Chromatin modifications associated with activation and repression of stevor multigene family of plasmodium falciparum during the trophozoite stage

Mushunje, Annals Tatenda

2016

Mushunje, A. T. (2016). Chromatin modifications associated with the activation and repression of stevor multigene family of plasmodium falciparum during the trophozoite stage. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/69222>

<https://doi.org/10.32657/10356/69222>

**CHROMATIN MODIFICATIONS ASSOCIATED
WITH ACTIVATION AND REPRESSION OF
STEVR MULTIGENE FAMILY OF *PLASMODIUM*
FALCIPARUM DURING THE TROPHOZOITE
STAGE**

ANNALS TATENDA MUSHUNJE

Prof. Peter Rainer Preiser
Supervisor

A thesis submitted to
School of Biological Sciences
Division of Molecular Genetics and Cell Biology
In partial fulfillment of the Doctorate degree
NANYANG TECHNOLOGICAL UNIVERSITY

2016

ACKNOWLEDGEMENTS

When I first started this Ph.D., I thought it would be a breeze. I remember writing my personal statement for my graduate application and noting down that I was aware that a Ph.D. would be mentally, socially and financially challenging. But the reality hit me too soon. After four years, I am drained, but at the same time, I would not trade these years for anything. I have met and interacted with some of the most amazing people. I survived these four years because of the following people and I am richly indebted to them.

I have to especially thank my NTU lab members, Makthar, Amy, Aniweh, Annie, Anthony, Karthik, Kripa, Sally, Sam, and Xue Yan and Jenny. Makhtar helped greatly with teaching me culturing of *P. falciparum*. These are skills which I have used these past four years. He was also instrumental in setting my mind on how to best cope with my studies and indirectly mentored me on how to push through when all else screams ‘run’. Annie, the lab aunty or ‘G’ provided the much-needed laughs which were welcome distractions to the everyday science. She always had a kind and encouraging word for me. I know you prayed much for me Annie and I appreciate the time you invested in me. XueYan provided much-needed *stevor* talks. You were such an inspiration especially having listened to your Ph.D. story and how challenging it was and how you pulled through. Same as Sally. Thank you guys for showing me that, I just had to pull through and it would be okay in the end. I am now submitting, yay. Amy, thanks for the practical experience in animal work and again for having a listening ear. You gave me the opportunity to work on that project of yours and now, you can google me up lol. Regina, you came halfway through but still your expertise on the stats and analysis of genome-wide data, helped put some things into perspective. Becky, we had many laughs over lunch. Lunch was the highlight of the day where we would chill and just chat about life and where to go next. You taught me the gift of

efficiency and not to be cooped up in the lab all the time. Much thanks for proofreading my thesis. I wish you well in your next phase of your life. To all of you, the karaoke sessions were a blast. We should continue with them if I stay in Singapore. Yaw and Kripa, we have gone through this together and it has been nice having batch mates going through the same drama as I. You know I have mad love for both of you. Anthony and Omar, you guys have been fantastic. I have thought of calling it quits a few times but after having a conversation with you guys and after much chastisement, I would be greatly encouraged to continue. Omar the random chat sessions removed the boredom of the lab. To the entire lab, my Ph.D. life has been made considerably light by the scientific discussions, invaluable advice, and scientific critique during lab presentations and all the general laughter in the lab.

Other massive thanks to personal in Prof ZB's lab. In particular Archana and Bernard for help with ChIP-chip technique and for the advice on how to proceed with data analysis. Thanks to Julie for the confusing statistics talks. One day I will be a professional like you. Kirk and Sachel for help with basic microarray interpretation. One has to thank their Prof and I take this opportunity to thank mine. I thank him for his guidance, advice, encouragement, and understanding. Thanks, Prof.

My parents and family have been my rock. Their prayers mean so much to me. They have always guided me in the choices that I have made and are always willing to listen to my frustrations but always urging me on. Without them, I would not be here. This is pretty much for you guys especially to my brother Alec who encouraged me to take this route.

Thanks to NTU for this chance to pursue my dreams. Lastly, thanks be to God for His grace and mercy towards me and for there always in the dark times.

ABSTRACT

The *stevor* multigene family is one of three gene families (including *var* and *rifin*) postulated to play a role in antigenic variation in *Plasmodium falciparum*. Recent work has shown how the protein is involved in cytoadherence, rosetting and possibly merozoite invasion showing the proteins involvement in the pathology of malaria. The parasite has 35+ *stevor* genes and expresses 1-3 of these at a parasite level. This tight level of gene control is not unique to *stevor* but to the *vars* and *rifins* as well. In this study, we aimed to characterize histone modifications associated with transcription activation and repression of *stevor* using ChIP-qPCR. We studied four *stevors*: a transgenic *hdhfr-gfp stevor* and three *stevors* expressed naturally. Our study shows that different *stevor* genes show histone-specific regulation and subtle differences in their regulation. H3/H4 acetylation reactions are involved in activation of *stevor* genes and are primarily enriched in the 5' UTR as well coding regions of all *stevors*. H3K4 methylation reactions likely play a supporting role in activation of *stevors* and are enriched in *stevor* specific positions along the 5' UTR and coding regions. H3K9/36 trimethylation reactions are enriched in the 3' end of the 5' UTR and coding regions and are involved in silencing. Interestingly we observed subtle differences between the epigenetic landscape of the *stevors* and we propose that the subtle differences observed between the epigenetic landscapes of the *stevors*, could in turn influence the transcription pattern of *stevors* and also could explain why *stevors* do not follow a strict mutual exclusion as observed for *var* genes.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	i
ABSTRACT.....	iii
TABLE OF CONTENTS.....	iv
LIST OF FIGURES	xii
LIST OF TABLES	xvi
LIST OF ABBREVIATIONS.....	xvii
CHAPTER 1	1
INTRODUCTION	1
1.1 PLASMODIA.....	1
1.2 <i>P. FALCIPARUM</i>	1
1.2.1 Lifecycle	1
1.2.2. Pathogenesis and clinical symptoms	3
1.2.3. Natural Immunity	4
1.2.4 Antigenic variation	5
1.3 MALARIA AND CURRENT EFFORTS AT ERADICATION	6
1.3.1 Global malaria problem	6
1.3.2 Preventative measures	7
1.3.3 Treatment measures/therapies	8
1.3.4. Vaccines.....	9

1.4 MULTIGENE FAMILIES OF <i>P. FALCIPARUM</i>	10
1.4.1 The <i>var</i> gene family	10
1.4.2 The <i>rifin</i> (<i>rif</i>) gene family	14
1.4.3 The <i>stevor</i> gene family	16
1.5 VARIANT ANTIGEN MULTIGENE FAMILIES IN OTHER PLASMODIUM SPECIES	19
1.5.1 The <i>cir</i> gene family	20
1.5.2 The <i>yir</i> gene family	22
1.5.3 The <i>bir</i> gene family	22
1.6 EXPRESSION PROFILES OF THE <i>P. FALCIPARUM</i> MULTIGENE FAMILIES	23
1.7 GENE REGULATION IN <i>P. FALCIPARUM</i>	25
1.7.1 Basal transcription machinery	25
1.8 EPIGENOME OF <i>P. FALCIPARUM</i>	26
1.8.1 The importance of epigenetic regulation	26
1.8.2 Chromatin-mediated epigenetic mechanism	30
1.8.3 Histone modifications and their function	31
1.8.4 <i>P. falciparum</i> modifications associated with transcription activation	33
1.8.5 <i>P. falciparum</i> modifications associated with transcription repression	35
1.8.6 Other elements involved in transcriptional repression in <i>P. falciparum</i>	36
1.8.7 Histone variants	36

1.8.8 Nucleosome density and regulation of transcription	38
1.9 TRANSCRIPTIONAL ACTIVATION OF MULTIGENE FAMILIES	39
1.9.1 Co-regulation of multigene families.....	39
1.9.2 What is known: <i>var</i> and <i>rifin</i>	41
AIMS.....	46
1) Develop targeted tools to study epigenetic marks of active and silent <i>stevors</i>	46
2) Study <i>stevors</i> in their natural expression profiles	46
CHAPTER 2	47
MATERIALS AND METHODS.....	47
2.1 CELL CULTURE	47
2.1.1 Culture and synchronization.....	47
2.2 GENERATION OF TRANSGENIC PARASITES	47
2.2.1 DNA/vector constructs	47
2.2.2 Cloning of inserts into vectors.....	49
2.2.3 Ligation.....	50
2.2.4 Transformation of competent cells	51
2.2.5 Screening of positive colonies	51
2.2.6 Transfection of <i>P. falciparum</i> and selection of transfectants.....	52
2.2.7 Checking for integration.....	53
2.3 Chromatin immunoprecipitation (ChIP) coupled to qPCR.....	54

2.3.1 Chromatin immunoprecipitation.....	54
2.4 RT-PCR.....	60
2.4.1 RNA extraction.....	60
2.4.2 RNA to cDNA conversion.....	61
2.4.3 <i>Stevor</i> primer set.....	61
2.4.4 RT-PCR run.....	64
2.5 MICROSCOPY	65
2.6 LIMITING DILUTION.....	65
2.7 SOUTHERN BLOT	66
2.8 WESTERN BLOT.....	67
2.9 PHYLOGENY ANALYSIS.....	68
2.10 SEARCH FOR OVER-REPRESENTED SEQUENCES	69
2.11 MICROARRAYS	69
2.11.1 CDNA synthesis	69
2.11.2 CDNA labeling and microarray hybridization	70
2.11.3 Microarray data analysis.....	70
CHAPTER 3	72
RESULTS	72
PART 3A: <i>STEVOR</i> INTERGRATED LINE.....	72
3A.1 GENERATION OF <i>STEVOR</i> MUTANT PARASITE LINES.....	72

3A.1.1 Introduction to plasmid design and rationale.....	72
3A.1.2 PCR and construction of stevor plasmids	75
3A.1.2.2 cloning experiments.....	78
3A.1.3 Transfections.....	83
3A.1.4 Addition of drug against the negative selection cassette	84
3A.1.5- Summary of transfection studies	87
3A.1.6- Conclusion and Discussion	88
PART 3B: STUDIES USING TRANSGENIC <i>STEVOR</i> LINE	92
3B.1 INTRODUCTION: STEVOR INTEGRATED PARASITES.....	92
3B.2 ANALYSIS OF STEVOR TRANSCRIPTION IN DIFFERENT <i>pBstevor</i> INTEGRATED CLONES	96
3B.3 ANALYSIS OF CLONE AF5	99
3B.3.1 Conditional expression of <i>hdhfr-gfp</i>	99
3B.3.2 Live cell imaging and real-time PCR analysis.....	99
3B.4 VALIDATION OF ANTIBODIES USED FOR STUDIES	104
3B.4.1 Validation of PCR primers for <i>stevor</i> epigenetic analysis.....	110
3B.5 EPIGENETIC LANDSCAPE OF STEVOR.....	111
3B.5.1 H3K9 acetylation and H3K4 trimethylation are enriched in 5' UTR of 3D7_1254100 ^{ON}	111
3B.5.2 H3K4me2 enriches in the coding region of 3D7_1254100 ^{ON} <i>stevor</i>	115

3B.5.3 H3K9me3 and H3K36me3 are selectively enriched in the 3' ends of 5' UTR and coding region of 3D7_1254100 ^{OFF}	117
3B.5.4 H4K20 trimethylation does not show specific differences between 3D7_1254100 ^{ON} and 3D7_1254100 ^{OFF}	121
3B.5.5 H4 acetylation does not correlate to activation or repression of 3D7_1254100.....	123
3B.5.6 H3K79 trimethylation and H4K20 monomethylation are devoid at <i>stevor</i> loci.....	125
3B. 5.7 PF3D7_1254100 endogenous <i>stevor</i> shows similar enrichment for silencing marks	128
3B.6 CHAPTER CONCLUSION AND DISCUSSION	130
CHAPTER 3C.....	137
STUDIES WITH ENDOGENOUS <i>STEVORS</i>	137
3C.1 INTRODUCTION	137
3C.2 TRANSCRIPTIONAL ANALYSIS OF 3D7 ISOGENIC CLONES	137
3C.3 EPIGENETIC MAP FOR 3D7_1040200.....	141
3C.3.1 Validation of PCR primers.....	141
3C. 3.2 H3K9 and H4 acetylation differentially associate with active 3D7_1040200.....	142
3C.3.3 H3K9me3 enriches in coding region of 3D7_1040200 ^{OFF}	145
3C.3.4 H3K4 di and trimethylation, H3K36 trimethylation and H4K20 trimethylation show no specific differences between 3D7_1040200 ^{ON} and 3D7_1040200 ^{OFF}	146
3C.3.5 Conclusion for PF3D7_1040200	151
3C.4 EPIGENETIC LANDSCAPE FOR 3D7_0617600.....	153

3C.4.1 H3 and H4 acetylation associate with 3D7_0617600 ^{ON}	154
3C.4.2 H3K4 di and trimethylation enriched for 3D7_0617600 ^{ON}	159
3C.4.3 H3K9/36 trimethylation enrich in 3D7_0617600 ^{OFF}	162
3C.4.4: H4K20 trimethylation shows differences between clones 6.2A and 3.2C	165
3C.4.5 Conclusion for 3D7_0617600	166
3C.5 EPIGENETIC LANDSCAPE FOR PF3D7_1372500	170
3C.5.1 H3K9/H4ac enrich in the coding regions of 3D7_1372500 ^{ON}	171
3C.5.2 H3K4 methylation enriches in the coding regions of 3D7_1372500 ^{ON}	175
3C.5.3 H3K9/36 trimethylation enriches in the coding regions of 3D7_1372500 ^{OFF}	177
3C.5.4 H4K20 trimethylation enrichment same for 3D7_1372500 ^{ON} and 3D7_1372500 ^{OFF}	180
3C.5.5 Conclusion for 3D7_1372500	181
3C.6 H3K9/36 trimethylation associate with silent stevor genes	183
CHAPTER 3D	187
STEVOR 5' UTR BIOINFORMATIC ANALYSIS	187
3. D.1. Promoter analysis and overall conservation of stevor promoters.....	187
3. D.2: stevor 5' UTR group into two separate groups	188
3. D.3 3 major sequences are overrepresented in stevor 5' UTR regions	195
CHAPTER 4	199
FINAL CONCLUSION AND DISCUSSION.....	199
4.1 Introduction	199

4.2 Stevors show specific selectivity in terms of regulatory repressive marks	201
4.3 H3K36 trimethylation modification	204
4.4 H3/H4 acetylation reactions associate with active stevors	208
4.5 <i>Stevor</i> H3K4 methylation activation marks are enriched at different genomic regions and may potentially play supporting roles.	213
4.6 Stevors grouping does not change epigenetic profile	219
4.7 Conclusion.....	219
CHAPTER 5	223
FUTURE WORK.....	223
SUPPLEMENTARY	226
PUBLICATIONS AND CONFERENCES.....	243
REFERENCES	244

LIST OF FIGURES

Figure 1.1 The life cycle of <i>Plasmodium falciparum</i> . (Copied from Cowman et al. 2012).	2
Figure 1.2 Countries with ongoing transmission of malaria as of 2013.	7
Figure 1.3: All var genes have 2 exons with a conserved intron located near the 3' end.....	12
Figure 1.4: Pictorial representation of PfEMP1 protein and its adhesive domains.	12
Figure 1.5: Schematic depiction of <i>STEVAR</i> protein.	17
Figure 1.6: Timing of transcription and expression of 3D7 multigene families.....	25
Figure 1.7: Diagram highlighting the modifications associated with silent and active <i>var</i> genes.	44
Figure 3.1: Schematic of desired integration event.	74
Figure 3.2: Positive-Negative plasmids used for cloning created by Vector NTI.	75
Figure 3.3: Gel run of digested pCC4 plasmid.	76
Figure 3.4: Characteristics of chosen <i>stevors</i> :	78
Figure 3.5: Successful PCR of 5' and 3' UTR fragments as well of <i>bsd</i> and <i>hdhfr</i>	80
Figure 3.6: Complementary restriction enzyme for insert fragments.	81
Figure 3.7: PCR of ligated products.	82
Figure 3.8: PCR to check for the presence of transfectants.	83
Figure 3.9: PCR to check for integration in transfectants.....	86
Figure 3.10: Integrated <i>stevor</i> line	93
Figure 3.11: Southern blot of limiting dilution clones from pBstevor.	95
Figure 3.12: A representative analysis of gDNA control in the RNA.	97
Figure 3.13: Analysis of transcription levels of the <i>stevor</i> multigene family in AF5, BC10, BC2, AC9 and AB4, integrated clones.	98
Figure 3.14: Validation of AF5 ^{ON} and AF5 ^{OFF} systems	101

Figure 3.15: Transcription analysis of AF5 ^{ON} and AF5 ^{OFF}	103
Figure 3.16: Human peptides are conserved in Plasmodium H3/4 proteins.....	105
Figure 3.17: Western blot for histone modification proteins.....	106
Figure 3.18: Histone modifications enriched on the seryl-tRNA synthetase gene.....	108
Figure 3.19: Epigenetic mapping of silent <i>var2csa</i> in clonal lines.....	109
Figure 3.20: Schematic showing position and specificity of the 12 PCR primers used.	110
Figure 3.21 H3K9 acetylation levels at 3D7_1254100 <i>stevor</i> gene.	112
Figure 3.22: H3K4 trimethylation levels at different primer positions.	113
Figure 3.23 H3K4 dimethylation levels in 3D7_1254100 ^{OFF} and 3D7_1254100 ^{ON}	115
Figure 3.24: H3K9 trimethylation levels in 3D7_1254100 ^{OFF} and 3D7_1254100 ^{ON}	118
Figure 3.25: H3K36 trimethylation levels in 3D7_1254100 ^{OFF} and 3D7_1254100 ^{ON}	119
Figure 3.26: H4K20 trimethylation levels at 3D7_1254100 ^{OFF} and 3D7_1254100 ^{ON}	121
Figure 3.27:H4K8 acetylation levels at different primer positions.....	124
Figure 3.28: H4K12 acetylation levels at different primer positions.....	124
Figure 3.29: H4K20 monomethylation levels at different primer positions.	126
Figure 3.30: H3K79 trimethylation levels at different primer positions.	127
Figure 3.31: The expression profile of the endogenously expressed PF3D7_1254100 <i>stevor</i> ...	129
Figure 3.32: A summary of modifications associated with regulation of 3D7_1254100 <i>stevor</i> .	134
Figure 3.33: <i>Stevor</i> transcriptional analysis of 3D7 clones.	139
Figure 3.34: Relative expression profiles of PF3D7_1040200 (A) PF3D7_0617600 (B) and PF3D7_1372500 in 6.2A, 3.2C and AF5 clones.	140
Figure 3.35: Schematic and specificity of PCR reactions for PF3D7_1040200:	141
Figure 3.36: H3K9 acetylation levels in 3D7_1040200 on and off states.	142

Figure 3.37: H4K8 acetylation levels in 3D7_1040200 on and off states.	143
Figure 3.38: H4K12 acetylation levels in 3D7_1040200 on and off states	144
Figure 3.39: H3K9 trimethylation levels in 3D7_1040200 on and off states.	145
Figure 3.40: H3K4 trimethylation levels in 3D7_1040200 on and off states.	146
Figure 3.41: H3K4 dimethylation levels in 3D7_1040200 on and off states.	147
Figure 3.42: H3K36 trimethylation levels in 3D7_1040200 on and off states.	149
Figure 3.43: H3K20 trimethylation levels in 3D7_1040200 on and off states.	150
Figure 3.44: A summary of modifications associated with regulation of 3D7_1040200 <i>stevor</i>	152
Figure 3.45: Schematic representation and specificity of primers used for ChIP analysis of 3D7_0617600.	153
Figure 3.46: H3K9 acetylation enrichment when PF3D7_0617600 is on and off.	154
Figure 3.47: H4K8 acetylation levels across 3D7_0617600 ^{ON} and 3D7_0617600 ^{OFF}	156
Figure 3.48: H4K12 acetylation levels across 3D7_0617600 ^{ON} and 3D7_0617600 ^{OFF}	157
Figure 3.49: H3K4 trimethylation levels across 3D7_0617600 ^{ON} and 3D7_0617600 ^{OFF}	160
Figure 3.50: H3K4 dimethylation levels across 3D7_0617600 ^{ON} and 3D7_0617600 ^{OFF}	161
Figure 3.51: H3K9 trimethylation levels across 3D7_0617600 ^{ON} and 3D7_0617600 ^{OFF}	162
Figure 3.52: H3K36 trimethylation levels across 3D7_0617600 ^{ON} and 3D7_0617600 ^{OFF}	164
Figure 3.53: H4K20 trimethylation levels across 3D7_0617600 ^{ON} and 3D7_0617600 ^{OFF}	165
Figure 3.54: A summary of modifications associated with regulation of 3D7_0617600 <i>stevor</i>	169
Figure 3.56: H3K9 acetylation levels across 3D7_1372500 ^{ON} and 3D7_1372500 ^{OFF}	171
Figure 3.57: H4K8 acetylation levels across 3D7_1372500 ^{ON} and 3D7_1372500 ^{OFF}	173
Figure 3.58: H4K12 acetylation levels across 3D7_1372500 ^{ON} and 3D7_1372500 ^{OFF}	174
Figure 3.59: H3K4 dimethylation levels across 3D7_1372500 ^{ON} and 3D7_1372500 ^{OFF}	175

Figure 3.60: H3K4 trimethylation levels across 3D7_1372500 ^{ON} and 3D7_1372500 ^{OFF}	176
Figure 3.61: H3K9 trimethylation levels across 3D7_1372500 ^{ON} and 3D7_1372500 ^{OFF}	177
Figure 3.62: H3K36 trimethylation levels across 3D7_1372500 ^{ON} and 3D7_1372500 ^{OFF}	179
Figure 3.63: H3K20 trimethylation levels across 3D7_1372500 ^{ON} and 3D7_1372500 ^{OFF}	180
Figure 3.64: Modifications associating with activation and transcription of 3D7_1372500.....	182
Figure 3.65: Silent <i>stevor</i> assays for H3K9/36 trimethylation and H3K9 acetylation	184
Figure 3.66: Multiple sequence alignment of <i>stevor</i> 5' UTR region.....	189
Figure 3.67: Neighbor-joining analysis of <i>stevor</i> 750bp upstream regions.....	190
Figure 3.68: Minimum evolution analysis of <i>stevor</i> 750bp upstream regions.	191
Figure 3.69: Top three significant sequences in the 5' UTR regions of <i>stevor</i>	195
Figure 3.70: Position of H3K9/36 trimethylation enrichment for PF3D7_0617600, PF3D7_1040200, PF3D7_1254100 and PF3D7_1372500.....	203
Figure 3.71: Position of H3K9 acetylation enrichment for PF3D7_0617600, PF3D7_1040200, PF3D7_1254100 and PF3D7_1372500	209
Figure 3.72: Position of H4 acetylation enrichment for PF3D7_0617600, PF3D7_1040200 and PF3D7_1372500	210
Figure 3.73: Position of H3K4 methylation enrichment for PF3D7_0617600, PF3D7_1040200, PF3D7_1254100 and PF3D7_1372500	214
Figure 3.74: General model showing histones modifications associated with activation and repression of <i>stevor</i>	222
Sequence 1- Supplementary Figure 1	233
Sequence 2- Supplementary Figure 2	236
Sequence 3-Supplementary Figure 3	240

LIST OF TABLES

Table 1.1: Table was tabulated using information from plasmodb.	29
Table 1.2: Histone post translational modifications identified and confirmed in eukaryotes.	32
Table 2.1: Cloning primers used for all constructs in the PCC vectors.....	48
Table 2.2: PCR parameters for PCC constructs.....	49
Table 2.3 Primers used to amplify 3 way insert ligation.	51
Table 2.4: Primers used to check for integration.	53
Table 2.5: Antibodies used for pull down.....	56
Table 2.6: List of <i>stevor</i> primers used for the ChIP-qPCR reactions.	59
Table 2.6b: List of control primers used for the ChIP-qPCR reactions.....	60
Table 2.7: <i>Stevor</i> primers used for RT-PCR:.....	64
Table 2.8: Antibody dilutions used for Western blot.....	68
Table 2.9: Parameters used for sequence search.....	69
Table 3.1: Summary of constructs targeted for <i>stevor</i> integration:	87
Table 3.2: Chromosomal positions of the five silent <i>stevor</i>	184
Table 3: Chromosomal data for <i>stevors</i> in <i>stevorupsB</i>	194
Table 3.4: <i>stevor</i> grouped based on presence or absence of sequences in the 5' UTR regions..	197
Supplementary Table 1: <i>Stevor</i> TSS from Karmodiya et al. (2015).....	227
Supplementary Table 2: KO data from Jiang et al (2013) for <i>stevor</i> at 30hr post-invasion.....	228
Supplementary Table 3:Primer efficiencies of primers used for ChIP-QPCR studies	230
Supplementary Table 4: Primer efficiency of <i>stevor</i> genes and genes used as controls.....	232
Supplementary Table 5: Sequencing results of 5'UTR PCR products.	232

LIST OF ABBREVIATIONS

Abbreviation used	Full form
5-FU	5 Fluoro-cytosine
ACT	Artemisinin-based combination therapy
ApiAP2	Apicomplexan AP2
<i>bir</i>	<i>Berghei</i> interspersed repeats
BSA	Bovine serum albumin
<i>Bsd</i>	Blasticidin
CD	Cytodeaminase
cDNA	Complementary DNA
ChIP-qPCR	Chromatin immunoprecipitation –quantitative polymerase chain reaction
CIDR	Cysteine-rich inter-domain regions
<i>cir</i>	<i>Chabaudi</i> interspersed repeats
CR1	Complement receptor 1
CSA	Chondroitin sulphate A
CSP	Circumsporozoite protein
CTF	Chromatin-related transcription factors
DAPI	4',6-diamidino-2-phenylindole
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytidine triphosphate

dGTP	Deoxyguanosine triphosphate
dTTP	Deoxythymidine triphosphate
DBL	Duffy-binding-like
DTT	Dichlorodiphenyltrichloroethane
<i>epa</i>	Epithelial adhesion
E-selectin	Endothelial selectin
FISH	Fluorescence <i>in situ</i> hybridization
gDNA	Genomic deoxyribonuclease acid
GFP	Green fluorescent protein
GTF	General transcription factors
<i>G. lamblia</i>	<i>Giardia lamblia</i>
H3K9ac	Histone 3 lysine 9 acetylation
H3K4me2/3	Histone 3 lysine 4 di and trimethylation
H3K9me3	Histone 3 lysine 9 trimethylation
H3K9me3	Histone 3 lysine 36 trimethylation
H4ac	Histone 4 acetylation
H4K8ac	Histone 4 lysine 8 acetylation
H4K12ac	Histone 4 lysine 12 acetylation
H4K20me3	Histone 4 lysine 20 trimethylation
HAT	Histone acetyltransferases

HDAC	Histone deacetylases
<i>hdhfr</i>	Human dihydrofolate reductase
HMTs	Histone methyltransferases
HRP	Horseradish peroxidase
IDC	Intra-erythrocytic developmental cycle
IgG	Immunoglobulin G
iRBC/pRBC	Infected red blood cell/parasitized Red blood cell
ICAM-1	Intercellular adhesion molecule 1
IRS	Indoor residual spraying
ITN	Insecticide treated mosquito
MC	Maurer's clefts
MEE	Mutual exclusive element
M-MuLV	Moloney murine leukemia virus
<i>Msg</i>	Major surface glycoprotein
NaOH	Sodium hydroxide
PBS	Phosphate buffered saline
<i>P. berghei</i>	<i>Plasmodium berghei</i>
<i>P. chabaudi</i>	<i>Plasmodium chabaudi</i>
<i>P. falciparum</i>	<i>Plasmodium falciparum</i>
<i>P. knowlesi</i>	<i>Plasmodium knowlesi</i>

<i>P. malaria</i>	<i>Plasmodium malaria</i>
<i>P. ovale</i>	<i>Plasmodium ovale</i>
<i>P. vivax</i>	<i>Plasmodium vivax</i>
<i>P. yoelli</i>	<i>Plasmodium yoelli</i>
PfEMP1	Plasmodium falciparum erythrocyte membrane protein 1
PFMC-2TM	<i>Plasmodium falciparum</i> maurer's cleft
	two transmembrane domain proteins
PCR	Polymerase chain reaction
PROSC	Plasmodium reactive oxidative stress complex
RBC	Red blood cell
qPCR	Quantitative polymerase chain reaction
<i>Rifin</i> /RIFIN	Repetitive interspersed family
RNA	Ribonucleic acid
RNA poIII	RNA polymerase II
RPMI	Roswell park memorial institute
RT	Reverse transcriptase
<i>Stevor</i> /STEVOR	Sub-telomeric variant open reading frame
SICAvar	Schizont infected cell agglutination variant antigen
SDS	Sodium dodecyl sulfate
STF	Specific transcription factor

<i>T. brucei</i>	<i>Trypanosoma brucei</i>
TAP	Transcription-associated proteins
TF	Transcription factor
TK	Thymidine kinase
TNF- α	Tumor necrosis factor- α
TRiC	T-complex protein-1 (TCP1) ring complex
upsA-E	Upstream sequence A-E
VES	<i>Var</i> expression site
vsg	Variant surface glycoprotein
WHO	World health organization
<i>yir</i>	<i>yoelli</i> interspersed repeats

CHAPTER 1

INTRODUCTION

1.1 PLASMODIA

Plasmodia are obligate eukaryotic protozoan parasites and the causative agents of malaria CDC. The commonly known pathogenic forms in humans are *Plasmodium falciparum* (*P. falciparum*), *P. vivax*, *P. malaria*, *P. ovale* and *P. knowlesi*. The parasites are passed from human to human via the female mosquito of the genus *Anopheles*. *P. knowlesi* which normally infects monkeys in South East Asia has recently been reported to cause serious infections in humans (Cox-Singh et al. 2008; Figtree et al. 2010). *P. falciparum* and *P. vivax* account for the major cases of malaria globally. *P. falciparum* is the most pathogenic accounting for over 80% of deaths, most which occur in Africa. *P. vivax*, however, has a broader geographic distribution as it can develop in the *Anopheles* mosquito at cooler climates and higher altitudes (Jones & Good 2006; WHO 2015). In fact, infections outside Africa are mainly caused by *P. vivax*. Although less fatal compared to *P. falciparum*, the infection is debilitating. In addition, it can form hypnozoites (a dormant form of the parasite) that can cause relapses and act as a potential parasite reservoir.

1.2 *P. FALCIPARUM*

1.2.1 Lifecycle

P. falciparum has a complex life cycle in both the mosquito vector (mosquito stage) and the mammalian host (liver and blood stage) (Figure 1.1). The cycle begins with an infected mosquito injecting sporozoites into the host during a blood meal. Sporozoites migrate to the liver and invade the hepatocytes, where they go through several cycles of asexual division finally maturing

into merozoites. About 30000 merozoites rupture from the hepatocytes and quickly invade erythrocytes (RBC's). In the RBCs, the parasites go through distinct stages of growth: namely the ring, trophozoite, and schizont. The trophozoite stage is the feeding and growth stage characterized by metabolic maturation. The schizont stage is characterized by multiple rounds of asexual division to give rise to 16-32 merozoites which burst from the pRBC and invade fresh RBC. This process in pRBC development is known as the intra-erythrocytic developmental cycle (IDC) or the blood stage and is responsible for clinical symptoms of malaria.

Figure 1.1 The life cycle of *Plasmodium falciparum*. (Copied from Cowman et al. 2012).

Under certain conditions, some merozoites develop into male and female gametes which are taken up by the mosquito where they fuse and sexually reproduce forming zygotes. The zygote matures to an oocyst which will give rise to thousands of sporozoites. These sporozoites migrate

to the salivary glands of the mosquito and can be injected into the vertebrate host during the next blood meal.

1.2.2. Pathogenesis and clinical symptoms

Clinical symptoms of malaria are associated with the IDC stage of the parasite and include flu-like symptoms like fever, headaches, vomiting, joint aches, sweat chills and sometimes diarrhea. It is at this stage during the rupturing of the iRBC, that a lot of the parasites proteins are potentially exposed to the host immune system and are targets for an immune response. During the IDC, particularly during the early trophozoite stage, the parasite extensively modifies the iRBC by insertion of its proteins on the surface (Maier et al. 2008), to create new permeation pathways enabling the parasite to establish its own nutrient supply. Some of these proteins act as ligands and bind host cell receptors in the microvasculature allowing the parasite to adhere/sequester to host tissue. This characteristic is closely associated with virulence, and in *P. falciparum* is mostly linked with the adhesive properties of the variable erythrocyte surface antigens *P. falciparum* erythrocyte membrane protein 1 (PfEMP1) (Pasternak & Dzikowski 2009; Yipp et al. 2003) and to some extent STEVOR (Niang et al. 2014) and RIFIN (Goel et al. 2015), discussed in more detail later.

Sequestration helps the parasite avoid splenic clearance (Engwerda et al. 2005) as the sequestered parasites are not in the circulation (del Portillo et al. 2012). However, this has a negative effect on the host. Sequestration has been linked to cerebral malaria, which accounts for at least 30% of malaria deaths in infants and non-immune adults. Postmortem studies of patients that die from cerebral malaria show obstruction of their brain capillaries by pRBC (MacPherson et al. 1985; Pongponratn et al. 1991; Silamut et al. 1999). Cytokines like TNF- α released in response to the parasite have the effect of inducing expression of ICAM-1, one of the ligands

used by PfEMP1 for sequestration (Sherf et al. 1998). This sequestration might lead to inflammatory reactions that result in swelling of the brain. In this scenario, the inappropriate host immune response seems to contribute to an observed fatality (Langhorne et al. 2008).

Sequestration in other organs leads to obstruction of blood flow to these organs. This obstruction can promote blood clots, can damage the endothelium and can cause local inflammation leading to organ failure (Miller et al. 2002; Trossaert et al. 1991; Schofield & Grau 2005). In pregnant women sequestration of parasites in the umbilical cord endothelium via the CSA receptor (Salanti et al. 2004; Beeson et al. 2008), can have detrimental effects on the growing fetus often resulting in stillborn or premature births due to lack of oxygen and nutrients (Beeson & Duffy 2005).

Other severe/complicated events are anemia, due to the cyclic hemolysis of pRBC and cytokine-induced impairment of erythropoiesis, renal failure, hypoglycemia, hypoxia and lactic acidosis, promoting metabolic acidosis. The parasite can also form rosettes, where an iRBC can interact with non-infected RBC. This is thought to serve the parasite in two ways: i) The iRBC is shielded from the immune system and ii) when the iRBC bursts, the released merozoites can quickly invade the surrounding cells greatly reducing detection by the immune system and increasing parasite burden and risk of severe malaria.

1.2.3. Natural Immunity

Natural immunity against malaria is maintained by antibodies of the IgG subtypes. Sterilizing immunity is rarely achieved but protective immunity does develop slowly (Langhorne et al. 2008). One of the reasons why natural immunity takes so long to develop could be because of the number of antigens that the host needs to target and develop antibodies against. Not only this, but

the host also needs to produce protective antibodies against different strains of *P. falciparum* found in a particular geographical area. In immune subjects, clinical attacks and mortality risks due to cerebral malaria are greatly reduced in individuals between the ages of 2-5 years (anti-disease immunity). From adolescence onwards, severe malaria episodes are rare, although mild episodes can be experienced (anti-parasitic immunity protects against high-density parasitemia and risk of severe malaria) (Snow et al. 1997; Snow & Marsh 1998)(Marsh & Snow 1997). These seemingly healthy individuals however often have parasite loads that would prove lethal to malaria-naïve persons (Doolan et al. 2009).

Longitudinal studies show that natural immunity is i) achieved and effective in adults who have gone through uninterrupted lifelong exposure to parasites ii) is lost when the exposure to parasites stops, iii) is species specific, and iv) could be stage specific (Doolan et al. 2009). Points I and ii are interesting as they argue that the presence of parasites and regular exposure is needed to maintain a protective immunity (Snow et al. 1997). This agrees with the finding that normal memory B-cells are never produced in response to malaria but rather atypical B cells are formed (Weiss et al. 2009; Portugal et al. 2013), that probably produce short-lived plasma cells.

1.2.4 Antigenic variation

Apart from sequestration to avoid splenic clearance, *plasmodia* have developed other ways to evade the immune system (Reeder & Brown 1996) and to cause chronic infections, especially in places where transmission is seasonal. One such mechanism is antigenic variation, whereby the pathogen varies expression of its surface-exposed proteins (Deitsch et al. 1997; Kyes et al. 2007). There are two mechanisms of antigenic variation. One is typically seen in viruses like Influenza where spontaneous mutations lead to slow antigenic drift and antigenic shift (rapid and sudden appearance of new strains) (Wikramaratna et al. 2013). Another type is mostly seen in parasites,

including *trypanosomes* (Borst et al. 1998; Borst et al. 1997), *plasmodium* (Niang et al. 2009; Recker et al. 2011; Smith et al. 1995; Kyes et al. 1999), and *borrelia* (Stoenner et al. 1982). This involves evading the host immunity by changing the surface coat (antigen switching) of expressed surface proteins. In the latter example, antigenic variation involves multigene families in the pathogens. *Trypanosomes* do this by mutually expressing one of their thousand variant surface antigens (*vsg*) (Pays & Nolan 1998; Michels 1984) and *borrelia* uses *vlsE* genes (Zhang & Norris 1998; Dresser et al. 2009; Coutte et al. 2009) to achieve the same purpose. In both cases one member of the family is expressed while the rest are kept silent, thus minimizing chances of exhausting the antigen repertoire.

In *P. falciparum*, antigenic variation involves differential control of multigene families giving rise to hypervariable surface antigens during the course of an infection (Ralph et al. 2005; Stockdale et al. 2008; Noble et al. 2013). Using this mechanism the host immune system is continually confronted with a changing population of epitopes, making it difficult or impossible for the host to control or eliminate the pathogen. This works to the parasites favour, as continued persistence increases the likelihood of transmission.

1.3 MALARIA AND CURRENT EFFORTS AT ERADICATION

1.3.1 Global malaria problem

Malaria, dubbed “the poor man's” disease, continues to be a worldwide burden affecting mainly the poor tropical and sub-tropical countries (Figure 1.2). Per a 2014 WHO report on malaria, there were at least 198 million cases of malaria and an estimated 584000 deaths in 2013. Globally, half the world's population is still at risk of contracting the disease. Most of the mortalities (90%) occur in Sub-Saharan Africa, with 78% occurring in children under 5 years.

Figure 1.2 Countries with ongoing transmission of malaria as of 2013.

Figure reproduced from WHO malaria report (2014).

1.3.2 Preventative measures

Efforts to eradicate malaria have met with some success. Programs for vector control implemented by the WHO, such as the use of insecticide-treated mosquito nets (ITN) and indoor residual spraying (IRS) of DDT against mosquitoes, reduced and even eradicated transmission in some parts of the world (WHO 2014). Use of DDT for IRS has led to the emergence of some resistant strains of plasmodium, highlighting the need to use other insecticides with different modes of action. IRS had achieved greater success in endemic countries in sub-Saharan Africa, where the percentage of the threat-risk population protected by at least one vector control method increased by at least 48%.

Use of antimalarial drugs as prophylaxis has proven beneficial in children and pregnant women in endemic areas. Drugs given include pyrimethamine, chloroquine, proguanil, pyrimethamine-

dapsone, and sulfadoxine-pyrimethamine. A few high profile studies looking at benefits specifically towards pregnant women in endemic countries (parts of Africa and Thailand) showed a reduced risk of developing severe anemia, decreased antenatal and placental parasitemia and an average increase in birth weight when compared with placebo or no drug controls (Radeva-Petrova et al. 2014).

1.3.3 Treatment measures/therapies

With the growth of chloroquine-resistant *P. falciparum* strains, the current gold standard for malaria treatment is a combination of artemisinin with other drugs (mefloquine and primaquine) termed artemisinin-based combination therapy (ACT). As of 2003, 79 out of 87 countries where malaria is endemic are using ACT as a first line treatment option (WHO Malaria report 2014). However, in some parts of Asia (Cambodia, Laos, Myanmar, Thailand and Vietnam) resistance to artemisinin drugs has been reported (Amaratunga et al. 2012; Breman 2012; Noedl et al. 2008; Dondorp et al. 2009) creating a gap in effective therapeutic options. Understanding the molecular mechanism of this resistance is crucial, as the resistance threatens the future of malaria treatment and control. Studies to elucidate the mechanism of resistance have implicated mutations in the Kelch locus (Mbengue et al. 2015; Straimer et al. 2014). Mok et al, through a transcriptomic study of more than 1000 patients' samples with acute malaria, showed that the Kelch mutations lead to an increase in the expression of unfolded protein response (UPR) pathways (Mok et al. 2015). These pathways involve the major *plasmodium* reactive oxidative stress complex (PROSC) and cytoplasmic T-complex protein-1 (TCP1) Ring Complex (TRiC) chaperone complexes which could allow the parasites to protect themselves against artemisinin effects. More work still

needs to be done to fully understand the mechanisms of resistance as this knowledge will aid future drug discovery.

