Identification and Characterization of Insulator Elements and Proteins in *Plasmodium falciparum*

A Thesis

submitted to

Indian Institute of Science Education and Research Pune in partial fulfilment of the requirements for the BS-MS Dual Degree Programme

by

Ghadage Kalyan Pradip (20181175)

Indian Institute of Science Education and Research Pune

Dr. Homi Bhabha Road,

Pashan, Pune 411008, INDIA.

April 2023

Supervisor: **Dr. Krishanpal Karmodiya**

Assistant Professor, Department of Biology, IISER Pune

From May 2022 to Mar 2023

INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH PUNE

Certificate

This is to certify that this dissertation entitled '**Identification and Characterization of Insulator Elements and Proteins in *Plasmodium falciparum***' towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents study/work carried out by Ghadage Kalyan Pradip at Indian Institute of Science Education and Research under the supervision of Dr. Krishanpal Karmodiya, Department of Biology, during the academic year 2022-2023.

Dr. Krishanpal Karmodiya
Assistant Professor, IISER Pune

Committee:

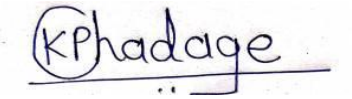
Dr. Krishanpal Karmodiya
Assistant Professor,
Department of Biology, IISER Pune
(Supervisor)

Dr. Mridula Nambiar
Assistant Professor,
Department of Biology, IISER Pune
(TAC member)

This thesis is dedicated to my beloved Parents and Family!!

Declaration

I hereby declare that the matter embodied in the report entitled '**Identification and Characterization of Insulator Elements and Proteins in *Plasmodium falciparum***' are the results of the work carried out by me at the Department of Biology, Indian Institute of Science Education and Research, Pune, under the supervision of Dr. Krishanpal Karmodiya and the same has not been submitted elsewhere for any other degree.



Dr. Krishanpal Karmodiya
Assistant Professor,
Biology Department, IISER Pune
Date: 10 April 2023

Ghadage Kalyan Pradip
20181175
BS-MS Student, IISER Pune
Date: 10 April 2023

Table of Contents

Abstract 7

Acknowledgments 8

Contributions 9

Chapter 1 Introduction 10

Chapter 2 Materials and Methods 18

Chapter 3 Results 26

3.1 ChIP sequencing analysis shows differential enrichment of chromatin modification on *var* gene intron 26

3.2 In-silico analysis to identify the putative insulator-like proteins in *Plasmodium falciparum* 32

3.3 Insulator-like motif interacts with the proteins from *P. falciparum* nuclear extract 36

3.4 Cloning and generation of transfection lines for the selected proteins of interest 41

3.5 PfHP1 as the possible insulator-like protein and study of its interactome 43

Chapter 4 Discussion 46

Chapter 5 References 50

List of Tables

Table 1: Primers for the pull-down of insulator-like proteins 19

Table 2: Set-up of concatenation reaction 21

Table 3: Primers for the PCR amplification of genes 22

Table 4: List of H3K9ac+ and H3K9ac- genes 30

Table 5: List of the insulator proteins and their homolog in *P. falciparum* 33

Table 6: List of SMC proteins 35

Table 7: List of the proteins obtained as hits in pull-down experiment 39

Table 8: Status of cloning of the genes and transfection lines 42

Table 9: List of the proteins obtained as hits in PfHP1 IP experiment 44

List of Figures

Figure 1: The life cycle of *Plasmodium falciparum* 11
Figure 2: A schematic presentation of var genes 12
Figure 3: Nuclear organization of *var* genes 13
Figure 4: Epigenetic modifications on clonally variant multicopy genes 14
Figure 5: PfHP1 ChIP sequencing showing enrichment on chromosome 3 17
Figure 6: Occupancy of histone modification on *var* genes PF3D7_1200400 and PF3D7_0800300 28
Figure 7: Heatmap for the enrichment of H3K9ac and PfHP1 protein on *var* genes 29
Figure 8: Average enrichment of H3K9ac, H4K8ac and PfHP1 protein on H3K9ac+ and H3K9ac- *var* genes 30
Figure 9: Plot for the occupancy of PfTRZ and PfHP1 on *var* genes 35
Figure 10: Motif analysis highlighting motifs and their location on *var* gene intron 37
Figure 11: Validation of the NE preparation. 39
Figure 12: Number of protein hits in biotin-oligo pull-down experiment. 40
Figure 13: Pull-down with concatenated DNA schematic. 40
Figure 14: ChIP sequencing analysis for PfGBP2, PfISWI and PfHMGB1 42
Figure 15: Hi-C and ChIP sequencing analysis for chromosome 7 44

Abstract

Malaria, a life-threatening disease is becoming a major concern on a global scale due to the resistance to the available drug treatment and the lack of an effective vaccine. *Plasmodium falciparum*, the deadliest *Plasmodium* species responsible for severe malaria, exhibits antigenic variation through mutually exclusive expression and timely switching of *var* genes which is responsible for the enhanced pathological complexity during the infection and longevity of the infection. The exact regulatory mechanism of antigenic variation is still unclear, although research to date suggests the potential roles of characteristic positioning of *var* genes on chromosomes, their subnuclear organization, epigenetic modifications, and non-coding RNAs in the expression of *var* genes in a mutually exclusive manner. In this study, we are identifying the putative insulator-like proteins in *Plasmodium falciparum* and investigating the prospect of insulators serving as the possible regulators involved in the mutually exclusive expression of *var* genes. Through the in-silico approach and wet lab experiments, we have identified the putative insulator-like proteins and validating their role through qPCR followed by overexpression of the proteins of interest. Overall, we are also aiming to identify the putative chromatin architectural proteins in *Plasmodium falciparum*. Collectively, this study identifies the putative insulator-like proteins in *Plasmodium falciparum* and provides future directions to characterize their role in *var* gene expression and 3D genome organization.

Acknowledgments

I would want to express my gratitude to everyone who contributed to the successful completion of my MS Project. I want to start by expressing my gratitude to my supervisor, Dr. Krishanpal Karmodiya, for allowing me to work on this project, for offering his advice and patient support throughout the project, and for his invaluable insights into the design and direction of the project. A sincere vote of thanks to Dr. Mridula Nambiar for agreeing to be my TAC member and providing suggestions for the project's improvement and progress.

I would like to express my heartfelt gratitude to Mamatharani D V for her guidance and continuous assistance at every stage of the project. I appreciate her contribution to developing and putting our many ideas into action. Thanks a lot for always being there. I am extremely grateful to all the members of Krish lab for their help, support, and love as well as for maintaining a very healthy and research-friendly atmosphere in the lab. It was great working with every one of you and learning so much. I would like to thank the Mass-spectrometry facility at IISER Pune and the biology department of IISER Pune for providing the required facilities.

I would like to thank my family for their unconditional love, motivation, and support that encouraged me to reach to this point. A special thanks to all my friends for their moral support and encouragement that helped me to keep going and do my best.

Contributions

Contributor name	Contributor role
KG, MDV, KK	Conceptualization Ideas
KG, MDV	Methodology
KG	Software
KG	Validation
KG	Formal analysis
KG	Investigation
KG	Resources
KG	Data Curation
KG	Writing - original draft preparation
KG, MDV, KK	Writing - review and editing
--	Visualization
MDV, KK	Supervision
KK	Project administration
KK	Funding acquisition

Authors' contributions: Kalyan Ghadage (KG), Mamatharani D V (MDV), Krishanpal Karmodiya (KK)

1. Introduction

Malaria is a mosquito-borne parasitic infectious disease caused by the protozoan of the *Plasmodium* genus with a major prevalence in the tropical and subtropical areas of the globe. The emergence of parasites resistant to the gold standard artemisinin-based combination therapies has raised serious concerns regarding the treatment of malaria and restricting the spread of the disease (Dhorda et al., 2021). According to World Malaria Report, 2020, malaria infection led to ~240 million cases and around 6 lakh deaths making malaria a major concern worldwide (Report, 2020). *P. falciparum, P. vivax, P. malariae, P. knowlesi* and *P. ovale* are the five *Plasmodium* species that are capable of infecting humans among which the *P. falciparum* is the causative agent of severe malaria in humans. The parasite *Plasmodium falciparum* transits between two hosts for the completion of its life cycle, the invertebrate host female anopheline mosquito and humans being the vertebrate host. In humans, malaria infection begins when during a blood meal the infected mosquito injects the infective form of parasites known as sporozoites into the healthy human body. The injected sporozoites find their way to invade liver cells (hepatocytes) through the bloodstream. In hepatocytes, parasites multiply asexually to form merozoites. The merozoites, when released into the bloodstream, invade red blood cells (RBCs) and undergo an intraerythrocytic developmental cycle (IDC) which involves three different stages namely ring, trophozoite and schizont. During the developmental transition from the ring stage to schizont, parasites undergo multiple asexual divisions to form 16-24 merozoites. Rupturing of these schizonts releases the merozoites which can invade fresh RBCs and continue the IDC of the parasite leading to the increase in the parasitemia inside the host body. IDC is the symptomatic stage of the malaria infection during which the infected individual shows symptoms like fever, chills, headache, muscle aches, fatigue and other complications in case of severe malaria (Wassmer et al., 2015). To complete the lifecycle of the parasite, some of the parasites in the IDC take the fate of sexual forms termed male and female gametocytes through the process of gametocytogenesis. Mosquitoes during the blood meal ingest the gametocytes. The process of maturation of gametocytes and formation of zygote occurs in the midgut of the mosquito and further develops to form sporozoites (Fig. 1). This completes the lifecycle of the parasite (Andrews et al., 2009).

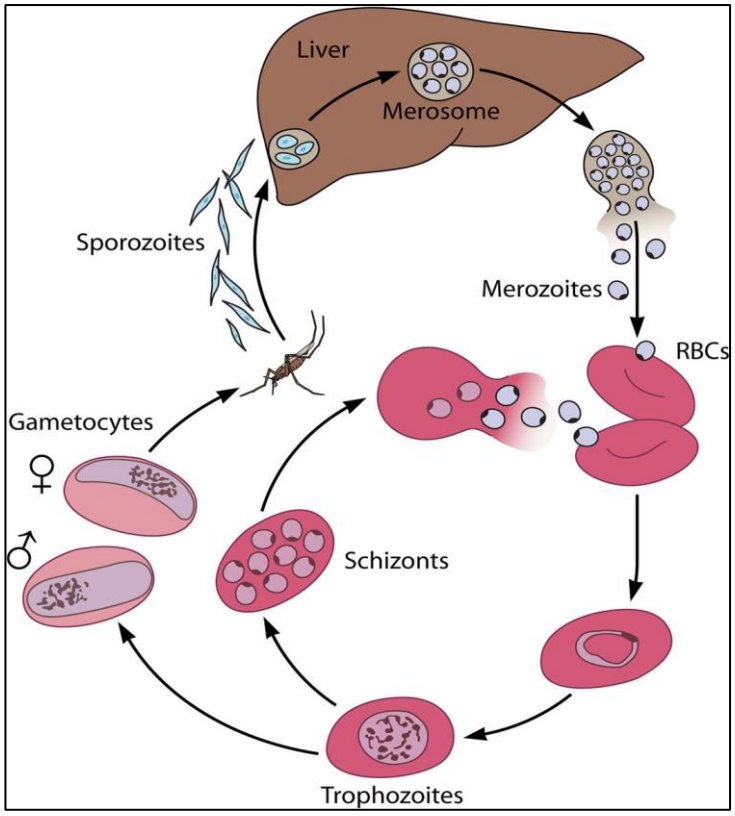


Figure 1: The life cycle of *Plasmodium falciparum*. An illustration representing the intraerythrocytic, sexual and liver stages of the life cycle (Adapted from Guggisberg et. al., 2014).

During IDC through the insertion of parasite-specific proteins into the RBC membrane, parasites are able to modify the surface of infected RBCs (iRBCs). *Plasmodium falciparum* erythrocyte membrane protein 1 (PfEMP1), a protein encoded by *var* genes, is a dominantly expressed protein on the surface of iRBCs (Ryan et al., 2013). The *var* gene family consisting of nearly 60 genes codes for this protein (Gardner et al., 2002). Depending on the 5' upstream regions and the orientation of the gene on chromosomes *var* genes are classified into 5 classes, namely, upsA, upsB, upsC, upsD and upsE. There are two exons in each var gene, which are separated by an intronic sequence (Fig. 2). Exon1 is highly polymorphic while the exon2 and intron are conserved in all the *var* genes. While the part of the protein encoded by exon 2 stays intracellular, the part encoded by exon 1 is expressed outside the RBC surface. The expression of PfEMP1 on the RBC surface leads to the phenomena of cytoadherence and rosette formation by the iRBCs (Analysis, 1985; Carison et al., 1989). Through these phenomena parasites remain sequestered in the small blood vessels, obstructing the blood flow and could damage the associated tissues leading to severe pathologies. Cytoadherence and sequestration help parasites circumvent splenic clearance of the iRBCs and escape the filtration and destruction process from the host body (Miller et al., 2002, 2010). The host immune system can generate a strong

adaptive immune response against the expressed PfEMP1 protein as the membrane insertion of this foreign protein makes it a primary antigen (Miller et al., 2002; Scherf et al., 1998). However, the parasites have evolved a strategy to circumvent the adaptive immune response generated by immune system of the host. Although the parasite possesses about 60 var genes, only one of them is expressed at a time while the others are kept transcriptionally silent; this is referred to as the mutually exclusive expression of *var* genes (Scherf et al., 1998). The parasites also possess the memory of which *var* gene is being expressed so the parasite population inside the host or in the in-vitro-grown culture majorly expresses a single *var* gene. When the host-body generated immune response against the expressed PfEMP1 targets the parasite population and reduces the parasitemia to a significant level, a small number of parasitic populations switches the expression to one of the silent *var* genes leading to the transcription and expression of a variant PfEMP1 and keeps the earlier expressed gene silent (Milner, 2018). Because of this, the immune response generated against earlier expressed PfEMP1 stands ineffective and the parasitic populations expressing a different PfEMP1 protein undergo clonal expansion increasing again the parasitic burden on the host body. The ~60 variant PfEMP1 proteins are the antigenic repertoire of the parasite, their mutually exclusive expression and timely switching help parasites to escape the different stress conditions as well as the host immune response continuing the chronic and persistent infection inside the host body.

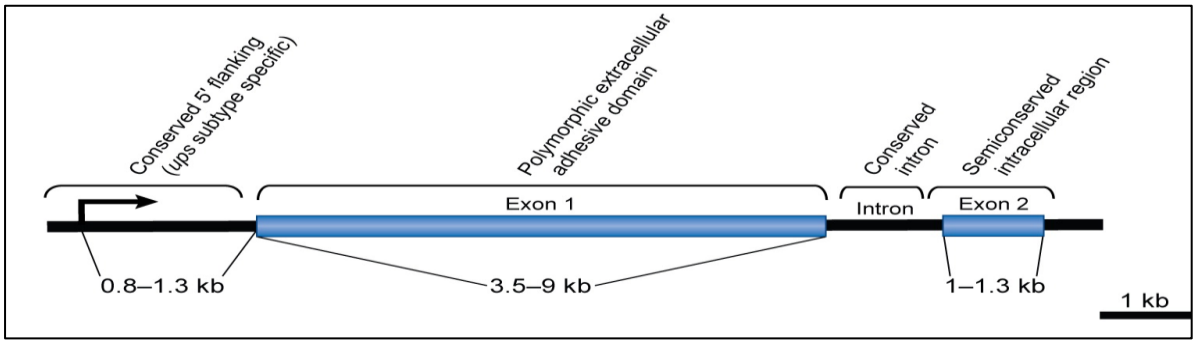


Figure 2: A schematic presentation of *var* genes. The figure demonstrates the structure of *var* genes with two exons and an intron as well as indicates the variable lengths of them in all *var* genes.

