and Processed Mapping Statistics

Library	**Number of Reads from Sequencing (counting R1 and R2)**	**Fully Processed Reads**
AST H3K4me3		
Astro D3 H3K4me3 rep 2	7,664,552	5,512,160
Astro D3 H3K4me3 rep 1	5,937,454	4,390,214
Astro D5 H3K4me3 rep 1	46,125,430	30,322,160
Astro D5 H3K4me3 rep 2	29,935,398	20,274,772
Astro D7 H3K4me3 rep 1	176,984,942	102,658,276
Astro D7 H3K4me3 rep 2	47,781,266	35,232,388
AST H3K9me3		
Astro D3 H3K9me3 rep 1	12,315,682	3,936,146
Astro D3 H3K9me3 rep 2	7,513,964	909,406
D5 Astro H3K9me3 rep 1	177,093,030	100,493,350
Astro D5 H3K9me3 rep 2	54,346,770	19,664,278
Astro D7 H3K9me3 rep 1	39,573,190	1,386,498
Astro D7 H3K9me3 rep 2	49,713,060	2,103,174
AST H3K14ac		
Astro D3 H3K14ac rep 1	34,646,068	1,614,748
Astro D3 H3K14ac rep 2	11,995,862	1,090,630
Astro D5 H3K14ac rep 1	34,816,536	2,816,608
Astro D5 H3K14ac rep 2	16,138	430
Astro D7 H3K14ac rep 1	201,226,784	21,170,896
Astro D7 H3K14ac rep 2	39,045,236	764,792
HFF H3K4me3		
HFF D5 H3K4me3 rep 1	47,907,366	24,149,898
HFF D5 H3K4me3 rep 2	48,799,930	26,986,142
HFF D7 H3K4me3 rep 1	42,532,454	25,756,902
HFF D7 H3K4me3 rep 2	66,628,274	27,569,604
HFF H3K9me3		

HFF D5 H3K9me3 rep 1	157,854,086	6,394,264
HFF D5 H3K9me3 rep 2	52,150,952	2,870,786
HFF D7 H3K9me3 rep 1	47,530,140	7,222,556
HFF D7 H3K9me3 rep 2	39,224,434	3,396,030
HFF H3K14ac		
HFF D5 H3K14ac rep 1	40,471,880	525,792
HFF D5 H3K14ac rep 2	54,871,198	1,958,578
HFF D7 H3K14ac rep 1	43,863,548	4,540,612
HFF D7 H3K14ac	11,254,346	7,785,056
BZ (Bradyzoites)		
Bradyzoites H3K4me3 rep 1	20,885,616	13,893,968
Bradyzoites H3K4me3 rep 2	8,235,090	4,188,596
Bradyzoites H3K9me3 rep 1	35,410,340	5,314,072
Bradyzoites H3K9me3 rep 2	12,361,054	329,304
Bradyzoites H3K14ac rep 1	24,653,984	576,726
Bradyzoites H3K14ac rep 2	4,103,898	61,884
Bradyzoites H3K14ac rep 2 (Re-Seq)	1,313,534	22,200

