Regulation of gene expression during DNA damage response in *Plasmodium falciparum*

Gupta, Devendra Kumar

2016

Gupta, D. K. (2016). Regulation of gene expression during DNA damage response in *Plasmodium falciparum*. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/67323>

<https://doi.org/10.32657/10356/67323>

Regulation of gene expression during DNA damage response in *Plasmodium falciparum*

DEVENDRA KUMAR GUPTA

SCHOOL OF BIOLOGICAL SCIENCES

A thesis submitted to the Nanyang Technological University in partial fulfillment of the
requirement for the degree of Doctor of Philosophy.

2016

Acknowledgement

First and foremost, I would like to express my sincere gratitude to my Ph. D. supervisor Associate Professor Zbynek Bozdech for his encouragement, support and guidance throughout my research. I thank him for providing an excellent scientific environment, stimulating discussions and for having faith in me. I would also like to thank my doctoral committee members Professor Daniela Rhodes and Dr. Eugene Makeyev for their valuable suggestions. I am thankful to SBS, NTU, for providing all the facilities and the financial support throughout my PhD.

I am thankful to all the former and present lab members of Zbynek Bozdech Lab for providing the cordial research environment. I am very grateful to Dr. Archana Gupta for assisting me in the data analysis. I would also like to thank Dr Vineet Kumar and Alok Patra for their friendship and continuous help that accelerated my research and thesis progress.

Lastly, I want to express my gratitude and love to my parents, sisters, and my wife Dr. Vijit Dalal for their incessant love and prayers which gave me the strength to finish this PhD thesis.

Abstract

The genomic integrity of *Plasmodium* parasite remains under constant threat from exogenous and endogenous genotoxic stresses throughout the developmental stages of the parasite complex life cycle. Here, we investigated the responses of *P. falciparum* to DNA damaging agents to assess its DNA repair capacity on one side and the mutational effect of DNA damage on the other. Using DNA microarrays, we have characterized a global transcriptional response of *P. falciparum* to methyl methanesulphonate (MMS), which alkylates DNA bases and causes strand breaks. Alongside, we studied the epigenetic regulation of DNA damage response in the parasite. Our results showed that MMS causes up-regulation of DNA repair and DNA repair-related pathways. MMS-induced gene expression is accompanied by hyperacetylation of histone lysine residues H4K8Ac, H4K16Ac and deacetylation of H3K56Ac and H3K9Ac. Intriguingly, the induction of the DNA repair machinery, initially observed in the drug susceptible parasite strains (3D7 and D6), could not be detected in two drug-resistant parasites (Dd2 and W2). Also, our genome sequencing results identified point mutations in 18 DNA repair genes that were exclusively present in Dd2 and W2. These latter strains were previously shown to exhibit a mutator phenotype also termed as accelerated resistance to multiple drugs (ARMD). Although a link between the DNA repair machinery and ARMD phenotype has been suggested previously, our results provide mechanistic evidence that the ARMD phenotype is potentially caused by a defect in a sensing/signaling pathway(s) that triggers the DNA repair machinery upon DNA damage. Further, Artemisinin, the primary antimalarial drug for *P. falciparum* infections, has been long debated for its mechanism of action. Here, we elucidated that MMS and artemisinin induced transcriptional up-regulation of repair machinery and altered the chromatin structure in a

similar manner at multiple stages. Taken together, artemisinin elicits DNA damage response in parasite, which is similar to the MMS mediated response.

Differential gene expression in *Plasmodium* confers flexibility to the changing environments as well as providing the parasite with indispensable tools for survival and pathogenesis. The exact mechanism of such transcriptional differences in *Plasmodium* gene expression is unknown. Although few focused studies on sub-telomeric virulence gene families have indicated the role of chromatin in regulating the differential gene expression. Our objective was to investigate the role of epigenetic mediated mechanism in the differential gene expression in two *P. falciparum* strains. By comparing the gene expression and histone modification patterns (H3K9Ac and H3K56Ac) between Dd2 and T996 strains, we found that most differentially expressed genes associated with differential H3K56 acetylation. Further, we characterized one of the AP2 domains containing transcription factor (putative), PF14_0271, which was found to be both differentially acetylated and differentially expressed between *P. falciparum* strains. Using endogenous tagging and knock-out approach, our results suggest that PF14_0271 is a schizont specific transcriptional regulator.

Table of contents

Acknowledgement.....	i
Abstract.....	ii
Table of content.....	iv
Conferences presentation.....	x
List of figures.....	xi
List of tables	xiv
Table of abbreviations.....	xv
1 Introduction.....	1
1.1 The Disease.....	2
1.2 <i>P. falciparum</i> life cycle.....	2
1.3 Genome, transcriptome and proteome of <i>P. falciparum</i>	4
1.4 Transcriptional control in <i>P. falciparum</i>	6
1.5 Cis regulatory motifs and trans-acting factors.....	7
1.6 DNA damage and repair pathways.....	8
1.6.1 Direct repair.....	9
1.6.2 Excision repair pathways	10
1.6.2.1 Base excision repair (BER).....	10
1.6.2.2 Nucleotide excision repair (NER).....	13
1.6.2.3 Mismatch repair (MMR).....	13
1.6.3 Double strand break repair pathways.....	15

1.7 Association of defective DNA damage response and drug resistance in <i>P. falciparum</i>	18
1.8 The role of chromatin dynamics in <i>Plasmodium</i> gene regulation	19
1.8.1 Chromatin architecture.....	21
1.8.2 Histone and its modifications.....	21
1.8.3 Chromatin modifiers associated with histone acetylation and methylation	23
1.9 Role of chromatin dynamics in DNA damage response.....	24
1.9.1 Chromatin remodeling complex.....	24
1.9.2 Histone modifications in DNA damage response (DDR)	25
1.9.2.1 Histone phosphorylation.....	27
1.9.2.2 Histone acetylation.....	27
1.9.2.3 Histone methylation.....	29
1.10 Role of chromatin modifications in preventing spontaneous mutation.....	29
1.11 Mechanism of action of artemisinin, the first line antimalarial drug.....	30
1.12 Transcriptional heterogeneity in <i>Plasmodium</i>	32
1.12.1 Variant expression of invasion related proteins.....	33
1.12.2 Variable/ Differential gene expression: role in pathogenesis.....	33
1.12.3 Epigenetic basis of known differentially expressed genes.....	34
1.13 Objective of thesis.....	36
Chapter 2 Materials and Methods.....	39
2.1 Parasite culture.....	40
2.2 Drug treatments.....	41

2.3 RNA isolation.....	41
2.4 cDNA synthesis by Reverse transcription.....	41
2.5 cDNA labeling and microarray hybridization.....	42
2.6 Microarray data analysis.....	42
2.7 Chromatin immunoprecipitation.....	43
2.8 Western Blot/ Immunoblot analysis.....	44
2.9 Immunofluorescence assay (IFA).....	45
2.10 Quantitative Real time PCR (RT-PCR).....	46
2.11 Endogenous tagging of PF14_0271.....	46
2.11.1 Generation of recombinant plasmid (PF14_0271-mCherry-HA plasmid)	46
2.11.2 Transfection and verification of integration in the parasite genome.....	47
2.12 High-throughput genome sequencing by Illumina Mi-seq method.....	47
Chapter 3 DNA damage response in <i>Plasmodium falciparum</i>: transcriptional and epigenetic events following DNA damage in malaria parasite.....	49
3.1 Introduction.....	50
3.2 Results.....	52
3.2.1 Global transcriptional response to DNA damaging agents.....	52
3.2.2 Effect of MMS on transcription of malaria DNA repair genes.....	58
3.2.2.1 Base excision repair (BER)	58
3.2.2.2 Mismatch repair (MMR)	61
3.2.2.3 Nucleotide excision repair (NER)	61
3.2.2.4 Double strand break repair pathways (DSBR)	63

3.2.3 DNA damage sensors.....	64
3.2.4 Role of chromatin dynamics in DNA damage response in malaria.....	66
3.2.5 Genome wide distribution of H3K9Ac and H4K8Ac during DNA damage in malaria.....	68
3.2.6 DNA damage induces genome wide spread of H4K8 acetylation whereas H3K9 acetylation preferentially marks the DNA repair genes.....	72
3.2.7 Antimalarial, artemisinin induce DNA damage response in <i>P. falciparum</i>	75
3.2.8 Transcriptional response of clinical doses of artemisinin	76
3.2.9 Effect of clinical dose of artemisinin on parasite histone modifications.....	78
3.3 Discussion.....	79
3.3.1 Transcriptional response to DNA damaging agent, MMS.....	81
3.3.2 DNA damage induces H4K8 hyperacetylation and H3K9 deacetylation.....	82
3.3.3 Artemisinin induce DNA damage response in <i>P. falciparum</i>	83
3.4 Summary.....	86
Chapter 4 ARMD parasites displaying mutator phenotype have defective DNA repair.....	89
4.1 Introduction.....	90
4.2 Results.....	92
4.2.1 Transcriptional differences between ARMD (Dd2) and non-ARMD (3D7) parasite during DNA damage.....	92
4.2.2 Functional classification of MMS induced genes in 3D7 and Dd2: Trophozoite and schizont stage.....	94
4.2.3 Identification of defective DNA repair machinery in ARMD parasites.....	99

4.2.4 Epigenetic mechanism associated with ARMD phenotype.....	103
4.2.5 Genetic mechanisms associated with development of ARMD “mutator” phenotype..	105
4.3 Discussion.....	110
Chapter 5 Chromatin mediated regulation of differential gene expression in <i>Plasmodium falciparum</i>.....	117
5.1 Introduction.....	118
5.2 Results.....	119
5.2.1 Identification of differentially expressed genes between <i>P. falciparum</i> strains.....	119
5.2.1.1 Pearson method.....	119
5.2.1.2 Euclidean distance method.....	121
5.3 Unraveling the epigenetic landscape of differentially expressed genes.....	123
5.3.1 H3K9 acetylation is not associated with differentially transcribed genes.....	123
5.3.2 Euclidean distance and differential H3K9 acetylation.....	124
5.3.3 Association of H3K56 acetylation with differential transcription: Euclidean distance method.....	126
5.3.4 Epigenetic marking of H3K56Ac over differentially expressed genes.....	128
5.4 Molecular characterization of putative transcription factor PF14_0271 which was found to be both differentially acetylated and differentially expressed between <i>P. falciparum</i> strains.....	131
5.4.1 Endogenous tagging of AP2 domain containing transcription factor PF14_0271...	132
5.4.2 Endogenous protein detection and intracellular localization of PF14_0271.....	134
5.4.3 Genome wide identification of PF14_0271 binding site.....	135

5.4.4 Transcriptional analysis of PF14_0271 knock-out parasites.....	139
5.4.4.1 PF14_0271 knock out has no effect on global stage specific expression in <i>P. falciparum</i>	140
5.4.4.2 Identification of differentially expressed genes which were altered in PF14_0271 knock-out parasites cell line.....	142
5.5 Discussion.....	144
5.5.1 Unraveling the epigenetic landscape of differentially expressed genes.....	144
5.5.2 Molecular characterization of putative transcription factor which was found to be both differentially acetylated and differentially expressed.....	147
Chapter 6 General discussion and conclusion.....	151
Chapter 7 References.....	159

Conference presentations:

- 11th BioMalPar Meeting held on 11-13 May 2015 in Heidelberg, Germany. (Poster presentation).
- Annual Molecular Parasitology Meeting held on 14-18 Sept 2014 in Woods Hole, USA. (Poster presentation)
- Gordon Research Conference (GRC) on Genomic instability held on 06-11 July 2014 in Hong Kong. (Poster presentation)
- 7th Asian Pacific Organization of Cell Biology (APOCB) congress held on 24-27 February 2014 in Singapore. (Poster presentation)

Publication:

- **Gupta DK**, Patra AP, Lei Z, Gupta AP and Bozdech Z: **DNA damage regulation and its role in drug-related phenotypes in the malaria parasites.** *Scientific Reports* **6**, 23603 (2016).

List of Figures

Figure #	Description	Page #
1.1	Overview of <i>P. falciparum</i> life cycle	3
1.2	The transcriptome and proteome of the IDC of the <i>P. falciparum</i>	6
1.3	Eukaryotic cells have three major excision repair pathways	11
1.4	Major pathways to repair double strands breaks (DSBs)	17
1.5	Summary of the post translational modifications (PTMs) identified in the <i>P. falciparum</i>	20
1.6	Relation between <i>var</i> gene expression and various chromatin marks	35
3.1	The source of DNA damage and different mechanism of DNA repair in <i>P. falciparum</i>	51
3.2	Global transcriptional response of <i>P. falciparum</i> to DNA damaging agents	54
3.3	Functional enrichment analysis of genes which were up-regulated by MMS	57
3.4	DNA damage affects the overall abundance of histone modifications during the <i>P. falciparum</i> IDC	68
3.5	The Global nucleosome occupancy of H3K9Ac, H3K56Ac and H4K8Ac during DNA damage in <i>P. falciparum</i>	70
3.6	Distribution of histone marks along parasite genes	71
3.7	Correlation analysis between MMS induced gene expression and histone marks	73
3.8	Association between differential expression and altered histone marks occurrence during MMS treatment	74
3.9	Correlation analysis between MMS induced gene expression and H3K56Ac	75

3.10	Artemisinin and MMS induce similar transcriptional response	77
3.11	Effect of clinical dose of artemisinin and other antimalarial drugs on parasite histone modifications	80
3.12	The events of epigenetic regulation of DNA damage response in eukaryotes	87
4.1	Global transcriptional response of <i>P. falciparum</i> ARMD (Dd2) and non-ARMD (3D7) to DNA damaging agents	93
4.2	Correlation between biological replicates for 3D7 and Dd2	95
4.3	MMS induced differential expression of genes	96
4.4	GSEA enriched genes: Differentially expressed DNA repair genes 3D7 vs. DD2	98
4.5	GSEA enriched genes: Differentially expressed DNA repair genes 3D7 vs. DD2	100
4.6	Global transcriptional response of <i>P. falciparum</i> ARMD (Dd2, W2) and non-ARMD (3D7, D6) to DNA damaging agents	102
4.7	Effect of MMS on ARMD and non-ARMD parasite chromatin	104
4.8	DNA polymorphism in DNA damage sensors between ARMD and non-ARMD parasites	111
5.1	IDC transcriptome of Dd2 and T996	120
5.2	Overview of transcriptional variation between Dd2 and T996	122
5.3	Association of H3K9Ac with differential expression	125
5.4	Association of H3K56 acetylation with differential transcription	127
5.5	The differentially enriched H3K56c profiles and differentially expressed profiles between two strains for PF14_0740 and PF14_0271	129
5.6	Distribution of H3K56 acetylation over the 108 oligos whose differential H3K56Ac correlates with differential expression	131
5.7	Endogenous tagging of AP2 domain containing transcription factor PF14_0271	133

5.8	Endogenous protein detection of PF14_0271 in schizont stage	135
5.9	Endogenous intracellular localization of PF14_0271 in schizont stage	136
5.10	Genome wide identification of PF14_0271 binding site	138
5.11	Functional enrichment analysis on these 642 genetic loci which were associated with PF14_0271	139
5.12	PF14_0271 knock out has no effect on global stage specific expression in <i>P. falciparum</i>	141
5.13	Identification of differentially expressed genes which were altered in PF14_0271 knock-out parasites cell line	143
5.14	Identification of differentially expressed genes altered in PF14_0271 knock-out parasites cell line	145
6.1	Overview of the <i>P. falciparum</i> DNA damage pathways and their transcriptional properties under MMS-induced stress	153

List of Tables

Table #	Description	Page #
1.1	Summary of PTMs identified on <i>P. falciparum</i> histones	22
1.2	List of histone modifications that are associated with DNA damage response in yeast and mammals	26
2.1	List of various antibodies used in immunoblot analysis	45
3.1	List of DNA excision repair genes which were significantly differentially expressed after 1 hour, 3 hour and 6 hour of MMS treatment	59
3.2	Comparison of core components of nucleotide excision repair (NER) among yeast, humans and <i>P. falciparum</i>	62
3.3	List of double strand breaks (DSB) repair genes which were significantly differentially expressed by MMS	64
3.4	Comparison of DNA damage sensors among mammals, yeast and <i>Plasmodium</i>	65
4.1	SNP's identified in Dd2, W2 and D6 clones	106
4.2	SNP's identified in DNA repair genes among Dd2, W2 and D6 clones	107
4.3	List of SNP located in DNA repair genes of ARMD parasites (Dd2 and W2)	109
4.4	The comparison of drug resistant (ARMD) and drug sensitive (Non-ARMD) <i>P. falciparum</i> clones	114

List of Abbreviations

Abbreviation	Extended form
BER	Base excision repair
cDNA	Complementary DNA
ChIP-on-chip	Chromatin immunoprecipitation with microarrays
co-IP	Co-immunoprecipitation
CGH	Comparative genomic hybridization
DBDs	DNA binding domains
DSBR	Double strand break repair
FIRE	Finding Informative Regulatory Elements
GFP	Green fluorescence protein
GTFs	General transcription factors
GSEA	Gene set enrichment analysis
hpi	Hours post invasion
HR	Homologous recombination
IDC	Intra-erythrocytic developmental cycle
mRNA	Messenger RNA
MMR	Mismatch repair
MMS	Methyl methane sulphonate
NER	Nucleotide excision repair
NHEJ	Non homologous end joining
ORFs	Open reading frames
PBS	Phosphate buffered saline

PIC	Pre-initiation complex
PCC	Pearson correlation coefficient
PTM	Post translational modifications
RBCs	Red blood cells
RNAi	RNA interference
RNA pol II	RNA polymerase II
STFs	Specific transcription factors
TAPs	Transcription associated proteins
TBP	TATA binding protein
TFs	Transcription factors
TSS	Transcription start site
UTR	Untranslated region

Chapter 1

Introduction

1.1 The Disease

Malaria is one of the most deadly diseases associated with human civilization since the origin of mankind. Hippocrates, a Greek physician, regarded as the father of medicine, was the first to describe the symptoms of the disease (1). The disease is transmitted to humans by the bite of a female *Anopheles* mosquito carrying malaria parasites. The malaria parasite was discovered in 1880 by Alphonse Laveran, a French army surgeon, who primarily reported the parasite from the blood of a patient infected with malaria (2). The malaria parasite, *Plasmodium*, is a protozoan which belongs to phylum Apicomplexa and is regarded as one of the most successful intracellular parasite belonging to this phylum. There are 5 *Plasmodium* species, *P. falciparum*, *P. vivax*, *P. malariae*, *P. knowlesi* and *P. ovale* which are known to cause infection in humans. *P. falciparum* is the most notorious and is responsible for the maximum number of deaths associated with malaria. The infection caused by *Plasmodium* is not just restricted to humans, but there are other *Plasmodium* species, which cause diseases in a broad range of organisms which include birds, lizards and monkeys (3).

Approximately 200-300 million clinical cases are reported every year and on average, 1 million people are killed annually. In 2012, malaria was responsible for 627,000 number of deaths and most of the casualties were children below the age of five years (4). Although, the overall toll caused by malaria has declined over recent years due to the initiation of several malaria eradication programs by WHO, however, complete elimination of malaria does not seem feasible with the current intervention strategies and the increased incidence of drug resistance; hence additional tools are required (WHO, 2009).

1.2 *P. falciparum* life cycle

The life cycle of *Plasmodium falciparum* is very complex, which involves two hosts, human and an arthropod vector and comprises three stages: The mosquito stage, the human liver stage and blood stage. Malaria infection begins when a *Plasmodium*-infected mosquito injects

sporozoite into the human host during a blood meal (**Figure 1.1**). These infective sporozoites travel to the liver by passing through the phagocytic Kupffer cells and then invade the hepatocytes. Inside the hepatocytes, sporozoites undergo first asexual multiplication and produce merozoites. These merozoites are released from the liver and then enter the bloodstream, where they invade red blood cells (RBCs).

Figure 1.1 Overview of *P. falciparum* life cycle. The life cycle of the parasite comprises three stages. The mosquito stage (sexual cycle), liver stage and the blood stage, the last two stages are asexual stages in humans. The blood stage is known as the intraerythrocytic development cycle (IDC). The schematic of life cycle of malaria is reproduced with permission from Cowman and Crabb *et al.* (6).

The merozoites initiate a second round of asexual multiplication known as the intra-erythrocytic developmental cycle (IDC) inside RBC, which lasts for approximately 48 hours. During this periodic IDC cycle, merozoites develop into three morphologically distinct stages: ring, trophozoite and schizont. After every round of the IDC, RBC ruptures to release

16-32 daughter merozoites, which invade fresh RBC to continue their multiplication cycle. These repetitive cycles of invasion and rupturing cause most of the pathological conditions associated with malaria. In the blood, some of the IDC stages differentiate into the sexual male or female gametocytes, which are taken up by the female *Anopheles* mosquito during a mosquito bite. In the gut of the mosquito, the gametocytes mature into male and female gametes and fuse to produce a zygote that transforms into an ookinete. Another round of multiplication occurs in the ookinete, which further matures into an oocyst that contains the sporozoites. The rupture of oocyst releases infective sporozoites, which subsequently migrate to the salivary gland, and are ready to infect humans during the next blood meal (5,6).

1.3 Genome, transcriptome and proteome of *P. falciparum*

The genome of *P. falciparum*, the pathogen responsible for the most severe form of malaria was decoded in 2002 and was the groundwork finding which paved the way for the current functional genomics studies on malaria. The size of *P. falciparum* genome is 22.8 mega-bases (Mb) encompassing 14 linear chromosomes, a circular plastid like genome and a linear mitochondrial genome. The malaria genome has nearly 5,400 predicted protein-encoding genes, and more than 60% of these have no sequence homology to other known organisms.

The most striking attribute of the *P. falciparum* genome is its high content of A-T nucleotides (overall 80.6%), which increases to 90% in introns and intergenic regions. The average length of genes in *P. falciparum* is 2.3kb after removing introns, which is considerably larger than in the other organisms in which the average gene lengths ranges from 1.3 to 1.6kb (7). Another, interesting fact about the *P. falciparum* genome is that it lacks genes coding for the RNA interference machinery, including *dicer* and *argonaute* (8). Also, so far, the transposable elements or retrotransposons have not been detected in the genome of the malaria parasite (7).

The completion of the genome sequence of *P. falciparum* identified so many unique open reading frames (ORFs), hence, to gain insights into the function of these genes and regulation of gene expression, many genome wide microarray studies were carried out. The first detailed transcriptome analysis of the asexual IDC cycle using *P. falciparum* specific DNA microarray unveiled that about 60% of the genome is transcriptionally active during the IDC and most of the genes follow a cyclic expression pattern that resembles a “just in time” production strategy in which the expression of a particular gene concurs with a need for the function of that particular gene (**Figure 1.2**). For instance, DNA replication occurs in the early schizont stage and correlates well with the maximal accumulation of transcripts related to DNA replication machinery (9). Such programmed transcriptional cascade is not just restricted to the asexual stage but has also been reported for sporozoite and the gametocyte stages (10,11).

The high-throughput proteomics approach has led to the identification of most of the proteins encoded by hypothetical ORFs in the genome of *P. falciparum*. Comparative proteome analysis between four different developmental stages of *P. falciparum*: merozoite, sporozoite, trophozoite and gametocyte resulted in the identification of a distinct proteome for every stage and only 6% of the proteins were found to be overlapping between the different stages (12). Also, the timing of maximal RNA abundance does not correlate with timing of maximal protein abundance in the IDC cycle (13). Similar to the RNA expression profiles, most proteins also display cyclic abundance profiles with single peak during the IDC cycle. However, the protein level peaks at 11 hours after the peak of the corresponding transcript (14). This translational delay highlights a complex relationship between the transcript and the actual protein levels, involving many factors, post-transcriptionally and post-translationally (15). Such weak correlation between the transcript and the protein levels has also been reported for other eukaryotes including *Saccharomyces cerevisiae* (16).

Figure 1.2 The transcriptome and proteome of the IDC of the *P. falciparum*. The figure represents the transcriptome of *P. falciparum* asexual IDC cycle. The genes are clustered based on the timing of expression and the associated metabolic pathways across the 48hr lifecycle. Hence, each row represents the mRNA abundance profile of each gene. Red representing high levels of transcripts and green represent low levels of transcripts with variations of intensities in between. The transcriptome figure is reproduced from Bozdech *et al.* (9).

1.4 Transcriptional control in *P. falciparum*

Transcription initiation in *Plasmodium* is similar to the conventional eukaryotic monocistronic model which involves the formation of pre-initiation complex (PIC) at the canonical bipartite promoters. The key components of PIC including all 12 subunits of RNA polymerase II and a handful of conserved general transcription factors (GTFS) have been

identified in the parasite genome. Most of the unidentified GTFs belong to TFIID family, whose members share a conserved histone fold domain. The binding site of the TATA binding protein, PfTBP, the only member of TFIID family, which directly interacts with DNA, has been reported to lie a few hundred base pairs upstream to the transcription start site (TSS). The unique AT-rich content of the parasite genome also leads to an increase in the occurrence of TATA box-like sequences in promoters, thus leading to multiple TSS (17,18).

1.5 Cis-regulatory motifs and trans-acting factors

Promoter deletion mapping coupled with reporter assay has demonstrated that cis-regulatory motifs in *Plasmodium* are present a few hundred to thousand bases upstream to TSS and can enhance or repress promoter activity (17). This suggests that the architecture of *Plasmodium* promoters is similar to the eukaryotic bipartite promoters, which consist of core promoter regulating a basal level of transcription and upstream DNA elements controlling maximal transcriptional activity or silencing activity. With the availability of genomic and transcriptomic datasets for *P. falciparum*, there is a rapid surge in developing *in silico* tools for identifying cis-regulatory elements. The algorithm FIRE (Finding Informative Regulatory Elements) is one such tool; it utilizes the transcriptome data to search for any sequences enriched in the upstream 1kb sequences of co-regulated genes and has uncovered a wealth of putative DNA regulatory motifs in the parasite genome (19). It has been reported that only 1.3% of the parasite genome is represented by transcription-associated proteins (TAPs), in comparison to an average of 5.7% in a similar size genome of other eukaryotes (20). So far, only PfMyb1 (*P. falciparum* Myb1,) and ApiAP2, a lineage-specific family, consisting of 26 specific transcription factors (STFs), that contains the AP2 DNA Binding Domain had been identified (21,22). Hence, the current findings suggest that *P. falciparum* genome is highly underrepresented in transcription factors, which can be attributed to the atypical AT-richness composition, the complex life cycle involving multiple hosts, the repercussion of

evolutionary expansion of *P. falciparum* genome from other eukaryotes or inadequacy of the current bioinformatics tool.

In summary, it seems that specific transcription initiation in *P. falciparum* does not play an important role in gene regulation as in other eukaryotes. This hypothesis further supported by overrepresentation of CCCH-type zinc finger proteins in the genome, which is associated with mRNA stability (20) and the presence of full complement of chromatin modifiers and post-translational modifications (PTMs). All these information emphasizes the role of post-transcriptional and post translational control in gene expression in *P. falciparum* (23). The role of chromatin in gene regulation will be summarized in the section 1.8.

1.6 DNA damage and repair pathways

The genomic integrity of parasite is always under constant threat from various exogenous and endogenous agents because of the parasite complex life cycle which involves both vertebrate and invertebrate as hosts. According to an estimate, the eukaryotic genome is challenged with more than 10,000 DNA lesions per day and therefore, eukaryotes possess extensive repertoires of DNA repair enzymes, which can recognize and repair the DNA lesions (24). The main sources of genetic assaults to *Plasmodium* are replications errors, reactive oxygen species (ROS) generated during heme metabolism, by-products from host cellular respiration, host immune system, free radicals generated by antimalarial drugs such as artemisinin, etc. (25,26). DNA damage can be detrimental for the parasites as it will hinder DNA replication and transcription and would eventually lead to mutations or parasite death. In order to maintain genomic stability and replication fidelity, parasites need to repair their damaged DNA, hence similar to other eukaryotes, parasites also possess homologues of most of the DNA repair genes (7). Eukaryotic cells have the ability to detect and activate a DNA lesion specific cellular response which termed as the DNA damage response (DDR) (27).

DNA damage response is a highly orchestrated and regulated cellular response which involves a network of proteins that can be classified as sensors, mediators/transducers and effectors. Sensors are proteins which can detect any DNA aberrations, mediators/transducers are proteins which can successfully transmit the information to cell and effectors can respond to DNA damage by regulating chromatin structure, cell cycle, and DNA repair machinery (27). DDR comprises various complementary networks of DNA repair machineries, each of which can detect and repair a distinct category of DNA aberration (28). These repair machineries are classified into three broad categories:

- 1) Direct repair by reversal of DNA damage
- 2) Excision repair pathways
- 3) Double-strand break repair pathways

1.6.1 Direct repair

Direct repair is a single step repair process in which, the repair enzyme directly reverses the altered DNA base to a normal base without any intermediate steps. Most prominent examples are O6-methylguanine DNA methyltransferase (MGMT), AlkB homologue (ALKBH) and DNA photolyases. MGMT directly transfers the methyl group from O6-guanine (guanine having an extra methyl group on its oxygen atom at position 6th in the ring) to its cysteine residue which results in the irreversible inactivation of the enzyme (29). AlkB homologues are α -ketoglutarate and iron-dependent dioxygenase enzymes that directly removes the methyl group present on modified DNA bases by oxidative demethylation (30). Both MGMT and ALKBH family members are not conserved in the *P. falciparum* genome. Surprisingly, the parasite genome encodes a homologue of DNA photolyase (PF3D7_0513600) which is activated by ultraviolet irradiation and is generally absent in the placental mammals (7). No

study has been carried out so far to test the *in vitro* or *in vivo* activity of this repair enzyme during the parasite asexual or sexual stages.

1.6.2 Excision repair pathways

The excision repair pathways catalyzes the repair of damaged DNA in five distinct consecutive steps: damage sensing and recognition, dual incision of the phosphodiester backbone around the damaged site, removal of the DNA aberration, resynthesis of the DNA sequence and filling the gap by DNA polymerase and sealing of the broken ends with DNA ligase (28). Eukaryotic cells have three major excision repair systems (**Figure 1.3**):

- 1) Base excision repair (BER)
- 2) Nucleotide excision repair (NER)
- 3) Mismatch repair (MMR)

1.6.2.1 Base excision repair (BER)

The BER mode of DNA repair can recognize a variety of damaged substrate such as oxidized DNA bases, simple alkylated DNA bases, spontaneous abasic sites generated during replication, uracil residues in DNA and apurinic/apyrimidinic (AP) sites generated by DNA glycosylases. It has been estimated that about 10,000 spontaneous abasic sites (AP) are produced by a single cell every day and out of four bases, the chances of spontaneous base loss is high for guanine and adenine (33). The *Plasmodium* genome is highly AT-rich (almost 80%) and hence it has been suggested that spontaneous loss of adenine bases from parasite DNA will normally result in the formation of a large amount of AP sites in the genome, which can be repaired by the BER machinery (34).

The first step for the BER machinery is the recognition of the DNA lesion by DNA glycosylases, which cleave the N-glycosidic bond between the damaged base and sugar

Figure 1.3 Eukaryotic cells have three major excision repair pathways. **A)** The base excision repair (BER) pathway is initiated by the recognition of DNA lesion by DNA glycosylases (AAG), which generates apurinic/apyrimidinic (AP) sites. These AP sites are further processed by the AP endonucleases (APE), which incise the 5' phosphodiester bond to generate the 3'-OH and 5'-deoxyribose phosphate group (5'-dRp). Further enzymatic activities of the DNA polymerase (Pol β) and the DNA ligase fill the gap and ligate the broken ends. **B)** XPC-RAD23B acts as a DNA damage sensor and as a scaffold for other repair factors involved in the NER (31). Subsequently, to DNA damage recognition, transcription factor IIH complex and XPA are recruited to the damaged site. Further, replication binding protein A (RPA) promotes the recruitment of endonucleases such as XPG and ERCC1-XPE to the damaged site (31,32). These endonucleases remove the bulky lesions and create a gap of 27-29 nucleotides. Subsequently, the general replication apparatus resynthesize the excised region. **C)** The mismatch repair (MMR) system is activated when the MutS α heterodimer and MutL α recognize to the base-base mismatches. Diffusion of MutS-MutL produces the nick around the mismatches, and facilitates the recruitment of exonuclease (EXO1), which excise the damaged region. Subsequently, the general replication apparatus resynthesize the excised region. This figure is adapted with permission from Fu *et al.* (30).

phosphate backbone resulting in the removal of damaged bases and generation of abasic, apurinic/apyrimidinic (AP) sites (**Figure 1.3a**). These AP sites are further processed by the AP endonucleases (APE), which incise the 5' phosphodiester bond to generate the 3'-OH and 5'-deoxyribose phosphate group (5'-dRp). Further enzymatic activities of the DNA polymerase (Pol β) and the DNA ligase fill the gap and ligate the broken ends (35). The *Plasmodium* genome has 4 homologues of DNA glycosylases: A/G specific adenine glycosylases, AAG (PF3D7_1129500), PfUDG Uracil DNA glycosylases (PF3D7_1415000), DNA-3-methyladenine glycosylases (PF3D7_1467100) and N-glycosylase/DNA lyase putative (PF3D7_0917100) (7). The parasite genome also encodes for a single putative AP endonuclease (PF3D7_0305600). Uracil DNA glycosylases (UDG) are those glycosylases that can recognize and remove the uracil residues from the DNA. Recently, the glycosylase activity of *Plasmodium* PfUDG, has been characterized and till date, it is the only glycosylase that had been shown for its *in vitro* enzymatic activity in the parasite (36).

The BER pathway can recognize the AP sites by two different modes: a DNA polymerase β dependent short patch mode and a PCNA dependent long patch mode. In short patch mode, there is a replacement of only one nucleotide which followed by the sealing of broken ends and this mode is most prevalent in mammals, yeast and *Xenopus*, whereas in long patch BER repair mode, cells can replace a 6-7 nucleotide stretch (35). It has been reported that *Plasmodium* can utilize only long patch BER to repair its AP sites and it is worth to mention here that the *Plasmodium* genome does not have the DNA polymerase β . In the same study, it was shown that the cell-free parasite extracts have the ability to replace uracil residues and can repair AP sites in different DNA substrates such as closed circular plasmids or linear double stranded DNA molecules (34).

1.6.2.2 Nucleotide excision repair (NER)

The nucleotide excision repair (NER) pathway is activated in response to large bulky DNA lesions that distort the helical structure and interferes with replication and transcription. Some of the helix-distorting lesions are cyclobutane pyrimidine dimers (CPD) or 6–4 pyrimidine–pyrimidone photoproducts (6-4PPs) which are induced by ultraviolet light (28). NER recognizes damaged DNA by two mechanisms: Global genome repair (GGR) and by transcription-coupled repair (TCR) (32). GGR scans the whole genome and removes helix-distorting lesions in the non-transcribed region of the genome, whereas TCR repairs the bulky lesion in the transcribed region that halts transcription and is activated in response to stalling of RNA polymerase II during the transcriptional elongation step (37). NER recognizes damage portion of DNA by DNA damage sensors XPC/Rad4 and HR23B/Rad23 (**Figure 1.3b**) (31). Homologues of both XPC and HR23B are encoded by the parasite genome: PF3D7_1432600 is the gene ID of XPC and PF3D7_1011700 of Rad23. The other core components of the NER machinery are also conserved in the parasite genome and discussed in the results section (Table 3.2). It has been reported that *Plasmodium* cell extracts can recognize, incise and repair the plasmid DNA that had been previously exposed to the UV light but till date no study has been done to analyze the NER repair capacity in the parasite, or the role of any putative enzymes belonging to the NER pathway (38).

1.6.2.3 Mismatch repair (MMR)

The mismatch repair system is an evolutionary conserved repair pathway that plays a vital role in maintaining the replication fidelity and mutation avoidance (39). The MMR system corrects two of the major replication associated errors: base-base mismatches, which are generated by the misincorporation of nucleotides by DNA polymerase, and the other is the insertion/deletion loops (IDL), which results from slippage during replication. The MMR system has also been known to be involved in many cellular processes such as the DNA

damage response signaling, programmed cell death, cell cycle regulation, meiotic recombination, somatic hypermutation, and class switching recombination (40). The defective MMR pathway has been associated with the predisposition to numerous cancers, microsatellite instability, DNA damage tolerance and drug resistance (41,42).

It has been reported that some bacterial strains with defective MMR repair have 1000 times higher mutation rate as compared to normal strains, which plays a key role in the evolution of antibiotic resistant bacteria (43). So far, only a few studies have been done to understand the role of defective mismatch repair in the development of drug resistance in *Plasmodium*, and will be reviewed in the later section.