The mutually exclusive expression of *var* genes involves a complex mechanism of regulation governed by multiple factors. Apart from the transcription factors regulating the transcription of a single *var* gene, there should be mechanisms mediating the

crosstalk and interaction between all the *var* genes such that only a single gene is actively expressed while the rest are silent. The *var* genes are specifically present at the telomeric, sub-telomeric and pericentromeric regions of the chromosomes (Rrubio et al., 1996). These regions are localized at the periphery of the nucleus and are marked by the presence of histone modifications H3K9me3 and H3K36me3 which are repressive in nature (Chookajorn et al., 2007; Cui and Miao, 2010). The binding of *Plasmodium falciparum* heterochromatin protein 1 (PfHP1) to the H3K9me3 marked regions leads to the heterochromatinization of those regions (Flueck et al., 2009; Pérez-Toledo et al., 2009). As reported by the Fluorescent in Situ Hybridization (FISH) experiments, the heterochromatinized *var* gene loci are present close to the nuclear periphery forming 6-7 clusters (Freitas-junior et al., 2000) (Fig. 3). The active *var* gene is also present at the nuclear periphery but is not part of the heterochromatic foci. The active *var* gene possibly moves away from the heterochromatic foci and relocates to the *var* gene expression site as proposed by the previous studies (Ralph et al., 2005). This suggests that the characteristic nuclear organization of *var* genes influences their transcriptional activity.

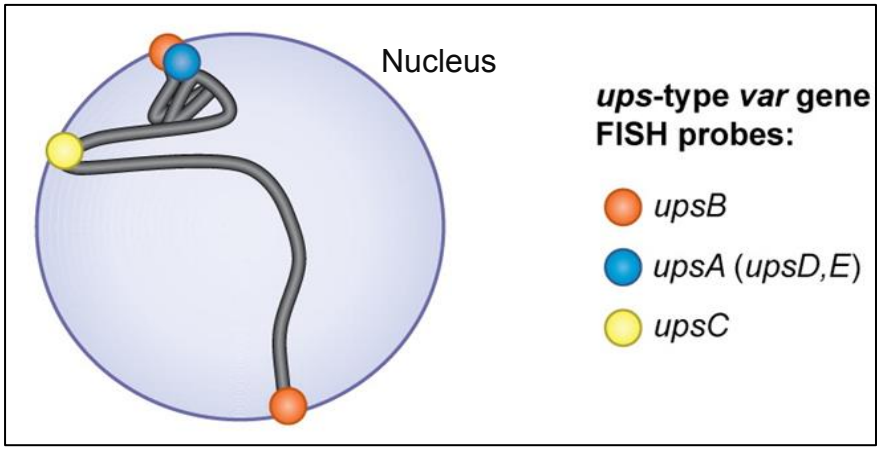


Figure 3: Nuclear organization of var genes. A schematic presentation of the positioning of *var* genes at the outer edge of the nucleus using FISH probes which differentiate the different types of *var* genes (Adapted from Scherf et al., 2008).

Epigenetic modifications are the covalent modifications on the DNA or the amino acids of histone proteins. These modifications make up the epigenome of the organism. It includes methylation of the DNA and the modifications such as methylation, acetylation and phosphorylation on the specific amino acid residues of histones. The presence of these modifications regulates the chromatin structure, accessibility of

DNA and in turn the process of gene expression. Similar to most eukaryotes, the gene expression in *Plasmodium falciparum* is regulated by the presence of specific epigenetic modifications as well as the recruitment of histone variants like H2A.z, H2B.z at the regulatory regions of the genes (Karmodiya et al., 2015; Petter et al., 2013). Histone modification trimethylation of histone 3 at lysine 9 (H3K9me3) and its reader PfHP1 protein, specifically mark the clonally variant multicopy (CVM) gene families, RIFIN, stevor, *var*, etc and the promoters of transcriptionally silent genes (Flueck et al., 2009). The promoters of active genes display acetylation of histone 3 at lysine 9 (H3K9ac) modification while the trimethylation of histone 3 at lysine 36 (H3K36me3) modification is associated with both active and silent genes (Jiang et al., 2013). Interestingly, the comprehensive epigenome analysis of *Plasmodium falciparum* indicates that the CVM genes have differential chromatin modifications compared to the housekeeping genes with the enrichment of modifications towards the 3' end of the genes (Karmodiya et al., 2015) (Fig. 4).

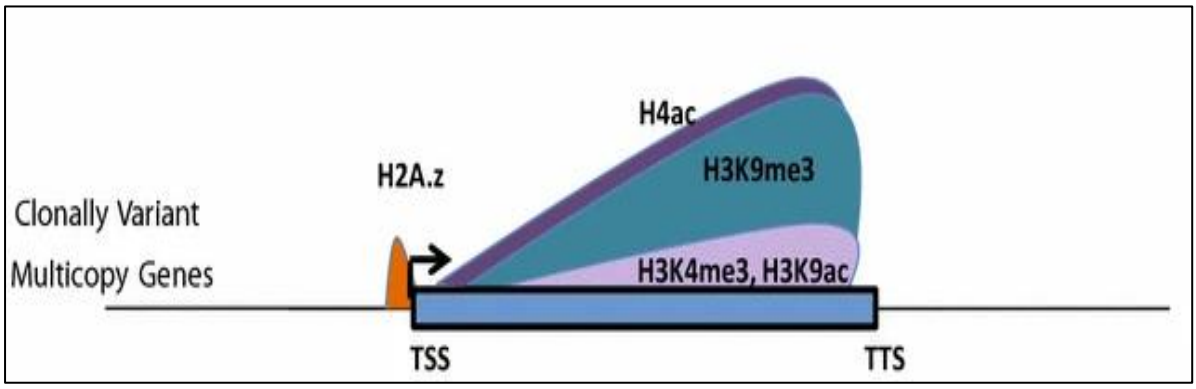


Figure 4: Epigenetic modifications on Clonally variant multicopy gene bodies. A schematic model depicting the occupancy of histone modifications H3K9me3, H3K4me3, H3K9ac, H4ac and histone 2 variant H2A.z on the CVM genes (Adapted from Karmodiya et. al., 2015)

Recent studies demonstrate the substantial production of antisense transcripts in *P. falciparum* which could possibly have a role in controlling gene expression through the recruitment of specific chromatin remodeling complexes, transcription factors, etc (Siegel et al., 2014). The intronic region of the var genes has been shown to exhibit bidirectional promoter activity, resulting in the generation of sense and antisense transcripts (Epp et al., 2009). While the sense RNA is transcribed from the intron of all the *var* genes regardless of their transcription status, at the late IDC stages, the

antisense RNA is transcribed only from the intron of the actively expressed *var* gene during the early stage of IDC, the time corresponding to the expression of the active *var* gene. The antisense transcript is nearly 2 kB long having the conserved intronic sequence at the 5' end while the 3' end includes the exon 1 region which is polymorphic (Amit-Avraham et al., 2015). The antisense and sense transcription from the intron and the enrichment of histone modifications at the intronic region (3' end), as stated above, could possibly regulate how *var* genes are expressed.

The genome of higher eukaryotes shows complex 3D organization inside the micrometer sized nucleus with the presence of strictly separated active (euchromatin) and inactive (heterochromatin) compartments (Jerković et al., 2020; Misteli, 2020). The genome is organized as topologically associated domains (TADs), which are self-interacting territories in which the chromatin inside the TAD exhibits stronger physical interactions with chromatin inside the TAD than outside (Dixon et al., 2012). TADs are formed by the looping of chromatin at the boundary/insulator elements via the binding of specific associated proteins (Chetverina et al., 2017). Insulator elements are the short cis-regulatory DNA elements to which the specific proteins called insulator proteins bind and this interaction is involved in the proper regulation of gene expression and higher order genome organization. The possible functions executed by insulators in regulating gene expression are enhancer-blocking and barrier function, the role through which insulators were first discovered in *Drosophila melanogaster* (Umbetova et al., 1991). While acting as barriers, insulators restrict the spread of chromatin modifications dividing euchromatin and heterochromatin, demarcate their boundaries and form the genome compartments. In enhancer-blocking, the loop formation by the interaction between insulator proteins and insulator elements prevents the contact between the promoter of a gene and the enhancer, a distal regulatory element that enhances gene expression through interaction with the promoter of a gene (Ali et al.; Heger and Wiehe, 2014; Kurbidaeva, 2021). This inhibition of enhancer-promoter interaction could drastically downregulate the gene expression or silence the gene. CCCTC-binding factor (CTCF) is one of the well-studied insulator proteins in *Drosophila* and mammals, which demarcates the boundaries of TADs through its interaction with cohesin complex mediating the regulation of 3D chromatin organization and expression of genes (Glenn et al., 2017; Rao et al., 2014).

Hi-C studies in *Plasmodium falciparum* at different life-cycle stages indicate that the genome of the parasite has an intricate 3D organization as domain-like structures which could be influenced by the location of *var* genes inside the nucleus (Mcgovern et al., 2019). Chromatin immunoprecipitation followed by sequencing (ChIP-sequencing) results for H3K9me3 histone modification and PfHP1 protein represent that these repressive signatures are restricted only to the specific parts of the genome, the telomeric, subtelomeric and sometimes pericentromeric regions of the chromosomes (Fraschka et al., 2018) (Fig. 5). It is unknown how this distribution of histone modifications is contained to a particular region of the genome. Surprisingly, in *P. falciparum* no insulator protein has been reported to date which could possibly govern these characteristic features of the genome. However, earlier research has indicated that the var genes contain a hypothetical insulator-like element. Having the implications of pairing between promoter and intron of the *var* genes through the previous studies (Dzikowski et al., 2007; Frank et al., 2006), in the study by Avraham et. al., 2012, authors performed motif analysis for the 5' upstream region and the intronic region of the var genes. In the analysis, they noted that these regions have conserved 8-12 bp TG-rich motifs. The three identified motifs are Motif 1: TGTATGTATGTG, Motif 2: TGTGTATATGTG and Motif 3: TGTATGTG. The three motifs form a conserved consensus motif of 18bp, TGTGTATATGTATGTGTG (Avraham et al., 2012). The Electrophoretic mobility shift assays (EMSA) performed with the motif showed that they are associated with nuclear protein complexes but none of the proteins were identified. Through their experiments, they hypothesized that this motif in the promoter and intron could mediate the pairing between them and hence it was considered an insulator-like motif, as the pairing through the motif might be insulating the *var* gene from its expression. These insulator-like motifs are enriched at the ends of the *var* gene intron, the boundary of the intron-exon junction. Intriguingly, the motifs are present in the opposite orientation at the ends of the intron, a feature conserved in known efficiently loop forming insulators. Considering this, here we hypothesize that, the insulator-like motif at the ends of *var* gene intron might undergo pairing and loop formation via the binding of specific insulator-like proteins. This kind of loop formation could be involved in the activation or silencing of var genes in coordination with the antisense and sense transcription mediated from the *var* gene intron. Along with this, as suggested by the epigenetic studies, the presence of specific chromatin modifications at the 3' end of var genes could be influencing the formation

of the loop. In a way, it could be the interplay governed by the insulator-like proteins, transcription from the intronic region and the chromatin modifications which is regulating the expression of *var* genes in a mutually exclusive manner.

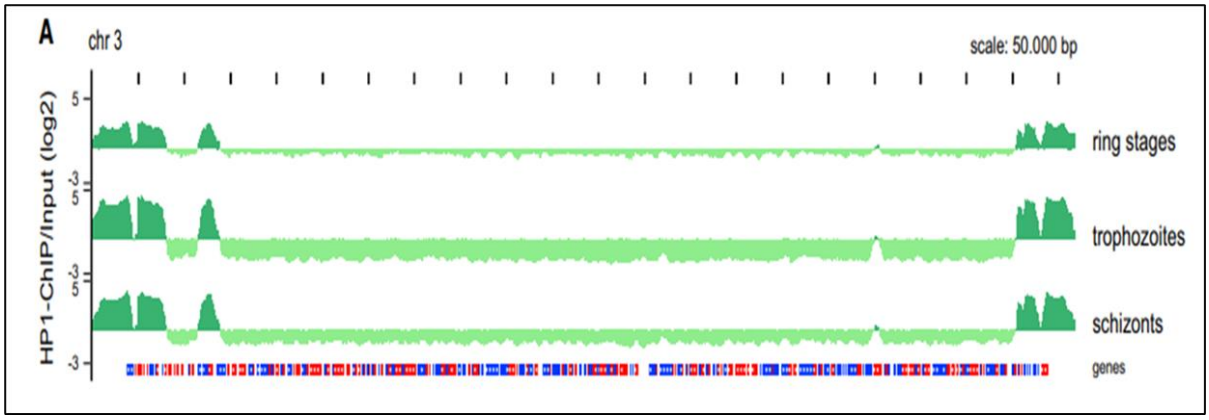


Figure 5: Enrichment of PfHP1 protein on chromosome 3. IGV image showing the enrichment of PfHP1 protein on chromosome 3 of *P. falciparum* parasites at the IDC stages ring, trophozoites and schizont. All three stages show occupancy of PfHP1 at the telomeric regions (Adapted from Freschka et. al., 2018).

In this study, we are interested in the investigation and study of the possible insulator-like proteins in *Plasmodium falciparum*. First, we analyzed the available ChIP-sequencing data for activating and repressive chromatin modifications in *P. falciparum* to study the epigenetic landscape on the *var* genes and analyze how the *var* genes differ in that. We started the search for possible insulator-like proteins with the homology study of the proteins in *P. falciparum* with the already known and characterized insulator proteins in other model organisms. Further, through an in-vitro pull-down experiment approach, we have identified the possible insulator-like proteins that could interact with the earlier hypothesized insulator-like motif in *var* gene introns which may be involved in a loop formation. We are also trying to find if the putative candidates for insulator-like activity bind the insulator-like motif present in the *var* gene introns, by the overexpression of proteins in in-vitro culture-grown parasites. It would be fascinating to examine the consequences of insulator-like protein overexpression on the mutually exclusive expression of *var* genes. We hope that this study sheds light on the unexplored aspects of the mechanism regulating the expression of *var* genes in mutually exclusive manner and would also highlight the role of architectural chromatin organization regulators in *P. falciparum*.

2. Materials and Methods

2.1. Parasite culture

Plasmodium falciparum 3D7 (Pf3D7) parasites were cultured in vitro in RPMI1640 medium added with 0.5 % AlbuMAX II, 25 mM HEPES (buffering agent), 11mM Glucose, 100 µM hypoxanthine, 1.77 mM sodium bicarbonate ($NaHCO_3$) and 58.6 µg/ml gentamicin sulfate at 37 °C. Cultures were split and expanded/harvested after every two days to sustain parasitemia around 5%. Hematocrit was kept at 2.5 - 3% by supplementing newly washed fresh human RBCs. Parasitemia was tracked every day using slide preparation and Giemsa staining of blood smear.

2.2. Parasite Harvest

30 ml of Pf3D7 culture was spun down at 4000rpm for 4 minutes(min) at 4^0C to pellet down RBCs. Wash given with 1X PBS at 4000rpm for 5 minutes at 4^0C. RBC pellet mixed with 7ml of 0.16% saponin solution diluted with 3ml of 1X PBS. Vortex for 15 seconds. Incubated at 37^0C for 15 min. Centrifuge at 6000rpm for 15 min at 4^0C. The parasite pellet was obtained and treated with 3 washes of 1X PBS till the supernatant becomes clear. Pellet stored at –80^0C till use. The volumes of PBS and saponin were changed depending on the volume of culture used for harvest.