Supplemental Table 2. AST Day 3 and Day 5

Seqnames	Called	Fold	p.value	FDR	Gene	Short_Name	Description
TGME49_chrIX	Day 3	8.398748758	1.66E-09	1.72E-06	TGME49_266920		"3'5'-cyclic+nucleotide+phosphodiesterase+domain-containing+protein"
TGME49_chrIX	Day 3	7.985297722	1.23E-08	9.16E-06	TGME49_290310		hypothetical+protein
TGME49_chrII	Day 3	7.875297801	2.21E-08	1.28E-05	TGME49_221490		cell+cycle+regulator+protein
TGME49_chrXI	Day 3	7.46814627	1.63E-07	4.94E-05	TGME49_217182		RNA-dependent+RNA+polymerase+RDP
TGME49_chrII	Day 3	7.462106027	1.71E-07	4.94E-05	TGME49_222060		hypothetical+protein
TGME49_chrXI	Day 3	7.122749985	7.18E-07	0.0001576940432	TGME49_309980		dynein+heavy+chain+family+protein
TGME49_chrXII	Day 3	6.38181288	7.51E-06	0.0009371063941	TGME49_219738		hypothetical+protein
TGME49_chrXI	Day 3	6.327244484	7.69E-06	0.0009371063941	TGME49_215960		hypothetical+protein
TGME49_chrIb	Day 3	6.089234876	1.20E-05	0.001304361307	TGME49_207410		tetratricopeptide+repeat-containing+protein
TGME49_chrVIII	Day 3	5.928719163	1.53E-05	0.001475383072	TGME49_269050		hypothetical+protein
TGME49_chrVIII	Day 3	5.14329205	3.87E-05	0.003050472299	TGME49_233450	SRS1	SAG-related+sequence+SRS29A
TGME49_chrXII	Day 3	5.07584759	2.50E-05	0.002138232336	TGME49_308090	ROP5	rhoptry+protein+ROP5
TGME49_chrXI	Day 3	5.026698159	4.24E-05	0.003249866663	TGME49_216710		transporter,+major+facilitator+family+protein
TGME49_chrVI	Day 3	4.941420257	4.60E-05	0.003472675383	TGME49_243382		hypothetical+protein
TGME49_chrIX	Day 3	4.819397877	5.12E-05	0.003809339842	TGME49_291880		hypothetical+protein
TGME49_chrVIII	Day 3	4.530165943	6.49E-05	0.004550770763	intergenic		
TGME49_chrV	Day 3	4.470678528	6.80E-05	0.004651077053	TGME49_285865		hypothetical+protein

TGME49_chrVIIb	Day 3	4.3880423	7.28E-05	0.004860216484	intergenic		
TGME49_chrVIII	Day 3	3.31432563	0.0001608893763	0.009069503043	TGME49_231991		hypothetical+protein
TGME49_chrIa	Day 3	2.556753896	0.0002812226123	0.01331205584	TGME49_293680		hypothetical+protein
TGME49_chrVIII	Day 3	2.381729724	0.0003201617109	0.0143713968	TGME49_233500	TPI-II	triose-phosphate+isomerase+TPI-II
TGME49_chrVIIa	Day 3	2.369212204	6.88E-05	0.004651077053	TGME49_205060		hypothetical+protein
TGME49_chrIa	Day 3	2.22816919	7.73E-09	6.71E-06	TGME49_295920		hypothetical+protein
TGME49_chrXI	Day 3	2.218661526	0.0003656234557	0.01573389532	TGME49_313235		hypothetical+protein
TGME49_chrIX	Day 3	2.078507288	0.0004107773162	0.01684186996	TGME49_289740		hypothetical+protein
TGME49_chrVIIb	Day 3	1.982241027	0.0004467679234	0.01789477367	TGME49_260460		oxidoreductase,+putative
TGME49_chrX	Day 3	1.897438894	0.0004780915171	0.01885926159	TGME49_236990		beta-ketoacyl+synthase,+N-terminal+domain-containing+protein
TGME49_chrXII	Day 3	1.830364235	1.19E-13	6.22E-10	TGME49_250710	MIC10	microneme+protein+MIC10
TGME49_chrVI	Day 3	1.354842114	0.0004390497144	0.01772195242	TGME49_242055		"DEAD/DEAH+box+helicase+domain-containing+protein"
TGME49_chrVIIa	Day 3	1.348971252	3.03E-08	1.31E-05	TGME49_203300		hypothetical+protein
TGME49_chrXII	Day 3	1.329467704	1.37E-05	0.001402122388	TGME49_218920		proteasome+subunit+beta+type,+putative
TGME49_chrVI	Day 3	1.306501092	3.52E-08	1.41E-05	TGME49_239910		cyclin-dependent+kinase
TGME49_chrVI	Day 3	1.286732187	1.01E-05	0.001140793406	TGME49_239795		hypothetical+protein
TGME49_chrIII	Day 3	1.277544985	0.0008995383041	0.031225973	TGME49_252890		hypothetical+protein
TGME49_chrX	Day 3	1.267223743	8.45E-07	0.0001692231094	TGME49_212160		hypothetical+protein