Much of the understanding of the MMR pathway is based on the studies of *Escherichia coli* mismatch MutS-MutL system. Several components of the MMR machinery have been identified in the *Plasmodium* genome based on their sequence homology with *E. coli*, humans and yeast. These includes three MutS homologues, two MutL homologues, two PCNA genes and one RPA gene, which is a homologue of single strand binding protein (SSB) and similarly to the other eukaryotes, the parasite genome does not have MutH homologue (7). The MMR pathway is initiated when heterodimer MSH2-MSH6 or MSH2-MSH3 recognizes and binds to base-base mismatch or IDL loops (**Figure 1.3c**). The three MutS homologues of parasite are PfMSH2-1, PfMSH2-2 and PfMSH6. It has been reported that MSH3 is not conserved in parasite genome (44). In *C. elegans* and *D. melanogaster*, only one heterodimer of MSH2-MSH6 initiates MMR response as MSH3 is also not conserved in these two organisms (45). The loss of MSH3 from the parasite genome may also suggest the role of the heterodimeric complex PfMSH2-PfMSH6 in the initiation of MMR pathway. In eukaryotes, MutL is known to interact physically with the MutS and enhances its mismatch recognition affinity. Although the MutL homologue (PfMLH/PF3D7_1117800) of *Plasmodium* has been

recently characterized for its *in vitro* ATPase and endonuclease activity, but its role in the damage recognition and its interaction with the MutS homologue is unknown (46).

1.6.3 Double-strand break repair pathways (DSBR)

DNA double-strand breaks (DSBs) are the most lethal form of genetic aberrations, or chromosomal aberrations, which if unrepaired can lead to cell death and if misrepaired can result in chromosomal anomalies. DSBs occur when the phosphodiester bonds are broken in the two complementary strands concurrently. Some of the well-known exogenous agents that can induce DSBs are ionizing radiations (gamma-rays, X-rays), chemical mutagens (hydroxyurea, aphidicolin), DNA alkylating agents (methyl methanesulphonate), DNA cross-linking agents (cisplatin, mitomycin C) and topoisomerase inhibitors (etoposide, camptothecin) (47). Endogenous DSBs result from a variety of sources, including cellular metabolism, such as reactive oxygen species (ROS) mediated DSBs, which are generated as by-products of the aerobic respiration. It has been estimated that the dividing human cells experiences, almost 50 DSBs /cell cycle and the majority of the spontaneous DSBs originate from DNA replication related processes such as stalling of replication or collision of the replication fork with the transcriptional machinery (47,48). Two distinct repair machineries have evolved in eukaryotes for recognizing and repairing the DNA double strands breaks: Homologous recombination repair (HR) and non-homologous end-joining (NHEJ). In the HR pathway, the damaged DNA is repaired using intact sister chromatid as the template to reproduce the missing information and hence, considered as the error-free mode of recombinational repair. In the NHEJ pathway, the broken ends are simply ligated together without any involvement of homologues sequences and hence, regarded as an error-prone mode of the DSB repair. The NHEJ mode is initiated with the Ku heterodimer proteins (Ku70 and Ku80) bind to the broken ends and act as the platform to recruit the mediator complex (DNA ligase IV-XRCC1) (**Figure 1.4a**). These proteins also protect the single-stranded

broken ends from the resection/nuclease activities (28). The *Plasmodium* genome does not encode for most of the homologues of NHEJ, for example, homologues of Ku70, Ku80, DNA ligase IV, XRCC1 have not been identified till date, neither by biochemical analysis nor by bioinformatics mining (7). It has been shown that the parasite utilizes alternative end-joining pathway for DSB repair, instead of canonical NHEJ pathway (49). *P. falciparum* utilizes DSB repair pathways for driving some of the specialized functions, including antigenic variation and copy number variation (25).

HR-mediated double strand break repair (DSBR) is initiated by binding of the MRN/MRX exonuclease complex (Mre11-Rad50-NBS1/Xrs2) to the damaged site (**Figure 1.4b**). The MRN complex consists of two evolutionary conserved proteins: Mre11 and Rad50, whereas NBS1/Xrs2 is poorly conserved (50). The homologues of Mre11 and Rad50 are conserved in *Plasmodium* species, whereas Nbs1/Xrs2 homologue is not found in the parasite genome (7). The MRN complex processes the 5' end of DSB via its 5' to 3' exonuclease activity and generates the 3' end overhangs, which subsequently acts as a substrate for the single-stranded binding protein, RPA. The accessory repair proteins, such as Rad52, Rad54 and BRCA2 that work together in replacing RPA with Rad51 on the single-stranded DNA (ssDNA). Rad51 and Rad54 play a pivotal role in the HR pathway. These proteins catalyze the strand invasion and strand annealing steps of the HR pathway (51). *Plasmodium* Rad51 protein, (PfRad51), Rad54 (PfRad54) and RPA (PfRPA1, PfRPA2) have been identified and biochemically characterized. PfRad54 has double stranded DNA (dsDNA) dependent ATPase activity, whereas PfRAD51 has ssDNA dependent ATPase activity *in vitro*. It has been shown that PfRad54 physically interacts with the PfRad51 and enhances PfRad51 mediated association between ssDNA and its homologous dsDNA (52,53). It has been experimentally shown that a parasite with defective HR functions, failed to repair its DSBs, shows hypersensitivity to DNA damaging agents and eventually dies. Mice infected with transgenic

Figure 1.4 Major pathways to repair double strands breaks (DSBs). The two distinct repair pathways in eukaryotes for repairing the DNA DSBs: Homologous recombination repair (HR) and non-homologous end-joining (NHEJ). **A)** The NHEJ mode is initiated with the Ku heterodimer proteins (Ku70 and Ku80) bind to the broken ends, which then recruit the DNA-dependent protein kinase (DNA-PKcs). DNA-PKcs catalyze the phosphorylation of the endonuclease, Artemis, which recruit the mediator complex (DNA ligase IV-XRCC1). **B)** The HR-mediated DSB repair is initiated by binding of the MRN/MRX exonuclease complex (Mre11-Rad50-NBS1/Xrs2) to the damaged site. The MRN complex resects the 5' end of DSB and generates the 3' end overhangs, which subsequently acts as the substrate for accessory repair proteins, such as RPA, Rad52 and Rad54. Rad51 and Rad54 proteins catalyze the strand invasion and strand annealing steps of the HR pathway. Further, the general replication apparatus resynthesize and ligate the DNA. This figure is adapted with permission from Downs *et al.* (55).

Plasmodium having dominant negative PfRad51 mutant forms, exhibit reduce parasite load and survive for longer as compared to mice infected with wild *Plasmodium* with intact

PfRad51. This study demonstrated that parasite exclusively relies on the HR repair pathways for DSBs repair (54). Although, it has been reported previously that DNA damage can induce up-regulation of PfRad51 and PfRad54, but how does HR repair machinery recognizes DSBs in the parasite has yet to be studied (52). The MRN sensory complex is conserved in the parasite genome, but its functional role in the DNA damage sensing is still unknown.

1.7 Association of defective DNA damage response and drug resistance in *Plasmodium falciparum*

The DNA repair system protects the genome from deleterious DNA lesions and mutations, but also acts as a robust mechanism to acquire drug resistance. One of the central mechanisms of the drug resistance is the loss of the MMR pathway, which results in cells being unable to recognize damaged DNA and, hence not able to initiate the DNA damage response. This results in the accumulation of a large number of mutations throughout the genome, including mutations in some of the key drug resistant genes. MMR-deficient cells are found to be non responsive to several chemotherapeutic drugs like etoposide, procarbazine, busulfan, cisplatin and doxorubicin (56). Some of the pathogenic bacterial strains, for example, *Neisseria meningitidis*, *Helicobacter pylori*, *E. coli* and *Streptococcus pneumoniae* exhibit hypermutation rate, which also results from their defective MMR machinery, contributing to the emergence of antibiotic resistant bacteria (43). Similarly, the role of MMR deficiency in drug resistance in other protozoa intracellular parasites such as *Trypanosoma cruzi* and *Toxoplasma gondii* have also been reported (57,58). Disruption of the MutS homolog, (a key DNA damage recognition protein of MMR complex) in *T. gondii* results in loss of the MMR that confers multidrug resistant phenotype to the parasite (58).

Regular exposures of different types of traditional antimalarial drugs lead to the selection of those parasites that not just show resistance to individual drugs, but also exhibit resistance to novel classes of drugs whose structure and mechanism of action are quite dissimilar from

the existing ones. Such mutator phenotypes have been referred as accelerated resistance to multiple drugs (ARMD) (59). One of the suggested mechanisms behind the emergence of ARMD phenotype is a defective DNA repair. It has been shown that the ability of DNA repair differs among different *Plasmodium* strains with varying degree of drug sensitivities. In one of the study, the efficiency of DNA repair in UV light induced DNA damage was found considerably reduced in highly resistant W2 strain, which is a known *Plasmodium* strain displaying ARMD phenotype (38). In another study, ARMD parasites, W2 and Dd2 were shown to have defective DNA mismatch repair system whereas, non-ARMD parasites HB3, D6 and 3D7 were having proficient mismatch repair systems (60).

Recently, it has been found that artemisinin resistant parasites from Southeast Asia have point mutations in the DNA repair machinery. Non-synonymous single nucleotide polymorphisms (SNPs) have been found in MMR proteins (PMS1 and MLH) and in Rad50, which is a part of MRN complex that detects DSBs (61). Although, these reports suggest an association between artemisinin resistance and DNA repair, the underlying mechanism of artemisinin resistance is not yet fully understood.

1.8 The role of chromatin dynamics in *Plasmodium* gene regulation

The 48-h *P. falciparum* asexual intraerythrocytic developmental cycle (IDC), which consists of multiple stages in RBC, is responsible for most of the pathological conditions associated with malaria. The *Plasmodium* expresses nearly 90% of its gene during the blood stage in a highly regulated ‘just-in-time’ manner (9,62). However, the paucity of recognizable transcription factors in the *Plasmodium* genome raises the question about the mechanism of controlled transcription cascade in the IDC cycle. The presence of a full complement of chromatin modifying proteins, conserved nucleosome architecture and a diverse array of post translational modifications in the *P. falciparum* genome, suggests a role of epigenetic in regulating the parasite gene expression (**Figure 1.5**) (23).

Figure 1.5 Summary of the post translational modifications (PTMs) identified in the *P. falciparum*. The malaria parasite genome encodes all the 4 core nucleosome histones, H2A, H2B, H3 and H4. The N- terminal tails of these histones have histone modifications such as acetylation (Ac), methylation (Me) and ubiquitylation (Ub). The histone-modifying enzymes catalyzing the histone acetylation (PfGCN5, PfMYST) and histone methylation (PfSET1, PfSET2, PfSET3, PfSET8, PfCARM1, PfPRMT1) have been identified in the *Plasmodium* genome and are shown here. This figure is reproduced with permission from Cui *et al.* (23).

In a recent study, it has been discovered that histone deacetylase (HDAC) inhibition causes transcriptional de-regulation of a large number of genes during the IDC affecting all three developmental stages. But in similar studies conducted on yeast and cancer cells, HDAC inhibition resulted in changes in no more than 2-17% of the genome (63). These findings also reflect that epigenetic contribute much more in regulating hard wired transcriptome during the IDC in *Plasmodium* than in yeast. Also, it has been identified that eight histone marks including H4K8Ac, H4K16Ac, H4Ac4, H3K56Ac, H3K9Ac, H3K14Ac, H3K4me3 and H4K20me1, displayed a positive association with transcription in *P.*

falciparum. Importantly, H4K8Ac mark was found primarily in the promoters of transcriptionally active genes, whereas other histone marks were mostly enriched near the 5' start site (64).

1.8.1 Chromatin architecture

Chromatin, the physiological template for all DNA-related processes, consists of the fundamental repeating unit, the nucleosome, consisting of approximately 160 bp of DNA wrapped around a histone octamer (23). The eukaryotic chromatin is not a static entity and it harbors dynamic changes from relaxed, euchromatin to transcriptionally silenced, heterochromatic state. Chromatin mediated epigenetic event imposes a profound effect on the accessibility of genes and their promoters to transcription factors, and therefore regulate transcription. The *Plasmodium* nucleosome distribution is restricted to coding regions, and their intergenic regions are generally devoid of nucleosome that may serve as a binding site for transcription factor (65). In addition, there is a strong correlation between active gene expression and nucleosome free regions at the transcriptional start site (TSS) and core promoter during the IDC (65,66). Nucleosome positioning shows the dynamic variations in the IDC cycle in *Plasmodium*, from a significant loosening of the nucleosome in the early stages of the life cycle, followed by compact nucleosome packaging at the later development stages (schizont stage) (65,67).

1.8.2 Histone and its modifications

The malaria parasite genome encodes all the 4 core nucleosome histones, H2A, H2B, H3 and H4 as well as variant histones, H3.3, H2AZ, CenH3 and H3Bv but lacks the linker histone H1 (23,68). The N-terminal tail of these histones have more than 44 post translational modifications (PTMs), some of which are unique to *Plasmodium* such as hyperacetylation of variant histone H2A.Z and H2Bv (**Table 1.1**).

Table 1.1 Summary of PTMs identified on *P. falciparum* histones.

Histones	PTM(s)*
H2A.....	N-term-ac, K3ac, K5ac
H2B.....	K112ub
H3.....	K4me, K4me2, K4me3 , K9ac , K9me, K9me3, K14ac, K14me, R17me, R17me2, K18ac, K23ac, K27ac, K36me3, K56ac , K79me3
H4.....	N-term-ac, R3me, R3me2, K5me, K5ac, K8ac , K12ac, K12me, K16ac , R17me, K20me , K20me2, K20me3
H2A.Z.....	N-term-ac, K11ac, K15ac, K19ac, K25ac, K28ac, K30ac, K35ac
H2Bv.....	N-term-ac, K3ac, K8ac, K13ac, K14ac, K18ac, T85ph
H3.3.....	K4me, K4me2, K4me3, K9ac, K14ac, R17me, R17me2, K18ac, K23ac, K27ac

Adapted from Cui *et al.*, 2010 (69)

*bold indicates those histone modifications, which show strong association with transcription across the parasite asexual life cycle (64).

Histone modifications such as acetylation, methylation, phosphorylation, sumoylation and ubiquitylation etc, along with incorporation of variant histones are thought to have a role in regulating the chromatin structure (23). Acetylation of histone tails at specific residues leads to chromatin decondensation and makes the chromatin more accessible to transcription whereas histone methylation can lead to both transcription activation and repression depending upon the specific residues involved (70).

In *Plasmodium*, the amount of histone modifications presumed to be associated with the euchromatin state is higher compared to vertebrate host (64). It has also been reported that the occupancy of the variant histone, H2A.Z has been found to colocalised with the active gene marks H3K9Ac and H3K4me3 in the intergenic regions during the entire IDC cycle (71). In contrast, H3K9me3, a heterochromatin marker of model eukaryotes, is not found outside the

silenced subtelomeric region of *Plasmodium*, indicating that H3K9me3 is not a mark for transcriptional inactivity in the parasite (72). Taken together, the epigenomic profiling is highly dynamic and complex during the course of 48 hour IDC cycle.

The combinatorial or sequential combination of different histone modifications make the ‘histone code’ and the outcome of a particular code, depends at the recognition by effector proteins (73). Such effector proteins are large multi-protein complex having evolutionally conserved domains such as bromodomain, chromodomain, PHD fingers etc. Many effector proteins in *Plasmodium* such as histone acetyltransferase, PfGCN5 and methyltransferase PfSET1 have bromodomains that can recognize acetylated lysine residues (23). Similarly, heterochromatin protein PfHP1, specifically reads the H3K9 trimethylation marks on the silenced var gene loci, via its conserved chromodomain (23).

1.8.3 Chromatin modifiers associated with histone acetylation and methylation

Histone acetylation by histone acetyltransferases (HATs) and deacetylation by histone deacetylases (HDACs) are key components of the epigenetic machinery, which regulates transcription in eukaryotes. PfGCN5 is a well characterized histone acetyltransferase enzyme in *P. falciparum*. It acetylates H3K9, 14 and to a lesser extent H4 and has been shown that inhibition of PfGCN5 by curcumin and anacardic acid leads to growth inhibition (69). PfMYST is a HAT, which belongs to the MYST family of histone acetyltransferase and similar to the other eukaryotes, it has been shown to acetylate lysine residue 5,8,12 and 16 on H4 (74).

P. falciparum has five putative histone deacetylase enzymes (HDACs), which covers all three classes of HDACs. Out of all, class III members PfSir2A and PfSir2B are extensively studied in the parasite. PfSir2A catalyzes the NAD-dependent deacetylation of H3K9Ac, H3K14Ac and H4K16Ac and also plays an important role in maintaining the heterochromatin

state in subtelomeric regions and controls the mutually exclusive expression of *var* genes (75).

Two main classes of methyltransferases are histone lysine methyltransferases (HKMTs) and protein arginine methyltransferases (PRMTs). HKMTs are characterized by a SET methyltransferase domain and there are 10 proteins identified in the *P. falciparum* having this SET domain. PfSET1,-2,-3, and-8 methylate at H3K4, H3K36, H3K9, and H4K20, whereas PRMT methylates at H4R3me1 and 2 (69). The parasite genome also encodes two classes of histone demethylases: lysine-specific demethylase 1 (LSD1) and Jumonji-C domain (JmjC) containing histone demethylases (JHDMs). PfJmjC1 and PfJmjC2 are the lysine-specific demethylases that catalyze the demethylation of H3K9me and H3K36me (76).

1.9 Role of chromatin dynamics in DNA damage response (DDR)

In eukaryotic cells, DNA damage recognition, signaling and repair occur in the context of chromatin (77). DNA damage results in relaxation of chromatin at damaged foci; this loosening of chromatin allows the DNA repair factors and transcription modulators to gain access into the damaged site (78). Two main mechanisms that drive the alteration in chromatin architecture during DNA damage are 1) chromatin remodeling complex dependent alteration and 2) covalent modifications of histone tails.

1.9.1 Chromatin remodeling complex

The ATP-dependent chromatin remodelers alter the nucleosome packaging and positioning around the damaged site, which expedites the access of DNA damage sensors and other repair related proteins (79-81). These remodelers can modulate the chromatin accessibility by nucleosome eviction or by promoting histone exchange at the damaged site. SWR1 complex is believed to be responsible for facilitating histone exchange by replacing canonical histone H2A with histone variant H2AZ near the DSB site (82,83). TIP60/NuA4 complex is a multi-subunit chromatin remodeling complex, which plays a significant role in the DDR. Studies

performed on yeast and mammalian cells, have reported that DSBs induced either by the site-specific endonucleases or genotoxic agents results in the active recruitment of TIP60 complex to the damaged site (79). This complex has two main catalytic subunits: 1) its TIP60 subunit has the HAT activity and is responsible for DNA damage induced hyperacetylation of lysine residues of H4 and H2A and 2) p400 ATPase subunit which facilitates H2AZ, a histone variant, deposition at the damaged site (82). Hyperacetylation of H4 in chromatin domains, together with p400 ATPase mediated H2AZ exchange results in “nucleosome disassembly” and “nucleosome-free” region surrounding the damaged site (82).

The mammalian TIP60/NuA4 complex has 16 subunits and only three carries the catalytic activities (82). Homologues of these three catalytic subunits have been identified by sequence alignments in the *Plasmodium* genome (23,84). These are: 1) TIP60 histone lysine acetyltransferase (PfMYST), 2) p400 ATPase (SWI2/SNF2 ATPase) and 3) RuvB helicase (PfRuvB helicase).

PfMYST is a HAT, which belongs to the MYST family of histone acetyltransferase and similar to the other eukaryotes, it has been shown to acetylate lysine residues 5,8,12 and 16 on H4 (74). The parasite genome possesses at least 11 SWI2/SNF2 ATPase motor proteins which are homologues of p400 ATPase (23). Similarly, three homologues of RuvB proteins have been identified in the parasite genome (84). The preservation of catalytic subunits in the parasite reflects the conserved evolutionary functioning of NuA4/Tip60 complex, but studies need to be undertaken to demonstrate its role in DNA damage/DSB repair.

1.9.2 Histone modifications in DNA damage response (DDR)

Chromatin modulation by post translational modifications of histone tails, which includes acetylation, phosphorylation, methylation, etc. play a significant role in several DNA metabolic processes including DNA replication, transcription, repair and recombination. Here, we reviewed the role of histone modifications in DDR in eukaryotes and compared

Table 1.2 List of histone modifications that are associated with the DNA damage response in yeast and mammals.

Histone	<i>S. cerevisiae</i>		Mammals		<i>P. falciparum</i>	
	Modification	Enzyme	Modification	Enzyme	*Modification	Enzyme
H2A / H2AX	pS129 ⁽⁸⁵⁾	Mec1 and Tel1	pS139	ATM/ATR kinase	pS120	ND
H3	K9ac ⁽⁵⁵⁾	GCN5	K9ac	GCN5	K9ac	PfGCN5
H3	K56ac ⁽⁵⁵⁾	GCN5	K56ac	p300	K56ac	ND
H3	K79me ⁽⁸⁶⁾	Dot1	K79me	DOT1L	K79me	ND
H3	K4me3 ⁽⁸⁶⁾	SET1	K4me3	SET1	K4me3	PfSET1
H4	K5ac ⁽⁵⁵⁾	Esa1	K5ac	TIP60	K5ac	PfMYST
H4	K8ac ⁽⁵⁵⁾	Esa1	K8ac	TIP60	K8ac	PfMYST
H4	K12ac ⁽⁵⁵⁾	Esa1	K12ac	TIP60	K12ac	PfMYST
H4	K16ac ⁽⁵⁵⁾	Esa1	K16ac	TIP60	K16ac	PfMYST
H4	K20me2 ⁽⁸⁶⁾	MMSET	K20me2	MMSET	K20me2	PfSET8

(The symbols p, ac and me corresponding to histone modification: phosphorylation, acetylation and methylation respectively.) (ND - not defined)

*Histone marks identified in the *P. falciparum*, but their roles in DNA damage regulation have not been studied. Information of histone modifications and enzymes for *P. falciparum* were taken from previous publications (24,87).

with the *Plasmodium* chromatin biology in order to identify the potential area, which needs to be explored in the parasite DNA damage response (Table 1.2).

1.9.2.1 Histone phosphorylation

One of the first histone modification that is induced within minutes after DNA damage or DSBs is the phosphorylation of H2A, at serine (S) 129 in yeast or at serine 139 in the histone variant H2AX in higher eukaryotes. The kinases responsible for this phosphorylation are ATM kinase, ATR kinase and DNA-dependent protein kinase (DNA-PKs), which belongs to phosphatidylinositol-3 kinase-related kinase (PIKK) family (88). The recognition motif of PIKK kinases is SQ motif (Ser-Gln) at the C-terminal of H2A/H2AX. H2A/H2AX phosphorylation occurs on the serine residue of the SQ recognition motif, which is the fourth residue from the C-terminal tail of either H2A or H2AX and, following SQ residue, there is one acidic residue (D/E) and then another hydrophobic residue (L/I/Y). This SQ motif is evolutionary conserved and has been identified in mammals, fruit flies, chick peas, yeast, *Giardia* and *Arabidopsis* (89). Recently, this conserved SQ (E/D) motif has been identified in the H2AX histone variant, of another, Apicomplexan obligate parasite, *T. gondii*. Similarly to the other eukaryotes, DNA damage also induces H2AX phosphorylation in *T. gondii* (90).

The *Plasmodium* genome encodes four canonical histones (H2A, H2B, H3 and H4) and four histone variants (H2A.Z, H3.3, CenH3 and H2B.Z), but doesn't encode histone variant H2AX. Bioinformatic analysis reveals that *Plasmodium* H2A (PF3D7_0617800) has the conserved SQ sequence, near the C-terminal end. Recently, a phosphorylation site has also been identified at serine-121 position on this SQ sequence by a sensitive mass spectrometry (91). Further studies need to be carried out in order to elucidate the role of this conserved sequence in the DNA damage response in the malaria parasite.

1.9.2.2 Histone acetylation

H4 acetylation is one of the well-documented histone modifications known to play a role in DDR (79). MOF, Esa1 and Tip60 are members of MYST family of histone lysine acetyltransferases (HATs), which acetylates H4 and have been reported to play crucial roles

in DNA damage induced repair (79). Tip60 is a HAT subunit of NuA4 complex, which is known to acetylate lysine residues 5, 8, 12 and 16 on H4. The loss of H4 acetylation either by depletion of Esa1, a yeast homologue of Tip60 or mutation in the lysine residues led to defective DNA repair in yeast (92). H4K16 is acetylated by two HATs: Tip60 and MOF. It has been found that cells deficient in MOF activity have reduced H4K16Ac levels and have defective DNA repair (93). The MRN complex mediates the recruitment of both ATM kinase and Tip60 HAT to the DSB site. Tip60 acetylates lysine residues of H4 and facilitates chromatin decondensation, and activates ATM kinase (79). The role of Tip60 in DNA damage response in Apicomplexan parasites has also been reported. TgMYST-B, a Tip60 homologue of *T. gondii* has been found to mediate the ATM kinase dependent protection from DNA damage induced by alkylating agent, MMS (94). Recently, PfMYST, a Tip60 *Plasmodium* homologue, has also been studied in the context of its role in DNA damage response. PfMYST overexpression confers parasite enhance protection from DNA damaging agents MMS and camptothecin (74). However, the underlying molecular mechanism that can bridge the association between the role of PfMYST and DNA damage repair in the malaria parasite need to be investigated.

H3K56Ac is another conserved histone mark that plays a substantial role in maintaining genomic stability. H3K56Ac was found to be induced in response to DNA damaging agents such as MMS, hydroxyurea and camptothecin and induced marks were found to be colocalised with other repair factors at the DNA repair site (95,96). Yeast cells deficient for H3K56 acetylation show higher susceptibility to DNA damaged induce death, which reflects the significant role of H3K56Ac in DDR (97). Using CHIP-chip assay, it has been shown that MMS induced DNA damage resulted in the enrichment of H356Ac on several genes of DNA damage repair pathways (98). In another report in mammalian cells, DNA damage induced by MMS and hydroxyurea resulted in global enhancement of H3K56Ac and H4K16Ac levels

(99). Interestingly, a contradictory study showed that the DNA damage resulted in a reduction of H3K9 and H3K56 acetylation in mammalian cells (100).

1.9.2.3 Histone methylation

Recently, H4K20me₂ and H3K79me are the two histone lysine methylations, that have been shown to play an essential role in maintaining genomic integrity (83). Although, levels of these two histone marks are not induced during DNA damage repair but their basal levels on chromatin act as a docking site for recruitment/binding of repair proteins (101). Loss of histone methyltransferase SET8, which mediates H4K20 methylation, prevents the active recruitment of repair proteins 53BP1/Crb2 in humans/yeast (102). Similarly, H3K79me has also been reported to mediate the recruitment of 53BP1 at DSBs by acting as a docking site for binding of Tudor domain of 53BP1 (101). It has been suggested that alteration in chromatin architecture during DNA damage by histone acetylation and chromatin remodelers uncovers these constitutively expressed methylation marks that may act as a scaffold for recruitment/binding of repair proteins (81). H3K4me₃, a highly conserved euchromatic mark which is known to be associated with active gene transcription, has recently been found to be important in maintaining genomic stability in yeast. The loss of H3K4me₃ in yeast resulted in impaired DNA repair and cells displayed hypersensitivity to DNA damaging agents (103). These three methylated histone marks have been identified in the *Plasmodium* chromatin, but their role in parasitic DNA damage response needs to be examined (23).

1.10 Role of chromatin modifications in preventing spontaneous mutation

Chromatin dynamics plays a significant role in regulating the physiological processes such as replication, transcription, recombination and DNA repair. However, the role of chromatin in the context of spontaneous mutagenesis is poorly studied. Recently, it has been found in human cells that H3K36me₃ plays a profound role in preventing spontaneous mutations by regulating the mismatch repair system (MMR) (104). Cells depleted of H3K36

trimethyltransferase SETD2, showed characteristics of mutator phenotype, which includes elevated spontaneous mutations and microsatellite instability that are the hallmark of MMR deficiency (104). In another report, it has been suggested that H3K56 acetylation and deacetylation coordinates with MMR system in preventing generation of point mutations (105). In another genome-wide comparative study, it was found that chromatin architecture, in particular, up-regulation of a heterochromatic mark H3K9me3 was strongly associated with mutation rate variability in human cancer cells (106).

1.11 Mechanism of action of artemisinin, the first line antimalarial drug

Artemisinin is the most potent antimalarial drug established as the first line of defense against multiple drug-resistant strains of *P. falciparum* (107). The antimalarial activity of artemisinin is due to the presence of a characteristic endoperoxide bridge that is crucial for its activity (108). Elucidation of its mode of action will aid in the development of artemisinin derivatives with better efficacy, improved anti-parasitic activity and will also facilitate in the understanding of artemisinin resistance mechanisms. Two types of mode of action of artemisinin have been postulated: 1) Heme or ferrous ion dependent bioactivation of artemisinin or 2) Heme independent artemisinin action that involves direct inhibition of parasite calcium pump, PfATP6.

Hemoglobin digestion takes place inside the parasitic digestive vacuole, which releases a large amount of heme or ferrous ion which has been suggested as a prerequisite requirement for the bioactivation of artemisinin (26). It has been reported that ferrous iron from heme catalyzes the cleavage of the endoperoxide bridge of artemisinin. This ring opening results in the generation of artemisinin-derived free radicals or ROS (26). It has been shown that the parasite lacking falcipain-2, a cysteine protease, (which is required for hemoglobin digestion) becomes less susceptible or sensitive to artemisinin (109). Artemisinin is more active in later

stages of the parasite asexual cycle (trophozoite/ schizont) which coincides with peak hemoglobin uptake and digestion (109).

An alternative mode of artemisinin action has been proposed, which is heme-independent and involves a direct inhibition of *Plasmodium* orthologue of sarco-endoplasmic reticulum membrane calcium ATP-ase (SERCA) (110). *P. falciparum* has only one SERCA, known as PfATP6ase (PF3D7_0106300). Artemisinin and its structural homologue, thapsigargin, a well-known mammalian SERCA inhibitor was found to inhibit PfATP6ase. It was reported that both compounds irreversibly inhibited PfATP6ase activity, when expressed in *Xenopus laevis* oocytes (111). In another study, using a computational approach, it was postulated that mechanism of action of artemisinin involves both heme-dependent activation as well as direct binding and inhibition of the calcium pump, PfATP6 (112).

Artemisinin also possesses anti-proliferative activities; hence it has also been explored as an anti-cancer agent. It has been reported that free radicals or ROS derived from artemisinin cause DNA damage, particularly DSBs in cancer cells (113). Artemisinin induced DNA damage results in activation of the DDR signaling cascade, which includes activation of ATM/ATR kinase, H2AX phosphorylation and induction of DSB repair processes such as homologous recombination (HR) and non homologue end joining (NHEJ) (113,114). Recently, it has been suggested that artemisinin induces oxidative stress, which may also result in DNA damage in *P. falciparum* (115). In the same study, they analyzed artemisinin induced DNA damage in the parasite using the comet assay. Until now, this was the only report that suggests DNA damage could be one of the potential modes of action of artemisinin (115). Further studies are needed to carry out to understand the dynamics of artemisinin induce DNA damage response in the malaria parasite.

1.12 Transcriptional heterogeneity in *Plasmodium*

Most of the genes that expressed in the *Plasmodium* IDC cycle are highly regulated and expressed only when their products are needed (9). There is a conservation of such transcriptional cascade or gene expression profile across the cycle in genetically different parasite lines (116). The transcriptional variability in *Plasmodium* has been reported for the subtelomeric gene families, such as *var*, *rif*, *stevor* and *Pfmc-2TM*, which are highly polymorphic and show clonally variant gene expression (117). Clonally variant gene expression is defined as the variation in the gene expression among the subclones. Recently, it has been shown that there are extensive transcriptional differences among the subclones during the asexual cycle in the different clonal population of 3D7, 7G8, D10 and HB3 strains. About 70% of the genes, which show differential expression among the three genetically different strains 7G8, D10 and HB3, are also differentially expressed within clonal parasite populations (118). Hence, there is a significant overlap in the gene families differentially expressed between strains and within the isogenic clonal population. In malaria, most of the differentially expressed genes found are subtelomeric and encode proteins exported to red blood cell membrane including exported HSP40-type cochaperones (*dnaj*), predicted exported protein kinases (*fikk*), exported proteins of unknown function (such as *hyp*, *phist*, and *gbp* families) (118). Furthermore, differential expression is also detected in gene families of lipid metabolism (such as acyl-CoA synthetases, acyl-CoA binding proteins, and lysophospholipases), transcription factors (ApiAP2) and host cell invasion (*clag*) (118).

In order to survive and promote virulence, *P. falciparum* exports proteins to the erythrocyte surface. These exported proteins have to traverse many membranous networks including the parasitophorous vacuole, which encompasses all asexual stages inside infected RBC and the erythrocyte membrane. It has been shown that a conserved pentameric sequence, PEXEL motif (for *Plasmodium* export element), plays a vital role in protein export

into the erythrocyte (119). Most of the exported proteins belong to subtelomeric gene families, and the PEXEL motif is specific and conserved across the *Plasmodium* genus. Clonally variant gene families like *var*, *rifin*, *stevor* are well characterized members of *P. falciparum* exportome whereas *hyp*, *fikk* and *phist* are recently discovered exported proteins of unknown function (120).

1.12.1 Variant expression of invasion related proteins

The invading merozoites express different parasitic ligands and use multiple redundant pathways to infect the RBC. Erythrocyte binding antigen (EBA) family and the reticulocyte homology (PfRh) family are parasitic ligands that play crucial roles in the parasite invasion (121). Members of both these families undergo variant gene expression in both the culture adapted parasites and the field isolates. It has been demonstrated that the parasite expressing primarily EBA-175, EBA-140, and PfRH1 together will express comparably low levels of PfRH2a, PfRH2b and PfRH4 and vice versa (122). The members of rhoptry protein 1/cytoadherence-linked asexual gene (RHOPH1/CLAG) family that have the potential role in erythrocyte invasion also exhibit differential expression (123).

1.12.2 Variable/ Differential gene expression: role in pathogenesis

Clonally antigenic variation is the major survival strategy evolved in many mammalian pathogens to evade the host immune surveillance. *P. falciparum* erythrocyte membrane protein 1 (PfEMP1) is a major variable antigen that is encoded by *var* gene family, a multigene family having 60 members and in each parasite, only one *var* gene is expressed and the rest are transcriptionally silenced (124). Such strict mutually exclusive expression of *var* gene enables the parasite to escape from recognition by host immune system and hence, plays a central role in the malaria pathogenesis (117). Also, PfEMP1 mediated cytoadherence of infected RBC prevents the splenic clearance of the malaria pathogens which causes diseases like cerebral malaria and pregnancy-associated malaria (125). Variable gene

expression of invasion related proteins also equally contribute to the severity of malaria. It allows different EBA and PfRH proteins to interact with different erythrocyte receptors. Hence, varying the expression of these invasion related proteins confer flexibility to the parasite to invade erythrocyte with different stages of maturity (122). Recently, it has been shown that variant gene expression can mediate the parasite adaptability to changes in the environment. The parasite cell lines grown under conditions of periodic heat shock simulating the fluctuations during the recurrent fever in malaria, demonstrated higher resistance to heat shock, indicating that the transcriptional variations lead to phenotypic plasticity (118). Taken together, variant gene expressions confer the following advantages to the parasite, 1) evasion of host immune system and cytoadherence, 2) host cell invasion and 3) adaptation to stress.

1.12.3 Epigenetic basis of known differentially expressed genes

The dynamic histone modification patterns around the transcriptional start site (TSS) of *var* gene (PfEMP1) have been proposed to be one of the key factors governing the differential gene expression of *var* genes (126). The 5'upstream region of the single expressed *var* gene is enriched in H3K9Ac, H3K4me3, and H3K4me2 in the ring stage, where it is transcriptionally active (**Figure 1.6**). On the other hand, in the later developmental stages of the IDC, when its expression is poised (temporarily repressed), the upstream regions are enriched in H3K4me2 only. Therefore, it is believed that H3K4me2 acts as a cellular memory mark that allows reactivation of the same marked gene in the next cycle (126). In comparison, the 5' flanking regions as well as the coding region of the silenced *var* genes are enriched with H3K9me3 and depleted in H3K9Ac, H3K4me3, and H3K4me2 (126,127). The enrichment of the heterochromatin mark, H3K9me3, remains unchanged during the whole cycle in silenced *var* genes (126).

Similarly to the *var* gene, variant gene expression among *rif* gene family members has also been reported to be associated with the dynamic epigenetic marking of H3K9Ac (128). Chromatin mediated variant gene expression is not exclusive to *var* and *rifin* gene families but has also been reported for genes encoding invasion related proteins. Differences in expression of PfrH4 protein, an invasion protein, between two strains is found to be well correlated with differences in H3K9 acetylation and H3K9me3 (129).