2.3. Preparation of Pf3D7 nuclear extract (NE)

Harvested parasite pellets were resuspended in 5X volume of lysis buffer (10mM HEPES pH 7.9, 10mM KCl, 0.1mM EGTA, 0.1 mM EDTA, 1mM Dithiothreitol (DTT), 1X protease inhibitor cocktail (PIC, Roche) and 10% IGPEAL) and incubated for 30 min at 4^0C/ice. Complete parasite lysis and homogenization were achieved by providing 200 strokes in a pre-chilled Dounce homogenizer. The lysate was centrifuged at 14000 rpm for 10min at 4^0C. The supernatant represents the cytoplasmic extract which was isolated and stored at –80^0C, while the intact nuclei are pelleted down. 3 times filtered 1X PBS wash was given to the nuclei pellet with centrifugation at 14000 rpm/5min/4^0C. The nuclei pellet resuspended in 3X volume of high salt nuclear extraction buffer (20mM HEPES pH 7.9, 800mM NaCl, 1mM EGTA, 1mM EDTA, 1X PIC, 1mM DTT) and incubated with vigorous shaking at 4^0C for 30 min. Centrifuge for 30 min at 14000rpm/10min/4^0C, the supernatant represents nuclear extract which was isolated and stored at –80^0C. The pellet was again

resuspended in low salt nuclear extraction buffer (20mM HEPES pH 7.9, 400mM NaCl, 1mM EGTA, 1mM EDTA, 1X PIC, 1mM DTT) and in the same way for high salt extraction, the nuclear lysate was isolated and stored at -80^{0}C. The nuclear extraction was validated by western blot and probing the membrane with an anti-PfHP1 antibody. The concentration of the protein in the lysate was estimated by Bradford assay.

2.4. Primers used for the pull-down of the proteins interacting with biotin-tagged insulator-like motif

No.	Primer Name	Description	Primer Sequence
1	KK595	biotin_*var*_insulator_FP	AATGTGTATATGTATGTGTGAAAA
2	KK596	biotin_srambled_FP	AAATGTATGGTATTGTGTGTAAAA
3	KK598	*var*_insulator_RP	TTTTCACACATACATATACACATT
4	KK599	scrambled_RP	TTTTACACACAATACCATACATTT
5	KK840	*var*_conact_FP	GACTTTATTGTGTATATGTATGTGTGTTCTT TGTATGTATGTGTTCTTTGTGTATATGTGA
6	KK841	biotin_*var*_insulator_FP	AATGTGTATATGTATGTGTGAAAA
7	KK842	scr_concat_FP	GACTTTAGTTTGGTTAACACATGTTGTTTTT AAGTTTGTTTTGGGTGTGATTTGTTTGATT
8	KK843	biotin_scr_FP	AAGTATGTTGTTTAATGTGGAAAA
9	KK844	*var*_conact_RP	TAAAGTCTCACATATACACAAAGAACACATA CATACAAAGAACACACATACATATACACAA
10	KK845	biotin_*var*_insulator_RP	TAAAGTCTTTTCACACATACATATACACATT
11	KK846	scr_concat_RP	TAAAGTCAATCAAACAAATCACACCCAAAA CAAACTTAAAAACAACATGTGTTAACCAAAC
12	KK847	biotin_scr_RP	TAAAGTCTTTTCCACATTAAACAACATACTT

Table 1: List of the primers used in the pull-down experiment. FP: forward primer, RP: Reverse primer

2.5. Identification of proteins associated with insulator-like motif in *var* gene intron (Protocol adapted from Gurung et al., 2006; Jutras et al., 2012)

BS buffer, 5X: 50 mM HEPES, 25 mM $CaCl_2$, 250 mM KCl, 60% (v/v) glycerol

THES buffer: 50 mM Tri-Cl, pH 7.5, 10 mM EDTA, 20% (w/v) sucrose, 140 mM NaCl, 0.7% (v/v) PIC, 0.1% (v/v) Phosphatase Inhibitor Cocktail

B/W buffer, 2X: 10 mM Tris-Cl, pH 7.5, 2 M NaCl, 1 mM disodium EDTA

BS/THES binding washing buffer: 13.3 ml THES buffer, 6 ml BS buffer, 10.7 ml NFW, using a 0.2-µm filter, filter sterilize the final mixture

Elution buffer: 25 mM Tris-Cl, pH 7.5, Varying concentration of NaCl: 100 mM, 200 mM, 300 mM, 500 mM, 750 mM, and 1 M NaCl

TE buffer: 10mM Tris-Cl pH 8, 1mM EDTA

Primer annealing: We had ssDNA for the *var* insulator-like motif and a scrambled sequence. An equal volume of respective primers (100µM) was mixed in TE buffer to a final concentration of 10 µM. Annealing reaction was carried out with the thermocycling conditions: 98^0C for 2 min, followed by 60 cycles with -1.3^0C decrement in temperature for each cycle of 1 min, infinite hold at 10^0C. The annealing of primers was confirmed by an agarose gel run.

Biotinylated-DNA probe binding with Streptavidin agarose beads (Sigma):

Carefully and slowly mix the settled beads with suspension. Take 140ul of the whole suspension (70µl of beads). Centrifuge at 1200rpm/2min to discard the supernatant. 3 times wash given with 500µl of 2X B/W buffer. Finally beads resuspended in 300µl of 2X B/W buffer. Add 60ug of biotinylated DNA probe and incubate with rolling for 45 min. Centrifuge, and remove supernatant. Probe-bead complex washed thrice with 400µl TE buffer. Two times wash was given with 500µl BS/THES buffer and a final wash with 500µl BS/THES buffer supplemented with 10µg/ml salmon sperm DNA.

DNA probe-bead complex incubation with NE:

Add 200µl of BS/THES buffer and 300µg of NE supplemented with 100µg of salmon sperm DNA as competitor DNA. Incubation with rolling at 4^0C for 1hr. Centrifuge and save the flowthrough. Three times wash with 500µl of BS/THES buffer supplemented with 10µg/ml salmon sperm DNA. Two times wash with 500µl of BS/THES buffer.

Elution of the bound proteins:

Add 120µl of 100mM elution buffer. Roll for 3-5min at RT. Centrifuge and save the elution. Repeat the elution step with elution buffers of NaCl concentrations, 300mM, 500mM, 750mM and 1M. Store the elutions at -80^{0}C till to be processed further.

Note: All the centrifugation steps were done at 1200rpm/2min at 4^{0}C.

2.6. Identification of proteins associated with insulator-like motif in *var* gene intron (Protocol adapted from Flueck et al., 2010)

Phosphorylation and primer annealing:

Phosphorylation of the 5' end and annealing of the primers was done in a single reaction. The phosphorylation reaction was set up according to NEB protocol, non-radioactive phosphorylation with T4 PNK. Thermocycling conditions: 37^{0}C for 30min, 98^{0}C for 2 min, followed by 60 cycles with -1.3^{0}C decrement in temperature for each cycle of 1 min, infinite hold at 10^{0}C. The annealing of primers was confirmed by an agarose gel run.

Reaction setup for concatenation of biotin-tagged and non-biotinylated primers to get multiple repeats of insulator-like motif:

dsDNA (300pmole biotinylated and 900pmole non-biotinylated DNA)	200µl
ATP (10mM)	5µl
T4 DNA ligase buffer	30µl
T4 DNA ligase (NEB)	10µl
NFW	55µl

Table 2: Reaction set-up used for the concatenation reaction.

Ethanol precipitation of DNA:

To the DNA, add 0.1 volumes of 3M sodium acetate and 2.5-3 volumes of ice-cold 100% ethanol. Vortex to mix thoroughly. Precipitate at –20^{0}C for 1hr/ -80^{0}C overnight. Centrifuge at full speed for 30min/4^{0}C. Wash the pellet twice with 500µl ice-cold 70% ethanol. Dry the pellet and resuspend it in the appropriate volume of NFW.

Pull-down Protocol:

First, the NE with ~600-700mM salt concentration was diluted to ~160-200 mM salt concentration by adding an appropriate volume of dilution buffer (20mM pH 7.9, 1mM EDTA, 0.025mM $ZnCl_2$, 2mM $MgCl_2$, 1X PIC, 2mM DTT, 0.1% Triton-X 100, 30µg/ml salmon sperm DNA). Gently mix the settled streptavidin agarose beads suspension and take 130µl of the whole suspension. Centrifuge and discard the supernatant. Beads to be used for preclearing were washed with nuclear extraction buffer while the beads to be used for DNA binding were washed with TE buffer three times. Incubate diluted NE with beads for 3hr at RT on a rotating wheel platform. Incubate 420µl of 10µM biotinylated dsDNA with beads for 3 hr on a rotating wheel platform. Save the diluted NE after preclearing. Wash the beads-DNA complex with 400 µl of TE buffer 3 times. Overnight incubation of NE with DNA-coated beads at 4^0C on a rotator. Beads-DNA-protein complex washed twice with 1 ml of wash buffer 1 (20mM HEPES pH 7.9, 160mM KCl, 1mM EDTA, 0.025mM $ZnCl_2$, 2mM $MgCl_2$, 1X PIC, 2mM DTT, 0.1% Triton-X 100, 20µg/ml salmon sperm DNA). A final wash was given with 1 ml of wash buffer 2 (same composition as wash buffer 1 except salmon sperm DNA). Finally, two elutions each of 150µl were taken in elution buffer (20mM HEPES pH 7.9, 2M KCl, 2mM EDTA, 20% Glycerol). Elutions were stored at –80^0C.

2.7. Primers used for the PCR amplification of genes of possible insulator-like proteins

No.	Primer Name	Description	Primer sequence	Tm (^{0})
1	KK853	PF3D7_1132300_NABP_FP	GAAATATATCACCTAGGATGGATATGAATTTCGGATTCTGTGAT	61
2	KK854	PF3D7_1132300_NABP_RP	TACTAGATCCGCTAGCTGGAATACTTATTATGTCCTCATATTG	
3	KK855	PFGBP2_PDC2_Infusion_FP	GAAATATATCACCTAGGATGTCGATGGAAAATAATAGCC	60
4	KK856	PFGBP2_PDC2_Infusion_RP	TACTAGATCCGCTAGCTTCGTTATTTTCATAATTAACTTTTAATATATATCC	
5	KK857	PFALBA1_PDC2_Infusion_FP	GAAATATATCACCTAGGATGAAGAAAGATAGAGAACCAATAGACG	63

6	KK858	PFALBA1_PDC2_Infusion_RP	TACTAGATCCGCTAGCTGATAAAGCTCTTCCACCTCTGAATCC	
7	KK859	PFALBA3_PDC2_Infusion_FP	GAAATATATCACCTAGGATGGCAAGCACAGAAGAAGTATCTC	
8	KK860	PFALBA3_PDC2_Infusion_RP	TACTAGATCCGCTAGCGTTTGCTACAAAATCTGGGTGT	62
9	KK861	PF3D7_0709500_NABP_FP	GAAATATATCACCTAGGATGGATAATATGGATGAAGAG	
10	KK862	PF3D7_0709500_NABP_RP	TACTAGATCCGCTAGCATATGATGGTATGAAAACATCCG	56
11	KK863	PFHMGB2_PDC2_Infusion_FP	GAAATATATCACCTAGGATGGCTTCAAAATCTCAAAAGAAAG	
12	KK864	PFHMGB2_PDC2_Infusion_RP	TACTAGATCCGCTAGCTTCTTGATTTTTCTTTCTATATTCTTC	58
13	KK865	PFALBA2_PDC2_Infusion_FP	GAAATATATCACCTAGGATGCCGGGTAGCACAAAG	
14	KK866	PFALBA2_PDC2_Infusion_RP	TACTAGATCCGCTAGCATTTCTTTTTTCATATCTTCTATTATTCCACATG	63
15	KK867	PFALBA4_PDC2_Infusion_FP	GAAATATATCACCTAGGATGGAAAATGATAAAAAACATAACCAA	
16	KK868	PFALBA4_PDC2_Infusion_RP	TACTAGATCCGCTAGCTTGCTCATTGTGATTGAAAACTTG	61

Table 3: List of primers required for PCR and cloning purposes. FP: Forward primer, RP: Reverse primer

2.8. PCR Amplification, Restriction Digestion and Cloning

For the cloning purpose, genes were PCR amplified using the MRA cDNA library and Phusion DNA polymerase at the respective annealing temperatures. PCR amplified gene product and pDC2-cam-mRFP-2A-GFP-bsd-attP (mentioned as pDC2 hereafter) overexpression plasmid were digested using restriction enzymes NheI-HF and AvrII. After overnight ligation reaction, the transformation was carried out in E-coli DH5α cells. Following the alkaline lysis of the selected colonies, screening and clone confirmation was done by PCR and restriction digestion. Finally, the clone was confirmed by sequencing.

2.9. Transfection of the clones in *P. falciparum* parasites

After midi preparation, 100µg of cloned plasmid ethanol precipitated and dissolved in 10µl TE buffer. Before transfection, it was dissolved in 100µl of P3 primary buffer. Prior to transfection, Pf3D7 culture was maintained at 5-7% parasitemia and strictly synchronized by performing percoll density gradient centrifugation 2-3 times. At the time of transfection, the segmented schizonts obtained after percoll centrifugation were mixed with 100µl of P3(primary) solution mix. The Lonza Nucleofector 4D machine was used for nucleofection, with the FP 158 pulse program. The cells were quickly moved to a T25 flask containing 5 ml of media and 200 µl of fresh RBCs and they were then kept at 37^0C for 1.5 hr in a shaker incubator. 24 hr post-transfection drug (Blasticidin S HCl) selection was started. The culture was supplemented with fresh media for the first 6 days, and after that media changed on alternate days. 2 weeks post-transfection, the culture split into two with the addition of fresh RBCs. On the fourth-week post-transfection 100µl of fresh RBCs were added to the culture. The appearance of transgenic lines was checked three times a week via blood smear.

2.10. Immunoprecipitation (IP) of PfHP1

The expression and purification of PfHP1 protein were done by Mamatha. For the IP, protocol from Thermofisher Scientific for IP using Dynabeads protein A was followed.

Beads wash buffer: 1X PBS + 0.02% Tween-20

Immunoprecipitation wash buffer: 25mM Tris pH 7.9, 5mM $MgCl_2$, 10% glycerol, 100mM KCl, 0.1% IGEPAL, 0.3mM DTT

Elution buffer: 50mM Glycine pH 2.8.

Preparing Dynabeads (Invitrogen): Resuspend Dynabeads in the vial (vortex for more than 30 sec or rotate for 5 min). Transfer 50µl Dynabeads to a tube by placing the tube on a magnet separate the beads and remove the supernatant. Two washes were given with beads wash buffer.

Bind antibody: Add antibody to the Dynabeads and incubate at RT for 1 hour with rotation. Remove supernatant and one wash given with bead wash buffer.

Immunoprecipitate the target antigen: NE of *P. falciparum* resuspended in Dynabeads-antibody complex. Incubate overnight at 4^0C with rotation. Next day, after

centrifugation flowthrough was stored and 3 times wash was given with immunoprecipitation wash buffer.

Elute the target antigen: Add 30µl of the elution buffer and resuspend the Dynabeads-antibody-antigen complex gently with the pipette. Incubate with rotation for 2 min at RT. Save the elution.

2.11. Mass Spectrometry

The elutions obtained after the biotin-tagged oligo pull-down experiment were processed for mass spectrometry. First, for reduction, 100mM DTT in TEAB buffer was added in elutions such that its final concentration is 10mM. Incubation in a thermomixer at 60^0C/45min at 800rpm. Samples were then processed with 100mM iodoacetamide (IAA) for alkylation, such that its final concentration is 10mM. Incubation at room temperature for 15 min in dark. Next, the samples were treated with 1µg of trypsin with an overnight incubation at 37^0C at 800rpm mixing. To stop the trypsin action 50% formic acid (FA) is added such that its final concentration is 1%. Desalting was done for the samples using Zip-tips made from desalting C18 membrane. Finally, the elutions were taken using 50µl of 60% acetonitrile in 0.1% Trifluoroacetic acid and vacuum dried in Centrivap. Samples were then sent for mass spectrometry (SCIEX QTOF). Results were analyzed using Protein Pilot software.

2.12. Motif analysis

MEME Suite was used to search for the motifs. Intron sequences of the 60 *var* genes were used as input database. Parameters used were: Maximum number of motifs: 5, Minimum motif width: 6, Maximum motif width: 24, Minimum sites per motif: 2 and Maximum sites per motif: 60

2.13. ChIP Sequencing Data Analysis

FastQC was used to evaluate the raw fastq sequencing files' quality, and Trim_Galore was employed to clear adapter sequences and low-quality reads. Reads were mapped to the reference genome of *P. falciparum* 3D7 using Bowtie2. SAMtools was used to convert the generated SAM files to BAM files. The MACS2 software was used for peak calling using input BAM and ChIP files for the particular case. Using IGV software the peaks were visualized and density graphs were generated in SeqMINER. Morpheus (Broad Institute) was used for generating heat maps.