1.3.4. Vaccines

With the growing resistance to gold standard drugs, there is an urgent need for other approaches to prevent or treat malaria. Thus, vaccines are being considered as an option, both for prevention as well as reducing transmission. However, to date, efforts at vaccine development have proven challenging partly due to i) the complexity of the parasite and its ability to evade host immunity, ii) lack of a full understanding of the complex interplay between the host immunity and the parasite. Current vaccine development strategies target three areas, the sporozoite/liver stage, the blood stage and the transmission stage.

RTS, S/AS01, a vaccine against *P. falciparum*, is the only vaccine that has progressed to phase 3 clinical trials. This is an anti-sporozoite vaccine made up of the central repeating region of a circumsporozoite protein (CSP) fused to recombinant hepatitis B surface antigen. Phase 3 trials recruited more than 15000 infants (6-12 weeks) and children (5-17 months) from different geographical areas in Africa. Vaccine efficacy against clinical and severe malaria in the 5-17 months group was 45%-47% and 55% respectively (Partnership 2011; Partnership 2011). In the infant group efficacy was 33% for all malaria episodes and 37% against severe malaria (The RTS 2012). This was in addition to the continued use of impregnated bed nets. An 18-month follow-up showed that efficacy in the infant age group had dropped significantly to 26% whereas in the 5-17 month age group efficacy was still around 46% for clinical malaria and 35% for severe malaria (The RTS, S Clinical Trials Partnership 2013). A booster shot given 18 months after third vaccine dose prolonged efficacy overall. Even though the efficacy over time

dropped, this is the only vaccine to have reached the advanced clinical trials. In places where malaria transmission is high, the vaccine could help substantially in reducing the number of cases (RTSS Clinical Trials Partnership 2015).

1.4 MULTIGENE FAMILIES OF *P. FALCIPARUM*

The three major antigenic variant families in *P. falciparum* are *var*, *rifin*, and *stevor*. Other smaller families include *surfin* (Winter et al. 2005; Chan et al. 2014) *phist* (*plasmodium* helical interspersed subtelomeric) and *hyp* families 1-17 (Sargeant et al. 2006) and *pfmc-2tm* (*P. falciparum* Maurer's clefts two transmembrane) (Sam-Yellowe et al. 2004).

1.4.1 The *var* gene family

The *var* gene family, coding for PfEMP1 (Su et al. 1995; Baruch et al. 1995; Baruch et al. 1996; Smith et al. 1995), is the best-characterized family in *P. falciparum*. There are about 60 *var* copies located at subtelomeric regions of all fourteen chromosomes in the 3D7 clone of *P. falciparum* used in the laboratory (Gardner et al. 2013; Rubio et al. 1996; Fischer et al. 1997; Hernandez-Rivas et al. 1997). Four *var* clusters are located internally on chromosome 4, 7, 8 and 12 (Gardner et al. 2013; Thompson et al. 1997). The *var* genes have been divided into 3 major subgroups (upsA-C) based on location and the 5' upstream promoter-like sequences (Lavstsen et al. 2003; Voss et al. 2000; Kraemer & Smith 2003). UpsA and upsB *var* are located in the subtelomeric regions whereas UpsC *var* are located internally (Lavstsen et al. 2003). Two other *var* genes with unique upsD and upsE type 5' UTR's are known (Lavstsen et al. 2003). The *var* genes have a 2 exon structure with a hypervariable adhesive exon 1 which codes for the extracellular regions of the PfEMP1 proteins and a semi-conserved exon 2 which codes for the

intracellular region (Su et al. 1995; Hernandez-Rivas et al. 1997) (Figure 1.3). The intracellular region anchors the protein to the RBC membrane.

PfEMP1 is one of the targets of the immune response against *P. falciparum* (Warimwe et al. 2009; Bull et al. 1998; Bull & Marsh 2002) and is associated with the pathogenic traits of *P. falciparum* (Welsh & Fujinami 2007; Kemp et al. 1992). It is expressed on the iRBC surface (Leech et al. 1984; Aley et al. 1984) and is thought to mediate cytoadhesion of iRBC to other uninfected RBC (rosetting) (Rowe et al. 1997; Rowe et al. 1995; Albrecht et al. 2011) or to the microvascular endothelium (Scherf et al. 1998; Yipp et al. 2003; Su et al. 1995; Chen et al. 2000). Cytoadherence to microvasculature occurs via several Duffy-Binding-Like (DBL, α - ϵ) domains as well as Cysteine-Rich Interdomain Regions (CIDR, α - γ) (Figure 1.4) coded for by exon 1. These different domains are adhesive and bind to different host receptors including CD36, complement receptor 1, ICAM-1, CSA, blood group antigen A, thrombospondin, and heparin sulfate to name a few (Barragan et al. 2000; Fernandez et al. 1998; Baruch et al. 1995; Baruch et al. 1999; Sherman et al. 2003; Q Chen et al. 1998; Baruch et al. 1996).

Var genes have been observed to be transcribed in a mutually exclusive manner such that at a single parasite level, only one *var* gene is active. Using RT-PCR and nuclear run-on assays, Chen et al. and Scherf et al. showed that during the early ring stage multiple *var* transcripts are detected in a single parasite but as the parasite matures (trophozoite stage) only a single mature transcript encoding the single cytoadherence variant is observed (Scherf et al. 1998; Chen et al. 1998).

Figure 1.3: All *var* genes have 2 exons with a conserved intron located near the 3' end.

Exon 1 is highly polymorphic between *vars* and codes for the extracellular adhesive domain of PfEMP1 which has various DBL and cysteine-rich interdomain regions. DBL domain 1 has short conserved sequence motif shared between all *vars*. Exon 2 codes for the intracellular domain also known as the acidic terminal segment (ATS). (Figure from (Stockdale et al. 2008).

Figure 1.4: Pictorial representation of PfEMP1 protein and its adhesive domains.

Presumably, the other transcripts are degraded. However, Duffy et al. reported observance of multiple *var* transcripts even at the trophozoite stage (Duffy et al. 2002) in parasites that had been selected for a specific binding phenotype. Each domain is hypervariable between the different PfEMP1 (Figure from (Kyes et al. 2001). One possible explanation could be that in the Chen and Scherf studies, the ring expressed *var* genes were below the detection level during the trophozoite stage and were therefore not detected. Transcription of more than one *var* loci and expression of more than one PfEMP1 molecule have also been observed in transgenic lines as

well as in wild-type parasites (Noviyanti et al. 2001; Joergensen et al. 2010; Dzikowski et al. 2007) suggesting that in some parasites this mutual exclusive property is lost. Though the above results might appear to be contradictory, better-controlled experiments have shed light on the mutually exclusive property of the *var* genes. It is now clear that the *var* promoter is responsible for dictating the states of activity (Voss et al. 2006; Voss et al. 2007; Dzikowski et al. 2006; Jose Juan Lopez-Rubio 2007; Witmer et al. 2012).

Dzikowski et al. and Voss et al. showed that transfection of an integrated or episomally expressed *var* promoter driving a drug selectable cassette, led to the shutdown of endogenously transcribed *var* genes (Dzikowski et al. 2006; Voss et al. 2006). These experiments highlighted that these drug-induced promoters could impact on the pathway of mutually exclusive transcription (Dzikowski et al. 2006; Voss et al. 2006). In Dzikowski et al. experiments, removal of the drug from the parasites, led to a decreased expression of the drug induced *var* promoter and re-expression of the previous endogenous *var* or a new *var* gene (Dzikowski et al. 2007). Importantly this shut down of the endogenous *var* promoters is an intrinsic property and not dependant on the production of a PfEMP1 protein, therefore not a negative feedback mechanism, as in the experiments the *var* gene had been replaced by a drug cassette and did not produce PfEMP1 (Dzikowski et al. 2007). The work of Voss and Dzikowski has also been observed by other groups. Witmer et al. also showed that only *var* gene promoters could control the mutually exclusive property and the promoters of other gene families did not have this characteristic (Witmer et al. 2012).

Even though the parasite has 60 copies of the *var* genes, allelic recombination between subtelomeric *var* genes located close to each other or on other chromosomes is thought to be responsible for the large diversity of *var* genes seen within and between laboratory and field

strains (Freitas-Junior et al. 2000; Taylor et al. 2000; Deitsch et al. 1999; Ward et al. 1999; Kraemer et al. 2007; Frank et al. 2008; Duffy et al. 2009). The nature and extent of this recombination was recently alluded to by Claessens et al (2014). By performing an *in vitro* evolution experiment and tracking *var* gene mutations by way of whole genome sequencing they found that, compared to the whole genome, *var* genes underwent highly concentrated *de novo* structural variations in exon 1. These ectopic recombination events involved inversions, duplications, translocations and the recombined *var* genes were always in frame. The investigators also calculated a recombination rate of 0.2% in iRBC *in vitro* per life cycle. This high rate of recombination potentially suggests that millions of new antigenic structures are generated per day in an infected individual (Claessens et al. 2014). This could translate to even more in the context of an active immune system in the field and could also help explain why it takes years to develop a protective immune system and a sterilizing immunity is never achieved. *Var* genes expression is also reset by the mosquito vector. Analysis of *var* genes from 18 volunteers, who had been infected with the same parental line, showed that they all expressed dominantly identical *vars* from group B and group A. The transcriptional profile between the parasites that were used to infect mosquitoes and *Plasmodia* recovered from infected volunteers was different (Bachmann et al. 2016). The authors conclude that the *var* gene expression had been reset by the mosquito and propose a method whereby new parasites entering the human blood express a specific *var* repertoire and this is modified by the host-parasite interactions.

1.4.2 The *rifin* (*rif*) gene family

The repetitive interspersed family (*rifin*) was first described in 1988 by Weber (Weber 1988). It is the largest multigene family in *P. falciparum* with approximately 160 members in the 3D7 clone. *Rifins* are located sub-telomerically on all 14 chromosomes in close association with *var*

genes in clusters within 50 kb of the telomeres (Kyes et al. 1999; Gardner et al. 2013). Each *rifin* gene has two exons with a short intron. Exon 1 encodes the signal peptide and exon 2 encodes most of the 35-45kDa clonally variant RIFIN protein (Cheng et al. 1998; Kyes et al. 1999; Gardner et al. 2013). The protein has a surface exposed N-terminus which is semi-conserved and rich in cysteine amino acids (Kyes et al. 1999). The C-terminal part is composed of hydrophobic amino acids and anchors the protein in the membrane (Goel et al. 2015).

RIFINs have been characterized as RIFIN A and RIFIN B which differ from each other by the presence or absence of 25 amino acids in the semi-conserved domain (Joannin et al. 2008; Petter et al. 2007). 70% of RIFINs belong to the A group and are trafficked to the surface of the iRBC, via Maurer's clefts whereas RIFIN B accumulate in the parasite (Petter et al. 2007; Joannin et al. 2008). Parasites can express more than one RIFIN (Fernandez et al. 1999; Petter et al. 2007) though there are dominant RIFINs which are expressed at the population level in different laboratory strains (Cabral & Wunderlich 2009) suggesting some kind of transcriptional memory.

Although the biological importance of RIFINs on the surface of iRBC was not fully known, roles in pathology via cytoadherence (using ligand CD31) and rosetting had been suggested (Kyes et al. 1999; Fernandez et al. 1999). Recently, Goel et al confirmed the rosetting and cytoadherence roles of RIFINs. In their work, they transfected CHO cells with group A and B RIFINs and analyzed their ability to bind RBCs. Only CHO with type A RIFINs bound to RBCs in large numbers and these preferentially only bound to RBCs with blood group A compared to group O antigens. In an *in-vivo* study in rats, sequestration independent of PfEMP1 by RIFIN proteins was also observed, (Goel et al. 2015) strongly suggesting the involvement of RIFINs in the pathology and severity of the disease. Their expression in the gametocyte stage (Petter et al. 2008; Petter et al. 2007; Wang et al. 2010; Mwakalinga et al. 2012), as well as in merozoites

(Petter et al. 2007; Mwakalinga et al. 2012), and sporozoites stages (Wang et al. 2010) suggests multiple distinct functions or a key function in antigenic variation during the parasites different developmental stages.

RIFINs also elicit a substantial humoral immune response (Fernandez et al. 1999) . In this sand are thought to most likely be a target of protective immunity against malaria (Abdel-Latif et al. 2002; Abdel-Latif et al. 2003). A recent study confirmed the presence of neutralizing antibodies against RIFINs (Tan et al. 2016). In this study a portion of the LAIR1 gene encoding the extracellular collagen-binding domain of more than 100 amino acids inserted itself between the V and the DJ segments. This inter-chromosomal transfer of sequence into a rearranged heavy chain gene led to the production of broad antibodies which could recognize RIFIN genes.

1.4.3 The *stevor* gene family

Stevor (subtelomeric variable open reading frame) genes are subtelomerically located clustered with *var* and *rif* genes on all 14 chromosomes (Cheng et al. 1998; Gardner et al. 2013). There are about 40 *stevor* genes in the haploid genome of the 3D7 clone. *Stevor* genes have 2 exons (Figure 1.5). Exon 1 is smaller and codes for the signal peptide. Exon 2 is much larger (about 1kb in size) and codes for the 30-40kDa semi-conserved domain, hypervariable as well as a cytoplasmic region of STEVOR (Cheng et al. 1998). The protein has 1 transmembrane domain (Bachmann et al. 2015; Joannin et al. 2011).

The expression pattern of STEVOR provides some interesting insights. STEVOR are expressed in all stages of the parasite development, including the surface of iRBC during the IDC (Niang et al. 2014; Niang et al. 2009) and, apical tips of merozoites (Blythe et al. 2008; Khattab et al. 2008; Khattab & Meri 2011) suggesting a role in invasion.

Figure 1.5: Schematic depiction of STEVOR protein.

The short exon 1 codes for the signal peptide whereas the long exon 2 codes for the majority of the protein as well the TM domain. Sequence similarities between the different regions of the STEVORS are shown with the hypervariable region having 47% sequence similarity between the STEVOR. New data supports the presence of only 1 TM, suggesting the semi-conserved region is exposed to the host immune system.

Expression of STEVOR during the merozoite stage could also be a mechanism employed by the parasite to evade the host immune system during invasion and establish a long-lasting persistent infection. In work by Niang et al. anti-STEVOR antibodies inhibited *in vitro* invasion of merozoites (Niang et al. 2014), supporting the role of STEVOR in invasion. *Stevor* expression is also observed during sporozoites (McRobert et al. 2004) and gametocyte stages (Sharp et al. 2006; McRobert et al. 2004). During the IDC *stevors* also undergo two waves of transcription: first during the trophozoite stage and then during the merozoite stage. The *stevor* variants expressed at both stages is the same (Blythe et al. 2008). Like the other gene families *stevor* do show antigenic variation and expression switching though the rate of switching is thought to be lower than that of *vars* (Lavazec et al. 2007a; Lavazec et al. 2007b). Unlike *var* genes, *stevor* do not exhibit strict mutual exclusive expression, but rather a small subset is transcribed even at a single cell level (Lavazec et al. 2007a), but some single cells do show single *stevor* transcription (Kaviratne et al. 2002).

The expression of STEVOR on the iRBC surface has also suggested a number of putative roles. An inconclusive clinical study on children showed that sera from one child could not identify STEVOR in sera of another child (Schreiber *et al* 2007). This could imply that STEVOR on the surface of RBCs mediates antigenic variation. Additional support for the importance of STEVOR in the host comes from the fact that culture-adapted parasites lose their need for STEVOR. In culture, less than 30% of cells express STEVOR compared to field isolates where more than 90% of cells express STEVOR (Blythe et al. 2008). This supports the argument that STEVOR proteins may have a role in host-parasite interactions and immune evasion.

Other studies on the surface expression of STEVOR on iRBC have suggested that it might play a role in sequestration of late stage parasites. This hypothesis was indirectly supported by Sanyal, who showed that surface expression of STEVOR impacted the mechanical properties of iRBCs, such that the iRBC becomes more rigid (Sanyal et al. 2012). They hypothesize that this could further enhance sequestration of iRBC in microvessels, further increasing disease severity. This work was confirmed by Niang et al. who showed that STEVOR is indeed an RBC binding protein and it binds RBC via Glycophorin C, the binding correlating with the amount of GPC on the RBC (Niang et al. 2014). This binding could be abolished by use of STEVOR antibodies. STEVOR, like RIFINs and PfEMP1, is able to rosette and can do so independently of PfEMP1 (Niang et al. 2014). Overall it is clear that STEVOR plays an important role in parasite biology and pathology.

1.5 VARIANT ANTIGEN MULTIGENE FAMILIES IN OTHER *PLASMODIUM* SPECIES

Genome sequencing projects of *Plasmodium* species have led to whole genome and comparative studies between the different *Plasmodium* species. All *plasmodium* species have 14 linear chromosomes and their genomes range from 20-25Mb coding for approximately >5000 genes (Gardner et al. 2013). 60% of genes are orthologous between *P. falciparum* and other *Plasmodium* species, with unique species-specific genes located in the subtelomeric regions of the chromosomes (Carlton et al. 2005). These genes are multi-gene families and code for immune-dominant antigens likely involved in host-parasite interactions.

A conserved multigene family is present in *P. vivax* (human host), *P. knowlesi* (monkey and human host), *P. berghei*, *P. yoelli* and *P. chabaudi* (rodent hosts). The variant antigenic genes in these species are *vivax* interspersed repeats (*vir*) in *P. vivax* (Bernabeu et al. 2012; Lopez et al. 2013; del Portillo et al. 2001), *knowlesi* interspersed repeats (*kir*) in *P. knowlesi* (Corredor et al. 2004; Figtree et al. 2010; Pain et al. 2008), *chabaudi* interspersed repeats (*cir*) in *P. chabaudi* (Hall & Carlton 2005; Fischer et al. 2003), *berghei* interspersed repeats (*bir*) in *P. berghei* (Carlton et al. 2002; Hall et al. 2005) and *yoelli* interspersed repeat (*yir*) in *P. yoelli* (Janssen et al. 2004; Carlton et al. 2002). These gene families have a conserved intron/exon structure (short first exon, long polymorphic second exon (80% of the coding region) and short third exon) and splice junction sequences (Janssen et al. 2002).

Probabilistic modeling, protein structural predictions and gene structure analyses on *vir*, *cir*, *bir*, and *yir*, gene families showed that they shared structural homologies with the *rif* and *stevor* multigene families of *P. falciparum* (Janssen et al. 2004). This has led to the classification of *rif*,

stevor, *bir*, *cir*, *yir* and *vir* as a large variant gene superfamily called Plasmodium interspersed repeats, *pir* (Janssen et al. 2004). It is important to highlight that PfEMP1 is unique to *P. falciparum* (Janssen et al. 2002) with some homology to *P. knowlesi* SICAvar genes (Corredor et al. 2004; Korir & Galinski 2006). Furthermore, PfEMP1 and SICA proteins are much larger than the proteins products of *pir* genes; hence the PIR's are referred to as small antigen variant antigen molecules.

The PIRs have a conserved domain distribution indicating a functional relation. Also, the predicted secondary structure, as well as the topographical position of the variable and conserved regions, is the same with some indication that they share similar tertiary structures. The position of the hypervariable region is also conserved (Janssen et al. 2004). As the proteins are thought to be under immune selection, the conservation in structure and variable domain position suggests that PIR's interact with the host immune system in a similar manner. The conservation of PIR across human and rodent species provides a unique opportunity to study the biological role of these genes in an in vivo mouse model. Observations in the rodent parasites would allow us to make predictions concerning interactions of *P. vivax* or *P. falciparum* with its human host. I will give a brief description of *cir*, *bir* and *yir* families.

1.5.1 The *cir* gene family

P. chabaudi is good rodent parasite to study sequestration as well as infection as it causes both acute and chronic phases of malaria (McLean et al. 1982; Brugat et al. 2014). *P. chabaudi* is known to undergo organ sequestration, which is influenced by the immune system, (Brugat et al. 2014), possibly using host receptor CD36 (Mota et al. 2000), and the CIR proteins are potentially responsible for this phenotype. In the sequenced *P. chabaudi* strain, there are 198 copies of *cir* genes (Otto et al. 2014). *cir* genes are located in the subtelomeric regions of the chromosomes

(Fischer et al. 2003). Microarrays and RNA sequencing studies using synchronized cultures show that 1) in a parasite population about 40% of *cirs* are transcribed during the IDC cycle, 2) there are dominant *cirs* which are expressed at peak infection and 3) specific *cirs* are transcribed at different stages of the blood stage cycle (Lawton et al. 2012), suggesting different functional properties. Some *cirs* have also been shown to undergo antigenic switching at least in laboratory mice and different *cir*-expressing parasites accumulate in different tissues of infected mice (Ebbinghaus & Krücken 2011; Cunningham et al. 2010).

Experiments using *chabaudi* have shown how the mosquito vector plays a central role in resetting/reprogramming expression of virulence genes. It had been observed that serial passage of parasites led them to become more virulent and more pathogenic and it was hypothesized that the mosquito could potentially reverse/modify this phenotype. This hypothesis was formally proven by (Spence et al. 2013). In their experiments, they observed that infecting mice with serially passaged parasites over time led to severe disease and hyperparasitemia in the mice, whereas mosquito transmission of the same serially passaged parasites, led to a low-grade, chronic infection with minimal pathology. RNA-seq showed that the difference between the two types of parasites was more pronounced in the expression of *pir* genes, *cirs*, in *P. chabaudi*. Serially passed parasites expressed a restricted number of *cirs*, whereas mosquito parasites expressed a high proportion of *cirs*. The conclusion was that the host immune system modifies the expression of these virulent *cir* genes, leaving a small repertoire which is better adapted to the host, promotes survival and chronic infections. Mosquito uptake and gamete formation in mosquito resets this ‘programme’ and allows the new parasites which will be injected into a host to develop their own ‘programme’.

CIR proteins have been grouped into subfamily A and B based on sequence similarity (Lawton et al. 2012). Recent work shows that different CIR's localize to the surface of iRBC and merozoites, strongly suggesting interaction with the host (Yam et al. 2016). In the same work, the authors using recombinant CIR proteins show that CIRs bind to mouse RBCs linking them to rosetting or potentially invasion (Yam et al. 2016).

1.5.2 The *yir* gene family

The *yir* multigene family in *P. yoelli* has 838 *yir* genes which is the largest variant antigen family within *Plasmodium* species (Carlton et al. 2002). Expression of *yir* is similar to other *pirs*, with a broad range of *yirs* expressed during the IDC, but surprisingly 1-3 *yirs* expressed at a parasite level (Cunningham et al. 2009). There is a stage-specific expression of certain *yirs* during the IDC (Cunningham et al. 2009). Again the expression of *yirs* has been shown to occur during other developmental stages including sporozoites and gametocytes (Carlton et al. 2002), suggesting a multifunctional role during parasite development. Clonal infections carried out by Cunningham et al. demonstrated that *yirs* undergo transcriptional switching but this is not accompanied by epigenetic memory as no dominant YIR expression was apparent (Cunningham et al. 2009). Hence, antigenic variation in *P. yoelli* involves simultaneous exposure of different YIR to the immune system.

1.5.3 The *bir* gene family

The *bir* gene family has 180 copies (Hall & Carlton 2005). Similarly, to *cirs* and *yirs*, *birs* are expressed in a stage-specific manner, including the IDC, gametocyte, oocyst, ookinete and sporozoite stages (Hall et al. 2005). *P. berghei* parasites go through CD36 mediated sequestration in the lungs, as well as other tissues such as the brain, placenta, and adipose tissue

(Franke-Fayard et al. 2005). In humans, PfEMP1 is known to sequester via CD36, hence this sequestration to CD36 by *berghei* indicates the importance of non-PfEMP1 ligands in parasite pathology. The sequestration of *P. berghei* is needed for the survival of the parasite, as reduced CD36 sequestration is detrimental to the parasites growth *in vivo* (Fonager et al. 2012). Also, *P. berghei* undergoes antigenic variation during the course of an infection (Janssen et al. 2002).

1.6 EXPRESSION PROFILES OF THE *P. FALCIPARUM* MULTIGENE FAMILIES

Due to the role of variant multigene families in host-parasite interactions, as well as in immune evasion, the regulation of variant multigene families in *P. falciparum* has become an important area of research. More in-depth understanding regarding the regulation of these genes would provide us with new insights and strategies on how to perturb their switching, which could, in turn, lead to new vaccination strategies. Currently, most of the work in this field has focused on *P. falciparum*.

Earlier transcriptomic studies on the IDC cycle of *P. falciparum* using microarray showed that during the IDC there was a developmental up-regulation of specific mRNAs (Mamoun et al. 2001). During the ring stage genes involved in protein synthesis (ribosomal proteins, translation factors) peaked, followed at the trophozoite stage by genes involved in metabolism, especially genes for proteins involved in glycolysis. The trophozoite is known as the most metabolically active stage. Adhesion/invasion genes peak at the schizont stage when the parasite is preparing for invasion. A general shut-off of gene expression happens at the end of the IDC with a small

set of genes, including a number of protein kinases, turned on (Mamoun et al. 2001). This tight regulation of gene expression was also observed by other groups (Bozdech et al. 2003; Le Roch 2003). Such regulation extends to the *var*, *rifin* and *stevor* gene families which show an ordered pattern of transcription (Figure 1.6). *Var* is transcribed first, soon after merozoites invade the RBC and transcription continues till the early trophozoite stage. *Rif* is transcribed next for a period of about 10 hours between early trophozoite to middle trophozoite stage (18-27 hours post-invasion). In parasite culture, several *rifin* are transcribed while the rest of the members are transcriptionally silent. Recent RNA-seq data suggests that, at the RNA level, a single *rifin* is dominantly transcribed while the rest are transcribed at low levels (Goel et al. 2015), suggesting the presence of a relaxed mutual exclusion mechanism. *Stevor* is transcribed after *rifin* and starts around 22 – 32 hours post invasion. A parasite can transcribe between 1-3 *stevor* (Kaviratne et al. 2002; Lavazec et al. 2007b). Though *rifin* and *stevor* do not show strict mutual exclusion in culture, at a population level the transcription profile is relatively stable, suggesting the presence of a coordinated regulatory mechanism.

Figure 1.6: Timing of transcription and expression of 3D7 multigene families

The *var* gene family is transcribed first from the ring stage to around 24 hours after invasion. The expression of *var* genes shares the same temporal profile with *rifin*. *Stevor* peaks transcription at the trophozoite stage. The proteins ones made persist throughout the IDC cycle.

1.7 GENE REGULATION IN *P. FALCIPARUM*

1.7.1 Basal transcription machinery

In eukaryotes, transcription is mediated by complexes of RNA polymerases and a variety of different factors which serve to make the process more stringent and well regulated. The *P. falciparum* genome codes for genes involved in basal transcription machinery which includes TATA-binding protein (TBP), general transcription factors (GTFs), and transcription-associated proteins (TAPs) (Gardner et al. 2013; Callebaut et al. 2005; Bischoff & Vaquero 2010). The GTF's include transcription factor (TF) II subunits (TFIIA-E) which assemble and help to

stabilize the polymerase preinitiation complex (PIC) structure. The GTF's and TAPs are affected by binding of specific transcription factors (STFs) that direct the transcription of specific genes in response to stimuli, and hence regulate the rate of transcription.

Though *P. falciparum* codes for GTF's and TAPs, specific transcription factors (STFs) were thought to be lacking. However, the discovery of 27 Apicomplexan lineage-specific AP2 family (Balaji et al. 2005; De Silva et al. 2008) and PfMyb1 transcription factors (Gissot et al. 2005) as well as 1 Prx Regulatory Element-binding protein (PREBP) (Komaki-Yasuda et al. 2013) is challenging this view. Two AP2 TFs have been characterized. The first PfSIP2 binds to a sequence which is conserved and found in upstream regions of subtelomeric *var* genes as well as within telomere-associated repeat elements (Flueck et al. 2010). In the same study, the authors show that PfSIP2 is involved in end-chromosome biology i.e. maintenance of the telomeric ends and in *var* gene silencing (Flueck et al. 2010). The second which is AP-2G has been shown to be essential for gametocytogenesis *in vitro* (Kafsack et al. 2014) and *in vivo* (Sinha et al. 2014) as well as being epigenetically regulated by HP1 (Brancucci et al. 2014). PfMyb1 proteins are crucial for cell cycle regulation, as knockout parasites have cell cycle defects and die in later stages of the life cycle (Gissot et al. 2005). Only one PREBP protein has been characterized, and it is thought to activate transcription during the later stages of IDC (Komaki-Yasuda et al. 2013).

1.8 EPIGENOME OF *P. FALCIPARUM*

1.8.1 The importance of epigenetic regulation

Though there seems to be a shortage of STFs in *P. falciparum*, it does contain a rich set of chromatin-related transcription factors (CTFs) (Volz et al. 2010; Aravind et al. 2003; Iyer et al.

2008; Deitsch 2005) (see also Table). CTFs include proteins involved in nucleosome structure (core histones and their variants), proteins involved in nucleosome/chromatin compactness and proteins involved in the histone code (chromatin modifying proteins). *Plasmodium* contains histone acetyltransferases (HATs), histone deacetylases (HDACs), histone demethylases (HDMTs) and histone methyltransferases (HMTs) involved in epigenetic regulation (Joshi et al. 1999; Carrozza et al. 2003; Fan et al. 2004; Sautel et al. 2007; Cui et al. 2008; Cui & Miao 2010; Bischoff & Vaquero 2010; Volz et al. 2012; Chahal et al. 2010; Mukherjee et al. 2008; Tonkin et al. 2009; Freitas et al. 2005; Duraisingh et al. 2005; Merrick et al. 2010; Miao et al. 2010). In addition, the genome contains proteins involved in remodeling of chromatin and proteins that read the marks laid down by HATs and HMTs. These remodeling or effector proteins include SWI2/SNF2 ATPases, proteins with bromodomains and chromo domains, Royal superfamily of proteins, and 14-3-3 proteins (Hossain et al. 2010; Cui & Miao 2010; Duffy et al. 2012; Flueck et al. 2009) (see also Table 1.1).

Gene ID	Product description
PF3D7_1142700	methyltransferase, putative
PF3D7_0818200	14-3-3 protein (14-3-3I)
PF3D7_1362100	14-3-3 protein (14-3-3II)
PF3D7_1422900	14-3-3 protein, putative
PF3D7_1323300	acetyltransferase, GNAT family, putative
PF3D7_1033700	bromodomain protein 1 (BDP1)
PF3D7_1212900	bromodomain protein 2, putative (BDP2)
PF3D7_0110500	bromodomain protein, putative
PF3D7_1234100	bromodomain protein, putative
PF3D7_1475600	bromodomain protein, putative
PF3D7_0727300	DNA (cytosine-5)-methyltransferase (DNMT)
PF3D7_1220900	heterochromatin protein 1 (HP1)
PF3D7_1118600	histone acetyltransferase (MYST)
PF3D7_0823300	histone acetyltransferase GCN5 (GCN5)
PF3D7_0809500	histone acetyltransferase subunit NuA4, putative
PF3D7_1020700	histone acetyltransferase, putative
PF3D7_0416400	histone acetyltransferase, putative (HAT1)
PF3D7_0925700	histone deacetylase 1 (HDAC1)
PF3D7_1008000	histone deacetylase 2 (HDA2)
PF3D7_0711400	histone deacetylase complex subunit SAP18, putative (SAP18)
PF3D7_1472200	histone deacetylase, putative (HDA1)
PF3D7_0506600	histone deacetylase, putative, pseudogene
PF3D7_0617800	histone H2A (H2A)
PF3D7_0320900	histone H2A variant, putative (H2A.Z)
PF3D7_1105100	histone H2B (H2B)
PF3D7_0714000	histone H2B variant (H2B.Z)
PF3D7_0610400	histone H3 (H3)
PF3D7_0617900	histone H3 variant, putative (H3.3)
PF3D7_1105000	histone H4 (H4)
PF3D7_1227800	histone S-adenosyl methyltransferase, putative
PF3D7_0811500	histone-arginine methyltransferase CARM1, putative (CARM1)
PF3D7_1221000	histone-lysine N-methyltransferase, H3 lysine-4 specific (SET10)
PF3D7_1214200	histone-lysine N-methyltransferase, putative (SET5)
PF3D7_1355300	histone-lysine N-methyltransferase, putative (SET6)
PF3D7_0809900	JmjC domain containing protein (JmjC1)
PF3D7_0602800	JmjC domain containing protein (JmjC2)
PF3D7_1211600	lysine-specific histone demethylase 1, putative (LSD1)

PF3D7_0916600	methyltransferase, putative
PF3D7_1437000	N-acetyltransferase, putative
PF3D7_1426200	protein arginine N-methyltransferase 1 (PRMT1)
PF3D7_1361000	protein arginine N-methyltransferase 5, putative (PRMT5)
PF3D7_0629700	SET domain protein, putative (SET1)
PF3D7_0827800	SET domain protein, putative (SET3)
PF3D7_0910000	SET domain protein, putative (SET4)
PF3D7_1115200	SET domain protein, putative (SET7)
PF3D7_0403900	SET domain protein, putative (SET8)
PF3D7_0508100	SET domain protein, putative (SET9)
PF3D7_1328800	transcriptional regulatory protein sir2a (SIR2A)
PF3D7_1451400	transcriptional regulatory protein sir2b (Sir2B)
PF3D7_1322100	variant-silencing SET protein (SETvs)

Table 1.1: Table was tabulated using information from plasmodb.

This table is not exhaustive and shows some of the annotated chromatin modifying proteins in the *P. falciparum* genome. These include HATS, SET proteins, histone acetyltransferases, bromodomain-containing proteins etc.

Evidence for chromatin-mediated epigenetic mechanisms playing a significant role in gene control is supported by a study carried out by Chahal and colleagues. In this study, inhibition of HDACs (which leads to hyperacetylation) by apicidin, a molecule with anti-proliferative properties, led to the transcriptional deregulation of 59% ring genes, 33.8% trophozoite and 51.5% schizont genes (Chahal et al. 2010). This deregulation was a result of disrupting the histone modifications H3K9ac, H4K8ac, H4K3me3, and H4ac. In a similar study, treatment of parasites with TSA, SAHA, and 2-ASA-9, all HDAC inhibitors, led to an overall altered gene expression of between 2–21% of genes in trophozoite stage parasites after 2 hours of drug exposure (Andrews et al. 2012). The prolonged exposure led to parasite death but when the same parasites were washed and returned back to culture, 70% of altered genes resumed their normal gene expression dynamics, suggesting a very transient system of gene expression (Andrews et al.

2012). Other studies also show how H3K9ac and acetylation levels are important in modulation gene expression (Cui et al. 2007; Cui et al. 2008; Srivastava et al. 2014). These studies suggest that epigenetic mechanisms play a substantial role in *P. falciparum* gene control.

1.8.2 Chromatin-mediated epigenetic mechanism

Chromatin is a physiological substrate for a number of cellular events including DNA replication, repair, and transcription. It is a dynamic structure and can exist in two forms. The transcriptionally active form, termed euchromatin, allows for proteins involved in transcription to access the DNA template (Groth et al. 2007). The transcriptionally silent form termed heterochromatin is more compact and less accessible to DNA-binding proteins. *P. falciparum* chromosomes have been shown to have both euchromatin and heterochromatin regions (Stockdale et al. 2008). The transition between the two forms is mediated by molecular mechanisms like histone replacement, reversible histone modifications, reversible DNA methylation and subnuclear localization (Kouzarides 2007). Recently by LC/ES-MS, it was thought that *P. falciparum* gDNA lacked DNA methylation (Choi et al. 2006) but current studies do show the presence of DNA methylation (Templeton et al. 2004; Ponts et al. 2013; Pollack et al. 1991).

The functional/basic unit of chromatin is a nucleosome, which is made up of 155bp DNA wrapped twice around 8 core histones; 2 H2A, 2 H2B, 2 H3 and 2H4 (Luger et al. 1997). Histone fold domains maintain histone/histone and histone/DNA interactions and additional, well-ordered structure elements extending from this motif. H1 histone further stabilizes the structure. The *P. falciparum* genome codes for these core conserved histones as well as four variant histones H2A.Z, H2B.v, H3.3, and CenH3 (Miao et al. 2006; Trelle et al. 2009; Bischoff &

Vaquero 2010). The nucleosome is one of the most stable structures *in vivo* because of numerous stabilizing interactions between the histones and the DNA. The nucleosome structure repeats to form a 30nm fiber (a higher order structure), with the aid of H1 protein. The 30nm fiber is further packaged into metaphase chromosomes by chaperones and scaffold proteins (Li et al. 2007). *P. falciparum*' genome, however, does not have the gene coding for H1 protein and it is thought that the chromatin is generally more relaxed (Coleman & Duraisingh 2008; Parseghian & Hamkalo 2001).

1.8.3 Histone modifications and their function

With the discovery of nucleosomes came the observation that they could be repressive for transcription (Knezetic et al. 1986; Lorch et al. 1987; Han et al. 1988; Laybourn & Kadonaga 1991). This was a logical conclusion because processes that used DNA as a template had to overcome the barrier presented by the nucleosome structure. Early experiments using cell-free systems to compare the rate of RNA polII transcription between naked DNA and nucleosomal DNA showed a reduction or complete inhibition of transcription in nucleosomal DNA (Knezetic & Luse 1986; Lorch et al. 1987). However, removal of nucleosomes from the DNA restored transcription (Han et al. 1988; Han et al. 1988) This suggested the existence of a mechanism *in vivo* that can disrupt histone-DNA interactions to allow the transcription machinery access to DNA and initiate transcription.

Work by Allfrey and colleagues highlighted the role of histone modifications in regulating chromatin structure (Allfrey et al. 1964). In their experiments, they observed that when they added acetic anhydride to histones in the context of nucleosomes, the rate of transcription became comparable to that of naked DNA i.e. there was less inhibition (Allfrey et al. 1964). The more acetylated the histones, the higher was the rate of transcription. In another experiment

using 15-day chicken erythrocytes, Hebbes et al. observed that the active α -d *globin* locus always precipitated with acetylated active chromatin (Hebbes et al. 1994). These and other early experiments (Sealy & Chalkley 1978), suggested that histone modifications (Table 1.2) could alter chromatin structure to favor transcription.

Histone modifications can act by changing the net charge of nucleosomes, thereby loosening inter-intranucleosomal interactions and inhibiting formation of higher order structures. This has been observed *in vitro* for H4K16ace modification (Hong et al. 1993; Shogren-Knaak et al. 2006). The change in net charge possibly destabilizes the nucleosome allowing its eviction from the template during RNA polymerase II transcription (Li et al. 2007).

Histone	PTM(s)
H2A	N-term-ac, K3ac, K5ac, K9ac, K119ub
H2B	K5me1, K12ac, S14, K15ac, K16ac, K20ac, K46ac, K120ac, K120ub, K112ub,
H3	K4me1, K4me2, K4me3, T3, K9ac, K9me, K9me3, K14ac, K14me, R2me, R8me, R17me, R17me2, K18ac, K23ac, R26me, K27ac, K27me1, K27me2, K27me3, S28, K36me1, K36me2, K36me3, K56ac, K79me1, K79me2, K79me3, K91ac
H4	N-term-ac, S1, R3me, R3me2, K5me, K5ac, K8ac, K12ac, K12me, K16ac, R17me, K20me1, 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

Table 1.2: Histone post translational modifications identified and confirmed in eukaryotes.

The black colored ones have been confirmed in *P. falciparum*: K-lysine. Ac- acetylation, me1, 2, 3- mono, di, tri methylation, ph- phosphorylation, R –arginine, un- ubiquitination, digit- position modified. (Table modified from Cui & Miao 2010)

Modifications also act by recruiting specific ‘effector’ proteins; the ‘effector’ proteins recruit or stabilize chromatin-related proteins facilitating downstream events such as transcription.

Examples of effector proteins include ATPase complexes such as SWI/SNF, SAGA/SLIK that use ATP to remodel chromatin (Côté et al. 1994; Grant et al. 1997; Deindl et al. 2013; Clapier & Cairns 2009; (Carlson & Laurent 1994; Burns & Peterson 1997) Narlikar et al. 2013). They also include proteins such as GCN5 (Kuo et al. 1996; Grant et al. 1997), P/CAF (Schiltz et al. 1999), p300/CBP (Bannister & Kouzarides 1996; Ogryzko et al. 1996; Yang et al. 1996), Bdf1 (Lygerou et al. 1994), all having a bromodomain that binds to acetylated residues. Bdf1 can bind and recruit TFIID to promoter regions (Duffy et al. 2014) whereas p300/CBP is unique in that it has multiple other domains including a domain that binds nuclear hormone receptors, a domain that recruits transcription factors such as cFOS, PCAF, TFIIB and the TATA box binding protein and RNA polII to the promoters of genes (Turner 2007).

1.8.4 *P. falciparum* modifications associated with transcription activation

Mass spectrometry of *P. falciparum* histones show that the genome has more active marks than silent ones and the most abundant are methylation and acetylation (Table 1.2) (Trelle et al. 2009; Salcedo-Amaya et al. 2009; Gupta et al. 2013), much like the genome of unicellular eukaryotes like yeast (Garcia et al. 2007). The genome is largely constitutively acetylated (Miao et al. 2006). An earlier study investigating the pattern of modifications during the maturation of the parasite from rings to schizonts showed that during the ring stages the ‘active’ marks, H3K9ac and H3K4me3, are distributed evenly across the genes with minimal increments on active as opposed to inactive genes (Salcedo-Amaya et al. 2009). During the trophozoite to schizont stage, however, H3K9ac is specifically enriched in the 5’ ends of active genes and correlates with gene activity (Cui et al. 2007; Salcedo-Amaya et al. 2009; Miao et al. 2006). This suggests that H3K9ac is stage specific, highly dynamic and tightly regulated. H3K4me3, on the other hand, follows the

same profile as H3K9ac, peaking during the trophozoite stage but later during the schizont stages, its enriched on both silent and active genes in the euchromatic regions (Bártfai et al. 2010).