Also, mutually exclusive expression of *clag* 3.1 and *clag* 3.2 genes are regulated at the epigenetic level (130). Silencing of *clag* 3.1 or *clag* 3.2 gene is correlated with the localization of H3K9me3, a heterochromatin mark, whereas activation of either *clag* 3 gene is associated with the euchromatic enrichment of H3K9Ac and H3K4me3 in the 5'upstream regions (123).

Figure 1.6 Relation between *var* gene expression and various chromatin marks. When *var* gene is switched on in the ring stage, the 5'flanking region near to the exon 1 is enriched with H3K9Ac, H3K4me3 and H3K4me2. During the poised stage, it remains enriched with H3K4me2 only but in both stages this region remains depleted for H3K9me3. However, in transcriptionally silenced state, out of 4 PTMs mentioned above, only H3K9me3 marks the exon1 in all the developmental stages. Adapted with permission from Lopez-Rubio *et al.* (126).

1.13 Objective of thesis

The parasite genetic integrity is under constant threat from various exogenous and endogenous stresses that can induce DNA aberrations, which can jeopardize the normal functioning of cells and can also alter the fidelity of the genetic transmission to the daughter cells (25). For this, eukaryotic cells developed a complex DNA damage repair system. The malaria parasite genome encodes the majority of DNA repair proteins, but most of these repair factors remain uncharacterized biochemically. Hence, we are still lacking much of the understanding of how the parasite responds to the damaged DNA. The paramount significance of studying DNA damage repair in the parasite is highlighted by the fact that it drives two of the basic parasitic functions like antigenic variation and copy number variation (25). The antigenic variation and copy number variation are the two indispensable tools of malarial pathogenesis. Understanding, DNA damage response in parasite may open new avenues for identifying novel drug targets. Here, for the first time, we studied the genome-wide transcriptional and epigenetic response of *P. falciparum* to DNA damaging agents.

This PhD project had three main objectives:

1) DNA damage response in *P. falciparum*: transcriptional and epigenetic events following DNA damage in the malaria parasite. First, using DNA microarrays, we have characterized the global transcriptional response of *P. falciparum* to methyl methanesulphonate (MMS), which alkylates DNA bases and causes strand breaks. DNA damage recognition and repair occur in the context of chromatin. In eukaryotes, it has been established that chromatin plays a significant role in regulating DNA damage response (77). Till now, the role of chromatin dynamics in DNA damage response in *P. falciparum* has not been explored. Hence, alongside, we studied the epigenetic regulation of DNA damage response in the parasite. Further, artemisinin, the primary antimalarial drug for *P. falciparum* infections has been long debated for its mechanism of action (26). In cancer cells, it has been

reported that artemisinin induces DNA damage and elicits various DNA repair pathways (113). Our objective was to identify whether artemisinin can induce DNA damage response in the parasite. Hence, we characterized the parasite transcriptional and epigenetic response to physiological doses of artemisinin and other antimalarial drugs.

2) Association of defective DNA damage response and mutator phenotype/drug resistance in *P. falciparum*

The rapid origin and spread of antimalarial drug resistance in Southeast Asia has been suggested due to the presence of hypermutable parasites (59). It has been demonstrated that these hypermutable parasites (ARMD) have higher tendency to develop/acquire drug resistance to novel drugs (59). The existence of such mutator phenotypes that exhibit elevated mutation rate is quite frequent in bacterial cells, yeast and mammalian cells (43). One of the most common mechanisms associated with the development of mutator phenotype is alterations in proteins belonging to the DNA repair and DNA replication machinery (56). Previously, it has been suggested that the DNA repair machinery in the malaria parasites can also contribute to acquiring the mutator phenotype (60). But, the underlying genetic and biochemical mechanisms responsible for the mutator phenotype in *P. falciparum* are still not understood. Hence, we decided to assess the transcriptional and epigenetic responses of mutator/ARMD (Dd2 and W2) and non-ARMD parasites (3D7 and D6) in the presence of DNA damaging agents. Our primary objective was to identify whether ARMD parasites possess transcriptionally active DNA repair machinery or not when challenged with the genotoxic agents. Further, we carried out whole genome sequencing of these parasites to understand the genetic basis of accelerated resistance to multiple drugs phenotype (ARMD).

3) Chromatin mediated regulation of differential gene expression in *P. falciparum*

Our third objective was to investigate the role of epigenetic mediated mechanism in the differential gene expression in *P. falciparum*. To achieve this, we compared the gene

expression and histone modification patterns (H3K9Ac and H3K56Ac) across the IDC cycle between Dd2 and T996 strains. Further, we characterized one of the AP2 domains containing transcription factor (putative), PF14_0271, which was found to be both differentially acetylated and differentially expressed between *P. falciparum* strains.

Chapter 2

Materials and Methods

2.1 Parasite culture

P. falciparum Dd2, W2, 3D7 and T996 strains were cultured in accordance with the established protocol for continuous culturing of *P. falciparum* (131). The parasites were grown in the human red blood cells in the RPMI 1640 media (Gibco, USA) supplemented with 10mg/l gentamycin (Gibco, USA), 2g/l sodium bicarbonate (Sigma, Germany), 0.25% Albumax II (Gibco, New Zealand) and 0.1 mM hypoxanthine (Sigma, Germany). The Parasite culture was kept at 2% hematocrit, with 5- 10% parasitemia in the sterile tissue culture flasks. The pH of the culture media was maintained at pH 7.4. Parasite cultures incubated at 37°C under the controlled atmosphere of 5% CO₂, 5% O₂, and 90% N₂. Parasites were synchronized by two successive treatments with the 5% D-Sorbitol (Sigma, USA) in a period of 10-12 hours. Blood smears stained with Giemsa (Sigma, Germany) were made to observe the different stages of the IDC cycle under the microscope and to calculate the parasitemia.

2.2 Drug treatments

The four drugs used in our study for different treatment experiments were: MMS (0.05%), cisplatin (100µM), etoposide (100µM) and TSA (IC₉₀/50nM). MMS is a water-soluble drug and hence prepared in the water. Cisplatin was prepared in DMF, whereas etoposide and TSA dissolved in DMSO. Drug treatments were carried out in the synchronized *P. falciparum* 3D7 strain at the trophozoite stage. For all drug treatment experiments, the parasite culture was kept at 5% parasitemia and 2% hematocrit.

The four antimalarial drugs studied were artemisinin (1µM), chloroquine (30µM), mefloquine (5µM) and pyrimethamine (5µM). Chloroquine was dissolved in the water, and the rest three drugs were dissolved in DMSO.

2.3 RNA isolation

RNA isolation was performed as previously described (132). Briefly, synchronized parasite cultures were harvested by the centrifugation at 2500 rpm for 5 minutes and then washed with PBS twice to obtain parasites infected RBC pellets. These infected cell pellets were immediately frozen in the liquid nitrogen and stored at -80°C until the next step of the RNA isolation. To lyse the cell pellet, TRIzol (Invitrogen) was added, for rings stage 10 pellet volumes were added and for all other stages, 20 pellet volumes were added, followed by centrifugation at 4500 rpm for 10 min. Upon centrifugation, the aqueous supernatant layer (containing RNA) was mixed with isopropanol (Fisher Scientific) and kept at -20°C overnight for the RNA precipitation. Following precipitation, another round of centrifugation was performed and after careful removal of the supernatant, a pellet containing the RNA appeared. The resulted pellet was washed with the 70% ethanol (Merck), air-dried at room temperature and re-suspended in nuclease-free water (Promega, USA) and stored at -80°C. Before proceeding for the microarray, the integrity of the RNA sample was checked by 1% agarose gel electrophoresis.

2.4 cDNA synthesis by reverse transcription

4µg of 1:1 mixture of random primers: N9 and oligo-dT25 oligonucleotides (Proligo) and 12µg of total extracted RNA was put together for initial priming. Equal amounts of dATP, dCTP, dGTP, dTTP (Fermentas), 5-aminoallyl-dUTP (Biotium, USA) and RevertAid H Minus M-MuLV reverse transcriptase (Fermentas) were added to the above mixture and incubated at 42°C for reverse transcription. Followed by, purification of the aa-dUTP-containing cDNA using MinElute PCR Purification Kit (Qiagen, USA). Equal amounts of RNA from 6-time points corresponding to the 8-hour interval stages throughout the IDC of

the 3D7 strain were mixed together to obtain a common reference pool, which subsequently used for all the microarray hybridizations.

2.5 cDNA labeling and microarray hybridization

Reverse transcribed cDNA samples were labeled with the fluorescent cyanine dyes CyTM3 or CyTM5 (Amersham Biosciences) by incubating the respective dyes in 0.1M NaHCO₃ for 2 hours. Subsequently, the labeled samples were purified using MinElute PCR purification kit (Qiagen). The time course samples from both the strains were labeled with the Cy5, and the common reference pool was labeled with the Cy3. Equal amount of Cy3 labeled sample and Cy5 labeled reference pool were mixed with 20X SSC, 1M HEPES and 10% SDS, followed by boiling this hybridization mixture at 100°C for 5min and then cooling it at room temperature for an additional 5 minutes. All samples were prepared under the identical conditions in the dark. cDNA hybridizations of Cy3 and Cy5 labeled samples were carried out at 65°C for 16-18 hours over the Maui hybridization system (BioMicro Systems, USA). After hybridizations, arrays were washed with SSC and scanned using GenePix 4000B (Axon Instruments, USA). For ChIP-chip, microarray hybridization was carried out at 63.5°C using either Agilent hybridization system (Agilent technology) or over the Maui hybridization system (BioMicro Systems, USA).

2.6 Microarray data analysis

Microarray images produced after scanning were analyzed using GenePix Pro 6.0 program (Axon Instruments). The resulted microarray data was normalized based on the LOWESS (locally weighted scatterplot smoothing) normalization method and then filtered for the signal intensity more than background intensity plus two times the standard deviation of background intensity for each Cy channel, in Acuity 4.0 software (Axon Instruments, USA). The software then generated the log-transformed ratios of Cy5 intensity to Cy3 intensity. Hierarchical

clustering analysis was performed on these log-transformed ratios using Gene Cluster program version 2.11 and results were visualized using the Java Treeview version 1.60 software (<http://rana.lbl.gov/EisenSoftware.htm>). Further analyses of the log-transformed ratios were carried out using Pearson's correlation and Euclidean distance method.

2.7 Chromatin immunoprecipitation

Chromatin immunoprecipitation was performed as previously described (64). Briefly, synchronized parasite cultures were lysed with saponin (0.1%), and the resulted pellet was washed with the PBS three times. Isolated parasites pellets were treated with 0.5% final concentration of formaldehyde for 10 min at room temperature to promote crosslinking between the DNA and protein components of the chromatin. After 10 minutes, glycine solution (0.125M) was added to quench the crosslinking, which was followed by two washing step. Parasites were incubated with the lysis buffer (25mM Tris-HCl pH 7.8, 1mM EDTA, 0.25% Nonidet P-40, protease inhibitor cocktail, 2mM PMSF) on ice for 1 hour and then lysed using dounce homogenizer. After homogenizing, the nuclei pellet was obtained by centrifugation, which was further dissolved in the SDS containing buffer (1% SDS, 10mM EDTA, 50mM Tris-HCl pH 8.0) and then incubated for 15 minutes on the ice. Following incubation, nuclei were sonicated 8 times at 25% amplitude (10 seconds pulse on, 50 seconds off) and centrifuged at 1300 rpm for 10 minutes. The resulted supernatant from the centrifugation contains the DNA fragments in the range of 200 to 1000 bp. ChIP dilution buffer (0.01% SDS, 1.1% triton X-100, 1.2mM EDTA, 16.7mM Tris-HCl pH 8.0, 167mM NaCl) was added to the supernatant. From the mixture, an aliquot was kept as input DNA and the rest was subjected to chromatin immunoprecipitation as described by Millipore ChIP Assay Kit (#17-295). The concentration of the antibody was kept at 1: 200 dilutions (H3K9Ac, H4K8Ac and H3K56Ac specific antibody). Both immunoprecipitated DNA and

input DNA (sonicated genomic DNA) was exponentially amplified using the random primers 5' GTTTCCTCAGTCAGGATCNNNNNNNNN 3' and 5' GTTTCCTCAGTCACGATC 3' as described by Gupta et al (133). Following amplification, the amplified DNA was purified using MinElute PCR purification kit (Qiagen) and eluted in 15µl of elution buffer. The input pool DNA was prepared by mixing equal amounts of input DNA samples from 6 time points of the Dd2 and T996 strain. For microarray, the input pool was labeled with Cy3, whereas immunoprecipitated DNA with Cy5. Additionally, for ChIP-chip with MMS treated samples, the input was made by mixing equal amounts from MMS treated and untreated samples. To perform microarray with the amplified DNA, 3 microgram of the amplified and input DNA were used for labeling with Cy5 or Cy3 dye. The labeling was done in the dark for 3 hours.

2.8 Western Blot/ Immunoblot analysis

MMS treated synchronized parasite culture was harvested using 0.1% saponin (Sigma). Saponin lysis was carried out for 5 minutes at room temperature and followed by washing the parasites with ice cold PBS (3 times PBS wash). Isolated parasite pellet was suspended in SDS sample buffer at 100°C for 10 minutes. After that, samples were centrifuged at 13,200 rpm for 10 minutes, and the clear supernatant was collected. The supernatant was total protein lysates. Total protein lysate was analyzed on 15% SDS-PAGE. Following, SDS-PAGE, proteins were transferred from the acrylamide gel into the nitrocellulose membrane. The membrane was incubated with the primary antibody (overnight incubation, at 4°C) and washed thrice with TBS-T. Followed by, incubation with the horseradish peroxidase conjugated secondary antibody (room temperature incubation for 2 hours). Table 2.1 lists, the various antibodies and their respective dilutions used in our experiments. All the antibodies used previously validated in *P. falciparum* (64)

Table 2.1 List of various antibodies used in immunoblot analysis

Antibody	Dilutions
H4	1:4000
H4K8Ac	1:4000
H4K16Ac	1:4000
H4Ac4	1:4000
H3K9Ac	1:3000
H3K56Ac	1:2000
H3K4me3	1:3000
HDAC1	1:2000
Actin	1:3000
mCherry	1:1000
horseradish peroxidase conjugated anti-mouse/rat/rabbit	1:4000

2.9 Immunofluorescence assay (IFA)

Blood smears were made for both 3D7 and Dd2 transfections on the microscopic glass slides. The smear was air dried for 5 minutes and then fixed with of 4% paraformaldehyde and 0.0075% glutaraldehyde for half an hour. Further, smear was permeabilized by incubating for 10 minutes in 0.1% Triton-X 100. Subsequently, parasite smear was quenched by incubating with 0.001% sodium borohydride (NaBH₄) for fifteen minutes for three times. To avoid non-specific binding, smears were then treated with the blocking solution (5% bovine serum

albumin) for 1 hour. The smear was incubated with primary antibody anti-mCherry (1:50, Abcam) and anti-H4K8Ac (1:1000, Millipore). The incubation for the primary antibody was carried out at 4°C, overnight incubation. Followed by, incubation with secondary antibody conjugated with Alexa Fluor 488 or 594 for 2 hours at the room temperature in the dark. To stain parasite nuclei, smear was incubated with DAPI dye for 3 minutes and followed by, final PBS washing step (3 times, each time for 5 minutes). Lastly, smear was air-dried and mounted with a single drop of VECTASHIELD mounting media (Vector Laboratories, USA.).

2.10 Quantitative real-time PCR (RT-PCR)

Validations of microarray results for selected genes were done by quantitative real-time PCR (Roche) according to the manufacturer instructions. Briefly, 20ng cDNA was mixed with 0.5µM of each of forward and reverse primers of selected genes, with 10µL of 2X SYBR green from Roche. The total reaction volume was of 20 µL and was prepared in the 96 well plates. Each cDNA sample was amplified in triplicates and to check specific amplification, no-template control was also included. Further, only those samples were included in the data analysis, which showed a single peak in the melting curve. Relative fold change in the RNA expression was computed by $2^{-\Delta\Delta C_t}$ method (134). The reference control gene used was PF3D7_1218600 (arginyl-tRNA synthetase).

2.11 Endogenous tagging of PF14_0271

2.11.1 Generation of recombinant plasmid (PF14_0271-mCherry-HA plasmid)

PF14_0271 was tagged with a mCherry-HA tag at 3' end by the single crossover homologous recombination. To achieve this tagging, 917 base pair fragment from 3' end of PF14_0271 was amplified from the genomic DNA obtained from 3D7 and Dd2 strains of *P. falciparum* by PCR. The primer pair used for amplification was: 5' primer AGCTCCCGGGAAGAAGATATGGAAGATAGTAAATGTAGAG and 3' primer

AGCTCTGCAGATTATCAAGGAGATGATTATTATTGTTATG. The sequence underlined in 5' primer, and 3' primer represent for XmaI and PstI restriction sites respectively. This amplified PCR product was cloned into the pTCherryHA vector between XmaI and PstI site.

2.11.2 Transfection and verification of integration in the parasite genome

Transfection of recombinant plasmid into the synchronized rings stage was performed as described by Methods in Malaria Research, 2008 (MR4). Transfection of the plasmid into 3D7 and Dd2 strains of *P. falciparum* was done using electroporation. Briefly, 100µg of recombinant plasmid (100µL volume) was mixed in the 2X cytomix solution (300µL). Subsequently, this plasmid-cytomix solution was put together with 250µL of infected erythrocytes (20-25% early rings). Transfection was carried out in Bio-Rad Gene Pulser and the conditions used for electroporation were 320 V, 95 µF, and 0 Ω. Subsequently, the transfected parasites were grown for 10-12 hours under normal culture condition and after 10-12 hours, parasites were kept under the drug pressure (2.5 nM WR99210). After, 14-21 days of continuous drug pressure, parasites were started to appear in the culture. The genomic DNA of the transfected parasites was isolated and analyzed for verification of the stable integration. The integration of the recombinant plasmid into the genome of 3D7 and Dd2 strain was analyzed using the primer pair: forward primer 5'CACCATGAATGATATGAATG GTTCATC 3' and reverse primer 5'TGGAACATCGTAAGGATACGCATAATC 3'.

2.12 High-throughput genome sequencing by Illumina Mi-seq method

Genome was sequenced for three *P. falciparum* strains (Dd2, W2 and D6) using whole-genome sequencing Illumina Mi-seq method as described by Mok *et al.* (135). Briefly, two microgram of genomic DNA was obtained from each of the parasite's strains and was sequenced. The sequenced data was aligned to the reference *Plasmodium* genome (3D7

genome, PlasmoDB). Only those reads with a mapping score of >30 were included. SNPs were identified in the sequences using Genome Analysis Toolkit software (GATK).

Chapter 3

DNA Damage Response in *P. falciparum*: Transcriptional and Epigenetic Events following DNA Damage in the Malaria Parasite

This work has been published in *Scientific Reports*.

Devendra Kumar Gupta *et al.* "DNA damage regulation and its role in drug-related phenotypes in the malaria parasites" *Sci Rep.* 2016.

3.1 Introduction

Maintenance of genetic integrity is essential for the proper functioning of a cell and is also vital for successful transmission of genetic information to daughter cells. Given, *P. falciparum* complex life cycle, which involves multiple hosts and numerous physiological/developmental stages, the parasite's cellular DNA is under constant threat from various exogenous and endogenous stresses (25). Major sources of genetic assaults to *Plasmodium* include replication errors, reactive oxygen species (ROS) generated during heme metabolism, by-products from the host cellular respiration, free radicals generated by antimalarial drugs such as artemisinin, etc. (25,136) (**Figure 3.1**). To protect its genetic integrity, eukaryotic cells have developed a highly orchestrated and conserved DNA damage response, which can initiate different repair pathway depending on the type of DNA lesion (137). These repair pathways can be classified into two categories: (1) The excision repair pathway, which includes nucleotide excision repair (NER), base excision repair (BER), mismatch repair (MMR) and (2) double strand break repair pathway (DSBR), which includes homologous recombination repair (HR) and nonhomologous end joining (NHEJ). Genomic sequencing has identified most of the homologues of genes belonging to these different repair pathways in *P. falciparum* (7). However, the events involved in the response of a cell to DNA damage, and their regulation have not been studied in the parasite.

It has been shown that the parasite mostly utilizes the homologous recombination repair (HR) system to repair its double-strand breaks, which are the most deleterious forms of genetic aberrations, which in turn can lead to cell death (54). It is believed that the parasite lacks the machinery for canonical NHEJ repair system (7). Only a few studies were done to understand the functioning of repair mechanisms in the parasite such as characterization of PfRAD51, PfRAD54, which are key components of DSBR and few others focused

predominantly in investigating the role of mismatch repair machinery in the emergence of drug resistance in the parasite (52,60).

Figure 3.1 *The source of DNA damage and different mechanism of DNA repair in P. falciparum.* Major sources of genetic aberrations are replication errors, reactive oxygen species (ROS) generated during heme metabolism, by-products from host cellular metabolism, free radicals generated by artemisinin, etc. The major mechanism of DNA repair in the malaria parasite is excision repair pathway, which includes nucleotide excision repair (NER), base excision repair (BER), mismatch repair (MMR) and double strand break repair pathway (DSBR), which includes homologous recombination repair (HR) and nonhomologous end joining (NHEJ).

DNA damage recognition and repair occur in the context of chromatin. In eukaryotes, it has been established that chromatin plays a significant role in regulating the DNA damage response (77). Following DNA damage, histone modifications and chromatin remodeling complex alter the chromatin architecture, which loosens the chromatin at the damaged site and facilitates the accessibility of repair machinery (78). In particular, H2A/H2AX phosphorylation and H4 acetylation play a significant role in the recruitment of various DNA

repair factors such as ATM kinase, ATR kinase, Rad51, BRAC1, TP53, TIP60 HAT, etc. (79). Most of the histone modifications such as acetylation, phosphorylation, methylation, etc. and chromatin remodeling complex have been identified in the malaria parasite (23,138). Until now, the role of chromatin dynamics in DNA damage response in the *P. falciparum* has not been explored.

In this study, we characterized the transcriptional response to DNA damaging agents in the malaria parasite, in order to understand how the parasite repair machinery responds as well as studied how different histone modifications are affected once the parasite is exposed to genotoxic agents. We also explored the clinical relevance of the DNA damage response in the parasite. Artemisinin, the primary antimalarial drug for *P. falciparum* infections, has been long debated for its mechanism of action. Here, we study the artemisinin induced DNA damage response in the parasite and characterized the transcriptional and epigenetic changes induced by the drug.

3.2 RESULTS

3.2.1 Global transcriptional response to DNA damaging agents

In this study, we investigated the genome-wide response of DNA damaging agents such as methyl methanesulfonate (MMS), cisplatin, etoposide and a histone deacetylase inhibitor, trichostatin A (TSA). MMS is a well-known DNA damaging agent which alkylates DNA bases and causes strand breaks including double strand breaks (139). Cisplatin is an effective anticancer drug, and its primary mode of action involves inhibition of DNA replication. Cisplatin promotes covalent bond formation between its platinum atom and nitrogen, N7 of purine bases, which results in the formation of the intrastrand crosslink. This cisplatin-DNA adduct halts DNA synthesis and interferes with the transcription. The most common modes of repair of cisplatin-induced intrastrand crosslink's are nucleotide excision repair (NER) (140). Etoposide is a known anti-tumor agent, which inhibits topoisomerase II and induces the DNA

damage mediated cell death via apoptosis (141). Etoposide treatment in mammalian cells induces DNA damage response that includes H2AX phosphorylation, recruitment of ATM kinase and MDC1 to H2AX site (142).

In *P. falciparum*, it has been previously shown that 100 μ M of cisplatin treatment for 6 hours caused sequence specific cisplatin- DNA adduct formation, which induced parasite death (143). Similarly, in another report, 100 μ M of etoposide treatment induced site-specific chromosomal cleavage in the parasite (144). Similarly, treatment of parasite with MMS (0.05%) resulted in DNA damage, which had been validated using comet assay and also caused up-regulation of PfRAD51, PfRAD54 genes, which belong to the HR pathway, which repairs double-strand breaks (52).

To elucidate the molecular mechanism of DNA damage response and its regulation, rather than characterizing the individual genes of different repair pathways, we used a genome-wide gene expression profiling approach to comprehensively capture the cellular transcriptome of the parasite. Synchronized trophozoite stages (32-36 hour post invasion) of 3D7 parasites were treated with DMSO/DMF (vehicle control), MMS (0.05%), cisplatin (100 μ M), etoposide (100 μ M) and TSA (50nM) for six hours (**Figure 3.2a**). Parasites were treated with the drug concentrations that are known to cause either DNA damage or shown to interact with the DNA in *P. falciparum* (52,143,144). Furthermore, to differentiate a non-specific DNA damage response, parasites were treated with a histone deacetylase HDAC inhibitor, TSA, whose mode of action is not DNA damage (63). Since, DNA replication begins at the late trophozoite stage and peaks at an early-mid schizont; we treated at the late trophozoite stage and harvested the parasites after six hours in early schizont stage in order to analyze the maximal activity of the DNA repair machinery.

Following treatment, cells were harvested for total RNA extraction and were analyzed by the *P. falciparum* DNA microarray (see Materials and Methods Chapter 2). Drug treatment

Figure 3.2 Global transcriptional response of *P. falciparum* to DNA damaging agents. **A)** *P. falciparum* 3D7 parasites were treated with four chromatin perturbations MMS (0.05%), Cisplatin (100 μ M), Etoposide (100 μ M), and TSA (IC90) at trophozoite stage for 6 hours. **B)** Heat map representing up (red) and down (green) regulated genes induced by drug treatment for two replicates. **C)** The number of genes differentially induced by >2 fold for each drug shown graphically. **D)** Venn diagram representing the genes which altered either by MMS or TSA treatment or commonly induced genes.

experiments were done in triplicates, and the microarray results were highly reproducible (Pearson's correlations between biological replicates were, 0.6-0.8, P-value < 0.05). The average numbers of genes analyzed/detected for MMS, cisplatin, etoposide and TSA treatments were 3779, 4105, 3727 and 3862 genes respectively. Hierarchical clustering representing the drugs induced differentially expressed genes in malaria parasite shown in

Figure 3.2b. The color scale represented alterations in the levels of mRNA abundance for each gene. Red represents the transcriptional up-regulation of a gene and green represents the under-expression of a gene as compared to the untreated reference. This clustering revealed that MMS and TSA were able to induce significant changes in the mRNA expression of a large number of genes. In order to identify drug induced differential gene expression, we considered only those genes which had at least >2- fold change in expression levels and those changes in gene expression were consistent among three replicates as analyzed using student's t-test ($p\text{-value} < 0.05$). **Figure 3.2c** depicts the number of genes which were differentially expressed after 6 hours of treatment by these agents. Using a >2-fold change cut-off ($p\text{-value} < 0.05$), which resulted in the following: MMS induced alteration of 1170 genes (30% of the total transcript detected), cisplatin and etoposide altered the expression of only 79 and 146 genes respectively whereas HDAC inhibitor TSA was able to alter the expression of 448 genes respectively (11.6%). Out of 1170 genes, MMS induced the expression of 523 genes whereas 647 genes were down-regulated. In order to differentiate a non-specific DNA damage response, parasites were also exposed to HDAC inhibitor, TSA for 6 hours simultaneously. There was a small but statistically significant overlap between genes induced by MMS and TSA with 42 and 99 being commonly up and down regulated, respectively (**Figure 3.2d**). It suggests that the effect of these 2 drugs on the parasite transcriptome is different from each other.

In order to elucidate the cellular machinery targeted by these agents, functional pathway analysis was carried out on the differentially expressed genes. The functional pathway analysis comprises of 1261 gene sets of Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) and Malaria Parasite Metabolic Pathways (MPMP) as described previously (9). Only pathways with binomial $p\text{-value} < 0.05$ were considered to be statistically significant. The results of pathway enrichment analysis revealed that many

cellular functions were affected by MMS and TSA. Among the 143 commonly co-regulated genes, there is a significant representation of factors of host-parasite interactions such as “merozoite invasion” and “PfEMP1 domain architecture”. In the case of MMS-induced gene expression, the biological pathways which were significantly over-represented comprise of protein biosynthesis, palmitome proteins, ribosomal assembly genes, chaperones, nuclear genes destined for mitochondria and apicoplast and glycolysis (**Figure 3.3**). Most of these functionalities could also represent indirect responses to the stress conditions. Nonetheless, one of the most predominant functional groups induced by MMS was a set of genes associated with the DNA repair machinery that reflects its presumed mode of action. This supports previous studies showing that treatment of *P. falciparum* parasites with MMS (0.05%) for six hours results in DNA double strands breaks and up-regulation of two factors of the HR pathway PfRAD51, PfRAD54 (52). Here, we show that MMS induces a broad DNA damage response including most of the factors of both the excision repair machinery and double-strand break repair machinery (DSBR) (**Figure 3.3**). In comparison, there was a distinct set of functional groups, which were significantly repressed by MMS treatment and mainly consist of genes involved in merozoites invasion.

The other two DNA damaging agent's cisplatin and etoposide were not able to induce the expression of the DNA repair machinery and also had a moderate or low effect on mRNA abundance of a large number of genes (**Figure 3.3**). One possible explanation could be that these drugs have general cytotoxic effects that may lead to lose of cell viability without causing any dramatic transcriptional changes in the malaria parasite. Such observations have been previously documented, when the trophozoite parasites exposed to lethal doses of antifolate, WR99210, for 6 hours and 12 hours. It lead to a loss of 50% of cell viability without exhibiting any transcriptional changes in the parasite (145).

Figure 3.3 The functional enrichment analysis of genes which were up-regulated by MMS. Up-regulated genes tested for functional over-representation by hypergeometric testing using MPMP, KEGG and GO pathways and highlights significantly enriched pathways (p-value < 0.05). Further, the expression pattern of 29 excision and 6 DSB repair genes which were up-regulated by MMS are shown along with the effect of cisplatin, etoposide, and TSA. The color scale used to represent the variations in the transcript abundance shown at the bottom of the figure.

3.2.2 Effect of MMS on the transcription of malaria DNA repair genes

In order to understand a time-dependent response of MMS induced DNA damage on the parasite transcriptome with a focus on the DNA repair genes, we exposed the trophozoite stage of the parasites with 1 hour, 3 hours and 6 hours of 0.05% MMS. Crucially, at least 6 hours of MMS exposure was required for induction of enough DNA repair genes, which can be, significantly overrepresented ($p\text{-value} < 0.05$) by the excision repair machinery and double-strand break repair machinery (DSBR) in the parasite. As with the shorter duration of MMS exposure, it did not lead to a significant transcriptional response. **Table 3.1** summarizes the time-dependent effect of MMS on genes belonging to excision repair pathways. Excision repair pathways consist of nucleotide excision repair (NER), base excision repair (BER) and mismatch repair (MMR). In particular, 9 genes of BER, 17 genes of NER, 4 genes of MMR and 7 genes of DSBR were found to be significantly up-regulated in three independent replicates after 6 hours of MMS exposure ($p\text{-value} < 0.05$).

3.2.2.1 Base excision repair (BER)

DNA glycosylases recognize altered/damaged DNA bases and initiate BER by generating an abasic site, which further processed by AP endonucleases to produce a base gap. The *Plasmodium* genome has four homologues of DNA glycosylases (7). Two DNA glycosylases were found to be up-regulated by MMS after 6 hours of exposure: A/G specific adenine glycosylases (PF3D7_1129500), and N-glycosylase/DNA lyase (PF3D7_0917100). Further, the expression of one AP endonucleases (exodeoxyribonuclease III putative, PF3D7_1430600) and three helicase enzymes (DEAD box helicase, RuvB-like helicase 2 and RuvB-like helicase 3) were also induced by MMS in the parasite.

Table 3.1 List of DNA excision repair genes which were significantly differentially expressed after 1 hr, 3 hr and 6 hr of MMS treatment (p-value < 0.05)*

Gene Id	Description (Base Excision Repair)	MMS_1 hr	MMS_3 hr	MMS_6 hr
PF3D7_1106000	RuvB-like helicase 2			1.49
PF3D7_1129500	A/G-specific adenine glycosylase, putative			1.1
PF3D7_1362200	RuvB-like helicase 3			0.95
PF3D7_1429900	ATP-dependent DNA helicase, putative			-0.49
PF3D7_1430600	Exodeoxyribonuclease III, putative			0.94
PF3D7_0310500	DEAD box helicase, putative			0.83
PF3D7_0504400	ATP-dependent helicase, putative			0.99
PF3D7_0604600	DNA helicase, putative			1.72
PF3D7_0910500	DNA repair protein REV1, putative			-0.65
PF3D7_0918600	DNA helicase, putative			0.72
PF3D7_0820000	Snf2-related CBP activator, putative		-0.83	-0.77
PF3D7_1363200	Bifunctional polynucleotide phosphatase		-1.15	-0.34
PF3D7_0917100	N-glycosylase/DNA lyase, putative		0.95	1.26

*The number in the MMS column represents the log ratios of gene expression under MMS exposure after 1, 3 and 6 hours. Genes were considered differentially expressed if MMS induced changes in gene expression were consistent among three replicates as analyzed using student t-test (p-value < 0.05).

Gene Id	Description (Nucleotide Excision Repair)	MMS_1 hr	MMS_3 hr	MMS_6 hr
PF3D7_1432600	Conserved <i>Plasmodium</i> protein, unknown function		1.18	1.05
PF3D7_1011700	DNA repair protein RAD23, putative			0.97
PF3D7_1408400	DNA-repair helicase, putative			0.7
PF3D7_1037600	DNA repair helicase RAD25, putative	1.21	1.27	1.6
PF3D7_1201500	XPA binding protein 1, putative		0.47	1.6
PF3D7_0409600	Replication protein A1, large subunit			0.66
PF3D7_1304100	DNA ligase I			0.81
PF3D7_1368800	DNA repair endonuclease, putative			0.42
PF3D7_1314900	TFIIH basal transcription factor subunit, putative			-0.88
PF3D7_1010200	D2/M7 helicase, putative			0.89
PF3D7_1441900	Transcription factor TFIIH complex subunit Tfb5, putative			-1.57
PF3D7_1463200	Replication factor C3			0.79
PF3D7_0107800	Double-strand break repair protein MRE11, putative			0.86
PF3D7_0111300	Replication factor c protein, putative			-0.3
PF3D7_0605800	DNA repair protein RAD50, putative			0.37
PF3D7_0908000	P1 nuclease, putative			-1.6
PF3D7_0934100	DNA excision-repair helicase, putative			0.94
PF3D7_1235900	Pre-mRNA-splicing factor SYF1, putative			1.03
PF3D7_1250800	DNA repair protein rhp16, putative			1.15
PF3D7_0219600	Replication factor C subunit 1, putative		1.0825	1.58
PF3D7_1244200	Transcription factor Tfb2, putative		1.1065	1.8

Gene Id	Description (Mismatch Repair)	MMS_1 hr	MMS_3 hr	MMS_6 hr
PF3D7_0726300	DNA mismatch repair protein PMS1, putative			0.37
PF3D7_0725000	Exonuclease I, putative		-0.52	-1.61
PF3D7_1427500	DNA mismatch repair protein MSH2, putative	0.42		0.92
PF3D7_0204600	Conserved <i>Plasmodium</i> protein, unknown function		1.66	1.97
PF3D7_0204800	3'-5' exonuclease, putative		1.51	1.22

3.2.2.2 Mismatch repair (MMR)

The mismatch repair machinery (MMR) is involved in removing mismatches arisen from the incorporation of wrong bases during replication, and the mismatches caused by alkylating agents. Four key genes of MMR were found to be up-regulated in the parasite due to MMS exposure. The parasite genome has 3 mutS homologues, which can recognize the mismatches. These are PfMSH2-1, PfMSH2-2 and PfMSH6 (7). MMS induced up-regulation of only PfMSH2-1. Similarly, MutL homologue, known as PfPMS1 (PF3D7_0726300) was also up-regulated. In addition, there were two exonucleases, which were found to be up-regulated as early as three hours of MMS exposure. One of them possesses 5' to 3' exonuclease activity (PF3D7_0204600), whereas the other one has 3' to 5' exonuclease activity (PF3D7_0204800).

3.2.2.3 Nucleotide excision repair (NER)

In **Table 3.2**, I outlined the core components of NER and a bioinformatics comparison of the NER orthologues among yeast, mammals and *P. falciparum*. The orthologs of *P. falciparum* were obtained from earlier reports (31) and the sequences of *P. falciparum* genes belonging to NER repair pathway were taken from PlasmoDB database. NER eliminates the bulky lesions that can distort the helical structure of DNA and hence, plays a crucial role in

Table 3.2 Comparison of the core components of nucleotide excision repair (NER) among yeast, humans and *P. falciparum*.