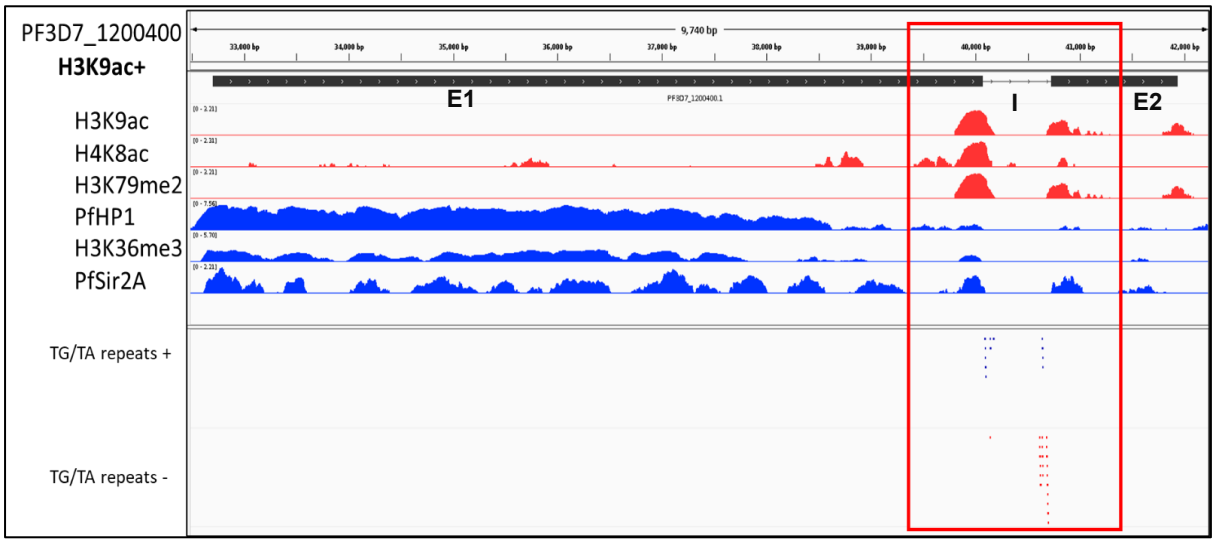
3. Results

3.1. ChIP sequencing analysis shows differential enrichment of histone modifications at the intronic region of *var* genes

The epigenome plays an important role in governing the spatiotemporal expression of genes. The presence of activating or repressive chromatin modifications determines the transcriptional state of a gene. In *P. falciparum* the epigenomic studies suggest that the active histone modifications positively correlate with the active transcription of the gene while the heterochromatinization via PfHP1 silence the gene expression of the telomeric and subtelomeric genes. To inspect the epigenomic landscape in *P. falciparum*, we analyzed the in-house generated paired-end ChIP-sequencing data for H3K9ac, H4K8ac, H3K79me2, H3K36me3 histone modifications and, PfHP1 and PfSir2A (histone deacetylase) proteins for the parasite population harvested at 18 hrs post RBC invasion i.e. the mature ring or early trophozoite stages. H3K9ac and H4K8ac are the activating histone modifications while PfHP1 marks the repressed genes. The role of H3K79me2 modification is not very clear. H3K36me3 is associated with both active and silent genes but it is mainly involved in the repression of CVG gene families. Analysis in Integrative Genomics Viewer (IGV) showed that PfHP1 protein is enriched at the telomeric regions of all the chromosomes and marks all the *var* gene loci. As we were specifically interested in *var* genes, we systematically analyzed this ChIP-sequencing data in IGV and seqMINER to understand the histone modifications landscape on each of the ~60 *var* genes. Along with the IGV analysis, we obtained the read density values for H3K9ac and PfHP1 protein on each *var* gene using seqMINER and plotted it as a heatmap in Morpheus (Online software by Broad Institute) (Fig. 6). In the analysis we observed that most of the *var* genes (44 in total) show enrichment of PfHP1 protein on the entire *var* gene body while almost no enrichment of H3K9ac and only sparse occupancy of H4K8ac. H3K36me3 modification and PfSir2A protein showed an enrichment pattern similar to PfHP1 (Fig. 6B and Fig. 7B). Interestingly, 17 of the remaining *var* genes showed enrichment of H3K9ac and other activating histone modifications at the junction of exon1 and intron as well as intron and exon2 (Fig. 6A and Fig. 7A). Along with this, PfHP1 protein is depleted from the intron as well as the exon2 of these *var* genes. Hence, we classified the *var* genes as H3K9ac+ and H3K9ac- *var* genes based on either the presence or absence of H3K9ac at the boundaries of the intron (Table 4). Out of the total 61 *var*

genes considered for this analysis, 17 are H3K9ac+ while 44 are H3K9ac- ones. Not all the 17 H3K9ac+ *var* genes showed equivalent enrichment of activating histone modifications. PF3D7_1200600 (*var*2csa), a *var* gene candidate responsible for severe placental malaria, had the highest enrichment of H3K9ac while the others showed lower and varying peak values for H3K9ac at the intron-exon junctions in IGV. Next, we plotted the average read density values for H3K9ac, H4K8ac modifications and PfHP1 protein at 1 kB region on either side of the transcription start site (TSS), intron start, intron end and transcription termination site (TTS) of the H3K9ac+ and H3K9ac- *var* genes. These plots also indicate the specific enrichment of H3K9ac and H4K8ac at the intron start and end of H3K9ac+ *var* genes along with the depletion of PfHP1 while the opposite results are observed for H3K9ac- *var* genes (Fig. 8). Followed by these peculiar observations, we performed the literature study to find out what is known about the *var* gene introns or their role in the mutually exclusive expression of *var* genes. We came across a study by Avraham et al. 2012, wherein through the MEME motif analysis they had noticed the presence of an insulator-like motif at the ends of *var* gene intron. Next, in IGV we simultaneously observed for an insulator-like motif and the repeats of TG/TA along with visualizing the bedgraph files of ChIP-sequencing. Through this analysis, we noticed that the insulator-like motifs are present at the ends of the intron, the region occupied by histone modifications in H3K9ac+ *var* genes (Fig. 6). This differential enrichment of chromatin modifications at the intronic region of *var* genes is a novel observation which we are reporting for the first time. Considering the bidirectional promoter activity of the *var* gene introns and its possible role in the mutually exclusive expression of *var* genes, we hypothesized that the insulator-like motifs and differential enrichment of histone modifications at the region surrounding them could regulate the promoter activity of the intron, *var* gene expression and antigenic *var*iation through some yet unidentified mechanism. This observation highlights the importance of insulator-like motifs and promotes the study further to identify possible insulator-like proteins.

A)



B)

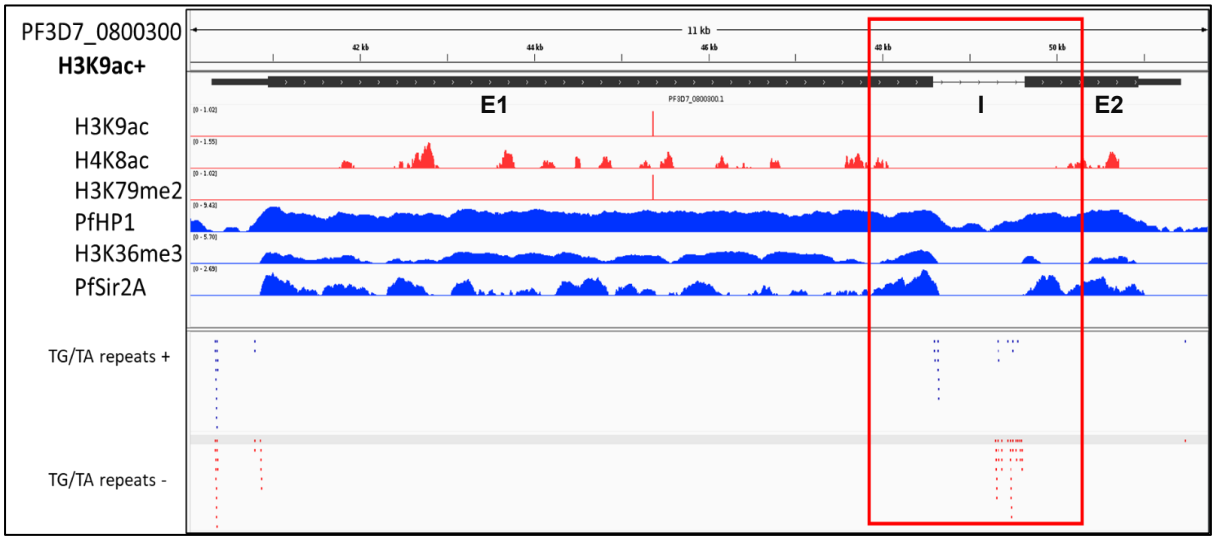
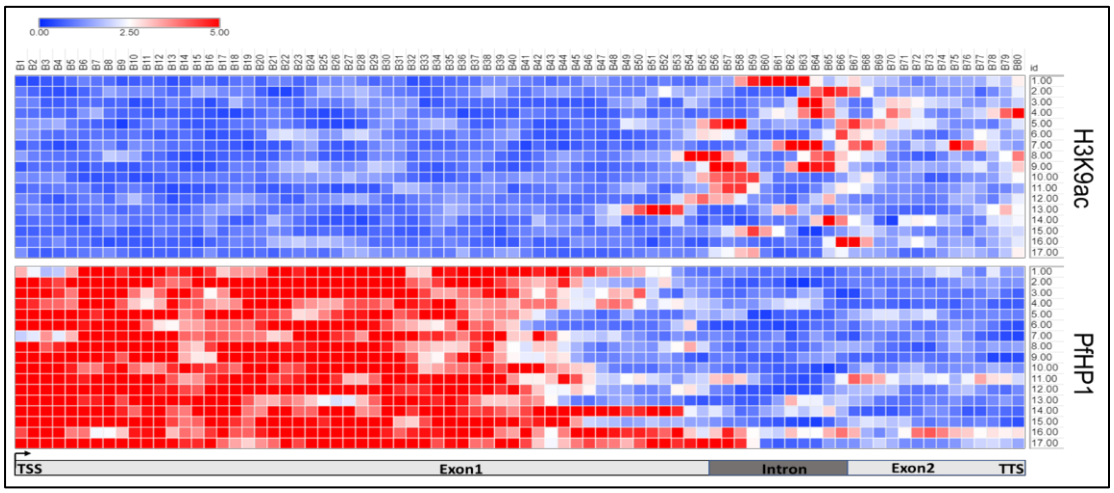


Figure 6: Occupancy of histone modifications on var genes PF3D7_1200400 and PF3D7_0800300. IGV images displaying the occupancy of H3K9ac, H4K8ac, H3K79me2, H3K36me3 histone modifications and PfHP1, PfSir2A proteins on A) Pf3D7_1200400, an H3K9ac+ *var* gene and B) Pf3D7_0800300, an H3K9ac- *var* gene. These are the representative images for H3K9ac+ and H3K9ac- *var* genes. Shown below in the images is the location of insulator-like motifs/TG-TA repeats on the *var* genes. The red box highlights the intronic region of genes. Arrow indicates the direction of gene transcription. E1: Exon1, I: Intron and E2: Exon2.

A) H3K9ac+ var genes



B) H3K9ac- var genes

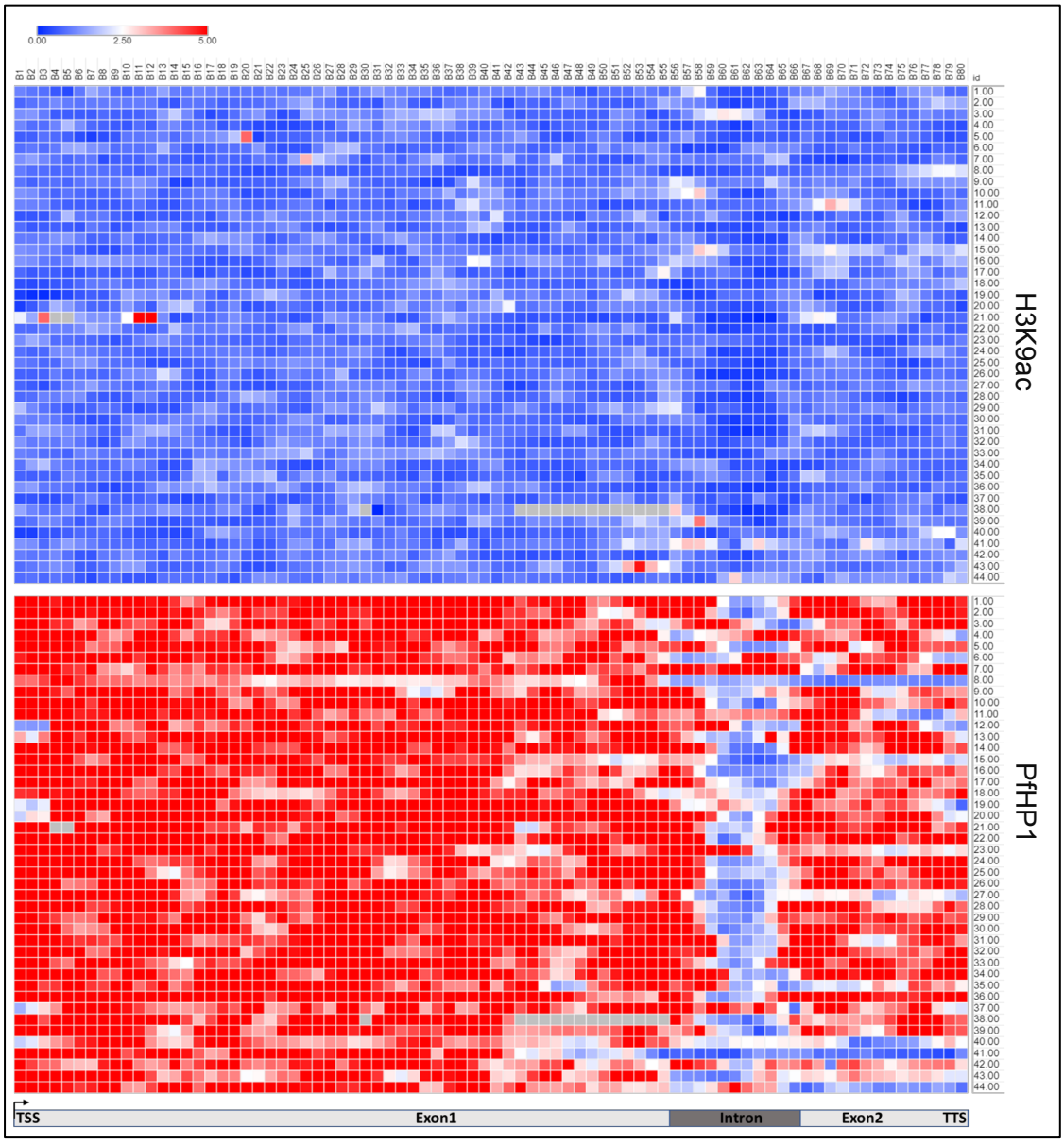


Figure 7: Heatmap representation of PfHP1 and H3K9ac enrichment on var genes. Heatmap demonstrating the enrichment of H3K9ac modification and PfHP1 protein on (A) 17 H3K9ac+ and (B) 44 H3K9ac- *var* genes. Each gene is divided into 80 bins and the average density value for the bin is used for plotting the heatmap. Blue indicates the lowest enrichment while red indicates the maximum enrichment of histone modification/protein at the given bin.