A comprehensive study by Gupta et al. investigating 12 (H3 and H4) acetylation and methylation histone modifications gives further evidence for the role of epigenetic regulation in *P. falciparum* (Gupta et al. 2013). In this study more than 80% of the parasites genome is enriched in H4K5ac, H4K12ac, H3K14ac, H4K8ac, H3K4me3, H3K56ac, and H3K9ac, suggesting a very euchromatic organization as also reported by other studies (Bártfai et al. 2010; Karmodiya et al. 2015). H4R3me2, H4K20me1 and H4K16ac associate with 65-80% of the genome and H4ac, H4K20me3 and H3K79me3 with less than 60% of the genome(Gupta et al. 2013).

8 of these modifications, the acetylations minus H4ac and H4K12ac, plus H3K4me3 and H4K20me1 correlated well with transcription. The modifications associating with transcription were enriched in the 5' UTR of genes (H4K8ac) or in the 5' termini of coding regions with the enrichment more pronounced in the trophozoite to schizont stages (Gupta et al. 2013). This is in agreement with their role as transcription inducers, as during the trophozoite stage the cells are undergoing DNA replication and massive transcription events. The other four marks H4K20me3, H4R3me2, H4K12ac and H3K79me3 showed no association with transcription (Gupta et al. 2013). Karmodiya and colleagues using ChIP-seq coupled to RNA-seq, further dissected these modifications and they show that active histone modifications including H3K4me2, H3K4me3, H3K9ac and H3K14ac (active marks) cluster together possibly indicating that they exert a cumulative action presumably for efficient gene expression. Based on RNA-seq, results they also grouped H4ac, H4K20me3, H3K79me3, H3K27ac, H3K36me3, H3K4me1 modifications and the histone variant H2A.Z,as marks for active and poised chromatin in *P. falciparum* (Karmodiya et al. 2015).

1.8.5 *P. falciparum* modifications associated with transcription repression

Repressive marks occupy defined positions in *P. falciparum*. H3K9me3 and H3K36me3 mark heterochromatin regions. These modifications have a defined distribution in the subtelomeric regions of the chromosomes and in specific intrachromosomal locations where variant antigenic families are located (Salcedo-Amaya et al. 2009; Lopez-Rubio et al. 2009; Flueck et al. 2009; Jiang et al. 2013). The presence of H3K9me3 in the subtelomeric regions keeps this region in a facultative heterochromatic state via recruitment of the heterochromatin protein 1 (HP1) (Flueck et al. 2009; Pérez-Toledo et al. 2009). HP1 has a chromodomain which recognizes and binds specifically to H3K9me3. It forms homodimers, enabling it to aggregate methylated nucleosomes and thus form heterochromatin (Pérez-Toledo et al. 2009). H4K20me3, another silencing marker, shows different genome distribution as compared to H3K9me3. It does not have a sharp and defined localization as H3K9me3 or H3K36me3; instead, it is more ubiquitously distributed but is absent from telomeric repeats (Lopez-Rubio et al. 2009). The exact role of H4K20me3 in *P. falciparum* has not been well studied.

In a genome-wide study by Karimodiyana and colleagues, they propose H3K36me2 as a global transcription repressor for *P. falciparum* (Karimodiyana et al. 2015). When they did a correlation between H3K36me2 and gene expression, they observed a negative correlation of -0.79 correlation, suggesting that this modification is a repressor of gene activity (Karimodiyana et al. 2015). They then compared the histone occupancy of H3K36me2 in the wild type 3D7 against an H3K36 methyltransferase knockout (KO) line (Jiang et al. 2013). The least expressed genes with the highest H3K3me2 occupancy in the wild type showed high up-regulation in the KO line. Interestingly the group that actually created the KO line, do not mention anything about

H3K36me₂, but only to state that it was not involved in repression of variant surface antigens (*var*, *rifin*, and *stevor*) (Jiang et al. 2013). Perhaps this was because in their work they focused more on H3K36 trimethylation and its repression of *var* genes (Jiang et al. 2013) instead of the global transcription.

1.8.6 Other elements involved in transcriptional repression in *P. falciparum*

The nuclear organization of *P. falciparum* is such that it has active and repressive compartments denoted by different histone marks (Issar et al. 2009). The nuclear periphery is repressive in nature because this is where the heterochromatic regions of the chromosome telomeric ends cluster (Duraisingh et al. 2005; Freitas-Junior et al. 2000; Scherf et al. 2001). *P. falciparum* chromosome ends have a 20-50kB heterochromatic non-coding region comprised of six different repetitive polymorphic elements termed, telomere-associated repeat elements 1 to 6 (TARE 1-6). The histone deacetylase enzymes (HDAC1-3) and *P. falciparum* silent information regulator 2 (PfSir2) A and B, are required for maintenance and silencing (by promoting chromatin compaction) of this heterochromatin region (Duraisingh et al. 2005; Tonkin et al. 2009; Freitas et al. 2005). Sir2 to this end maintains a histone hypoacetylation gradient spanning 50kB from the telomere end into the chromosome coding region, which promotes heterochromatin formation (Freitas et al. 2005; Stockdale et al. 2008). Deletion of SIR2 leads to hyperacetylation of the region and transcription of more than one *var*. The affected *var*, are type A *var*. In these KO there is an increased expression of typeA1 *rifins* (Freitas et al. 2005)

1.8.7 Histone variants

Histone variants are slightly more unstable compared to the core histones. Their incorporation into nucleosomes reduces the nucleosomes stability creating a transcription-permissive

environment (Raisner & Madhani 2006; Li et al. 2007). *P. falciparum* has 4 histone variants, H2A.Z, H2B.Z, H3.3, and CenH3 (Gardner et al. 2013; Miao et al. 2006). H2A.Z and H2B.Z have been the most studied ones. Bartfai & Hoeijmakers et al. (2010) using an antibody raised against the N-terminus of the H2A.Z protein, performed chromatin immunoprecipitation experiments which showed that H2A.Z was preferentially and stably enriched in euchromatic intergenic regions throughout the IDC (Bartfai et al. 2010). When compared to the canonical H2A histone, this was found to enrich in the coding regions, showing a different enrichment pattern between these two histones. Interestingly enrichment of H2A.Z in the intergenic regions co-localized perfectly with H3K4me3 and H3K9ac modifications suggesting that perhaps the H3K4me3 and H3K9ac modifications are actually deposited on this histone variant (Bartfai et al. 2010). An expansion of this study looking at the H2B.Z variant, showed a similar enrichment pattern as H2A.Z (Hoeijmakers et al. 2013; Petter et al. 2013) suggesting that these two variants probably pair up to form a nucleosome subtype as both histones co-localize in the promoter regions as well.

A unique feature of *P. falciparum* genome is that the intergenic regions can be up to 80% AT-rich (Gardner et al. 2013). The fact that *P. falciparum* intergenic regions are enriched in H2A.Z and H2B.Z histone variants has led to the hypothesis that perhaps during evolution, the AT-rich genome of the parasite shaped the nucleosome landscape to preferentially enrich for these histones. In other eukaryotes, H2A.Z is enriched on nucleosomes positioned only around the nucleosome-free region surrounding the transcriptional start site and not the 1.5kb of intergenic regions as seen in *P. falciparum*. An interesting observation comes from these studies. The enrichment of H2A.Z and H2B.Z in the intergenic regions coincides with the euchromatic nature of the *P. falciparum* genome. This suggests that these less stable histone variants help in

maintaining this open chromatin structure allowing for transcription as they carry gene activation permissive marks (Hoeijmakers et al. 2013; Trelle et al. 2009).

1.8.8 Nucleosome density and regulation of transcription

Studies on the nucleosomal landscape of *P. falciparum* have shown conflicting results. Genome-wide nucleosome mapping studies done by Westenberger et al. suggest that the coding regions of genes are nucleosome rich and the telomeric regions have the highest nucleosome density, consistent with the more condensed nature of telomeres. Conversely, the intergenic regions of *P. falciparum* are nucleosome-deficient and therefore less dense. This is different from eukaryote models such as yeast which have a nucleosome-free region (NFR) around the transcription start site (TSS) region but condensed intergenic regions (Westenberger et al. 2009). Their study goes to suggest that the unique nature of *P. falciparum* has perhaps hindered mapping of TSS using nucleosome position techniques as there is no defined nucleosome-free region in the intergenic region as in other eukaryotes. Moreover, in other eukaryotes, the TSS has strongly positioned nucleosomes (-1 and +1 nucleosomes) around it with downstream nucleosomes showing less pronounced nucleosome positioning. However, in *P. falciparum* strongly positioned nucleosomes are not found around the TSS but rather at the start and end of the coding regions. The difference in intergenic nucleosome content between *P. falciparum* and other eukaryotes perhaps can be attributed to *P. falciparum* high AT content in these regions. However, these results need to be evaluated in light of another study, which reported a comparable nucleosomal occupancy between the intergenic and coding regions of *P. falciparum* using an optimized method which accounts for the highly AT-rich intergenic regions, (Bartfai et al. 2010). In light of this data, the nucleosomal landscape of *P. falciparum* might have to be reviewed.

Whether the nucleosomal landscape is similar or not, it appears it still affects transcription significantly. For example, genes involved in basal transcription machinery, tRNA synthesis, mRNA splicing and ribosomal maturation have low nucleosome occupancy (Westenberger et al. 2009). During the trophozoite stage where maximal transcription happens the nucleosome occupancy is 2 fold lower in coding and intergenic regions and there is increased abundance of RNA polII as evidenced by mass spectrometry (Bunnick et al. 2014). There is a strong inverse correlation between the levels of nucleosomes in the promoter regions of highly expressed genes, with highly expressed genes having open chromatin organization. Progression to the schizont stage leads to nucleosome repackaging of the genome in preparation of parasite egress and re-invasion (Bunnick et al. 2014). Overall, the evidence suggests that perhaps due to a lack of a variety of STFs, the parasite uses nucleosome occupancy levels, histone modifications, and histone variants to fine-tune the level of transcription of genes by carefully regulating the accessibility of promoter regions to general transcription factors and the RNA polymerase II machinery.

1.9 TRANSCRIPTIONAL ACTIVATION OF MULTIGENE FAMILIES

1.9.1 Co-regulation of multigene families

It has been suggested that *var*, *rifin*, *stevor* and *Pfmc-2TM* family members are co-regulated and share common activation factors (Howitt et al. 2009). This assumption is natural considering that the families are located in the same regions and are activated and silenced in a clonally variant manner.

As for sharing a common activation factor, when Howitt and colleagues expressed episomal plasmids with multiple *var* gene promoters it led to downregulation of *rifin*, *stevor* and even *Pfmc-2TM* (Howitt et al. 2009). In this study, the downregulation was strongest for *upsAvar* subtypes and *rifinA* possibly because these genes lie in head to head orientation and selection of *upsA var* genes lead to changes in *rifinA* gene. In their DNA-FISH experiments, active *var* promoters and active *rifin* promoters co-localize at a subnuclear expression site (Howitt et al. 2009). PfSir2A regulates subtelomeric *upsAvar* genes (Tonkin et al. 2009) and parasite lines where PfSir2A is knocked down show an increase in *upsAvar* and *rifinA* transcription (Duraisingh et al. 2005). Thus, it seems *upsAvar* and *rifinA* may be co-regulated. Whether this translates to other types of *rifin* genes is unknown but unlikely (Witmer et al. 2012).

In contrast to the *upsAvar* and *rifinA* possible link, a study by Sharp et al. which investigated transcription co-regulation of *var* and *stevor* showed no co-regulation. When *var* genes were selected for binding phenotypes, the same *stevor* were expressed before and after selection even though the *var* gene expression shifted (Sharp et al. 2006) suggesting *var* expression does not influence *stevor* expression. Cabral et al. 2009 also showed how *rifins* were not regulated with *var* genes. Now we know that only a subset does co-regulate with *var* genes. These *rifins* which do are positioned in a head to head orientation with *upsAvar* genes. So co-regulation seems to be limited to *upsAvar* and of *rifinA* genes.

Witmer et al. also looked at the possible co-regulation of these families and confirmed the non-co-regulation of the gene families except possibly for *upsAvar* and *rifinA* genes. Their study also investigated whether the *rifin*, *stevor*, and *pfmc-2TM* gene promoters could be involved in controlling mutual exclusion like the *var* promoters. The conclusion of their study was that, only the *var* promoters had this intrinsic function and that other gene families probably used a

different mechanism for their regulation (Witmer et al. 2012). This confirmed studies in the past also looking at *var* promoters (Voss et al. 2006; Gannoun-Zaki et al. 2005; Vázquez-Macías et al. 2002). Kyes et al. also show that *var* genes are potentially regulated at the transcription initiation stage (Kyes et al. 2007) and the *cis*-elements here are crucial for the regulation. Unlike the Howitt et al. study, Witmer et al. had another construct with a calmodulin promoter. In the parasites, transfected with this construct, they also observed a down-regulation of the multigene families, suggesting that just the presence of drug led to a fitness mechanism which led to a shutting down of the virulence genes as they are not needed in an *in vitro* culture (Witmer et al. 2012). So, in contrast to the Howitt's study, this extra control plasmid suggests that what had been observed in the Howitt's study was a drug effect.

1.9.2 What is known: *var* and *rifin*

Most studies focusing on gene regulation of *P. falciparum* have focused on the *var* gene family and the mechanism leading to allelic exclusion. A complex network of regulatory elements involving, the *var* intron (Calderwood et al. 2003; Gannoun-Zaki et al. 2005) *cis* elements, histone modifications and epigenetic memory (Dzikowski et al. 2007; Chookajorn et al. 2007; Chookajorn et al. 2008; Lopez-Rubio et al. 2007; S. Kyes et al. 2007) chromatin remodeling (Duraisingh et al. 2005; Freitas et al. 2005; Zhang et al. 2011) and long non-coding RNAs (Amit-Avraham et al. 2015; Sierra-Miranda et al. 2012; Bright & Winzeler 2011). The important role played by the *var* promoter in controlling allelic exclusion was shown by Voss et al. In their transfections experiments, a *var* transgene under drug pressure was able to impact on the allelic exclusion mechanism leading to expression of the transgene and down-regulation of endogenous *var* genes. Upon removal of the drug, however, the transgene was switched off. This suggested that the *var* promoter is able to mediate nucleation and spreading of stably inherited

heterochromatin (Voss et al. 2006). They also confirmed work by Ralph et al. which highlighted the role of chromatin remodeling in the removal of activated *var* from a repressive environment to a transcriptional permissive environment (Ralph & Scherf 2005).

Brancucci et al. expanded on this earlier work by identifying *cis*-elements involved in *var* gene regulation. These *cis* elements are in the conserved promoter sequences of the *var* genes and include an upstream activation sequence (UAS) located -1401 to -1217bp upstream of the TSS, which when deleted inhibit transcriptions (Brancucci et al. 2012). In the same study they also identified a conserved 101bp mutual exclusive element (MEE) in 40 of the 60 *var* genes, which they hypothesize is bound by an unknown regulatory factor and is involved in singular *var* gene expression because a deletion of this region or the MEE abolished the competition between the transgene with endogenous *var* genes (Brancucci et al. 2012). In their model, interactions between the site where active *var* genes localize and the MEE mediates the exchange of the repressive histone marks with active histone marks. This suggests that histone modifications are potentially the last layer of regulation, as they modulate chromatin such that the non-expressed *var* genes are not accessible to DNA-binding factors necessary for transcription.

Tham et al. also identified the TSS of *rifin* genes located in the functional promoter region (Tham et al. 2007). The researchers also identified two repressive *cis* elements CGCACAACAC (-870 to -860) and TATGCatgATT (-546 to -535) which are bound by nuclear protein factors expressed in different stages of the IDC (Tham et al. 2007). They hypothesize that these *cis* elements are involved in differential repression of *rifin* and could repress *rifin* transcription early

during the ring stage and later after the trophozoite stage. BLAST analysis revealed that these repressive elements are associated with *rifin* genes.

As already alluded to, *var* and *rifin* promoters are conserved and are all capable of being actively transcribed. The fact that not all the genes in the gene families are transcribed, suggests another mechanism other than *cis* elements is at play. Histone modifications have a potential to modulate chromatin as previously stated which in turn has a role in regulating mono-allelic expression of multigene families in other organisms; *vsgs* of *trypanosomes* (Sullivan et al. 2006; Cross et al. 1998), mating type loci of *saccharomyces cerevisiae* (Rine et al. 1979; Haber 1998; Huang 2002), mammalian immunoglobulin genes (Xu & Feeney 2009; Sakamoto et al. 2012; Liu et al. 2007; Bevington & Boyes 2013) .

In *P. falciparum* histone modifications, do regulate expression of gene families. Lopez-Rubio et al. mapped the epigenetic marks on an active *var* gene (*var2csa*) using chromatin immunoprecipitation (ChIP) technique (Figure 1.7) and highlighted the importance of histone modifications in *var* gene activation (Lopez-Rubio et al. 2007). The *var2csa* gene is structurally distinct from other *var* genes (Vázquez-Macías et al. 2002; Salanti et al. 2003) and generalization to other *vars*, should be made with care. The ChIP experiments showed that stable silent *var* genes are marked by H3K9me3 spreading from the 5' region to the coding regions during the whole IDC (Lopez-Rubio et al. 2007) (Lopez-Rubio et al. 2007).

During the ring stage, the active *var* gene loses the 5' H3K9me3 mark and is enriched for H3K9ac, H3K9me2, and H3K4me3 in the 5' region.

Figure 1.7: Diagram highlighting the modifications associated with silent and active *var* genes.

The 5' flanking regions of *var* genes seems to be important for mutual exclusive expression. *Var* genes are switched on in the ring stage and the active *var* is decorated with modifications at histone 3 K9 and K4 residues at its 5' flanking region. In this state, the H3K9me3 marker is removed from both the TSS regions and exon 1. H3K4me2 marker seems to mark the *var* to be transcribed in the next cycle i.e. poised state. Silent *var* genes are kept in the repressed state by the H3K9me3 mark which covers both the 5' flanking region, as well as the coding region (Figure from Lopez-Rubio *et al* 2009- permission granted).

During the trophozoite and schizont stages, when *var* expression has been repressed, the H3K4me3 and H3K9ac are lost but the H3K4me2 is maintained, presumably as a mark for reactivation of this gene (Lopez-Rubio *et al.* 2007). This was an expansion of earlier work done by Chookajorn *et al.* in which they showed H3K9me3 enriched in the coding region of a silent drug selectable cassette downstream of a *var* promoter (Chookajorn *et al.* 2007). In its active state, the coding region was devoid of the H3K9me3 mark. Thus Lopez-Rubio's work extended the study to include the upstream region of the *var* genes. H3K36me3 (Jiang *et al.* 2013) and H4ac (Miao *et al.* 2010) have also been shown to regulate *var* gene expression.

Sir2A protein also seems to have a crucial role in silencing of telomere located *var* genes as the protein associates with *var* genes in the promoter region (Freitas et al. 2005). Activated *var* genes lose association with PfSir2 allowing the genes to be acetylated in their promoter regions. DNA fluorescence *in situ* hybridization (FISH) experiments probing for active and inactive *var* genes and studies using transgenic parasites with one or multiple active *var* promoters showed that active promoters localized at a specific site in the nuclear periphery (termed *var* expression site; VES) (Dzikowski et al. 2007; Chookajorn et al. 2007; Voss et al. 2006) separate from the silent *var* gene cluster. Recently, Zhang et al. implicated the cytoskeleton and filamentous (F) actin as having a crucial role in repositioning the active *var* loci from the nuclear periphery to VES and also in switching of *var* genes (Zhang et al. 2011).

Studies have also been carried out on the *rifin* gene family to map the histone code associated with active and silent *rifin* genes. Earlier ChIP-chip work showed that silent *rifin* are enriched with H3K9me3 (Salcedo-Amaya et al. 2009; Lopez-Rubio et al. 2009) and that Sir2A was involved in silencing and activation of *rifin* genes (Freitas et al. 2005). This work was further continued using ChIP experiments coupled with quantitative (q) PCR. This work showed that active *rifin* had H3K9ac and silent *rifin* genes were enriched with H3K9me3 with some silent *rifin* having no modifications (Cabral et al. 2012). However, in this study, due to technical difficulties with designing specific primers in the 5' UTR regions of the 6 *rifins* studied (out of 160), they only focused their mapping to the coding regions. Currently, no studies have been completed for *stevor*, in terms of studying its epigenetic footprint, or whether any *cis* elements are responsible for its regulation. Determining whether the footprint is the same as seen for *var* or *rifin* is important to understand the regulation of these genes.

AIMS

The broad aim of this project was to study the regulation of the *stevor* variant antigen family. The transcriptional regulation of *stevors* is unknown and yet they are important in pathology as well as in immune evasion. As yet, there are no tools available to help in studies of these genes. This is unlike the *var* gene family. Specific *var* gene expressing populations can be selected for by use of panning assays as the receptors for most *var* genes are known (Scherf et al. 1998). This system is not available for *stevor*, making studies in this field challenging.

1) Develop targeted tools to study epigenetic marks of active and silent *stevors*

This part of the work used a more targeted approach to study *stevor* regulation. A transgenic line was used which can allow manipulation of *stevor* expression, allowing us to have a line where a known *stevor* is active ‘on’ and that same *stevor* is ‘off’ in the other line. ChIP-qPCR was then performed on chromatin precipitated from the ‘on’ and ‘off’ cultures and enrichment compared (Chapters 3A and 3B).

2) Study *stevors* in their natural expression profiles

Clonal lines expressing different *stevor* were used for this work. By q-PCR, we monitored the expression levels of the 40 *stevors*. Clonal lines were then chosen for the on and off expression of the *stevors*. As in aim 1, we did chromatin immunoprecipitations looking at different modifications (Chapter 3C).

CHAPTER 2

MATERIALS AND METHODS

2.1 CELL CULTURE

2.1.1 Culture and synchronization

This method is used for all the parasites used in this study. Unless otherwise stated, cells were used in the trophozoite stage. Parasites were cultured in human red blood cells in RPMI 1640 media supplemented with 50mg/l gentamicin, 2g/l sodium bicarbonate, 0.25% Albumax II, and 0.1 mM hypoxanthine. For healthy parasites, the cultures were maintained at a low parasitemia of 5-6% at 37°C. To synchronize the cells, the late stage parasites were separated by 68% percoll gradient. The top layer with late stage parasites was washed and returned to culture and monitored for ring appearance. At ring appearance, the parasites were treated with 5% D-Sorbitol for 10 minutes to kill any late stage parasites that had not yet ruptured. Another sorbitol treatment was repeated in next cycle for tight synchronization of cells. Cultures were harvested when they reached the trophozoite stage. Parasite stages and parasitemia were checked by thin blood smears stained with Giemsa and viewed under the light microscope.

2.2 GENERATION OF TRANSGENIC PARASITES

2.2.1 DNA/vector constructs

P. falciparum 3D7 strain parasites were used for transfections. pCC1 and pCC4 vectors were used as templates for cloning (A kind gift from Prof. Cowman). 10µg pCC1 plasmid was

3D7_1040200 5' F (1.5kb)	(SpeI) GATC <u>cactagt</u> ATGTGAAATATATTATATTATGTTAATTATTTTATCTTTT
3D7_1040200 5' R (1.5kb)	(ApaI) TATA <u>gggccc</u> TTTTAATTCAAAAAATAATATATTATATTCAAATTG
3D7_1040200 3' F (1kb)	(EcoRI) ACGT <u>gaattc</u> TCGAAATTTTAGAATATATATGATGTTACTTAAG
3D7_1040200 3' R (1kb)	(NcoI) TACT <u>ccatgg</u> TAAAATTTTCATAAATTGTGTTAAAATTTTG
3D7_1149900 5' F (1.5kb)	(SpeI) GATC <u>cactagt</u> ATTTATTTTTTCCCTTTTAAATAAAATATGAA
3D7_1149900 5' R (1.5kb)	(ApaI) TATA <u>gggccc</u> TTTTAATTCAAAAAATAATATGTTATATTCAAATTG
3D7_1149900 3' F (1kb)	(HindIII) GGCT <u>aagctt</u> TATATTTTTGAAAATATGAATTATGTAATTTGT
3D7_1149900 3' R (1kb)	(EcoRI) ACGT <u>gaattc</u> GGAAAAGTTATAATATATATTATATGAATTGCATAT
3D7_1254100 5' F (1.2kb)	(SpeI) GATC <u>cactagt</u> TAAATTAACATAATCTTACCGTAAATGTTATTATAA
3D7_1254100 5' R (1.2kb)	(ApaI) TATA <u>gggccc</u> TTTTAATTCAAAAAATAATATGTTATATTCAAATTTTA
3D7_1254100 3' F (1kb)	HindIII) GGCT <u>aagctt</u> TGTATTTTAAATATATGAATTATATAAAGTTGAATT
3D7_1254100 3' R (1kb)	(NcoI) TACT <u>ccatgg</u> CATTAATATATTTTAACTTTTAAATATAAAAGAAAAGTAATTAT
3D7_0102100 5' F (1.5kb)	(SpeI) GATC <u>cactagt</u> CCTCTAAATTATCTAATACATTATATAATTCTTCCAG
3D7_0102100 5' R (1.5kb)	(ApaI) TATA <u>gggccc</u> CTAAAAAGATAAAGAGGAATATACATAAATATATATATATATATATAT
3D7_0102100 3' F (1kb)	(NcoI) TACT <u>ccatgg</u> TAAATATAAAATAGAAATAAAATTATAATAAAATATATATTTATATTTTATTTT
3D7_0102100 3' R (1kb)	(EcoRI) ACGT <u>gaattc</u> TGTATTACTTAAATATATGAAACATATTACTTTTAAATAATATAC
hDHFR F (0.567kb)	(ApaI) GTC <u>Agggccc</u> ATGCATGGTTCGCTAAAC
hDHFR R (0.567kb)	(EcoRI) GTC <u>Agaattc</u> TTAATCATTCTTCTCATATACTTCAAA
hDHFR R (0.567kb)	(HindIII) GGCT <u>aagctt</u> TTAATCATTCTTCTCATATACTTCAAA
hDHFR R (0.567kb)	(NcoI) TACT <u>ccatgg</u> TTAATCATTCTTCTCATATACTTCAAA
BSD F (0.399kb)	(ApaI)TTA <u>Agggccc</u> ATGCATAAGCCTTTGTCTCAAGAAGA
BSD R (0.399kb)	(HindIII)GGC <u>Caagc</u> TTAGCCCTCCCACACATAACCA
BSD R (0.399kb)	(EcoRI)GGCT <u>gaattc</u> TTAGCCCTCCCACACATAACCA

Table 2.1: Cloning primers used for all constructs in the PCC vectors.

The table lists all the primers used for cloning purposes. The restrictions sites are also included in small letter and underlined.

digested with SpeI and EcoRI or SpeI and NcoI (3D7_1040200 and 3D7_1149900 construct) (Fast Digest). Positive constructs from each cloning procedure were verified by restriction digest analysis and automated DNA sequencing (Axil and 1st base, Singapore).

PCR components

Component	Reaction volume 50µl
10X Buffer for KOD Hot Start DNA Polymerase	5 µl
25 mM MgSO ₄	3 µl
dNTPs (2 mM each)	5 µl
Forward Primer (10µM) and Reverse primer (10 µM)	1.5µl
Template DNA (50ng/µl)	1µl
KOD Hot Start DNA Polymerase (1 U/µl)	1µl
PCR Grade Water	33.5µl

PCR programme

Polymerase activation	95°C for 2 min	
Denature	95°C for 20s	} X 10
Annealing	56°C for 20s	
Extension	68°C for 20 s/kb	
Denature	95°C for 20s	} X 10
Annealing	51°C for 20s	
Extension	68°C for 20 s/kb	
Extension	68°C for 5 min	

Table 2.2: PCR parameters for PCC constructs.

For PCR using of *bsd* and *hdhfr* from plasmid vector, the template DNA was reduced to 20ng/µl and the total combined cycling time was 15cycles. For amplification of 1.5kb fragments use an annealing temperature of 49°C.

2.2.2 Cloning of inserts into vectors

PCR products were purified using the QIAquick PCR purification Kit (Qiagen, Germany). 1µg of the purified products was digested with appropriate enzymes using Fast Digest enzymes (Fermentas). Digested pCC1 plasmid was purified using Qiagen gel extraction kit and further

purified by MinElute PCR kit. The digested PCR products were also purified by the same kit. All products were eluted in 30µl of elution buffer.

Component	Reaction volume 50 µl
10X Fast Digest™ Buffer	5µl
PCR product	1-2µg (variable)
Fast Digest enzyme 1	2µl
Fast Digest enzyme 2	2µl
Water, nuclease-free	up to 50µl

5' UTR fragments were digested with SpeI/ApaI. 3' UTR fragments were digested with EcoRI/NcoI or HindIII/EcoRI for 3D7_1149900. *Bsd* and *hdhfr* were digested with ApaI/HindIII/EcoRI. These were all digested in a 50µl volume. Digestion of pCC1 digest was as above but in 100µl volume and 5µl of enzymes. Digest products were mixed gently and spun down for a few seconds. The reactions were incubated for 4 hours at 37°C in PCR machine.

2.2.3 Ligation

Fast ligase, ATP-dependant T4 DNA ligase (NEB) was used to ligate inserts first. The inserts included digested 5' UTR, 3' UTR, and drug cassette. The molar ratio for this was 1:1:1 for 5 mins at room temperature. An overnight incubation in cold room did not help with efficiency of ligation. The ligated 3 piece insert became the template for PCR using the following primers. The PCR was done using the 5' UTR F primer and the 3' UTR R primer. The PCR program was the same as used to amplify the 5' UTR regions with an extension of elongation time to 1min 20secs. The resulting product from the PCR would be digested with appropriate enzymes and purified. The digested product was then ligated into the linearized vector. Ligation reaction was in the molar ratio of 3:1, vector: insert for 5 minutes at room temperature in 10µl reaction volume

3D7_1040200 5' F	(SpeI) GATC <u>cactagt</u> ATGTGAAATATATTATATTATGTTAATTATTTTATCTTTT
3D7_1040200 3' R	(NcoI) TACT <u>ccatgg</u> TAAAATTTTCATAAATTGTGTTAAAATTTG
3D7_1149900 5' F	(SpeI) GATCactagtATTTATTTTTCCTTTTAAATAAAATATGAA
3D7_1149900 3' R	(EcoRI) ACGT <u>gaattc</u> GGAAAAGTTATAATATATATTATATGAATTGCATAT
3D7_1254100 5' F	(SpeI) GATC <u>cactagt</u> TAAATTAACATAATCTTACCGTAAATGTTATTATAA
3D7_1254100 3' R	(NcoI) TACT <u>ccatgg</u> CATTAATATATTTTAACTTTTAATATAAAAGAAAAGTAATT AT

Table 2.3 Primers used to amplify 3 way insert ligation.

*3D7_0102100 did not amplify therefore not included in the analysis. Restrictions site are shown in small letters and underlined.

2.2.4 Transformation of competent cells

Ligated insert: vector fragments (4µl) were transformed into 50µl of thawed TOP10 (*E.coli*) chemically competent cells and incubated on ice for 30minutes. For transformation, the bacteria cells were heat shocked at 42°C for 30secs and then incubated again on ice for 2 minutes. 1ml pre-warmed LB broth was added to the mixture and incubated for 1 hour with shaking at 37°C. 200µl cell suspension was plated on ampicillin plates and incubated overnight at 37°C. A negative plate with only the digested ligated with water was also plated.

2.2.5 Screening of positive colonies

Colonies from plates were picked and incubated overnight in 3 ml LB media (mini-prep). The plasmids from these miniprep cultures were extracted using Favorgen miniprep kit following manufacturer's protocol. Extracted plasmid was restriction digested using enzymes flanking insert (SpeI and EcoRI/NcoI). Colonies which showed expectant band sizes after gel run were

sent for sequencing to confirm sequence. Positive colonies with correct sequences were further grown to obtain high plasmid yield concentration and used for transfection (**Section 2.2.6**)

2.2.6 Transfection of *P. falciparum* and selection of transfectants

High concentrations of pure plasmid were obtained using EndoFree and Promega Maxiprep kits performed according to manufacturer's protocol. The quality of plasmid was checked by running on an agarose gel to make sure there was no gDNA contamination and that the plasmid was neither degraded nor smeary. Culture with predominant ring stage (at least 10%) parasites was used for transfection. The culture was spun down and the pellet used for transfection. 150µg of plasmid (dissolved in 100µl sterile water) was used for each ring stage transfection. To this, 2x cytomix was added to a final volume of 400µl. 200µl of the pellet culture was added to the plasmid/cytomix mixture and transferred to an electroporation cuvette (0.2cm). Electroporation parameters were 0.31kV, 950 µF and maximum capacity (Biorad, gene Pulser II). After electroporation, parasites were added back to prewarmed 12 ml culture with 400µl fresh RBC. 5nM WR99210 (Jacobus Pharmaceuticals) or 2.5µg/ml blasticidin (Invivogen) was added 6 hours after transfection. Transfectants were selected under drug pressure and were seen in blood smears after 31 days for the constructs with blasticidin. Transfections were repeated twice for *hdhfr* (WR99210) constructs. The second transfection was done 30 days after the initial transfection.

Negative selection was done with 5-FC drug (Invivogen) at a concentration of 1µM, 10µM, and 20µM. In the Maier et al. (2006) paper, 1µM was the optimum concentration when selecting with WR99210. I tried these three ranges as I was using blasticidin and the optimum concentration with blasticidin is not known. I also include the 1µM concentration in my assays.

The parasites were grown for a month on this drug before genomic DNA was extracted and PCRs for integration carried out.

2.2.7 Checking for integration

Integration was assessed by PCR analyses. Initial PCR for integration used F1 primers at three 2 step temperatures were 50°C & 47°C, 55°C & 53°C and 63°C & 59°C. The PCR products in the event of a successful integration were expected to be a 2kb band. For the second PCR integration check primers, F1 to F4 were used and PCR parameters used were the same as those in 2.2.3 (annealing was changed to 50°C and 47°C and an elongation step set at 1min 20sec).

Steavor		Forward primer	Reverse primer
PF3D7_125 4100	F 1	CTGCAAATATTTATAATAATGTTACA ATGATTG	CCTTGTATTATAAACATAAAATTGACAA AACAT
PF3D7_125 4100	F 2	ACTATTATGATATATATATGGTTTACA CCCATAACAG	GATATACATTGACACCAGTGAAGATGC
PF3D7_125 4100	F 3	CTGACTTGTATCGTCGCGATC	TTAATAACAATACGATTAATTGGTGATA GG
PF3D7_125 4100	F 4	GCTTTGTTTTTGTTTTTATGTTATTTTT TTAATATT	CAACATATTTCTGTGTTGTAAATTATTTA CAAAT
PF3D7_104 0200	F 1	GCAAGCTCTGACTACATCTATTTATGC	GTGCATTATAAAAAAATAAATATATAT ATTACACTCG
PF3D7_104 0200	F 2	CAAGAAAAATTGTTGTTAGTTTCACA ATAG	GATATACATTGACACCAGTGAAGATGC
PF3D7_104 0200	F 3	CTGACTTGTATCGTCGCGATC	GTTTAATATAAAAAATTGAATAAATGGTT TTAATAAAAT
PF3D7_104 0200	F 4	GGATACTACATGAACTATAGAAACA GTGTATTTT	CCTGTGTTGTAAATTATATTTAAATATA TAAATCATAG
Cytodeamin ase (CD)		ATGGTGACAGGGGGAATGG	TTAAACACAGTAGTATCTGTCACCAAAG TCA

Table 2.4: Primers used to check for integration.

Primer pairs were designed to give bands of 2kb in length (except CD primers which gave a band of 1kb). PCR parameters for CD were 50°C and 47°C and an elongation step set at 30secs.

2.3 Chromatin immunoprecipitation (ChIP) coupled to qPCR

2.3.1 Chromatin immunoprecipitation

Saponin lysis and crosslinking steps

All washing steps were done using the table top Eppendorf centrifuge using the following parameters: 5-minute spin at 3214 g (Eppendorf centrifuge 5810R) with break 5 and acceleration 9.

0.15% saponin (MP Biochemicals) dissolved in 1xPBS was used to lyse trophozoite synchronized cells. The volume of the cells was usually 6ml per 50ml falcon tube. To this volume, 24ml (or 4x volume of the cell pellet) of 0.15 saponin, was added and incubated for 4 minutes but no more than 5 minutes. After lysis, cells were washed with pre-warmed 1x PBS at 37°C 2x or until the supernatant was clear (to achieve this sometimes a third wash was necessary). 36.5-38% formaldehyde (Sigma) was used for crosslinking at a final working concentration of 0.5% for 10 minutes with slight swirling after every 2 minutes. After 10 minutes, 1.67M glycine (Promega) at a final concentration of 0.125M glycine was added to the cross-linked material to stop the crosslinking reaction. The material was then washed with cold 1x PBS. The pellet was snap frozen in liquid nitrogen and then kept in the -80°C freezer or directly processed.

Homogenization and Sonication steps

Parasites were incubated on ice with 1.5ml lysis buffer for 60 minutes and homogenized by lysing with a dounce homogenizer for 200 strokes. To obtain nuclei pellet a centrifugation step was done and the pellet was dissolved in 1.5ml SDS-containing buffer and incubated on ice for 15 minutes. Following incubation, nuclei were sonicated for 8 minutes at 25 % amplitude (10 sec

pulse on, 50 sec off) to give a range of between 300 to 1000bp. The sonication was done in 200µl aliquots which were later pooled together. This was centrifuged and to the supernatant was added 13.5 ml ChIP dilution buffer. A small aliquot of this was decrosslinked, purified and the concentration took by nanodrop (this procedure allowed me to work backward and quantitate the amount of ChIP DNA/µl in the original sonicated chromatin). Roughly 2ml of trophozoite iRBC at 10% parasitemia gives about 7µg of chromatin for a pull-down.

Immuno-precipitation steps

The chromatin in ChIP dilution buffer was pre-cleared by incubation with rotation with 600µl Protein A agarose/Salmon Sperm DNA beads (Millipore) at 4°C for 60 minutes. After the 60 minutes, this was spun down at 4°C for 5 minutes at 129 g (Eppendorf centrifuge 5810R). The supernatant (pre-cleared chromatin/ChIP buffer mixture) was transferred to a different tube and the beads discarded. For each ChIP pull down 4µg of antibody was used per ≈7µg of ChIP material in a 1.5ml volume (Table 2.5). 10% of this volume (150µl) was set aside in fridge as input and was not subjected to any pull-down. The antibody: chromatin mixture was incubated overnight in a cold room with rotation at 10 rpm (Labnet labroller II- Bio laboratories).

Washing steps and decrosslinking steps

Washing steps were done in cold room with 10 minutes of rotation for each wash and a 5-minute spin at 4°C at 129 g. Buffers should be cold.

50µl Protein A agarose/Salmon Sperm DNA beads prewashed with 0.15% BSA (Sigma) in ChIP dilution buffer was added to each of the immunoprecipitated samples and further incubated for an additional 60 minutes still at 4°C. Samples were then spun down at 4°C for 5 minutes at 1000rpm. The supernatant was aspirated and the pellet was washed

Antibody	Company	Code	Concentration	Amount for pull-down
H3	Abcam	ab1791	100 µg at 0.9 - 1 mg/ml	4µl
H3K9ac	Abcam	ab10812	100 µg at 1 mg/ml	4µl
H3K36me3	Cell signaling	9763S	100 µg at 1 mg/ml	4µl
H3K4me2	Abcam	ab7766	100 µg at 1 mg/ml	4µl
H3K4me3	Cell signaling	9797S	100 µg at 1 mg/ml	4µl
H3K79me3	Abcam	ab2621	100 µg at 0.9 - 1 mg/ml	4µl
H3K9me3	Upstate	07-442	100 µg at 1 mg/ml	4µl
H4K12ac	Abcam	ab46983	100 µg at 1 mg/ml	4µl
H4K20me1	Abcam	ab9051	100 µg at 1 mg/ml	4µl
H4K20me3	Abcam	ab9053	100 µg at 1 mg/ml	4µl
H4K8ac	Abcam	ab15823	100 µg at 1 mg/ml	4µl
IgG control	Abcam	ab171870	100 µg at 1 mg/ml	4 µl

Table 2.5: Antibodies used for pull down.

2X with Low salt wash (150 mM NaCl)

3X with High salt wash (500 mM NaCl)

1X with Lithium Chloride wash

2X with TE wash (at room temperature)

After the last wash, 250µl elution buffer was added to the pelleted protein A agarose/antibody/histone complex. This was vortexed briefly to mix and incubated at room temperature for 20 minutes with rotation. This was spun down and the supernatant (eluate) carefully transferred to a 1.5 ml eppendorf. The elution step was repeated and elutes combined (total volume $\approx$ 500µl).