TGME49_chrX	Day 3	1.208064757	1.03E-09	1.34E-06	TGME49_226010		pterin-4-alpha-carbinolamine+dehydratase
TGME49_chrVIII	Day 3	1.145375986	1.70E-08	1.10E-05	intergenic		
TGME49_chrXII	Day 3	1.122769866	9.93E-08	3.23E-05	TGME49_218740		membrane+protein,+putative
TGME49_chrX	Day 3	1.096352467	1.72E-06	0.0002977116634	TGME49_227335		hypothetical+protein
TGME49_chrXII	Day 3	1.051409069	3.76E-06	0.0005928198678	TGME49_246170		"ARID/BRIGHT+DNA+binding+domain-containing+protein"
TGME49_chrII	Day 3	1.044745302	0.001199065476	0.0378395996	TGME49_222075		hypothetical+protein
TGME49_chrIII	Day 5	-1.081474488	1.83E-10	3.17E-07	intergenic		
TGME49_chrXI	Day 5	-1.236796286	2.84E-05	0.002346338841	TGME49_309930		melibiase+subfamily+protein
TGME49_chrIII	Day 5	-1.358464217	1.89E-07	5.18E-05	TGME49_298820		hypothetical+protein
TGME49_chrII	Day 5	-2.571930629	3.31E-11	8.62E-08	TGME49_297150		MORN+repeat-containing+protein
TGME49_chrIII	Day 5	-5.545455912	3.99E-06	0.0006111342762	TGME49_252640		P-type+ATPase+PMA1

Supplemental Table 3. AST Day 5 and AST Day 7

Seqnames	Called	Fold	p.value	FDR	Gene	Short_Name	Description
TGME49_chrIX	Day 5	3.433387817	2.24E-31	1.19E-27	TGME49_290310		hypothetical+protein
TGME49_chrII	Day 7	-1.304338532	5.65E-06	0.007479583656	TGME49_222080		hypothetical+protein

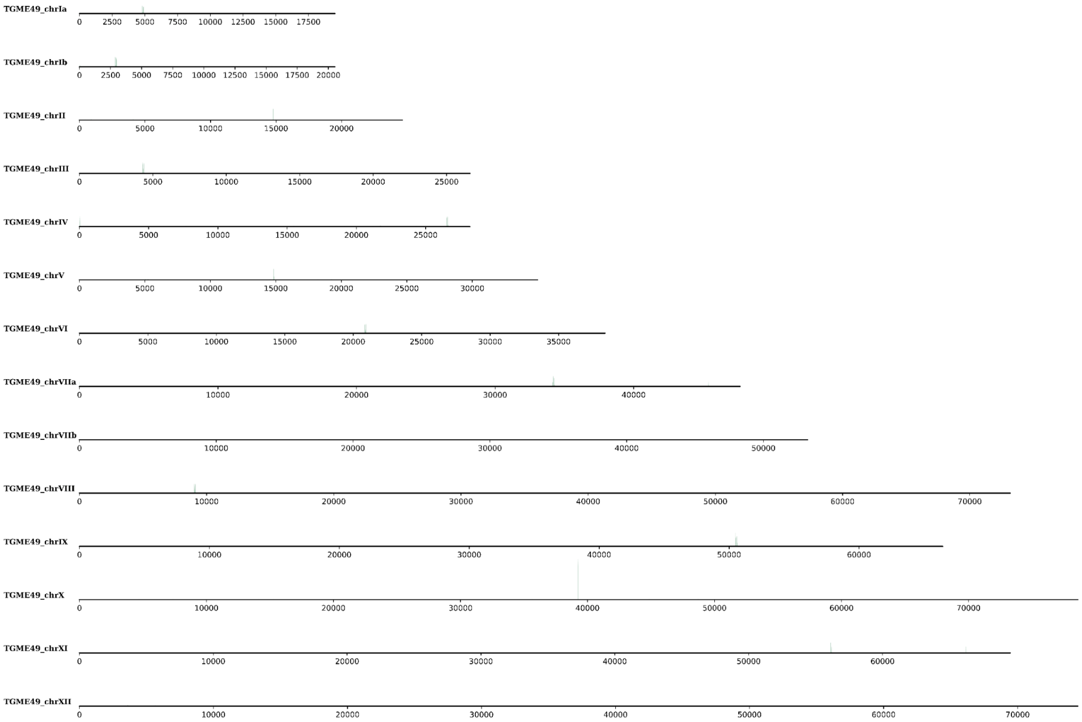
Supplemental Table 4. HFF Day 5 and HFF day 7