		<i>S. cerevisiae</i>	<i>H. sapiens</i>	<i>P. falciparum</i>	Homologues in <i>P. falciparum</i>	MMS 6 hr*
Function in NER	Protein					
DNA damage sensors	XPC	Rad4	XPC	PF3D7_1432600	1	Up-regulated
	HR23B	Rad23	hHR23B	PF3D7_1011700	1	Up-regulated
TFIIH complex/ helicases	XPB	Rad25	XPB	PF3D7_1037600	1	Up-regulated
	XPB	Rad3	XPB	PF3D7_1408400	1	Up-regulated
DNA damage verification	XPA	RAD14	XPA	PF3D7_1201500 PF3D7_1235900 PF3D7_0917700 PF3D7_1349800	4	PF3D7_1201500 PF3D7_1235900
	RPA1	Rfa1	RPA1	PF3D7_0409600 PF3D7_0904800	2	PF3D7_0409600
	RPA2	Rfa2	RPA2	absent		
	RPA3	Rfa3	RPA3	absent		
Structure specific endonucleases	XPG	Rad2	XPG	PF3D7_0206000	1	
	ERCC1	Rad10	ERCC1	PF3D7_0203300	1	
	XPF	Rad1	XPF	PF3D7_1368800	1	Up-regulated

*Genes were considered to be up-regulated; if MMS induced changes in gene expression were consistent among three replicates as analyzed using student t-test (p-value < 0.05).

preventing mutagenesis. XPC-HR23B acts as a DNA damage sensor and as a scaffold for other repair factors involved in the NER (31). Subsequently to DNA damage recognition, transcription factor IIH complex is recruited to the damaged site. XPB and XPD are the two

components of this general TFIIH complex, which have helicase activity and are responsible for the strand separation. XPA is a DNA binding protein, which is recruited for DNA damage verification. Further, replication binding protein A (RPA), which is a single-strand DNA binding protein, promotes the recruitment of structure-specific endonucleases like XPG and ERCC1-XPF to the damaged site (31,32). These endonucleases remove the bulky lesions and create a gap of 27-29 nucleotides. Subsequently, the general replication apparatus resynthesize the excised region. MMS treatment in the parasite causes up-regulation of DNA damage sensors: XPC-HR23B complex, TFIIH helicase XPB and XPD, DNA damage verification proteins (XPA and RPA) and one of the endonucleases XPF.

3.2.2.4 Double strand break repair (DSBR)

MMS induced DNA damage caused significant up-regulation of double-strand break repair (DSBR) in the parasite. 7 genes of the DSBR were found to be up-regulated by six hours of MMS treatment. DSBR pathways are error-free recombination pathways, which can repair double-strand breaks (DSB). DSBs are considered as the most lethal genetic aberrations which can lead to cell death if remain unrepaired. Genome mining led to the identification of most of the components of DSBR/HR in *Plasmodium* (7). **Table 3.3** summarizes the genes belonging to DSBR pathways, which were differentially induced in the parasite by DNA damaging agent MMS. The initial step of DSBR/HR is the recognition of the damaged DNA by the MRN exonuclease complex (146). In general, MRN complex consists of MRE11, Rad50 and Nbs1. The parasite genome encodes a single copy of each MRE11 and Rad50 but lacks the homologue of Nbs1. Our results showed that MMS induced the overexpression of both MRE11 and Rad50. Also, it significantly up-regulates RPA, which is a single-strand DNA binding protein that promotes stability of the processed DNA and facilitates the recruitment of Rad51 and Rad54. MMS also caused the up-regulation of Rad51, which is a

key nucleoprotein that plays a vital role in strand invasion and strand annealing steps of the DSB repair pathway.

Table 3.3 List of double-strand break repair (DSBR) genes, which were significantly differentially expressed by MMS (*p*-value < 0.05)

Gene Id	Description (Double Strand Break Repair)	MMS_1 hr	MMS_3 hr	MMS_6 hr*
PF3D7_1368800	DNA repair endonuclease, putative			0.42
PF3D7_1328200	Conserved <i>Plasmodium</i> protein, unknown function			-1.02
PF3D7_1107400	DNA repair protein RAD51			0.55
PF3D7_1010200	DNA2/NAM7 helicase, putative			0.89
PF3D7_0918600	DNA helicase, putative			0.72
PF3D7_0605800	DNA repair protein RAD50, putative			0.37
PF3D7_0409600	Replication protein A1, large subunit			0.66
PF3D7_0107800	Double-strand break repair protein MRE11, putative			0.86

*The number in the MMS column represents log ratios of gene expression under MMS exposure after 1, 3 and 6 hours. Genes were considered to be differentially expressed if MMS induced changes in gene expression were consistent among three replicates as analyzed using student t-test (*p*-value < 0.05).

3.2.3 DNA damage sensors

The DNA repair machinery has dedicated set of enzymes that can sense and initiate a DNA lesion specific repair pathway. These damage recognition proteins are known as DNA damage sensors (**Table 3.4**) (146). A/G specific adenine glycosylases, putative (PF3D7_1129500) is an example of DNA damage sensors for BER in the parasite. It has a characteristic helix-hairpin-helix DNA binding motif that is a signature motif for DNA repair proteins (33).

Table 3.4 Comparison of DNA damage sensors among mammals, yeast and *Plasmodium*.

	Repair Mechanisms	Mammals	Yeast	<i>Plasmodium falciparum</i>	Number of homologues in <i>P. falciparum</i>	MMS induced upregulation in <i>P. falciparum</i> **
Excision repairs	Base excision repair (BER)	DNA glycosylase	DNA glycosylase	PF3D7_1129500 PF3D7_1415000 PF3D7_1467100 PF3D7_0917100	Four	PF3D7_1129500 PF3D7_0917100
	Nucleotide excision repair (NER)	XPC-hHR23B complex	Rad4-Rad-23 complex	PF3D7_1432600 (XPC) PF3D7_1011700 (Rad23)	XPC - 1 Rad23 - 1	PF3D7_1432600 (XPC) PF3D7_1011700 (Rad23)
	Mismatch repair (MMR)	hMutS α (MSH2-MSH6) hMutS β (MSH2-MSH3)	MSH1-MSH3, MSH6	PfMSH2-1 (PF3D7_1427500), PfMSH2-2 (PF3D7_0706700), PfMSH6 (PF3D7_0505500)	MSH2 - 2 MSH6 - 1 MSH3- Absent	PfMSH2 (PF3D7_1427500)
Double strand break repairs	Homologous recombination (HR)	MRE11-RAD50-NBS1 (MRN complex)	Mre11-Rad50-Xrs2 (MRX complex)	PF3D7_0107800 (MRE11) PF3D7_0605800 (Rad50)	MRE11 - 1 PfRAD50 - 1 NBS1 - Absent	PF3D7_0107800 (MRE11) PF3D7_0605800 (Rad50)
	Non homologous recombination end joining (NHEJ)	Ku 70/80 complex	Ku 70/80 complex	Alternative NHEJ	Absent	
	Single strand break (SSBR)	Replication binding protein -1(RPA-1)	Rfa1	PF3D7_0409600 PF3D7_0904800	Two	PF3D7_0409600
Direct repair	AlkBH	AlkB homologue family	AlkB homologue family	ND	ND	
	DNA photolyase (UV induced photo reactivation repair)	Absent in placental mammals	DNA photolyase (Phr1)	PF3D7_0513600 (DNA photolyase)	One	PF3D7_0513600
	O-6 methylguanine DNA methyltransferase (MGMT)	MGMT	MGMT	ND	ND	

ND - Not defined, ** genes significantly up-regulated (p-value < 0.05)

This motif is known to mediate non-sequence specific DNA interaction. Similarly, MRN exonuclease complex acts as a DNA damage sensor for DSB repair and consists of three components: MRE11, Rad 50 and Nbs1. Only MRE11 has the DNA binding domain which can recognize the damaged DNA. Furthermore, XPC-h23B is a damage recognition protein for the NER, whereas mutS heterodimer (MSH2-MSH6) act as a sensor for MMR. **Table 3.4** summarizes the effect of MMS on each class of DNA damage sensor and a comparison of these proteins among yeast, humans and *Plasmodium*.

3.2.4 Role of chromatin dynamics in DNA damage response in malaria

Chromatin alteration by histone modifications during DNA damage regulates the accessibility of DNA repair factors and other regulatory proteins (79). The role of H4 acetylation, H3 acetylation and H2A/H2AX phosphorylation in DNA damage response is well documented in yeast, mammals and more recently in *T. gondii* but in *Plasmodium*, it has not been investigated (82,86,90). Here, we wished to study the effect of DNA damage on histone modifications in *P. falciparum*. For this, the trophozoite stage (28-32 hpi) parasites were treated with the same panel of drugs used for the transcriptional profiling with the same concentrations and duration of exposure (above). Simultaneously, trophozoite parasites were treated with inhibitory concentrations (IC)-90 doses of artemisinin (50nM) and chloroquine (70nM). Subsequently, we utilized an immunoblotting analysis to study the abundance of several histone marks that were previously shown to play a role in the *P. falciparum* chromatin remodeling (64). These include histone 3 Lysine 9 acetylation (H3K9Ac), H3K56Ac, H3K4me3 (lysine 4 trimethylation) and H4K8Ac, H4K16Ac and H4KAc4 (K4, 8, 12, 16 acetylations). Specific antibodies for HDAC enzyme, histone 4 and actin were used as loading controls (**Figure 3.4a**).

Our results revealed that unlike cisplatin and etoposide, MMS caused increases in acetylation at H4K8 and H4K16 and (likely) as a results of that H4KAc4. On the other hand,

H3K9Ac and H3K56Ac were reduced after the 6hr MMS treatment. There were no detectable changes in the level of the studied histone marks following treatment with IC90 doses of artemisinin and chloroquine. It has been documented that HDAC1 inhibition by apicidin and TSA leads to induction of H3 and H4 acetylation (147). Similarly, we also observed the global hyperacetylation of H3 and H4 prior to TSA treatment. Additionally, there were no changes in the protein levels of HDAC1, Histone 4 and actin across all the treatments.

Further, to understand the dynamics of DNA damage recovery, we treated the parasite with MMS for 6 hours, followed by washing in order to remove the drug and then transferred the parasite into fresh culture medium lacking MMS, so the parasite could repair its damaged DNA. We collected various time points from 30 minutes to 12 hours after washing off MMS (**Figure 3.4b**). Our results revealed that the increased levels of H4K8Ac and the decreased levels of H3K9Ac remained detectable even 12 hours after MMS was removed from the culture; past the 6 hour treatment (**Figure 3.4b**). This suggests that the DNA damage repair in *P. falciparum* is a relatively slow process in which the studied histone modifications play a role as sensors/effectors throughout the DNA repair process (79). The MMS-induced effect on chromatin is distinct from TSA inhibition of HDAC that results in increases of H4K8Ac, H4K16Ac, H3K9Ac and H3K56Ac that was previously shown (147,148). The TSA-induced H4 and H3 hyperacetylations are reversible as early as 2 hours after drug removal, which also contrasts the effect of MMS on chromatin remodeling (147).

Taken together, our results highlighted two major outcomes 1) H3K9Ac and H4K8Ac regulate chromatin architecture during DNA damage response in the parasite. 2) DNA damage induces hyperacetylation of lysine residues on H4. This may reflect that the functional role of H4 in DNA damage is evolutionary conserved in *Plasmodium* similar to the other species (79).

Figure 3.4 DNA damage affects the overall abundance of histone modifications during the *P. falciparum* IDC. **A)** 3D7 parasites were treated with all drugs at the trophozoite stage for 6 hours and the overall abundance of individual histone modification was measured by western blotting. The MMS treatment resulted in reduction of H3K9Ac and H3K56Ac and increase of H4K8Ac, H4K16Ac and H4KAc4. Conversely, TSA induced hyperacetylation of all studied H3 and H4 acetylations. Cisplatin and etoposide as well as artemisinin and chloroquine at IC-90 concentrations had no effect on the overall abundance on the histone modification compared to *P. falciparum* grown under normal condition or treated with the carrier solvent, DMSO. Specific antibodies for HDAC enzyme, histone 4 and actin were used as loading controls. **B)** Recovery dynamics of H3K9Ac and H4K8Ac were assessed by first treating with MMS for 6 hrs and subsequently drug was washed off and samples were collected after 0.5 hr, 1 hr, 2 hr, 6 hr and 12 hr.

3.2.5 Genome wide distribution of H3K9Ac and H4K8Ac during DNA damage in malaria

Our immunoblot results suggested that H3K9Ac and H4K8Ac were the most sensitive histone marks altered during DNA damage in the parasite. Hence, in order to further identify the genome-wide distribution of these altered marks, we carried out chromatin immunoprecipitation coupled with DNA microarray (ChIP-chip). The main objective was to

identify chromosomal regions associated with H3K9Ac and H4K8Ac that are responsive to DNA damage. From the previous result section, our transcriptional analysis revealed that 6 hours was the shortest exposure time that leads to induction of MMS mediated DNA damage response in parasite. Hence, trophozoite stage parasites were treated with 0.05% MMS for 6 hours and subsequently, ChIP-chip was performed as described in section “Materials and methods.” Our *Plasmodium* DNA array consists of 15,518 microarray probes/genetic loci (4,447 intergenic probes, IGR and 11,071 open reading frame probes, ORF) which represent most of the intergenic and coding sequences of the parasite genome. In addition to H3K9Ac and H4K8Ac, we also explored the occurrence of H3K56Ac and for better statistical interpretation; we repeated our experimental analysis three times for all the histone marks. The total number of probes/genetic loci which was differentially acetylated (induced/reduced) for H4K8Ac, H3K9Ac and H3K56Ac were 2689, 528 and 2527 respectively (p-value <0.05). From these, 60% (1,592), 46% (243) and 52% (1,318) of genetic loci displayed a significant increase in H4K8ac, H3K9Ac and H3K56Ac occupancy in the presence of MMS (**Figure 3.5**). These results complemented our immunoblot findings, which showed that MMS treatment leads to global induction of H4K8 acetylation and reduction in the H3K9 acetylation.

In general, the intergenic regions displayed higher levels of histone acetylation regardless of the global acetylation levels in presence of MMS. At a finer level, H4K8Ac, H3K9Ac, H3K56Ac marks were enriched around intergenic regions and transcriptional start site (TSS). The loss of acetylation was observed within the coding regions with progressively decreasing occupancy towards the 3' termini of the genes for all three histone marks (**Figure 3.6a**). In contrast, the peak distribution for H4K8Ac was in the intergenic regions and H3K9Ac, H3K56Ac was at the start of the gene (**Figure 3.6b**). It has been previously shown that H4K8Ac is predominantly enriched at IGRs, whereas H3K9Ac and H3K56Ac are localized

Figure 3.5 The Global nucleosome occupancy of H3K9Ac, H3K56Ac and H4K8Ac during DNA damage in *P.falciparum*. The percentage of the number of probes which were differentially acetylated (p value < 0.05) for each histone mark was shown. For each histone mark, the number of probes enriched for the intergenic regions (IGRs) and open reading frames (ORFs) were also shown.

mainly within the 5' regions of the coding regions (133). Our result suggests that the MMS mediated DNA damage induced acetylations at regions which overlap with genetic loci that are associated with these histone marks at putative promoter regions during the *P. falciparum* life cycle under normal growth conditions. Gene set enrichment analysis (GSEA) on genes that associated with MMS induced H4K8 hyperacetylation, identified various cellular processes such as ribosomal structure and assembly, nucleosome assembly, protein folding, histone acetylation and possible pathway for apoptosis (p-value < 0.05, FDR < 0.25) (Supplementary table 3.1). Although, MMS induced global H3K9 deacetylation, but our ChIP-chip results identified certain genetic loci, which remained acetylated suggesting a physiological importance of these regions during DNA damage. GSEA analysis on these H3K9 acetylated regions revealed four functional groups: ribosomal structure, nucleosome assembly, kinetochore organization and mismatch repair machinery.

Figure 3.6 Distribution of histone marks along the parasite genes. The horizontal axis is the oligonucleotide probe position relative to the ATG start codon. Position -1 bp to -1000 bp represent the IGRs, and position 0 bp to >5000 bp represent the ORFs. The vertical axis is the number of histone marks associated genetic loci normalized to the number of probes which represent that particular region on the genome. **A)** The graph represents the genetic loci which have reduced ChIP signal in presence of MMS (P value < 0.05). Loss of acetylation was observed across gene body, and the peak of deacetylation was at 3' end of genes for all three histone marks. **B)** The graph represents the genetic loci which have increased ChIP signal in presence of the MMS (P value < 0.05). H4K8Ac, H3K9Ac, H3K56Ac marks were enriched around the intergenic regions and transcriptional start site (TSS). The peak distribution for H4K8Ac was in the intergenic regions and for H3K9Ac and H3K56Ac was at the start of the gene.

3.2.6 DNA damage induces genome wide spread of H4K8Ac whereas H3K9Ac preferentially marks the DNA repair genes

H3K9Ac and H4K8Ac are considered as transcription-linked histone marks in many biological systems, including yeast, mammals and recently, it was also shown for *P. falciparum* (64). Moreover, H4K8Ac is regarded as the most important histone mark in the malaria parasite, as, it is the only histone mark that was found to be enriched in the promoter regions of transcriptionally active genes (64). To study the effect of the MMS induced histone modifications on transcription; we correlated the ChIP-chip and RNA expression results. Interestingly, there was a sharp bimodal distribution with the positive and negative correlations between the MMS induced gene expression and histone acetylations (P-value <0.05) (**Figure 3.7**). Further, the overlap between up-regulated/down-regulated genes (>2-fold, p-value < 0.05) due to MMS treatment and with those that had altered histone marks (p-value < 0.05) is shown in **figure 3.8a** and **3.8b**. In particular, there were 533 genes whose altered transcription (>2-fold, p-value < 0.05) correlated significantly with alterations in occupancy of H4K8Ac (p-value 1.83×10^{-22} , hypergeometric test). From these, there were 177 genes in which the increase in H4K8Ac coincided with transcriptional up-regulation. Pathway enrichment analyses identified that these genes belong to several cellular processes of protein metabolisms, including ribosome assembly, chaperones/protein folding, as well as other biosynthetic pathways, including mitochondria (“*metabolic nuclear genes destined to mitochondria*”) and folate biosynthesis (**Figure 3.8a**). In yeast, MMS induces up-regulation of chaperones and also facilitates trafficking of proteins into mitochondria, which constitutes the indirect responses to DNA damage (149,150). Our results suggest that this process is evolutionarily conserved and that some of the indirect transcriptional responses induced by MMS are linked with H4K8Ac. For H3K9Ac, we identified 122 genes (132 genetic loci, p-value 1.75×10^{-5} , hypergeometric test) whose abundance correlated with the expression.

Figure 3.7 Correlation analysis between MMS induced gene expression and histone marks. To study the association between the expression and altered histone marks, we correlated ChIP-chip dataset and RNA expression dataset obtained from MMS treated trophozoite parasites. It resulted in a sharp bimodal distribution with the positive and negative correlations between MMS induced gene expression and histone acetylations (P-value <0.05) which is shown here.

Out of the 122 overlapping genes, 37 showed positive correlation (p-value < 0.05). Strikingly, this gene set is enriched for factors of the DNA repair and DNA replication associated pathways (**Figure 3.8b**). In mammals, it is the H3K56Ac that correlates with the expression of the DNA damage repair genes under stress conditions (96). However, our results showed that in *P. falciparum*, the occupancy of H3K56Ac was not linked with DNA repair machinery but instead with genes encoding S-Glutathionylated and mitochondrial proteins (**Figure 3.9**). Hence, our results suggest that *Plasmodium* may have undergone an evolutionary diversion where regulation of DNA repair was “taken over” by H3K9Ac. Intriguingly, this occurs despite the fact that MMS causes an overall reduction of H3K9Ac levels in *P. falciparum*. In future studies, it will be interesting to explore the molecular mechanisms that facilitate this chromatin-linked induction of the DNA damage repair genes.

Figure 3.8 Association between differential expression and altered histone marks occurrence during MMS treatment. For H4K8ac (**A**) and H3K9Ac (**B**), we identified 533 and 122 overlapping genes. Hierarchical clustering and functional groups of these 533 and 122 genes are also shown, which are divided into 4 clusters depending upon on correlation between ChIP-chip and transcriptome dataset.

3.2.7 Antimalarial, artemisinin induces DNA damage response in *P. falciparum*

Artemisinin is the primary antimalarial drug which has been recommended for treatment against multiple drug resistant strains of *P. falciparum*. Although, artemisinin has been shown to induce a DNA damage response in cancer cells, its effect on the *Plasmodium* genomic DNA remains debatable (26). In order to provide the evidence for an artemisinin role in DNA damage, we wished to investigate similarities between artemisinin and MMS in terms of their effect on the DNA repair machinery in *P. falciparum*.

Enriched functional groups (p value < 0.05)

S-Glutathionylated proteins
Nuclear genes with mitochondrial signal sequences
Modification of proteins by adenylation and ADP-ribosylation
Maturation and export of 60S and 40S ribosomal subunits
Genes coding for components involved in ribosome assembly
CoA biosynthesis
Chaperone-assisted protein folding
Asparagine and Aspartate metabolism

Figure 3.9 Correlation analysis between MMS induced gene expression and H3K56Ac. The Venn diagram depicts the overlap between the 1327 genes were differentially expressed and 2689 genetic loci with differential occupancy of H3K56ac as a result of MMS treatment (P-value <0.05). 189 genes (out of 530 overlapping) showed positive correlation between MMS induced RNA expression and H3K56Ac is shown. The list indicates functional pathways enriched in the set of 189 genes (P value <0.05).

3.2.8 Transcriptional response of clinical doses of artemisinin

In our previous section, we have shown that MMS induced DNA damage resulted in transcriptional up-regulation of the DNA repair machinery. Moreover, we also showed that it caused profound alterations in histone modifications in the parasite. Artemisinin may induce DNA damage in the malaria parasite giving positive results of an *in vitro* comet assay which analyze DNA damage (115). Hence, in order to characterize the transcriptional and epigenetic response of artemisinin, we treated the 3D7 trophozoite parasite with 1 μ M (micro molar) of artemisinin for 6 hours. Importantly, we used a physiological dose of 1 μ M for our treatment, which corresponds to the plasma concentration of the drug when administered to patients (151). Simultaneously, we treated the 3D7 trophozoite parasites with MMS (0.05%) for 6 hours.

Hierarchical clustering of RNA expression data identified 334 and 412 genes that were commonly up- and down-regulated, respectively, by both MMS and artemisinin (p-value < 0.05, FDR < 0.05) (**Figure 3.10**). Functional enrichment analysis revealed that in addition to several biological processes associated with protein metabolism and DNA replication, at the given treatment condition, artemisinin also induced genes of the DNA repair pathways. These include 8 genes of excision repair and 4 genes of DSB repair machinery. Interestingly, Rad50 and MRE11 that constitute the MRN exonuclease complex, which are regarded as DNA damage sensors of DSB repair were also up-regulated. Moreover, Rad51, which is another important repair protein of DSB repair pathway, was also up-regulated by artemisinin. Rad51 is a key nucleoprotein, which is required for the strand invasion and strand annealing steps of DSB repair (52). It has been also reported that artemisinin inhibits topoisomerase enzymes in cancer cells, which leads to generation of single and double-strand breaks and subsequently to cell death (26). Interestingly, artemisinin treatment in parasite also leads to under-expression of two DNA topoisomerase: PF3D7_1433500 and PF3D7_1365600.

Figure 3.10 Artemisinin and MMS induce a similar transcriptional response. Trophozoite stages of 3D7 parasite were treated with clinical doses of artemisinin (1uM) and MMS (0.05%) for 6 hours. **A)** Genes with similar differential transcription for both MMS and artemisinin were shown in the heat map. Experiment was done in triplicates (p value < 0.05, FDR < 0.05). **B)** Pathways which are commonly up-regulated in both are shown.

Previously, Natalang *et al.* have shown that 3 hours exposure to 780 nM artemisinin treatment at trophozoite stage altered the transcript abundance of 398 genes, none of which were involved in DNA repair. (152). In our analysis, we identified 1,506 genes that were differentially expressed due to artemisinin treatment among three replicates (P-value < 0.05). 190 genes detected were also altered in the Natalang *et al.* and our dataset (Pearson correlation of 0.79). Interestingly, some of the commonly up-regulated genes were 4 putative zinc-fingers proteins (PF3D7_1425600, PF3D7_0525000, PF3D7_1208800, and PF3D7_1019300) and 5 AP2 domains containing transcription factors (PF3D7_0622900, PF3D7_1317200, PF3D7_0516800, PF3D7_1107800, and PF3D7_1350900). It has also been reported that one of the postulated cellular targets of artemisinin is the SERCA type calcium pump PfATP6ase, and artemisinin inhibits this calcium pump (111). But in our expression analysis as well as in Natalang *et al.* results, the RNA expression of PfATP6ase remains unaffected by the artemisinin treatment.

The genetic basis of artemisinin resistance in south-east Asia is associated with mutations in the kelch-13 protein (PF3D7_1343700) (153). Intriguingly, our results showed that artemisinin treatment in 3D7 parasites led to significant up-regulation of kelch-13 protein (P-value 0.0004). In summary, the transcriptional response of artemisinin treatment with a clinical dose of 1µM is similar to that of MMS, which is a known DNA damaging agent.

3.2.9 Effect of physiological dose of artemisinin on the parasite histone modifications

The role of histone H4 acetylation in the DNA damage response is well documented (79). We also demonstrated that MMS induced DNA damage in the parasite resulted in the alterations of H4 and H3 acetylation. From our transcriptional analysis, we identified that both artemisinin and MMS can induce the DNA excision and DSB repair pathways. Hence, we reasoned that artemisinin mediated DNA damage may also alter the chromatin architecture similarly to the MMS. Therefore, we wished to study the effect of clinical doses of

artemisinin and other antimalarial drugs on the parasite chromatin. The four antimalarial drugs studied were artemisinin (1 μ M), chloroquine (30 μ M), mefloquine (5 μ M) and pyrimethamine (5 μ M). Strikingly, our results demonstrated that both artemisinin and MMS induced hyperacetylation of H4K8 in trophozoite and schizont stages (**Figure 3.11**). This strongly suggests that H4 acetylation plays a significant role in the DNA damage response in the parasite. Contrary to this, the effect of H3K9ac was more complex. In trophozoite stage, both artemisinin and MMS caused deacetylation of H3K9, whereas both drugs showed no effect on H3K9Ac at schizont stage. The other three antimalarial drugs chloroquine, mefloquine and pyrimethamine showed no dramatic effect on the H3 and H4 acetylation in both the tested stages.

Overall, artemisinin and MMS induce similar transcriptional responses that include over-expression of DNA repair machinery genes. In addition, both agents alter the chromatin composition in a similar manner at the multiple stages, which involves H4K8 hyperacetylation. Taken together, our results suggest that artemisinin elicits DNA damage response in the parasite, which is similar to the MMS mediated response.

3.3 Discussion

Malaria is one of the most lethal human diseases, which is caused by an intracellular obligate parasite, *Plasmodium*. Out of five *Plasmodium* species that cause malaria, *P. falciparum* is responsible for the highest number of clinical cases and the highest number of deaths. According to the World Health Organization, WHO, half of the world is at risk to malaria. In 2012, malaria was responsible for 627,000 number of deaths and most of the casualties were children under the age of 5 years (4). Most alarmingly, the emergence of artemisinin resistance in Southeast Asia has hindered the progress of controlling malaria worldwide. Hence, there is an urgent requirement to identify new drug candidates and novel parasitic drug targets.

Figure 3.11 Effect of clinical dose of artemisinin and other anti-malarial drugs on the parasite histone modifications. Schizont (A) and Trophozoite stages (B) of parasites were treated with clinical dose of different anti-malarial drug for six hours: artemisinin (1 μ M), chloroquine (30 μ M), mefloquine (5 μ M) and pyrimethamine (5 μ M) and with MMS (0.05%).

Surprisingly, the DNA damage repair machinery and overall, the DNA damage response of the parasite remained mostly uncharacterized, in spite of their pivotal role in parasite cellular survival and malaria pathogenesis. The paramount significance of studying the DNA damage repair in the parasite is highlighted by the fact that it drives two of the basic parasitic functions such as antigenic variation and copy number variation (25). The antigenic variation and copy number variation are the two indispensable tools of the malaria pathogenesis.

The parasitic genetic integrity is under constant threat from various exogenous and endogenous stresses that can induce DNA aberrations, which can jeopardize the normal functioning of cells and can also alter the fidelity of the genetic transmission to daughter cells. Eukaryotic cells have evolved a complex DNA damage repair system. The excision

repair and the DSBR repair represent the two most common and robust repair mechanisms, which tackle most of the deleterious DNA lesions. The malaria parasite genome codes for the majority of the DNA repair proteins, but most of these repair factors remain uncharacterized biochemically. Hence, we still lack much of the understanding of how the parasite responds to the damaged DNA. Here, for the first time, we studied the global transcriptional and epigenetic response of *P. falciparum* to DNA damaging agents. The main objective of the study was to characterize the transcriptional and epigenetic signature of DNA damage in *P. falciparum*.

3.3.1 Transcriptional response to DNA damaging agent, MMS

Here, we demonstrated that when challenged with a DNA damaging agent, *P. falciparum* cells respond by the broad transcriptional changes (21% of genome), which include up-regulation of both DSBR and excision repair pathways. Similar types of responses have been reported for yeast, in which alkylating agents induced differential expression in 1/3 of the yeast genome (154). Previously, it has also been reported that the histone deacetylase inhibitor, apicidin, altered the RNA expression of 34% of genes in the trophozoite and 55% in the schizont stage (63). It is noteworthy that the global response to DNA damaging agents in yeast and mammals involves up-regulation of the DNA repair machinery and activation of cell cycle checkpoints. In addition, exposure of alkylating agents such as MMS, also elicit variety of other cellular functions such as protein metabolism, protein folding, cellular trafficking and drug metabolism (155). Our results showed that like in yeast, in *P. falciparum*, DNA damage leads to upregulation of genes involved in generic (indirect) stress responses such as protein synthesis and turnover as well as mitochondrial functions but also direct responses, involving genes of the DNA repair machinery. Our transcriptome profiling results showed that both DSBR and excision repair pathways were likely functional in the malaria parasite.

We observed that DNA damage sensors which sense a DNA lesion and initiate specific repair pathways as a result were also up-regulated in the parasite. These include the homologues of A/G specific adenine glycosylases, XPC-h23B, mutS heterodimer (MSH2-MSH6) and MRN exonuclease complex (MRE11-Rad50), which correspond to DNA damage sensors for BER, NER, MMR and DSBR repair respectively. Until now, neither the mechanism of DNA damage recognition nor the functional characterizations of these factors in the parasite have been explored. Here, we are providing the transcriptional evidence of the existence of these DNA damage sensors in the parasite.

3.3.2 DNA damage induces H4K8 hyperacetylation and H3K9 deacetylation

In eukaryotes, DNA damage recognition and repair occur in the context of chromatin (77). Covalent histone modifications and chromatin remodeling complexes modulate the chromatin architecture and hence, regulate the accessibility of the DNA repair machinery to the DNA damaged site (78). Here, for the first time, we addressed the role of chromatin dynamics in regulating the DNA damage response in the malaria parasite.

Our results illustrated that MMS mediated DNA damage in the parasite leads to a global wave of H4 acetylation, specifically at lysine residues 5, 8, 12 and 16. In contrary, MMS induced H3K9 deacetylation in the trophozoite stage of the malaria parasite. In yeast and mammals, one of the earliest events of DNA damage is the hyperacetylation of lysine residue of H4 by NuA4 histone acetyltransferase (NuA4 HAT) (82). NuA4 HAT is responsible for acetylation of H2A and lysine residues 5, 8, 12 and 16 of H4. Moreover, it has been shown that H4K16Ac is required for recruitment of repair factors, and the loss of H4K16Ac leads to a defective DNA damage response in yeast and mammals (79). Further, it has been reported that mutation in any of the four lysine residues (K5, K8, K12 and K16) of H4 would make cells more sensitive to DNA damaging agents and result in impaired DNA damage response (92). Therefore, H4 acetylation is regarded as hallmarks of eukaryotic DNA damage. MMS

induced H4 hyperacetylation in the parasites suggest that there is a conservation of molecular function of the NuA4 enzyme in the malaria parasite. It has been demonstrated that DNA damage in mammalian cells led to global reduction in the levels of H3K9 and H3K56 acetylation (100). We also observed the dramatic decrease in the levels of H3K9Ac and comparatively, lesser reduction in the H3K56Ac.

The ChIP-chip results revealed that MMS induced H3K9Ac and H4K8Ac was predominantly enriched in the intergenic regions and, thus are possibly associated with transcription. This is in contrast with histone deacetylase inhibitors such as apicidin and TSA that show very limited positive correlation between gene expression and altered acetylation induced by their effect in *P. falciparum* as well as mammals (148,156). 29% of the total up-regulated genes induced by MMS were associated with the H4K8Ac occurrence. These genes belong to cellular processes such as ribosome assembly, chaperones/protein folding, nuclear genes destined to mitochondria, folate biosynthesis and s-glutathionylated proteins. The presence of H4K8Ac on chaperones and nuclear genes destined to mitochondria suggests the role of H4K8Ac in transcriptional regulation of these genes. In contrast, the increase of H3K9Ac occurs preferentially at DNA repair genes that were also induced as a result of MMS-mediated DNA damage. Under normal growth conditions, H3K9Ac and H4K8Ac are selectively enriched on actively transcribed genes belonging to metabolism and ribose assembly related functions in the parasite (64). In the future, it will be important to investigate the molecular mechanism of H3K9Ac and H4K8Ac dynamics in transcriptional up-regulation during DNA repair.

3.3.3 Artemisinin induce DNA damage response in *P. falciparum*

Artemisinin is a sesquiterpene lactone having a characteristic endoperoxide bridge. The endoperoxide bridge of artemisinin is essential for its antimalarial activity. The mechanism of action of artemisinin is debatable (107). It is believed that the heme group released during

hemoglobin digestion cleaves this endoperoxide bridge. This heme dependent activation of artemisinin results in the generation of free radicals which mediate the parasitic death (26). Artemisinin has also been used as an anti-tumor agent. In cancer cells, it has been reported that artemisinin induces DNA damage and elicits HR and non-homologous end joining (NHEJ) repair pathways (113,114). Recently, it has been demonstrated that artemisinin induces DNA damage in the parasite that is mediated by oxidative stress (157). Here, for the first time, we studied the artemisinin induced DNA damage response in the parasite by characterizing the parasite transcriptional and epigenetic response, which was similar to the response caused by a DNA damaging agent (MMS).

In this study, we challenged the late trophozoite stage of 3D7 parasites with 1 μ M of artemisinin in order to study the parasite transcriptional and epigenetic responses. The peak median plasma concentration of artemisinin in patients is around 900 nM to 1 μ M (151). Hence, we used the clinical dose of artemisinin in order to capture the parasite response at a physiological dose. Our results showed that artemisinin, an antimalarial and MMS, a DNA damaging agent co-regulates 14% (746 genes) of the *Plasmodium* genome. In other words, both artemisinin and MMS altered the transcriptional abundance of 14% (746 genes) of the *Plasmodium* genome. Importantly, the cellular functions, which were altered, represent the possible drug targets. Both agents caused transcriptional up-regulation of the parasite excision repair and DSB repair pathways. In addition, both MMS and artemisinin induced expression of a group of genes, which are imported to mitochondria/apicoplast or have mitochondrial/apicoplast localization sequences. DNA damage is known to affect cellular trafficking of proteins and causes dynamic redistribution of proteins from one sub-cellular compartment to another. In yeast, DNA damage facilitates import of many proteins into mitochondria from the nucleus, which includes pro-apoptotic proteins such as p53, several DNA repair proteins, cell cycle regulatory proteins, etc. (158). It has been demonstrated that

DNA damage induced by the oxidative stress or by MMS results in trafficking of Ntg1, a N-glycosylase enzyme, which acts as a DNA damage sensor for base excision repair into mitochondria (159). “Chaperone and their regulators” represent another interesting functional group induced by both agents. MMS is a broad spectrum-alkylating agent that methylates various cellular targets, including specific nucleotides in DNA, RNA and side-chains in proteins (30). Similarly, artemisinin has also been reported to mediate alkylation of various proteins such as hemoglobin, cytochrome c, catalase, etc. (26). The up-regulation of chaperones genes suggest that MMS/artemisinin mediated alkylation may have altered the structure of proteins, which resulted in the misfolding of proteins. Such induction of chaperones proteins were previously reported in yeast when challenged with DNA damaging agents such as MMS and Ionizing radiations (149).