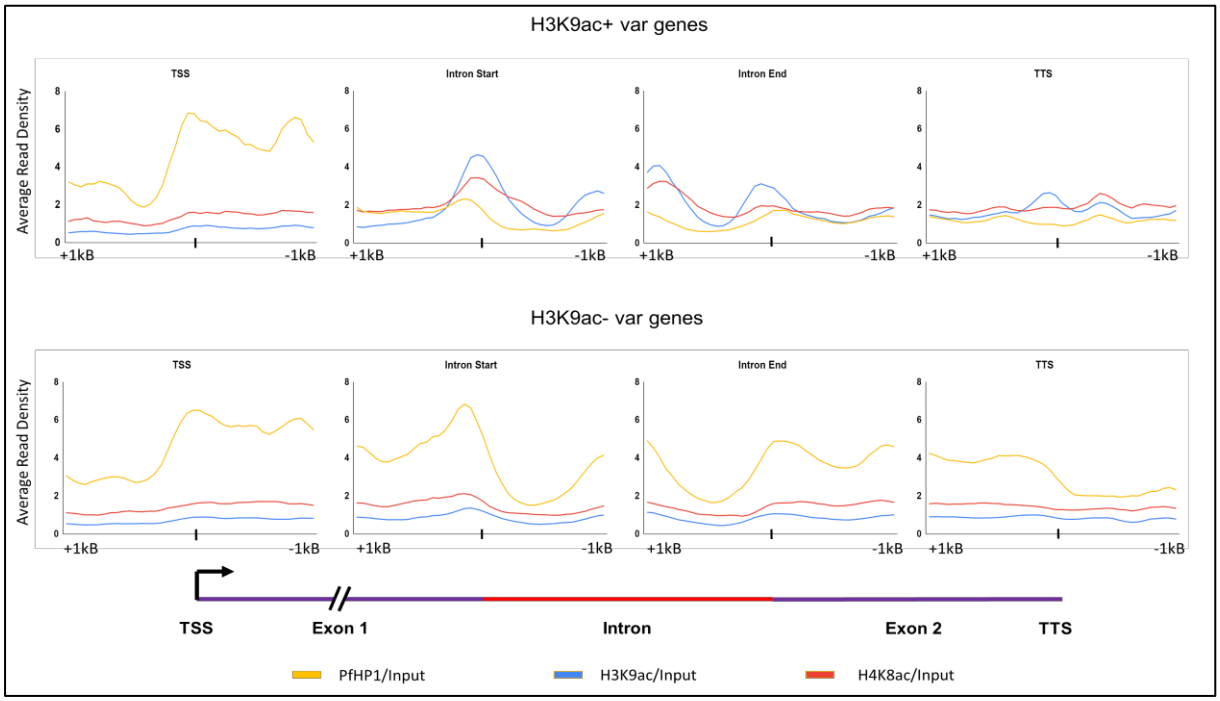


Figure 8: Average enrichment of H3K9ac, H4K8ac and PfHP1 protein on H3K9ac+ and H3K9ac- *var* genes. Plots showing the average read density of H3K9ac, H4K8ac and PfHP1 protein at 1 kB region on either side of the TSS, intron start, intron end and TTS of H3K9ac+ and H3K9ac- *var* genes. The average read density is normalized over ChIP-sequencing input (control used in ChIP-sequencing).

H3K9ac+ *var* genes	**H3K9ac- *var* genes**		
PF3D7_0100100	PF3D7_0100300	PF3D7_0632800	PF3D7_1150400
PF3D7_0200100	PF3D7_0115700	PF3D7_0700100	PF3D7_1219300
PF3D7_0300100	PF3D7_0223500	PF3D7_0711700	PF3D7_1240300
PF3D7_0400100	PF3D7_0324900	PF3D7_0712000	PF3D7_1240400
PF3D7_0400400	PF3D7_0412400	PF3D7_0712300	PF3D7_1240600

PF3D7_0425800	PF3D7_0412700	PF3D7_0712400	PF3D7_1240900
PF3D7_0500100	PF3D7_0412900	PF3D7_0712600	PF3D7_1255200
PF3D7_0800100	PF3D7_0413100	PF3D7_0712800	PF3D7_1300100
PF3D7_0800200	PF3D7_0420700	PF3D7_0712900	PF3D7_1300300
PF3D7_0809100	PF3D7_0420900	PF3D7_0733000	PF3D7_1373500
PF3D7_0900100	PF3D7_0421100	PF3D7_0800300	
PF3D7_0937800	PF3D7_0421300	PF3D7_0808600	
PF3D7_1100100	PF3D7_0426000	PF3D7_0808700	
PF3D7_1100200	PF3D7_0600200	PF3D7_0833500	
PF3D7_1200100	PF3D7_0600400	PF3D7_0937600	
PF3D7_1200400	PF3D7_0617400	PF3D7_1000100	
PF3D7_1200600	PF3D7_0632500	PF3D7_1041300	

Table 4: List of all H3K9ac+ and H3K9ac- *var* genes.

3.2. In silico analysis to identify the possible insulator-like proteins in *Plasmodium falciparum*

To date, no insulator protein has been identified in *P. falciparum*, but the genome organization shows well-structured euchromatin and heterochromatin compartmentalization as well as *var* genes possess hypothetical insulator-like motif raising the possibility of the presence of insulator-like proteins regulating higher-order genome organization and highly controlled gene expression. To get an idea about the possible insulator-like proteins in *P. falciparum*, we decided to first find the homologs of already known and characterized insulator proteins from other model organisms. If the homologs exist, they may be conserved during evolution and carry out similar roles in *P. falciparum* and could be the putative insulators. To proceed, we first conducted a literature search to compile a list of the known insulator proteins and examine their functions (Table 5). Insulators were first discovered in *Drosophila melanogaster* following research on the white eye mutation that resulted from the position effect variegation phenomena. Subsequent to that, many insulator elements and associated insulator proteins were found in *Drosophila*, and their insulator activities have since been thoroughly investigated through a number of studies. Because of this, our list of insulator proteins majorly includes proteins from *Drosophila melanogaster* while some proteins are from humans and yeast. However, several of the insulators discovered in *Drosophila* that are involved in insulator function appear to be conserved in mammals. Followed by the literature search, we performed a BLAST search against the *P. falciparum* proteome database for each insulator protein. A major hit in the BLAST search is a telomere-repeat binding zinc finger protein (PfTRZ), which is a distant homolog of the transcription factor TFIIIA. This protein was found to be binding to the telomeric repeats, 5S rDNA loci and sub-telomeric *var* genes (Bertschi et al., 2017). We reanalyzed the available ChIP-sequencing data for PfTRZ and found that it is present on all the *var* genes, and its distribution is similar to the distribution of PfHP1 on H3K9ac- *var* genes. We specifically examined the enrichment of PfTRZ on H3K9ac+ and H3K9ac- *var* genes and didn't find any difference (Fig. 9).

Interestingly, *P. falciparum* doesn't have the homolog of CTCF, a major insulator in higher eukaryotes. Along with this, the homologs of other insulator proteins and CTCF-interacting proteins are absent in *P. falciparum*. Bromodomain proteins were obtained as hits for the BLAST against BRD1 protein which is thought to be the possible

insulator-like protein in *Arabidopsis thaliana*. Other proteins like GCN5, ATP-dependent RNA helicase, general transcription factor IIIC polypeptide 5, and nucleoporin were obtained as hits, but the query coverage and percentage identity are very low. However, homologs of the Structural Maintenance of Chromosomes (SMC) complex (cohesin and condensin) subunits are present in *P. falciparum*. Current studies in higher eukaryotes are implying that the SMC complexes could function as the interactors of insulators or act as insulators in regulating higher-order genome organization and gene expression. In *P. falciparum*, the role of cohesin and condensin complexes in mitosis, chromosome segregation and condensation are being studied but their role in genome organization has not been studied so far (Table 6). In conclusion, the in-silico analysis identifies PfTRZ as the major hit and SMC complexes and PfTRZ protein could possibly be the putative candidates for insulator-like activity in *P. falciparum*.

No.	**Protein**	**Species**	**Homolog in *P. falciparum***	**Query coverage (%)**	**% identity**
1	**Su(Hw)**	*Drosophila melanogaster (Drosophila)*	Telomere repeat-binding zinc finger protein (PfTRZ)	35	27
2	**CTCF**	*Drosophila*	-	-	-
3	**BEAF-32**	*Drosophila*	-	-	-
4	**Ibf1**	*Drosophila*	-	-	-
5	**Ibf2**	*Drosophila*	-	-	-
6	**ZIPIC/CG7928**	*Drosophila*	PfTRZ	41	27.52
7	**Pita**	*Drosophila*	PfTRZ	35	25
8	**Dwg/Zw5**	*Drosophila*	PfTRZ	37	29
9	**GAF**	*Drosophila*	-	-	-
10	**Cp190**	*Drosophila* (Bridging insulator elements)	-	-	-
11	**Mod/mdg4**				
12	**Elba1**	*Drosophila*	-	-	-

13	**Elba2**	*Drosophila*	-	-	-
14	**Elba3**	*Drosophila*	-	-	-
15	**Shep**	*Drosophila*	-	-	-
16	**TFIIIC**	*S. cerevisiae, S.pombe*	General transcription factor IIIC, polypeptide 5	13	35.23
17	**Polycomb protein**	*Drosophila*	-	-	-
18	**YY1**	*Homo Sapiens*	-	-	-
19	**dTopors**	*Drosophila*	RING zinc finger protein	4	30.19
20	**Prdm5**	*Homo Sapiens*	PfTRZ	46	36.36
21	**Chromator**	*Drosophila*	Conserved Plasmodium protein, Unknown function	17	24.85
22	**DREF**	*Drosophila*	-	-	-
23	**L(3)mbt**	*Drosophila*	-	-	-
24	**-CAP-H(2)**	*Drosophila*	Condensin complex subunit 2, putative	9	27.71
25	**Barren**	*Drosophila*	Condensin complex subunit 2, putative	9	27.71
26	**Rm62 (ATP dependent RNA helicase)**	*Drosophila*	Many hits ATP dependent RNA helicase	~60	-
27	**Bptf (interactor with CTCF)**	*H. sapiens*	PfGCN5, Bromodomain proteins	2	38.82
28	**RAP1**	*S. cerevisiae*	-	-	-
29	**REB1**	*S. cerevisiae*	-	-	-
30	**Nup98**	*Homo sapiens*	Nucleoporin, Nup221 putative	7	37.91
31	**Insv**	*Drosophila*	-	-	-
32	**BRD1 (Proposed)**	*Arabidopsis thaliana*	Bromodomain protein 1, Bromodomain protein 2	21	31

Table 5: List of the insulator proteins discovered and studied in other model organisms with their homologs in *Plasmodium falciparum* along with the obtained query coverage and percentage identity.

1	**SMC2 condensin**	*P. berghei*	**SMC2**
2	**SMC4 condensin**	*P. berghei*	**SMC4**
3	**SMC1 (cohesin)**	*P. falciparum*	**SMC1**
4	**SMC3 (cohesin)**	*P. falciparum*	**SMC3**

Table 6: Condensin and cohesin complex subunits in *Plasmodium falciparum* and *Plasmodium berghei* (causative agent in rodents).

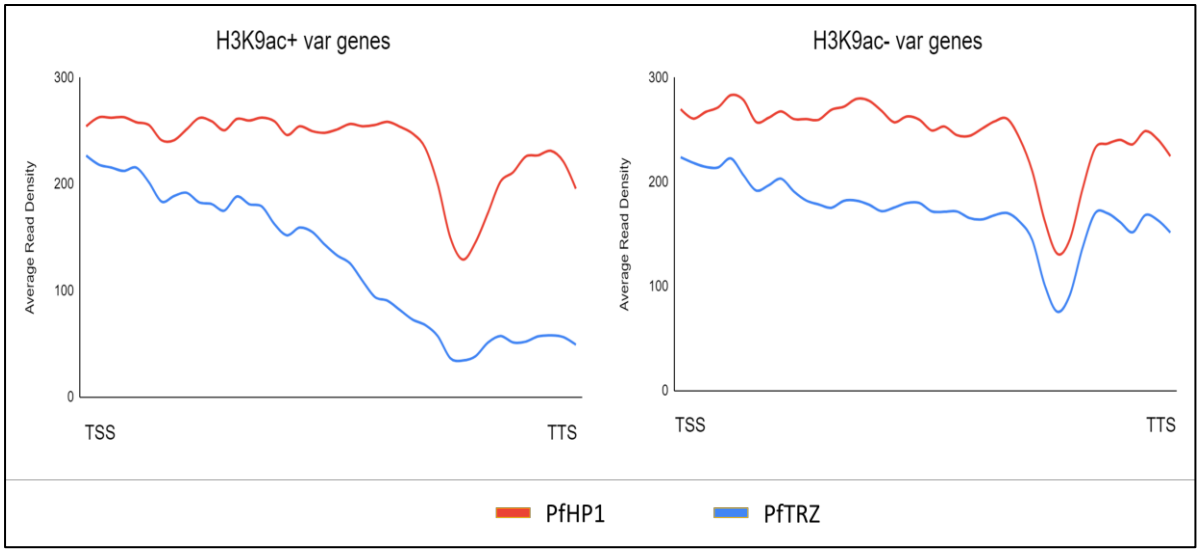


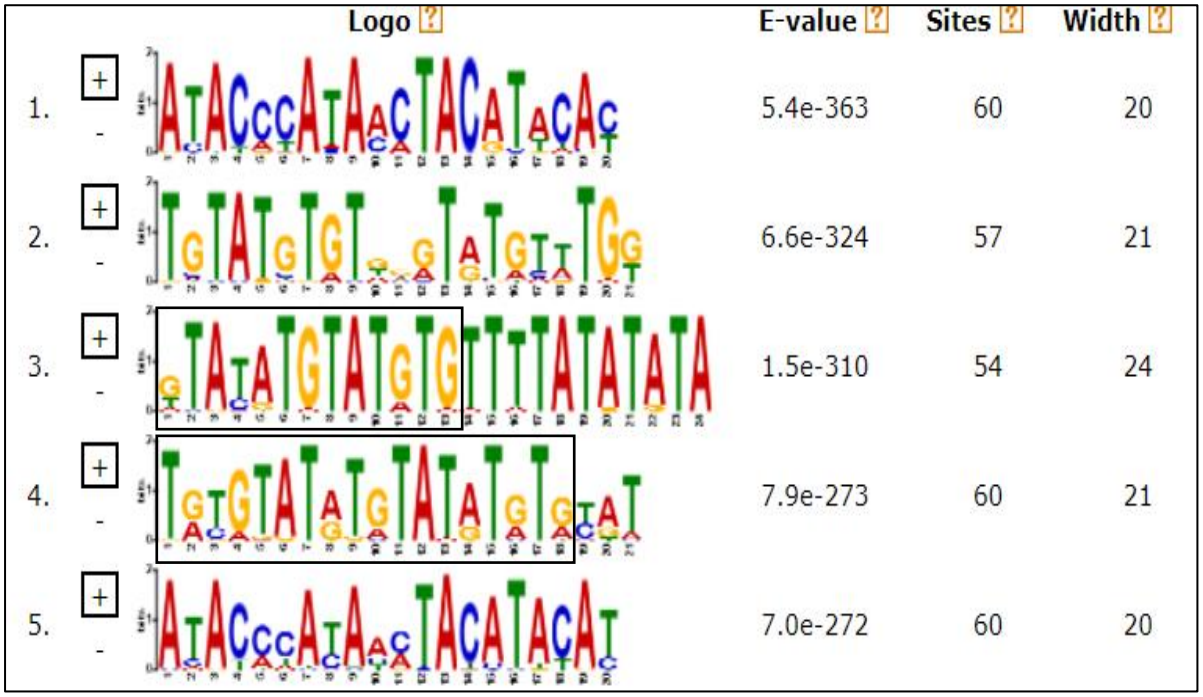
Figure 9: Occupancy of PfTRZ and PfHP1 on *var* genes. Plots showing the enrichment of PfTRZ and PfHP1 proteins on the H3K9ac+ and H3K9ac- *var* gene bodies. TSS and TTS mark the transcription start and termination sites respectively. PfTRZ ChIP-sequencing data is acquired from Bertschi et. al., 2017 and reanalyzed. In-house generated PfHP1 ChIP-sequencing data used for analysis.