ChIP DNA purification

20µl 5M NaCl was added to combined eluate (≈500µl) and 4µl 5M NaCl was added to the 10% input. This was heated at 65°C overnight. The next day, 1µl Proteinase K (20mg/ml-Thermo Scientific) was added to all samples and incubated at 37°C for 60 minutes. Decrosslinked DNA was purified using QIAquick PCR Purification Kit (Qiagen) following manufacturer's protocol and DNA was eluted in a 150µl volume of water.

ChIP-qPCR steps and data analysis

1µl for both input and immune-precipitated material was used for each PCR run. Because a 10% of the input had been used in the reaction, this was adjusted by subtracting 3.333 cycles from the input Ct value. All *stevor* oligos in each immunoprecipitation were then calibrated to the input DNA signal using the formula $2^{\Delta Ct = (Ct \text{ oligo Ab} - Ct \text{ oligo input})}$, where Ab is the respective antibody used for ChIP (Cabral et al. 2012; Petter et al. 2013). The plotted enrichment was calculated as the ratio of immunoprecipitated Ab oligo signal/immuno-precipitated H3 oligo signal (Lopez-Rubio et al. 2007; Cabral et al. 2012; Petter et al. 2013). H3 was used as to normalize between the biological replicates and also as an internal control for the experiments (Srivastava et al. 2014). Statistical difference between the ‘off’ and ‘on’ line was calculated using students t-test. Primers used for ChIP assays are in Table 2.1. All ChIP experiments were run with the IgG control but this was not plotted on graphs.

Polymerase activation	95°C for 2 min	
Denature	95°C for 10s	} X 40
Denature	95°C for 10s	
Annealing	51°C for 15s	
Extension	60°C for 40s	

Gene		Forward primer	Reverse primer
0201300 5'3	A	ACCCATAATAACTATTGTTCTACAC	AGTTTCGTCATTCTTATAATTTATTT
0201300 5'4	B	AAATAAATTATAAGAATGACGAACT	TTTTACATACATAATGTGGTAAC
0201300 cod1	C	GTTACCACATTATGTATGTAAAA	TGTCAATTATTTCTTTTAGTTCTGG
0201300 cod2	D	CCAGAACTAAAAGAAATAATTGACA	CTGATTTGTCATCATCATTGGGTACC
0201300 cod3	E	GGTACCCAAATGATGATGACAAATCAG	ATCAGGAGCACCATTTCACAGG
0617600 5'1	A	ATACTCAATGTCACATATGTGG	TCTATTCACCCATATAAAATGCTC
0617600 5'2	B	GAGAGCATTTTATATGGGTGAA	GTGTAGAACAATAGTTGTTATGGT
0617600 5'3	C	ACCATAACAACTATTGTTCTACAC	TAATCTAATGTTAAAAAATATATAAATATG
0617600 5'4	D	CATATTTATATATTTTTTTTAACATTAGATTA	CCAAAAATATAATTTTTATATTTCAAATTG
0617600 cod1	E	AACGCTTAAAAGAGGGATGT	CCATATCAGTAGCAACGGAAG
0617600 cod2	F	AACCTGTAGGTGAACATGGTA	GGGTACATCCCTCTTTTAAGC
0617600 cod3	G	ACAATCATTACAGTTTCAATATT	TACCATGTTACCTACAGGTT
0617600 cod4	H	ATGAACATATATTACGTTAAAA	AATATTGAAACTGTAATGATTGT
0700400 5'1	A	GAATATAAATTACAGTATCTCACTA	GATTTTTCTTTTACTTTATTCC
0700400 5'2	B	GGAATAAAGTAAAAAGAAAAATC	CTAAAAACCGTTCTCTCATATG
0700400 5'3	C	CATATGAGAGAACGGTTTTAAG	CCTTAATTCAAAAAATAATATCTT
0700400 cod1	D	GGATGAATATATATCACGTTA	GTCATTATGATAATGCGGATTG
0700400 cod2	E	CAATCCGCATTATCATAATGAC	ATTTGGGTATATCCCGAC
0700400 cod3	F	GTCGGGAATATACCCAAAT	GATTCATAAGTAAGCATATTAAT
0832400 5'1	A	GAGAACATTTTAAATTACATAATT	CGGAACTAACAACAAATTTTCTTG
0832400 5'2	B	CAAGAAAATTTGTTGTTAGTTCCG	AGCTTTTCTTAAAGTAATTCTCTC
0832400 5'3	C	GAGAGAATTACTTTAAGAAAAGCT	CAATAACATTTTAAAGGTAATATG
0832400 cod1	D	CATATTACCTTAAATGTTATTG	GATTTTGTGTTTGTGCTAAGAG
0832400 cod2	E	CTCTTAGCACAAACACAAAATC	TCAGAACATTGCGATCTTGA
0832400 cod3	F	TCAAGATGCGAATGTTCTGA	CATTGCACTAGCAGAAGTTG
0900900 5'1	A	AGAGAATTTTCTATCATATAAGGCT	TCATAATAGCCCTGATGATTTT
0900900 5'2	B	AAAATCATCAGGGCTATTATGA	ATTTATGACTATGTTTCATT
0900900 5'3	C	TTTTTTTAAATATTAATGAAC	TGCTAATGTATTTATTAATAAAGG
0900900 cod1	D	CCTTTTAAATAAATACATTAGCA	GCCCCGACTTCAACATATGAT
0900900 cod2	E	ATCATATGTTGAAGTCGGGC	TGCAATTCTTAGGCACTTCT
0900900 cod3	F	GCTACTAAAGCTGCTCTCAT	AGTAGAACCAAGCTGCAAT
1040200 5'1	A	CATTATGTTGTATTATTTGAAAAGC	TCAGTTTTGTTGTTAAGCAT
1040200 5'2	B	ATGCTTAACAACAAAACCTGA	CCTTCAAATACATATATAAATGTGTATAACAAT AA

1040200 5'3	C	TTATTGTTATACACATTTATATATGTATTGAA GG	CCATATTCATTTAATATTTAAAAA
1040200 5'4	D	TTTTTTTTTTTAAATATTAAATGAATATGG	GTATTAATCTAATTCATAAAAAATGATATG
1040200 cod1	E	GATGTATTACCTTAAATGA	GATTTTATCGTTGTTCTTGCG
1040200 cod2	F	CGCAAGGAACAACGATAAAATC	TAGCATGATTTTGTACCA
1040200 cod3	G	TGGTGACAAAAATCATGCTA	GGCACTAATGACTGAACAAG
1040200 cod4	H	G TTCAGTCATTAGTGCCTTT	AGCTAAATCAGAAGCATGAGT
1254100 5'1	A	ACTACTTTAAGAACTTCTATTC	ATTTTATATTGCTGTGTATGAA
1254100 5'2	B	TTCATACACAGCAATATAAAAT	TATATATGCATAATTAATAGGATATTC
1254100 5'3	C	GTAATAAGCTAATATATAATTGCTTTG	CTTTTATAGTTTATTTATTTTGTTC
1254100 5'4	D	GATACAATAGAGAGAATTTTTTATCATAG	TATAAATGTGTATAACAATAGC
1254100 5'5	E	GCTATTGTTATACACATTTATA	GAAATTTGTTGTAATTTATG
1254100 5'6	F	CATAAAATTACAACAAAATTC	GTTCTGGGACACAGCGAC
1254100 5'7	G	GTAATAAGCTAATATATAATTGCTTTG	CATATAAATGTGTATAACAATAGC
1254100 cod1	H	ACCTAAATATGTTATTGTTTACCTT	CGTTGCTCTTTTGGTCTTG
1254100 cod2	I	CAAGACCAAAAGAGCAACG	CCCGACTTCAACATATGACA
1254100 cod3	J	TGTCATATGTTGAAGTCGGG	GCACGCACTATAAGTTGTTT
1254100 cod4	K	GAACAACCTTATAGTGCCTGC	ACATAAATGTTTCTTGCATTCATG
1254100hdhfr 1	H	GCAAGAACGGGGACTACCC	AGGAATGGAGAACCAGGTCTTC
1254100hdhfr 2	I	GAGAACTCAAGGAACCTCCACAAG	AAACAGAACTGCCACCAACTATCC
1254100hdhfr 3	J	GGAGGAGAAAGGCATTAAGTACAAAT	TATGTTGCATCACCTTCACCCCT
1254100hdhfr 4	K	TCAATGCTTTGCGAGATACCC	GCACGTGTCTTGTAGTTCCCG
1254100hdhfr 5	L	TACCTGTCCACACAATCTGCC	CATCCATGCCATGTGTAATCC
1254600 5'2	A	CATCATGAGGGCTATTATGATA	CTTTAATTTACGATAATATTCATTT
1254600 5'3	B	AAATGAATATTATCGTAAATTAAG	CGTGATATATATTCATTTTAATTC
1254600 cod1	C	GAATTAATAATGAATATATATCACG	GGATTGCGTTTGTGCTAAA
1254600 cod2	D	TTTAGCACAAACGCAATCC	TTTTGCTAATTTACACCATGA
1254600 cod3	E	TCATGGTGTGAAATTAGCAAAA	GCTTGTTTCAAGAACTGCAA
1372500 cod1	A	TTGTCACAAAATGTATACAGAT	GGGTTTCTCTGTTTCTGCTA
1372500 cod2	B	TAGCAGAAAACAGAGAAACCC	GTGTACATATCAGATTTTGC
1372500 cod3	C	GCAAAATCTGATATGTACAC	GGGCTTGGGTCAATTTAGTA

Table 2.6: List of stevor primers used for the ChIP-qPCR reactions.

Primers were initially run on input DNA and checked for specificity by running a melting curve analysis. Primers showed here showed single peaks in the analysis. Not all primers showed good amplification efficiency when running on a diluted series (Supplementary Table 3). Primers amplify using Biorad SsoFast Green Supermix or the KAPA ;syber mix.

GENE		Forward primer	Reverse primer
Var2 5'1 A	A	TGCTTCATAAATAAAAACATGCAA	TTCTTGTTTTTCTAAAAATACCTCTCTT
Var2 5'2 B	B	CACGACATTAACAATACATGCAGA	TGTCATTGCATTACACACAAA
Var2 5'3 C	C	GGACAAAATGGATAGTACAAGCAC	TTTAAGTAATCTCCATCACCTCCA
Var2 cod1	D	F AAATGCGACAAATGTAAATCTG	GGGGTAAACCTATGGTGTGTTAGC
Var2 cod2	E	ACCCCAATCGGAAGGTAAGT	TTCATAGCTTCTAGCGCCTT
Var2 cod3	F	TGCCCAAAACCTTTACCATCCC	ACCCTAACGCCAACGCAATACC
seryl-tRNA synthetase		AAGTAGCAGGTCATCGTGGTT	TTCGGCACATTCTTCCATAA

Table 2.6b: List of control primers used for the ChIP-qPCR reactions.

These primers were not validated apart from blasting in the plasmodium database to check for specificity. PCR condition used were the same as for the other *stevor* ChIP-qPCR primers

2.4 RT-PCR

2.4.1 RNA extraction

Centrifugation for all steps is at 4°C for 40 minutes at 12857 g (Eppendorf centrifuge 5810R), acceleration 9 and break 3.

RNA was prepared using Trizol reagent (Ambion, Life tech). A 500µl pellet of parasites at between 5-10% parasitemia was used for RNA extraction. This pellet was washed once with 1x PBS. 10x the volume of Trizol to pellet volume was added to the pellet in the fume hood and incubated for 10 minutes with constant shaking. The Trizol/parasite mixture was either snap frozen in liquid nitrogen and then stored in -80°C freezer or processed further. In the case of freezing, the pellet was thawed in 37°C water bath and the further processed. 0.2ml chloroform (Merck) per 1ml Trizol was added to the Trizol/parasite mixture and mixed again thoroughly. When completely mixed the mixture turns a brown color. This was then centrifuged. This centrifugation step separates the mixture into 3 phases. The top aqueous phase containing the RNA was carefully pipetted out and stored in a separate 15ml falcon tube. Care was taken not to

disturb the organic layer. Analytical grade isopropanol (Merck) was added to the RNA, mixed and incubated in -20°C freezer overnight. After the overnight incubation, RNA was centrifuged. A white pellet is noticeable after this step. The isopropanol was changed for 75% ethanol and again RNA was centrifuged. After this step, the RNA pellet was allowed to dry (takes between 20-30 minutes). When pellet was dry, it was dissolved in 50µl RNase/DNase free water. RNA was then treated to DNase treatment (Qiagen) further purified using the Qiagen gDNA elimination columns and RNA cleanup columns (Qiagen) according to manufacturer's protocol. This was done to further reduce gDNA contamination.

2.4.2 RNA to cDNA conversion

1.5µg of RNA was converted to cDNA. The reactions were primed with a mixture of oligo-dTs as well as random hexamers (Sigma) in 20µl reaction volume for 5 minutes at 65°C (Master cycler Gradient, Eppendorf). The reactions were set to go to 22°C gradually. After this initial step, 1µl RNase out (Invitrogen), 2µl BSA (10mg/ml NEB), 2µl 0.1M DTT, 20µl 5x strand buffer and 1µl superscript II was added to each reaction and incubated for 2 hours in a 50µl reaction volume (added 30µl more). After the first hour, an additional 1µl of superscript II was further added and reaction ran for an extra hour. The samples were then diluted 3x and 1µl was used for each RT-PCR. A no-reverse transcriptase control was also run in parallel using the same amount of RNA minus the superscript II enzyme.

2.4.3 *Stevor* primer set

Stevor specific primers published by (Sharp et al. 2006) were used for q-PCR (Table 2.2). These primers were checked for specificity by blasting them against the Plasmodb genome. Some primers had to be redesigned due to the recent annotation and better sequence coverage of the

genome. 2 other *stevor* primers (PF3D7_0700700 and PF3D7_0832000) were added as they were not in the Sharp set (colour coded purple). To calculate primer efficiency, a serial dilution of gDNA covering 5 magnitudes was made and used a template. Using gDNA all *stevors* showed amplification. Each primer pair was amplified in each diluted sample in duplicates using the standard curve program on the ABI Prism FAST 7500 machine. The Applied Biosystems Master Mix was used for the *stevor* primers.

PCR program used for *stevor* cDNA q-PCR was as follows

Polymerase activation	95°C for 2 min	
Denature	95°C for 10s	} X 40
Denature	95°C for 10s	
Annealing	55°C for 15s	
Extension	60°C for 30s	

New <i>stevor</i> ID	Forward	Reverse
PF3D7_0101800	TGATGCCCTTGCTAGTTATGC	CAGAAGTTGCAGCAGGTACTGT
PF3D7_0102100	GGTTTGGCAAAGGCAAAATA	CAAGCTGAACTACCAGCTTCAA
PF3D7_0114600	CAGCTATTCAAGCAGGTGCTAA	CAATACCACAACCTCCAGGAA
PF3D7_0115400	GGCTGCCTTTGATACCTTGA	TGTACCGCCTGCAAACATAG
PF3D7_0200400	GCCCTAAGTGCTGTTGCTTC	CACCAACAACACATGCTTTCA
PF3D7_0200900	CTGCTGCCATTGCTACCTTT	AGCAGCACCAGATGCACATA
PF3D7_0201300	GCAAAATCTGCTGCCCTTAC	ACCATTTGCACAGGTTCCAC
PF3D7_0221400	TAAATTTACGAAGGCGCTTG	CCAGAGCTGCAATACCATAA
PF3D7_0222800	TGCTGCTGTCACTTCTAGCTTT	CTACACCTCCTGCTGCAGTAAC
PF3D7_0300400	AAAGGTGCTGCTATTTCTACCG	CAGCAGCACAACCTACTGTACC
PF3D7_0324600	TGCATGCTGCTAAAGTTGCT	TTGGTTGCACAAAGAACTGACT
PF3D7_0400800	TGGCAGCTACCAAAGCTACA	TCAGGACCCCAAGCTGTAATA
PF3D7_0401500	AAATGCGCATCCTCTATCACT	AGTTCCGACCCCATCACTAA
PF3D7_0402600	TTTATGGTATTAGCGCTGCAAG	GCGCTTTCTATTGATGTTTGAG
PF3D7_0425500	TCATCATCCCAATTGACGAA	CTTCATATTCGACGGACGTT
PF3D7_0425500	AATGCGTTGCCCTCTATATC	TCATATTCGACGGACGTTAT
PF3D7_0500600	GCAAAGGTTCTTGCTGGTGT	GCTATTCTCATAGCATCCCAATG
PF3D7_0617600	CTGCACCTTATGATGCTTGTGT	AAACAGTTCCGGCCATATCA
PF3D7_0631900	GTTGATGCCATCCTTCCTGT	ACGCTTCGGTTGCTTTAAGA
PF3D7_0700400	GCAAAAGCTGCTATCCTTGG	ACACCACCTGATGCACTAACTG
PF3D7_0700700*	CCTGTTGGTTTATGGTCTCC	ATGTTTTCGTCTGTTGCGTA
PF3D7_0732000	TTGCAGAAGTGCCTAAGAATTG	GCACTAACAGTAGTTGCTTCTCCT
PF3D7_0832000*	ACAGAATTCGCTGGTAATGT	AACTTGCAGCAAAGAACTC
PF3D7_0832400	AAGTGAGATGTTCCCGTGGT	TGGACAGTAAAGTGCGCAAG
PF3D7_0832600	AATTGTCCCAAACCTTGTC	AGCCTTCGTAGCCATTGAAC
PF3D7_0832900	TACGTGCGCATCCCATATTA	TTCCGACTAAATCACCAGCA
PF3D7_0900900	TTGCAGAAGTGCCTAAGAATTG	CATAGCAGCACTTGCAGCAG
PF3D7_0901600	GGCTGCACTTGCTTACTTTTC	TAGCACATGTTGCACCTCCT
PF3D7_1000800	TGGGGCTCCTACATTAAGTCA	CACGACAATTGCCATACCAT
PF3D7_1040200	AGGTGCTCTTGCTGAGTATGC	CAAATTTCAGTGCCTGCTTG
PF3D7_1100700	AGAACGTTGTGTTGGAGGTGT	GGAAGCCATAAACCAACAGG
PF3D7_1149900	TGACGTTGCTGCCTTGAATA	ACCAACGCCTGCTTGAATAG
PF3D7_1254100	TGCGTGCAAATCCTCTATCA	AAACAAGCTGTACTACCTGC
PF3D7_1254300	GCTAAAACGGCTGCCCTAA	CACCACTTGCACAAAATCCA
PF3D7_1254600	TCGGGCATGAGTCCCAATTT	TTAGGGCAGCCGTTTTAGCA
PF3D7_1300900	GCCGTTGCTTCTCTTGTTTATT	ATCAGTAGCACCGGCAAGAG
PF3D7_1372500	CCTTACTAAATTGACCCAAGC	CTTTTCTTCTTTTACGCAGC
PF3D7_1372800	CTGCTAAAACGTGTGCGTCT	ATTTGCAGCAGGCAAACTA
PF3D7_1400700	CTTCCGTTGGACCACCTTAT	CTTGCAGCAGTTCCGGTTAT
PF3D7_1479500	GCAAAAGTTGCTGTCATTGG	GAGCCTGCTGCATTAAGTGA

PF3D7_1479900	CATGAAGGCTGTTGCTGATTAT	ATCAGTAGCACCCCTCAACACAT
seryl-tRNA synthetase	AAGTAGCAGGTCATCGTGGTT	TTCGGCACATTCTTCCATAA
Hsp80	CACCGAATTACTCCGATTCCA	TCCGATCATTTTCGTCTCCTGA
Actin	AGCAGCAGGAATCCACACA	TGATGGTGCAAGGGTTGTAA

Table 2.7: *Stevor* primers used for RT-PCR:

Purple colored *stevors* with asterisks are *stevors* with primers not included in the Sharp primer set. Also included are the primers used as internal controls (tRNA, actin and hsp80).

2.4.4 RT-PCR run

Real-time PCR was carried out in duplicate/triplicate using ABI Prism FAST 7500 sequence detector (Applied Biosystems). Primer concentration used was 1 μ M of gene specific primers and 1X SYBR Green PCR-master mix (Applied Biosystems) in 20 μ l volumes. Each plate was run with these control genes, *hsp80*, *actin* and *seryl-tRNA synthetase* (Salanti et al. 2003; Sharp et al. 2006). From transcript data in Plasmodb (<http://www.plasmodb.org>), these genes fulfilled the criteria for housekeeping genes as they have consistent transcript expression patterns throughout the parasite life cycle. In addition to this, each run had the *seryl-tRNA synthetase* and *hsp80* genes, run in a no-reverse transcriptase control (sham cDNA) as controls for genomic contamination. Δ CT values for each individual primer pair were obtained by subtracting the CT value of the *stevor* primers from the CT value of the control *seryl-tRNA synthetase* (User bulletin 2, Applied Biosystems, <http://www.appliedbiosystems.com>). The formula $2^{\Delta CT}$ was used to convert the Δ CT values to relative copy numbers (Lavazec, Sanyal & Templeton 2007; Lavazec et al. 2006; Dzikowski et al. 2006). All reactions were analyzed with Microsoft Excel.

tRNA, hsp80, and actin were all used as internal controls for each run, however, data was normalized to tRNA sythetase in line with published data.

2.5 MICROSCOPY

Live cell imaging was done using trophozoite stage cells. The parasites were washed twice using incomplete media to remove any debris from culture. Washing conditions were 3 minutes at 2300rpm at acceleration 9 and brake 1. After each wash, the supernatant was aspirated out to leave the pellet. 5 μ l volume of this pellet was taken, mixed with 500 μ l of complete media and incubated for 10 minutes with DAPI stain (1/500 dilution) to stain the nuclei. Cells were viewed immediately between slide and coverslip under a Zeiss epifluorescence microscope using a 100X apochromatic objective and differential interference contrast (DIC) for the bright field. 75 cells were counted from the culture with drug and the culture without the drug. Cells were counted independently by lab member.

2.6 LIMITING DILUTION

The aim of this strategy is to end up with an average of half a parasite per 100 μ l. Ring cells were synchronized by sorbitol and re-cultured. At the schizont stage, cells were put on a rocker to ensure each RBC is invaded by a single merozoite. Thin smears were done to confirm the absence of multiple infections (more than one parasite per RBC) and cells were diluted to at least 0.4% parasitemia. A 100 μ l aliquot of this diluted culture was seeded in a 96 well plate and monitored after every two days. Column alternate wells were seeded as shown below (black stripped wells). On the 6th day, each well was split into two (second Figure) and observance continued till parasites started to appear. On average it took between 10-15 days for parasite appearance.

Initially, 100µl is seeded in an alternate column well as shown. After 6 days each well is diluted and split in two as shown below.

2.7 SOUTHERN BLOT

Genomic DNA was isolated from *pBstevor* line clones and untransfected 3D7 parasitized erythrocytes using genomic DNA purification kit (Fermentas). 5µg of gDNA from each sample was digested with PvuII (HF-NEB) overnight. Digested samples were run on 0.8% agarose gel and transferred onto Hybond N+ membrane (Amersham) in 10xSSC buffer overnight by

capillary action. The membranes were washed with distilled water and then UV cross-linked prior to hybridization (Ambion hybridization buffer). DNA probe targeting a 500bp region of the *hdhfr-gfp* gene was used with the incorporation of DIG-dUTPs (NEB). AP-conjugated anti-DIG Ab and CSPD substrate (Roche) was then used and the membranes were scanned using the Typhoon Scanner (GE Healthcare).

2.8 WESTERN BLOT

Antibody dilutions used in western blot are shown in Table 2.4

Nuclear fraction sample from *P. falciparum* was obtained from Dr. Yam. 2µl was run on a 12% acrylamide gel, in total 9 lanes loaded for each antibody. The gel was run on 100V for 60 mins till the dye front reached the bottom of the gel. The acrylamide gel was transferred to a nitrocellulose membrane and a wet transfer was done at 100V for 1 hour. Ponceau S staining was done for 2 mins to validate transfer, then washed off with 1X PBS. This was also done because the membrane was then cut into strips and the lanes were visible after the Ponceau stain. The cut membranes were blocked for 1hour at RT with 5% milk powder in 1x PBS w/0.5% Tween 20. After 1 hour, the blocking buffer was washed off and the primary antibodies (histone antibodies) were diluted in the 1 ml blocking buffer (Table 2.8) and incubated O/N at 4 degrees followed by 3 10 minute washes with 1x PBS with 0.5% Tween 20. Secondary antibody were diluted in fresh blocking buffer (Table 2.8- 1:5000 dilution) and incubated for 1hr at RT and washed 3x in 1xPBS with 0.5% Tween 20, each washing lasting 10 mins. For detection, membranes were incubated with 600uL chemiluminescent substrate (Pierce Chemiluminescence Pico Kit (Pierce) for 5 min at room temperature. Excess solution was dabbed off. Membranes were developed on

film using Kodak film for general purpose or Amersham hyper film TM ECL (GE healthcare) for sensitive blots.

Primary Antibody	Dilution	Secondary antibody Dilution
H3	1:1000	Anti rabbit hrp 1:5000
H3K36me3	1:500	Anti rabbit hrp 1:5000
H3K4me2	1:1000	Anti rabbit hrp 1:5000
H3K4me3	1:1000	Anti rabbit hrp 1:5000
H3K9me3	1:500	Anti rabbit hrp 1:5000
H4K12ac	1:1000	Anti rabbit hrp 1:5000
H4K20me3	1:500	Anti rabbit hrp 1:5000
H4K8ac	1:1000	Anti rabbit hrp 1:5000
H3K9ac	1:1000	Anti rabbit hrp 1:5000

Table 2:8: Antibody dilutions used for Western blot

2.9 PHYLOGENY ANALYSIS

750bp of the sequence from the ATG, as well as 800bp of the coding sequence from the 40 annotated stevor, were retrieved from plasmodb. The 40 sequences were saved in a WordPad as fasta format and uploaded to the MEGA6 software where the file was converted to a MEGA working format. The sequences were aligned in the software using MUSCLE. The aligned file was then used for analysis. I used the parameters by other groups (Joannin et al 2008). I constructed the tree using Phylogeny Reconstruction and Neighbor-joining statistical method. The test of phylogeny was Bootstrap method with 1000 replications. Model/method was p-distance and gaps or missing data was treated as complete deletions. Analysis using the same aligned file was done again using Minimum Evolution statistical method. Test of phylogeny was again 1000 bootstrap replicates. The model/method used was again p-distance with complete

deletion of missing gaps. For this analysis, the ME heuristic method used was Close-Neighbor-Interchange (CNI) and Initial tree for ME was obtained by Neighbor-Joining.

2.10 SEARCH FOR OVER-REPRESENTED SEQUENCES

The 750 5' UTR sequences were used again. These were uploaded into MEME (multiple Em for motif elicitation) <http://meme.sdsc.edu/meme/meme.html> (Bailey and Elkan. 1994) software. MEME is a motif search tool designed to identify over-represented sequences in any given sequence using an expectation maximization algorithm. For the search of any overrepresented sequences the following parameters were used

Motif Site Distribution	ANR: Any number of sites per sequence
Site Strand Handling	Sites may be on either strand
Maximum Number of Motifs	3
Motif E-value Threshold	no limit
Minimum Motif Width	6
Maximum Motif Width	50
Minimum Sites per Motif	2
Maximum Sites per Motif	50

Table 2.9: Parameters used for sequence search

This gave three sequences as that was the set maximum specified.

2.11 MICROARRAYS

2.11.1 CDNA synthesis

12µg of RNA was mixed with random primers and oligo DT (Proligo) and incubated at 65°C for 10 minutes and then kept on ice. A master mix with 5-aminoallyl-dUTP (Biotin, USA)

supplemented with dATP, dCTP, dGTP, dTTP (Fermentas), and Revert Aid H Minus MMuLV Reverse transcriptase enzyme (Fermentas) and the buffer was added to each reaction and incubated for 2 hours at 42°C for reverse transcription. Residual RNA was removed by addition of EDTA and NaOH and incubating for 10 minutes. RNA was also DNase treated. The reaction was cleaned by MinElute PCR Purification Kit (Qiagen, USA) and eluted in the elution buffer. This was also done for the reference RNA from 8 3D7 time points covering the whole IDC. Equal amounts from each time point were mixed together to make 12µg and subjected to reverse transcription.

2.11.2 CDNA labeling and microarray hybridization

The following steps were carried out in the dark. For microarray hybridization, 2µg of sample cDNA was labeled with fluorescent cyanine dye CyTM3 and 2µg of reference cDNA was labeled with CyTM5 (Amersham Biosciences) and incubated with 0.1M NaHCO₃ for 2 hours. A purification step using MinElute PCR Purification Kit (Qiagen) was carried out and the samples eluted in 13µl of elution buffer. Cy3 labeled sample and Cy5 labeled reference pool were mixed together and 20X saline-sodium citrate (SSC), 1M HEPES and 10% SDS were added to the mixture, boiled for 5 minutes at 100°C and then cooled. Hybridizations of Cy3 and Cy5 labeled samples were carried out at 65°C for 14-16 hr over the Maui hybridization system (BioMicro Systems). GenePix 4000B and GenePix pro 6.0 scanner (Axon Instruments, USA) were used to scan slides.

2.11.3 Microarray data analysis

After scanning the resulting microarray data was normalized based on Lowess normalization method and then filtered for signal intensity of more than background intensity plus 2 times the

standard deviation of background intensity for each Cy channel, in Acuity 4.0 software (Axon Instruments, USA). The software produced the log transformed ratios of Cy5 intensity to Cy3 intensity. Further analysis was carried out by hierarchical clustering on the log transformed ratios using Gene Cluster program version 2.11 and results were visualized using Java Tree view version 1.60 software (<http://rana.lbl.gov/EisenSoftware.html>).

CHAPTER 3

RESULTS

PART 3A: *STEVOR* INTEGRATED LINE

3A.1 GENERATION OF *STEVOR* MUTANT PARASITE LINES

3A.1.1 Introduction to plasmid design and rationale

In order to study the regulation of *stevor*, we wanted to have a system which would allow us to switch on a *stevor* of interest and this *stevor* would be active even after multiple rounds of replication. Unlike for the *var* genes where a selection of unique *var* gene expression is possible, it is not yet possible to select for expression of unique *stevors* based on adherence properties. Even if this were possible for *stevors*, we know from work on other gene families (Salanti et al. 2003; Sherf et al. 1998a) and on *stevor* (Lavazec et al. 2007) that, eventually the selected *stevor* would shift to the expression of other *stevors* over time because they switch. Therefore to circumvent this problem of switching and to help distinguish the parasite chosen promoter I chose to take a transgenic approach. In this approach, I would create a transgenic parasite line where a *stevor* promoter would drive expression of a drug selectable cassette (Figure 3.1). To do this, two homologous regions flanking the targeted *stevor* gene are used for cloning purposes (labeled as homology region in Figure 3.1, step 1). The two homology arms will facilitate double crossover integration (step 2) of the drug cassette at the targeted *stevor* such that the drug cassette will be under the 5' UTR of an endogenous *stevor* (step 3). In a successful integration event, the targeted *stevor* is knocked out and replaced by the drug marker (step 3). In such a

system, applying drug pressure would artificially switch on the *stevor* promoter activating it, and removing the drug pressure would lead to deactivation of the promoter.

Such systems have been used in studies of *var* genes (Dzikowski et al. 2007; Dzikowski et al. 2006; Voss et al. 2006). In those experiments targeting a drug selectable cassette downstream of a *var* promoter and applying resistance drug led to the activation of the *var* promoters. In our design, it was important that the construct be integrated into the genome. This would capitalize on any regulatory elements that naturally exist to switch on and off the *stevor* genes as previous studies had shown that histone modifications controlling gene regulation appear to occur in flanking non-coding regions (Cui et al. 2007; Lopez-Rubio et al. 2007; Swamy et al. 2011). Thus, I designed constructs that would allow for direct integration at a *stevor* locus as shown in Figure 3.1 schematic. This would achieve two aims;

- 1) Allow tracking of the active promoter and '*stevor*' as the parasites go through rounds of asexual development. This is possible because applying the drug to the culture would lead to activation of the promoter so after each round of development the active *stevor* promoter is known.
- 2) To test whether *stevor* undergoes mutual exclusion expression. Studies using isogenic clones, show clones expressing a single *stevor* (Sanyal et al. 2012; Lavazec et al. 2007b) suggesting the existence of a mechanism controlling the expression of a single *stevor*. Earlier studies in this area have used microarrays to answer this question. We wanted to use q-PCR to answer this question as q-PCR is not based on reference sample but rather a quantification of transcripts.

To overcome the laborious and challenging transfections in *P. falciparum*, I wanted to use a plasmid with both positive (*bsd* and *hdhfr*) and negative selection markers (yeast enzyme cytosine deaminase (CD)). Such a plasmid reduces the time needed to enrich for homologously integrated parasites by minimizing the need for cycling drug treatment off and on and allows for quicker integration events (Duraisingh et al. 2002; Maier et al. 2006; Maier et al. 2008). Plasmids used for this are shown in Figure 3.2.

Figure 3.1: Schematic of desired integration event.

In this schematic, two homologous arms flanking the targeted *stevor* gene are used for cloning purposes. In the cloning procedure the two arms, flank a drug marker. In these experiments, 1.5kb 5' UTR and a 1kb 3' UTR regions were used. The two homology arms will facilitate double crossover integration of the drug cassette at the targeted *stevor* such that the drug cassette will be under the 5' UTR of an endogenous *stevor*. In a successful integration event, the targeted *stevor* is knocked out and replaced by the drug marker.

3A.1.2 PCR and construction of *stevor* plasmids

Figure 3.2: Positive-Negative plasmids used for cloning created by Vector NTI.

Shown for each plasmid are the two drug cassettes. pCC1 (top panel) has *hdhfr* whereas pCC4 (bottom panel) has *bsd*. Each of the positive drug cassettes is under the control of the calmodulin promoter, with histidine rich protein 2 3' UTR acting as the terminator. The negative selection cassette has a heat shock 86 protein promoter driving the expression of the cytochrome gene with *P. berghei* used as a terminator. Both plasmids have 2 multiple cloning sites (MCS); MS1- *Sac*II to *Afl*II and MCS2- *Eco*RI to *Avr*II.

pCC1 has a human *dihydrofolate reductase* (*hdhfr*) as the positive selection and pCC4 has blasticidin (*bsd*) as the positive selection and both have CD as the negative selection marker. To create a linear cloning vector, the pCC1 positive selection cassette (*hdhfr*, CAM promoter & *hrp2* 3') was removed by enzyme digestion using *SpeI* and *NcoI/EcoRI* enzymes. The resulting digested product was run on a 0.8% agarose gel for 4 hours at 80V (Figure 3.3). The top band (in the red box) was gel extracted, purified and quantified using a nanodrop. This became the cloning vector.

Figure 3.3: Gel run of digested pCC4 plasmid.

SpeI and *NcoI/EcoRI* fast digest enzymes were used for digestion of the pCC1 plasmid. The digestion led to the removal of the positive cassette (*hdhfr*, CAM promoter, and *hrp*) on bottom band. The top band 6kb, (which is boxed in red) was the used for cloning purposes. The arrow shows a schematic of the vector and the restriction sites on either end.

Few *stevors* were chosen to maximize chances of getting successful cloning and also of getting integration. Four *stevor* genes, 3D7_1040200, 3D7_1254100, 3D7_0102100 (a pseudogene) and 3D7_1149900, were used for cloning purposes. Different criteria were used for targeting *stevors* to work with including distance in relation to the telomeric ends and expression profiles (Figure 3.4) as well as information in publications.

Distance from the telomere- 3D7_0102100 is a pseudogene located at 96000 bp from the telomere. The other three *stevor* are located at least 2000000 bp from the telomere with the exception of 3D7_1040200 which is located 1600000bp from the telomere. The telomeric regions are more heterochromatic and this could potentially lead to inaccessibility of the locus. Therefore, I also chose *stevors* that were more internally located on the chromosomes to have a varied profile.

Expression profiles- 3D7_0102100 peaks in ring stage which is different from other *stevors*. 3D7_1254100 and 3D7_1149900 show moderate to low expression levels during the IDC. To add to this 3D7_1254100 was shown to be expressed on merozoites surface as well during trophozoite by an earlier publication so this was chosen as it showed a varied expression as well (Khattab & Meri 2011). 3D7_1040200 is strongly expressed in the IDC stage. 3D7_1040200 has also been observed to be expressed by other studies suggesting that it's a dominantly expressed *stevors* or that it has a strong promoter (Sharp et al. 2006). Hence the *stevors* all showed different expression profiles which were a good selection.

In addition to this 2 of the *stevors*, 3D7_0102100 and 3D7_1040200 episomally expressed had been used to study *stevor* rosetting (Niang et al. 2014). In this study, however, they had used a non-*stevor* promoter and used the *stevor* coding region for cloning. The study showed that 3D7_1040200 is also exposed on the surface of the iRBC using antibodies against 3D7_1040200. So, different *stevors* showing different characteristics were chosen for cloning.

Figure 3.4: Characteristics of chosen *stevors*:

The figure depicts the genomic location of 3D7_1040200, 3D7_1254100, 3D7_1149900 and 3D7_0102100. 3D7_1040200 shows most expression during the early trophozoite stage (24 hours). 3D7_1254100 and 3D7_1149900 are transcribed minimally and 3D7_0102100 is transcribed more during the schizont to ring transition (data from plasmodb).

3A.1.2.2 cloning experiments

Two fragments, a 5' UTR 1.5kb upstream of ATG codon and 3' UTR 1kb region downstream of TAA codon flanking the endogenous targeted *stevor* were used as homology arms. This was to

facilitate a double homologous crossover, which would knock out the endogenous *stevor* gene replacing it with a drug cassette as shown in the schematic in Figure 3.1. PCR reactions of the 5' UTR (1.5 kb or 1.3kb for 3D7_1254100) and 3' UTR (1kb) fragments of the different *stevors* were performed and the correct amplification product sizes were checked by running PCR products on 1% agarose gel (Figures 3.5A-C). 3D7_0102100 failed to amplify (after a few attempts) and was no longer included in the subsequent reactions (Figure 3.5A lanes 6 and 7). The 3' UTR fragments of 3D7_1040200, 3D7_1254100, and 3D7_1149900 amplified with good efficiency (lanes 3, 5 and 9 respectively). The 5' UTR fragments for 3D7_1040200 (lane 2), 3D7_1254100 (lane 4), 3D7_1149900 (lane 8) amplified poorly but subsequent optimizations using lower temperature (annealing temperature of 49° C for 40secs) led to good amplification (Figure 3.5B). Primers were also designed to target the *bsd* and *hdhfr* genes. PCC4 was used as a template to amplify the *bsd* gene and PCC-1 was used to amplify the *hdhfr* gene. Amplification was checked by running on a gel (Figure 3.5C). PCR cycle used was the same as the PCR cycle for the *stevor* 3' UTR regions (2 step PCR with 56 ° C and 51 ° C as annealing temperatures- see M and M). All primers were designed such that they had complimentary restriction termini on each fragment for ligation purposes (Figure 3.6) to allow for simultaneous ligation. This method was chosen as it's faster to clone the fragments because it's a one-step process whereas sequential digesting and cloning into vector takes a longer time. The problem with this method, however, is that there a mixed population of ligated forms (see Figure 3.7) in minute quantities, therefore, a PCR always has to be performed to get enough quantities for digestion and cloning.

Figure 3.5: Successful PCR of 5' and 3' UTR fragments as well of *bsd* and *hdhfr*

A: 5' and 3' UTR PCR products of 4 genes: Lane 1 DNA marker: The genes in order are 3D7_1040200 (Lanes 2-3), 3D7_1254100 (Lanes 4-5), 3D7_0102100 (Lanes 6-7), and 3D7_1149900 (Lanes 8-9). **Lane 2, 4, 6 & 8 - *stevor* gene 5' UTR 1.5kb products; lanes 3, 5, 7&9- 3' UTR *stevor* gene 1kb products.** Lanes 6 and 7 are supposed to be the PCR products of 3D7_0102100.

B: PCR products of 3 *stevor* genes: Lane 1 DNA marker; Lane 2-4 *stevor* genes. **5' UTR 1.5kb products**-the genes in order are 3D7_1040200 (lane 2), 3D7_1149900 (lane 3) and 3D7_1254100 (lane 4). Note for 3D7_1254100, an upstream region of 1.3kb (lane 4) was amplified due to difficulties in designing specific primers past the 1.3kb position. This region has 100% homology with two other *stevors*. The best annealing temperature for the 5' UTR regions was 49°C for 40 seconds. At higher temperatures of above 51, the efficiency of amplification reduced.

C: PCR of *bsd* and *hdhfr* genes: *Bsd* gene is 399bp long and *hdhfr* gene is 567bp long.

Figure 3.6: Complementary restriction enzyme for insert fragments.

The terminus of each PCR product was designed such that they had complementary restriction sites. In the schematic diagram, targeted *stevor*, 3D7_1040200 is shown. All three *stevor* had the same design except 3D7_1149900 which had EcoRI as the restriction site on the reverse primer. With this design, all products can be ligated together eliminating the need to ligate each fragment separately into the vector (An et al. 2010).