Recently, several factors of the DNA repair machinery have been linked with artemisinin resistance *in vivo* and *in vitro* (61,160). Interestingly, PfRad50 was one of the proteins identified in these studies having a non-synonym single nucleotide polymorphism (SNP). PfRAD50 is part of MRE11 complex that recognizes double-strand breaks. Our results demonstrate that artemisinin induces transcriptional up-regulation of PfRAD50. These results indicate that PfRAD50 could be one of the important cellular targets of artemisinin, which may be associated with artemisinin resistance.

Further, we tested whether; the antimalarial mode of action of artemisinin is mediated by alterations in histone modifications in the parasite. In addition to artemisinin, we tested physiologically relevant concentrations of three antimalarial drugs, including chloroquine, mefloquine and pyrimethamine. In cancer cells, it has been reported that chloroquine induces rapid degradation of histone deacetylase, which results in histone hyperacetylation (161). Rahim *et al.*, have also demonstrated that anti-tumor effect of quinolone drugs involves H3 and H4 hyperacetylations (162). Our results show that artemisinin was the only antimalarial

drug which induced profound chromatin alterations in both trophozoite and schizont stage. Interestingly, the effect of artemisinin and MMS were identical on the parasite chromatin. Both these agents mediate hyperacetylation of H4K8 in trophozoite and schizont stage. Similarly, both drugs had no effect on H3K9Ac at schizont stage, but both mediate the dramatic reduction in the levels of H3K9Ac at the trophozoite stage.

3.4 Summary

Following DNA damage, it has been suggested that eukaryotic chromatin undergoes a dynamic and reversible decondensation around the damaged site that facilitates the accessibility of DNA repair factors (77,78). Chromatin structure can be altered by covalent modifications of histone tails and by ATP-dependent chromatin remodeling complex. It has been firmly established that both histone modifications and chromatin remodelers play significant roles in the DNA damage response (82). Here, first, we summarize the key events of the epigenetic regulation of DNA damage response in eukaryotes and subsequently, discussing our results in *P. falciparum* (**Figure 3.12**).

In eukaryotes, one of the earliest events during DNA damage is the phosphorylation of the serine residue at C-terminal of H2A/H2AX (81). ATM/ATR/DNA-PKs kinases are the serine-threonine kinases belonging to the PI3K kinase family, which are responsible for H2A/H2AX phosphorylation (**Figure 3.12a**). The recognition motif of PI3K kinase is SQ motif (Ser-Gln) at C-terminal of H2A/H2AX (88). Following phosphorylation, NuA4 histone acetyltransferase complex (NuA4-HAT) is recruited at the DNA damaged site. The substrates of NuA4-HAT are H2A, H4 and ATM kinase (79). Mammalian NuA4-HAT complex has 16 subunits, and only 3 have catalytic activities (MYST enzyme, RuvB helicase enzyme and p400 ATPase) (82). NuA4 mediated H4 acetylation (K5, 8, 12 and 16) are important for the DNA damage response. It has been shown that mutations in any of these four lysine residues lead to defective DNA repair in yeast/mammalian cells (92). Moreover, the two chromatin

Figure 3.12 Model for epigenetic regulation of DNA damage response in the eukaryotes. A) Following H2A/H2AX phosphorylation, NuA4-HAT complex is recruited at the DNA damaged site. Mammalian NuA4-HAT complex has 16 subunits and only 3 have catalytic activities (MYST enzyme, RuvB helicase enzyme and p400 ATPase). RuvB helicase and p400 ATPase loosens the chromatin by decreasing histone-histone and histone-DNA interactions. Taken together, H4 hyperacetylation and chromatin loosening by the chromatin remodelers enhance the accessibility of the DNA repair proteins to the DNA damaged foci. B) The DNA damage response in *P. falciparum*. Primarily, the role of H2A phosphorylation in DNA damage regulation has not been studied. Secondly, *Plasmodium* genome encodes all the three catalytic components of NuA4 HAT complex. Both MMS and artemisinin induced hyperacetylation of H4 lysine residues: K5, 8, 12 and 16. Moreover, both induced transcriptional up-regulation of excision repair and DSB repair machinery in the parasite.

remodelers RuvB helicase and p400 ATPase also contribute in the DNA repair by decreasing the histone-histone and histone-DNA interactions, which loosens the chromatin (82,163). Altogether, H4 hyperacetylation and chromatin loosening by the chromatin remodelers enhance the accessibility of the DNA repair proteins, transcriptional factors and cell cycle regulators to the DNA damaged foci.

The characteristic H2A phosphorylation and the enzyme ATM kinase have not been identified in the *Plasmodium* so far. Interestingly, our bioinformatics analysis revealed that *Plasmodium* H2A (PF3D7_0617800) has a conserved SQ sequence near its C-terminal end. The identification of the SQ motif is exciting, which is a potential target of phosphorylation. Further studies needed to be carried out in order to elucidate the role of this conserved sequence in DNA damage response in the parasite. Recently in *T. gondii*, which is another member of apicomplexan, the role of H2AX phosphorylation in DNA damage repair has been established (90). The *Plasmodium* genome encodes all three catalytic components of NuA4 HAT complex (**Figure 3.12b**). These are: 1) TIP60 histone lysine acetyltransferase (PfMYST), 2) p400 ATPase (SWI2/SNF2 ATPase) and 3) RuvB helicase (PfRuvB helicase). Our results show that both MMS and artemisinin induced hyperacetylation of H4 lysine residues: K5, 8, 12 and 16. Additionally, both induced transcriptional up-regulation of excision and DSB repair pathways in the parasite. Taken together, DNA damage is likely a part of antimalarial mode of action of artemisinin given the fact that its effect on *P. falciparum* includes inductions of the DNA damage repair mechanisms and histone acetylations in the similar fashion to MMS; presumably as a result of DNA damage. Further in-depth studies needed to be undertaken to understand the different aspects of DNA damage response in the malaria parasite.

Chapter 4

ARMD Parasites Displaying Mutator Phenotype have Defective DNA Repair

4.1 Introduction

The malaria parasite, *P. falciparum* has become resistant to most of the known antimalarial drugs. Interestingly, Southeast Asia is the epicenter for the origin of most of the antimalarial drug resistance. For instance, the emergence of chloroquine resistance was first reported in Cambodia in the early 1960, followed by pyrimethamine resistance and subsequently sulphadoxine resistance (164-166). More recently, the geographical origin of artemisinin resistance was also established to be in Western Cambodia (167).

The underlying mechanism of antimalarial drug resistance is not fully understood. Until now, all drug resistance phenotypes have been associated with DNA polymorphisms; typically non-synonymous mutations in drug targets and/or drug transporters (168). It is further believed that additional point mutations will enhance the strength of the resistance. For instance, *P. falciparum* harbouring K76T mutation at codon position 76 in *P. falciparum* chloroquine resistance transporter gene (PfCRT) contributes to the development of resistance to chloroquine (169). Similarly, the resistance to the anti-folate drug pyrimethamine, is associated with a single point mutation at codon 108 in the dihydrofolate reductase (*dhfr*) gene, with additional mutations, 51I and 59R, enhancing the strength of the resistance (170). Resistance to sulfadoxine, another anti-folate drug is associated with mutations in the codon 437, 540 and 581 of the dihydropteroate synthetase (170). Moreover, non-synonym point mutation C580Y, in the propeller domain of kelch-13 protein is associated with the artemisinin resistance (171).

It has been suggested that a rapid emergence of the antimalarial drug resistance phenotypes in Southeast Asia is linked with increased mutation capacity of some parasite strains (mutator phenotype), known as accelerated resistance to multiple drugs (ARMD) (59). It has been demonstrated that these hypermutable parasites have higher tendency to develop/acquire drug resistance to novel drugs. In the study conducted by Rathod *et al.*,

various parasite clones were challenged with the two new antimalarial drugs (5-fluoroorotate and atovaquone) in order to study the occurrence of frequency of drug resistance (59). After three weeks of continuous exposure, parasite clones W2 from Indo-china become resistant to both novel drugs, whereas a clone D6 from West Africa remained susceptible to both the antimalarial agents. The ARMD parasites may harbor diverse genetic backgrounds, which may provide additional advantage in developing the actual drug resistance mechanisms (59). Such mutator phenotypes have also been found in bacteria, yeast as well as mammalian cells (172-174). Higher mutation rate provides a selective advantage because it enhances the adaptability of mutator cells to a variety of stress conditions. For example, the bacterial cells with the higher mutation rate generate more antibiotic-resistant conferring mutations as compared to non-mutator bacteria (43). Generally, the mutator phenotypes results from alterations in proteins belonging to the DNA repair and the DNA replication machinery (41,56).

Mismatch repair machinery (MMR) plays a significant role in maintaining the genomic stability and development of drug resistance phenotype. DNA mismatch repair deficient cells can be also unresponsive to several chemotherapeutic drugs such as etoposide, procarbazine, busulfan, cisplatin and doxorubicin (56). Some of the pathogenic bacterial strains eg. *N. meningitidis*, *H. pylori*, *E. coli*, and *S. pneumoniae* exhibit hypermutation rate, which results from their defective MMR machinery (43). Similarly, the defective DNA repair machinery confers resistance to cisplatin and anti-parasitic drug monensin in other protozoa intracellular parasites *T. cruzi* and *T. gondii* (58).

Previously, three studies have been carried out in malaria, trying to understand the link between the defective DNA repair and drug resistance, ARMD phenotype. In one report in 2004, the efficiency to repair ultraviolet light induced DNA damage was found to be considerably reduced in the highly resistant W2 strain (38). In 2007, Bethke *et al.*, showed

that the disruption of MMR repair protein, MSH2-2, in *P. berghei* had no effect on the development of drug resistance (44). In another study in 2011, ARMD parasites W2 and Dd2 were shown to have defective MMR whereas, non-ARMD parasites HB3, D6 and 3D7 were having the proficient MMR activities (60).

These studies suggest that ARMD parasites have reduced DNA repair activities as compared to non-ARMD parasites. However, the underlying genetic and biochemical mechanisms responsible for the differences are still not understood. Here for the first time, we assessed the transcriptional and epigenetic response of ARMD and non-ARMD parasites in the presence of the DNA damaging agent. Further, we carried out whole genome sequencing of these parasites in order to understand the genetic basis of the ARMD phenotype.

4.2 Results

4.2.1 Transcriptional differences between ARMD (Dd2) and non-ARMD (3D7) parasites during DNA damage

In order to assess the DNA repair activities of ARMD and non-ARMD parasites in presence of the DNA damaging agent, we decided to initiate our experiment with a chloroquine resistant parasite, Dd2 (ARMD) and a chloroquine sensitive parasite, 3D7 (non-ARMD). The geographical origin of Dd2 strain was Indo-China region of Southeast Asia, whereas 3D7 was originated from Netherland. We hypothesized that a defective DNA damage response is responsible for the development of mutator phenotype associated with the Dd2 parasites (ARMD), whereas non-ARMD parasite 3D7 has a proficient DNA repair activity. In order to test this hypothesis, we challenged both 3D7 and Dd2 parasites with a DNA damaging agent, MMS for 2 and 6 hours. We performed our treatment in the two developmental stages: trophozoite (24-28 hpi) and schizont stages (34-38 hpi) in both 3D7 and Dd2. Subsequent to

Figure 4.1 Global transcriptional response of *P. falciparum* ARMD (Dd2) and non-ARMD (3D7) to DNA damaging agents. *P. falciparum* 3D7 and Dd2 parasites were treated with MMS (0.05%) at trophozoite stage and schizont stage for 2 and 6 hours. The numbers of genes differentially induced by > 2 fold (P value < 0.05) are shown graphically for trophozoite (**A**) and schizont (**B**) stage. The red and the green histograms represent upregulated and downregulated genes, respectively.

treatment, total RNA was isolated and the transcriptional analysis was carried out as described in “materials and methods”. As established in the previous chapter, that MMS

induces DNA damage response in the *P. falciparum* (3D7 strain), which includes up-regulation of both DNA excision and DSB repair. Our main objective with this study was to assess whether the DNA damaging agent can also induce transcriptional up-regulation of DNA repair machinery in the ARMD parasite, Dd2. We found that overall, 6 hours of MMS (0.05%) treatment leads to large transcriptional changes in both 3D7 and Dd2. The effect of MMS was profound in both the trophozoite and schizont stage. In 3D7, MMS exposure altered the transcriptional abundance of 1169 genes (22% of genome) and 659 genes (12.3%) by >2-fold in the trophozoite and schizont stage correspondingly (**Figure 4.1a**). Similarly, in Dd2, MMS modulated the transcriptional levels of 666 genes (12.5% of genome) and 741 genes (14%) by >2-fold in the trophozoite and schizont stage respectively (**Figure 4.1b**). In trophozoite stage, there were 123 genes that were commonly modulated by MMS in both parasites strains and 187 genes which were commonly altered in the schizont stage.

We also assessed the biological reproducibility of the microarray data by calculating the Pearson correlation coefficient between replicates. **Figure 4.2a** and **Figure 4.2b** represent the scatter plot demonstrating the tight positive correlation between the two replicates for trophozoite stage (0.89 for 3D7 and 0.91 for Dd2 replicate) and schizont stage respectively (0.89 for 3D7 and 0.91 for Dd2 replicate).

4.2.2 Functional classification of MMS induced genes in 3D7 and Dd2: Trophozoite and schizont stage

In order to visualize the MMS induced transcriptional differences between 3D7 and Dd2, we compared the log2 fold change data of expression of 3D7 and of Dd2. In the schizont stage, there was a positive correlation between 3D7 and Dd2 dataset (R value 0.58), but in the trophozoite stage, 3D7 distribution was dramatically different from Dd2 fold change dataset (R value 0.14) (**Figure 4.2a**). These results suggest that gene expression in the trophozoite stage exhibit much more transcriptional variations when challenged with a DNA damaging

Figure 4.2 Correlation between biological replicates for 3D7 and Dd2. The horizontal and vertical axis represents the log2 fold change data. Scatter plot demonstrated strong positive correlation between 2 replicates for both strain. **A) Trophozoite stage:** Scatter plot demonstrating the tight positive correlation between 2 replicates (R value 0.92 for 3D7 replicate and 0.7 for Dd2 replicate). To visualize the MMS induced transcriptional differences between 3D7 and Dd2, we compared the log2 fold change data of expression of 3D7 and of Dd2. In trophozoite stage, 3D7 distribution was dramatically different from Dd2 fold change dataset (R value 0.14). **B) Schizont stage:** Scatter plot demonstrating the positive correlation between 2 replicates (R value 0.8 for 3D7 replicates and 0.9 for Dd2 replicates). To visualize the MMS induced transcriptional differences between 3D7 and Dd2, we compared the log2 fold change data of expression of 3D7 and of Dd2. In schizont stage, there was positive correlation between 3D7 and Dd2 dataset (R value 0.58).

agent, MMS. In the schizont stage, overall, we detected 584 genes, which were differentially expressed between ARMD (Dd2) and non-ARMD (3D7) parasite in the presence of MMS (p-value <0.05). The RNA expressions of 315 genes in 3D7 were significantly higher than the corresponding genes in Dd2 and similarly, there were 281 genes in Dd2, which display higher expression as compared to their expression in the 3D7. In trophozoite stage, MMS induced differential transcription of 1416 genes between 3D7 and Dd2, in which, 843 genes were up-regulated in 3D7 and 573 genes were up-regulated in Dd2 (p-value <0.05) (**Figure 4.3**)

Figure 4.3 MMS induced differential expression of genes. MMS induced differentially expressed genes between the ARMD (Dd2) and non-ARMD (3D7) parasites (P value < 0.05) at trophozoite and schizont stage. Shown are the mean-centered expression log₂ ratios for these genes. Each column represents a fold change log₂ data. The color scale used to represent variations in transcript abundance is shown at the bottom of the figure.

To assess the physiological significance of the genes that were induced by MMS we used gene set enrichment analysis (GSEA), which reveals statistically enriched biological

functions in the given input dataset (175). We carried out GSEA enrichment analysis on MMS treated 3D7 and Dd2 samples separately. Our objectives were 1) to identify the

A

B

Figure 4.4 GSEA enriched genes: Differentially expressed DNA repair genes 3D7 vs. DD2. The gene expression patterns of differentially expressed genes at trophozoite stage are shown. **A)** 18 genes of base excision repair (BER), **B)** 10 genes of nucleotide excision repair (NER), **C)** Nine genes of mismatch repair (MMR) and **D)** Eight genes of double strand break repair pathway (DSBR) were significantly up-regulated in 3D7 as compared to their expression in Dd2 strain.

overlapping functional groups that were induced in both 3D7 and Dd2; 2) to identify parasite strain specific functions and 3) to evaluate whether DNA damage induces enrichment of the DNA repair functions in the drug resistant, Dd2 parasites.

GSEA results revealed that MMS treatment altered various physiological pathways that were similar in both 3D7 and Dd2. For instance, in trophozoite stage, MMS induced up-

regulation of various pathways such as modifications of proteins by adenylation and ribosylation, generation of polyphosphates in both 3D7 and Dd2. In schizont stage, MMS induced the up-regulations of additional pathways such as ribosome assembly genes, genes involved in pre-mRNA processing, genes targeted for mitochondrial import, chaperones regulators and chaperones assisted protein folding. It is worth to mentioning here that in yeast and mammals, DNA damage induced by MMS results in the overexpression of protein folding genes, genes involved in protein modification and also induce alterations in the cellular trafficking of proteins (149). Significant up-regulation of these pathways in both 3D7 and Dd2 suggests that these genes are commonly modulated by MMS in various organisms. Our GSEA results showed that MMS treatment in drug sensitive 3D7 strain, resulted in the up-regulation of DNA repair machinery and genes associated with DNA replication in both trophozoite and schizont stage. In contrast, MMS was not able to induce significant enrichment of DNA repair machinery in mutator Dd2 parasites in both the trophozoite and schizont stage.

Further, we investigated the transcriptional differences in the individual DNA repair genes found to be enriched in GSEA results. In trophozoite stage, 18 genes of BER, 10 genes of NER, 9 genes MMR and 8 genes of DSBR pathways were significantly up-regulated in 3D7 as compared to their expression in the Dd2 strain (**Figure 4.4**). Similarly, the data shown in **Figure 4.5a** represents the comparison of the 34 DNA repair genes that were differentially induced by MMS in 3D7 and Dd2 at schizont stage. Further, differential expressions of 4 genes were validated by quantitative RT-PCR (**Figure 4.5b**).

4.2.3 Identification of defective DNA repair machinery in the ARMD parasites

Next, we want to test whether the observed transcriptional deficiency of DNA repair machinery was restricted to Dd2, or whether this trait is present in other *P. falciparum* strains. Hence, we assessed the transcriptional response to MMS in two more *Plasmodium* strains

(D6 and W2) that have the varying degree of drug susceptibilities and originated from different geographical regions. Overall, we tested four *P. falciparum* strains that exhibit differences in the frequency to acquire drug resistance when exposed to antimalarial drugs for longer duration (59). The parasite strains 3D7 and D6 originated from “Netherland” and “Sierra Leone” and remain drug susceptible even after continuous selection on antimalarial drugs, hence considered as non-ARMD parasites. In contrast, Dd2 and W2 which originated from Southeast-Asia (Indo-China) and had displayed higher frequency in acquiring the novel drug resistance are known as hypermutable ARMD parasites (59). Therefore, we wished to characterize the transcriptional differences in the DNA repair machinery among *Plasmodium* strains with varying degree of drug susceptibility. Hence, we challenged the trophozoite stage of ARMD (W2, Dd2) and non-ARMD (3D7, D6) parasites with MMS (0.05%) for 6 hours. For this analysis, we took extra care to make sure that overall experimental condition such as parasitemia and developmental stage of the IDC cycle remained identical for all the four strains.

Overall, MMS was able to modulate the expression of 1,093 (20%), 1,014 (19%), 1,065 (19%) and 1,018 (19%) genes in 3D7, D6, W2 and Dd2 respectively (>2 -fold, p -value < 0.05) as shown in **Figure 4.6a**. In order to identify MMS induced differentially expressed genes, we compared the global transcriptional abundance level between the two ARMD and non-ARMD parasites. Using this approach, we identified 431 genes that were over-expressed in the non-ARMD and 295 genes that were up-regulated in ARMD parasites (>1.5 -fold, p -value < 0.05) (**Figure 4.6b**). Functional enrichment analysis on these 431 up-regulated genes identified that numerous cellular processes such as DNA excision repair machinery, DNA replication associated genes, histone acetylation and genes associated with nucleosome assembly and regulation were up-regulated in non-ARMD parasites as compared to the ARMD (**Figure 4.6c**). Both MMR and NER repair pathways were found significantly up-

Figure 4.6 Global transcriptional response of *P. falciparum* ARMD (Dd2, W2) and non-ARMD (3D7, D6) to DNA damaging agents. A) The number of genes differentially induced by > 2 folds (P value < 0.05) for each parasite clones are shown graphically. **B)** MMS induced differentially expressed genes between the ARMD and non-ARMD parasites (P value < 0.05) at trophozoite stage. Shown are the mean-centered expression log2 ratios for these genes. Each column represents a fold change log2 data. **C)** Functional enrichment analysis of the 431 up-regulated genes in non-ARMD parasites (3D7 and D6) is shown.

regulated in non-ARMD parasites (3D7 and D6) as shown in **Figure 4.6c**. However, these repair genes were not induced by MMS in the ARMD parasite strains. Taken together, our results demonstrate that ARMD parasites have deficient DNA repair machinery, whereas non-ARMD parasites have proficient DNA repair machinery.

4.2.4 Epigenetic mechanism associated with the ARMD phenotype

4.2.4.1 MMS induces similar alterations in chromatin architecture in ARMD and non-ARMD parasites

The ARMD parasites can acquire resistance to multiple unrelated drugs with different chemical structure and mechanism of action (59). It has been suggested that multiple mechanisms are associated with the development of drug resistance phenotype. Genetic and epigenetic mechanisms have been shown to play key roles in the rapid acquisition of drug resistance in cancer cells (176,177).

Epigenetic mechanisms play an important role in regulating the gene expression pattern in the malaria parasite (64). Additionally, epigenetic mechanisms are responsible for regulating the clonally antigenic variation in *P. falciparum* (178). The antigenic variation is an indispensable tool of malaria pathogenesis by which parasites evade the host immune surveillance. Moreover, these mechanisms also regulate the expression of genes involved in the erythrocyte invasion and sexual differentiation (179). In the previous chapter we demonstrated that MMS-induced DNA damage caused significant alterations in the parasite epigenetic code, specifically in H3 and H4 acetylation. Here, we wished to understand whether MMS induces similar or differential response on the chromatin in parasites having different drug susceptibilities.

To explore the differences in the histone marks, we treated ARMD and non-ARMD parasites with MMS (0.05%) for two and six hours and investigated the alterations in H4 and H3 marks. We performed the experiment in two developmental stages: trophozoite and

schizont. Our immunoblot results suggested that in spite of the differences in drug sensitivity, both ARMD (Dd2) and non-ARMD (3D7) parasites showed similar alterations in the global chromatin landscape. This includes the hyperacetylation of lysine residues of H4 (K 5,8,12 and 16) in both trophozoite and schizont stage. In addition, the loss of H3K9 and H3K56 acetylation was observed only in the trophozoite stage. Interestingly, most of the effected histone marks were altered even with two hours of MMS exposure. In schizont stage, the levels of H3K9Ac, H3K56Ac and H3K4me3 remain unaffected in both 3D7 and Dd2. (Figure 4.7).

Figure 4.7 Effect of MMS on ARMD and non-ARMD parasite chromatin. 3D7 and Dd2 parasites were treated with MMS (0.05%) at trophozoite (A) and schizont stage (B) for 6 hours and followed by protein isolation. Total protein extracts was analyzed by SDS-PAGE and various histones 3 and histone 4 marks were detected by immunoblot analysis.

As the method used does not give quantitative differences in the histone marks levels between ARMD and non-ARMD parasite, further careful investigations will be needed in order to analyze the subtle epigenetic events that may be associated with the drug resistance phenotype in the malaria parasite. In summary, our results suggest that DNA damage induced similar chromatin alterations in both 3D7 and Dd2, and epigenetic mechanisms may not play a large role in regulating the acquisition of drug resistance phenotype in the parasites.

4.2.5 Genetic mechanisms associated with the development of ARMD “mutator” phenotype

The DNA repair system protects the genome from deleterious DNA lesions, but also acts as a robust mechanism to acquire drug resistance. In yeast, bacteria as well as mammals, mutations in the DNA repair genes underlay mutator phenotypes, which can subsequently result in the rapid emergence of drug resistance (172,173).

One of the important mechanisms associated with defective DNA repair is genetic polymorphism such as single nucleotide polymorphism (SNP). These point mutations in the DNA damage repair genes can lead to impaired DNA repair functions. It has been described in the bacterial cells that SNPs in the MMR repair proteins such as MutL and MutS confer mutator phenotypes and promote the occurrence of antibiotic resistance (180,181). Similarly, mutations in the DNA repair genes MSH2, PMS1 and Rad 50 lead to increase in the frequency of drug resistance in the pathogenic fungus, *Candida albicans* (182). In humans, a point mutation in the MMR gene hPMS1, leads to reduced MMR efficiency that contributes to the development of mutator phenotype (183). In our previous section, we illustrated that ARMD parasites (W2 and Dd2) have a reduced or defective DNA damage response. In other words, these parasites might have increased DNA damage tolerance levels. Hence, we want to assess whether these DNA repair defects are associated with mutations in the DNA repair machinery in the ARMD parasites.

In order to identify genetic differences between ARMD (Dd2, W2) and non-ARMD (3D7, D6) parasites, we carried out high-throughput whole-genome sequencing using the Illumina Mi-seq platform. The sequence data was aligned to the reference *Plasmodium* genome (3D7 genome) as described by Mok *et al* (135). Compared to the 3D7 reference strain, we identified 6550, 7231 and 7253 non-synonymous SNPs (nsSNPs) within coding regions of 2099, 2089 and 2046 genes in the Dd2, W2 and D6 strains, respectively (**Table 4.1**).

Table 4.1 SNP's identified in Dd2, W2 and D6 clones.

	Dd2	W2	D6
Intron-SNP	1523	1628	1497
Exon-s-SNP	2900	2812	2973
Exon-ns-SNP	6550 (2099 genes)	6725 (2088 genes)	6278 (1942 genes)
Total Intra-SNP	10973	11165	10748

Reference genome : 3D7

Map score > 30

The table shows the total number of single nucleotide polymorphism detected within intron and exon (synonymous and non-synonymous SNPs) among D6, Dd2 and W2. The reference genome is 3D7. (ns-SNP is non- synonymous single nucleotide polymorphism, s-SNP is synonymous single nucleotide polymorphism). Only those reads were selected with mapping score >30.

As our objective was to identify polymorphisms in the DNA repair genes, hence, for further analysis we focussed only on exonic non-synonymous SNPs (nsSNPs) in the DNA repair genes. We identified 57 DNA repair genes that have nsSNPs in at least one strain (**Table 4.2**). These repair genes belonging to various repair pathways, including MMR, BER, NER and DSBRR repair.

Table 4.2 SNPs identified in the DNA repair genes among Dd2, W2 and D6 clones These are 57 DNA repair genes which have non-synonymous SNPs within coding region. The reference genome is 3D7.

Gene ID	Description	Dd2	W2	D6
PF3D7_0726300	DNA mismatch repair protein PMS1, putative	18	13	4
PF3D7_1244200	transcription factor Tfb2, putative	11	11	10
PF3D7_1303800	conserved <i>Plasmodium</i> protein, unknown function	7	6	2
PF3D7_1368800	DNA repair endonuclease, putative	7	6	8
PF3D7_1368800	DNA repair endonuclease, putative	6	6	8
PF3D7_1037000	DNA polymerase zeta catalytic subunit, putative	5	5	5
PF3D7_0509500	conserved <i>Plasmodium</i> protein, unknown function	5	5	0
PF3D7_0514100	ATP-dependent DNA helicase UvrD	5	5	5
PF3D7_0310500	DEAD box helicase, putative	4	4	3
PF3D7_0504200	ATP-dependent helicase, putative	4	4	0
PF3D7_0605800	DNA repair protein RAD50, putative	4	2	12
PF3D7_0820000	Snf2-related CBP activator, putative	3	3	2
PF3D7_1455300	conserved <i>Plasmodium</i> protein, unknown function	3	3	1
PF3D7_0910500	DNA repair protein REV1, putative	3	3	2
PF3D7_1316900	conserved <i>Plasmodium</i> protein, unknown function	2	2	0
PF3D7_0710400	nucleotide excision repair protein, putative	2	2	2
PF3D7_1363500	DNase I-like protein, putative	2	2	1
PF3D7_1449700	exosome complex exonuclease RRP6, putative	2	2	2
PF3D7_1471600	conserved <i>Plasmodium</i> protein, unknown function	2	2	3
PF3D7_0204800	3'-5' exonuclease, putative	2	2	1
PF3D7_0409600	replication protein A1, large subunit	2	2	0
PF3D7_0411900	DNA polymerase alpha	2	2	0
PF3D7_0513600	deoxyribodipyrimidine photo-lyase, putative	2	2	9
PF3D7_0630300	DNA polymerase epsilon, catalytic subunit a	2	2	3
PF3D7_0908000	P1 nuclease, putative	2	2	0
PF3D7_1250800	DNA repair protein rhp16, putative	2	2	3
PF3D7_1331100	DEAD box helicase, putative	1	1	2
PF3D7_1343400	DNA repair protein RAD5, putative	1	1	0
PF3D7_1304100	DNA ligase I	1	1	1
PF3D7_1362500	exonuclease, putative	1	1	3
PF3D7_1314900	TFIIH basal transcription factor subunit, putative	1	1	0
PF3D7_0710100	conserved <i>Plasmodium</i> protein, unknown function	1	1	0
PF3D7_0803400	DNA repair protein RAD54, putative	1	1	1
PF3D7_1003700	conserved <i>Plasmodium</i> protein, unknown function	1	1	0
PF3D7_1010200	DNA2/NAM7 helicase, putative	1	1	0
PF3D7_1106000	RuvB-like helicase 2	1	1	0
PF3D7_1332100	conserved <i>Plasmodium</i> protein, unknown function	1	1	2
PF3D7_1337300	exoribonuclease, putative	1	1	1
PF3D7_1411400	plastid replication-repair enzyme	1	1	2

PF3D7_1429900	ATP-dependent DNA helicase, putative	1	1	0
PF3D7_1430600	exodeoxyribonuclease III, putative	1	0	0
PF3D7_1432600	conserved <i>Plasmodium</i> protein, unknown function	1	1	1
PF3D7_1449700	exosome complex exonuclease RRP6, putative	1	2	2
PF3D7_0107800	double-strand break repair protein MRE11, putative	1	1	0
PF3D7_0111300	replication factor c protein, putative	1	1	1
PF3D7_0206000	DNA repair endonuclease, putative	1	1	1
PF3D7_0216000	DEAD/DEAH box helicase, putative	1	1	2
PF3D7_0305600	AP endonuclease , putative	1	1	0
PF3D7_0416400	histone acetyltransferase, putative	1	1	3
PF3D7_0505500	DNA mismatch repair protein MSH6, putative	1	1	2
PF3D7_0604600	DNA helicase, putative	1	1	1
PF3D7_0614800	endonuclease III homologue, putative	1	1	0
PF3D7_0628600	conserved <i>Plasmodium</i> protein, unknown function	1	1	0
PF3D7_0917100	N-glycosylase/DNA lyase, putative	1	1	3
PF3D7_0934100	DNA excision-repair helicase, putative	1	1	0
PF3D7_1215700	conserved <i>Plasmodium</i> protein, unknown function	1	1	1
PF3D7_1235900	pre-mRNA-splicing factor SYF1, putative	1	1	5

Further, we identified 18 DNA repair genes, which have SNPs exclusively in the ARMD parasites (Dd2 and W2) (**Table 4.3**). Surprisingly, out of 18, four DNA repair genes had been previously reported to have same point mutation and were found to be associated with artemisinin resistance in Southeast Asia (61). These were PF3D7_0509500 (conserved *Plasmodium* protein, unknown function), PF3D7_0710100 (conserved *Plasmodium* protein, unknown function), PF3D7_1106000 (RuvB-like helicase 2) and PF3D7_1343400 (DNA repair protein RAD5, putative) (61). In addition, there were also ARMD-specific polymorphisms in the genes encoding “sensors” of DNA damage. This includes homologues of MRE11 (PF3D7_0107800)) and AP lyase (PF3D7_0305600) belonging to DSBR and BER pathway, which showed polymorphisms in the ARMD strains compared to non-ARMD.

Table 4.3 List of SNPs located in the DNA repair genes of ARMD parasites (Dd2 and W2).

These were 18 DNA repair genes which have non-synonymous SNP within the coding region of ARMD parasites. Only one repair gene (PF3D7_0509500) had 4 SNPs whereas remaining 17 genes had only one SNP per gene.

Gene id	Description	SNP position	Non-ARMD		ARMD	
			3D7*	D6	W2	Dd2
PF3D7_1429900	ATP-dependent DNA helicase, putative	1178968	T	T	A	A
PF3D7_0305600	AP endonuclease (DNA-[apurinic or apyrimidinic site])	277369	G	G	A	A
PF3D7_0504200	ATP-dependent helicase, putative	168159	A	A	G	G
PF3D7_1316900	conserved <i>Plasmodium</i> protein, unknown function	701120	C	C	T	T
PF3D7_1003700	conserved <i>Plasmodium</i> protein, unknown function	169791	C	C	G	G
PF3D7_0628600	conserved <i>Plasmodium</i> protein, unknown function	1182311	A	A	T	T
PF3D7_0509500	conserved <i>Plasmodium</i> protein, unknown function	401713	A	A	G	G
PF3D7_0509500	conserved <i>Plasmodium</i> protein, unknown function	401714	G	G	A	A
PF3D7_0509500	conserved <i>Plasmodium</i> protein, unknown function	401724	A	A	T	T
PF3D7_0509500	conserved <i>Plasmodium</i> protein, unknown function	401725	A	A	G	G
PF3D7_0934100	DNA excision-repair helicase, putative	1345021	T	T	A	A
PF3D7_0411900	DNA polymerase alpha	528768	T	T	A	A
PF3D7_1010200	DNA2/NAM7 helicase, putative	408872	T	T	A	A
PF3D7_0107800	double-strand break repair protein MRE11, putative	323034	G	G	A	A
PF3D7_0614800	endonuclease III homologue, putative	616527	C	C	G	G
PF3D7_0908000	P1 nuclease, putative	370170	T	T	C	C
PF3D7_0409600	replication protein A1, large subunit	453720	A	A	C	C
PF3D7_1314900	TFIIH basal transcription factor subunit, putative	633490	G	G	C	C
PF3D7_0710100	conserved <i>Plasmodium</i> protein, unknown function	460942	C	C	A	A
PF3D7_0509500	conserved <i>Plasmodium</i> protein, unknown function	401707	G	G	A	A
PF3D7_1343400	DNA repair protein RAD5, putative	1718368	T	T	G	G
PF3D7_1106000	RuvB-like helicase 2	257027	T	T	C	C

SNPs located in four genes previously identified to be associated with artemisinin resistance in Southeast Asia are highlighted in red (61). (*3D7 is the reference genome).

Each individual repair pathway has a dedicated set of proteins that can recognize a specific DNA lesion and initiates DNA repair. These recognition proteins are known as DNA damage sensors. Further, we want to investigate DNA polymorphisms in the DNA damage sensors between ARMD and non-ARMD parasites. **Figure 4.8** shows the distribution of SNPs among sensory proteins. Overall, we found genetic variations within each class of DNA damage sensors. Strikingly, we identified extensive polymorphism in the *P. falciparum* homologue of PMS1 protein (PF3D7_0726300) in ARMD parasites, which potentially involved in the MMR repair. Interestingly, artemisinin resistant parasites from Southeast Asia were also found to carry SNPs in the PMS1 gene (61). It has also been reported that additional point mutations enhance the strength of the drug resistance to antimalarial drugs (184). One hypothesis, which we postulate is that these large numbers of mutations in the MMR protein, PMS1, may lead to a defective MMR repair and may facilitate the acquisition of mutator phenotype. Overall, this result suggests the existence of diverse genetic variations among different sensory proteins in *Plasmodium* strains with varying degree of drug susceptibility.

Taken together, defective DNA damage response such as impaired MMR repair mechanism will lead to a higher recombination rate. Importantly, a higher recombination rate will facilitate the adaptability of the ARMD parasites to antimalarial drugs. In other words, DNA repair deficient parasites exhibit higher mutation rate, which may generate more mutations in the known drug resistance or drug transporter genes. Although further experimental studies are needed to investigate the role of these mutations in the ARMD phenotype of *P. falciparum*, here we showed their occurrence in the parasite cells that failed to respond to DNA damage induced by MMS.