3.3. Insulator-like motif interacts with the proteins from *P. falciparum* nuclear extract

Earlier studies have found the conserved motifs in the intronic region of the *var* genes and based on this insight we were also interested in critically analyzing the nucleotide sequence of these regions of the *var* genes to search for motifs. The intronic regions of *var* genes are highly AT-rich with almost 80-90% AT content. The interesting feature of the intronic region is the presence of repeats of TR (R is any purine base Adenine or Gaunine) at both ends of the intron. We performed the MEME (Multiple Em for Motif Elicitation) motif analysis for the intronic sequences of 60 *var* genes. We tried to generate motifs of varying lengths but specifically, the MEME motif search to identify the motifs of length 20-25 base pairs provided us with five different motifs (Fig. 10A). Motifs 1 and 5 are mostly the same in the nucleotide sequence except at the two base positions while motifs 2, 3 and 4 majorly have repeats of TG-TA. Figure 10B highlights the location of the motifs on the intron of *var* genes and all the motifs show clear enrichment at the ends of the intron. Motifs 1 and 5 are present in almost every *var* gene but they are present only at the end of the intron and hence are not of interest to study from the perspective of insulator-like elements and proteins. Though it could be possible that this motif has some role in the regulation of *var* gene expression. Among these motifs, motifs 3 and 4 caught our attention as they are similar to the insulator-like motif reported earlier and are present in the opposite orientation at the ends of the intron of all the *var* genes. As our analysis also identifies the presence of the motif identical to the insulator-like motif reported earlier, we decided to find the proteins interacting with the motif through in-vitro experimental approaches. The experimental plan involves the pull-down of the *P. falciparum* nuclear lysate proteins binding to the biotin-tagged DNA motif cross-linked to the streptavidin agarose beads. *P. falciparum* nuclear lysate was prepared and validated by probing for PfHP1 in a western blot (Fig. 11). Using a 5' biotin-tagged insulator-like motif (biotin-TGTGTATATGTATGTGTG), streptavidin agarose beads and *P. falciparum* NE we performed pull-down of the proteins interacting with the motif. As a control we carried out pull-down with a scrambled sequence (biotin-ATGTATGGTATTGTGTGT), a DNA sequence having the same nucleotides as the insulator-like motif but in a shuffled order. This experiment was performed three times. The first time, we followed a protocol adapted from Jutras et. al., 2012, while the next two times we followed the protocol adapted from Flueck

et. al., 2010. With the first pull-down experiment, we didn't get any specific protein as hits apart from histone 4 and elongation factor-1 alpha. With the repetition of the pull-down experiment for the second time we obtained 32 proteins as hit with the insulator-like motif and 12 proteins with the scrambled sequence (Fig. 12). In both pull-down with control and insulator-like motif, the top hits are RNA-binding proteins, single-stranded DNA binding proteins, Replication factor A1 and some histones. The major specific hits followed by the above ones for the pull-down with an insulator-like motif were chromatin-binding proteins and ALBA family proteins. After excluding the RNA-binding, translational and splicing associated proteins and the proteins common in pull-down with the scrambled sequence, we were left with the 11 proteins of interest (Table 7). Two of them are putative nucleic-acid binding proteins, four ALBA family proteins, architectural proteins PfGBP2, PfHMGB2 and ISWI chromatin remodeling complex ATPase (PfISWI), FanCJ-like helicase and protein SIS1. These proteins could be the major interactors of the insulator-like motif and could be involved in the insulator-like activity.

A)



B)

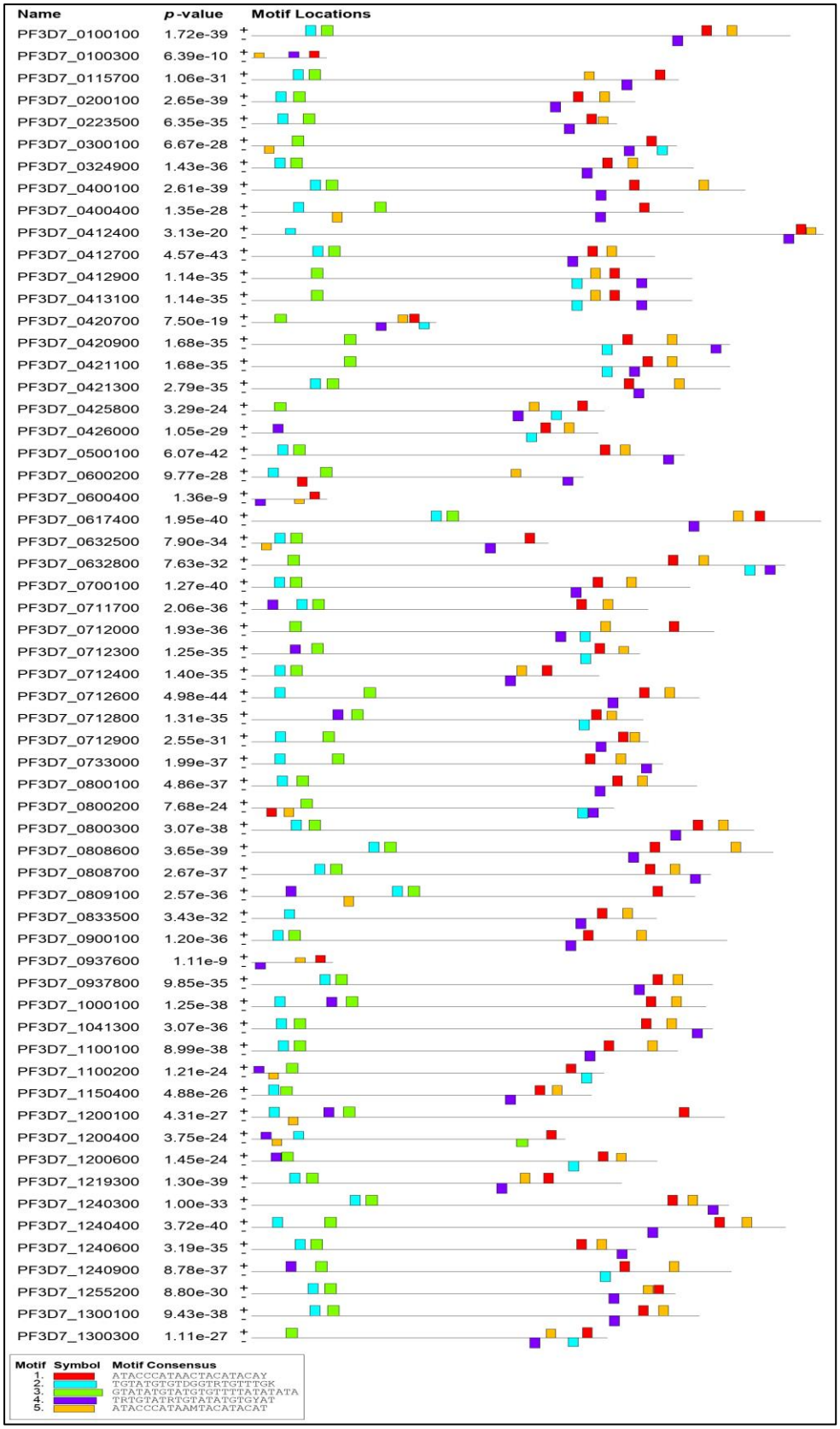


Figure 10: Motif analysis for the intronic region of *var* genes. A) The list of the motifs as identified in the MEME motif analysis with the E-value, number of sites and the width of the motifs. B) A map showing the location of the five motifs identified using

MEME motif analysis on the intron of *var* genes. The strand orientation of DNA is indicated by + and -. If the motif is present on the + strand it is displayed above the line while if it is on the – strand it is shown below the line.

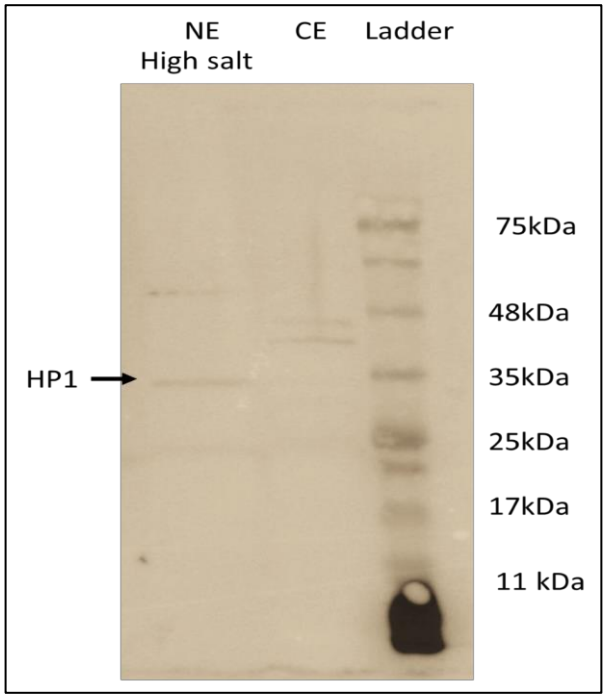


Figure 11: Validation of the NE preparation. Western blot image for the probing of PfHP1 in nuclear and cytoplasmic (NE and CE) extracts respectively. The molecular weight of the PfHP1 protein is 31kDa. N=2

Protein ID	Protein name	Peptide count
PF3D7_1132300	**nucleic acid binding protein, putative**	9
PF3D7_0709500	**nucleic acid-binding protein, putative**	8
PF3D7_1346300	**DNA/RNA-binding protein Alba 2 (PfALBA2)**	3
PF3D7_1006800	**G-strand-binding protein 2 (PfGBP2)**	8
PF3D7_0817900	**high mobility group protein B2 (PfHMGB2)**	7
PF3D7_0814200	**DNA/RNA-binding protein Alba 1 (PfALBA1)**	4
PF3D7_1347500	**DNA/RNA-binding protein Alba 4 (PfALBA4)**	2
PF3D7_1006200	**DNA/RNA-binding protein Alba 3 (PfALBA3)**	4
PF3D7_0624600	ISWI chromatin-remodeling complex ATPase	2
PF3D7_1408400	FANCJ-like helicase, putative	2
PF3D7_0213100	protein SIS1	2

Table 7: List of the proteins of interest shortlisted after the pull-down experiment with the biotin-tagged insulator-like motif along with the peptide count in the mass spectrometry analysis.

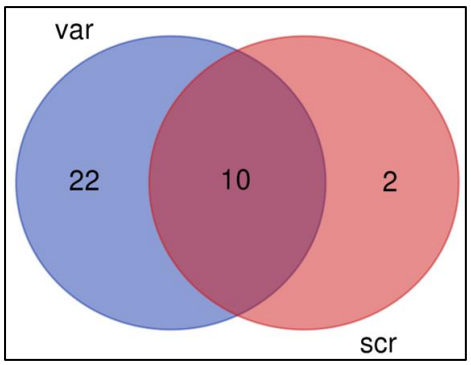


Figure 12: Number of protein hits in biotin-oligo pull-down experiment. Venn diagram highlighting the number of proteins obtained as hits in pull-down with the scrambled and insulator-like motif DNA. Blue: *var* insulator-like sequence, Red: scrambled (scr) sequence.

As reported in our motif analysis study and earlier reports, insulator-like motifs are present as repeats at the ends of *var* gene intron. As a result, we reasoned that having multiple repeats of the insulator-like motif might make pull-down experiments more effective and aid in the discovery of potential insulator-like protein candidates through their effective binding to the repeats. We had DNA having a 5' biotinylated consensus insulator-like motif and a non-biotinylated DNA element with 3 repeats of the insulator-like motif. We ligated these DNA elements to get concatenated DNA product having multiple repeats of motifs and we then carried out a pull-down experiment for the third time (Fig. 13). With this approach of pull-down, we didn't get any hit in the mass spectrometry analysis for both the pull-down with scrambled sequence and insulator-like motif. We used a low amount of biotin-tagged DNA in this case, and because of this pull-down might not have worked. Finally, we have the proteins obtained in the second pull-down as the major interactors of the insulator-like motif.

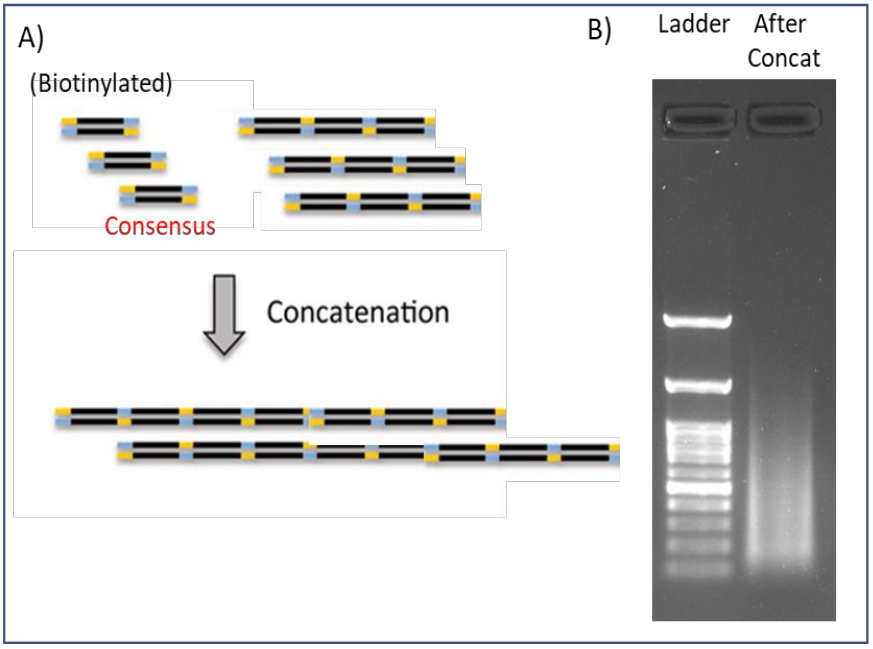


Figure 13: Pull-down with concatenated DNA schematic. A) A schematic of the concatenation of biotinylated and non-biotinylated DNA used for pull-down purposes. B) A gel run of the concatenated product after the overnight ligation reaction.

3.4. Cloning and generation of transfection lines for the selected proteins of interest

From the 11 proteins of interest obtained in the in-vitro pull-down approach the ALBA proteins, PfGBP2, PfISWI and the homolog of PfHMGB2 (PfHMGB1) have already been studied in *P. falciparum* and are known to be chromatin-interacting proteins. The knockouts of PfGBP2 and PfHMGB1 are shown to be affecting the mutually exclusive expression of *var* genes indicating their role in the regulatory mechanism. We reanalyzed the available ChIP-sequencing data for PfGBP2, PfISWI and PfHMGB1. PfISWI and PfHMGB1 show enrichment exclusively on the exons of all the var genes. Though the peak value in IGV for PfGBP2 enrichment is less it shows enrichment specifically at the intron-exon junction of *var* genes (Fig. 14). This makes PfGBP2 an interesting candidate to study further. We concentrated on the 8 proteins (the first 8 proteins in bold in Table 7) to study their role as insulator-like proteins, taking into account their peptide counts in the mass spectrometry analysis and a review of the literature on them. Next, we wanted to validate whether these proteins bind to the motif in the intronic region of *var* genes in the in-vitro culture-grown *P. falciparum* parasites. For this, we decided to overexpress these proteins by cloning the genes of interest in the plasmid, pDC2-cam-mRFP-2A-GFP-bsd-attP, an overexpression plasmid specific to *P. falciparum*. By restriction-digestion cloning, the gene of interest is cloned downstream to the calmodulin promoter and will have a GFP tag at the C-terminal. With this GFP-tagged overexpressed protein, we intend to perform ChIP-qPCR to validate the localization of the protein at the insulator-like sequence in the parasite genome. Along with this, at the same time, we could monitor the effect of overexpression of these putative insulator-like proteins on the expression of *var* genes using qPCR. We have successfully cloned 6 out of the 8 genes and have transfected the clones in *Plasmodium falciparum* 3D7 parasites. Currently, we are monitoring the transfected culture lines for the appearance of the transfected parasites. Table 8 provides the status of the clones and transfection lines for the 8 proteins of interest.