Seqnames	Called	Fold	p.value	FDR	Gene	Short_Name	Description
TGME49_chrIX	Day 7	1.95E+00	4.05E-12	1.07E-08	TGME49_290310		hypothetical+protein
TGME49_chrXII	Day 7	1.9469782	1.61E-09	1.42E-06	TGME49_245770		hypothetical+protein
TGME49_chrVIII	Day 7	1.199339944	2.66E-11	3.53E-08	TGME49_268985		hypothetical+protein
TGME49_chrXII	Day 7	1.201652508	2.75E-07	0.0001145757234	TGME49_277230		hypothetical+protein
TGME49_chrXII	Day 7	9.146444466	1.41E-11	2.49E-08	TGME49_245980		hypothetical+protein
TGME49_chrX	Day 7	4.295915905	3.09E-13	1.64E-09	TGME49_215210		hypothetical+protein
TGME49_chrXI	Day 7	2.75010713	2.14E-06	0.0006675626406	TGME49_313235		hypothetical+protein
TGME49_chrIa	Day 5	-6.49E-01	5.20E-06	1.38E-03	TGME49_295110	ROP7	rhoptry+protein+ROP7
TGME49_chrIV	Day 5	-1.30E+00	2.94E-10	3.11E-07	TGME49_317705		"enoyl-CoA+hydratase/isomerase+family+protein"
TGME49_chrII	Day 5	-1.31E+00	3.40E-07	1.29E-04	TGME49_297150		MORN+repeat-containing+protein

Supplemental Figure 1. Astrocyte K4 Day 3

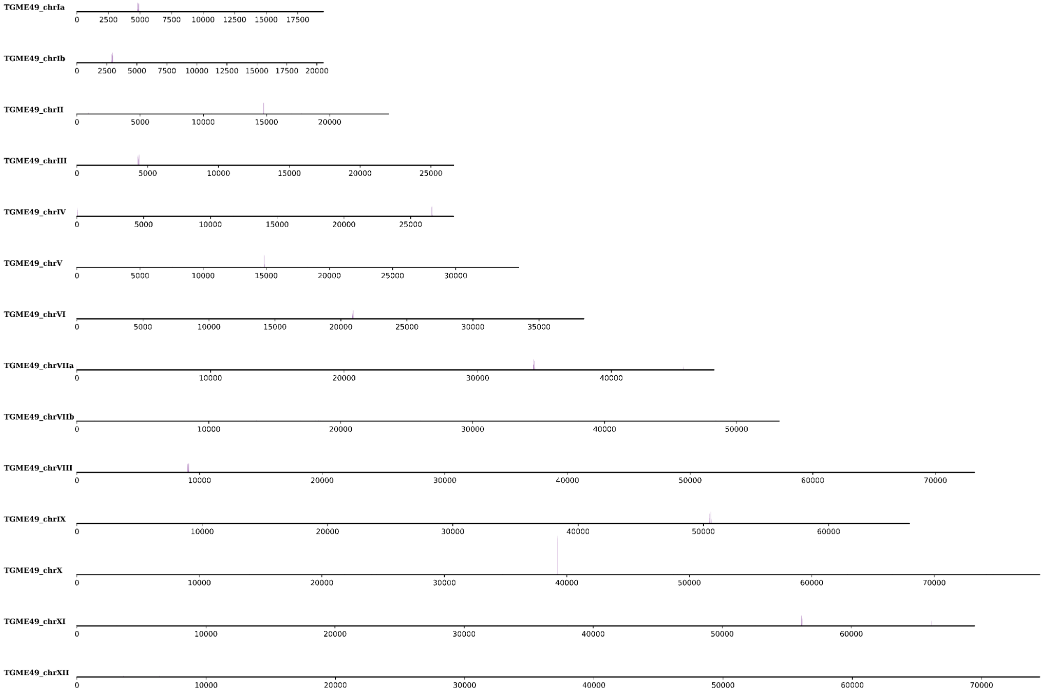
Supplemental Figure 2. Astrocyte K4 Day 5