4.3 Discussion

Defects in DNA repair machinery were demonstrated to play significant roles in mutator phenotypes in various organisms such as yeast, mammals, bacteria, *T. brucei* and *T. gondii*

Figure 4.8 DNA polymorphisms in the DNA damage sensors between ARMD and non-ARMD parasites. There are three different types of repair mechanisms: DNA excision repair (BER, MMR, NER), double strand break repair (DSBR) and direct repair (DNA photolyase). Each repair mechanism has different set of sensory proteins which detect DNA damage. For example RAD23 and XPC are the DNA damage recognition proteins for NER repair mechanism. Each box have the distribution of SNPs in the DNA damage sensory proteins for a specific repair pathway among D6, 3D7 (non-ARMD) and Dd2, W2 (ARMD).

(172-174). In malaria, it has been suggested that ARMD parasites have reduced DNA repair capacity as compared to the wild type (non-ARMD) parasites (38). However, the underlying genetic and biochemical mechanisms responsible for the ARMD phenotype are still not understood. Here for the first time, we assessed the transcriptional and epigenetic response of mutator/ARMD (Dd2 and W2) and non-ARMD parasites (3D7 and D6) in presence of DNA damaging agent. Further, we carried out whole genome sequencing of these parasites in order to understand the genetic basis of accelerated resistance to multiple drugs phenotype (ARMD).

4.3.1 Identification of defective DNA repair machinery in the ARMD parasites

Our main goal was to characterize the transcriptional differences in the DNA repair machinery among *Plasmodium* strains with varying degree of drug susceptibility. Our results showed that the MMS treatment in drug sensitive 3D7 strain caused the up-regulation of the DNA repair machinery at both trophozoite and schizont stages. In contrast, the DNA damaging agent, MMS was not able to induced significant enrichment of the DNA repair machinery in the mutator, Dd2 parasites in both stages. Further, we showed that both MMR and NER repair machinery was found to be significantly up-regulated in the non-ARMD parasites (3D7 and D6) as compared to the ARMD parasites (Dd2 and W2).

Taken together, our results suggest that the ARMD parasites have a deficient DNA repair machinery, whereas the non-ARMD parasites have proficient DNA repair machinery. In other words, the mutator parasites display higher DNA damage tolerance.

4.3.2 Epigenetic mechanisms associated with the ARMD phenotype

The acquisition of drug resistance phenotype is mediated by several factors and multiple mechanisms. It has been reported that the genetic and epigenetic mechanisms contribute significantly in the acquisition of the drug resistance phenotype in mammalian cells and yeast (105). Recently, it has been found in human cells that a histone mark, H3K36me3, plays a profound role in preventing spontaneous mutations by regulating the MMR repair system (104). In yeast, it has been suggested that the H3K56 acetylation and deacetylation coordinate with the MMR system in preventing spontaneous mutagenesis (105). In another study, it has been found that a heterochromatic mark, H3K9me3 was strongly associated with mutation rate variability in human cancer cells (106). Further, it has been demonstrated that the epigenetic modifications regulate differential expression of key drug resistance genes. For example, DNA methylation and H3 acetylation were found to be responsible for the transcriptional up-regulation of multiple drug resistance gene (MDR1) in the human cancer

cells, which led to the acquisition of drug resistance phenotype (177). Genome-wide DNA cytosine methylation has also been found to be associated with drug resistance in humans (185).

However, the role of chromatin in the context of spontaneous mutagenesis has not been investigated in the malaria parasite. Hence, we wished to understand the global role of histone modification changes in the acquisition of drug resistance. In our previous results, we found that the ARMD and non-ARMD parasites show differences in the genome-wide transcriptional response to DNA damage. Here, we wanted to study, how chromatin responds to the DNA damage in ARMD and non-ARMD parasites. To achieve this, we challenged both the ARMD (Dd2) and non-ARMD parasites (3D7) with DNA damaging agent, MMS. We observed that in spite of differences in drug susceptibility, both ARMD and non-ARMD parasites exhibited similar alterations in the global chromatin landscape. This includes the hyperacetylation of lysine residues of H4 (K 5, 8, 12 and 16) and loss of H3K9 acetylation. As we are not able to assess the quantitative differences in histone marks level between ARMD and non-ARMD parasite. Hence, further careful investigations are needed to study the role of chromatin-mediated mechanisms in the development of drug resistance phenotype. In summary, our results suggest that DNA damage induces similar chromatin alterations in both 3D7 and Dd2.

4.3.3 Genetic mechanisms associated with the development of ARMD “mutator” phenotype

Lastly, we explored the genetic mechanisms that may contribute in the development of mutator phenotype. Single point mutations in the DNA repair genes of mismatch repair (MMR) and homologous recombination (HR) have been shown to confer the mutator phenotype in several eukaryotic organisms (173). These point mutations in the DNA damage repair genes can lead to impaired DNA repair functions.

Table 4.4 The comparison of drug resistant (ARMD) and drug sensitive (Non-ARMD) *P. falciparum* clones.

	ARMD (Dd2, W2)	Non-ARMD (3D7, D6)
Geographical origin	South east Asia (Indo-china)	Netherland (3D7) Sierra Leone (D6)
Drug susceptibility Patterns	Higher frequency to acquire drug resistance	Lower frequency to acquire drug resistance
a) Chloroquine b) Pyrimethamine c) Quinine d) Sulfadoxine	Resistant Resistant Resistant Resistant	Sensitive Sensitive -/sensitive(D6) Resistant(3D7)/Sensitive (D6)
DNA damage response*		
a) Transcriptional response	DNA repair machinery was not up-regulated or parasite possess increased DNA damage tolerance mechanisms	DNA repair machinery is proficient: DNA damage induces up-regulation of DNA excision and DSB repair
b) Epigenetic response	Alterations of H3 and H4 lysine acetylation	Alterations of H3 and H4 lysine acetylation
Genetic mechanisms	Point mutation in 18 DNA repair genes (RAD5, MRE11, PfRuvB helicase)	-

In our previous section, we illustrated that the ARMD (W2 and Dd2) parasites, have reduced or defective DNA damage response (at the transcriptional level). Additionally, DNA damage induced similar chromatin alterations in both 3D7 and Dd2. Therefore, we assessed whether these DNA repair defects were associated with mutations in the DNA repair machinery in these ARMD parasites.

Our genome sequencing results identified point mutations in 18 DNA repair genes that were exclusively present in the mutator parasites (Dd2 and W2). Four of the DNA repair genes had been previously reported to carry same mutations and were found to be associated with artemisinin resistance in Southeast Asia (61). These were PF3D7_0509500 (conserved

Plasmodium protein, unknown function), PF3D7_0710100 (conserved *Plasmodium* protein, unknown function), PF3D7_1106000 (RuvB-like helicase 2) and PF3D7_1343400 (DNA repair protein RAD5, putative). In addition, some of the polymorphism was in genes whose product was involved in recognizing DNA damage such as MRE11 (putative) and AP lyase (putative) belonging to DSBR and BER pathway respectively. Our results suggest that point mutations in these DNA repair genes can lead to a defective DNA damage response such as impaired mismatch repair, which may lead to a higher recombination rate. Importantly, higher recombination rate can facilitate the adaptability of ARMD parasites by enhancing the genetic variability. In other words, DNA repair deficient parasites exhibit higher mutation rates, which may generate more mutations in the known drug resistance, or drug transporter genes.

Taken together, we demonstrated that the mutator parasites (Dd2 and W2) have defective DNA repair machinery and have mutations in key DNA repair proteins belonging to both excision and DSBR repair pathways (**Table 4.4**). Further investigations are needed to validate the role of these mutations in the development of mutator phenotype in the malaria parasites.

Chapter 5

Chromatin Mediated Regulation of Differential Gene Expression in *Plasmodium falciparum*

5.1 Introduction

The apicomplexan parasite, *P. falciparum*, is well adapted to flourish in fluctuating environments such as the mosquito midgut and salivary glands, human liver cells and red blood cells. However, there are other obstacles that parasites must confront, which mainly include the host immune response and drug pressure. Differential gene expression in *Plasmodium* plays a significant role in the adaptability of the parasite to changing environment as well as contributing in the processes related to evasion of the host immune surveillance and host cell invasion. The exact mechanism of such transcriptional variations in the *Plasmodium* gene expression is not known. Our objective was to investigate the role of chromatin in the differential expression between two laboratory adapted field isolates of *Plasmodium*. The experimental studies were sub-divided into three main goals:

1) Identification of differentially expressed genes between *P. falciparum* strains

To identify genes, which were differentially expressed between *P. falciparum* strains, Dd2 and T996, we performed whole genome transcriptome analysis of the asexual IDC cycle by DNA microarray and then we compared the gene expression profiles across the IDC between two strains.

2) Unraveling the epigenetic landscape of differentially expressed genes

We employed the ChIP-chip technique to generate epigenetic maps for histone modifications (H3K9Ac and H3K56Ac) across the IDC in the Dd2 strain and compared these results to our past work on T996 strain (64). By such comparative analysis, we could distinguish differentially enriched histone marks between the two strains. Further, results from the ChIP-chip were correlated with the transcriptome analysis. The overall objective was to find the association between differentially expressed genes and differentially enriched histone marks between the two strains.

3) Molecular characterization of a putative transcription factor which was found to be both differentially acetylated and differentially expressed.

Our results suggest that most of the transcriptional variations observed between Dd2 and T996 strain were associated with the differential H3K56Ac. Our analysis identified 97 genes whose differential expression correlated with the differential H3K56Ac. Some of these genes are PfHDAC1 (histone deacetylase 1), ApiAp2 transcription factor, putative zinc finger containing proteins, putative ubiquitin protein ligase E3 and erythrocytic binding antigen-165 (EBA-165). Further, we have characterized one of the AP2 domains containing transcription factor that has not been studied previously. The PlasmoDB gene ID of this putative transcription factor is PF14_0271 (PF3D7_1429200).

5.2 Results

5.2.1 Identification of differentially expressed genes between *P. falciparum* strains

5.2.1.1 Pearson method

The main objective was to uncover the epigenetic mechanism behind the transcriptional variations between the two *P. falciparum* strains, Dd2 and T996. To achieve this, first, we identified differentially expressed genes by comparing their IDC transcriptome. The transcriptome for Dd2 and T996 strains were generated by collecting RNA samples at five different time points of the IDC cycle. These time points corresponding to 10 hours post-invasion (hpi), 18hpi, 26hpi, 34hpi and 42hpi, which represents the different developmental stages (rings, trophozoite and schizont) of the IDC cycle.

Of the 5224 *Plasmodium* genes represented on the microarray, we obtained the IDC time course expression profiles for 4344 genes in both strains, which corresponding to 83% of the *Plasmodium* genome. In this comparative study, all the genes belonging to the hyper-variable gene families such as var, rifin, stevor, etc. were removed from the data analysis. The major reason for exclusion of these polymorphic genes was to focus on the differential transcription

caused by the epigenetic mechanisms rather than due to the sequence polymorphisms. The stage-specific expression of the genes in the IDC cycle was reported to be highly conserved among the *Plasmodium* strains (116). Similarly, when we compared the time-course gene expression profiles between Dd2 and T996 using Pearson's correlation coefficient (PCC), we found that about 4125 genes showed a positive correlation of more than 0.4 out of 4344 common genes. This indicates that 94% genes showed similar expression profiles between Dd2 and T996, whereas only 67 genes showed negative correlation of less than 0.4, this means that, only 1.5% genes exhibit distinct expression profiles (**Figure 5.1**). This suggests that the timing of maximal expression of a gene during the IDC cycle is essentially identical in the two strains. Thus, it reflects that the transcriptional regulatory machinery is highly conserved in the genetically and geographically distinct strains, Dd2 and T996.

Figure 5.1 IDC transcriptome of Dd2 and T996. A) Heat maps of the Dd2 and T996 depicting the mRNA expression profiles of 4344 genes along the developmental IDC cycle. Red represents over expression of the gene, and green represents under expression of the gene. Each row represents mRNA abundance of a single gene. The identical profile between Dd2 and T996 signifies the conservation of transcriptional regulatory machinery. B) Histogram showing the distribution of Pearson's correlation between the gene expression profiles from Dd2 and T996. 94% genes out of 4344 had Pearson's correlation of more than 0.4.

5.2.1.2 Euclidean distance method

Transcriptional variations could be attributed to the differences in the timing of maximal abundance of transcripts, or due to differences in the overall transcript levels across the IDC between the two strains, Dd2 and T996. Since we observed that there is a strong conservation of the timing of expression during the IDC (based on PCC results), we further investigated the differences in the gene expression profiles based on the transcript abundance across the IDC between these strains. Euclidean distance is a distance measure tool that is used to compute the similarity or dissimilarity between gene expression profiles. We used Euclidean distance as a tool to identify differences in expression profile across IDC in context with differences in the level of transcript abundance between two strains. Higher Euclidean distance means the expression profiles are dissimilar between two and vice versa.

We calculated the Euclidean distance between the expression profiles for all 4344 genes and analyzed the top 10% of genes (434 genes having highest Euclidean distance greater than 1.8) by hierarchical clustering. Similarly, the bottom 10% of genes (another 434 genes with lowest Euclidean distance) were also analyzed using the same approach. Hierarchical clustering revealed that these top 10% genes (>1.8 Euclidean distance) indeed represented the most differentially expressed genes and bottom 10% genes with lowest Euclidean distance represented genes with identical expression profiles between these strains (**Figure 5.2a and 5.2b**). PF14_0740, a *Plasmodium* exported protein of unknown function, is one such example that was found to be differentially expressed between Dd2 and T996. PF14_0740 is expressed during the early stages of the IDC in Dd2 whereas it is not expressed in T996 at any stages/time points and was identified by the top 10% criteria of Euclidean distance (**Figure 5.2c**). Similarly, PF10_0320, represents an example whose expression is highly identical across the IDC in both the strains and belongs to the bottom 10% genes having lowest Euclidean distance (**Figure 5.2d**).

Figure 5.2 Overview of the transcriptional variations between Dd2 and T996. A) Heat map of the top 10% genes having largest Euclidian distance between expression profiles of Dd2 and T996 and were found to be most differentially expressed. B) Heat map of genes exhibiting identical expression profiles with lowest Euclidean distance between two strains. C) The time course expression profile of PF14_0740, which was found to be highly differentially expressed between Dd2 and T996 using top 1 % criteria of Euclidean distance. D) The time course expression profile of PF10_0320, which was found to be highly similar which belongs to the bottom 10% of genes having lowest Euclidean distance. Expression levels (Log2 ratio of mRNA expression relative to the reference pool) are plotted against the time points which corresponding to different developmental stages of the IDC cycle.

5.3 Unraveling the epigenetic landscape of differentially expressed genes

After identifying the differentially expressed genes, our next goal was to evaluate the differences in the nucleosome occupancy of the histone modifications (H3K9Ac and H3K56Ac) in Dd2 and T996 strains. These modifications were chosen for the investigation because temporal changes in the occupancy of these PTMs correlate with the dynamic changes associated with the IDC transcriptome (133).

5.3.1 H3K9Ac is not associated with differentially transcribed genes

ChIP-chip experiments were performed on the formaldehyde cross-linked Dd2 strain using *P. falciparum* specific long oligonucleotides DNA microarray as described in “material and method”. Like the transcriptome analysis (above), the ChIP-chip analysis for both strains, Dd2 and T996, were carried out at the same time point samples so that the two datasets would be compared directly. Our *Plasmodium* DNA array consists of 15,518 microarray probes/genetic loci (4,447 intergenic probes, IGR and 11,071 open reading frame probes, ORF) which represent 5363 genes of the *Plasmodium* genome. In this microarray, there are many genes whose coding regions are represented by more than one probe with an average spacing between probes of 1.1Kb.

To uncover differentially enriched H3K9Ac marks, we compared log₂ ratio (IP/input) data of 5 time points from H3K9Ac ChIP-chip analysis of Dd2 with the H3K9Ac data from T996 using Pearson’s correlation method. To ensure the statistical significance of the results, each ChIP-chip dataset was obtained by averaging three independent experiments. Further, during data mining only those probes were included in which at least 66% of the data were present throughout the time course experiment. As an additional criterion for filtering, only genetic loci (represented by the microarray probes) whose occupancy of H3K9Ac exhibited a change of at least 1.5 fold across the IDC were included in the analysis. The total number of overlapping probes enriched with the H3K9Ac from Dd2 and T996 analysis that passed these

conditions was 1713. Out of 1713, 1077 oligos (about 63%) showed positive PCC of more than 0.4 and only 53 oligos (about 3%) had the negative correlation of less than 0.4. These results demonstrated that overall two datasets were positively correlated with each other, which indicate that the genome-wide occupancy of H3K9Ac is essentially conserved among different *Plasmodium* strains.

5.3.2 Euclidean distance and differential H3K9Ac

In the next stage, we determined the differential H3K9Ac using the Euclidean distance method as described above. We calculated the Euclidean distance between H3K9Ac profiles for the two strains, for all the 8240 detected probes (no fold change across the IDC). The top 10% probes (824 probes/769 genes) having the largest distance between their ChIP-chip profiles representing the most differentially H3K9 acetylated genetic loci and were used for analysis. In order to find differentially H3K9Ac genetic loci, which might be associated with the differential expression, this group of genetic loci were overlapped with the top 10% of the genes that were found to be differentially expressed by Euclidean distance (**Figure 5.3a**). Overlapping of these two datasets, identified about 82 genetic loci (73 genes) but the overlap was not statistically significant with the p-value of 0.4761 (binomial distribution test). Moreover, hierarchical clustering revealed that the differences in acetylation and expression corresponding to these overlapped probes were not significant (**Figure 5.3b**). Further, the Mann-Whitney U test was used to calculate the statistical significance of this overlap. The result from this test also showed that the correlation between differential H3K9 acetylation and differential expression for these loci was not statistically significant, as compared with the correlation between differences in H3K9 acetylation and differences in expression for all the detected 8011 genetic loci or probes (these 8011 probes were considered as the reference or background) (p-value is 0.36) (**Figure 5.3c**).

Figure 5.3 Association of H3K9Ac with differential expression. **A)** Venn diagram showing significant overlaps between the two groups. One group represents top 10% differentially acetylated genetic loci/ probes and other group comprises the top 10% differentially expressed genes. Overlapping genes had a p-value (binomial value) 0.4761. **B)** Hierarchical clustering of 82 probes (73 genes) representing the differences in H3K9Ac and of the corresponding differentially expressed genes. **C)** Density plot of correlation between differential H3K9Ac and differential expression. Distributions of reference 8011 probes are shown in black color and of 82 probes are shown in red color.

In conclusion, the temporal pattern of H3K9Ac occupancy across the genome during the IDC cycle is conserved between the two genetically distinct *Plasmodium* strains. Euclidean distance method revealed that there is no significant correlation between the differential H3K9 acetylation and differential transcription. Thus, our results indicate that differential transcription between Dd2 and T996 is likely not associated with the dynamic changes of H3K9 acetylation occupancy.

5.3.3 Association of H3K56Ac with differential transcription: Euclidean distance method

In the next part, we calculated the Euclidean distance between each H3K56Ac profile for two strains and then sorted profiles based on the greatest to the smallest distance. The top 10% profile with the greatest distance between ChIP-chip profiles represents the most differentially acetylated genetic loci. In order to determine the correlation between the differentially H3K56Ac and differential expression, this group of top 10% probes (949 probes/849 genes) was overlapped with the top 10% of genes (474 genes) that were found to be differentially expressed by the Euclidean distance. The overlap of the two datasets was statistically significant (p-value 0.0261) (**Figure 5.4a**). Hierarchical clustering of the overlapped genetic loci (108 probes/97 genes) revealed that these loci were differentially acetylated and differentially expressed between the two strains (**Figure 5.4b**). Differential H3K56 acetylation and differential expression of these 108 genetic loci were correlated using Pearson correlation method. About 49% of the overlapped probes (53/108) had positive Pearson correlation (≥ 0.4 PCC), 64% of the overlapped probe (68/108) had positive Pearson correlation (≥ 0.2 PCC), whereas only 19% (21/108) showed negative correlation of less than 0.4. The statistical significance of such a correlation was analyzed and the correlation was found significant for these 108 genetic loci. Further, their distribution was skewed towards positive correlation in comparison to the correlation of the reference 9247 probes (Mann-Whitney U test, p-value 0.0016) (**Figure 5.4c**). Thus, these results indicated that the differential expression of these 97 genes was linked with differential H3K56 acetylation.

Figure 5.4 Association of H3K56Ac with differential transcription. **A)** Venn diagram showing significant overlaps between the two groups. One group represents top 10% differentially acetylated genetic loci/ probes and other group comprises the top 10% differentially expressed genes. **B)** Hierarchical clustering of these 108 probes which are differentially acetylated and of the corresponding 97 differentially expressed genes. **C)** Density plot of correlation between differential H3K56Ac and differential expression. Distributions of reference 9247 probes are shown in black color and of 108 probes are shown in red color.

PF14_0740, a *Plasmodium* exported protein (hyp17) and PF14_0271, a putative transcription factor containing an AP2 domain are the two representative examples having a high correlation between differential H3K56Ac and differential expression. The genetic loci which were found to be differentially marked by H3K56Ac and associated with the ON and OFF state of these two genes occur in 5' flanking region, whereas most of the other differentially H3K56 acetylated loci were localized at the 3' end of the gene (discussed in section 5.3.4). The H3K56Ac enrichment profiles (shown as bars) for both Dd2 and T996 are plotted on one

graph and similarly, the expression profiles (shown as line) for both the strains are plotted on the same graph. For both the genes, our results showed that the H3K56Ac enrichments were high in the earlier stages (TP1-TP3) and then decreased across the cycle in Dd2 (blue color bars), whereas enrichment of this mark markedly reduced at all the time points in the T996 (red color bars). The mRNA expression profiles were also identical for both the genes. The mRNA levels peaked during the early stages and then decreased across the cycle in Dd2 (blue color line), whereas both the genes were not expressed in T996 at all stages (orange-red color line) (**Figure 5.5a and 5.5b**). Pearson correlation between the differential H3K56Ac and differential expression is 0.896 and 0.806 for PF14_074 and PF14_0271, respectively.

5.3.4 Epigenetic acetylation marking of H3K56 over differentially expressed genes

H3K9Ac, H3K56Ac, H4K8Ac, H4K16Ac, H3K4me3, H4Ac4, H4K20me1 and H4K20me3 are eight histone modifications, whose dynamic association with chromatin correlates with transcription (64). Out of all transcription-linked histone modifications mentioned above, only H4K8Ac is predominantly enriched at the putative promoters and the others seven are found to be localized to varying degrees along the gene body with higher preference to the beginning of the coding sequence (133).

In order to determine the regions in the genes which were associated with the H3K56Ac, we investigated the distribution of H3K56Ac on the 108 probes, whose differential H3K56Ac was correlated with the differential expression. We found that most of these probes mapped within the coding region, with the peak of enrichment at the 3' end of the gene bodies (**Figure 5.6a**). Strikingly, in these genetic loci intergenic regions were comparatively deficient in H3K56Ac and instead it was mainly the loci close to the 3' of the differentially expressed genes that showed differential H3K56Ac occupancy. The presence or absence of the H3K56Ac on these differentially acetylated genetic loci may influence the transcriptional status of the corresponding genes.

Figure 5.5 The differentially enriched H3K56Ac profiles and differentially expressed profiles between two strains for PF14_0740 and PF14_0271. **A)** PF14_0740: On the left panel is the distribution of H3K56Ac over the differentially enriched locus across five time points and in the right panel is the expression profile of the gene across five time points in Dd2 (blue color) and in T996 (red color) **B)** PF14_0271: On the left panel is the distribution of H3K56Ac over the differentially enriched locus across five time points and in the right panel is the expression profile of the gene across five time points in Dd2 (blue color) and in T996 (red color). PCC between differential acetylation and differential expression is 0.896 and 0.806 for PF14_074 and PF14_0271, respectively.

Previously, it has been described that differential expression of erythrocyte binding antigen 165 (EBA-165) correlates with the differential H3K9/K14 acetylation in the 5' UTR regions. It was shown that EBA-165 is expressed in the Dd2-NM strain at 44 hpi of the IDC

cycle, where its expression is correlated with the H3K9/K14 acetylation marks. On the other hand, in the Dd2 strain, this gene is not expressed at 44 hpi, which coincides with the depletion of the acetylated histone marks (186). Here in our analysis, we found that a genetic locus within EBA-165 coding region was differentially acetylated for both the histone marks, H3K9Ac and H3K56Ac, and was located 3268 base pair downstream to the start codon, which is near to the 3' end of the gene. Similarly, we found that EBA-165 was expressed only in the later stages in T996 strain (TP5, which corresponds to 42 hpi) and its expression correlated with the H3K56/H3K9Ac enrichment along the genetic locus. On the contrary, EBA-165 was not expressed in Dd2 at all time points and the corresponding H3K56/H3K9Ac marks were markedly reduced compared to T996 (**Figure 5.6b and 5.6c**).

In our analysis, some of the interesting genes whose differential expression correlated with the differential H3K56Ac includes the putative histone deacetylase (PfHDAC1), ApiAp2 transcription factor, putative zinc finger containing proteins, putative ubiquitin protein ligase E3, putative cytoplasmic dynein intermediate chain, putative clathrin-adaptor medium chain and EBA-165. In the enriched dataset, there were ten exported proteins of unknown functions, some of them belonging to *hyp* and *phist* gene families that are recently discovered exported protein families. Most exported proteins are sub-telomeric in nature and have a characteristic PEXEL motif in their N-termini, which represents a hallmark for protein export in malaria (187). Malaria parasite export proteins to the erythrocyte surface in order to survive and promote virulence. The differential expression within *var* genes, an exported gene family encoding PfEMP1, has been known to contribute in malaria pathogenesis by immune evasion and cytoadherence. Similarly, invasion-related gene families (erythrocyte binding antigen, EBA family and the reticulocyte homology, PfRh family) have also been reported to exhibit differential gene expression (123).

Figure 5.6 *Distribution of H3K56 acetylation along the IGR and ORF of 108 oligos whose differential H3K56Ac correlates with differential expression. A)* Analysis of H3K56Ac enriched regions and their association with expression in Dd2 and T996. On y axis is the fraction of total number of oligos on chip which were positively correlated to transcription and x axis represents position in the genes. **B)** Expression profiles of EBA-165 across five time points are shown for Dd2 (blue color line) and T996 strain (red color line). **C)** Distribution of H3K9Ac and H3K56 acetylation over the same genetic locus within EBA-165 in Dd2 (blue bars) and in T996 strain (red bars).

5.4 Molecular characterization of a putative transcription factor, PF14_0271, which is found to be both differentially acetylated and differentially expressed between *P. falciparum* strains

Our genome wide association results identified that differential expression of PF14_0271 is epigenetically regulated by H3K56Ac. PF14_0271 is a putative transcription factor belonging to ApiAP2 family of transcription factors. The ApiAP2 proteins have a characteristic AP2-

integrase DNA binding domain, homologous to plant's AP2 containing transcription factors. In *Plasmodium*, there are 27 members in the AP2 family which are highly conserved across the *Plasmodium* species. The number of AP2 domains in *Plasmodium* proteins varies from a single domain to tandem AP2 domains. In general, each AP2 domain is consists of 60 amino acids. PF14_0271 has a single AP2 DNA binding domain (from amino acid 51 to 109) and encodes for 1186 amino acid protein (22). In order to elucidate the function of this uncharacterized transcription factor, we endogenously tagged the protein. With this endogenous tagging approach we identified stage specific expression of PF14_0271, and further we carried out genome-wide analysis of the tagged protein by performing ChIP-chip.

5.4.1 Endogenous tagging of AP2 domain containing transcription factor, PF14_0271

One of the important reasons for employing the endogenous tagging approach to study PF14_0271 over recombinant over-expression strategy was to characterize its functions under the influence of its native promoter and enhancer. The PF14_0271 protein was tagged at the 3' end with mCherry-HA epitope by single-crossover homologous recombination. The plasmid vector used for endogenous tagging was pTCherryHA vector. The targeting vector, pTCherryHA, consists of a drug selection cassette known as hDHFR. A 917 base pair region from the 3'end of PF14_0271 gene was cloned in frame into the pTCherryHA vector. This 917 bp sequence of PF14_0271 did not contain a stop codon and was followed by the mCherry-HA tag. The stop codon was present in the plasmid after mCherry-HA tag. The recombinant plasmid was then transfected into 3D7 and Dd2 parasites. The transfected parasites were grown under the WR99210 drug pressure and the parasites started to appear after 18 days of transfection. The plasmid was successfully integrated into the 3D7 and Dd2 parasites genome as verified by PCR shown in **Figure 5.7**.

Figure 5.7 Endogenous tagging of AP2 domain containing transcription factor PF14_0271.

A) The PF14_0271 protein was tagged at 3' end with mCherry-HA epitope by single crossover homologous recombination. The targeting vector, pTCherryHA, consists of a drug selection cassette, known as hDHFR. A 917 base pair region from the 3'end of PF14_0271 gene was cloned in frame into the pTCherryHA vector. This 917 bp sequence of PF14_0271 was without stop codon and was followed by mCherry-HA tag. The stop codon was present in the plasmid after mCherry-HA tag. B) PCR verified the integration in 3D7 and Dd2 transfected cell lines.

5.4.2 Endogenous protein detection and the intracellular localization of PF14_0271

After generating the endogenously tagged parasite cell line we wanted to characterize the endogenous protein levels of PF14_0271 in 3D7 and Dd2 parasites across the 48 hours intra-erythrocytic developmental cycle (IDC). To achieve this, we harvested protein samples from rings, trophozoites and schizonts stages from both 3D7-PF14_0271 and Dd2-PF14_0271 transfected cell lines. Immunoblot analysis was performed on these protein samples and the PF14_0271 endogenously tagged protein was probed using an anti-mCherry antibody. The expected protein band was of 170 KDa which consists of 140 KDa of PF14_0271 protein and 30 KDa of mCherry-HA tag. Our immunoblot results identified the single band of the desired molecular weight in the schizont stage of both 3D7 and Dd2 transfected cell lines (**Figure 5.8**). PF14_0271 was not observed in the ring and trophozoite stage in Dd2, but a faint protein band was identified in the ring stage of the 3D7 parasites.

To study the intracellular localization of PF14_0271, we performed immunofluorescence assays (IFA) of the PF14_0271 tagged transfected cell lines. Our immunoblot results showed that the PF14_0271 protein was expressed mostly in the schizont stage, hence we carried out IFA analysis only at the synchronized schizont stage. Similar to the western blot detection, IFA results showed the fluorescence signal of the PF14_0271 protein at the schizont stage. The fluorescence signal was observed in the nuclei of schizont parasites in both, 3D7 and Dd2, transfected cell lines (**Figure 5.9**). In our IFA analysis, we also tested the localization of H4K8Ac in the 3D7 transfected cell line as a control. The IFA results revealed that H4K8Ac was localized around the nuclear periphery, which is similar to the distribution shown in previous studies (188). Taken together, our result shows that PF14_0271 protein is most likely a schizont stage nuclear protein.

Figure 5.8 Endogenous protein detection of PF14_0271 in schizont stage. Protein samples from ring, trophozoite and schizont stage from both 3D7-PF14_0271 and Dd2-PF14_0271 transfected cell lines were harvested. Western blot using anti-mCherry antibody identified single band of desired molecular weight (170KDa) in schizont stages of both 3D7 and Dd2 transfected cell lines. PF14_0271 was not observed in rings and trophozoite stages in Dd2 but a faint protein band was identified in rings stage of 3D7 parasites.

5.4.3 Genome wide identification of PF14_0271 binding site

Our next objective was to identify the chromosomal binding sites of PF14_0271, transcription factor in the *P. falciparum*. Our western blot and localization results suggested that this protein is schizont specific and hence, we carried out ChIP-chip experiment in the synchronized schizont stage of the 3D7 transfected cell line. The localization of this mCherry-tagged PF14_0271 transcription factor (putative) in the nucleus of the schizont and

Figure 5.9 Endogenous intracellular localization of PF14_0271 in the schizont stage.

Immunofluorescence assays (IFA) were carried out at synchronized schizont stages using anti-mCherry antibody (Alexa Flour 594). The fluorescence signal was observed in the nuclei of schizont parasites in both 3D7 and Dd2 transfected cell lines. DAPI stain was used to stain nuclei as seen in blue color. Merge was the overlay of DAPI signal and signal from anti-mCherry antibody (Alexa Flour 594). The localization of H4K8ac in 3D7 transfected cell line was shown as a control.

being an AP2 protein which is the most important family of transcription factors in the parasite genome, suggested that this protein might play an essential role in transcriptional regulation. To identify regions bound by mCherry-HA tagged PF14_0271, cross-linked parasites were immunoprecipitated with anti-mCherry antibody. To rule out false-positive signal by anti-mCherry antibody, we carried out ChIP-chip on vector-control parasites, simultaneously. These vector-control parasites were transfected with empty mCherry-HA vector. To identify the PF14_0271 targets, we compared the log₂ ratio (IP/input) of the PF14_0271-tagged cell line with the log₂ ratio (IP/input) of the vector-control cell line. To ensure the statistical significance of results, only those genetic loci (represented by the

microarray probes) whose positive enrichment was consistent among three replicates (p-value < 0.01) were included into the analysis. Our analysis identified 642 genetic loci (corresponding to 583 genes) that were associated with the PF14_0271 binding. Further, we mapped the distribution of these 642 genetic loci with respect to their gene position. Out of 642, only 80 genetic loci were mapped in the intergenic regions, whereas remaining 562 loci were mapped within the gene body occurring from 5' start site to 3' end with varying degree of enrichment (**Figure 5.10a**). Further, we analyzed the distribution pattern of these enriched genetic loci with respect to their transcript abundance across the developmental cycle. We identified that the genetic loci associated with PF14_0271 mostly belonged to the rings and schizont specific genes. 34.5% of the detected genetic loci associated with the PF14_0271 protein were expressed in the schizont stage and 33% of the genetic loci were expressed in the rings stage. Some of the genes which were expressed in the trophozoites, merozoites and gametocytes were also found to be associated with the PF14_0271 binding (**Figure 5.10b**).

In the next step, we carried out functional enrichment analysis on these 642 genetic loci. Our analysis identified several functional groups that were associated with the PF14_0271 binding (**Figure 5.11**). The majority of the enriched functional groups were schizont specific genes, which belong to either cytoskeleton proteins such as centrosome proteins, actin filaments, molecular motor prototypes or DNA replication and repair such as homologous recombination. Some of the other enriched functional groups were metabolism related such as inositol phosphate metabolism, glutamate metabolism and ubiquinone metabolism. Few other interesting enriched groups were phosphoproteome, proteins containing histone post translational modification-binding modules and molecular markers of resistance to antimalarial drugs.

In summary, ChIP-chip results revealed that PF14_0271 binding occurred within both intergenic and coding regions of the targets genes in the parasite genome. The majority of the

genes which were found to be associated with the PF14_0271 binding, were rings and schizonts stage specific.

Figure 5.10 Genome wide identification of PF14_0271 binding site. A) To identify PF14_0271 targets, we compared log2 ratio (IP/input) of PF14_0271-tagged cell line with log2 ratio (IP/input) of vector-control cell line. On y axis is the fraction of total number of probes on chip whose positive enrichment were consistent among three replicates (p value < 0.01) and x axis represents position in the genes. **B)** Genome wide association of PF14_0271 binding in different stages of the IDC cycle. Here bar graph represents distribution of genetic loci which were associated with PF14_0271 binding in genes which are stage specific. For example, we identified that 34% of the detected PF14_0271 binds to genetic loci which are expressed in the schizont stage.

Figure 5.11 Functional enrichment analysis on these 642 genetic loci which were associated with PF14_0271. Over-representation of genes is determined by the comparison of the ratio of number of genes in each functional pathway to total number of genes, and the ratio of number of genes belonging to each functional group in the cluster to the total number of genes in the cluster. The ratios for each functional group with a p-value (binomial value) less than 0.05 were considered for the analysis. (*In cluster* is represented by the number of genes in pathway/total number of genes whereas *in genome* is represented by the number of genes in pathway and cluster/number of genes in cluster). Schizont specific genes are highlighted in the black box, which includes cytoskeleton proteins such as centrosome proteins, actin filaments, molecular motor prototypes or genes related to DNA replication and repair such as homologous recombination.

5.4.4 Transcriptional analysis of PF14_0271 knock-out parasites

In a large genome wide study in 2008, Maier *et al.* were able to disrupt 53 genes using double crossover recombination strategy (189). These genes were classified as non-essential for the parasite during the asexual RBC stage. In their study, out of the 53 genes that were successfully knocked out, PF14_0271 was also one of the candidate genes. The objective of

their study was to identify the parasite genes which were exported to the surface of infected RBC. We obtained the PF14_0271 knock-out cell line and the wild-type cell line from Alan Cowman and Alex Maier *et al.* (Walter and Eliza Hall Institute, Australia). The wild-type parasite line has a native endogenous copy of PF14_0271 intact. The CS2 strain of the parasite was used for transfection in the study conducted by the Maier *et al.* (189).