The ligated material was used as a template for further PCR reactions (Figure 3.7) using primers flanking the outer parts of the total ligated insert (Refer to Figure 3.4 for schematic). This was necessary to have enough material for digestion and subsequent ligation with the vector. Using PCR primers flanking the ligated insert would give the following band sizes 3D7_1040200 *bsd* (2899bp), 3D7_1040200 *hdhfr* (3067bp), 3D7_1254100 *bsd* (2699bp), 3D7_1254100 *hdhfr* (2867), and 3D7_1149900 *hdhfr* (3067bp) (Figure 3.7). The correct band sizes for each construct are shown by a white arrow in Figure 3.7. In theory using the primers flanking the ligated insert

should have given expected band sizes only as shown for PCR amplification of 3D7_1040200 *bsd/hdhfr* 2839bp/3064bp (lanes 2/3).

Figure 3.7: PCR of ligated products.

Lane1: DNA markers, Lane 2-3D7_1040200 *bsd* (2899bp), Lane 3-3D7_1040200 *hdhfr* (3067bp), Lane 4-3D7_1254100 *hdhfr* (2867), Lane 5-3D7_1254100 *bsd* (2699bp), and Lane 6-3D7_1149900 *hdhfr* (3067bp). From the gel pictures, amplification of smaller intermediate ligated products is seen. ▽ Denotes correct band size in the lane.

However as can be seen from Figure 3.7, as well as amplification of the correct insert sizes, smaller fragments also amplified. The origin of these smaller intermediate fragments is unknown and they could be inserts with wrong ligation orientation. This phenomenon has been observed by other groups who have tried to ligate numerous fragments before performing a PCR (An et al. 2010). The PCR products were purified and digested using *SpeI* and *NcoI/EcoRI* and ligated with the digested pCC1 vector. Ligated products were transformed into competent cells and plated on ampicillin agar plates. Chosen colonies were incubated in LB medium and screened for correct ligation events by restriction digest of the insert. Positive colonies were verified by sequencing and expanded by maxi-prep.

3A.1.3 Transfections

Transfections were carried out as described in the methods section. Transfections with *hdhfr* and selected using WR21990 drug were unsuccessful, no parasites were observed in the culture until day 60. This transfection was repeated twice in a space of 30 days apart. Both efforts were a failure as no transfections came up. The *bsd* plasmids (3D7_1254100 and 3D7_1040200) were successful, with parasites appearing after 31 days. To verify that the parasites were true transfectants, genomic DNA (gDNA) was extracted and PCR was performed using primers against *bsd* fragment (Figure 3.8). The presence of the correct 399bp band in the gDNA from transfectants alone and not in untransfected parasite gDNA confirmed the presence of transfectants.

Figure 3.8: PCR to check for the presence of transfectants.

Primers were against the *bsd* of the plasmid which gives a band of 399bp. Lane 1= DNA markers, Lane 2=untransfected 3D7 gDNA, Lane 3= 3D7_1040200 transfected gDNA and Lane 4= 3D7_1254100 transfected gDNA.

3A.1.4 Addition of drug against the negative selection cassette

Having confirmed the presence of transfectants, the negative drug was added to the culture and the parasites monitored for growth. Maier et al. used the 5-FU drug at a concentration of 1 μ M and also continually selected for positive parasite transfectants using the WR99210 (Maier et al. 2006; Duraisingh et al. 2002). Therefore 1 μ M is the optimum 5-FU concentration to use. In their study, however, the DD2 strain of *P. falciparum* was used and I was using 3D7 strain. As well as using the optimum 5-FU concentration of 1 μ M, I also used 10 μ M and 20 μ M as I was using blasticidin as opposed to WR99210 and I was also using a different parasite line. Parasites were treated with the three 5-FC concentrations in conjunction with 2.5 μ g/ml blasticidin (Maier et al. 2006; Duraisingh et al. 2002), as well as 5-FC alone. 5-FC was used alone as a control for unknown combined blasticidin-5-FU effects.

When treating with the negative selection drug, all cultures showed an initial slow growth but they started growing as expected. I did not observe the death of parasites in any of the cultures at these concentrations. I grew the cultures on the negative selection drug for one month before checking for possible integration. The rationale was that the parasites were not dying because all of them had potentially integrated and the cytoadhesin (CD) gene had been excised out. Integration primers were designed to amplify a region in the *bsd* gene and upstream of the 5' UTR *stevor* cloning region. A schematic of the primer design is shown in Figure 3.9A. This primer would give a PCR band size of 2kb if PCR was successful.

The PCR did not show any bands, suggesting the absence of integration (Figure 3.9B) even after culturing the transfectants in the presence of the 5-FC for a month. The absence of bands in this gel run does not reflect the failure of integration as I did not use a positive control. It could be that the particular template I used was not pure enough or the enzyme I used did not work. We

decided to check for the presence of the cytochrome oxidase gene to indicate the presence of episomes and or integration (Figure 3.9C). To check for the presence of the cytochrome oxidase gene, we designed primers spanning the gene. The PCR to check for cytochrome oxidase gene was positive in both transfectants lines (Figure 3.9C). This suggested that the plasmid was still present and that no integration had taken place. In the event of integration, the CO gene would not have amplified unless the integration was a single integration event. Hence to check for single integration event, I designed primers both covering the 5' and 3' UTR position of both transfectants (Figure 3.9D). F1 primers = F primer is 200bp outside the 5' UTR region forward primer in combination with a reverse primer targeting the middle of the *bsd* gene. F2 primer = F primer targeting *bsd* and R primer 600bp outside the R primer of the 3' UTR region. F3 primer = R primer in *bsd* and F primer located at least 200bp from F primer used in 5' UTR cloning. F4 primer = F primer located in *bsd* gene and R primer located 300bp away from R primer used for 3' UTR cloning.

By PCR no evidence of integration was observed, despite using primers that were designed to cover both the 5' and 3' end of the cloning constructs (Figure 3.9E). To validate that the PCR worked, a PCR using primers against the *bsd* gene was carried out. This showed a band of expected size, ruling out a problem with the PCR reaction (Figure 3.9E lanes 12 and 13). This suggested that integration had not occurred but rather the plasmid was still episomally expressed. A southern blot would have been able to show whether an integration event had occurred. However, during this time another group managed to integrate a *stevor* (Witmer et al. 2012) which worked, therefore I did not pursue creating my own transgenic line but used their integrated *stevor* line.

Figure 3.9: PCR to check for integration in transfectants

A: Schematic of primer design to check for the presence of integration.

B: 1st PCR to check for integration event. PCR was carried out using primers in 3.9A. A 2 step PCR using 2 different temperatures was performed. Lanes 2, 7, 12=F1 primers designed for 3D7_1040200 *bsd*. Lanes 3, 8, 13= F1 primers designed for 3D7_1254100 *bsd*. Lanes 4, 9, 14; 3D7_1040200=F1 primers using untransfected gDNA and Lanes 5, 10 & 15; 3D7_1254100=F1 primers in untransfected 3D7 gDNA. **Expected band sizes 2kb but no bands were observed.**

C: PCR of CD fragment of the negative selection cassette. Lane 1: 3D7_1040200 *bsd*, Lane 2: 3D7_1254100 *bsd* and Lane 3: DNA markers. **Expected size is 1kb in the event of a successful PCR.**

D: Schematic design of multiple PCR primers to check for integration event. Primers targeted both the 5' UTR and 3' UTR regions. **Product sizes will be 2kb for all primer pairs. PCR conditions included using an annealing temperature of 50°C and 47°C and elongation set at 1min 20sec for 30 cycles.**

E: PCR check for integration using multiple primers. Multiple primers were designed to check for either single crossover integration at both the 5' and 3' end of the plasmid. The latter lanes (12 and 13) show the presence of the *bsd* gene in the genome of the parasites transfected suggesting the presence of the plasmid. Lanes 2-5= F1 to F4 for 3D7_1040200 *bsd* transfectants. Lanes 6-10 = F1 to F4 for 3D7_1254100 *bsd* transfectants. Lane 12 and 13= *bsd* PCR band for 3D7_1040200 and 3D7_1254100 respectively. Lanes 1 and 10 = DNA markers. **PCR conditions mentioned in Figure 3.9d.**

3A.1.5- Summary of transfection studies

Efforts to integrate a construct at targeted *stevor* loci were unsuccessful and no integration event was detected. This could be because of several reasons. Integration events in *P. falciparum* are difficult as the construct must cross four membranes to get to the nucleus. What was surprising was that, after treating the parasites with the negative drug selection, the parasites survived but still we could not detect any integration. We managed to get an integrated line and due to time pressure, we used the *pBstevor* line and froze our own transfected parasites. Table 3.1 shows the progress we made with each construct as a summary of our efforts.

Gene	PCR of 5' UTR	PCR of 3' UTR	Drug cassette	PCR of ligation	Transfection success	Negative selection	Integration	Episomal expression
3D7_0102100	no	no						
3D7_1040200	yes	yes	<i>bsd</i>	yes	yes	yes	no	yes
3D7_1040200	yes	yes	<i>hdhfr</i>	yes	no			
3D7_1149900	yes	yes	<i>hdhfr</i>	yes	no			
3D7_1254100	yes	yes	<i>bsd</i>	yes	yes	yes	no	yes
3D7_1254100	yes	yes	<i>hdhfr</i>	yes	no			

Table 3.1: Summary of constructs targeted for *stevor* integration:

The table summarizes the constructs that we initially started with and at what stage of the process I got to. Only the *bsd* constructs were successful and reached the transfection stage. But these did not integrate and were still episomally expressed.

3A.1.6- Conclusion and Discussion

As part of my Ph.D., I was interested in answering a few questions regarding *stevor* expression. One of the main questions was whether *stevor* is able to undergo mutual exclusion expression. I was interested in this question because studies show that at a single parasite level, some parasites express just one *stevor* (Kaviratne et al. 2002). In addition to this, another study had shown that episomal expression of *stevor* promoters driving a drug selectable marker could lead to a shutdown of endogenous *stevors* (Howitt et al. 2009). This study had used episomally expressed plasmids that exist as multiple copies and do not represent the natural chromosomal situation. I wanted to try and answer this question with a better experimental setup. To achieve this, I wanted to create a transgenic line where I would have a single integrated copy of a *stevor* promoter driving a drug cassette much like what has been done for the *var* genes (Voss et al. 2006; Dzikowski et al. 2007; Dzikowski et al. 2006).

My efforts to get this integrated line failed. The integration I was aiming for is a very rare event in *falciparum*. This is because 1) in *P. falciparum* circular plasmids are used for transfections as linear plasmids are thought to be degraded (Waterkeyn et al. 1999). The circular plasmids preferentially exist as episomal circles and double crossover homologous recombination integration happens spontaneously and occurs at low frequency (Maier et al. 2006; Duraisingh et al. 2002), 2) the introduced DNA has to cross at least four membranes before it gets to the parasites genome and 3) once in the parasite nucleus the plasmids do not segregate evenly to daughter cells hence some parasites lose the plasmids meaning very few parasites maintain the plasmid (Van Dijk et al. 1997; O'Donnell et al. 2001). All of these reasons contribute to the very low transfection efficiency usually observed. Transformants also grow at a slower rate and rarely

integrate. This is why I opted for the negative selection plasmids as these are supposed to greatly improve the efficiency (Maier et al. 2006; Duraisingh et al. 2002).

I started with 4 *stevors* and I managed to get episomal transfectants in only two *stevor*. In my efforts, I made use of two drug selectable markers, *hdhfr*, and *bsd*. The *hdhfr* constructs 3D7_1040200 *hdhfr*, 3D7_1149900 *hdhfr* and 3D7_1254100 *hdhfr* all did not give transfectants even after 60 days and redoing the transfections twice. When these constructs failed to come up, I switched from *hdhfr* to *bsd* but only used 3D7_1040200 and 3D7_1254100. These two *stevor* were chosen because a publication used 3D7_1254100, and their construct came up and integrated at the endogenous *stevor* locus (Witmer et al. 2012). This suggested that this locus was able to integrate and it was logical to reclone this particular *stevor*. 3D7_1040200 was chosen because in a number of publications this *stevor* is usually transcribed (Sharp et al. 2006; Blythe et al. 2008) (we also observed transcription of this gene in our clones) suggesting it's a dominantly transcribed *stevor* and also because we have antibodies to this *stevor*. So in the event of integration, we had tools for further studies.

For both 3D7_1040200 *hdhfr/bsd* and 3D7_1254100 *hdhfr/bsd* constructs, the same 5'UTR and 3' UTR regions were used. The *hdhfr* versions of the transfectants did not come up but the *bsd* versions did. Re-sequencing of the cloned *hdhfr* sequence did not show any mutations, eliminating problems with the *hdhfr* sequence. An explanation could be that the 5' UTR used was not of enough length to drive the *hdhfr* gene but this is highly unlikely as the same 5' UTR was able to drive expression of the *bsd* gene. More plausible explanations are that there was an insufficient provision of stable *hdhfr* transcripts to confer WR resistance; the cultures required a longer period of time before they came up or there were problems with the plasmid purification

giving a plasmid that was not pure. Apart from these supposed reasons, I cannot find a convincing way of why the *hdhfr* parasites did not come up.

For the *bsd* constructs that came up, I did not observe any integration by PCR analyses. I treated these with 5-FC the drug targeted against the negative selection cassette and did not observe any parasites clearing. One explanation is that integration failed because these *stevor* are essential and cannot be knocked out or that the loci of the selected *stevor* were inaccessible for recombination presumably because the heterochromatic environment where *stevor* are located. However, previous studies show that *stevor* are not essential in culture as only about 33% of parasites express STEVOR compared to field isolate where up to 90% of parasites express STEVOR (Blythe et al. 2008). Also, integration of 3D7_1254100 by another group implies that *stevor* integration is possible (Witmer et al. 2012), although their experimental setup was different from my setup and this could be a crucial difference. They used a plasmid with two drug cassettes *bsd* and *hdhfr*. They first selected the transfectants with blasticidin against the *bsd* genes and when they got stable transfectants they then used WR99210 against *hdhfr*. In this publication, they were studying the promoter activities of the multigene families but with this set up, they got integration.

Problems with the use of CD as a negative selection cassette have been published before (Duraishighn et al. 2002). In this work they did not observe parasite death and when they did a PCR of the transfectants genome, they obtained a band for CD gene. The authors do postulate that the parasites probably gained resistance to the 5-FC drug, not via mutation in the CD (as they sequenced the rescued plasmid) but possibly through one of the transporters. The same group, however, had double integration event with the plasmids that use thymine kinase (TK) as the negative selection cassette. Perhaps the TK suicide gene is more potent than the CD gene.

The plasmids I used, are supposed to be an improved version of the Duraisighn et al. 2002 plasmid, but I observed the same results as the original 2002 paper.

Failure to get or confirm integration and the availability of an integrated *stevor* line led me to stop pursuing these transfection studies. This prevented me from carrying out the initial studies I had wanted to do. I could have used the episomal transfectants but this had already been done (Howitt et al. 2009). And an even better experimental set up with controls showed that applying drug pressure on the parasites leads to a general decrease of endogenously expressed virulent genes (Witmer et al. 2012). This is supposedly a coping mechanism by the parasites as these multigene families are not essential in an *in vitro* system.

A major drawback with the experimental setup for this study was the lack of positive control. As it is, I do not know whether the integration failed because of gene essentiality, technical issues or the *stevor* actually did integrate but at different *stevor* loci as this is multigene family as observed for other *var* integration experiments (Dzikowski et al. 2007) A southern blot would have been able to show this as opposed to PCR analyses for integration. A better and proposed approach would be to use a gene from the Maier (RESA,2 or GARP genes) or Jiang (PFSETs) papers that had integration using the same plasmids as a positive control in parallel with *stevor* constructs (Maier et al. 2008; Jiang et al. 2013). This would at shed light on whether I was unlucky with the plasmids or a technical issue was responsible.

PART 3B: STUDIES USING TRANSGENIC *STEVOR* LINE

3B.1 INTRODUCTION: *STEVOR* INTEGRATED PARASITES

In our experiments, a *Plasmodium stevor* integrated line (Witmer et al. 2012) was used (*pBstevor*). These parasites had the plasmid integrated at the endogenous *stevor* locus. The plasmid had two drug selection cassettes; *bsd* and *hdhfr* (Figure 3.10A). The first cassette has the heat shock protein 86 (*hsp86* 5') promoter driving the *bsd* resistance gene and the *P. berghei* dihydrofolate reductase 3' (PbDT 3') terminator sequence. A 0.5kb *rep20* sequence was included in the plasmid as it helps tether the plasmid to the chromosomes and therefore to facilitate segregation of episomal plasmids to daughter cells during schizogony (O'Donnell et al. 2002). This helps maintain the episomes in the parasites. The second cassette had an *hdhfr/gfp* fusion gene under the control of a 2kb 3D7_1254100 region from the ATG (3D7_1254100 5'). The terminator for this cassette was the histidine-rich protein 2 3' sequence (*hrp2* 3'). The construct had a *var* intron as this is known to play role in *var* gene regulation and the authors wanted to keep all constructs for all the multigene families the same. The pBcam *stevor* line integrated at the endogenous *stevor* locus shown as PFL2610w (Figure 3.10B-D). PFL2610w is the old annotation for PF3D7_1254100. The plasmid backbone has two PvuII sites in the *bsd* and *hdhfr* gene which gives a band of 4890bp in the *pBstevor* line (Figure 3.10B).

Figure 3.10: Integrated *stevor* line

A: Schematic presentation of the pBcam *stevor* plasmid used for transfections. Hsp86 5' = heat shock protein 5' UTR sequence, bsd = blasticidin, PbDT 3' = plasmodium berghei dihydrofolate reductase 3' sequence, rep20 = repetitive element 20 (TARE 6), 3D7_1254100 5' = A 2kb upstream region from ATG, hdhfr/gfp = a fusion gene of human dihydrofolate reductase and green fluorescent protein, hrp2 3' = histidine rich protein 2 3' terminator sequence, var intron.

Green arrows = 5' UTR regions (promoter regions). Red boxes = 3' terminator sequences.

B-D: Schematic of integration event at the 3D7_1254100 (PFL2610w is the old annotation for 3D7_1254100) locus. This image from the paper published with the *stevor* line we used. The plasmid backbone has two PvuII sites in the bsd and hdhfr gene which gives a band of 4890bp in the pBstevor line (**B**). However, upon integration, a second band is observed at 3479bp (**C** and **D**). In the integrated state, the wild type 3D7_1254100 5' UTR drives expression of hdhfr fused to green fluorescent protein (gfp) (**D**). Also shown are the expected band sizes upon digestion with PvuII (Figure from Witmer et al. 2012).

However, upon integration, a second band is observed which 3479bp long (Figure 3.10D). This band comes from the disruption of the two PvuII sites at the PFL2605w locus (Figure 3.10C). In the integrated state, the wild type 3D7_1254100 5' UTR drives expression of *hdhfr* fused to green fluorescent protein (*gfp*). Also shown are the expected band sizes upon digestion with PvuII in the event of integration (Figure 3.10D). In the construct, applying the WR99210 drug will lead to activation of the *stevor* promoter and transcription of *hdhfr*, leading to its subsequent translation to produce the protein which confers resistance to the drug. However, the *pBstevor* parasites were un-cloned. Thus, we performed a limiting dilution step after initial validation of integration using Southern blot. The limiting dilution step is necessary for two reasons, i) to produce a parasite with only an integrated form of the plasmid and no episomal expression and ii) to clone distinct genetic lines from the established WR99210 (WR) resistant culture. These genetic lines will be homogenous as they arise from single parasites. Clones from the limiting dilution step were validated again by southern blot (Figure 3.11). Two bands were observed. The lower band (3.4kb) showing the integration of the construct at the 3D7_1254100 *stevor* locus and a top band (4.8kb) probed against the plasmid backbone, as this was a single crossover event. The specificity of the Southern blot is shown by the absence of bands in the 3D7 gDNA also probed with *hdhfr*. The Southern blot pattern I saw, was identical to the pattern seen by original authors after their southern blot (Witmer et al. 2012: Figure S1).

Figure 3.11: Southern blot of limiting dilution clones from pBstevor.

Clone numbers based on the well in which they grew on the 96 well plate. DNA was digested with High Fidelity PvuII from NEB and probed against *hdhfr*. No bands are seen in the 3D7 control. + (+) is gDNA from the parental parasites. The 4.8kb band represents the plasmid backbone and the 3.4kb band represents the *hdhfr* integrated form of the plasmid. The top band has a much greater intensity than the 3.4kb band suggesting multiple copies of the episomal plasmid as opposed to the integrated plasmid.

3B.2 ANALYSIS OF *STEVOR* TRANSCRIPTION IN DIFFERENT *pBstevor* INTEGRATED CLONES

To monitor *stevor* gene transcript levels and to examine the expression state of the rest of the *stevor* genes, RT-PCR was used with a modified *stevor* primer set (Sharp et al. 2006) technique was favored as it allows simultaneous monitoring of transcription levels of all *stevor* genes within the entire 3D7 parasite line. *Stevor* primers were validated by re-blasting them against the genome (plasmodb.org), amplifying them from gDNA, and confirming the presence of a single band. A melt-curve analysis was also performed at the end of each run to check for single peaks for each primer. The amplification efficiency for each *stevor* primer, as well as the housekeeping gene controls, was checked by amplifying each primer with serially diluted gDNA. The amplification efficiencies were all above 90%, showing that the amount of product doubled at each PCR cycle (Supplementary Figure 4). In all qPCR runs, tRNA was also run and used as an internal control and used to normalize the transcript levels (Niang et al. 2009; Cabral et al. 2012; Lavazec et al. 2007). I also ran hsp80 as an additional control and to further check gDNA contamination using a different primer but data was not normalized to this.

5 clones obtained by limiting dilution (Figure 3.10) were analyzed for *stevor* expression. To ensure the absence of the contaminating gDNA the samples were treated with DNase before converting to cDNA. In the RT-PCR run, a no reverse transcriptase control was also used for each clonal line. A representative of one such run is shown in Figure 3.12. No presence of transcripts is observed in the –RT samples. Figure 3.13 shows the results of *stevor* expression in the clonal lines from the limiting dilution analysis. Clone AF5 highly expressed another *stevor* (PF3D7_1040200) in addition to the *stevor* transgene (labeled as *hdhfr* on the graphs). Clone

BC10 showed expression of three other *stevors* (PF3D7_022800, PF3D7_1040200 and PF3D7_137200), but these just above the 0.2 relative units which we considered as the cutoff point for expression and silent (Cabral et al. 2012). The other three clones BC2, AC9 and AB4, did not express any other *stevor* apart from the *stevor* transgene (*hdhfr* on the graph). Limiting dilution clones that do not express *stevor* have been reported before (Lavazec, Sanyal & Templeton 2007; Niang et al. 2009) and this could be a reflection of earlier observations that in culture, *stevor* is expressed at lower levels than in when the parasite is a host (Blythe et al. 2008).

Figure 3.12: A representative analysis of gDNA control in the RNA.

Hsp80 and tRNA were run in RNA with and without reverse transcriptase. The graph shows the absence of gDNA contamination in the RNA used to convert to cDNA as no transcripts are observed in the –RT controls. All transcription values are normalized to tRNA with a value of 1.

Figure 3.13: Analysis of transcription levels of the *stevor* multigene family in AF5, BC10, BC2, AC9 and AB4, integrated clones.

These clones were a result of a limiting dilution cloning step from the *pBstevor* parasites of 3D7 parasites. RNA was harvested using Trizol and converted to cDNA. Q-PCR was performed at early trophozoite stage (26-30hpi). Shown is the average of two biological duplicates for clones. Only one clone expressed an additional *stevor* (AF5). All the clones did not show any additional *stevor* expression. *Seryl-tRNA synthetase* used for normalization for all expression. The *stevor* transgene is represented by *hdhfr* as the primer was designed against this region of the gene. Error bars = standard deviation.

3B.3 ANALYSIS OF CLONE AF5

3B.3.1 Conditional expression of *hdhfr-gfp*

Clone AF5 was chosen for further studies as it also expressed an additional *stevor* (PF3D7_1040200) at higher levels which could also be studied for epigenetic regulation. AF5 was grown and the culture was split into two. One-half of the culture was grown with 5nM of the drug WR99210 (WR). In this culture, the promoter of the gene is switched on to drive expression of the drug cassette. This culture is henceforth referred to as AF5^{ON}. The other half of the culture was grown without the drug. This culture is referred to as AF5^{OFF} as there is no pressure to express the *hdhfr* gene. Over time expression of the *hdhfr* cassette of the *stevor* transgene grown without drug should decrease as in the absence of drug pressure, the *stevor* transgene expression would lower over time (Dzikowski et al. 2006; Chookkajorn et al. 2007).

3B.3.2 Live cell imaging and real-time PCR analysis

Expression of the *stevor* 3D7_1254100 transgene in AF5^{ON} and AF5^{OFF} was monitored by live cell imaging and real-time PCR. Live cell imaging was possible as the *hdhfr* gene was fused to a *gfp* reporter tag. In the live cell imaging experiments, there is an obvious difference in the GFP fluorescence for the ‘AF5^{ON}’ and ‘AF5^{OFF}’ cultures (Figure 3.14A). The GFP signal in the ‘AF5^{ON}’ culture was of a brighter intensity compared to the GFP signal in the ‘AF5^{OFF}’ culture. In fact, in the ‘AF5^{OFF}’ culture, a higher proportion of the cells did not show any GFP signal or the signal, was similar to background levels. Cells expressing GFP or not expressing GFP were quantified and results are shown in Figure 3.14B. As can be seen from Figure 3.14B, not all cells in the AF5^{ON} culture expressed GFP but a significant proportion has an active promoter. 82% were seen to have good fluorescence signal suggesting an active promoter. Even the AF5^{OFF}

culture, GFP expressing cells are present (less than 20%) a significant drop compared to AF5^{ON}. Interestingly the GFP signal was very diffuse (Figure 3.14A); suggesting localization of the GFP protein is in the nucleus and regions of the cytoplasm. This is in line with the function of *dhfr* in DNA synthesis and the metabolism of certain amino acids and its localization in this region.

Hdhfr expression was also monitored by real-time PCR using a primer designed against the *hdhfr* fragment. This same primer was used in the ChIP-qPCR analysis (Part 3B). The real-time PCR results mirrored the results of the imaging experiments, as more *hdhfr* transcripts were observed in AF5^{ON} compared to AF5^{OFF} (3.14C). There was at least 5 fold reduction in expression of the transgene in the AF5^{OFF} compared to the AF5^{ON}. Interestingly, even after three months of growing the culture with no drug, *hdhfr* transcripts are still present in the cells, suggesting that complete shutdown of the 3D7_1254100 transgene takes a long time. This phenomenon is similar to what has been observed in other like experiments (Dzikowski et al. 2006) where the *var* transgene was still expressed after removal of the drug for 10 weeks. This could indicate that *stevor* promoters have low switching rates as alluded to by another study (Lavazec et al. 2007).

Figure 3.14: Validation of AF5^{ON} and AF5^{OFF} systems

A: Live cell imaging of AF5^{ON} and AF5^{OFF} culture parasites. Cultures were incubated with DAPI which stains the nucleus and then viewed under the epi-fluorescent microscope. The green fluorescence showed a diffuse pattern suggesting localization in the cytoplasm. When merged with the DAPI channel, there is co-localization of both stains. In the AF5^{OFF} culture, however, the GFP signal was markedly reduced, but the DAPI stain remained strong and bright. After merging the GFP and DAPI stains, only the DAPI stain is observed.

B: Quantitative analysis of GFP-positive cells in AF5^{ON} and AF5^{OFF}. Live imaging was done on 75 cells in AF5^{ON} and AF5^{OFF}. Cells were prepared as described in the methods section. 25 cells were counted over a three-cycle period of the parasite life cycle before the cells were harvested for ChIP-qPCR to make a total of 75 cells counted.

C: Real Time PCR (q-PCR) of 3D7_1254100^{ON} and 3D7_1254100^{OFF} parasites. Q-PCR was performed using primers designed against the *hdhfr* part of the fused gene (Primer H in ChIP-qPCR analysis). Values were normalized to *seryl-tRNA-synthetase*. This is a gene which fulfills the housekeeping criteria, as it is expressed throughout the blood stage cycle of the parasite uniformly apart from the ring stage where there is a lower expression. The graph shows the relative expression of the AF5^{ON} parasite against the AF5^{OFF}.

Several studies have shown that expression of an integrated trans-*var* gene by the addition of drug, leads to a down-regulation of endogenous *var* genes and only expression of the trans *var* gene (Voss et al. 2006; Chookajorn et al. 2008; Dzikowski et al. 2006). In two of the studies, they show that upon removal of drug pressure, the parasite population over time becomes heterogeneous in terms of *vars* expressed (Chookajorn et al. 2007; Dzikowski et al. 2006). I was interested in understanding whether I would observe the same phenomena with the transgenic *stevor* line. To emulate these studies, I used the transgenic AF5 clone and compared the *stevor* expression in the AF5^{ON} and AF5^{OFF} cultures. I found no differences in the expression pattern between the two cultures even after the AF5^{OFF} culture had been off the drug for more than 2 months (Figure 3.15). AF5^{ON} expressed PF3D7_1040200 and to a lesser level PF3D7_0832000, and there was no down-regulation of any *stevors* as these same *stevors* were expressed in AF5^{OFF}. This may suggest that *stevor* promoters do not exhibit mutually exclusive expression in agreement with the observations made by other studies. This result also suggests that *stevors* undergo limited switching.

Figure 3.15: Transcription analysis of AF5^{ON} and AF5^{OFF}.

The clones show expression of the same *stevor* regardless of whether it's under drug pressure or not as evidenced by the *hdhfr* expression levels in both systems. In both systems, PF3D7_1040200 is highly expressed. The white bars represent AF5^{ON} whereas the black bars represent the AF5^{OFF}. Plotted is the average of 3 biological replicates. Error bars = SEM. The samples for AF5^{OFF} were first collected after 2 months off the drug.

3B.4 VALIDATION OF ANTIBODIES USED FOR STUDIES

To study epigenetic regulation of *stevor*, I wanted to look at several modifications. The modifications studied in this thesis were H4K8ac, H3K9ac, H4K12ac, H3K4me2, H3K4me3, H3K9me3, H3K36me3, and H4K20me3. I also looked at H3K79me3 and H4K20me1 (antibodies were a kind gift from Prof ZB lab) these had been validated by western blot and IFA (Gupta et al. 2013). The modifications were chosen based on work that has been done in *P. falciparum* and because their role in gene regulation is known (Introduction section). H3 antibody was used in parallel to all other antibodies as an internal control and to normalize between biological replicates. Even though histone proteins are conserved in eukaryotes (Baxeavanis & Landsman 1996), I still blasted the human peptides the antibodies were raised against to the *plasmodium* H3 and H4 amino acid sequences (Figure 3.16). The figure shows the human peptides in red and the region in *Plasmodium* they match to in red. The peptides have corresponding matches in the Plasmodium H3/4 proteins. For H3K36me3 and H4K20me3, there was a one amino acid mismatch between the human and *falciparum* counterpart, but the mismatch was not at the modified amino acid residue.

H3K4 dimethylation and trimethylation peptide- **RT(me2/3K)QT**

P. falciparum H3 sequence- MA**RTKQT**ARKSTAGKAPRKQLASKAARKSAPISAGIKKPHRYRPGTVALREIRRYQKSTDLLI
RKLPFQRLVREIAQDYKTDLRFQSSAVMALQEAAEAYLVGLFEDTNLCAIHAKRVTIMPKDIQLARRIGERS

H3K9 acetylation peptide-**ARTKQTAR(acK)ST**

P. falciparum H3 sequence-M**ARTKQTARKST**AGKAPRKQLASKAARKSAPISAGIKKPHRYRPGTVALREIRRYQKSTDLLI
RKLPFQRLVREIAQDYKTDLRFQSSAVMALQEAAEAYLVGLFEDTNLCAIHAKRVTIMPKDIQLARRIGERS

H3K9 trimethylation peptide -**AR(me3K)S**

P. falciparum H3 sequence-MARTKQT**ARKS**TAGKAPRKQLASKAARKSAPISAGIKKPHRYRPGTVALREIRRYQKSTDLLI
RKLPFQRLVREIAQDYKTDLRFQSSAVMALQEAAEAYLVGLFEDTNLCAIHAKRVTIMPKDIQLARRIGERS

H3K36 trimethylation peptide -**GV(me3K)KP**

P. falciparum H3 sequence- MARTKQTARKSTAGKAPRKQLASKAARKSAPISA**GIKKP**HRYRPGTVALREIRRYQKSTDLLI
RKLPFQRLVREIAQDYKTDLRFQSSAVMALQEAAEAYLVGLFEDTNLCAIHAKRVTIMPKDIQLARRIGERS

H4K8 acetylation peptide- **GG(acK)GL**

P. falciparum H4 sequence- MSGRGK**GKGGL**GKGGAKRHRKILRDNIQGITKPAIRRLARRGGVKRISGLIYEEIRGVLK
VFLENVIKDSIMYTEHAKRKTVTAMDIVYSLKRQGRITLYGFGG

H4K12 acetylation peptide- **LG(acK)GG**

P. falciparum H4 sequence- MSGRGKGKG**GLGKGG**AKRHRKILRDNIQGITKPAIRRLARRGGVKRISGLIYEEIRGVLK
VFLENVIKDSIMYTEHAKRKTVTAMDIVYSLKRQGRITLYGFGG

H4K20 trimethylation peptide-**HR(me3K)VL**

P. falciparum H4 sequence- MSGRGKGKG**GLGKGGAKRHRKIL**RDNIQGITKPAIRRLARRGGVKRISGLIYEEIRGVLK
VFLENVIKDSIMYTEHAKRKTVTAMDIVYSLKRQGRITLYGFGG

Figure 3.16: Human peptides are conserved in *Plasmodium* H3/4 proteins.

The human peptide used for antibody production. There is sequence conservation between the human and *Plasmodium* H3/H4 peptides. The peptide sequences are available from the Abcam or Millipore websites.

A western blot was also run using nuclear extract from *P. falciparum* (Figure 3.17). The figure shows the western blot results and that the commercial antibodies recognize the histone modification in *P. falciparum*. We did observe two bands with the H4K8ac antibody. The nature of the top band is unknown but is probably a modified nuclear non-histone protein. This protein would be digested by the Proteinase-K used in ChIP protocol before DNA purification. The bands for H3K9me3 and H3K36me3 are faint in this blot. Previous work has shown that these two modifications are not abundant in the *P. falciparum* genome enriching in the subtelomeric regions only as opposed to the other modifications which are abundant (Lopez-Rubio et al. 2009; Jiang et al. 2013) and this could explain the faint bands observed. Expected band sizes are for H3 are 15kDa but usually, 17kDa bands are observed. H4 expected sizes are 11kDa but modifications and gel conditions affect this. Observed bands are usually above 13kDa.

Figure 3.17: Western blot for histone modification proteins.

A western blot was performed using the *P. falciparum* nuclear extract material. 2µl sample was used for the 12% blot. The same membrane was used and was cut into strips for incubation with specific antibody. The stripes were washed together and exposed at the same time. *This is not a blot to quantitate the amount of modification but rather to show that our commercial antibodies do recognize the modifications in *P. falciparum*.

Antibodies were also validated by looking at genes where their enrichment is known. For this, active gene seryl-tRNA synthetase and a silent *var* (*var2csa*) gene were used. Seryl-tRNA synthetase is a housekeeping gene transcribed across the IDC. As seen in Figure 3.18, the gene is devoid of modification associated with gene suppression in *P. falciparum*, H3K9 trimethylation, but is enriched for modifications associated with activation and modifications enriched in the euchromatic regions which include H3/H4 acetylation as well as H3K4 methylations (Salcedo-Amaya et al. 2009; Gupta et al. 2012; Karmodiya et al. 2015). We also observed H4K20 trimethylation highly enriched in this gene.

The enrichment of *var2csa* which is used as the model for virulence gene epigenetic footprint (Lopez-Rubio et al. 2007; Figure 1.7) was also determined. In a microarray analysis, this gene was found to be switched off in clones studied (Figure 3.19A). This gene in the silenced state is enriched for H3K9 trimethylation and devoid of H3K9 acetylation (Figure 3.19B). H3K36 trimethylation is devoid at this gene (comparable levels of IgG). In the original paper, they did not look at this modification so its enrichment is unknown. I also looked at H3K4 dimethylation and trimethylation. In the original paper, when the gene was silent, the H3K4 di/trimethylation modification profiles are similar to those in figure 3.19B clone 3.2C with a uniform distribution of the marks in the 5'UTR and coding region. I did observe differences of H3K4 di/trimethylation in clone 3.2C and 6.2A possibly pointing to clonal differences for this gene as the Seryl-tRNA synthetase showed the same profile for both clones.

Figure 3.18: Histone modifications enriched on the seryl-tRNA synthetase gene.

Pull-down assays were done in duplicates for the gene using a primer located in the coding region. The same primer was used in the analysis of RT-PCR. Silent marks are devoid at this gene locus. Orf = open reading frame. 6.2A and 3.2C are two isogenic clonal lines.

K9A- Histone 3 lysine 9 acetylation, K93- Histone 3 lysine 9 trimethylation, K8A- Histone 4 lysine 8 acetylation, K412A- Histone 4 lysine 12 acetylation, K36- Histone 3 lysine 36 trimethylation, K42- Histone 3 lysine 4 dimethylation, K43- Histone 3 lysine 4 trimethylation, K20- Histone 4 lysine 20 trimethylation

Figure 3.19: Epigenetic mapping of silent var2csa in clonal lines

A: PF0030c is switched off in the clonal lines used. Microarray analysis of the *var* genes in the clonal lines used in this thesis. This analysis is for the *var* genes only. During the trophozoite stage of the IDC, *var* genes are transcriptionally silent. In this analysis, the *var* genes are silenced with a few *var* showing up-regulation. Importantly var2csa- PFL0030c is transcriptionally silent in all four clones.

B: Histone modifications associated with silent var2csa. Primers used for this study are the same primers used in the original study. 2 primers in the original study were skipped to make 6 instead of 8 reactions. The analysis shows 2 biological replicates and the results are normalized to H3.

Var transcriptional start site = 1400kb from ATG.

3B.4.1 Validation of PCR primers for *stevor* epigenetic analysis

With the validation done, I looked at *stevor* genes and their enrichment. To start with the *stevor* epigenetic analyses the integrated line (PF3D7_1254100) was initially used. 12 specific PCR primers spanning both 5' UTR and the *stevor-hdhfr* transgene were designed (Figure 3.20A and 3.20B). Twelve PCR reactions were then performed on PF3D7_1254100 in both AF5^{ON} (PF3D7_1254100^{ON}) and AF5^{OFF} (PF3D7_1254100^{OFF}). The designed primers gave products ranging from 150-250bp.

A

B

Figure 3.20: Schematic showing position and specificity of the 12 PCR primers used.

A: Schematic representation of the position of PCR primers used for immunoprecipitated material. The primers overlapped in the 5' UTR region (●). In the coding region, the primers did not overlap each other. B: Primers were checked for specificity by blasting, melting curve analysis and gel of PCR products. PCR was performed using input material from the Sonicated samples. PCR products ranged from 150bp to 250bp. Each primer pair when run on the gel after amplification gives a single band. ● = primer overlap.

3B.5 EPIGENETIC LANDSCAPE OF STEVOR

3B.5.1 H3K9 acetylation and H3K4 trimethylation are enriched in 5' UTR of 3D7_1254100^{ON}

Acetylation of histone tails is thought to neutralize the protein-protein interactions between adjacent nucleosomes and also between the nucleosomes and the DNA wrapped around the histones. Three acetylated residues were studied in relation to *stevor* regulation. The first was H3K9 acetylation which correlates with active transcription. H3K9 acetylation is significantly enriched in the 5' UTR of the PF3D7_1254100^{ON} gene (Figure 3.21) with 4 regions showing significant difference (primer A: $p = 0.005878$, B: $p = 0.026852$, value primer D: $p = 0.013822$, primer F: $p = 0.000337$). In the coding region, only one primer position showed a significant difference (primer H: $p = 0.000666$). The enrichment of H3K9 acetylation in the 5' UTR regions suggests a role for this modification in modulating the local chromatin structure i.e. favouring transcription.

Figure 3.21 H3K9 acetylation levels at 3D7_1254100 *stevor* gene.

Distribution of H3K9 acetylation levels along the *stevor* gene in 'on' culture (white bars) and 'off' culture (black bars) during trophozoite stage. Data was normalized to H3. Data shows average of 4 biological replicates. Error bars= SEM. * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$ between the off and on gene, t-test. ● = primer overlap.

Figure 3.22: H3K4 trimethylation levels at different primer positions.

Distribution of H3K4 trimethylation levels along the *stevor* gene in 'on' culture (white bars) and 'off' culture (black bars) during trophozoite stage. Data was normalized to H3. Data shows average of best 3 biological replicates. Error bars= SEM. * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$ between the off and on gene, students t-test. ● = primer overlap.

H3K4me3 is known to associate with actively transcribed genes, as during transcription H3 at position K4 becomes methylated. This methylation has been hypothesized to be important in the recruitment of HATs, proteins that acetylate histone lysine residues (de la Crux et al. 2005). In our pull-down assays, H3K4 trimethylation was significantly enriched in the 5' UTR region (primer B: $p = 0.032609$, primer C: $p = 0.007615$, primer F: $p = 0.024329$, and primer G: $p = 0.018063$) (Figure 3.22). In the coding region primer H and I showed a significant difference (primer H: $p = 0.000337$ and primer I: $p = 0.032574$). Again like H3K9 acetylation, H3K4 trimethylation is significantly enriched in the 5' UTR regions of the active *stevor* gene as well as the coding regions. This enrichment is consistent with the hypothesis that H3K4 trimethylation is involved in the recruitment of HATs complexes although this has not been formally shown to happen in *P. falciparum*.