Supplemental Figure 3. Astrocyte K4 Day 7

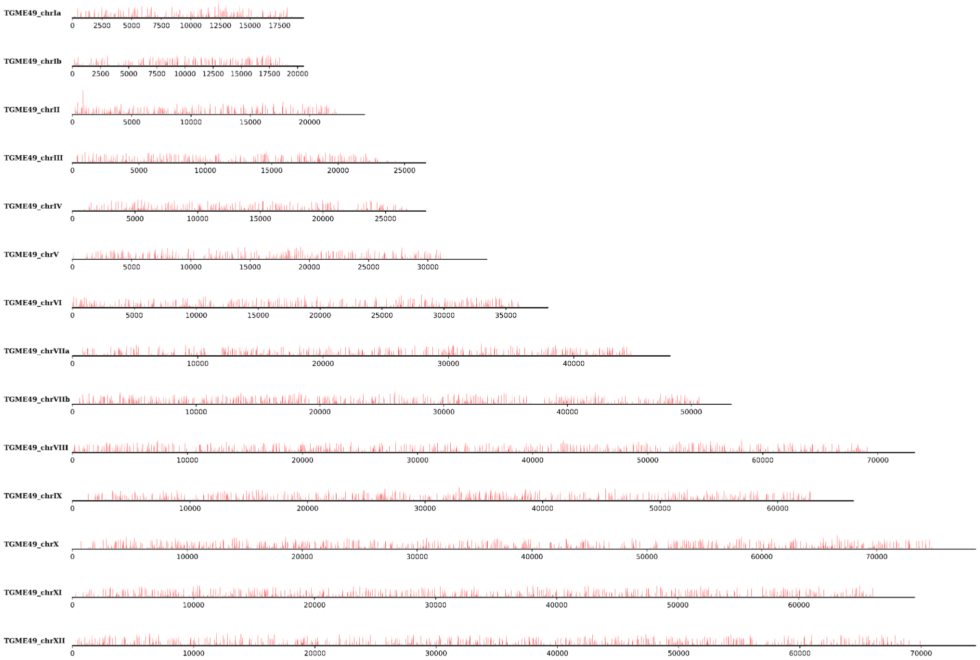


Supplemental Figure 4. Astrocyte K9 Day 3

Supplemental Figure 5. Astrocyte K9 Day 5

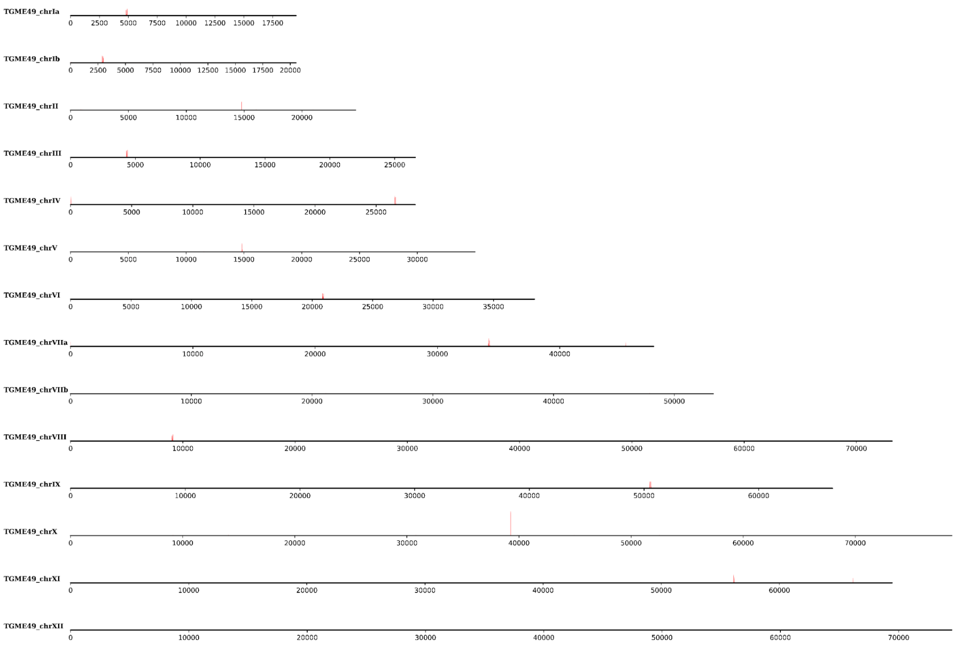