We collected six RNA samples after every eight hour interval covering the different developmental stages (ring, trophozoite and schizont) of the IDC cycle. We collected the RNA samples for both, PF14_0271 knock-out cell line and wild-type cell line and microarrays were performed on these RNA samples as described previously.

5.4.4.1 PF14_0271 knock-out has no effect on the global stage specific expression in the *P. falciparum*

Our objective was to characterize the differentially expressed genes in the absence of the AP2 containing transcription factor, PF14_0271. Those genes, whose expression could be altered in the absence of PF14_0271, would represent genes under the regulatory control of the AP2 transcription factor. Of the 5224 *Plasmodium* genes represented on the microarray, the IDC time-course expression profiles for 4214 genes were detected in both strains, which correspond to 80% of the *Plasmodium* genome. In our comparative study, all those genes belonging to the hyper-variable gene families such as *var*, *rifin*, *stevor*, etc. and the genes that showed DNA polymorphism were removed from the data analysis. The stage specific expression of genes in the IDC cycle was reported to be highly conserved among *Plasmodium* strains (116). When we compared the time-course of gene expression profiles between PF14_0271 knock-out and wild-type parasites using Pearson correlation coefficient (PCC), we found that about 3880 genes showed a positive correlation of more than 0.4. This indicates that 92% of the detected genes showed similar expression profiles between PF14_0271 knock-out and wild-type parasites. Only 61 genes showed negative correlation of

less than 0.4. This shows that only 1.4% of the detected genes exhibited distinct expression profiles (**Figure 5.12**). These results indicate that the timing of maximal abundance of a gene during the IDC is identical in both PF14_0271 knock-out and wild-type parasites. This suggests that PF14_0271 may not be involved in the transcriptional regulation of the stage-specific gene expression in the parasite.

Figure 5.12 *PF14_0271 knock out has no effect on global stage specific expression in P. falciparum.* **A)** Heat maps of wild-type and Pf14_0271-knock-out depicting the mRNA expression profiles of 4214 genes along the developmental IDC cycle (six time points). Red represents over expression of gene and green represents under expression of gene. Each row represents mRNA abundance of a single gene. The identical profiles between wild-type and PF14_0271 knock-out signify PF14_0271 plays no role in the transcriptional regulation of stage specific gene expression. **B)** Histogram showing the distribution of Pearson correlation between the gene expression profiles of wild-type and Pf14_0271-knock-out. Out of 4214 genes, 80% genes had the Pearson correlation of more than 0.4.

5.4.4.2 Identification of differentially expressed genes which were altered in PF14_0271 knock-out parasite cell line

In order to identify genes whose transcript abundances were altered due to the loss of function of PF14_0271, we compared the gene expression patterns across the IDC cycle for PF14_0271 knock-out parasites to those of wild-type parasites. With this approach, we identified 289 genes, which were differentially expressed across all six time points in PF14_0271 knock-out parasite cell line (p-value <0.05). Out of 289 genes, 138 genes were constantly up-regulated in all the six time points, whereas 151 genes were consistently down-regulated in all the time points (**Figure 5.13a**).

Functional enrichment analysis of these 289 differentially expressed genes revealed that up-regulated genes belonged to various pathways such as phosphoproteome of infected RBC, ubiquinone metabolism, other DNA binding proteins with AP2 domain. The pathways which were down-regulated comprised of following physiological functions: redox metabolism, post translational modifications SUMOylation, mRNA degradation and generation of polyphosphates. Functional group phosphoproteome of infected RBC was the most enriched group with lowest p-value < 0.00001 and consisted of 24 genes. Some of the interesting genes in this pathway included other transcriptional regulators such as the myb2 transcription factor (PF10_0327), transcription factor with AP2 domain(s) (PFF0670w), transcription factor, putative (PF10_0293), SNF2 helicase (PFF1185w), CCAAT-box DNA binding protein subunit B (PF11_0477), regulator of nonsense transcripts, (PF10_0057) and RNA-binding protein (PFD0775c) (**Figure 5.13b and 5.14a**). Our result suggests that these 24 genes of the functional group “phosphoproteome of infected RBC” could be potential targets of the PF14_0271. In other words, it proposes that functional knock-out of PF14_0271 leads to up-regulation of these 24 genes. Currently, with the data available, it is difficult to predict how PF14_0271 regulates the transcription of these up-regulated genes. Further, in order to

Figure 5.13 Identification of differentially expressed genes, which were altered in the PF14_0271 knock-out parasite cell line. **A)** To identify genes whose transcript abundances were altered due to the loss of function of PF14_0271, we compared the transcriptional profile across 6 time points for PF14_0271 knock-out to those of wild-type parasites shown as Venn diagram. With this approach, we identified 289 genes, which were differentially expressed across all the 6 time points in PF14_0271 knock-out parasites cell line (p value < 0.05). Out of 289 genes, 138 genes were consistently up-regulated in all 6 time points, whereas 151 genes were consistently down-regulated in all 6 time points as shown in the heat map. **B)** Functional enrichment analysis of 289 differentially expressed genes. Red bar indicates functional groups, which were up-regulated in knock-out parasites, whereas green bar indicates down-regulated genes. Phosphoproteome of infected RBC was the most enriched group with the lowest p value < 0.00001 (*).

identify that at which stage maximum number of genes would be affected due to PF14_0271 knockout. We compared the individual transcriptional levels of 6 time points for PF14_0271 knock-out parasites to those of wild-type parasites. (TP1 to TP6, where TP1 represents ring stage and TP6 represents schizont stage). Our analysis revealed that at the schizont stage (TP6), maximum numbers of genes were differentially expressed by >2-fold due to the loss of function of PF14_0271 (**Figure 5.14b**). About 330 genes were differentially expressed by >2-fold in knock-out parasites at the schizont stage. Out of 330 genes, 140 genes were up-regulated, whereas 190 genes were down-regulated in the knock-out parasites as compared to the wild-type parasites.

In summary, our western blot and localization results showed that the PF14_0271 protein is most likely a schizont stage specific protein. Further, genome wide analysis of the PF14_0271 binding site by ChIP-chip revealed that PF14_0271 showed maximal binding to the schizont, and ring stage specific genes. In addition, functional knock-out of PF14_0271 caused maximum transcriptional differences in the schizont stage. Overall, our results suggest that the PF14_0271, a transcription factor with an AP2 DNA binding domain, is a schizont specific transcriptional regulator.

5.5 Discussion

5.5.1 Unraveling the epigenetic landscape of differentially expressed genes

The life cycle of *P. falciparum* is very complex, which involves two hosts, the human and an arthropod vector and comprises of the three stages: mosquito stage, the human liver stage and the blood stage. There are several obstacles that the parasite must confront, which mainly include host immune response and drug pressure. Differential gene expression in *Plasmodium* plays a significant role in the adaptability of the parasite to the changing environment as well as providing the parasite with indispensable tools for survival and pathogenesis (118). Due to the dearth of recognized transcription factors, epigenetic mechanisms have been regarded as

Figure 5.14 Identification of differentially expressed genes altered in the PF14_0271 knock-out parasite cell line. A) Relative transcript abundance of 24 genes of phosphoproteome of infected RBCs that were consistently up-regulated in all the 6 time points in knock-out with respect to the wild-type. **B)** Distribution of differentially expressed genes due to PF14_0271 knock-out across the six RNA samples collected after every 8 hour interval covering the different developmental stages (Rings, trophozoite and schizont) of IDC cycle. Y-axis represents the number of genes, which were differentially expressed >2 fold and X-axis represents 6 time points or stages of the IDC. At schizont stage (TP6), maximum numbers of genes were differentially expressed by >2 fold due to the loss of function of PF14_0271.

one of the important transcriptional regulators in the parasite “just in time” gene expression pattern (23). Previously, epigenetic mechanisms have been shown to be associated with the variant gene expression in the sub-telomeric gene families such as *var* and *rifin* gene families (123). Additionally, chromatin-mediated variant gene expression has also been reported for genes encoding invasion-related proteins such as PfRH4 and clag 3 genes (87). Moreover, it has been demonstrated that many clonally variant gene expression families were associated with two heterochromatin marks, H3K9me3 and a heterochromatin protein, HP1 (118).

Here, we studied the epigenetic mechanisms which were associated with the genome-wide transcriptional variations between *P. falciparum* strains. Rather than studying the epigenetic regulation of variant gene expression in few target genes, we decided to understand the global epigenetic map responsible for the transcriptional variations in *P. falciparum*. Hence, first we identified genes which were differentially expressed between Dd2 and T996 using genome-wide transcriptional survey. To achieve this we used Euclidean distance as a tool to identify differences in the expression profiles across the IDC in context with the differences in the level of transcript abundance between Dd2 and T996. Using this approach we identified 434 genes, which were differentially expressed between the two strains (>2-fold).

After identifying differentially expressed genes, the next part of the project was to evaluate the differences in the nucleosome occupancy of histone modifications (H3K9Ac and H3K56Ac) in the two strains. These euchromatic marks were chosen for the investigation because the temporal changes in the occupancy of these PTMs correlated highly with the dynamic changes associated with the IDC transcriptome (64). Our main objective was to study the association between differential gene expression and nucleosome occupancy of euchromatic histone marks (H3K9Ac and H3K56Ac). In general, our results revealed that most of the transcriptional variations in *P. falciparum* strains were associated with H3K56Ac levels and not with the H3K9Ac levels. We identified 108 genetic loci (97 genes) in which

differences in gene expression correlated with the H3K56Ac levels. It has been reported that H3K9Ac and H3K56Ac are predominantly localized at the 5' start site of the active genes (64). But in our results, we found that most of these 108 genetic loci mapped within the coding region with the peak of enrichment at the 3' end of the gene bodies.

Much of our understanding of epigenetic-mediated variant gene expression in *P. falciparum* is limited to the sub-telomeric gene families and few other invasion-related ligands (87,178). In our results, we identified additional genes whose differential expression was associated with differential H3K56Ac occupancy. Some of the interesting genes were putative histone deacetylase (PF14_0690), two putative transcription factors (PF14_0271 and PFE0415w), a putative zinc finger containing protein (PF11_0357), a putative ubiquitin protein ligase E3, a putative cytoplasmic dynein intermediate chain, a putative clathrin-adaptor medium chain and the erythrocytic binding antigen-165 (EBA-165) and ten exported proteins of unknown functions. In summary, our results suggested that most of the transcriptional variations observed between Dd2 and T996 strains were associated with differential H3K56Ac.

5.5.2 Molecular characterization of a putative transcription factor which is both differentially acetylated and differentially expressed

Most of the genes expressed in the *Plasmodium* IDC cycle are highly regulated and are expressed only when their products are needed (9). However, the underlying mechanism of such a precise transcriptional regulation is not fully understood. The 27 members ApiAP2 family represent the most important transcriptional regulators of the *Plasmodium* genome. It has been shown that the members of the AP2 family are expressed across the different stages of the IDC cycle (22). Recently, some of the AP2 proteins were shown to play a major role in the developmental stages, for instance, AP2-L is required for the liver stage of parasite infection (190), AP2-O is required for the mosquito ookinete stage (191) and PfAP2-G

(PFL1085w) is required for the gametocyte formation (192). In addition, PFF0200c (PfSIP2) regulates the transcriptional silencing of *var* genes at the later developmental stages of the IDC cycle (193). In our study, we identified a putative ApiAP2 transcription factor, PF14_0271, whose differential expression was associated with differential H3K56Ac between *P. falciparum* strains.

In order to elucidate the functions of this uncharacterized transcription factor, we have endogenously tagged the protein at the 3' end with the mCherry-HA tag by a single crossover recombination approach. Our western blot and localization results showed that the PF14_0271 protein is most likely a schizont stage specific protein. The localization of the mCherry-HA tagged PF14_0271 putative transcription factor in the nucleus of the schizont stage suggests that this protein may play an essential role in the transcriptional regulation. Further, genome wide analysis of the PF14_0271 binding site by ChIP-chip revealed that the PF14_0271 binding occurred within both intergenic and coding regions of the target genes. Most of the genes which were found to be associated with the PF14_0271 binding were ring and schizont stage specific. Our analysis identified several functional groups, which were associated with the PF14_0271 binding. Majority of the enriched functional groups were schizont specific genes that belong to either cytoskeleton proteins such as centrosome proteins, actin filaments, molecular motor prototypes or DNA replication and repair proteins. In addition, PF14_0271 binding occurred in genes that belong to ubiquinone metabolism, phosphoproteome of *P. falciparum* infected RBCs and molecular markers of resistance to antimalarial drugs. Further, functional knock-out of the PF14_0271 resulted in the maximum transcriptional differences in the schizont stage.

Because the functional knock-out of PF14_0271 was performed in the CS2 strain and the genome-wide identification of the PF14_0271 binding site by ChIP-chip was done in the 3D7 strain, it was difficult to directly compare the two datasets. Nevertheless, it is important to

mention that the knock-out of the PF14_0271 resulted in the up-regulation of genes belonging to the phosphoproteome of infected RBC and ubiquinone metabolism. Moreover, ChIP-chip results of PF14_0271 binding also identified these two functional groups. It is tempting to speculate that the binding of the PF14_0271 results in the transcriptional repression of genes belonging to these two groups. Overall, our results suggest that the PF14_0271, a transcription factor with AP2 DNA binding domain, is a schizont specific transcriptional regulator.

Chapter 6

General Discussion and Conclusions

The DNA repair machinery as well as the overall DNA damage response in *P. falciparum* are believed to contribute to many crucial biological functions in the parasites such as antigenic variation and copy number variation (25,194). Hence, full characterization of these processes will be of great value for understanding the general biology of the parasite. Here, we demonstrated that when challenged with a DNA damaging agent, *P. falciparum* cells respond by broad transcriptional changes (21% of genome), which include up-regulation of both DSBR and excision repair pathways. Similar type of responses were reported in yeast where alkylating agents induced differential expression in 1/3 of the yeast genome (154). Like in yeast, in *P. falciparum* DNA damage leads to upregulation of genes involved in generic (indirect) stress responses such as protein synthesis and turnover as well as mitochondrial functions but also direct responses, involving genes of the DNA repair machinery. Many of the canonical DNA repair genes are conserved in the *Plasmodium* genome and in this study we are providing the transcriptional evidence of their functional involvement. This included upregulation of enzymes that are believed to sense DNA lesions and initiate specific repair pathways (146). In *P. falciparum*, up-regulated enzymes include the homologues of A/G specific adenine glycosylases, XPC-h23B, mutS heterodimer (MSH2-MSH6) and MRN exonuclease complex (MRE11-Rad50) that correspond to DNA damage sensors for BER, NER, MMR and DSBR repair, respectively (**Table 3.4**). MMS has also been shown to induce overexpression of three *P. falciparum* DSBR repair genes, PfRAD51, PfRAD54 and PfRPA1 (52). Here, we demonstrated that the MMS-mediated DNA damage resulted in induction of the entire DNA repair machinery present in the *Plasmodium* genome (**Figure 6.1**). Although a single alkylating agent can generate different types of DNA lesions, each DNA lesion is recognized and repaired by a distinct pathway (195). The two most common DNA adducts generated by MMS are N7-methyl guanine (7 me G) and O6-methyl guanine (O6 me G) (139). N-alkyl lesions are mostly repaired by BER, NER and direct repair enzyme AlkB

Figure 6.1 Overview of the *P. falciparum* DNA damage pathways and their transcriptional properties under MMS-induced stress. The initial N- and O-alkyl lesions of *P. falciparum* DNA caused by MMS leads initially to transcriptional induction of nucleotide excision repair (NER), base excision repair (BER) and mismatch repair (MMR). The subsequent induction of the double strand break repair (DSBR), namely the homologous recombination repair (HR), suggests that the MMS induced O-alkyl lesions was not be fully repaired by MMR. The transcriptional induction encompasses the majority (if not all) components of the *P. falciparum* DNA damage repair. Other components of eukaryotic DNA repair pathways such as nonhomologous end joining (NHEJ), or enzymes of direct repair (ALkBH and MGMT) are not present in the *P. falciparum* genome (dashed line outline). Although extensive DSBs could lead to cell death, six hour treatment with 0.05% MMS was likely counteracted by the DNA repair factors and did not lead to upregulation of putative factors of *P. falciparum* cell death. *** indicates statistically significant functional enrichment of genes of particular pathway amongst genes up-regulated as a result of MMS-induced DNA damage. This outline is based on bioinformatics sequence analysis and the transcriptional results presented in this study.

homologue, whereas O-alkyl lesions are commonly repaired by MMR and a distinct direct repair enzyme known as MGMT (30,195). Particularly, O-6meG has higher tendency to form mispairs with thymine following replication (30). The O-6 meG:T mispairs are recognized by mutS (MSH2-MSH6), which leads to activation of the MMR pathway (196). If O6-meG:T mispairs remain unrepaired it leads to a futile cycle, which results in the collapse of the replication fork and subsequently induction of double-strand breaks (DSBs) (139). This leads either to apoptosis (if the damage is too extensive) or the DSBs are repaired by either HR/NHEJ pathway (30,139,195). Our results demonstrate that in *Plasmodium*, MMS treatment induces excision repair (NER, BER, and MMR) and DSBR (HR). As expected, NHEJ and/or several unique direct repair enzymes (MGMT, AlkBH) are not involved in repair mechanisms as most of their components are missing in the *Plasmodium* genome (7,197).

Our results also illustrate that MMS mediated DNA damage leads to significant change in histone modifications, specifically H4K8Ac, H3K9Ac and H3K56Ac. In yeast and mammals, DNA damage response is initiated by hyperacetylation of H4 by the MYST histone acetyltransferase (MYST-HAT) (82). It was also shown that mutations in any of the four H4 lysine residues impede DNA repair in yeast (92). In agreement with this, the *Plasmodium* homologue PfMYST-HAT acetylates H4K5, 8, 12 and 16 and its overexpression enhance protection from DNA damaging agents, MMS and camptothecin (74). This suggests that the role of H4 in DNA damage is evolutionary conserved in the malaria parasite. The ChIP-chip results revealed that MMS induced H3K9Ac and H4K8Ac were predominantly enriched in the intergenic regions and possibly associated with transcription (**Figure 3.8a**). This is in contrast with histone deacetylase inhibitors, apicidin and TSA, that show very limited positive correlation between gene expression and altered acetylation induced by their effect in *P. falciparum* as well as mammals (63,156). In particular, MMS-induced H4K8Ac occurred

at the transcriptionally up-regulated genes that constitute a putative generic response to DNA damage (protein turnover, mitochondria). Conversely, the increase of H3K9Ac occurred preferentially at the DNA repair genes that were also induced as a result of MMS-mediated DNA damage. Under normal growth conditions, H3K9Ac and H4K8Ac are selectively enriched on actively transcribed genes belonging to metabolism and ribosome assembly related functions in *P. falciparum* (64). In the future, it will be important to investigate the molecular mechanism of H4K8Ac and H3K9Ac dynamics in transcriptional up-regulation during DNA repair.

Artemisinin appears to induce DNA damage in *P. falciparum* presumably as a result of oxidative stress generated by the drug (157). This is analogous to the effect in cancer cells where artemisinin engages DSBs (113,114). Here, we observed that at a physiologically relevant dose (1 μ M), artemisinin elicits similar transcriptional and epigenetic response that was induced by MMS. Importantly, both agents induced transcriptional up-regulation of parasite excision repair and DSBs. Additionally, we observed that both agents could induce expression of genes imported to the mitochondria/apicoplast. This is consistent with the results in yeast where DNA damage results in import of different proteins such as pro-apoptotic proteins (p53), several DNA repair and cell cycle regulatory proteins into mitochondria from the nucleus, (158,159). Previously, Natalang *et al.* have shown that three hour exposure to 780 nM of artemisinin treatment at the trophozoite stage altered transcript abundance of 398 genes, none of which is involved in DNA repair (152). Here, we showed that six hour treatment of *P. falciparum* cells by both MMS and artemisinin is needed before gene expression and histone modification characteristics of DNA repair are detectable. Only then, the DNA repair machinery is upregulated together with another ~ 1,500 genes (P-value < 0.05). In natural infections during treatment, artemisinin reaches up to 1 μ M concentrations in ~40 minutes and then it is rapidly eliminated with a half-life of ~20 minutes (198). This

pharmacokinetic profile may or may not provide sufficient time for artemisinin to cause DNA damage. Intriguingly, the kelch-13 protein (PF3D7_1343700) that has been associated with artemisinin resistance (153) was found to be upregulated in our artemisinin perturbation experiments (P-value 0.0004). Also, several factors of the DNA repair machinery have been linked with artemisinin resistance *in vivo* and *in vitro* (61,160). The results here open a plausible suggestion that DNA damage contributes to the “real-life” artemisinin mode of action and with that, the artemisinin resistance phenotype that is presently spreading through Southeast Asia (153,199,200).

Defects in DNA repair machinery were demonstrated to play significant roles in the mutator phenotypes in various organisms such as yeast, mammals, bacteria, *T. brucei* and *T. gondii* (43,57,58,173,174). Here, we showed that the ARMD (mutator) phenotype of *P. falciparum* might be underlined by differential transcriptional regulation of genes involved in the DNA repair machinery when the parasites were under the DNA damaging stress. In particular, MMS was able to induce expression of both MMR and NER repair machinery genes in the non-ARMD parasites, but failed to up-regulate these genes in the ARMD parasites. This is consistent with the previous results, which suggest that defective DNA repair might contribute to the emergence of the ARMD parasites in Southeast Asia (38,60). Although in our experiments, we did not link this phenomenon with (any) changes in histone modifications, further careful investigations are required to confirm if the ARMD phenotype is also involved in epigenetic factors. Additionally, polymorphisms in 18 DNA repair genes were identified, which might lead to impaired DNA repair functions. This suggestion is based on the fact that point mutations in mismatch repair genes lead to defective DNA repair that confers a mutator phenotype in yeast and bacteria (43,173). Further investigations are needed to validate the role of these mutations in the development of the mutator phenotype in the malaria parasites.

Lastly, in chapter 5, we identified that most of the transcriptional differences between *P. falciparum* strains were associated with H3K56Ac. Further, we characterized one of the AP2 domains containing putative transcription factor PF14_0271, which was found to be both differentially acetylated and differentially expressed between *P. falciparum* strains. By endogenously tagging method, we showed that the PF14_0271 is a schizont stage specific protein and localized to the nucleus. Further, genome-wide analysis of the PF14_0271 binding site by ChIP-chip revealed that it showed maximal binding to the schizont and ring stage specific genes. In addition, the functional knock-out of PF14_0271 induced maximum transcriptional differences in the schizont stage. Overall, our results suggest that the PF14_0271, a transcription factor with an AP2 DNA binding domain is a schizont stage-specific transcriptional regulator.

Chapter 7

References

1. Hempelmann, E. and Krafts, K. (2013) Bad air, amulets and mosquitoes: 2,000years of changing perspectives on malaria. *Malaria Journal*, **12**, 232.
2. Cox, F. (2010) History of the discovery of the malaria parasites and their vectors. *Parasites & Vectors*, **3**, 5.
3. Qari, S.H., Shi, Y.P., Pieniazek, N.J., Collins, W.E. and Lal, A.A. (1996) Phylogenetic Relationship among the Malaria Parasites Based on Small Subunit rRNA Gene Sequences: Monophyletic Nature of the Human Malaria Parasite, *Plasmodium falciparum*. *Molecular Phylogenetics and Evolution*, **6**, 157-165.
4. WHO. (2012). Geneva: World Health Organization.
5. Vaughan, A.M., Aly, A.S.I. and Kappe, S.H.I. Malaria Parasite Pre-Erythrocytic Stage Infection: Gliding and Hiding. *Cell Host & Microbe*, **4**, 209-218.
6. Cowman, A.F. and Crabb, B.S. (2006) Invasion of Red Blood Cells by Malaria Parasites. *Cell*, **124**, 755-766.
7. 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.
8. Baum, J., Papenfuss, A.T., Mair, G.R., Janse, C.J., Vlachou, D., Waters, A.P., Cowman, A.F., Crabb, B.S. and de Koning-Ward, T.F. (2009) Molecular genetics and comparative genomics reveal RNAi is not functional in malaria parasites. *Nucleic Acids Research*, **37**, 3788-3798.
9. Bozdech, Z., Llinas, M., Pulliam, B., Wong, E., Zhu, J. and DeRisi, J. (2003) The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS Biol*, **1**, E5.
10. Tarun, A.S., Peng, X., Dumpit, R.F., Ogata, Y., Silva-Rivera, H., Camargo, N., Daly, T.M., Bergman, L.W. and Kappe, S.H.I. (2008) A combined transcriptome and proteome survey of malaria parasite liver stages. *Proceedings of the National Academy of Sciences*, **105**, 305-310.
11. Young, J.A., Fivelman, Q.L., Blair, P.L., de la Vega, P., Le Roch, K.G., Zhou, Y., Carucci, D.J., Baker, D.A. and Winzeler, E.A. (2005) The *Plasmodium falciparum* sexual development transcriptome: A microarray analysis using ontology-based pattern identification. *Molecular and Biochemical Parasitology*, **143**, 67-79.
12. Florens, L., Washburn, M.P., Raine, J.D., Anthony, R.M., Grainger, M., Haynes, J.D., Moch, J.K., Muster, N., Sacci, J.B., Tabb, D.L. *et al.* (2002) A proteomic view of the *Plasmodium falciparum* life cycle. *Nature*, **419**, 520-526.

13. Le Roch, K.G., Johnson, J.R., Florens, L., Zhou, Y., Santrosyan, A., Grainger, M., Yan, S.F., Williamson, K.C., Holder, A.A., Carucci, D.J. *et al.* (2004) Global analysis of transcript and protein levels across the *Plasmodium falciparum* life cycle. *Genome Research*, **14**, 2308-2318.
14. Foth, B.J., Zhang, N., Chahal, B.K., Sze, S.K., Preiser, P.R. and Bozdech, Z. (2011) Quantitative time-course profiling of parasite and host cell proteins in the human malaria parasite *Plasmodium falciparum*. *Molecular & Cellular Proteomics*.
15. Greenbaum, D., Colangelo, C., Williams, K. and Gerstein, M. (2003) Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biology*, **4**, 117.
16. Griffin, T.J., Gygi, S.P., Ideker, T., Rist, B., Eng, J., Hood, L. and Aebersold, R. (2002) Complementary Profiling of Gene Expression at the Transcriptome and Proteome Levels in *Saccharomyces cerevisiae*. *Molecular & Cellular Proteomics*, **1**, 323-333.
17. Horrocks, P. and Lanzer, M. (1999) Differences in nucleosome organization over episomally located plasmids coincides with aberrant promoter activity in *P. falciparum*. *Parasitology International*, **48**, 55-61.
18. Coleman, B.I. and Duraisingh, M.T. (2008) Transcriptional control and gene silencing in *Plasmodium falciparum*. *Cellular Microbiology*, **10**, 1935-1946.
19. Elemento, O., Slonim, N. and Tavazoie, S. (2007) A Universal Framework for Regulatory Element Discovery across All Genomes and Data Types. *Molecular cell*, **28**, 337-350.
20. Coulson, R.M.R., Hall, N. and Ouzounis, C.A. (2004) Comparative Genomics of Transcriptional Control in the Human Malaria Parasite *Plasmodium falciparum*. *Genome Research*, **14**, 1548-1554.
21. Gissot, M., Briquet, S., Refour, P., Boschet, C. and Vaquero, C. (2005) PfMyb1, a *Plasmodium falciparum* Transcription Factor, is Required for Intra-erythrocytic Growth and Controls Key Genes for Cell Cycle Regulation. *Journal of Molecular Biology*, **346**, 29-42.
22. Balaji, S., Babu, M.M., Iyer, L.M. and Aravind, L. (2005) Discovery of the principal specific transcription factors of Apicomplexa and their implication for the evolution of the AP2-integrase DNA binding domains. *Nucleic Acids Research*, **33**, 3994-4006.
23. Cui, L. and Miao, J. (2010) Chromatin-mediated Epigenetic Regulation in the Malaria Parasite *Plasmodium falciparum*. *Eukaryotic Cell*.
24. Peterson, C.L. and Côté, J. (2004) Cellular machineries for chromosomal DNA repair. *Genes & Development*, **18**, 602-616.

25. Lee, A.H., Symington, L.S. and Fidock, D.A. (2014) DNA Repair Mechanisms and Their Biological Roles in the Malaria Parasite *Plasmodium falciparum*. *Microbiology and Molecular Biology Reviews*, **78**, 469-486.
26. O'Neill, P.M., Barton, V.E. and Ward, S.A. (2010) The Molecular Mechanism of Action of Artemisinin—The Debate Continues. *Molecules*, **15**, 1705-1721.
27. Ciccia, A. and Elledge, S.J. (2010) The DNA Damage Response: Making it safe to play with knives. *Molecular cell*, **40**, 179-204.
28. Ataian, Y. and Krebs, J.E. (2006) Five repair pathways in one context: chromatin modification during DNA repair This paper is one of a selection of papers published in this Special Issue, entitled 27th International West Coast Chromatin and Chromosome Conference, and has undergone the Journal's usual peer review process. *Biochemistry and Cell Biology*, **84**, 490-494.
29. Samson, L., Derfler, B. and Waldstein, E.A. (1986) Suppression of human DNA alkylation-repair defects by *Escherichia coli* DNA-repair genes. *Proceedings of the National Academy of Sciences*, **83**, 5607-5610.
30. Fu, D., Calvo, J.A. and Samson, L.D. (2012) Balancing repair and tolerance of DNA damage caused by alkylating agents. *Nat Rev Cancer*, **12**, 104-120.
31. de Laat, W.L., Jaspers, N.G.J. and Hoeijmakers, J.H.J. (1999) Molecular mechanism of nucleotide excision repair. *Genes & Development*, **13**, 768-785.
32. Hanawalt, P. (2002) Subpathways of nucleotide excision repair and their regulation. *Oncogene*, **21**, 8949-8956.
33. Nilsen, L., Forstrøm, R.J., Bjørås, M. and Alseth, I. (2012) AP endonuclease independent repair of abasic sites in *Schizosaccharomyces pombe*. *Nucleic Acids Research*, **40**, 2000-2009.
34. Haltiwanger, B.M., Matsumoto, Y., Nicolas, E., Dianov, G.L., Bohr, V.A. and Taraschi, T.F. (2000) DNA Base Excision Repair in Human Malaria Parasites Is Predominantly by a Long-Patch Pathway†. *Biochemistry*, **39**, 763-772.
35. Robertson, A.B., Klungland, A., Rognes, T. and Leiros, I. (2009) DNA Repair in Mammalian Cells. *Cellular and Molecular Life Sciences*, **66**, 981-993.
36. Suksangpleng, T., Leartsakulpanich, U., Moonsom, S., Siribal, S., Boonyuen, U., Wright, G. and Chavalitsheewinkoon-Petmitr, P. (2014) Molecular characterization of *Plasmodium falciparum* uracil-DNA glycosylase and its potential as a new anti-malarial drug target. *Malaria Journal*, **13**, 149.
37. Marteijn, J.A., Lans, H., Vermeulen, W. and Hoeijmakers, J.H.J. (2014) Understanding nucleotide excision repair and its roles in cancer and ageing. *Nat Rev Mol Cell Biol*, **15**, 465-481.

38. Trotta, R.F., Brown, M.L., Terrell, J.C. and Geyer, J.A. (2004) Defective DNA Repair as a Potential Mechanism for the Rapid Development of Drug Resistance in *Plasmodium falciparum*†. *Biochemistry*, **43**, 4885-4891.
39. Marti, T.M., Kunz, C. and Fleck, O. (2002) DNA mismatch repair and mutation avoidance pathways. *Journal of Cellular Physiology*, **191**, 28-41.
40. Jiricny, J. (2006) The multifaceted mismatch-repair system. *Nat Rev Mol Cell Biol*, **7**, 335-346.
41. Karran, P. (2001) Mechanisms of tolerance to DNA damaging therapeutic drugs. *Carcinogenesis*, **22**, 1931-1937.
42. Shah, S.N., Hile, S.E. and Eckert, K.A. (2010) Defective Mismatch Repair, Microsatellite Mutation Bias, and Variability in Clinical Cancer Phenotypes. *Cancer Research*, **70**, 431-435.
43. Chopra, I., O'Neill, A.J. and Miller, K. The role of mutators in the emergence of antibiotic-resistant bacteria. *Drug Resistance Updates*, **6**, 137-145.
44. Bethke, L., Thomas, S., Walker, K., Lakhia, R., Rangarajan, R. and Wirth, D. (2007) The role of DNA mismatch repair in generating genetic diversity and drug resistance in malaria parasites. *Molecular and Biochemical Parasitology*, **155**, 18-25.
45. Denver, D.R., Feinberg, S., Estes, S., Thomas, W.K. and Lynch, M. (2005) Mutation Rates, Spectra and Hotspots in Mismatch Repair-Deficient *Caenorhabditis elegans*. *Genetics*, **170**, 107-113.
46. Tarique, M., Satsangi, A.T., Ahmad, M., Singh, S. and Tuteja, R. (2012) *Plasmodium falciparum* MLH is schizont stage specific endonuclease. *Molecular and Biochemical Parasitology*, **181**, 153-161.
47. Mehta, A. and Haber, J.E. (2014) Sources of DNA Double-Strand Breaks and Models of Recombinational DNA Repair. *Cold Spring Harbor Perspectives in Biology*.
48. Vilenchik, M.M. and Knudson, A.G. (2003) Endogenous DNA double-strand breaks: Production, fidelity of repair, and induction of cancer. *Proceedings of the National Academy of Sciences*, **100**, 12871-12876.
49. Kirkman, L.A., Lawrence, E.A. and Deitsch, K.W. (2014) Malaria parasites utilize both homologous recombination and alternative end joining pathways to maintain genome integrity. *Nucleic Acids Research*, **42**, 370-379.
50. You, Z., Chahwan, C., Bailis, J., Hunter, T. and Russell, P. (2005) ATM Activation and Its Recruitment to Damaged DNA Require Binding to the C Terminus of Nbs1. *Molecular and Cellular Biology*, **25**, 5363-5379.
51. Li, X. and Heyer, W.-D. (2008) Homologous recombination in DNA repair and DNA damage tolerance. *Cell research*, **18**, 99-113.

52. Gopalakrishnan, A.M. and Kumar, N. (2013) Opposing Roles for Two Molecular Forms of Replication Protein A in Rad51-Rad54-Mediated DNA Recombination in *Plasmodium falciparum*. *mBio*, **4**.
53. Bhattacharyya, M.K. and Kumar, N. (2003) Identification and molecular characterisation of DNA damaging agent induced expression of *Plasmodium falciparum* recombination protein PfRad51. *International Journal for Parasitology*, **33**, 1385-1392.
54. Roy, N., Bhattacharyya, S., Chakrabarty, S., Laskar, S., Babu, S.M. and Bhattacharyya, M.K. (2014) Dominant negative mutant of *Plasmodium* Rad51 causes reduced parasite burden in host by abrogating DNA double-strand break repair. *Molecular Microbiology*, **94**, 353-366.
55. Downs, J.A., Nussenzweig, M.C. and Nussenzweig, A. (2007) Chromatin dynamics and the preservation of genetic information. *Nature*, **447**, 951-958.
56. Fink, D., Aebi, S. and Howell, S.B. (1998) The role of DNA mismatch repair in drug resistance. *Clinical Cancer Research*, **4**, 1-6.
57. Augusto-Pinto, L., Teixeira, S.M.R., Pena, S.D.J. and Machado, C.R. (2003) Single-Nucleotide Polymorphisms of the *Trypanosoma cruzi* MSH2 Gene Support the Existence of Three Phylogenetic Lineages Presenting Differences in Mismatch-Repair Efficiency. *Genetics*, **164**, 117-126.
58. Garrison, E.M. and Arrizabalaga, G. (2009) Disruption of a mitochondrial MutS DNA repair enzyme homologue confers drug resistance in the parasite *Toxoplasma gondii*. *Molecular Microbiology*, **72**, 425-441.
59. Rathod, P.K., McErlean, T. and Lee, P.-C. (1997) Variations in frequencies of drug resistance in *Plasmodium falciparum*. *Proceedings of the National Academy of Sciences*, **94**, 9389-9393.
60. Castellini, M.A., Buguliskis, J.S., Casta, L.J., Butz, C.E., Clark, A.B., Kunkel, T.A. and Taraschi, T.F. (2011) Malaria drug resistance is associated with defective DNA mismatch repair. *Molecular and Biochemical Parasitology*, **177**, 143-147.
61. Miotto, O., Almagro-Garcia, J., Manske, M., MacInnis, B., Campino, S., Rockett, K.A., Amaratunga, C., Lim, P., Suon, S., Sreng, S. *et al.* (2013) Multiple populations of artemisinin-resistant *Plasmodium falciparum* in Cambodia. *Nat Genet*, **45**, 648-655.
62. Otto, T.D., Wilinski, D., Assefa, S., Keane, T.M., Sarry, L.R., Böhme, U., Lemieux, J., Barrell, B., Pain, A., Berriman, M. *et al.* (2010) New insights into the blood-stage transcriptome of *Plasmodium falciparum* using RNA-Seq. *Molecular Microbiology*, **76**, 12-24.