3B.5.2 H3K4me2 enriches in the coding region of 3D7_1254100^{ON} *stevor*

Figure 3.23 H3K4 dimethylation levels in 3D7_1254100^{OFF} and 3D7_1254100^{ON}.

Distribution of H3K4 dimethylation levels along the *stevor* gene in 'on' culture (white bars) and 'off' culture (black bars) during trophozoite stage. Data was normalized to H3. Data shows an average of 3 biological replicates. This was the only modification where only 3 biological replicates were performed. Error bars= SEM. * = $p < 0.05$, ** = $p < 0.01$, between the off and on gene, t-test. ● = primer overlap.

H3K4 dimethylation (H3K4me2) modification also associates with actively transcribed genes.

This is a general feature of this modification as it has been also been observed in humans, yeast

and *drosophila* (Pokholok et al. 2005; Kim et al. 2007; Araki et al. 2009; Schübeler et al. 2004; Bernstein et al. 2002; Bernstein et al. 2005). This modification has also been linked to epigenetic memory and is thought to allow for a permissive chromatin state where genes are in a poised state of activation (Ng et al. 2003; Lopez-Rubio 2007; Santos-Rosa et al. 2002; Schneider et al. 2004).

ChIP analyses for this modification showed it to associate significantly with the 3D7_1254100^{ON} *stevor* gene in the coding region (Figure 3.23) where there was a statistical difference between 3D7_1254100^{ON} and 3D7_1254100^{OFF} *stevor* (primer H: p =0.013874, primer I: p = 0.018423, primer J: p =0.008219, primer K: p =0.03104) (Figure 3.23). In fact, there was a minimal enrichment in the 5' UTR of this gene, with primers B to D showing very minimal enrichment. Both H3K4 dimethylation and trimethylation are catalyzed by the PfSET10 enzyme in *P. falciparum* (Volz et al. 2012). The fact that the two modifications are enriched in different regions of the same gene suggests a different mode of recruitment for PFSET10. In yeast, an extensively studied organism, H3K4 dimethylation is enriched in the coding regions of genes as it is proposed to be involved in transcription elongation (Santos-Rosa et al. 2002; Kim et al. 2007). This suggests that H3K4 trimethylation and H3K4 dimethylation probably play different roles in the regulation of 3D7_1254100, with H3K4 trimethylation showing broad enrichment and H3K4 dimethylation associating with transcription elongation events.

3B.5.3 H3K9me3 and H3K36me3 are selectively enriched in the 3' ends of 5' UTR and coding region of 3D7_1254100^{OFF}

Previous analysis of the H3K9me3 modification in *P. falciparum* using ChIP-chip technique has linked its enrichment to be restricted to virulence gene families (*var*, *rifin*, *stevor*) located in the subtelomeric regions of the parasite chromosomes (Salcedo-Amaya et al. 2009, Flueck et al. 2009, Lopez-Rubio et al. 2009). This modification is thought to silence the expression of these genes and thus dictate their transcription. H3K9me3 as a mark for silenced genes is not a unique *P. falciparum* observation. This modification is also enriched for silent genes in other organisms, such as developmental genes in *Drosophila* and also on facultative heterochromatin regions of mammalian genes.

In our analysis, H3K9me3 was significantly enriched for 3D7_1254100^{OFF} (Figure 3.24) compared to the 3D7_1254100^{ON} (primer H: p = 0.03203, primer I: p=0.007026, primer J: p=0.012481, primer K: p=0.013022 and primer L: p=0.010338). The enrichment was more pronounced in the coding region (primers H-L) with minimal enrichment in the 5' UTR region, apart from primer locations B and E where 3D7_1254100^{OFF} was significantly off as compared to 3D7_1254100^{ON}. Moreover, H3K9 trimethylation covers the entire coding region and not just segments of it perhaps pointing to the role of nucleosomes in the coding region as regulators of transcription as well.

Figure 3.24: H3K9 trimethylation levels in 3D7_1254100^{OFF} and 3D7_1254100^{ON}.

Distribution of H3K9 trimethylation levels along the *stevor* gene in ‘on’ culture (white bars) and ‘off’ culture (black bars) during trophozoite stage. Data was normalized to H3. Data shows an average of best 3 biological replicates out of a total 5 replicates. Error bars= SEM. * = $p < 0.05$, ** = $p < 0.01$, between the off and on gene, t-test. ● = primer overlap.

Figure 3.25: H3K36 trimethylation levels in 3D7_1254100^{OFF} and 3D7_1254100^{ON}.

Distribution of H3K36 trimethylation levels along the *stevor* gene in 'on' culture (white bars) and 'off' culture (black bars) during trophozoite stage. Data was normalized to H3. Data shows an average of best 3 biological replicates from 4 replicates. Error bars= SEM. * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$ between the off and on gene, t-test. ● = primer overlap.

H3K36 trimethylation had a similar profile to H3K9 trimethylation (Figure 3.25) suggesting that these two modifications potentially cluster together. A study looking at the global enrichment of this modification showed that some *stevors* are enriched for H3K36 trimethylation (Jiang et al. 2013), while another group which studied the same modification did not come to this conclusion (Karmodiya et al. 2015). In the latter publication, they grouped H3K36 trimethylation as a modification that enriched for active or poised genes. Our study supports the observation of Jiang et al. This modification was more enriched for 3D7_1254100^{OFF}*stevor* as compared to 3D7_1254100^{ON} (Figure 3.25). The enrichment was again more specific in the coding region with significant enrichment in this region (primer H: p value= 0.031741, primer I; p-value = 0.042746, primer K; p-value = 6.07E-06 and primer L; p-value = 0.001644). In the 5' UTR, there is significant enrichment of this modification for 3D7_1254100^{OFF}, at primer positions A, D, and G. No enrichment was observed at primers B, E, and G for both 3D7_1254100^{ON} and 3D7_1254100^{OFF}.

3B.5.4 H4K20 trimethylation does not show specific differences between 3D7_1254100^{ON} and 3D7_1254100^{OFF}

Figure 3.26: H4K20 trimethylation levels at 3D7_1254100^{OFF} and 3D7_1254100^{ON}.

Distribution of H4K20 trimethylation levels along the *stevor* gene in ‘on’ culture (white bars) and ‘off’ culture (black bars) during trophozoite stage. Data was normalized to H3. Data shows an average of 2 biological replicates. Because the enrichment was the same between the ‘on’ and ‘off’ there was no need to repeat a third time. Error bars= STD. ● = primer overlap.

H4K20 trimethylation has been shown to be involved in multiple functions, including gene silencing (Nishioka et al. 2002), regulation of mitosis in metazoans cell cycle (Karachentsev et al. 2005), DNA checkpoint control in *Schizosaccharomyces pombe* (Sanders et al. 2004) and formation of pericentric heterochromatin (Schotta et al. 2004). Sautel et al. characterized this modification in *P. falciparum* and *Toxoplasma gondi*, and localized the modification to repetitive DNA and telomeres in *Toxoplasma* and to heterochromatin regions in *P. falciparum* (Sautel et al. 2007). Earlier work on H4K20 trimethylation, showed it to be enriched in the subtelomeric regions as well as euchromatic regions (Salcedo-Amaya et al. 2009). A recent study in *P. falciparum* showed H4K20 trimethylation, to negatively correlate to active genes, suggesting its potential role in gene silencing (Gupta et al. 2012). Jiang et al. (2013) showed that H4K20 trimethylation is enriched at silent *var* genes. In humans and *drosophila*, H4K20 trimethylation also negatively correlates with transcription (Wang et al. 2008; Nishioka et al. 2002) potentially linking this modification to transcription repression.

In our ChIP analyses, we could not correlate this modification with activation or silencing as the enrichment pattern was the same for 3D7_1254100^{ON} and 3D7_1254100^{OFF} (Figure 3.26). Again, as with the other studied modifications (except for H3K9 acetylation), more enrichment can be observed in the coding region of the *trans-stevor* gene. The fact that this modification is still enriched at this *stevor* gene suggests that it might be serving a non-activation or silencing role. In fact, H4K20me3 has been linked to maintenance of genomic integrity in mice (Oda et al. 2009) and *drosophila* (Sakaguchi & Steward 2007). The modification has also been linked to heterochromatin telomere maintenance (Schotta et al. 2004) as well as involvement in compaction of the genome at least in vitro (Lu et al. 2008; Everetts et al. 2013) and could possibly explain its enrichment at this *stevor* gene locus.

3B.5.5 H4 acetylation does not correlate to activation or repression of 3D7_1254100

H4K8 and 12 acetylation reactions are part of the H4ac modifications which also includes H4K5 and 16. H4ac as the other acetylation modifications correlates with transcription (Wang et al. 2012; Karmodiya et al. 2015). H4K8 acetylation was of interest because it has been shown to be enriched in the promoter regions of active genes in *P. falciparum* which appears to be a unique feature of this organism (Gupta et al. 2012). The enrichment of H4K8 acetylation between 3D7_1254100^{ON} and 3D7_1254100^{OFF} was the same for all analyzed regions except perhaps for position A, where there was more enrichment in the 3D7_1254100^{OFF} as compared to 3D7_1254100^{ON} (Figure 3.27). For H4K12 acetylation (Figure 3.28) there was also no significant enrichment between the primer positions for 3D7_1254100^{ON} and 3D7_1254100^{OFF} suggesting that this modification is not involved in activation or repression of 3D7_1254100 *stevor* gene.

Figure 3.27: H4K8 acetylation levels at different primer positions.

Distribution of H4K8 acetylation and IgG levels along the *stevor* gene in ‘on’ culture (white bars) and ‘off’ culture (black bars) during trophozoite stage. Data was normalized to H3. Data shows an average of 2 biological replicates. Error bars= SEM.

Figure 3.28: H4K12 acetylation levels at different primer positions.

Distribution of H4K12 acetylation and IgG levels along the *stevor* gene in ‘on’ culture (white bars) and ‘off’ culture (black bars) during trophozoite stage. Data was normalized to H3. Data shows an average of 2 biological replicates. Error bars= SEM.

3B.5.6 H3K79 trimethylation and H4K20 monomethylation are devoid at *stevor* loci

Two other histone modifications studied, H3K79 trimethylation and H4K20 monomethylation, did not enrich for any of the regions (Figures 3.29 and 3.30). In fact, their enrichment levels were slightly above or similar to the IgG negative antibody used. H3K79me3 was studied because of an earlier publication which localized this modification in the periphery of the nucleus in *P. falciparum* (Issar et al. 2009). Subtelomeric regions of the chromosomes localize to this region (Scherf et al. 2001; Freitas-Junior et al. 2000). In this subtelomeric region is where *stevor* is located (Gardner et al. 2002). In Issar et al. publication they hypothesized that the unique localization of this modification could potentially be a new subnuclear compartment of the parasite, as electron microscope images had shown it to be a euchromatic region (Issar et al. 2009). H4K20 monomethylation was of interest because of its previous correlation to transcription. However, no significant differences were observed from these modifications in our studies, in terms of *stevor* activation or repression and also their enrichment levels were similar to the IgG negative control used (Figure 3.29 and 3.30 IgG graphs).

Figure 3.29: H4K20 monomethylation levels at different primer positions.

Distribution of H420 mono-methylation and IgG levels along the *stevor* gene in 'on' culture (white bars) and 'off' culture (black bars) during trophozoite stage. Data was normalized to H3. Data shows an average of 3 biological replicates. Error bars= SEM. ● = primer overlap.

Figure 3.30: H3K79 trimethylation levels at different primer positions.

Distribution of H3K79 trimethylation and IgG levels along the *stevor* gene in 'on' culture (white bars) and 'off' culture (black bars) during trophozoite stage. Data was normalized to H3. Data shows an average of 3 biological replicates. Error bars= SEM. ● = primer overlap.

3B. 5.7 PF3D7_1254100 endogenous *stevor* shows similar enrichment for silencing marks

The transgenic line we used for these experiments had multiple copies of the plasmid integrated. Therefore we were looking at the average of events in both cells lines because of multiple copies of the plasmid. We were interested in using a non-integrated clonal line where the endogenous *stevor* was also silenced. We used clone 6.2A (Chapter 3.C) for this analysis and we concentrated our analysis on H3K9ac, H3K9/36me3 modifications as these have been well studied in terms of virulent genes in *P. falciparum* (Figure 3.31). For this analysis, we had to redesign new primers for the coding region as this was the endogenous *stevor* and not *hdhfr/gfp*. We used the same primers from the 5' UTR (primers B, D, E/F and G). Primers E/F was combined to give a larger PCR product.

Looking at the silent endogenous gene, the marks that associate with silent transgene namely H3K9/36me3, are enriched and concentrate more in the coding region (Figure 3.31) of the single copy gene. This was similar to what was observed for the transgene for these two marks (Figures 3.24 and 3.25). This was reassuring as this shows that the integrated line used is still a valid tool to use. H3K9/36me3 are also enriched in the 5'UTR at primers B and D (H3K36me3) and primer D (H3K9me3), in a similar trend what was seen for the transgene.

Figure 3.31: The expression profile of the endogenously expressed PF3D7_1254100 *stevor*.

For this analysis, we designed new primers for the coding region and used primers B, D, E/F and G from the transgene. Primer position refers to the primers used in the original transgene analysis. Primers B, D, E/F and G are in the 5' UTR regions. Primers H to K are in the coding region. IgG= Immunoglobulin control, H3K9ac = H3K9 acetylation, H3K9me3 = H3K9 trimethylation, H3K36me3= H3K36 trimethylation.

3B.6 CHAPTER CONCLUSION AND DISCUSSION

In this, the first study of its kind, active histone modifications associated with regulation of the *stevor* multigene family were identified. An integrated line, *pBcam/stevor*, where a drug selectable cassette had integrated at the 3D7_1254100 locus was used for this initial analysis of *stevor* epigenetic regulation. The *pBstevor* line had a 3D7_1254100 *stevor* 5' UTR (2kb upstream of ATG codon) driving an *hdhfr* gene fused to *gfp* (*hdhfr-gfp*). Thus the *hdhfr-gfp* was under the control of the endogenous *stevor* promoter. Limiting dilution step was performed and clone AF5 was chosen for further experiments as it also expressed another endogenous *stevor*, PF3D7_1040200. This second *stevor* gene would allow the study the epigenetic footprint of a different *stevor* in a 'natural' non- artificial system (Part 3C).

AF5 was grown under constant drug pressure, using WR99210, which was termed AF5^{ON} as the *stevor* transgene is transcriptionally active. Another batch of AF5 was grown without drug pressure and this was termed AF5^{OFF} as the *stevor* transgene is expected to be expressed at much lower levels. The 'on' and 'off' system was validated using live-cell imaging and quantitative PCR (qPCR) showed that AF5^{ON} had higher *hdhfr* transcripts as opposed to the off system. Indeed, there was more than a 5 fold reduction in the *hdhfr* transcript levels in AF5^{OFF} compared to the 'on' system. Using live cell imaging, AF5^{ON} showed higher levels of green fluorescence as compared to the AF5^{OFF} line.

I set out to perform immunoprecipitation assays using the AF5^{ON} and AF5^{OFF} cell lines. Different histone modifications that have been studied in other organisms as well as in *P. falciparum* were selected. These included H3K9, H4K8 and 12 acetylations, H3K4 di and trimethylation, H3K9, K36 and 79 trimethylation and H4K20 mono and trimethylation. This initial analysis allowed the grouping of the modifications based on their enrichment in the 3D7_1254100^{ON} and

3D7_1254100^{OFF} systems. This allowed prediction of the roles of these histone modifications with regards to *stevor* regulation.

Three modifications were associated with the 3D7_1254100^{ON} in AF5^{ON} system and are possibly involved in gene regulation and activation of the *stevor* genes. These modifications were H3K9 acetylation, H3K4 di, and trimethylation. This is the first time these three modifications have been associated with active *stevor*. Interestingly, these modifications were enriched at different regions, suggesting they play different roles in *stevor* regulation. H3K9 acetylation was significantly enriched in the 5' UTR region from a region 800bp before the ATG codon of the gene. This suggests that it keeps the chromatin in this region permissable for *stevor* transcription factors to bind. This observation is in line with the supposed role of H3K9 acetylation as a hallmark for active chromatin regions across all species (Schübeler et al. 2004; Bernstein et al. 2005; Wang et al. 2008; Schiltz et al. 1999; Kurdistani et al. 2004; Karmodiya et al. 2012; Karmodiya et al. 2015; Fukuda et al. 2006).

H3K4 trimethylation showed a preferential enrichment in the 5' UTR, as well as at the 5' end of the coding region with a higher enrichment status in the coding regions (400bp before ATG to 500bp after ATG codon). However, H3K4 dimethylation is enriched in the coding as opposed to the 5' UTR regions (200b before ATG to 800bp after ATG). Their enrichment in the active *stevor*, suggests that these H3K4 methylation marks are involved in some aspect of *stevor* activation or transcription of the mRNA. Studies show that H3K4 methylations are enriched on active genes though the direct roles for H3K4 methylation have not been fully dissected.

They have been suggested to serve secondary regulatory roles instead of having direct involvement in transcriptional regulation. H3K4 trimethylation is usually observed in the 5' UTR regions of active genes. Some studies have shown that recruitment of HATs like SAGA (Spt-Ada-Gcn5 acetyltransferase) and SLIK (SAGA-like), is dependent on the chromatin remodeling protein Chd1 (chromo-ATPase/helicase-DNA-binding domain 1) recognizing H3K4 modifications on genes (Bian et al. 2011; Pray-Grant et al. 2005). Chd1 is part of these HAT complexes. However, this has only been confirmed in humans and not in yeast (Flanagan et al. 2005; Sims et al. 2005). Another regulatory role would be the recruitment of RNA polII and modulation of nucleosome stability during RNA Polymerase II transit (Wang 2009; Chen et al. 2011). Indeed experiments using yeast show that, although these modifications are enriched in promoters of active genes, loss of the Set1 (which catalyzes methylation of H3K4) leads to very subtle changes in gene expression (Lenstra et al. 2011). Another study again in yeast suggests the role of H3K4 methylation in recruitment an ATPase that remodels chromatin for passage of RNA polymerase II (Santos-Rosa et al. 2003). These observations from yeast could, in part, explain the broad distribution of H3K4 methylation for 3D7_1254100

H3K9 trimethylation, as well as H3K36 trimethylation, were significantly enriched in 3D7_1254100^{OFF} as compared to 3D7_1254100^{ON} potentially linking these modifications as transcription repressors. The enrichment was significant in the coding region of the gene for both modifications (Figure 3.32). We also observed a similar trend when we looked at the endogenous *stevor* gene. H3K9 trimethylation has earlier been shown to be enriched at *stevors*, *vars*, *rifins* and other multigene families on a global scale (Lopez-Rubio et al. 2009; Salcedo-Amaya et al. 2009). We confirm these finding in this study.

H3K4 dimethylation, H3K9 trimethylation, and H3K36 trimethylation showed a preferred enrichment in the coding region in our analysis. This may suggest that histones in the coding region have a role in repression/activation of the *stevor-hdhfr-gfp* gene during the transcription elongation stage. Histones in the 5' end of coding regions have previously been shown to play a role in chromatin remodeling and gene regulation in *P. falciparum*. In their comprehensive analysis of histone modifications that drive regulation in *P. falciparum*, Gupta et al. (2012) also observed that most of the modifications they studied, except for H4K8 acetylation, were preferentially enriched in the 5' end of gene coding regions. These included H3K4 trimethylation, H4 acetylation, H3K56 acetylation and H3K9 acetylation. They hypothesize that preferred enrichment at the 5' end of coding regions could potentially highlight the role these nucleosomes have in chromatin regulation. This phenomenon is not unique to *P. falciparum*. In plants and yeast, most euchromatic modifications enrich in the 5' end of ORF (Liu et al. 2005), unlike in multicellular eukaryotes where gene regulation modifications preferentially enrich in the promoter or TSS regions (Li et al. 2008). From these results a schematic summarizing the findings is shown (Figure 3.32).

In the transgenic line used for this part of the thesis, the plasmid had integrated as multiple copies. In these multiple copies of the plasmid, some of the cassettes are silent and others are active. This suggests that the ChIP-qPCR results seen here for the 5' UTR as well as *hdhfr* gene reflect several identical sequences rather than those found at the single locus and I do not know if the modification profile at each drug cassette is identical.

Figure 3.32: A summary of modifications associated with regulation of 3D7_1254100 *stevor*.

Shown are the significant histone modifications correlating with activation and repression of this gene and the regions where they are enrichment as calculated from the ATG site. The gene was divided into 200bp regions as this is the average PCR product from the PCR primers used. *Note enrichment heights of the modifications not drawn to scale.

Rovira-Graells et al. did a comprehensive analysis of clones that had a single integration event. In their analysis, they do show that in a single integration event, which happens with multiple copies, some cassettes are active while others are off even in a culture where the drug is still applied (Rovira-Graells et al. 2015). Our contrast had a GFP tag and this allowed us to quantify the number of active cells in each line. In AF5^{ON} (82% green cells) fluoresced suggesting a significant number of cells in this line were transcriptionally active. About 18% of the cells did not fluoresce even though still under pressure suggesting they had become resistant to the drug or had mutations that had led to a non-functional GFP production or expressed GFP in very low levels like background levels. In the AF5^{OFF} line, 78% of the cells had their promoters off but some still fluoresced green. This suggested that some cells were still transcriptionally active. The RT-PCR result reflected this, where the relative expression of the transgene in AF5^{OFF} was still about 1 unit. In fact, using 0.2 as the cut off for expression (Calibri et al. 2012), the 3D7_1254100 transgene in AF5^{OFF} is 'on', but compared to the AF5^{ON} it is significantly off and this is reflected in the enrichments of the modifications that associate with active and silent genes.

The fact that we are looking at events on multiple copies of the same sequence and that in the lines we did not have 100% active cells in AF5^{ON} and 100% silent cells in AF5^{OFF} in no way negates the value of transgenic lines for the study epigenetic mechanisms in the field of malaria especially where variant multigene families are concerned. In a similar study to this, a *var* transgene was used to show that *vars* are enriched with H3K9 trimethylation when silent (Chookajorn et al. 2007). This was the first study to show this and to also that the H3K9 trimethylation mark is stably inherited. Other studies using transgenic lines have deciphered the epigenetic mechanisms of other multigene families involved in invasion (Cortés et al. 2007; Rovira-Graells et al. 2015). Therefore, transgene lines do have their place. In this study, when

the modifications associated with the silent transgene, H3K9/36 trimethylation, were assayed at the endogenous locus, I still observed a similar enrichment profile, with the modifications enriching more in the coding regions agreeing with the transgene (Figure 3.31).

In this part of the work as well, the modifications associated with active transcription observed were similar when compared to the tRNA synthetase, a housekeeping gene. H3K9 acetylation and H3K4 methylations were seen to enrich when the transgene was 'on' as is seen for the tRNA-synthetase where these modifications are also enriched. For tRNA synthetase, H3K9/36 trimethylation was devoid at the locus and these were enriched when the transgene was off in AF5^{OFF} and at the endogenous locus. Therefore, this work provided the groundwork to work from and suggests that this system is still an effective tool to work with.

In retrospect, a further improvement in this study, apart from using endogenously expressed genes or using a line with a single integrated copy, would have been to use a clonal line with a much lower expression of *hdhfr* like the AB4 clone (Figure 3.11). An even better approach would have been to perform a re-limiting dilution of the derived clonal lines as done by other studies and then assayed these re-cloned lines for a silent clone (Rovira-Graells et al. 2015). Then use this clone as the 'off' line, split it into two and grow the other half on WR99210 drug. This would not help in reducing the plasmid copy numbers, as the transgene would still be available as multiple copies but it would help with having a more uniform expression state for each line. Still, with all these flaws in the system, we still observed a similar profile for the endogenous gene when in the silent state. I then continued our analysis looking at endogenously expressed *stevors* (Chapter 3C).

CHAPTER 3C

STUDIES WITH ENDOGENOUS *STEVORS*

3C.1 INTRODUCTION

In the previous chapter, a parasite that had an integrated *stevor* line was used to study the histone modifications associated with active and silent *stevor*. Using this parasite line I also aimed to assess whether *stevor* undergoes mutual exclusion expression. Limiting dilution cloning was performed on the integrated parental line and from the resulting isogenic clones, clone AF5 was chosen for analysis (Chapter 3B-Figure 3.12). Clone AF5 endogenously expressed another *stevor*, 3D7_1040200 (Figure 3.12, AF5 panel). I was interested in studying the epigenetic footprint of 3D7_1040200 *stevor*. Additionally, I aimed to produce a parasite line where 3D7_1040200 *stevor* was silenced. Our other clonal lines from Figure 3.12 (BC10, BC2, AC9, and AB4) could have been used as they did not express 3D7_1040200. However, I wanted a clonal line that endogenously expressed a different *stevor* from 3D7_1040200 to have more *stevors* to study.

3C.2 TRANSCRIPTIONAL ANALYSIS OF 3D7 ISOGENIC CLONES

For this 3D7 isogenic clones previously produced in the lab were used. Samples were harvested at the trophozoite stage as previous work has shown that peak *stevor* expression occurs at this stage of the IDC (Kaviratne et al. 2002). All the analyzed clones expressed a subset of *stevor* (Figure 3.33), confirming earlier studies on *stevor* gene expression (Kaviratne et al. 2002; Sharp et al. 2006; Lavazec et al. 2007; Niang et al. 2009). Although, the subset of *stevors* expressed by

the different clones was the same, and the level of expression was variable. PF3D7_0617600, PF3D7_1100700 and PF3D7_1372500 were the expressed *stevor* in the different clones (Figure 3.33: 1.2A, 3.2B, 3.2C, 5B, 5.2A and 6.2A panels). These three *stevor* genes are different from other *stevor* genes which have been previously reported to be dominantly expressed by *Plasmodium falciparum* clones (Sharp et al. 2006; Lavazec et al. 2007; Sanyal et al. 2012) again confirming the clonally variant nature of laboratory cultured parasites. They are also different from the previous clonal lines (Figure 3.12).

From our analysis of the 3D7 isogenic clones (Figure 3.33) clones 6.2A and 3.2C were chosen because they both highly expressed 3D7_0617600 and 3D7_137200 compared to the other clones. 3D7_0617600 and 3D7_1372500 were both silent in the AF5 clone (Figure 3.12; AF5 panel, and Figure 3.34 panels B and C). 6.2A and 3.2C also transcribed 3D7_1040200 at very low levels. This would be the control for clone AF5 (Figure 3.12 and Figure 3.34A) where this same gene was highly expressed. Clone AF5 is the clone used in chapter 3B and has the integrated *stevor*. Clones 6.2A and 3.2C are 3D7 isogenic clones from our lab. For these three genes, we focused our analysis on H3K9 acetylation, H3K3 di and trimethylation, H3K9/K36 trimethylation, and H4 acetylation. We did not look at H3K79 trimethylation or H4K20 monomethylation as our earlier analysis showed them to be absent from *stevor* genes (Chapter 3B).

Figure 3.33: *Stevor* transcriptional analysis of 3D7 clones.

The analysis was performed at early trophozoite stage (26-30hpi) for 5 clones from limiting dilution. Results for 5 3D7 clones, 3.2C, 3.2B, 5.2A, 5B and 6.2A, are shown. All clones expressed the same subset of *stevors*, at different levels. Clones 6.2A and 3.2C expressed a *stevor* each the cutoff of 0.2 units. All transcripts were normalized relative to *seryl-tRNA synthetase* gene. Error bars are SEM.

Figure 3.34: Relative expression profiles of PF3D7_1040200 (A) PF3D7_0617600 (B) and PF3D7_1372500 in 6.2A, 3.2C and AF5 clones.

Panel **A** compares the expression of PF3D7_1020400 in clones 6.2A, 3.2C and AF5. In clone AF5 PF3D7_1020400 is switched on and is silenced in clone 6.2A and 3.2C. Panel **B** shows the expression pattern for PF3D7_0617600 in the same three clones with the gene expressed in 6.2A and 3.2C and silenced in AF5. Panel **C** shows the expression pattern for PF3D7_1372500. Results show an average of 3 biological q-PCR reactions using samples from trophozoite stage with *seryl-tRNA-synthetase* used as a control. Error bars are SEM.

3C.3 EPIGENETIC MAP FOR 3D7_1040200

3C.3.1 Validation of PCR primers

3D7_1040200 was highly expressed in the AF5 clone but was silent in 3.2C and 6.2A. Clone 6.2A was used as the silent control (Figure 3.34A). *stevor* PCR primers were designed spanning the coding region of the gene as well as a region 900bp from the ATG site as shown in plasmodb.org. This gave a total of 8 PCR regions (Figure 3.35 panels A and B).

A

B

Figure 3.35: Schematic and specificity of PCR reactions for PF3D7_1040200:

Panel A: The endogenous size of the 3D7_1040200 gene is around 1000bp. Primers were designed to cover at least 900bp of the gene and upstream region. Red dots: overlap region of primer pairs. B: Primers against the 5' UTR and coding regions of PF3D7_1040200 were designed with the primer blast function from NCBI. Sonicated material from the ChIP samples was used to perform PCR with the primers to ensure specificity. As can be seen, only one band is observed on the ethidium bromide gel. Lane symbols refer to primer position as shown in Figure 3.35A. ● = primer overlap.

3C. 3.2 H3K9 and H4 acetylation differentially associate with active 3D7_1040200

Figure 3.36: H3K9 acetylation levels in 3D7_1040200 on and off states.

This is the average of three independent cross linked samples. Two clones, 6.2A and AF5, were used for this analysis. In clone 6.2A, 3D7_1040200 is 'off' and in clone AF5, 3D7_1040200 is 'on'. Data was normalized to H3. Error bars= SEM. Primer A-D are in the 5' UTR region and primers E-H in the coding region. * = p value < 0.05, ** = p value < 0.01. ● = primer overlap.

Like our previous work which shows H3K9 acetylation is enriched when *stevor* gene is on, H3K9 acetylation is also enriched in 3D7_1040200^{ON} as opposed to 3D7_1040200^{OFF} (Figure 3.36). The enrichment was significant across the 5' UTR (primer A: p=0.004745, primer B: p=0.009455, primer C: p=0.001604, primer D: p=0.007228) as well as the coding region (primer E: p=0.000688, primer F: p=0.009484, primer G: p=0.00952, primer H: p=0.016709) highly suggestive of a very open chromatin structure throughout the gene structure.

Figure 3.37: H4K8 acetylation levels in 3D7_1040200 on and off states.

This is the average of three independent cross linked samples. Two clones, 6.2A and AF5, were used for this analysis. In clone 6.2A, 3D7_1040200 is 'off' and in clone AF5, 3D7_1040200 is 'on'. Data was normalized to H3. Error bars= SEM. Primer A-D are in the 5' UTR region and primers E-H in the coding region. * = p value < 0.05, ** = p value < 0.01. ● = primer overlap.

In contrast to the transgene, 3D7_1040200^{ON} gene H4K8 acetylation associated with the active gene and was significantly enriched in the coding regions (primer E: p=0.024839, primer F: p=0.046356, primer G: p=0.007505, primer H: p=0.012204) acetylation (Figure 3.37). The 5' UTR regions showed relatively similar enrichment profile between 3D7_1040200^{ON} and 3D7_1040200^{OFF}. This modification was shown to correlate with active genes in *P. falciparum*, but to be more enriched in the 5' UTR regions (Gupta et al. 2012). Enrichment in the coding regions could be a unique property of virulent genes.

Figure 3.38: H4K12 acetylation levels in 3D7_1040200 on and off states

This is the average of three independent cross-linked samples. Two clones, 6.2A and AF5, were used for this analysis. In clone 6.2A, 3D7_1040200 is 'off' and in clone AF5, 3D7_1040200 is 'on'. Data was normalized to H3. Error bars= SEM. Primers A-D are in the 5' UTR region and primers E-H in the coding region. * = p-value < 0.05. ● = primer overlap.

H4K12 acetylation (Figure 3.38), showed a significant enrichment at primer A: $p = 0.004085$ in the 5' UTR and primers G: $p = 0.00186$ and H: $p = 0.02814$ in the coding region when the gene was active. All the other positions showed no difference between on and off though the H4K12 acetylation levels were at least 2-fold when the gene was active.

3C.3.3 H3K9me3 enriches in coding region of 3D7_1040200^{OFF}

Figure 3.39: H3K9 trimethylation levels in 3D7_1040200 on and off states.

This is the average of three independent cross linked samples. Two clones, 6.2A and AF5, were used for this analysis. In clone 6.2A, 3D7_1040200 is ‘off’ and in clone AF5, 3D7_1040200 is ‘on’. Data was normalized to H3. Error bars= SEM. Primer A-D are in the 5’ UTR region and primers E-H in the coding region. * = p value < 0.05, ** = p value < 0.01. ● = primer overlap.

ChIP analysis of H3K9 trimethylation showed it to be enriched in 3D7_1040200^{OFF} (Figure 3.39).

The enrichment was significant in the coding region of the gene (primer E: p = 0.001108, primer F: p= 0.02051, primer G: = 0.03862, primer H: p=0.027141). In the 5’ UTR region, the only significant region was at primer C (p = 0.018003). The pattern of enrichment for this gene mirrors that of the transgene. This shows H3K9 trimethylation to associate with silencing of another *stevor* and agrees with previous studies in *P. falciparum* which shows *stevors* to be enriched for H3K9 trimethylation (Salcedo-Amaya et al. 2009; Lopez Rubio et al. 2009)

3C.3.4 H3K4 di and trimethylation, H3K36 trimethylation and H4K20 trimethylation show no specific differences between 3D7_1040200^{ON} and 3D7_1040200^{OFF}

Figure 3.40: H3K4 trimethylation levels in 3D7_1040200 on and off states.

This is the average of three independent cross linked samples. In clone 6.2A, 3D7_1040200 is 'off' and in clone AF5, 3D7_1040200 is 'on'. Data was normalized to H3. Error bars= SEM.. * = p value < 0.05, ** = p value < 0.01. ● = primer overlap.

Figure 3.40 shows the H3K4 trimethylation profile for PF3D7_1040200. Specific enrichment for 3D7_1040200 was observed at position A (p value = 0.00337) and B (p value = 0.010289) in the off line. The on line had enrichment at primer C (p value = 0.039697). At all the other regions, there was no specific enrichment between 3D7_1040200^{ON} and 3D7_1040200^{OFF}.

Figure 3.41: H3K4 dimethylation levels in 3D7_1040200 on and off states.

This is the average of three independent cross-linked samples. Two clones, 6.2A and AF5, were used for this analysis. In clone 6.2A, 3D7_1040200 is 'off' and in clone AF5, 3D7_1040200 is 'on'. Data was normalized to H3. Error bars= SEM. Primers A-D are in the 5' UTR region and primers E-H in the coding region. ● = primer overlap.

H3K4 dimethylation profile for 3D7_1040200 did not show any differences between 3D7_1040200^{ON} and 3D7_1040200^{OFF} (Figure 3.41) except at one primer position in the 5' UTR (primer A: $p=0.02688$). This is in sharp contrast to the *stevor* transgene in the previous chapter (Figure 3.23), as this modification was enriched in the coding region when the transgene was transcriptionally active. It is interesting to observe this because this potentially suggests that different *stevor* use different mechanisms for their activation. The only difference between the 3D7_1254100 transgene and 3D7_10400 is their coding segment and their chromosomal position. 3D7_1040200 is located on chromosome 10 whereas the *stevor* transgene is integrated at chromosome 12.

H3K36 trimethylation (Figure 3.42) did not show any significant differences between the 3D7_1040200^{ON} and 3D7_1040200^{OFF}. In the previous chapter, this modification had associated with the silent form of the gene, so this was rather surprising. When I looked at the Jiang et al work, they also observed an enrichment of this modification on silent and active *var* genes in the coding regions. It was in the silent *var* where they observed additional enrichment in the 5' UTR regions (Jiang et al. 2013).

H4K20 trimethylation (Figure 3.43) did not show any significant differences between the 3D7_1040200^{ON} and 3D7_1040200^{OFF}. This is a similar pattern as observed for the transgene (Figure 3.26). The potential role of H4K20 trimethylation being linked to maintenance of genome integrity and telomeric maintenance has already been described.

Figure 3.42: H3K36 trimethylation levels in 3D7_1040200 on and off states.

This is the average of three independent cross-linked samples. Two clones, 6.2A and AF5, were used for this analysis. In clone 6.2A, 3D7_1040200 is 'off' and in clone AF5, 3D7_1040200 is 'on'. Data was normalized to H3. Error bars= SEM. Primers A-D are in the 5' UTR region and primers E-H in the coding region. ● = primer overlap.

Figure 3.43: H3K20 trimethylation levels in 3D7_1040200 on and off states.

This is the average of three independent cross-linked samples. Two clones, 6.2A and AF5, were used for this analysis. In clone 6.2A, 3D7_1040200 is 'off' and in clone AF5, 3D7_1040200 is 'on'. Data was normalized to H3. Error bars= SEM. Primers A-D are in the 5' UTR region and primers E-H in the coding region. ● = primer overlap.

3C.3.5 Conclusion for PF3D7_1040200

The epigenetic landscape for 3D7_1040200 shows differences to those of the transgene and suggests a slightly modified epigenetic regulation. Only four modifications are strongly linked to the regulation of this gene (Figure 3.44). Our analysis shows H3K9 acetylation associated with active the gene as this was significantly enriched in 3D7_1040200^{ON} along the 5' UTR and coding regions as opposed to 3D7_1040200^{OFF}. This result was like PF3D7_1254100 transgene except for the transgene, H3K9 acetylation enriched only in the 5' UTR region. Unlike PF3D7_1254100 which did not show any H4ac enrichment, PF3D7_1040200 shows H4 acetylation enrichment in the coding region of 3D7_1040200^{ON}. This suggests that PF3D7_1040200 uses more than one acetylation modification for activation and H4 acetylation could be involved in transcription elongation as its enriched in the coding region.

H3K9 trimethylation was significantly enriched in 3D7_1040200^{OFF} compared to 3D7_1040200^{ON}. The enrichment was more significant in the coding region just as the transgene pointing to the nucleosomes in the coding regions as being responsible for the *stevor* silencing. Surprisingly H3K4 dimethylation did not show any significant differences between the 'on' and 'off' states suggesting it is not involved in transcriptional regulation of 3D7_1040200. H4K20 trimethylation also did not specific difference in 3D7_1040200^{ON} and 3D7_1040200^{OFF} line. This gene showed a significant enrichment of most modifications only at one position. For example, H3K4me3 at position G and, H3K4me2 at position A. This pattern is not consistent to make a logical conclusion. If the significance had been at the same position in all the modifications then postulations could be made about the region, but as it is, it seems as noise. Looking back at the 6.2A clone used as the 'off' control for this gene the level of transcription is just at the 0.2 cutoff value used. Perhaps this could influence the results of some of these modifications.

Figure 3.44: A summary of modifications associated with regulation of 3D7_1040200 *stevor*.

Shown are the histone modifications correlating with activation and repression of this gene and the regions where they are enrichment as calculated from the ATG site. The gene was divided into 200bp regions as this the average PCR product from the PCR primers used. Note enrichment heights of the modifications are not drawn to scale.

3C.4 EPIGENETIC LANDSCAPE FOR 3D7_0617600

3D7_0617600 was highly expressed in the 3D7 clone 6.2A. It was also expressed in the 3.2C clone. The same gene was transcriptionally off in clone AF5 (Figure 3.34B). The three clones were used to study epigenetic regulation of 3D7_0617600. For this, PCR primers giving PCR products of between 200-300bp were designed to target 900bp of the 5' UTR and 800bp of the coding region of the gene (Figure 3.45 A and B) as was done for 3D7_1040200 and 3D7_1254100.

A

B

Figure 3.45: Schematic representation and specificity of primers used for ChIP analysis of 3D7_0617600.

A: The endogenous size of the 3D7_0617600 gene is around 900bp. Primers were designed to cover at least 900bp of the gene. For the upstream region, a similar approach was chosen with primers designed to cover at least 900bp of this region as well. For this analysis, primers overlapped each other as denoted by blue circles after each primer region. **B:** Primers were designed with the primer blast function from NCBI. As can be seen, only one band is observed on the ethidium bromide gel. Lane symbols refer to primer position as shown in Figure 3.45A. ● = primer overlap.

3C.4.1 H3 and H4 acetylation associate with 3D7_0617600^{ON}

Figure 3.46: H3K9 acetylation enrichment when PF3D7_0617600 is on and off.

Three clones, 6.2A, 3.2C and AF5, were used for this analysis. In clones 6.2A and 3.2C, 3D7_0617600 is 'on' and in clone AF5, 3D7_0617600 is 'off'. Data was normalized to H3. Error bars= SEM. Primer A-D are located in the 5' UTR region and primers E-H in the coding region. * = p value < 0.05, **=p value <0.01. ● = primer overlap.