Supplemental Figure 6. Astrocyte K9 Day 7

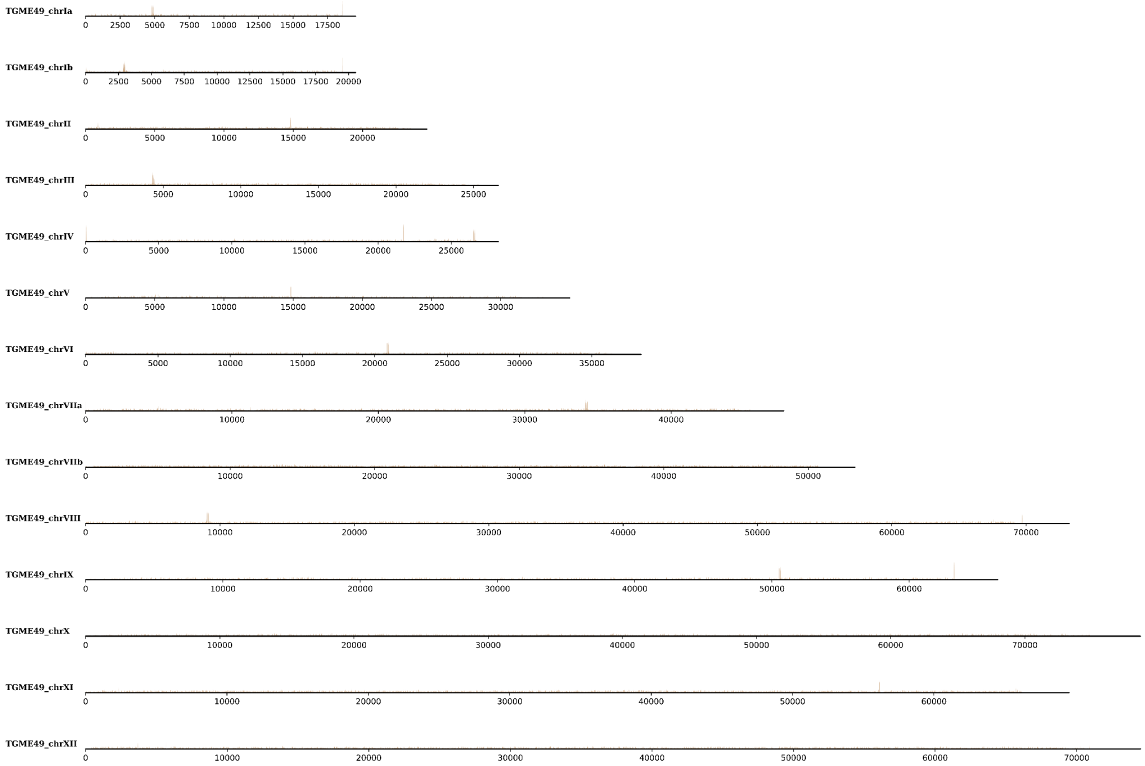


Supplemental Figure 7. HFF K4 Day 5

Supplemental Figure 8. HFF K4me3 Day 7



Supplemental Figure 9. HFF K9me3 Day 5



Supplemental Figure 10. HFF K9me3 Day 7