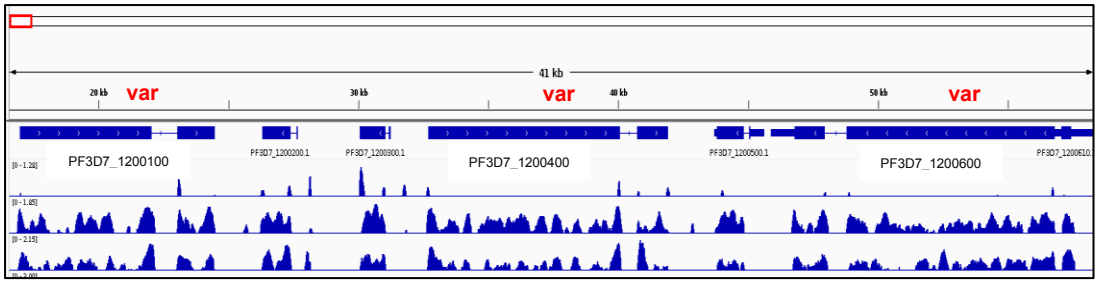


Figure 14: ChIP sequencing analysis for PfGBP2, PfISWI and PfHMGB1. IGV image showing the occupancy of PfGBP2, PfHMGB1, and PfISWI proteins on the 6 genes of chromosome 12. The three *var* genes PF3D7_1200100, PF3D7_1200400 and PF3D7_1200600 are indicated in the image. ChIP-sequencing data adapted from (Bryant et al., 2020; Gurung et al., 2006; Lu et al., 2013) and reanalyzed.

Gene	PCR amplification	Cloning	Transfection	Successfully transfected line
GBP2	Yes	Yes	Yes	No
NABP1	Yes	Yes	Yes	No
NABP2	Yes	Yes	Yes	No
ALBA1	Yes	No	No	No
ALBA2	Yes	Yes	Yes	No
ALBA3	Yes	Yes	Yes	No
ALBA4	Yes	Yes	Yes	No
HMGB2	No	No	No	No

Table 8: Status of the PCR, cloning and transfection lines of the proteins of interest.

3.5. PfHP1 and its interactome

In *P. falciparum* the spread of H3K9me3 modification is restricted only to the telomeric and subtelomeric regions. PfHP1 a reader of the H3K9me3 modification also occupies only those regions of the genome. A study in yeast, *Schizosaccharomyces pombe* reports that in the absence of HP1 protein swi6, the repressive mark H3K9me2 spreads into the euchromatic regions. This finding led to the conclusion that the swi6 protein serves as an insulator by restricting the spread of H3K9 methylation and defining the boundaries between euchromatin and heterochromatin by binding to H3K9me2 (Stunnenberg et al., 2015). In *P. falciparum*, the analysis of Hi-C data shows that chromosomes are divided into domain-like structures and the telomeric and intra-chromosomal *var* genes show strong interactions with each other (Mcgovern et al., 2019). The combined analysis of Hi-C data at 18 hr post-invasion stage and ChIP-sequencing data for PfHP1 at the same stage demonstrates that the boundaries of chromosome domains are formed at the regions where PfHP1 enrichment boundaries are present (Fig. 15). Also, the ChIP sequencing analysis for *var* genes indicates that the enrichment of activating histone modifications at the intronic regions is associated with the restriction of PfHP1 enrichment at that region. With these observations, we thought that in *P. falciparum*, PfHP1 itself or its association with other interacting proteins could be mediating the insulator-like activity and demarcating the chromatin boundaries. For this purpose, we performed immunoprecipitation of the PfHP1 protein to identify its interactome and find if the homolog of any already known insulator protein interacts with PfHP1. PfHP1 antibody generated in mice (Work done by Mamatharani D V) was used for immunoprecipitation and the pre-immune serum was used as a negative control. Table 9 lists the proteins that were identified as hits in the mass spectrometry analysis for PfHP1 immunoprecipitation after the protein hits identical in the negative control were taken out. Heterochromatin protein 1 was obtained as the top candidate indicating that the immunoprecipitation has worked successfully. The interesting hit in this immunoprecipitation experiment is high mobility group protein B2 (HMGB2) which is a known chromatin organizer protein in higher eukaryotes and is also obtained as a hit in the pull-down experiment with the insulator-like motif. HMGB2 could be an interesting candidate to study further in understanding the organization of chromatin boundaries and *var* gene expression in *Plasmodium falciparum*.

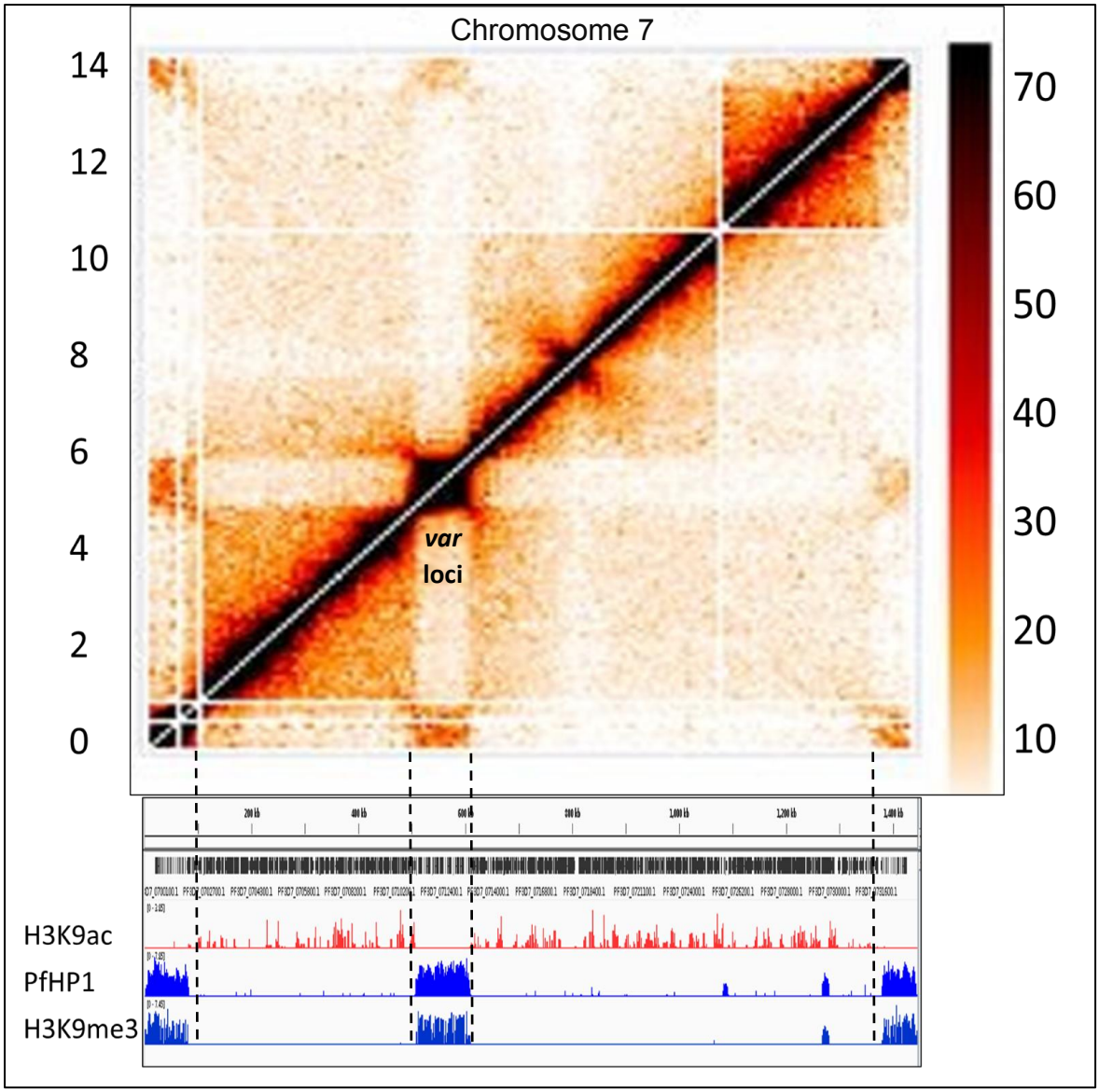


Figure 15: Hi-C and ChIP-sequencing analysis for chromosome 7. The Hi-C map for chromosome 7 (image adapted from Mcgovern et. al., 2019 and modified) and the IGV image including the enrichment of H3K9ac, H3K9me3 and PfHP1 protein on chromosome 7 of *P. falciparum*. In the Hi-C plot, chromosome start to end is from left to right and from bottom to top. The gradient shown at the right indicates the contact frequencies. Dashed lines demarcate the boundaries of PfHP1 occupancy and Hi-C domains.

Protein ID	Protein Name	Peptide count
PF3D7_1220900	Heterochromatin protein 1	9
PF3D7_1246200	actin I	5
PF3D7_0424600	Plasmodium exported protein (PHISTb)	2
PF3D7_1446600	centrin-2	2
PF3D7_0817900	High mobility group protein B2 (HMGB2)	2
PF3D7_0516200	40S ribosomal protein S11	2
PF3D7_0501000	Plasmodium exported protein, unknown function	2

Table 9: List of the proteins identified as interactors of PfHP1 protein in the immunoprecipitation experiment.

4. Discussion

Throughout the last two decades, researchers have been examining the mechanism underlying the mutually exclusive expression of *var* genes and *var* gene switching that controls the antigenic variation in *P. falciparum*. Although various components of the potential regulatory pathways have been validated and several hypotheses have been proposed, the precise mechanism underlying this phenomenon that supports parasite survival and infection persistence is yet unknown. In this study, we are focusing on the insulators, the possible regulators involved in the regulation of mutually exclusive expression of *var* genes, a factor that is not completely known. The ChIP-sequencing analysis reveals the differential enrichment of chromatin modifications at the locations of insulator-like motifs which enhanced our interest in the study of insulator-like proteins in *P. falciparum*. Through in-silico and pull-down experiments, we have identified the putative candidates which could have insulator-like activity. Currently, we are validating the role of these candidate proteins. In general, we are casting light on the potential insulator-like proteins in *P. falciparum*, which may function in a variety of other ways in addition to being genome architectural proteins that regulate the distinctive genome architecture.

According to the analysis of ChIP-sequencing data, activating and repressive chromatin modifications are differentially enriched at the intron-exon junctions of the *var* genes. On that basis, we classified *var* genes as H3K9ac+ and H3K9ac- *var* genes. The H3K9ac+ *var* genes may be actively transcribed and expressed given the positive correlation between histone modifications and transcription, however, this raises the question about the mutually exclusive expression of var genes. Possibly, the *var* gene having the highest enrichment of H3K9ac is being highly expressed in the background of comparatively very low transcription of remaining H3K9ac+ *var* genes. It's also possible that previously expressed *var* genes still contain H3K9ac traces on their introns, or that *var* genes with H3K9ac enrichment are clustered together inside the nucleus and are regulated differently. We will be performing RNA-sequencing of the same ChIP-sequencing samples and after that, we could get more idea about the expression of var genes showing differential enrichment of chromatin modifications. The most intriguing finding is the differential enrichment of histone modifications, particularly on the region surrounding the insulator-like motifs, which emphasizes the potential significance of the insulator-like motif in controlling the *var* gene expression.

This led to our curiosity about researching potential insulator-like proteins in *P. falciparum*.

We used two strategies for determining the potential insulator-like proteins. First, the in-silico analysis using BLAST and second the pull-down experiment with the proposed insulator-like motif which we also confirmed through our motif analysis. The proteins of interest from the in-silico analysis and top candidates in mass spectrometry followed by pull-down with the insulator-like motif are not common. PfTRZ emerged to be the major hit through in-silico analysis but was not detected in the pull-down as well as IP though it showed an enrichment pattern similar to PfHP1 on *var* genes. Possibly, because of the presence of the zinc-finger domain, a domain responsible for DNA binding in the majority of DNA-binding proteins, it is showing homology to most of the insulator proteins. Based on these results, as of now, we don't consider PfTRZ as a possible insulator-like protein. SMC complexes were also not detected in the pull-down and IP but based on the limited study of these proteins in *P. falciparum*, we can't comment much about their role as insulators. Additionally, based on the in-silico analysis, we could suggest that the *P. falciparum* might have a distinct set of insulator-like proteins compared to the insulator proteins in higher eukaryotes and the hits that we obtained in pull-down could be the potential insulator-like proteins.

Based on the available literature and the chromatin-binding feature we short-listed the proteins from the pull-down analysis to study them further from the perspective of insulator-like activity. According to a breast cancer study, FANCJ-like helicase is a DNA helicase implicated in the double-strand break repair pathway and interacts with Replication protein A (Gupta et al., 2007). The protein SIS1 is a crucial Hsp70 co-chaperone in yeast and resembles the bacterium protein DnaJ (Zhong and Arndt, 1993). The ISWI chromatin remodeling complex ATPase has earlier been studied in *P. falciparum*, and the available ChIP-sequencing data reveals occupancy exclusively on the exon 1 and exon 2 of *var* genes. Therefore, we were not majorly interested in these three proteins. G-strand binding protein 2 (PfGBP2), is a G-quadruplexes (G4s) binding protein in *P. falciparum*. G4s are enriched at the telomeric, subtelomeric and 5' upstream regions of *var* genes (Gurung et al., 2006). Also, the knockout of PfGBP2 protein in *P. falciparum* has been demonstrated to affect the expression of *var* genes making it interesting to study this protein. We obtained all the ALBA family proteins as hits in the pull-down. ALBA family proteins, PfALBA1, PfALBA2, PfALBA3 and

PfALBA4 are the DNA/RNA binding proteins and are known to bind the telomeres as well as the regulatory regions of *var* genes (Chêne et al., 2012; Vembar et al., 2015). Also, some of the PfALBAs are found to be associated with epigenetic modifiers. PfALBA3 interacts with the histone deacetylase PfSir2A and PfSet10, a methyltransferase while PfALBA4 is associated with the bromodomain protein PfBDP1 (Goyal et al., 2012; Josling et al., 2015). This enhances the chances that all the ALBA proteins and their crosstalk with epigenetic modifiers might have significance in the regulation of *var* gene expression through the deposition of required chromatin modifications. The homolog of the high mobility group protein B2 (PfHMGB2), PfHMGB1 is shown to be an architectural protein involved in genome organization in *P. falciparum* while in other organisms, HMGB homologs are known to be regulating the chromosome structure and gene expression through specific TAD formation (Lu et al., 2021; Voong et al., 2021). The top hits are the two putative nucleic-acid binding proteins, PF3D7_1132300 and PF3D7_0709500. The BLAST search for both proteins against the proteome of *Drosophila melanogaster*, humans and mouse (*Mus musculus*) gave no hit. The InterPro analysis to identify the putative domains in the proteins suggested that broadly proteins belong to the nucleic acid binding protein superfamily while the PF3D7_1132300 has the presence of replication protein-A related domain. In *Arabidopsis*, a replication protein A2, putative is shown be involved in the silencing of the genes via epigenetic regulation (Xia et al., 2006). All these interesting observations and previous studies focusing on the homologs or the proteins obtained as hits make it interesting to decipher the role of these proteins further.

It is important to note that we got PfHMGB2 as an overlapping hit in both the pull-down with the insulator-like motif and the immunoprecipitation of PfHP1, increasing the confidence to investigate the function of this protein since it has not been studied previously in *P. falciparum*. To increase our confidence in the outcomes and the identification of PfHP1 interactors, we intend to carry out immunoprecipitation with antibody cross-linking and are also working to improve our immunoprecipitation methodology.

Currently, in this project, we are at the stage of generating stably transfected overexpression cell lines for the selected 8 proteins. In *P. falciparum*, transfection involves the electroporation of iRBCs and the efficiency of transfection is low which requires multiple rounds of transfection to establish the stably transfected line. Also,

for successful transfection the plasmid DNA needs to cross four major barriers, the RBC plasma membrane, the parasitophorous vacuole membrane of the parasite, the parasite plasma membrane and the nuclear membrane which could significantly decrease the efficiency of transfection. We are doing multiple transfections rounds for all the proteins so that we could get the transfected lines as early as possible and proceed with the planned experiments.