ChIP analysis for 3D7_0617600 is shown in Figure 3.46 for both clones. In clone 3.2C, H3K9 acetylation significantly enriched in 3D7_0617600^{ON} at all the primer positions with the exception of primer H in the coding region. In clone 6.2A, however, we observe a significant enrichment at primer positions A and C in the 5' UTR and primer E in the coding region for 3D7_0617600^{ON}. The difference here could be a technical issue on the pull downs as the 3.2C pulldowns were done months after the 6.2A pull down or because the 6.2A H3K9 acetylation ChIP recoveries were much lower compared to 3.2C clone. We do actually observe higher H3K9 acetylation ChIP recovery in 3.2C clone (x-axis) for this gene. Again for a different *stevor*, H3K9 acetylation associates with an active gene.

We wanted to test if other acetylation marks, were significantly enriched when 3D7_0617600 was active. To test this, immunoprecipitation was carried out for both clones with antibodies against H4K8 and H4K12 acetylation modifications. Subsequently, PCR was performed to assess the levels of these two acetylation reactions in 3D7_0617600^{ON} and 3D7_0617600^{OFF} for both clones. ChIP-qPCR results are shown in Figure 3.47 and 3.48 for both clones. The H4K8 and 12 acetylation levels are highly enriched in the 3D7_0617600^{ON} compared to 3D7_0617600^{OFF} in both clones (Figure 3.47 and 3.48). H4K8 acetylation is specifically enriched in the 5' UTR regions for clone 6.2A (Figure 3.47 primer A to D). H4K8 acetylation enrichment levels in the coding region (primers E to H) are similar for 3D7_0617600^{ON} and 3D7_0617600^{OFF} in the 6.2A clone. In the 3.2C clone, the modification again significantly enriches across the gene body from 5' UTR to coding region. This suggests that this modification specifically enriches in the 5' UTR/ promoter/coding regions of this gene.

Figure 3.47: H4K8 acetylation levels across 3D7_0617600^{ON} and 3D7_0617600^{OFF}.

Three clones, 6.2A, 3.2C, and AF5, were used for this analysis. In clone 6.2A and 3.2C, 3D7_0617600 is 'on' and in clone AF5, 3D7_0617600 is 'off'. H4K8 acetylation shows specific enrichment across the 5' UTR of the active gene with little to no enrichment in the same region when the gene is silent. Data was normalized to H3. Error bars= STD. Primers A-D are located in the 5' UTR region and primers E-H in the coding region.* = p-value < 0.05, ** = p-value 0.01. ● = primer overlap.

Figure 3.48: H4K12 acetylation levels across 3D7_0617600^{ON} and 3D7_0617600^{OFF}.

Three clones, 6.2A, 3.2C, and AF5, were used for this analysis. In clones 6.2A and 3.2C, 3D7_0617600 is 'on' and in clone AF5, 3D7_0617600 is 'off'. Enrichment is observed when the gene is transcriptionally on except for primer position E where enrichment is observed when the gene is 'off'. Data was normalized to H3. Error bars= STD. Primers A-D are located in the 5' UTR region and primers E-H in the coding region. Error bars = STD. * = p-value < 0.05, ** = p-value 0.01, *** = p-value <0.001. ● = primer overlap.

H4K12 acetylation is enriched in 5' UTR (Figure 3.48) from primer A to C in the 5' UTR and primer E to G in the coding region for 3D7_0617600^{ON} for both clones. The result suggests an open chromatin for transcription in the 5' UTR extending towards the coding region. In the 6.2A clone, though, the higher levels of H4K12 acetylation are in the 5' UTR region and they reduce in the coding region. In the 3.2C clone, they acetylation levels are the same in both the 5' UTR and coding regions. It was interesting to observe this enrichment for these two modifications when the gene is on, as the transgene enrichment profiles for these two modifications did not correlate with activation or repression (Figures 3.27 and 3.28). The results of Figures 3.47 and 3.48 show that H4 acetylation significantly associates with 3D7_0617600^{ON}.

However, differences between the clones were observed especially for H3K9 acetylation and H4K8 acetylation in positions of enrichment. In clone 3.2C, enrichment was across the 5' UTR and gene body for both H3K9/H4K8 acetylation reactions, whereas in 6.2A clone, the same modifications enriched primarily in 5' UTR regions. The biological outcome would mostly be the same in all cases as there is a neutralizing of charges and the opening of chromatin (Kouzarides 2007) but these small differences could suggest that clonal differences potentially exist in terms of regulation and enrichment of modifications. Also, acetylation reactions are known to more dynamic compared to methylation reactions which are thought to be more stable. The small differences could also just be a reflection of the time when the samples were harvested. There was a 6 month period between harvesting the two samples and even though microscopically the cells were in the same stage, the timing could have been slightly different resulting in a slightly different epigenetic profile.

3C.4.2 H3K4 di and trimethylation enriched for 3D7_0617600^{ON}

The results for H3K4 di and trimethylation also show these modifications to be enriched on the 3D7_0617600^{ON} as opposed to the 3D7_0617600^{OFF} (Figure 3.49 and 3.50) in both 3.2C and 6.2A. H3K4 trimethylation modification has stayed constant for all three genes studied as it has consistently been enriched in active genes though at different genomic regions (Figure 3.22, 3.40 and 3.49). Enrichment levels are significant in the 5' UTR from primer position A to C and in the coding regions from primer E to F in 3D7_0617600^{ON} in clone 6.2A (Figure 3.49). A similar enrichment profile is seen for 3.2C clone, though at higher levels as compared to 6.2A but nonetheless the same enrichment profile is observed (Figure 3.49).

H3K4 dimethylation showed significant enrichment in 3D7_0617600^{ON} (Figure 3.50) as compared to 3D7_0617600^{OFF} in both clones as well. In clone 6.2A, two positions in the 5' UTR (primer B and C) showed significant enrichment in 3D7_0617600^{ON}. The majority of the coding region (primer position E to G) also showed significant differences in 3D7_0617600^{ON} compared to 3D7_0617600^{OFF}. This was the same for 3.2C except in the 5' UTR primer A, also showed significant enrichment.

Figure 3.49: H3K4 trimethylation levels across 3D7_0617600^{ON} and 3D7_0617600^{OFF}.

Two clones, 6.2A and AF5, were used for this analysis. In clone 6.2A 3D7_0617600 is 'on' and in clone AF5, 3D7_0617600 is 'off. H3K4 trimethylation shows specific enrichment across the 5' UTR of the active gene as well as the coding region. Minimal enrichment is observed when the gene is silenced. Error bars= STD. Primers A-D are located in the 5' UTR region and primers E-H in the coding region. Data was normalized to H3. Error bars = STD. * = p-value < 0.05, ** = p-value 0.01, *** = p-value 0.001. ● = primer overlap.

Figure 3.50: H3K4 dimethylation levels across 3D7_0617600^{ON} and 3D7_0617600^{OFF}.

Three clones, 6.2A, 3.2C, and AF5, were used for this analysis. In clones 6.2A and 3.2C 3D7_0617600 is 'on' and in clone AF5, 3D7_0617600 is 'off'. H3K4 dimethylation shows specific enrichment across the 5' UTR of the active gene as well as the coding region except for primer D and H. Data was normalized to H3. Error bars= STD. Primers A-D are located in the 5' UTR region and primers E-H in the coding region. Error bars = STD. * = p-value < 0.05, ** = p-value 0.01, *** = p-value 0.001. ● = primer overlap.

3C.4.3 H3K9/36 trimethylation enrich in 3D7_0617600^{OFF}

Figure 3.51: H3K9 trimethylation levels across 3D7_0617600^{ON} and 3D7_0617600^{OFF}.

Three clones, 6.2A, 3.2C and AF5, were used for this analysis. In clone 6.2A and 3.2C, 3D7_0617600 is 'on' and in clone AF5, 3D7_0617600 is 'off'. Data was normalized to H3. Error bars= STD. Primer A-D are located in the 5' UTR region and primers E-H in the coding region. * = p value < 0.05, ** = p value 0.01, *** = value < 0.01. ● = primer overlap.

The analysis for H3K9 trimethylation showed significant differences between the 3D7_0617600^{ON} and 3D7_0617600^{OFF} with more enrichment in 3D7_0617600^{OFF} at all positions except at position A, B and F in clone 6.2A (Figure 3.51). In the coding region, significant enrichment was observed on three out four PCR fragments (E, G, and H) in 3D7_0617600^{OFF} as compared to the 3D7_0617600^{ON}. This result mirrors the same result seen for PF3D7_1254100 (Chapter 3B). In the 5' UTR region only primers C and D show clear enrichment in 3D7_0617600^{OFF} in the 6.2A clone. Actually, at position A, we observed more but not significant, H3K9 trimethylation levels in 3D7_0617600^{ON} as compared to 3D7_0617600^{OFF} for clone 6.2A. The reason for this is not known. But in clone 3.2C the same primer is significantly enriched in 3D7_0617600^{OFF}. Clone 3.2C shows a similar enrichment profile as 6.2A with the modification significantly enriched in 3D7_0617600^{OFF}.

H3K36 trimethylation modification showed a similar pattern as H3K9 trimethylation for both clones. This modification correlates more with the silent gene, as there are more H3K36 trimethylation levels in 3D7_0617600^{OFF} as compared to 3D7_0617600^{ON} from primers C to F (Figure 3.52) which corresponds to 3' end of 5' UTR region and the 5' end of the coding region both clones. It is possible that enrichment at this region is sufficient to repress the transcription of the gene. Overall, 3D7_0617600 uses both H3K9 trimethylation and H3K36 trimethylation as modifications that associate with the silent form of the gene as observed for 3D7_1254100. Interestingly H3K9 trimethylation follows the same pattern as H3K36 trimethylation including positions of enriched primers suggesting that these two modifications cluster (primarily primers C to F). The pattern observed is similar to that observed for PF3D7_1254100 (Figures 3.24 and 3.25) where the same primers were enriched for H3K9 and 36 trimethylation and enrichment was more in the coding region.

Figure 3.52: H3K36 trimethylation levels across 3D7_0617600^{ON} and 3D7_0617600^{OFF}.

Clones, 6.2A and AF5, were used for this analysis. In clone 6.2A 3D7_0617600 is 'on' and in clone AF5, 3D7_0617600 is 'off'. Apart from primer A and B, H3K36 trimethylation shows specific enrichment across the 5' UTR of the active gene as well as the coding region. Error bars= STD. Primers A-D are located in the 5' UTR region and primers E-H in the coding region. Data was normalized to H3. Error bars = STD. * = p-value < 0.05, ** = p-value 0.01, *** = p-value < 0.001. ● = primer overlap.

3C.4.4: H4K20 trimethylation shows differences between clones 6.2A and 3.2C

Figure 3.53: H4K20 trimethylation levels across 3D7_0617600^{ON} and 3D7_0617600^{OFF}.

Clones, 6.2A and AF5, were used for this analysis. In clone 6.2A 3D7_0617600 is 'on' and in clone AF5, 3D7_0617600 is 'off'. Apart from primer A and B, H3K36 trimethylation shows specific enrichment across the 5' UTR of the active gene as well as the coding region. Error bars= STD. Primers A-D are located in the 5' UTR region and primers E-H in the coding region. Data was normalized to H3. Error bars = STD. * = p-value < 0.05. ● = primer overlap.

Our analysis for H4K20 trimethylation showed no differences in the 6.2A clone for 3D7_0617600^{ON} and 3D7_0617600^{OFF} (Figure 3.53) which is similar to the *stevor* studied thus far. But in clone 3.2C, a statistical difference is observed between 3D7_0617600^{ON} and 3D7_0617600^{OFF} with the modification enriched in the 5' UTR of 3D7_0617600^{ON} (primers B to D) and primer E in the coding region. This was a rather unexpected result. In personal communication with Dr. Gupta, they also observed H4K20 trimethylation correlating with *var* genes (unpublished data). This suggests that H4K20 trimethylation can positively correlate with variant genes. However, in the published manuscript of the same study, they observed a negative correlation with H4K20 trimethylation in euchromatic genes (Gupta et al. 2012) perhaps suggesting different modes of action of the modification in different genomic environments.

3C.4.5 Conclusion for 3D7_0617600

Our analysis for 3D7_0617600 involved looking at two clones 6.2A and 3.2C where this gene was active. The control for both clones was AF5 where the gene was silent. Data from the AF5 clone where this gene was transcriptionally silent was used for both 6.2A and 3.2C. For both clones, we observed a very similar pattern for enrichment for the modifications showing the robustness of the assay. Minor differences between the two clones were observed for the acetylation modifications, in terms of regions where the enrichment was observed. For the 3.2C clone, H3K9/H4ac enriched across the coding and 5' UTR regions, whereas in the 6.2A clone enrichment of the same was more pronounced in the 5' UTR regions. However, all the modifications correlated with 3D7_0617600^{ON}. This is the first study to correlate H4 acetylation reactions to an active *stevor* gene. Acetylation reactions are thought to be highly dynamic; therefore the minor differences could just be a reflection of this in the difference clones. I also

tried to harvest the cells at the same time point but the trophozoite stage is long, spanning 14 hours. So the minor differences in acetylation reactions could also be influenced by the exact time I collected the samples from the clones.

The rest of the modifications (H3K4me2/3, H3K9/36me3) showed a similar profile between the two clones. The profiles were almost identical for some these methylation reactions, with the same regions showing significant enrichment. H3K4me2/3 methylation correlated with 3D7_0617600^{ON} as opposed to 3D7_0617600^{OFF} and the enrichment spanned both 5' UTR and coding regions. This suggests that H3K4me2/3 methylations are involved in regulating this particular *stevor* gene. H3K9/36me3 correlated with 3D7_0617600^{OFF}, as was observed for the 3D7_1254100 transgene (Chapter 3B). This suggests these modifications are potentially involved in silencing of the *stevor*. Due to the resolution of the assay, I cannot say if the two modifications, though they show a clustering pattern, are located on the same histone in a nucleosome or they would be on different histones as the sonication method used for this assay gives a range of between 2-9 nucleosomes. This is the same for all assayed genes. Genome-wide studies of these two modifications do show that they enrich in the subtelomeric regions of the *P. falciparum* chromosomes so it is not surprising that they could cluster (Salcedo-Amaya et al. 2009; Jiang et al. 2013; Lopez-Rubio et al. 2009).

The only anomaly we encountered was the H4K20 trimethylation. All the other modifications showed good correlation between the two clones, showing the robustness of the technique, therefore this result cannot just be a technical issue between the pull downs. I did not expect to observe this result as this suggests that in the 3.2C clone H4K20 trimethylation does correlate with an active *stevor*. However, more work needs to be done, likely in a third clone, to validate the findings for H4K20 trimethylation.

Another interesting observation I made was that, in clone 3.2C, 3D7_0617600, was transcribed at lower levels than in 6.2A and yet the enrichment profile was similar for both clones. Previous studies have suggested that highly transcribed genes also have higher levels of activation marks (Cabral et al. 2012) but this was not observed in this study as the transcriptional levels did not correlate with modification levels.

This result had me thinking if the clonal variation is responsible for these small and sometimes major differences between clones. In work done by Rovira-Graells et al, they extensively analyzed the transcriptional profiles of isogenic clones grown in homogeneous conditions and observed very stark heterogeneity in terms of expression of genes involved in host-parasite interactions in different clones (Rovira-Graells et al. 2012). They concluded that this was an epigenetic phenomenon and proposed that this was the parasites way of preparing itself to quickly adapt to new environments – a bet-hedging strategy. Observing these minor differences between the H3K9 acetylation modification and now the H4K20 trimethylation made us wonder whether this was a biological event. More studies will need to be done to establish this. The results of 3D7_0617600 are summarized in Figure 3.54.

Figure 3.54: A summary of modifications associated with regulation of 3D7_0617600 *stevor*.

Shown are the histone modifications correlating with activation and repression of this gene and the regions where they are enriched as calculated from the ATG site. The gene was divided into 200bp regions as this was the average PCR product from the PCR primers used. Note enrichment heights of the modifications are not drawn to scale. H4K20 trimethylation is also included in this figure as a possible regulator of *stevor*. The question mark signifies the ambiguity about this modification.

3C.5 EPIGENETIC LANDSCAPE FOR PF3D7_1372500

Another *stevor* which was studied was PF3D7_1372500. This *stevor* gene was transcriptionally active in clones 6.2A and 3.2C and was transcriptionally silent in clone AF5. As before, we used these three clones for the analyses. Due to a 99% homology with PF3D7_0425500 in the 5' UTR region, it was impossible to design specific primers in this region for PF3D7_1372500. Hence focus was given to the coding region. A 750bp region of this primer was targeted (Figure 3.55 A) and specificity was shown by getting a single band on ethidium bromide gel (Figure 3.55B).

Figure 3.55: Schematic representation and specificity of primer positions used for ChIP analysis of 3D7_1372500.

A: The endogenous size of the PF3D7_1372500 gene is around 900bp. Primers were designed to cover at least 750bp of the gene. For this analysis, primers overlapped each other as denoted by blue circles after each primer region. **B:** Sonicated material from the ChIP samples was used to perform PCR with the primers to ensure specificity. As can be seen, only one band is observed on the ethidium bromide gel. Lane symbols refer to primer position as shown in Figure 3.55A.

3C.5.1 H3K9/H4ac enrich in the coding regions of 3D7_1372500^{ON}

Figure 3.56: H3K9 acetylation levels across 3D7_1372500^{ON} and 3D7_1372500^{OFF}.

Clones, 6.2A, 3.2C and AF5, were used for this analysis. In clones 6.2A and 3.2C, 3D7_0617600 is 'on' and in clone AF5, 3D7_0617600 is 'off'. Primer A-C are located in the coding region. Data was normalized to H3 and is 3 biological replicates. Error bars = SEM. * = p-value < 0.05, ** = p-value 0.01, *** = p-value < 0.001.

Like the other *stevor* genes which were studied in this thesis, H3K9 acetylation significantly enriches for 3D7_1372500^{ON} and this pattern is the same for both clones (Figure 3.56). The interesting thing we observe here is that H3K9 acetylation also extends into the coding region as has been observed for PF3D7_1040200 and PF3D7_0617600.

H4K8 acetylation correlated with 3D7_1372500^{ON} and was significantly enriched in clone 3.2C at all positions (Figure 3.57). In clone 6.2A, there was an enrichment in 3D7_1372500^{ON} but this was not significant between 3D7_1372500^{ON} and 3D7_1372500^{OFF}.

H4K12 acetylation, showed significant enrichment in both clones for 3D7_1372500^{ON} (Figure 3.58) again highlighting that this modification does correlate with activation of *stevor*.

Figure 3.57: H4K8 acetylation levels across 3D7_1372500^{ON} and 3D7_1372500^{OFF}.

Clones, 6.2A, 3.2C and AF5, were used for this analysis. In clones 6.2A and 3.2C, 3D7_0617600 is 'on' and in clone AF5, 3D7_0617600 is 'off'. Primer A-C are located in the coding region. Data was normalized to H3 and is 3 biological replicates. Error bars = SEM. * = p value < 0.05, ** = p value 0.01.

Figure 3.58: H4K12 acetylation levels across 3D7_1372500^{ON} and 3D7_1372500^{OFF}.

Clones, 6.2A, 3.2C and AF5, were used for this analysis. In clones 6.2A and 3.2C, 3D7_0617600 is 'on' and in clone AF5, 3D7_0617600 is 'off'. Primer A-C are located in the coding region. Data was normalized to H3 and is 3 biological replicates. Error bars = SEM. * = p value < 0.05, ** = p value 0.01, *** = p value 0.001.

3C.5.2 H3K4 methylation enriches in the coding regions of 3D7_1372500^{ON}

Figure 3.59: H3K4 dimethylation levels across 3D7_1372500^{ON} and 3D7_1372500^{OFF}.

Clones, 6.2A, 3.2C and AF5, were used for this analysis. In clones 6.2A and 3.2C, 3D7_0617600 is 'on' and in clone AF5, 3D7_0617600 is 'off'. Primers A-C are located in the coding region. Data was normalized to H3 and is 3 biological replicates. Error bars = SEM. * = p value < 0.05, ** = p value < 0.01, *** = p value < 0.001.

Our analysis of H3K4me2, showed this modification to be significantly enriched for 3D7_1372500^{ON} for 6.2A and 3.2C (Figure 3.59), suggesting that it potentially plays roles in *stevor* regulation.

Figure 3.60: H3K4 trimethylation levels across 3D7_1372500^{ON} and 3D7_1372500^{OFF}.

Clones, 6.2A, 3.2C and AF5, were used for this analysis. In clones 6.2A and 3.2C, 3D7_0617600 is 'on' and in clone AF5, 3D7_0617600 is 'off'. Primers A-C are in the coding region. Data was normalized to H3 and is 3 biological replicates. Error bars = SEM. * = p value < 0.05, ** = p value < 0.01.

H3K4 trimethylation was enriched for 3D7_1372500^{ON} for both clones (Figure 3.60). In clone 6.2A at primer B, there was no specific enrichment observed, however.

3C.5.3 H3K9/36 trimethylation enriches in the coding regions of 3D7_1372500^{OFF}

Figure 3.61: H3K9 trimethylation levels across 3D7_1372500^{ON} and 3D7_1372500^{OFF}.

Clones, 6.2A, 3.2C and AF5, were used for this analysis. In clones 6.2A and 3.2C, 3D7_0617600 is 'on' and in clone AF5, 3D7_0617600 is 'off'. Primers A-C are in the coding region. Data was normalized to H3 and is 3 biological replicates. Error bars = SEM. * = p value < 0.05, ** = p value < 0.01.

H3K9 trimethylation is seen to significantly enrich for 3D7_1372500^{OFF} for both clonal lines (Figure 3.61). As shown for other *stevor*, this modification associates silent *stevor*, strongly suggesting it as a repressor of *stevor* genes. It is interesting to note that, in the active form, 3D7_1372500 has lower methylation levels in 3.2C (not as low IgG not shown) as compared to 6.2A. This could point to potential clonal differences between the two clones in terms of absolute levels of modifications. This argument applies to PF3D7_0617600 as well.

H3K36 trimethylation like H3K9 trimethylation also significantly enriches for 3D7_1372500^{OFF} (Figure 3.62). This enrichment is observed for both 6.2A and 3.2C. Like for the H3K9 trimethylation, the 3.2C clone has lower H3K36 trimethylation levels (not as low as IgG, not shown), again potentially suggesting clonal differences in the amount of modifications present. Or this could likely be a timing issue as when the samples were harvested as mentioned before.

H4K20 trimethylation did not show any significant differences between the 3D7_1372500^{ON} and 3D7_1372500^{OFF} (Figure 3.63).

Figure 3.62: H3K36 trimethylation levels across 3D7_1372500^{ON} and 3D7_1372500^{OFF}.

Clones, 6.2A, 3.2C and AF5, were used for this analysis. In clones 6.2A and 3.2C, 3D7_0617600 is 'on' and in clone AF5, 3D7_0617600 is 'off'. Primers A-C are in the coding region. Data was normalized to H3 and is 3 biological replicates. Error bars = SEM. * = p value < 0.05, ** = p value 0.01, *** = p value < 0.001.

3C.5.4 H4K20 trimethylation enrichment same for 3D7_1372500^{ON} and 3D7_1372500^{OFF}

Figure 3.63: H3K20 trimethylation levels across 3D7_1372500^{ON} and 3D7_1372500^{OFF}.

Clones, 6.2A, 3.2C, and AF5, were used for this analysis. In clones 6.2A and 3.2C, 3D7_0617600 is 'on' and in clone AF5, 3D7_0617600 is 'off'. Primers A-C are located in the coding region. Data was normalized to H3 and is 3 biological replicates. Error bars = SEM.

3C.5.5 Conclusion for 3D7_1372500

A summary of the modifications studied is shown in Figure 3.64. The pattern we observed for 3D7_1372500 is like the pattern observed for PF3D7_0617600. We observed H3K9/H4ac significantly enriched when the gene was transcriptionally active. The additional point added by this gene is that activation marks are uniform distributed in the coding as well as the 5' UTR regions. This was alluded to by Cabral and colleagues in their work on rifins (Cabral et al. 2012). Similar to H3K9 acetylation, H3K4me2/3 methylation also correlated with the active gene in both clonal lines. yAnd the modifications associated with repression, H3K9/36me3, were also observed to correlate with the silent form of the gene. For this gene, though, the H4K20me3 mark, did not correlate with any activation or silencing of the gene just as observed for the transgene and PF3D7_1040200 gene. In plasmodb, PF3D7_1372500 is annotated as a pseudogene. The results observed from the study of this gene show that for *stevors*, the epigenetic footprint is the same for pseudogenes and *stevor* coding genes.

Figure 3.64: Modifications associating with activation and transcription of 3D7_1372500.

Shown are the histone modifications correlating with activation and repression of this gene and the regions where they are enriched as calculated from the ATG site. Only the coding region was analyzed for this gene. The gene was divided into 200bp regions as this was the average PCR product from the PCR primers used. Note enrichment heights of the modifications are not drawn to scale.

3C.6 H3K9/36 trimethylation associate with silent *stevor* genes

In our clonal lines, we had some *stevors* that were silent in all clones. In order to establish H3K9/36 trimethylation as silencing marks we analyzed 4 *stevor* that were silent in our clonal lines for these two marks (Figure 3.65). We also assayed H3K9 acetylation as this is mark most studied for variant genes (*var* and *rifin*) and is linked to active transcription. We compared our findings to IgG and plotted all antibody results for all *stevor* (Figure 3.65). For this analysis, we used clone 6.2A and the assay were done in duplicates. The 5' UTR regions are highlighted by a red box.

Analysis of H3K9ac modification on silent *stevor* shows that most *stevor* are devoid of this mark when silent. In our analysis three of the four *stevor* we looked at had H3K9ac levels comparable to IgG levels. These were PF3D7_0201300, PF3D7_0900900 and PF3D7_1254600 (Figure 3.65). This is not surprising as this modification is heavily linked to virulent gene expression in *P. falciparum* (Lopez-Rubio et al 2007; Cabral et al. 2012; Freitas-Junior et al. 2005). In other organisms, as well H3K9 acetylation correlates well with active gene transcription. This suggests that this modification is potentially involved activation of *stevor*. However, one *stevor* PF3D7_0700400 (Figure 3.65) did show H3K9ac levels. In its silent state, is still has H3K9 acetylation levels above the IgG background.

I tried looking at the gene-environment of these 4 *stevors* for an explanation of the observed differences in H3K9 acetylation enrichment (Table 3.2). The genes in black are the ones with no H3K9 acetylation and the one in red is the one with H3K9 acetylation. All the *stevor* are in the subtelomeric regions and they have genes that are enriched in H3K9me3 surrounding them per previous studies (Salcedo-Amaya et al 2009; Lopez-Rubio et al. 2009) hence local chromatin environments are not the cause of differences.

Figure 3.65: Silent *stevor* assays for H3K9/36 trimethylation and H3K9 acetylation

Enrichment profile for K9Aac, K93me3 and K36me3 across the genes for PF3D7_0201300, PF3D7_1254100, PF3D7_0700400 and PF3D7_0900900. The boxed region is the 5' UTR region. The assay was done in duplicates and H3 was used for normalization. IgG= Immunoglobulin control, H3K9ac = H3K9 acetylation, H3K9me3 = H3K9 trimethylation, H3K36me3= H3K36 trimethylation.

Plasmodb Id	Location	Distance from left telomere	Transcribed towards	Genes before	Genes after
PF3D7_0201300	tel	66529-67545 (+)	centromere	exported ←, rifin →	hyp10 →, hyp9 ←
PF3D7_0900900	tel	55074-56081 (-)	telomere	rifin ←, var exon 2 pseudo ←	rifin ←, rifin ←
PF3D7_1254600	tel	2221835-2222848 (+)	telomere	rifin →, rifin →	rifin ←, rifin ←
PF3D7_0700400	tel	36922-37927 (-)	telomere	rifin →, rifin ←	rifin ←, exported pseudogene ←

Table 3.2: Chromosomal positions of the five silent *stevor*.

The table shows the chromosomal environment of the four silent *stevor* that were analyzed. Plotted are the chromosomal positions of the *stevor*, their genomic distance, and the genes before and after the *stevor*. Tel = telomere

However, in literature, it is possible to have an acetylation mark on a silent gene/locus. Permissive chromatin regions can be enriched not only with H3ac but also H3K4 methylation. Presumably, the histone acetylation reaction is there to prime the locus for activation (Karmodiya et al. 2012). And this is possibly what we observe here that this *stevor* is permissible for activation. Another group also observed occupancies of H4 acetylation on a subset of least expressed genes, which were also enriched with H3K9me3, suggesting a novel role of these modifications in inactive genes (Karmodiya et al. 2015). It is possible that there is a group of *stevor* that are always permissive or poised for transcription and that other factors like how open the chromatin is, and availability of transcription factors influence which *stevor* are eventually transcribed. This could potentially be possible because during the trophozoite stage when *stevor* are expressed, the nucleosomal landscape of the genome changes drastically and there is a reduction in nucleosomal content as parasite prepares for this transcriptionally active stage (Ponts et al. 2010; Bunnik et al. 2014). This opening of the chromatin might potentially lead to a poised state of activation for the *stevor* genes.

Analysis of the silent *stevor* does show that when *stevor* are silent they do enrich for the H3K9 trimethylation as well as H3K36 trimethylation. H3K9 trimethylation shows a rather interesting profile for all the *stevor*. All the *stevor* PF3D7_0201300, PF3D7_1254600, PF3D7_0900900, PF3D7_0700400 do show a marked enrichment of this modification in the coding regions. This is interesting, as in a recent paper by Karmodiya et al. using ChIP-seq, do propose that H3K9 trimethylation modification is enriched in the coding regions of virulent genes and peaks at the 3' end of the coding region (Karmodiya et al. 2015). This is a similar pattern we have observed for all the other *stevor* studied in this thesis. These results do put the histones in the coding region for *stevor* as being important for gene regulation as was proposed by Gupta et al. (2012).

The pattern for H3K36 trimethylation is not so clear cut. Apart from PF3D7_0201300 where the modification enriches more in the coding region, the rest of the *stevor* do not show preferential enrichment either in the 5'UTR or coding regions but rather the enrichment is similar in both regions. In all cases the H3K36me3 levels are above the IgG, signifying the presence of the modification.

CHAPTER 3D

STEVROR 5' UTR BIOINFORMATIC ANALYSIS

3. D.1. Promoter analysis and overall conservation of *stevor* promoters

In terms of enrichment of modifications, PF3D7_1040200^{ON} showed the most difference from the other *stevor*. The only similarity the gene broadly shared with the other *stevor* was the enrichment of H3 and H4 acetylation in the 5' UTR. Unlike the other *stevor* however, PF3D7_1040200 did not show specific differences in enrichment patterns for H3K4 dimethylation and H3K36 trimethylation. I wanted to see if the differences between these observations were because of promoter differences because all the other *stevor* I studied had silencing marks enriched in the coding region (Chapter 3B and 3C).

For this, the upstream regions of the 40 *stevors* were aligned using MUSCLE in MEGA6 software and then analyzed to see if they would group into different groups as has been observed for *vars* and *rifin* genes. For these gene families, the grouping has a biological/functional relevance. RIFINs group into A and B based on the presence or absence of 25aa. Group A RIFINs are exported to the surface to the iRBC and involved in rosetting and pathology (Goel et al. 2015) and group B RIFINs accumulate in the Maurer's cleft (Petter et al. 2007; Joannin et al. 2008). *Vars*/*PfEMP1* also group into 3 major groups based on their 5' UTR regions and orientation of transcription as well as their genomic position (Lavstern et al. 2003). Group A *var* (Jensen et al. 2004; Rottmann et al. 2006) and group B *var* (Kaestli et al. 2006; Rottmann et al. 2006) have been correlated with severe malaria in children with group A *var* rarely expressed in vitro. Group C *var* are located centrally on the chromosomes and are correlated with asymptomatic malaria (Rottmann et al. 2006). For the *stevor* analysis, a 750bp region upstream

of the ATG was used, as this region gave us good bootstrap values when we constructed a phylogeny tree. An initial analysis was to align the promoter sequences of all 40 *stevor* (including pseudogenes) and to calculate the percentage homology of the *stevor* (Figure 3.66). *Stevors* have an overall conserved homology of 73%. The majority of them have high homology between themselves but there are *stevors* (in green) that have very low sequence homology to other *stevor*.

3. D.2: *stevor* 5' UTR group into two separate groups

The sequence similarity between these *stevor* ranges from 60-86%. There are *stevor* which fall below this homology threshold highlighted in yellow and green (Figure 3.66). The *stevor* in green fall below the 30% homology and include PF3D7_0102100, PF3D7_0832600, PF3D7_0700700 and PF3D7_1100700. This analysis already suggests that *stevor* 5' UTR sequences could group based on the sequence homology observed here. With these same sequences, I then used MEGA, phylogenic software, and tried to do some rudimentary *stevor* 5' UTR analysis initially using the Neighbor-joining (NJ) analyses as has been done for the *var* and *rifin* genes. The best result of this analysis with acceptable bootstrap values is shown in Figure 3.67. The phylogeny tree shows that *stevor* can be grouped into two main groups which have been termed *stevorupsA* and *stevorupsB*. This tree was producible using a different but similar phylogeny method; the Minimum-Evolution (ME) analysis method (Figure 3.68). Using the ME method, the separation into the two *stevor* groups as seen in Figure 3.67 is not clear cut, though for both analyses the *stevor* in each group are essentially the same. NJ and ME method are similar. They both rely on an initial clustering of the sequences and are both distance based methods.

Figure 3.66: Multiple sequence alignment of *stevor* 5' UTR region.

Multiple sequence alignment was done by T-coffee using the DNA homologous gene function. This function compares all the *stevor* sequences by aligning them and then looking for overall conservation between all the submitted inputs.

NJ analyses are a good heuristic method for estimating ME. NJ looks for similarities in the sequences provided. So, in essence, it looks at how similar the sequences are and takes the calculated distances between them to produce a tree. In this sense the longer the tree branches are, the more 'different' the sequences. ME uses a method with only a minor difference. So in ME, the tree that minimizes the lengths of the tree, which is the sum of the lengths of the branches, is regarded as the best estimate of the phylogeny (Tamura et al. 2013).

Figure 3.67: Neighbor-joining analysis of *stevor* 750bp upstream regions.

This evolutionary tree was produced using the Neighbor-joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches (Felsenstein 1985). All 40 *stevor* 5' UTR sequences were first aligned using MUSCLE and then subjected to an evolutionary analysis with the elimination of positions with gaps and/or missing data. *Stevors* group into 2 main clusters A and B. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013). *Red asterisks= pseudogenes.

Figure 3.68: Minimum evolution analysis of *stevor* 750bp upstream regions.

This evolutionary tree was produced using the Minimum-Evolution method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches (Felsenstein 1985). All 40 *stevor* 5' UTR sequences were first aligned using MUSCLE and then subjected to an evolutionary analysis with the elimination of positions with gaps and/or missing data. *Stevors* group into 2 main clusters A and B. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013). *Red asterisks= pseudogenes.

Further analyses into this separation showed that 6/9 of the annotated pseudogenes genes cluster in *stevorupsB* (Figure 3.67). These have been highlighted on the graphs by red asterisks and include PF3D7_0102100, PF3D7_1372500, PF3D7_0401500, PF3D7_1000800, PF3D7_0700700, and PF3D7_1100700. The rest of the pseudogenes are in *stevorupsA* (PF3D7_0200900, PF3D7_0114600, and PF3D7_0500600). As well as this, some of the *stevor* with very low or average sequence homology but not annotated as pseudogenes also group into *stevorupsB*. These include PF3D7_0832600 (29%), PF3D7_0402600 (65%), PF3D7_0425500 (67%) including the already mentioned pseudogenes. Two of the three pseudogenes which did not group in *stevorupsB* have good sequence homology with the rest of the *stevors* (PF3D7_0114600 (79%) and PF3D7_0200900 (84%).

One *stevor* PF3D7_0500600 did show a peculiar grouping. This is a pseudogene with low homology to the rest of *stevors* (45%) yet still groups with *stevorupsA* using the NJ analysis. However, in Figure 3.68 this *stevor* does actually group in the *stevorupsB* cluster. With ME analyses coding *stevor* with good homology to other *stevor* (PF3D7_1372800 77% and PF3D7_0832900 85%) are actually grouped in *stevorupsA* suggesting this is a better phylogeny analysis as opposed to the NJ method where these two *stevor* group in *stevorupsB* (Figure 3.67). The ME was thought of as the better of the two methods and all further explanations will use data in Figure 3.68.

Using ME suggests that the *stevors* in *stevorupsB* are either pseudogenes which have low sequence homology with other *stevor* or coding *stevor* with low or average sequence homology to other *stevor*. There are anomalies in this grouping where we have coding *stevor* with average to good homology to other *stevors* grouping in *stevorupsB* (PF3D7_0402600%, PF3D7_0425500

67%, PF3D7_0832000 65%). This suggests 5' UTR sequence is not solely a determinant in the grouping system and we propose to group these anomalies as *stevorupsB/A*.

I then looked at the genomic locations of the *stevorupsB* and *stevorupsB/A* to try to see if genomic positions could also be involved in the grouping. For this, I used plasmodb and the information compiled in the database and tabulated all the information (Table 3 A and B - these two tables are a continuation of the same genes, just broken up into two parts). The tables do show the chromosomal position of each of the *stevorupsB* or *B/A*, the orientation to which they are transcribed, the genes before and after each *stevor*, whether they are pseudogenes or coding genes, presence or absence of SP, size of the amino sequence and distance of the *stevor* as calculated from the left telomere. All *stevor* are located in the telomeric regions and no *stevor* are located in the chromosomal interior regions as seen for *var* and *rifin* genes. Nothing significant was observed for these *stevors* apart from the fact that 7 of the 12 *stevor* with a transcription orientation towards the centromere are *stevorupsB* (Table 3B). So the 5' UTR sequence and potentially transcription orientation are the major determinants of *stevor* grouping.

A

Plasmodb Id	Notes	SP	Size	Distance from left telomere
PF3D7_0425500←	Coding	present	304aa	1146672 - 1147697 (-)
PF3D7_1372500→	pseudogene	present	301aa	2,854,544 to 2,855,568
PF3D7_0832000←	Coding	present	303aa	1383349 - 1384377 (-)
PF3D7_1000800→	pseudogene	present	292aa	57968 - 58980 (+)
PF3D7_0402600←	Coding	present	301aa	159279 - 160314 (-)
PF3D7_0102100→	pseudogene	absent	281aa	95050 - 95899 (+)
PF3D7_0401500←	pseudogene	present	310aa	89153 - 90184 (+)
PF3D7_0832600←	Coding	present	317aa	1405835 - 1406999 (-)
PF3D7_1100700→	Pseudogene	present	287aa	56792 - 57840 (+)
PF3D7_0700700→	pseudogene	present	287aa	45559 - 46607 (+)
PF3D7_0500600←	Pseudogene	present	307aa	39443 - 40488 (-)
PF3D7_0221400→	Coding	present	307aa	860466 - 861504 (+)

B

Plasmodb Id	Chromo Position	Trans towards	Gene before	Genes after
PF3D7_0425500←	right chromo 4	centromere	phista→	rifin→
PF3D7_1372500→	right chromo 13	telomere	rifin→	AcylCoA→
PF3D7_0832000←	right chromo 8	centromere	phist (psi)→	rifin (psi)←
PF3D7_1000800→	left chromo 10	centromere	phista (psi)←	truncated var→
PF3D7_0402600←	left chromo 4	Telomere	rifin→	rifin←
PF3D7_0102100→	left chromo 1	centromere	phista (psi)←	RESA→
PF3D7_0401500←	left chromo 4	Telomere	rifin→	rifin←
PF3D7_0832600←	right chromo 8	centromere	rifin←	phista like psi→
PF3D7_1100700→	left chromo 11	centromere	phista←	PfMc-2TM→
PF3D7_0700700→	left chromo 7	centromere	exported (psi)←	Pfmc-2tm→
PF3D7_0500600←	left chromo 5	Telomere	rifin←	←rifin
PF3D7_0221400→	right chromo 2	Telomere	complement pseudogene←	Pfmc-2TM←

Table 3: Chromosomal data for *stevors* in *stevorupsB*.

A and B show the chromosomal information of the *stevor* genes grouping in *stevorupsB*. The two tables are of the same genes with B, a continuation of A. Information annotated about the *stevor* in plasmodb. Psi=pseudogenes, chromo= chromosome. ← → = transcription orientation either towards the telomeric ends or towards the centromere depending on the position of *stevor* whether it's in the left or right telomere regions.

3. D.3 3 major sequences are overrepresented in *stevor* 5' UTR regions

Using MEME software I tried to look for any significantly represented regions present in all the *stevor* with a p value of less than 0.5, ranging from 6bp to 50bp. I again used the 750bp region as before and three major regions are present in the *stevor* selected using this particular setting (Figure 3.69).

Figure 3.69: Top three significant sequences in the 5' UTR regions of *stevor*.

MEME program was used for the analysis. Parameters used were. Motif site distribution= any number of sites per sequence. Site strand handling=sites may be on either strand. Maximum number of motifs= 3. Motif E-value threshold=No limit. Minimum motif with = 6. Maximum motif with = 50. Minimum sites per motif= 2 and maxim sites per motif = 50.