63. Chaal, B.K., Gupta, A.P., Wastuwidyaningtyas, B.D., Luah, Y.-H. and Bozdech, Z. (2010) Histone Deacetylases Play a Major Role in the Transcriptional Regulation of the *Plasmodium falciparum* Life Cycle. *PLoS Pathog*, **6**, e1000737.
64. Gupta, A.P., Chin, W.H., Zhu, L., Mok, S., Luah, Y.-H., Lim, E.-H. and Bozdech, Z. (2013) Dynamic Epigenetic Regulation of Gene Expression during the Life Cycle of Malaria Parasite *Plasmodium falciparum*. *PLoS Pathog*, **9**, e1003170.
65. Ponts, N., Harris, E.Y., Prudhomme, J., Wick, I., Eckhardt-Ludka, C., Hicks, G.R., Hardiman, G., Lonardi, S. and Le Roch, K.G. (2010) Nucleosome landscape and control of transcription in the human malaria parasite. *Genome Research*, **20**, 228-238.
66. Ponts, N., Harris, E.Y., Lonardi, S. and Le Roch, K.G. (2011) Nucleosome occupancy at transcription start sites in the human malaria parasite: A hard-wired evolution of virulence? *Infection, Genetics and Evolution*, **11**, 716-724.
67. Westenberger, S., Cui, L., Dharia, N., Winzeler, E. and Cui, L. (2009) Genome-wide nucleosome mapping of *Plasmodium falciparum* reveals histone-rich coding and histone-poor intergenic regions and chromatin remodeling of core and subtelomeric genes. *BMC Genomics*, **10**, 610.
68. Miao, J., Fan, Q., Cui, L., Li, J., Li, J. and Cui, L. (2006) The malaria parasite *Plasmodium falciparum* histones: Organization, expression, and acetylation. *Gene*, **369**, 53-65.
69. Cui, L. and Miao, J. (2010) Chromatin-Mediated Epigenetic Regulation in the Malaria Parasite *Plasmodium falciparum*. *Eukaryotic Cell*, **9**, 1138-1149.
70. Jaenisch, R. and Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet*.
71. Bártfai, R., Hoeijmakers, W.A.M., Salcedo-Amaya, A.M., Smits, A.H., Janssen-Megens, E., Kaan, A., Treeck, M., Gilberger, T.-W., François, K.-J. and Stunnenberg, H.G. (2010) H2A.Z Demarcates Intergenic Regions of the *Plasmodium falciparum* Epigenome That Are Dynamically Marked by H3K9ac and H3K4me3. *PLoS Pathog*, **6**, e1001223.
72. Salcedo-Amaya, A.M., van Driel, M.A., Alako, B.T., Trelle, M.B., van den Elzen, A.M.G., Cohen, A.M., Janssen-Megens, E.M., van de Vegte-Bolmer, M., Selzer, R.R., Iniguez, A.L. *et al.* (2009) Dynamic histone H3 epigenome marking during the intraerythrocytic cycle of *Plasmodium falciparum*. *Proceedings of the National Academy of Sciences*, **106**, 9655-9660.
73. Strahl, B.D. and Allis, C.D. (2000) The language of covalent histone modifications. *Nature*, **403**, 41-45.
74. Miao, J., Fan, Q., Cui, L., Li, X., Wang, H., Ning, G., Reese, J.C. and Cui, L. (2010) The MYST Family Histone Acetyltransferase Regulates Gene Expression and Cell

- Cycle in Malaria Parasite *Plasmodium falciparum*. *Molecular microbiology*, **78**, 883-902.
75. Duraisingh, M.T., Voss, T.S., Marty, A.J., Duffy, M.F., Good, R.T., Thompson, J.K., Freitas-Junior, L.H., Scherf, A., Crabb, B.S. and Cowman, A.F. (2005) Heterochromatin Silencing and Locus Repositioning Linked to Regulation of Virulence Genes in *Plasmodium falciparum*. *Cell*, **121**, 13-24.
 76. Cui, L., Fan, Q., Cui, L. and Miao, J. (2008) Histone lysine methyltransferases and demethylases in *Plasmodium falciparum*. *International Journal for Parasitology*, **38**, 1083-1097.
 77. van Attikum, H. and Gasser, S.M. Crosstalk between histone modifications during the DNA damage response. *Trends in Cell Biology*, **19**, 207-217.
 78. Dinant, C., Houtsmuller, A. and Vermeulen, W. (2008) Chromatin structure and DNA damage repair. *Epigenetics & Chromatin*, **1**, 9.
 79. Sun, Y., Jiang, X. and Price, B.D. (2010) Tip60: Connecting chromatin to DNA damage signaling. *Cell cycle (Georgetown, Tex.)*, **9**, 930-936.
 80. Altaf, M., Saksouk, N. and Côté, J. (2007) Histone modifications in response to DNA damage. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, **618**, 81-90.
 81. Rossetto, D., Truman, A.W., Kron, S.J. and Côté, J. (2010) Epigenetic Modifications in Double-Strand Break DNA Damage Signaling and Repair. *Clinical Cancer Research*, **16**, 4543-4552.
 82. Xu, Y. and Price, B.D. (2011) Chromatin dynamics and the repair of DNA double strand breaks. *Cell Cycle*, **10**, 261-267.
 83. Shi, L. and Oberdoerffer, P. (2012) Chromatin dynamics in DNA double-strand break repair. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, **1819**, 811-819.
 84. Ahmad, M. and Tuteja, R. (2012) *Plasmodium falciparum* RuvB proteins: Emerging importance and expectations beyond cell cycle progression. *Communicative & Integrative Biology*, **5**, 350-361.
 85. Foster, E.R. and Downs, J.A. (2005) Histone H2A phosphorylation in DNA double-strand break repair. *FEBS Journal*, **272**, 3231-3240.
 86. Pandita, T.K., Kumar, R., Horikoshi, N., Singh, M., Gupta, A., Misra, H.S., Albuquerque, K. and Hunt, C.R. (2013) Chromatin Modifications and the DNA Damage Response to Ionizing Radiation. *Frontiers in Oncology*, **2**.
 87. Duffy, M.F., Selvarajah, S.A., Josling, G.A. and Petter, M. (2013) Epigenetic regulation of the *Plasmodium falciparum* genome. *Briefings in Functional Genomics*.

88. Durocher, D. and Jackson, S.P. (2001) DNA-PK, ATM and ATR as sensors of DNA damage: variations on a theme? *Current Opinion in Cell Biology*, **13**, 225-231.
89. Redon, C., Pilch, D., Rogakou, E., Sedelnikova, O., Newrock, K. and Bonner, W. (2002) Histone H2A variants H2AX and H2AZ. *Current Opinion in Genetics & Development*, **12**, 162-169.
90. Dalmaso, M.C., Onyango, D.O., Naguleswaran, A., Sullivan, W.J. and Angel, S.O. (2009) Toxoplasma H2A Variants Reveal Novel Insights into Nucleosome Composition and Functions for this Histone Family. *Journal of molecular biology*, **392**, 33-47.
91. Collins, M.O., Wright, J.C., Jones, M., Rayner, J.C. and Choudhary, J.S. (2014) Confident and sensitive phosphoproteomics using combinations of collision induced dissociation and electron transfer dissociation. *Journal of Proteomics*, **103**, 1-14.
92. Bird, A.W., Yu, D.Y., Pray-Grant, M.G., Qiu, Q., Harmon, K.E., Megee, P.C., Grant, P.A., Smith, M.M. and Christman, M.F. (2002) Acetylation of histone H4 by Esa1 is required for DNA double-strand break repair. *Nature*, **419**, 411-415.
93. Li, X., Corsa, C.A.S., Pan, P.W., Wu, L., Ferguson, D., Yu, X., Min, J. and Dou, Y. (2010) MOF and H4 K16 Acetylation Play Important Roles in DNA Damage Repair by Modulating Recruitment of DNA Damage Repair Protein Mdc1. *Molecular and Cellular Biology*, **30**, 5335-5347.
94. Vonlaufen, N., Naguleswaran, A., Coppens, I. and Sullivan, W.J. (2010) MYST Family Lysine Acetyltransferase Facilitates Ataxia Telangiectasia Mutated (ATM) Kinase-mediated DNA Damage Response in Toxoplasma gondii. *The Journal of Biological Chemistry*, **285**, 11154-11161.
95. Das, C., Lucia, M.S., Hansen, K.C. and Tyler, J.K. (2009) CBP/p300-mediated acetylation of histone H3 on lysine 56. *Nature*, **459**, 113-117.
96. Vempati, R.K., Jayani, R.S., Notani, D., Sengupta, A., Galande, S. and Haldar, D. (2010) p300-mediated Acetylation of Histone H3 Lysine 56 Functions in DNA Damage Response in Mammals. *Journal of Biological Chemistry*, **285**, 28553-28564.
97. Wurtele, H., Kaiser, G.S., Bacal, J., St-Hilaire, E., Lee, E.-H., Tsao, S., Dorn, J., Maddox, P., Lisby, M., Pasero, P. *et al.* (2012) Histone H3 Lysine 56 Acetylation and the Response to DNA Replication Fork Damage. *Molecular and Cellular Biology*, **32**, 154-172.
98. Vempati, R.K., Jayani, R.S., Notani, D., Sengupta, A., Galande, S. and Haldar, D. (2010) p300-mediated Acetylation of Histone H3 Lysine 56 Functions in DNA Damage Response in Mammals. *The Journal of Biological Chemistry*, **285**, 28553-28564.

99. Vempati, R. (2012) DNA damage in the presence of chemical genotoxic agents induce acetylation of H3K56 and H4K16 but not H3K9 in mammalian cells. *Molecular Biology Reports*, **39**, 303-308.
100. Tjeertes, J.V., Miller, K.M. and Jackson, S.P. (2009) Screen for DNA-damage-responsive histone modifications identifies H3K9Ac and H3K56Ac in human cells. *The EMBO Journal*, **28**, 1878-1889.
101. Downs, J.A. and Cote, J. (2005) Dynamics of Chromatin during the Repair of DNA Double-Strand Breaks. *Cell Cycle*, **4**, 1373-1376.
102. Dulev, S., Tkach, J., Lin, S. and Batada, N.N. (2014) *SET8 methyltransferase activity during the DNA double-strand break response is required for recruitment of 53BP1*.
103. Faucher, D. and Wellinger, R.J. (2010) Methylated H3K4, a Transcription-Associated Histone Modification, Is Involved in the DNA Damage Response Pathway. *PLoS Genet*, **6**, e1001082.
104. Li, F., Mao, G., Tong, D., Huang, J., Gu, L., Yang, W. and Li, G.-M. (2013) The Histone Mark H3K36me3 Regulates Human DNA Mismatch Repair through its Interaction with MutS α . *Cell*, **153**, 590-600.
105. Kadyrova, L.Y., Mertz, T.M., Zhang, Y., Northam, M.R., Sheng, Z., Lobachev, K.S., Shcherbakova, P.V. and Kadyrov, F.A. (2013) A Reversible Histone H3 Acetylation Cooperates with Mismatch Repair and Replicative Polymerases in Maintaining Genome Stability. *PLoS Genet*, **9**, e1003899.
106. Schuster-Bockler, B. and Lehner, B. (2012) Chromatin organization is a major influence on regional mutation rates in human cancer cells. *Nature*, **488**, 504-507.
107. Cui, L. and Su, X.-z. (2009) Discovery, mechanisms of action and combination therapy of artemisinin. *Expert Review of Anti-infective Therapy*, **7**, 999-1013.
108. Klayman, D. (1985) Qinghaosu (artemisinin): an antimalarial drug from China. *Science*, **228**, 1049-1055.
109. Klonis, N., Crespo-Ortiz, M.P., Bottova, I., Abu-Bakar, N., Kenny, S., Rosenthal, P.J. and Tilley, L. (2011) Artemisinin activity against *Plasmodium falciparum* requires hemoglobin uptake and digestion. *Proceedings of the National Academy of Sciences*, **108**, 11405-11410.
110. Krishna, S., Pulcini, S., Moore, C.M., Teo, B.H.-Y. and Staines, H.M. Pumped up: reflections on PfATP6 as the target for artemisinins. *Trends in Pharmacological Sciences*, **35**, 4-11.
111. Eckstein-Ludwig, U., Webb, R.J., van Goethem, I.D.A., East, J.M., Lee, A.G., Kimura, M., O'Neill, P.M., Bray, P.G., Ward, S.A. and Krishna, S. (2003) Artemisinins target the SERCA of *Plasmodium falciparum*. *Nature*, **424**, 957-961.

112. Shandilya, A., Chacko, S., Jayaram, B. and Ghosh, I. (2013) A plausible mechanism for the antimalarial activity of artemisinin: A computational approach. *Sci. Rep.*, **3**.
113. Li, P.C.H., Lam, E., Roos, W.P., Zdzienicka, M.Z., Kaina, B. and Efferth, T. (2008) Artesunate Derived from Traditional Chinese Medicine Induces DNA Damage and Repair. *Cancer Research*, **68**, 4347-4351.
114. Berdelle, N., Nikolova, T., Quiros, S., Efferth, T. and Kaina, B. (2011) Artesunate Induces Oxidative DNA Damage, Sustained DNA Double-Strand Breaks, and the ATM/ATR Damage Response in Cancer Cells. *Molecular Cancer Therapeutics*, **10**, 2224-2233.
115. Gopalakrishnan, A.M. and Kumar, N. (2014) Anti-malarial action of Artesunate involves DNA damage mediated by Reactive Oxygen Species. *Antimicrobial Agents and Chemotherapy*.
116. Llinás, M., Bozdech, Z., Wong, E.D., Adai, A.T. and DeRisi, J.L. Comparative whole genome transcriptome analysis of three *Plasmodium falciparum* strains. *Nucleic Acids Research*, **34**, 1166-1173.
117. Scherf, A., Lopez-Rubio, J.J. and Riviere, L. (2008) Antigenic Variation in *Plasmodium falciparum*. *Annual Review of Microbiology*, **62**, 445-470.
118. Rovira-Graells, N., Gupta, A.P., Planet, E., Crowley, V.M., Mok, S., Ribas de Pouplana, L., Preiser, P.R., Bozdech, Z. and Cortés, A. (2012) Transcriptional variation in the malaria parasite *Plasmodium falciparum*. *Genome Research*, **22**, 925-938.
119. Marti, M., Baum, J., Rug, M., Tilley, L. and Cowman, A.F. (2005) Signal-mediated export of proteins from the malaria parasite to the host erythrocyte. *The Journal of Cell Biology*, **171**, 587-592.
120. Sargeant, T., Marti, M., Caler, E., Carlton, J., Simpson, K., Speed, T. and Cowman, A. (2006) Lineage-specific expansion of proteins exported to erythrocytes in malaria parasites. *Genome Biology*, **7**, R12.
121. Lopaticki, S., Maier, A.G., Thompson, J., Wilson, D.W., Tham, W.-H., Triglia, T., Gout, A., Speed, T.P., Beeson, J.G., Healer, J. *et al.* (2011) Reticulocyte and Erythrocyte Binding-Like Proteins Function Cooperatively in Invasion of Human Erythrocytes by Malaria Parasites. *Infection and Immunity*, **79**, 1107-1117.
122. Cortés, A. (2008) Switching *Plasmodium falciparum* genes on and off for erythrocyte invasion. *Trends in parasitology*, **24**, 517-524.
123. Cortés, A., Carret, C., Kaneko, O., Yim Lim, B.Y.S., Ivens, A. and Holder, A.A. (2007) Epigenetic Silencing of *Plasmodium falciparum* Genes Linked to Erythrocyte Invasion. *PLoS Pathog*, **3**, e107.

124. Guizetti, J. and Scherf, A. (2013) Silence, activate, poise and switch! Mechanisms of antigenic variation in *Plasmodium falciparum*. *Cellular Microbiology*, **15**, 718-726.
125. Miller, L.H., Baruch, D.I., Marsh, K. and Doumbo, O.K. (2002) The pathogenic basis of malaria. *Nature*, **415**, 673-679.
126. Lopez-Rubio, J.J., Gontijo, A.M., Nunes, M.C., Issar, N., Hernandez Rivas, R. and Scherf, A. (2007) 5' flanking region of var genes nucleate histone modification patterns linked to phenotypic inheritance of virulence traits in malaria parasites. *Molecular Microbiology*, **66**, 1296-1305.
127. Freitas-Junior, L.H., Hernandez-Rivas, R., Ralph, S.A., Montiel-Condado, D., Ruvalcaba-Salazar, O.K., Rojas-Meza, A.P., Mancio-Silva, L., Leal-Silvestre, R.J., Gontijo, A.M., Shorte, S. *et al.* (2005) Telomeric Heterochromatin Propagation and Histone Acetylation Control Mutually Exclusive Expression of Antigenic Variation Genes in Malaria Parasites. *Cell*, **121**, 25-36.
128. Cabral, F.J., Fotoran, W.L. and Wunderlich, G. (2012) Dynamic Activation and Repression of the *Plasmodium falciparum* Gene Family and Their Relation to Chromatin Modification. *PLoS ONE*, **7**, e29881.
129. Jiang, L., López-Barragán, M.J., Jiang, H., Mu, J., Gaur, D., Zhao, K., Felsenfeld, G. and Miller, L.H. (2010) Epigenetic control of the variable expression of a *Plasmodium falciparum* receptor protein for erythrocyte invasion. *Proceedings of the National Academy of Sciences*.
130. Cortés, A., Carret, C., Kaneko, O., Yim Lim, B.Y.S., Ivens, A. and Holder, A.A. (2007) Epigenetic Silencing of Genes Linked to Erythrocyte Invasion. *PLoS Pathog*, **3**, e107.
131. Trager, W. and Jensen, J. (1976) Human malaria parasites in continuous culture. *Science*, **193**, 673-675.
132. Bozdech, Z., Mok, S. and Gupta, A. (2013) In Ménard, R. (ed.), *Malaria*. Humana Press, Vol. 923, pp. 189-211.
133. Gupta, A.P., Chin, W.H., Zhu, L., Mok, S., Luah, Y.-H., Lim, E.-H. and Bozdech, Z. (2013) Dynamic Epigenetic Regulation of Gene Expression during the Life Cycle of Malaria Parasite *Plasmodium falciparum*. *PLoS Pathog*, **9**, e1003170.
134. Livak, K.J. and Schmittgen, T.D. (2001) Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2- $\Delta\Delta$ CT Method. *Methods*, **25**, 402-408.
135. Mok, S., Liong, K.-Y., Lim, E.-H., Huang, X., Zhu, L., Preiser, P.R. and Bozdech, Z. (2014) Structural polymorphism in the promoter of *pfmpr2* confers *Plasmodium falciparum* tolerance to quinoline drugs. *Molecular Microbiology*, **91**, 918-934.
136. Radfar, A., Diez, A. and Bautista, J.M. (2008) Chloroquine mediates specific proteome oxidative damage across the erythrocytic cycle of resistant *Plasmodium falciparum*. *Free Radical Biology and Medicine*, **44**, 2034-2042.

137. Nowsheen, S. and Yang, E.S. (2012) The intersection between dna damage response and cell death pathways. *Experimental oncology*, **34**, 243-254.
138. Merrick, C.J. and Duraisingh, M.T. (2010) Epigenetics in Plasmodium: What Do We Really Know? *Eukaryotic Cell*, **9**, 1150-1158.
139. Kondo, N., Takahashi, A., Ono, K. and Ohnishi, T. (2010) DNA Damage Induced by Alkylating Agents and Repair Pathways. *Journal of Nucleic Acids*, **2010**, 543531.
140. Basu, A. and Krishnamurthy, S. (2010) Cellular Responses to Cisplatin-Induced DNA Damage. *Journal of Nucleic Acids*, **2010**, 16.
141. Karpinich, N.O., Tafani, M., Rothman, R.J., Russo, M.A. and Farber, J.L. (2002) The Course of Etoposide-induced Apoptosis from Damage to DNA and p53 Activation to Mitochondrial Release of Cytochrome c. *Journal of Biological Chemistry*, **277**, 16547-16552.
142. Soubeyrand, S., Pope, L. and Haché, R.J.G. (2010) Topoisomerase II α -dependent induction of a persistent DNA damage response in response to transient etoposide exposure. *Molecular Oncology*, **4**, 38-51.
143. Murray, V., Campbell, H.M. and Gero, A.M. (2011) Plasmodium falciparum: DNA sequence specificity of cisplatin and cisplatin analogues. *Experimental Parasitology*, **128**, 396-400.
144. Kelly, J.M., McRobert, L. and Baker, D.A. (2006) Evidence on the chromosomal location of centromeric DNA in Plasmodium falciparum from etoposide-mediated topoisomerase-II cleavage. *Proceedings of the National Academy of Sciences*, **103**, 6706-6711.
145. Ganesan, K., Ponmee, N., Jiang, L., Fowble, J.W., White, J., Kamchonwongpaisan, S., Yuthavong, Y., Wilairat, P. and Rathod, P.K. (2008) A Genetically Hard-Wired Metabolic Transcriptome in *Plasmodium falciparum* Fails to Mount Protective Responses to Lethal Antifolates. *PLoS Pathog*, **4**, e1000214.
146. McGowan, C.H. and Russell, P. (2004) The DNA damage response: sensing and signaling. *Current Opinion in Cell Biology*, **16**, 629-633.
147. Andrews, K.T., Gupta, A.P., Tran, T.N., Fairlie, D.P., Gobert, G.N. and Bozdech, Z. (2012) Comparative Gene Expression Profiling of P. falciparum Malaria Parasites Exposed to Three Different Histone Deacetylase Inhibitors. *PLoS ONE*, **7**, e31847.
148. Chaal, B., Gupta, A., Wastuwidyaningtyas, B., Luah, Y. and Bozdech, Z. (2010) Histone deacetylases play a major role in the transcriptional regulation of the Plasmodium falciparum life cycle. *PLoS Pathog*, **6**, e1000737.
149. Gasch, A.P., Huang, M., Metzner, S., Botstein, D., Elledge, S.J. and Brown, P.O. (2001) Genomic Expression Responses to DNA-damaging Agents and the Regulatory

- Role of the Yeast ATR Homolog Mec1p. *Molecular Biology of the Cell*, **12**, 2987-3003.
150. Tkach, J.M., Yimit, A., Lee, A.Y., Riffle, M., Costanzo, M., Jaschob, D., Hendry, J.A., Ou, J., Moffat, J., Boone, C. *et al.* (2012) Dissecting DNA damage response pathways by analysing protein localization and abundance changes during DNA replication stress. *Nat Cell Biol*, **14**, 966-976.
 151. Hien, T.T., Davis, T.M.E., Chuong, L.V., Ilett, K.F., Sinh, D.X.T., Phu, N.H., Agus, C., Chiswell, G.M., White, N.J. and Farrar, J. (2004) Comparative Pharmacokinetics of Intramuscular Artesunate and Artemether in Patients with Severe Falciparum Malaria. *Antimicrobial Agents and Chemotherapy*, **48**, 4234-4239.
 152. Natalang, O., Bischoff, E., Deplaine, G., Proux, C., Dillies, M.-A., Sismeiro, O., Guigon, G., Bonnefoy, S., Patarapotikul, J., Mercereau-Puijalon, O. *et al.* (2008) Dynamic RNA profiling in Plasmodium falciparum synchronized blood stages exposed to lethal doses of artesunate. *BMC Genomics*, **9**, 388.
 153. Miotto, O., Amato, R., Ashley, E.A., MacInnis, B., Almagro-Garcia, J., Amaratunga, C., Lim, P., Mead, D., Oyola, S.O., Dhorda, M. *et al.* (2015) Genetic architecture of artemisinin-resistant Plasmodium falciparum. *Nat Genet*, **47**, 226-234.
 154. Jelinsky, S.A., Estep, P., Church, G.M. and Samson, L.D. (2000) Regulatory Networks Revealed by Transcriptional Profiling of Damaged Saccharomyces cerevisiae Cells: Rpn4 Links Base Excision Repair with Proteasomes. *Molecular and Cellular Biology*, **20**, 8157-8167.
 155. Fry, R.C., Begley, T.J. and Samson, L.D. (2005) Genome-wide responses to dna-damaging agents. *Annual Review of Microbiology*, **59**, 357-377.
 156. Lopez-Atalaya, J.P., Ito, S., Valor, L.M., Benito, E. and Barco, A. (2013) Genomic targets, and histone acetylation and gene expression profiling of neural HDAC inhibition. *Nucleic Acids Research*, **41**, 8072-8084.
 157. Gopalakrishnan, A.M. and Kumar, N. (2015) Antimalarial Action of Artesunate Involves DNA Damage Mediated by Reactive Oxygen Species. *Antimicrobial Agents and Chemotherapy*, **59**, 317-325.
 158. Tembe, V. and Henderson, B.R. (2007) Protein trafficking in response to DNA damage. *Cellular Signalling*, **19**, 1113-1120.
 159. Griffiths, L.M., Swartzlander, D., Meadows, K.L., Wilkinson, K.D., Corbett, A.H. and Doetsch, P.W. (2009) Dynamic Compartmentalization of Base Excision Repair Proteins in Response to Nuclear and Mitochondrial Oxidative Stress. *Molecular and Cellular Biology*, **29**, 794-807.
 160. Takala-Harrison, S., Clark, T.G., Jacob, C.G., Cummings, M.P., Miotto, O., Dondorp, A.M., Fukuda, M.M., Nosten, F., Noedl, H., Imwong, M. *et al.* (2013) Genetic loci associated with delayed clearance of Plasmodium falciparum following artemisinin

- treatment in Southeast Asia. *Proceedings of the National Academy of Sciences*, **110**, 240-245.
161. Zhou, Q., Melkounian, Z.K., Lucktong, A., Moniwa, M., Davie, J.R. and Strobl, J.S. (2000) Rapid Induction of Histone Hyperacetylation and Cellular Differentiation in Human Breast Tumor Cell Lines following Degradation of Histone Deacetylase-1. *Journal of Biological Chemistry*, **275**, 35256-35263.
 162. Rahim, R. and Strobl, J.S. (2009) Hydroxychloroquine, chloroquine, and all-trans retinoic acid regulate growth, survival, and histone acetylation in breast cancer cells. *Anti-Cancer Drugs*, **20**, 736-745.
 163. Osley, M.A., Tsukuda, T. and Nickoloff, J.A. (2007) ATP-Dependent Chromatin Remodeling Factors and DNA Damage Repair. *Mutation research*, **618**, 65-80.
 164. Payne, D. (1987) Spread of chloroquine resistance in *Plasmodium falciparum*. *Parasitology Today*, **3**, 241-246.
 165. Roper, C., Pearce, R., Nair, S., Sharp, B., Nosten, F. and Anderson, T. (2004) Intercontinental Spread of Pyrimethamine-Resistant Malaria. *Science*, **305**, 1124.
 166. Mita, T., Venkatesan, M., Ohashi, J., Culleton, R., Takahashi, N., Tsukahara, T., Ndounga, M., Dysoley, L., Endo, H., Hombhanje, F. *et al.* (2011) Limited Geographical Origin and Global Spread of Sulfadoxine-Resistant dhps Alleles in *Plasmodium falciparum* Populations. *Journal of Infectious Diseases*, **204**, 1980-1988.
 167. Noedl, H., Se, Y., Schaefer, K., Smith, B., Socheat, D. and Fukuda, M. (2008) Evidence of artemisinin-resistant malaria in western Cambodia. *N Engl J Med*, **359**, 2619 - 2620.
 168. Hayton, K. and Su, X.z. (2004) Genetic and Biochemical Aspects of Drug Resistance in Malaria Parasites. *Current Drug Targets - Infectious Disorders*, **4**, 1-10.
 169. Sidhu, A.B.S., Verdier-Pinard, D. and Fidock, D.A. (2002) Chloroquine Resistance in *Plasmodium falciparum* Malaria Parasites Conferred by pfcrt Mutations. *Science (New York, N.Y.)*, **298**, 210-213.
 170. Anderson, T., Nkhoma, S., Ecker, A. and Fidock, D. (2011) How can we identify parasite genes that underlie antimalarial drug resistance? *Pharmacogenomics*, **12**, 59-85.
 171. Arieu, F., Witkowski, B., Amaratunga, C., Beghain, J., Langlois, A.-C., Khim, N., Kim, S., Duru, V., Bouchier, C., Ma, L. *et al.* (2014) A molecular marker of artemisinin-resistant *Plasmodium falciparum* malaria. *Nature*, **505**, 50-55.
 172. LeClerc, J.E., Li, B., Payne, W.L. and Cebula, T.A. (1996) High Mutation Frequencies Among *Escherichia coli* and *Salmonella* Pathogens. *Science*, **274**, 1208-1211.

173. Drotschmann, K., Clark, A.B., Tran, H.T., Resnick, M.A., Gordenin, D.A. and Kunkel, T.A. (1999) Mutator phenotypes of yeast strains heterozygous for mutations in the MSH2 gene. *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 2970-2975.
174. Jackson, A.L. and Loeb, L.A. (1998) The Mutation Rate and Cancer. *Genetics*, **148**, 1483-1490.
175. Subramanian, A., Tamayo, P., Mootha, V., Mukherjee, S., Ebert, B., Gillette, M., Paulovich, A., Pomeroy, S., Golub, T., Lander, E. *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*, **102**, 15545 - 15550.
176. Glasspool, R.M., Teodoridis, J.M. and Brown, R. (2006) Epigenetics as a mechanism driving polygenic clinical drug resistance. *British Journal of Cancer*, **94**, 1087-1092.
177. Chen, K.G., Wang, Y.C., Schaner, M.E., Francisco, B., Durán, G.E., Juric, D., Huff, L.M., Padilla-Nash, H., Ried, T., Fojo, T. *et al.* (2005) Genetic and Epigenetic Modeling of the Origins of Multidrug-Resistant Cells in a Human Sarcoma Cell Line. *Cancer Research*, **65**, 9388-9397.
178. Voss, T.S., Bozdech, Z. and Bártfai, R. (2014) Epigenetic memory takes center stage in the survival strategy of malaria parasites. *Current Opinion in Microbiology*, **20**, 88-95.
179. Coleman, Bradley I., Skillman, Kristen M., Jiang, Rays H.Y., Childs, Lauren M., Altenhofen, Lindsey M., Ganter, M., Leung, Y., Goldowitz, I., Kafsack, Björn F.C., Marti, M. *et al.* (2014) A Plasmodium falciparum Histone Deacetylase Regulates Antigenic Variation and Gametocyte Conversion. *Cell Host & Microbe*, **16**, 177-186.
180. de las Alas, M.M., Aebi, S., Fink, D., Howell, S.B. and Los, G. (1997) Loss of DNA Mismatch Repair: Effects on the Rate of Mutation to Drug Resistance. *Journal of the National Cancer Institute*, **89**, 1537-1541.
181. Denamur, E. and Matic, I. (2006) Evolution of mutation rates in bacteria. *Molecular Microbiology*, **60**, 820-827.
182. Legrand, M., Chan, C.L., Jauert, P.A. and Kirkpatrick, D.T. (2007) Role of DNA Mismatch Repair and Double-Strand Break Repair in Genome Stability and Antifungal Drug Resistance in Candida albicans. *Eukaryotic Cell*, **6**, 2194-2205.
183. Nicolaides, N.C., Littman, S.J., Modrich, P., Kinzler, K.W. and Vogelstein, B. (1998) A Naturally Occurring hPMS2 Mutation Can Confer a Dominant Negative Mutator Phenotype. *Molecular and Cellular Biology*, **18**, 1635-1641.
184. Petersen, I., Eastman, R. and Lanzer, M. (2011) Drug-resistant malaria: Molecular mechanisms and implications for public health. *FEBS Letters*, **585**, 1551-1562.

185. Chen, C.-C., Lee, K.-D., Pai, M.-Y., Chu, P.-Y., Hsu, C.-C., Chiu, C.-C., Chen, L.-T., Chang, J.-Y., Hsiao, S.-H. and Leu, Y.-W. (2015) Changes in DNA methylation are associated with the development of drug resistance in cervical cancer cells. *Cancer Cell International*, **15**, 1-9.
186. Duffy, M.F., Selvarajah, S.A., Josling, G.A. and Petter, M. (2012) The role of chromatin in Plasmodium gene expression. *Cellular Microbiology*, **14**, 819-828.
187. Horrocks, P. and Muhia, D. (2005) Pexel/VTS: a protein-export motif in erythrocytes infected with malaria parasites. *Trends in Parasitology*, **21**, 396-399.
188. Luah, Y.-H., Chahal, B.K., Ong, E.Z. and Bozdech, Z. (2010) A Moonlighting Function of *Plasmodium falciparum* Histone 3, Mono-Methylated at Lysine 9? *PLoS ONE*, **5**, e10252.
189. Maier, A.G., Rug, M., O'Neill, M.T., Brown, M., Chakravorty, S., Szeszak, T., Chesson, J., Wu, Y., Hughes, K., Coppel, R.L. *et al.* Exported Proteins Required for Virulence and Rigidity of Plasmodium falciparum-Infected Human Erythrocytes. *Cell*, **134**, 48-61.
190. Iwanaga, S., Kaneko, I., Kato, T. and Yuda, M. (2012) Identification of an AP2-family Protein That Is Critical for Malaria Liver Stage Development. *PLoS ONE*, **7**, e47557.
191. Yuda, M., Iwanaga, S., Shigenobu, S., Mair, G.R., Janse, C.J., Waters, A.P., Kato, T. and Kaneko, I. (2009) Identification of a transcription factor in the mosquito-invasive stage of malaria parasites. *Molecular Microbiology*, **71**, 1402-1414.
192. Kafack, B.F.C., Rovira-Graells, N., Clark, T.G., Bancells, C., Crowley, V.M., Campino, S.G., Williams, A.E., Drought, L.G., Kwiatkowski, D.P., Baker, D.A. *et al.* (2014) A transcriptional switch underlies commitment to sexual development in malaria parasites. *Nature*, **507**, 248-252.
193. 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.
194. Kanti Bhattacharyya, M., Norris, D.E. and Kumar, N. (2004) Molecular players of homologous recombination in protozoan parasites: implications for generating antigenic variation. *Infection, Genetics and Evolution*, **4**, 91-98.
195. Drabløs, F., Feyzi, E., Aas, P.A., Vaagbø, C.B., Kavli, B., Bratlie, M.S., Peña-Díaz, J., Otterlei, M., Slupphaug, G. and Krokan, H.E. (2004) Alkylation damage in DNA and RNA-repair mechanisms and medical significance. *DNA Repair*, **3**, 1389-1407.
196. Duckett, D.R., Drummond, J.T., Murchie, A.I., Reardon, J.T., Sancar, A., Lilley, D.M. and Modrich, P. (1996) Human MutS α recognizes damaged DNA base pairs containing O6-methylguanine, O4-methylthymine, or the cisplatin-d(GpG) adduct.

- Proceedings of the National Academy of Sciences of the United States of America*, **93**, 6443-6447.
197. Pegg, A.E. (2000) Repair of O6-alkylguanine by alkyltransferases. *Mutation Research/Reviews in Mutation Research*, **462**, 83-100.
 198. Dondorp, A., Nosten, F., Yi, P., Das, D., Phyto, A., Tarning, J., Lwin, K., Arie, F., Hanpithakpong, W., Lee, S. *et al.* (2009) Artemisinin resistance in *Plasmodium falciparum* malaria. *N Engl J Med*, **361**, 455 - 467.
 199. Ashley, E.A., Dhorda, M., Fairhurst, R.M., Amaratunga, C., Lim, P., Suon, S., Sreng, S., Anderson, J.M., Mao, S., Sam, B. *et al.* (2014) Spread of Artemisinin Resistance in *Plasmodium falciparum* Malaria. *The New England journal of medicine*, **371**, 411-423.
 200. Mok, S., Ashley, E.A., Ferreira, P.E., Zhu, L., Lin, Z., Yeo, T., Chotivanich, K., Imwong, M., Pukrittayakamee, S., Dhorda, M. *et al.* (2015) Drug resistance. Population transcriptomics of human malaria parasites reveals the mechanism of artemisinin resistance. *Science*, **347**, 431-435.