Through the overexpression of these proteins and the validation of their localization at the insulator-like motif, we will narrow down further to short-list the protein candidates which specifically bind to the motif and influence the expression of *var* genes. ChIP-sequencing with anti-GFP antibody for the parasite lines, overexpressing putative insulator-like proteins will provide information about the other genomic regions where the protein specifically binds and could control genome organization and transcription of the genes. After these preliminary confirmations, we could also knock out the possible insulator-like proteins and discern the effect on the mutually exclusive expression of *var* genes and the histone modification profile on the *var* genes. In the future, we are also interested in performing chromatin conformation capture (3C/Hi-C) experiments to investigate the loop formation at the intronic region of the var genes. Future experiments incorporating the aspects of antisense transcription, insulator proteins, epigenetic modifications and genome organization studies will help to work on the hypothesis that we have put forth and will contribute to revealing the unknown facets of the mutually exclusive expression of *var* genes.

5. References

Ali, T., Renkawitz, R., and Bartkuhn, M. ScienceDirect Insulators and domains of gene expression. Curr Opin Genet Dev *37*, 17–26. https://doi.org/10.1016/j.gde.2015.11.009.

Amit-Avraham, I., Pozner, G., Eshar, S., Fastman, Y., Kolevzon, N., Yavin, E., and Dzikowski, R. (2015). Antisense long noncoding RNAs regulate var gene activation in the malaria parasite Plasmodium falciparum. Proc Natl Acad Sci U S A *112*, E982–E991. https://doi.org/10.1073/pnas.1420855112.

Analysis, U. (1985). Ultrastructural Analysis.

Andrews, K., Tran, T., Wheatley, N., and Fairlie, D. (2009). Targeting Histone Deacetylase Inhibitors for Anti-Malarial Therapy. Curr Top Med Chem *9*, 292–308. https://doi.org/10.2174/156802609788085313.

Avraham, I., Schreier, J., and Dzikowski, R. (2012). Insulator-like pairing elements regulate silencing and mutually exclusive expression in the malaria parasite Plasmodium falciparum. Proc Natl Acad Sci U S A *109*. https://doi.org/10.1073/pnas.1214572109.

Bertschi, N.L., Toenhake, C.G., Zou, A., Niederwieser, I., Henderson, R., Moes, S., Jenoe, P., Parkinson, J., Bartfai, R., and Voss, T.S. (2017). Malaria parasites possess a telomere repeat-binding protein that shares ancestry with transcription factor IIIA. Nat Microbiol *2*, 1–12. https://doi.org/10.1038/nmicrobiol.2017.33.

Bryant, J.M., Baumgarten, S., Dingli, F., Loew, D., Sinha, A., Claës, A., Preiser, P.R., Dedon, P.C., Scherf, A., and Bryant, J.M. (2020). Exploring the virulence gene interactome with CRISPR / dCas 9 in the human malaria parasite. 1–22. https://doi.org/10.15252/msb.20209569.

Carison, J., Berzins, K., and Torii, M. (1989). Brief Definitive Report. *169*, 1835–1840.
.

Chêne, A., Vembar, S.S., Rivière, L., Lopez-Rubio, J.J., Claes, A., Siegel, T.N., Sakamoto, H., Scheidig-Benatar, C., Hernandez-Rivas, R., and Scherf, A. (2012). PfAlbas constitute a new eukaryotic DNA/RNA-binding protein family in malaria parasites. Nucleic Acids Res *40*, 3066–3077. https://doi.org/10.1093/nar/gkr1215.

Chetverina, D., Fujioka, M., Erokhin, M., Georgiev, P., Jaynes, J.B., and Ã, P.S. (2017). Prospects & Overviews Boundaries of loop domains (insulators): Determinants of chromosome form and function in multicellular eukaryotes. *1600233*, 1–11. https://doi.org/10.1002/bies.201600233.

Chookajorn, T., Dzikowski, R., Frank, M., Li, F., Jiwani, A.Z., Hartl, D.L., and Deitsch, K.W. (2007). Epigenetic memory at malaria virulence genes. *104*, 16–19. https://doi.org/10.1073/pnas.0609084103.

Cui, L., and Miao, J. (2010). MINIREVIEWS Chromatin-Mediated Epigenetic Regulation in the Malaria Parasite Plasmodium falciparum □. *9*, 1138–1149. https://doi.org/10.1128/EC.00036-10.

Dhorda, M., Amaratunga, C., and Dondorp, A.M. (2021). Artemisinin and multidrug-resistant Plasmodium falciparum–a threat for malaria control and elimination. Curr Opin Infect Dis *34*, 432. .

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). identified by analysis of chromatin interactions. Nature 0–4. https://doi.org/10.1038/nature11082.

Dzikowski, R., Li, F., Amulic, B., Eisberg, A., Frank, M., Patel, S., Wellems, T.E., and Deitsch, K.W. (2007). Mechanisms underlying mutually exclusive expression of virulence genes by malaria parasites. EMBO Rep *8*, 959–965. https://doi.org/10.1038/sj.embor.7401063.

Epp, C., Li, F., Howitt, C.A., Chookajorn, T., and Deitsch, K.W. (2009). Chromatin associated sense and antisense noncoding RNAs are transcribed from the var gene family of virulence genes of the malaria parasite Plasmodium falciparum. 116–127. https://doi.org/10.1261/rna.1080109.3.

Flueck, C., Bartfai, R., Volz, J., Niederwieser, I., Salcedo-Amaya, A.M., Alako, B.T.F., Ehlgen, F., Ralph, S.A., Cowman, A.F., Bozdech, Z., et al. (2009). Plasmodium falciparum heterochromatin protein 1 marks genomic loci linked to phenotypic variation of exported virulence factors. PLoS Pathog *5*. https://doi.org/10.1371/journal.ppat.1000569.

Flueck, C., Bartfai, R., Niederwieser, I., Witmer, K., Alako, B.T.F., Moes, S., Bozdech, Z., Jenoe, P., Stunnenberg, H.G., and Voss, T.S. (2010). A major role for the Plasmodium falciparum ApiAP2 protein PfSIP2 in chromosome end biology. PLoS Pathog *6*, e1000784. .

Frank, M., Dzikowski, R., Costantini, D., Amulic, B., Berdougo, E., and Deitsch, K. (2006). Strict pairing of var promoters and introns is required for var gene silencing in the malaria parasite Plasmodium falciparum. Journal of Biological Chemistry *281*, 9942–9952. https://doi.org/10.1074/jbc.M513067200.

Fraschka, S.A., Filarsky, M., Hoo, R., Preiser, P.R., Fraschka, S.A., Filarsky, M., Hoo, R., Niederwieser, I., and Yam, X.Y. (2018). Comparative Heterochromatin Profiling Reveals Conserved and Unique Epigenome Signatures Linked to Adaptation and Development of Malaria Resource Comparative Heterochromatin Profiling Reveals Conserved and Unique Epigenome Signatures Linked to Adaptation a. 407–420. https://doi.org/10.1016/j.chom.2018.01.008.

Freitas-junior, Â.H., Bottius, E., Pirrit, L.A., Deitsch, K.W., Scheidig, C., Guinet, F., Nehrbass, U., Wellems, T.E., and Scherf, A. (2000). Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of P . falciparum. *407*. .

Gardner, M.J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R.W., Carlton, J.M., Pain, A., Nelson, K.E., Bowman, S., et al. (2002). Genome sequence of the human malaria parasite Plasmodium falciparum. Nature *419*, 498–511. https://doi.org/10.1038/nature01097.

Glenn, B., Hilaire, S., Casellas, R., Lander, E.S., and Aiden, E.L. (2017). Cohesin Loss Eliminates All Loop Domains. Cell *171*, 305-309.e24. https://doi.org/10.1016/j.cell.2017.09.026.

Goyal, M., Alam, A., Iqbal, M.S., Dey, S., Bindu, S., Pal, C., Banerjee, A., Chakrabarti, S., and Bandyopadhyay, U. (2012). Identification and molecular characterization of an Alba-family protein from human malaria parasite Plasmodium falciparum. *40*, 1174–1190. https://doi.org/10.1093/nar/gkr821.

Gupta, R., Sharma, S., Sommers, J.A., Kenny, M.K., Cantor, S.B., and Brosh Jr, R.M. (2007). FANCJ (BACH1) helicase forms DNA damage inducible foci with replication protein A and interacts physically and functionally with the single-stranded DNA-binding protein. Blood, The Journal of the American Society of Hematology *110*, 2390–2398. .

Gurung, P., Gomes, A.R., Martins, R.M., Juranek, S.A., Alberti, P., Urbach, S., Gazanion, E., Guitard, V., and Paeschke, K. (2006). PfGBP2 is a novel G-quadruplex binding protein in.

Heger, P., and Wiehe, T. (2014). New tools in the box : An evolutionary synopsis of chromatin insulators. Trends in Genetics *30*, 161–171. https://doi.org/10.1016/j.tig.2014.03.004.

Jerković, I., Szabo, Q., Bantignies, F., and Cavalli, G. (2020). Higher-order chromosomal structures mediate genome function. J Mol Biol *432*, 676–681. .

Jiang, L., Mu, J., Zhang, Q., Ni, T., Srinivasan, P., Rayavara, K., Yang, W., Ribeiro, M.C., Scherf, A., Aravind, L., et al. PfSETvs methylation of histone H3K36 represses virulence genes in Plasmodium falciparum. 0–6. https://doi.org/10.1038/nature12361.

Josling, G.A., Petter, M., Oehring, S.C., Gupta, A.P., Dietz, O., Wilson, D.W., Schubert, T., Längst, G., Gilson, P.R., and Crabb, B.S. (2015). A Plasmodium falciparum bromodomain protein regulates invasion gene expression. Cell Host Microbe *17*, 741–751. .

Jutras, B.L., Verma, A., and Stevenson, B. (2012). Identification of Novel DNA-Binding Proteins Using DNA-Affinity Chromatography / Pull Down. 1–13. https://doi.org/10.1002/9780471729259.mc01f01s24.

Karmodiya, K., Pradhan, S.J., Joshi, B., Jangid, R., Reddy, P.C., and Galande, S. (2015). A comprehensive epigenome map of Plasmodium falciparum reveals unique mechanisms of transcriptional regulation and identifies H3K36me2 as a global mark of gene suppression. Epigenetics Chromatin *8*, 1–18. https://doi.org/10.1186/s13072-015-0029-1.

Kurbidaeva, A. (2021). Insulators in Plants : Progress and Open Questions.

Lu, B., Liu, M., Gu, L., Li, Y., Shen, S., Guo, G., Wang, F., He, X., Zhao, Y., Shang, X., et al. The Architectural Factor HMGB1 Is Involved in Genome Organization in the Human Malaria Parasite Plasmodium falciparum.

Mcgovern, K., Bunnik, E.M., Venkat, A., Shao, J., Mcgovern, K.E., Batugedara, G., and Worth, D. (2019). Comparative 3D genome organization in apicomplexan parasites. https://doi.org/10.1073/pnas.1810815116.

Miller, L.H., Baruch, D.I., Marsh, K., and Doumbo, O.K. (2002). The pathogenic basis of malaria. *415*, 673–679. .

Milner, D.A. (2018). Malaria Pathogenesis. 1–11. .

Misteli, T. (2020). The self-organizing genome: principles of genome architecture and function. Cell *183*, 28–45. .

Pérez-Toledo, K., Rojas-Meza, A.P., Mancio-Silva, L., Hernández-Cuevas, N.A., Delgadillo, D.M., Vargas, M., Martínez-Calvillo, S., Scherf, A., and Hernandez-Rivas, R. (2009). Plasmodium falciparum heterochromatin protein 1 binds to tri-methylated histone 3 lysine 9 and is linked to mutually exclusive expression of var genes. Nucleic Acids Res *37*, 2596–2606. https://doi.org/10.1093/nar/gkp115.

Petter, M., Selvarajah, S.A., Lee, C.C., Chin, W.H., Gupta, A.P., Bozdech, Z., Brown, G. V, and Duffy, M.F. (2013). H2A . Z and H2B . Z double-variant nucleosomes define intergenic regions and dynamically occupy var gene promoters in the malaria parasite Plasmodium falciparum. *87*, 1167–1182. https://doi.org/10.1111/mmi.12154.

Ralph, S.A., Scheidig-Benatar, C., and Scherf, A. (2005). Antigenic variation in Plasmodium falciparum is associated with movement of var loci between subnuclear locations. Proc Natl Acad Sci U S A *102*, 5414–5419. https://doi.org/10.1073/pnas.0408883102.

Rao, S.S.P., Huntley, M.H., Durand, N.C., and Stamenova, E.K. (2014). Article A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell *159*, 1665–1680. https://doi.org/10.1016/j.cell.2014.11.021.

Report, W.M. (2020). World malaria report 2020.

Rrubio, J., Thompson, J.K., and Cowman, A.F. (1996). The var genes of Plasmodium falciparum are located in the subtelomeric region of most chromosomes L1E. *15*, 4069–4077. .

Ryan, Cooper, and Tauer (2013). 済無No Title No Title No Title. Paper Knowledge . Toward a Media History of Documents *159*, 12–26. .

Scherf, A., Hernandez-Rivas, R., Buffet, P., Bottius, E., Benatar, C., Pouvelle, B., Gysin, J., and Lanzer, M. (1998). Antigenic variation in malaria: In situ switching, relaxed and mutually exclusive transcription of var genes during intra-erythrocytic

development in Plasmodium falciparum. EMBO Journal *17*, 5418–5426. https://doi.org/10.1093/emboj/17.18.5418.

Scherf, A., Lopez-Rubio, J.J., and Riviere, L. (2008). Antigenic variation in Plasmodium falciparum. Annu Rev Microbiol *62*, 445–470. https://doi.org/10.1146/annurev.micro.61.080706.093134.

Siegel, T.N., Hon, C., Zhang, Q., Scheidig-benatar, C., Martins, R.M., Sismeiro, O., Coppée, J., and Scherf, A. (2014). Strand-specific RNA-Seq reveals widespread and developmentally regulated transcription of natural antisense transcripts in Plasmodium falciparum.

Stunnenberg, R., Kulasegaran-Shylini, R., Keller, C., Kirschmann, M.A., Gelman, L., and Bühler, M. (2015). H3K9 methylation extends across natural boundaries of heterochromatin in the absence of an HP 1 protein. EMBO J *34*, 2789–2803. .

Umbetova, G.H., Belyaeva, E.S., Baricheva, E.M., and Zhimulev, I.F. (1991). Cytogenetic and molecular aspects of position effect variegation in Drosoplu ' la melanogaster. *2*, 55–61. .

Vembar, S.S., Macpherson, C.R., Sismeiro, O., Coppée, J., and Scherf, A. (2015). The PfAlba1 RNA-binding protein is an important regulator of translational timing in Plasmodium falciparum blood stages. 1–18. https://doi.org/10.1186/s13059-015-0771-5.

Voong, C.K., Goodrich, J.A., and Kugel, J.F. (2021). Interactions of HMGB Proteins with the Genome and the Impact on Disease.

Wassmer, S.C., Taylor, T.E., Rathod, P.K., Mishra, S.K., Mohanty, S., Arevalo-Herrera, M., Duraisingh, M.T., and Smith, J.D. (2015). Investigating the pathogenesis of severe malaria: a multidisciplinary and cross-geographical approach. Am J Trop Med Hyg *93*, 42. .

Xia, R., Wang, J., Liu, C., Wang, Y., Wang, Y., Zhai, J., Liu, J., Hong, X., Cao, X., and Zhu, J.-K. (2006). ROR1/RPA2A, a putative replication protein A2, functions in epigenetic gene silencing and in regulation of meristem development in Arabidopsis. Plant Cell *18*, 85–103. .

Zhong, T., and Arndt, K.T. (1993). The yeast SIS1 protein, a DnaJ homolog, is required for the initiation of translation. Cell *73*, 1175–1186. .

(2010). The New England Journal of Medicine Downloaded from nejm.org at UNL on October 8, 2015. For personal use only. No other uses without permission. From the NEJM Archive. Copyright © 2010 Massachusetts Medical Society. All rights reserved.