(Actual *stevor* specific sequences are in Supplementary Figures 1-3)

Stevor were analyzed based on the occurrence of these motifs and the results are tabulated in Table 3.4. *Stevor* that group in *stevorupsA* or the anomalies that were grouped in *stevorupsA/B* or B/A have all the motifs in their 5' UTR sequences. However, *stevor* that grouped in *stevorupsB* show a lack of these motifs (highlighted in yellow Table 3.4) with the exception of a few *stevor* that had motif 1 (PF3D7_0402600, PF3D7_0425500, PF3D7_0832000, PF3D7_1000800). The *stevor* which lack these motifs are the same *stevor* that are grouped in

*stevorups*B with the ME phylogeny analysis (Figure 3.68). Based on these findings we propose that the table represents a grouping of *stevor* into two main groups with a hybrid group.

<i>Stevor</i> gene	<i>stevorups</i>	Motif 1	Motif 2	Motif 3
PF3D7_0101800	A	yes	yes	yes
PF3D7_0115400	A	yes	yes	yes
PF3D7_0200400	A	yes	yes	yes
PF3D7_0201300	A	yes	no	yes
PF3D7_0222800	A	yes	yes	yes
PF3D7_0300400	A	yes	yes	yes
PF3D7_0324600	A	yes	yes	yes
PF3D7_0400800	A	yes	yes	yes
PF3D7_0617600	A	yes	yes	yes
PF3D7_0631900	A	yes	yes	yes
PF3D7_0700400	A	yes	yes	yes
PF3D7_0732000	A	yes	yes	yes
PF3D7_0832400	A	yes	yes	yes
PF3D7_0900900	A	yes	yes	yes
PF3D7_0901600	A	yes	yes	yes
PF3D7_1040200	A	yes	yes	yes
PF3D7_1149900	A	yes	yes	yes
PF3D7_1254100	A	yes	no	yes
PF3D7_1254300	A	yes	yes	yes
PF3D7_1254600	A	yes	yes	yes
PF3D7_1300900	A	yes	yes	yes
PF3D7_1400700	A	yes	yes	yes
PF3D7_1479500	A	yes	yes	yes
PF3D7_1479900	A	yes	yes	yes
PF3D7_1372800	A	yes	yes	yes
PF3D7_0114600	A/B	yes	yes	yes
PF3D7_0200900	A/B	yes	yes	yes
PF3D7_0401500	A/B	yes	yes	yes
PF3D7_0832900	B/A	yes	yes	yes
PF3D7_1372500	B/A	yes	yes	no
PF3D7_0102100	B	no	no	no
PF3D7_0221400	B	no	no	no

PF3D7_0402600	B	yes	no	no
PF3D7_0425500	B	yes	no	no
PF3D7_0500600	B	no	no	no
PF3D7_0832000	B	yes	no	no
PF3D7_0832600	B	no	no	no
PF3D7_1000800	B	yes	yes	no
PF3D7_0700700	B	no	no	no
PF3D7_1100700	B	no	no	no

Table 3.4: *stevor* grouped based on presence or absence of sequences in the 5' UTR regions.

The hybrid group has *stevor* that group in A but are pseudogenes with good sequence homology to other *stevor* (group A/B= PF3D7_0200900, PF3D7_0114600 and PF3D7_0401500) or coding *stevor*/pseudogenes grouped in B but with good sequence homology to the rest of the *stevor* and have all three motifs present (group B/A= PF3D7_0832900, PF3D7_1372500 (only has motif 1 and 2)). We propose that these motifs could have functional significance in the evolution of *stevor*, as all are lost in a group of *stevorupsB* genes.

This work was interesting as it shows that *stevor* 5' UTR sequences have functional information that can be used to group them. The grouping system is highly reproducible and is consistent with transcription orientation as well as sequence homology. The observation that the *stevorupsB* lack the three motifs seen in *stevorupsA* was an exciting result as it suggests that there must have been sequence losses in the expansion of this gene family. However, this is a rudimentary analysis and when I tried to blast these three motifs using the function in the MEME software I could not get any hits. I got many small hits of between 6-10 bp in length and this analysis is still ongoing. It would be interesting to see what the function of these motifs is in *stevor*, whether they are involved in some sort of regulation or there are involved in the timing of *stevor*

expression timing or other events. Other bioinformatics work needs to be carried out to extensively validate the result observed here, but I believe this analysis sets the stage for this. Of note is the fact that *stevorupsA* and *stevorupsB* are potentially regulated the same way epigenetically, as evidenced by the genes which we looked at in 3C. PF3D7_1040200 and PF3D7_0617600 group in the *stevorupsA* group whereas PF3D7_1372500 groups in *stevorupsB/A*. Analysis of these genes showed that they had similarities in the modifications that they associate with when active or silent. Acetylation as well H3K4 methylation reactions are enriched when the *stevors* are on. H3K9/36 trimethylation enriches when the genes are silent. Though the analysis is limited due to the lack of 5' UTR information for PF3D7_1372500, the results of PF3D7_0617600 showed a similar profile to that of PF3D7_1372500 in the coding regions.

CHAPTER 4

FINAL CONCLUSION AND DISCUSSION

4.1 Introduction

In this work, I set off with two aims. The first was to generate a transgenic parasite line with an integrated *stevor*. This was crucial as it's not possible to select for expression of certain/single *stevors* and also because of switching events between the *stevor*, having a transgenic parasite line would ensure the transcription of a known *stevor* at all times. In this approach, I would create a transgenic parasite line where a *stevor* promoter would drive expression of a drug selectable cassette. This line would have been used to study epigenetic regulation of *stevor* and other experiments. I failed to achieve this part of the aim as I failed to get an integrated *stevor* line. I could verify the presence of episomal transfectants but not of any integration. The reasons why this part of this work failed are less likely to do with an inaccessible loci or *stevor* essentiality as another group was able to get *stevor* integration (Witmer et al. 2013) and also previous studies do show that *stevor* are essential *in vivo* rather than *in vitro* (Lavazec et al. 2007; Blythe et al. 2008) so knocking them out is theoretically possible. However, integrations are known to be extremely rare events in *P. falciparum* occurring only by homologous recombination and taking up to months to occur. So the failure could be just misfortune on my part or the parasites needed longer time to come up. The greatest flaw, however, was the lack of a positive control. Without this, it is hard to ascertain what exactly happened, whether the failure was a technical issue on my transfection method. A positive control is highly encouraged if these studies are to be repeated. In other studies where they get integration they use a negative selection cassette with

thymidine kinase as opposed to cytoadhesin gene (Jiang et al 2013; Maier et al. 2008). Perhaps this is more potent than the cytoadhesin suicide gene used.

The second aim was to determine the epigenetic map of the *stevor* multigene family and study marks associated with ‘on’ and ‘off’ states of the genes. As mentioned in the introduction, histone modifications regulate gene transcription by regulating the chromatin environment (Lee et al. 1993; Wolffe & Hayes 1999; Li et al. 2007). Over the last few years, work on histone modifications and their roles in regulating chromatin have clearly shown the importance of these modifications in this biological process.

This part of the work was more successful. The analysis shows that broadly speaking, modifications which associate with active and silent *rifⁱⁿ* and *var* gene loci also associate with silent and active *stevor*. Unlike the *var* genes which undergo mutual exclusion, *stevors* do not exhibit this phenomenon. Rather what is observed is one *stevor* or a subset of *stevor* expressed simultaneously at a single parasite level. And in this work, by looking at multiple *stevor* in the same clones (PF3D7_0617600 and PF3D7_1372500), and *stevor* in different clones (PF3D7_1040200 and the transgenic line PF3D7_1254100), we show that the epigenetic footprint for the multiple *stevors* expressed in the same clone and multiple clones is similar. However, we also observed subtle differences between the three afore-mentioned *stevor* and PF3D7_1040200 in terms of histone modifications that associate with activation and the region of enrichment of the modifications that associate with repression. In this work we also made use of multiple clonal lines (6.2A and 3.2C) and direct comparison of two *stevor* in these two clonal lines showed the reproducibility of the data presented in this thesis and shows that the assay we used is robust. This gives confidence in the data presented here. I also performed some rudimentary bioinformatic analyses and show that *stevor* group into groups here named

stevorupsA and *stevorupsB*. Presence or absence of three 50bp sequences groups these *stevors* and most *stevor* annotated as pseudogenes or *stevor* with low sequence homology to other *stevor* fall into the *stevorupsB* group.

4.2 *Stevors* show specific selectivity in terms of regulatory repressive marks

It has been suggested that the default transcriptional state of the virulence genes is repression. This is crucial as the parasite needs to control the expression of these genes families so as not to exhaust the repertoire but gradually switch them when confronted with an active immune system. Hence the silencing modification marks are thought to dictate transcription. Analysis for *stevor* highlighted H3K9 trimethylation and H3K36 trimethylation as associating with silent *stevor* (Figure 3.70). This thesis shows that studied *stevor* when silent are enriched with H3K9 trimethylation in agreement with genome-wide studies from Salcedo-Amaya et al. (2009) and Lopez-Rubio et al. (2009). This finding was also shown when we looked at silent *stevor* (Figures 3.65 A to D) and show that these marks are enriched in silent *stevor* and that the H3K9 acetylation modification is absent from most silent *stevor* (Figure 3.65 A to C). H3K36 trimethylation has also been correlated to *stevor* silent *stevor* on a global scale, though there is conflicting data on the exact role of H3K36 trimethylation in *P. falciparum* (Jiang et al. 2013; Karmodiya et al. 2015). Unlike H3K9 trimethylation which decorates all forty *stevors*, H3K36 trimethylation apparently only decorates a few or subset of *stevors*. Indeed in our analysis, we observed three out of four *stevor* with on and off systems PF3D7_1254100, PF3D7_1372500 and PF3D7_0617600 (Figure 3.70) enriched in H3K36 trimethylation when silent.

In the other analysis of other four silent *stevor*, we also observed all four enriched for H3K36 trimethylation. When we looked closely at the enrichment status of H3K9/K36 trimethylation for PF3D7_1254100, PF3D7_0617600, and PF3D7_1372500, we noticed that they followed the same enrichment pattern more so for PF3D7_1254100 and PF3D7_1372500 (Figure 3.70).

For all genes, enrichment for both modifications was observed more in the coding region and about 400bp upstream of the ATG site for both modifications (5' UTR data not available for PF1372500 due to the problems designing primers in this region). Preferential enrichment of H3K36 trimethylation in the coding region was also observed for silent *stevor* (Figure 3.65: PF3D7_1254600 and PF3D7_0201300), but this is not a universal pattern for all *stevor* as some silent *stevor* have uniform levels of H3K36 trimethylation in the 5' UTR and coding regions (Figure 3.65-C and D - PF3D7_0900900 and PF3D7_0700400).

Based on the results of PF3D7_1254100, PF3D7_0617600, PF3D7_1372500, PF3D7_1254600, PF3D7_0700400, PF3D7_0201300 and the other silent *stevor*, nucleosomes located 3' end of the *stevor* 5' UTR and in the coding region enrich for modifications that associate with silent *stevor*. According to Bunnick et al. (2014) and Ponts et al. (2011), *P. falciparum* has the strongest positioned nucleosomes at the 5' end and 3' end of the ORF, though this fact has still to be proven as another study suggested that there were are no nucleosomal differences in the intergenic and coding regions of *P. falciparum* (Bartfai et al. 2010). Regardless of this conflict, Gupta et al. (2012) show how most of the modifications that associate with active genes preferentially enrich at the 5' end of coding regions. RNA data from Watannabe et al. (2004) and Karmodiya et al. (2015), maps most of the parasites genes, especially *stevor* genes TSS just at the ATG codon (Supplementary Table 1).

Figure 3.70: Position of H3K9/36 trimethylation enrichment for PF3D7_0617600, PF3D7_1040200, PF3D7_1254100 and PF3D7_1372500

This figure is made from all the data of all the *stevor* studied. With the exception of PF3D7_1040200, *stevor* generally show enrichment of H3K9 trimethylation in the coding region as well as about 500bp upstream of the ATG. *PF3D7_1040200 did not have any H3K36 trimethylation enrichment. The average silent *stevor* takes into account the four silent *stevor* looked at. These four *stevors* were silent in all clones analyzed and suggest they could be perpetually silent. These four had a profile of H3K9 trimethylation where it peaks in the mid gene to the 3' end.

Hence it is tempting to speculate that enrichment of modifications at the 5' end of ORF correlates with the presence of nucleosomes that supposedly provide the strongest resistance to RNA polymerase II. Enrichment of these marks in the 5' UTR shows similarities to *var* genes where these marks are also enriched in the 5' UTR regions.

The additional information we observed from *stevors* is that for some *stevor* H3K9 trimethylation enriches more in the coding regions and the enrichment levels are higher in this region compared to the 5' UTR regions (Figure 3.70- average silent *stevor*). The exception to this rule is PF3D7_1040200 where H3K9 trimethylation was clearly enriched in the 5' UTR (Figure 3.39). For this gene even though, Karmodiya et al. (2015) data suggest that the TSS is just at the ATG site (Supplementary Table 1), new TSS curation data currently being deposited in plasmodb from Adjalley et al. (personal communication) shows the TSS of PF3D7_1040200 to be upstream of the ATG codon. This could be one explanation for the position of the 5' UTR enrichment for H3K9 trimethylation. This could also suggest that silencing of this *stevor* is different from other *stevor*, which is an interesting observation.

4.3 H3K36 trimethylation modification

When I studied the H3K36 trimethylation mark, I also made use of data from Jiang et al. (2013). These researchers created a KO line of the PfSET responsible for methylating H3K36. In their work, they suggest that this modification is involved in silencing *var*, *rifins*, and *stevor* because in the KO line they observed expression of all *vars* and a subset of *rifins* and *stevor* when compared to the WT line. In the KO line, they observe 11 *stevor* that go from having absent (A) or unreliable transcripts in the WT to a detection (P) of the transcript in the KO line. Three of the *stevor* I studied had data from this study (Supplementary Table 2). In my analysis, 3D7_1254100 was significantly enriched with H3K36 trimethylation when it was silent for both the transgene

and endogenous gene. The KO data shows the presence of transcript (referred to as Marginal by the authors) in the KO line suggesting H3K36 trimethylation potentially correlates with the silent form of this gene as transcripts are detected in the KO line. 3D7_1040200 shows transcript in the KO line (A to P) suggesting H3K36 enriches for the silent gene, but in my studies, this modification had no significant difference when the gene was on and when it was silent. Lastly for 3D7_0617600, data from the KO shows the presence of transcripts in both the WT and KO line suggesting H3K36 trimethylation does not correlate with the silencing of this gene. I should mention that though transcripts were observed in both the KO line and WT line, in the KO line, the transcript levels are significantly higher (at least 10x) than in the WT. However, in my studies, this *stevor* showed H3K6 trimethylation when silent. This observed enrichment for PF3D7_1254100 and PF3D7_0617600 is reproducible using different clonal lines. Using this KO data set the *stevor* studied here, show differences from the KO line.

Even though in the Jiang et al. 2013 study, they could correlate H3K36 trimethylation with silent *vars*, *rifins*, and *stevor*, another study looking at the same modification report slightly different results (Karmodiya et al. 2015). In this later study they do observe H3K36 trimethylation at these virulent genes but at lower levels and do not correlate it with silencing of the virulence genes. They actually classify this modification as a modification which enriches for active/poised genes and correlate it with active transcription of the chromosomal core genes. By aligning their RNA-seq data with the KO line from Jiang et al., Karmodiya and colleagues observed a global increase in the genes that are enriched for H3K36 dimethylation and not H3K36 trimethylation. The authors emphasize more the role of H3K9 trimethylation in silencing virulence genes than the role of H3K36 trimethylation and they propose that the virulence genes are always in a state of poised expression.

Regardless of these two contradictory results, my analyses show that H3K36 trimethylation enriches for *stevor*. 3D7_0617600 and 3D7_1372500 both enrich for H3K36 trimethylation when silent in not one, but two clonal lines (6.2A and 3.2C). Also, the pattern of enrichment for this modification shows a similar profile to that of H3K9 trimethylation and this is known to associate with silent genes and recruits proteins like heterochromatin protein 1 (HP1), that compact chromatin leading to gene silencing (Pérez-Toledo et al. 2009). Regardless more work still needs to be done in the in *P. falciparum* to dissect the actual role of H3K36 trimethylation. If H3K36 trimethylation correlates with silencing as observed in this thesis, this is an interesting observation as it suggests that there is gene selectivity in terms of regulatory marks with some silent *stevors* associating with H3K9/36 trimethylation marks and other silent *stevors* associating with H3K9 trimethylation. Also, the region of enrichment for the marks is different for each gene. This could mean that *stevors* with the double repression marks would be less likely to be expressed than *stevors* with one repressive mark providing a mechanism for regulating *stevor* expression.

A different study also showed how this modification is enriched at infrequently transcribed long genes and how H3K36 trimethylation was needed for correct transcription of these genes. This potentially provides another role for H3K36 trimethylation at these variant genes, that of ensuring their proper transcription (Li et al. 2007) instead of just being associated with silencing. In this instance, knocking out the gene would lead to an expression of the gene families because the proper control mechanism would have been lost.

How H3K36 enriches for silent *stevor* is a mechanism unknown. As mentioned earlier, H3K36 deposition is tightly linked to RNA coupled transcription, and it is only deposited on recently transcribed genes, where it is thought to act as a suppressor of spurious transcription, and thus

H3K36 trimethylation has potential repressive properties (Carrozza et al. 2005; Lee & Shilatifard 2007). This hypothesis is partly supported by work again done in yeast which has shown that H3K36 trimethylation enhances the affinity of the repressive Imitation switch (IswIb) chromatin remodeling complex for nucleosomes, suppressing nucleosome mobility and turnover (Smolle et al. 2012; Maltby et al. 2012). Studies in yeast have also shown how Set2, the H3K36 methylase, is actually a regulatory module for the lysine deacetylase complex, Rpd3S, (Quan & Hartzog 2010; Carrozza et al. 2005; Tompa & Madhani 2007) and is thought to regulate its catalytic activity.

Though the IswIb chromatin remodeling and Rpd3S complexes have not been shown in *P. falciparum*, Jiang et al. and Ukaegbu et al. showed an association of the production of long non-coding RNAs (Jiang et al. 2013; Ukaegbu et al. 2014; Ukaegbu & Deitsch 2015) produced by the promoter activity of *var* gene introns (Epp et al. 2009; Calderwood et al. 2003; Su et al. 1995) with H3K36 trimethylation deposition on both active and silent *var* genes. It is known that in *P. falciparum* these *var* specific non-coding RNA (nc-RNA) are produced by RNA polIII, on both active and silent *var* genes, providing a mechanism for the possible recruitment of PfSET2 to *var* loci via interaction with RNA polIII, and the subsequent deposition of H3K36 trimethylation (Ukaegbu & Deitsch 2015). To further substantiate these findings, Jiang et al. showed how the presence of nc-RNAs correlates with the presence of H3K36 trimethylation (Jiang et al. 2013).

These studies explain how the H3K36 trimethylation mark is deposited on *var* genes. However, this does not explain how this modification is deposited on *stevor* genes. No studies have shown that *stevor* produce antisense RNAs or ncRNAs, and it is unlikely they do, as the *stevor* intron is on average less than 70bp compared to the 170bp to 1kb of *var* genes (Gardner et al. 2002). This then could potentially point to a different mechanism of recruitment of PfSETvs to silent *stevor*

genes. Studies do show the presence of long noncoding RNA's in the subtelomeric regions where the virulence genes are located (Broadbent et al. 2011; Sierra-Miranda et al. 2012). Whether these ncRNAs are involved in recruiting the PfSETvs methylase is unknown and could be an interesting area of study.

4.4 H3/H4 acetylation reactions associate with active *stevors*

If the default state of *stevor* is transcriptional repression, then the parasite needs to activate these genes as they perform important biological functions for the parasite. In this study H3 and H4 acetylation reactions associated with active *stevors*. Again, broadly speaking, these modifications are enriched in the 5' UTR and coding regions (Figures 3.71 and 3.72). With the new more robust data (use of clone 3.2C), both acetylation reactions associate with active *stevor*. We do however observe positional differences in H3K9 acetylation. In the transgene (Chapter 3B) and 3D7_0617600 (Chapter 3C-6.2A), H3K9 acetylation was more enriched in the 5' UTR and did not extend to the coding region but in all other *stevor*, the enrichment was from the 5' UTR to the coding region. In our analysis of 3D7_0617600, we observed an interesting observation. All the other modifications studied showed a perfect correlation in the two clones even at the sites of significant enrichment. This was different for the H3K9/H4K8ac reactions. In the 6.2A clone, H3K9/H4K8ac enriched only in the 5' UTR whereas in the 3.2C clone the modifications, enriched in the 5' UTR as well as coding regions of the active gene.

This could mean nothing biologically or could just reflect the variation in the clones or the timing when the cells were harvested as acetylation reactions are known to be highly dynamic. When H3K9 acetylation was analyzed in conjunction with silent *stevors* which were enriched for H3K9/36 trimethylation (Figure 3.65) this modification was devoid at these

Figure 3.71: Position of H3K9 acetylation enrichment for PF3D7_0617600, PF3D7_1040200, PF3D7_1254100 and PF3D7_1372500

H3K9 acetylation enrichment on all four stevors studied. The stevor do show slight positional enrichment with the transgene showing a more 5' UTR enrichment whereas all the other stevor showed an enrichment which extended from the 5' UTR to the coding regions. The 5' UTR information for PF3D7_1372500 is missing because of problems with PCR design in this region.

Figure 3.72: Position of H4 acetylation enrichment for PF3D7_0617600, PF3D7_1040200 and PF3D7_1372500

H3K4 acetylation enrichment on all four *stevors* studied. The *stevor* do show slight positional enrichment. PF3D7_0617600 has enrichment across the gene body as well 5' UTR regions, PF3D7_1040200 shows a localized enrichment. The transgene is not shown as we could not correlate H4 acetylation with it. The 5' UTR information for PF3D7_1372500 is missing because of problems with PCR design in this region.

loci, having levels comparable to the IgG control. This strongly suggests that H3K9 acetylation is important for *stevor* activation. This activation of these genes is important biologically for the parasite as *stevor* play roles in immune evasion as well as rosetting of the parasites. These all contribute to the pathology and survival of the parasite. Therefore the silencing marks (H3K9/36 trimethylation) need to be switched/exchanged for activation marks and we show H3/H4 acetylation as associating with active *stevor*. Another notable difference was seen with PF3D7_1372500 where H4K8 acetylation was only significantly enriched in 3.2C clone and not in the 6.2A clone. All the other modifications I looked at for this gene correlated well between 6.2A and 3.2C clonal lines suggesting that the difference cannot be just attributed to an experimental issue. PF3D7_1040200 also showed localized enrichment of H4 acetylation as opposed to the 5'UTR and coding regions as seen for 3D7_0617600 and 3D7_1372500 (Figure 3.72). Another anomaly we observed was the transgene where H4 acetylation did not correlate with activation or suppression. However, this line had a mixture of active and silent plasmids in both the AF5+ and AF5- lines as shown when we quantified the number of GFP-expressing cells in both lines. Hence what is observed is the average of modifications as the integrated gene was available in multiple copies. Therefore, it is entirely possible that the H4 acetylation was masked by having mixed cells in each line and that's why we could not correlate this modification with active transcription for the transgene as we did for the single copy genes.

Recently Karmodiya et al. reported on how H4 acetylation correlated with active virulence genes on a global scale (Karmodiya et al. 2015). Our study does not fully show this. Analyses of H4 acetylation marks (H4K8 and 12) did not give solid results for PF3D7_1254100 (therefore not shown Figure 3.72 but explained above) and H4 acetylation was only observed to associate with active expression for three of the *stevor* genes studied here, PF3D7_1040200 (localized

enrichment), PF3D7_0617600 and PF3D7_1372500 (Figure 3.72). It is noteworthy mentioning that, the antibody used in the Karmodiya study was against four H4 acetylation sites (H4K5/8/12/16) and in this thesis, only H4K8/12 were studied therefore a direct comparison of the two studies might not be suitable and could show conflicting results.

In terms of the downstream processes, though the different genes show different acetylation patterns, the downstream effects are likely to be the same as acetylation reactions neutralize the nucleosomal-DNA interactions allowing for the opening of chromatin.

The enrichment pattern of *stevor* is slightly different from that of the *var2csa* gene but shows similarities to other *var* genes. Acetylation reactions for *var2csa* *var* gene enrich mostly in the 5' UTR regions and level off in the coding regions when the gene is active (Lopez-Rubio et al. 2007). However, this gene is known to be structurally different from other *var* genes (Lavstern et al. 2003). In studies by other groups looking at different *var* genes, they find acetylation in the 5' UTR regions and the levels remain in the coding regions as well (Jiang et al. 2013). This suggests that the stricter mode of regulation seen for *var* genes which is necessary to maintain the mutually exclusive expression is maintained by additional features in the promoter, local chromatin environment and repositioning of active *var* to a different nuclear site for activation (Duraisingh et al. 2005; Freitas-Junior et al. 2005). This strict regulatory system is absent for the *stevor* genes as in our clonal lines multiple *stevor* are on and they are all acetylated at the same time which means the additional features that control *var* genes are absent for *stevor*.

4.5 *Stevor* H3K4 methylation activation marks are enriched at different genomic regions and may potentially play supporting roles.

PF3D7_1040200 showed H3K9 acetylation extending into the coding region in contrast to the other *stevor* (except PF3D7_0617600 in the 3.2C where H3K9 acetylation also extended the whole region). PF3D7_1040200 lacked H3K4 dimethylation and had a localized H3K4 trimethylation enrichment (Figure 3.73). This would suggest that the H3K9 acetylation is sufficient to maintain an open chromatin environment and H3K4 methylation modifications are not necessary for this gene. In contrast, PF3D7_1254100 only had had H3K9 acetylation levels in the 5' UTR (Figure 3.71). Therefore, for this gene, it would be essential to have another activating mark associating with it in the coding region, especially since the most stable nucleosomes are stationed in the coding regions (Bunnick et al. 2014). So for this gene, we observed H3K4 dimethylation enriched in the coding region and H3K4 trimethylation enriched 600bp upstream of the ATG to 400bp downstream of the ATG (Figure 3.73). For PF3D7_0617600, we observe H3K4 dimethylation dominating in the coding region as well as 600bp upstream of the ATG codon. H3K4 trimethylation enriches along the whole gene as well as the 5' UTR. These are the same regions where H3/H4 acetylation marks enrich, suggesting that this gene is decorated with these marks. Four histone marks, therefore, associated with this gene all through the gene body, again highlighting gene-specific regulatory marks between the *stevors*.

Figure 3.73: Position of H3K4 methylation enrichment for PF3D7_0617600, PF3D7_1040200, PF3D7_1254100 and PF3D7_1372500

H3K4 methylation enrichment on all four *stevors* studied. The *stevor* do show slight positional enrichment. PF3D7_0617600 has enrichment across the gene body as well 5' UTR regions; PF3D7_1040200 shows a localized enrichment spanning 200bp upstream and 200bp downstream of the ATG codon. The transgene showed clear separation of the two marks and PF3D7_1372500 showed clustering of the two in the coding region. The 5' UTR information for PF3D7_1372500 is missing because of problems with PCR design in this region.

For PF3D7_1372500 we observed the H3K4 methylation to be present in the coding region for both 6.2A and 3.2C. All in all, For PF3D7_1254100, PF3D7_0617600 and PF3D7_1372500, the H3K4 methylation reactions correlate with activation. This observation suggests that H3K4 methylation reactions with acetylation reactions are indirectly involved in some aspect of gene activation (PF3D7_1254100 and PF3D7_1040200) or also are directly involved (PF3D7_0617600 and PF3D7_1372500). As mentioned in Chapter 3B, some of these roles include stabilizing the RNA polymerase, modulating nucleosome stability during RNA Polymerase II transit, recruitment of HATs as well as inhibiting repressive marks to be deposited on active genes (Chen et al. 2011; Bian et al. 2011; Pray-Grant et al. 2005).

Lack of H3K4 dimethylation for PF3D7_1040200, also suggests that H3K4 dimethylation does not always associate with active *stevor*, much like what was observed for the *rifins* (Cabral et al. 2012). In *P. falciparum*, PFSET10, the enzyme that methylates H3K4, has been especially associated with the epigenetic memory of poised *var* genes. The evidence for this modification acting as a poised mark for *rifin* genes is not very strong (Cabral et al. 2012) and only highly expressed *rifin* genes seem to retain this modification as a marker of memory. This could argue that only *var* genes use H3K4 dimethylation to mark genes as the *var* genes exhibit strict mutual exclusion and they need to keep a memory of which single gene to activate in the next cycle. However, it needs to be highlighted that the evidence for H3K4 dimethylation being a marker for memory has only been shown for one *var* (Lopez-Rubio et al. 2007).

All the four *stevors* studied had H3K4 trimethylation modification when they were active (Figure 3.73). Again the differences among them were in the region of the enrichment. PF3D7_0617600 showed a preferential enrichment in the 5' UTR and coding regions, whereas PF3D7_1040200 and PF3D7_1254100, showed enrichment both in the 3' end of the 5' UTR as well as the coding

regions. The result for PF3D7_0617600 mirrors work of Salcedo-Amaya et al. (2009) and Bartfai et al. (2010) where H3K4 trimethylation enriched in the intergenic regions of active genes. However, the result for the other two *stevor* genes mirrors the work done by Gupta et al. (2012), where H3K4 trimethylation preferentially enriched in the coding regions, in particular, the 5' end of ORF. All these results suggest that the regulation of *stevors* is a complex process which raises the question of how the chromatin modifiers that write these marks are recruited to the promoter regions.

It should be noted that, although we are looking at the modifications individually, biologically they may all work together to produce a mark/signature that provides information on the downstream processes to occur. For example, it is becoming evident that in some instances H3K4 trimethylation is required for H3 acetylation (Zhang 2006). This hypothesis termed the 'histone code' states that different histone modifications work together in concert to give the final code, which is read by proteins that read these modifications and then translated to a biological outcome (Strahl & Allis 2002). It is the homeostatic totality of the modifications, brought about by a series of 'writing' and 'erasing' events performed by histone modifying enzymes and effectors, that is important for the biological outcome. Recently a study has highlighted the role of this homeostatic balance in activation and repression of genes in *P. falciparum*. Karmodiya et al. (2015) used ChIP-seq to comprehensively map the epigenome of *P. falciparum* as was done by Gupta et al. (2012). They uncovered a novel role for H3K36 dimethylation as a novel global repressor of genes in *P. falciparum*. What was perhaps interesting was that, by matching their ChIP-seq data to RNA-seq data, they observed that gene regulation was fine tuned by the ratio of activation marks (H3K9 acetylation, H3K4 di and trimethylation, H4K4 acetylation, H3K27 acetylation and H3K14 acetylation) to H3K36

dimethylation at a global scale, further providing evidence for a balance between the histone modifications in leading to gene regulation.

Interestingly our analysis of the activating or silencing marks showed that they were not mutually exclusive between the ‘off’ and ‘on’ culture. Thus we can conclude that activation or repression perhaps depends on a set threshold that maintains a homeostasis of the enrichment levels. How the control threshold is determined is not understood, although *P. falciparum* and other organisms have proteins that catalyze the reverse reactions of HATs or methylases (Cui & Liao 2010). These proteins are essential for the balance of activation and repression and for proper gene regulation (Shahbazian & Grunstein 2007).

Though there are some differences in the activation profiles of these multigene families the epigenetic landscape is rather similar in terms of what modifications associate with activation and repression. Acetylation reactions and H3/K4 methylation reactions associate with active *stevor*, *var* and *rifin* and H3K9/36 trimethylation associate with silent *stevor*, *var*, and *rifin*. This is an interesting observation considering Howitt et al. (2009) study where they propose that the three multigene families, *var*, *rifin* and *stevor* potentially shared a common activation factor. They all share same modifications for activation and repression. Other studies do not support the Howitt study conclusion (Witmer et al. 2012; Sharp et al. 2006; Freitas et al. 2005) instead showing that common activation factors could be shared between upsAvar and *rifin* Atype.

Another proposition is that there is a transcription permissible region in the nucleus where all these gene families are translocated and activated. This hypothesis is more plausible for *var* genes as FISH studies do show delocalization of an active *var* from the silent cluster of *vars* (Mok et al. 2008; Brodin et al. 2009; Ralph et al. 2005), though there is conflicting data for the

actual movement of the active *var* genes (Voss et al. 2006). Recent studies instead suggest that this supposed *var* expression site can house more than one *var* (Dzikowski et al. 2007; Jiang et al. 2013) and possibly other gene families (Howitt et al. 2009). A study by Volz et al. (2010) localized a number of proteins to the nuclear periphery which they hypothesized were chromatin regulators. Thus it seems this supposed *var* expression site would have all the necessary specific transcription factors as well as chromatin machinery for the activation of the multi-genes. The recruitment of *var* genes to this activation site has been well studied, with F-actin and the *var* intron playing roles (Zhang et al. 2011) but for *rifins* and *stevors* the mechanism of recruitment to this region if at all, is an unstudied area.

A recent study suggested a different mode of activation though for *stevors* and *rifins*. They argue for the *var* information of mutual exclusion expression is encoded in the promoter region. In their genome-wide studies, only the *var* promoter could silence endogenously expressed *var* (Witmer et al. 2012) whereas the other gene families (*rifin* and *stevor*) could not. They argue that for these gene families, activation and silencing could largely be dependent on the local chromatin environment i.e. the neighboring gene-environment. This is an interesting hypothesis, which warrants further study especially if the *rifins* and *stevors* do not also delocalize to different locations when active. The data in this thesis supports the role of histone modifications having a direct or indirect role in gene transcription. Another hypothesis would be that there are multiple expressions of these genes families and the parasite uses them to confuse the host to target them for antibody production while the important ligands are hidden from the host.

4.6 *Stevors* grouping does not change epigenetic profile

In this thesis, I was able to perform some rudimentary analysis of *stevors* and show that *stevors* group in two groups and that this grouping is based on number of parameters including whether the *stevors* is coding or a pseudogene, the orientation of its transcription and most importantly the presence or absence of particular motifs in the 5' UTR sequence. With this analysis, *stevors* were grouped into *stevorupsA* and *stevorupsB* with some hybrid *stevors* which had been grouped in either group but had characteristics of the other *stevors* in the other group. One of these *stevor* that I looked at PF3D7_1372500 grouped into the hybrid group *stevorupsB/A*. This *stevor* showed a similar epigenetic profile in the coding region as PF3D7_0617600, which grouped in A. This potentially suggests that this grouping of *stevors* does not lead to a major change in the epigenetic footprint of *stevors*. However, the data for this is limited, as primers specific for the 5' UTR could not be designed so this information is missing.

4.7 Conclusion

In conclusion, we show for the first time that H3/H4 acetylation reactions and H3K4 methylation reactions associate with active *stevors*. This data greatly enhances our understanding of how these virulence genes are regulated by the parasite. We also show how H3K4 methylation probably plays supporting roles in activation of genes. Perhaps the important observation is how slight differences in enrichment profiles for the different genes could be linked to the lack of mutually exclusive expression of *stevor* genes. Unlike the *var* genes, *stevors* do not seem to show a strict mode of regulation as there exist differences between them especially in enrichment patterns of H3K9/36 trimethylation marks. Some *stevors* have multiple silencing marks whereas others have only 1. This difference in silencing marks could regulate expression of the *stevors*. This was an interesting observation as other studies have shown that the parasite has a bet-

hedging strategy when it comes to expression of the host-parasite genes keeping them in a poised state of activation where it's possible to switch from one variant to another quickly when confronted by the host immunity or other environmental challenges (Rovira-Graells et al. 2012). This could potentially be done by histone modifications as we also observed H4K20 trimethylation correlating with activation in one of the clonal lines.

Though we found marks that associate with active *stevors*, this study like many other studies looking at histone modifications is correlative. It would be interesting to see if a knockdown of the genes responsible for the modifications we see affects the regulation of *stevors*. In this study, we utilized data from one such KO line (*PfSETvs* KO) and we could see some discrepancies not in terms of the overall conclusion but in the specific *stevors* that are enriched for this modification. KO data was missing for PF3D7_1372500. Unfortunately, the genes responsible for H3 and H4 acetylation and H3K9 trimethylation are essential in *P. falciparum* and cannot be knocked out. But the genes responsible for H3K4 trimethylation have been shown to be able to be knocked out (Jiang et al. 2013). Using such a line would allow us to determine whether the H3K4 methylations we observed is indeed necessary for *stevor* transcription or whether without them, *stevor* regulation is impaired. This would be ideal as in some cases like for 3D7_1040200, we could see H3K4 dimethylation enrichment when the gene was both on and off making it challenging to assign it a role for this gene.

Like all other chromatin experiments as well, this study is heavily dependent on the specificity of antibodies used. I tried to use ChIP-validated antibodies that have been used by other researchers. I did a bioinformatics analysis to show that the peptides used to for the antibodies are also conserved in *P. falciparum*. And I also ran a western blot to show that the commercial antibodies used do recognize the modifications in the *P. falciparum* genome and they do. I did observe a

second band of unknown origin in the H4K8 acetylation lane. Not sure of the nature of this band but could be a non-histone protein in the nucleus, as nuclear extract was used for the blot. This would have been digested by the Proteinase K before DNA purification, so I don't think the presence of the extra band, minimizes my result. I did, however, have a look at some of the blots on the Abcam web page submitted by other researchers and some of the blots do show extra bands. I also opted to use polyclonal antibodies to maximize the chances of the precipitation. However, with all these weaknesses I do still believe that the results from his thesis are valid and instrumental in increasing our knowledge of *P. falciparum*. We summarize our findings on histone modification regulating *stevors* transcription with a general model (Figure 3.74).

Figure 3.74: General model showing histones modifications associated with activation and repression of *stevor*.

The diagram is based on work done in this thesis. Note for the H3K9/36trimethylation schematic, we considered data for PF3D7_1254100, PF3D7_0617600, PF3D7_1372500 and PF3D7_1040200 separately as *stevor* off 1 and off 2. The proposed diagram shows H3K9/H4 acetylation as involved in *stevor* activation. These modifications are located in the 5' UTR/intergenic regions and were consistent for all three *stevors* studied. Other sets of activation marks include H3K4 di and trimethylation. These were spread out between the 5' UTR/Intergenic and coding regions. Silencing markers enriched more in the coding regions as depicted here and we propose that *stevor* is repressed at this region.

CHAPTER 5

FUTURE WORK

In this closing chapter, I examine some of the limitations of this work and future experiments that can be done to address these.

1. One of the hypotheses we suggested is that active *stevors* are located in a different nuclear region than silent *stevors*. This was based on work that has shown that active *var* and *rifins* co-localize at distinct subnuclear locations (Howitt et al. 2009, Voss et al. 2006, Freitas-Junior et al. 2005). This has not been shown for *stevors* though it seems reasonable to assume this. Using the current clones that we have it is now possible to address this question using techniques like RNA-FISH/ DNA-FISH. Clone AF5 used in this study has two active *stevors*, the transgene, and 3D7_1040200. These two *stevors* can be labeled with two different fluorophores. Fluorescence microscopy can then be used to visualize whether the two active *stevors*, co-localize. The second part would then to label the telomeric ends and again study to see if an active *stevor* detaches from the subtelomeric ends upon activation.
2. In this thesis, we show that H3K36 trimethylation associates with three out of four silent *stevors*. However, the mechanism for deposition of this mark on the *stevors* is unknown. This would be an interesting area of study as by knowing how these histone writers are deposited on the genes, it would greatly enhance our understanding of the parasite biology and could also provide novel avenues for novel therapies. Several long non-coding RNA's have been identified in the subtelomeric regions of the chromosomes and

some of these were correlated with H3K36 trimethylation on *var* genes. With the latest CRISPR technology now working in *Plasmodium*, creating a knockout line of these RNAs and seeing which are involved in silencing of *stevors* and if their presence correlates with H3K36 trimethylation could help in deciphering whether these RNAs are involved in this process or not. With the CRISPR technology, the integration experiments that I initially tried to do and failed can be revisited. *Stevor* integrated clones can help in studying other aspects of *stevor* regulation.

3. The interesting thing about this work was the realization that *stevors* can be grouped into separate groups based on their 5' upstream regions. The analysis done here could set the groundwork for discovering the *cis* elements involved in *stevor* regulation. As a starting point, promoter deletion experiments could be performed to identify the minimal promoter or regions needed for *stevor* activation. Also, experiments in which the motifs discovered here are deleted from the promoter would be instrumental in realizing the functional significance of these discovered motifs. One could also check to see if any of the sequences are targets of ApiAP2 transcription factors. Other analyses could include looking at the available data in the literature on the localization of different *stevors* and see how this data fits into the groups proposed here. Additional work could include checking common temporal profiles of the *stevors*. Though *stevors* are most commonly transcribed during the trophozoite stage of the IDC, *stevors* are also known to be expressed in the gametocyte, merozoite stages, and sporozoites stages. All this information could help in fine-tuning the in-silico analyses observed here.

4. Some of the modifications still need to be resolved in terms of *stevor* regulation. One of them is H4K20 trimethylation which associated with active PF3D7_0617600 in the 3.2C clone but not the 6.2A clone. Additional clones would be useful for this. In fact a third clone was initially grown for this the analysis of multiple clones, however, the clone was lost to contamination and there was not enough time to re-grow another and the replicates needed. Another modification that potentially requires more work is H3K36 trimethylation. And a proposed experiment is to create ‘in-house KO lines” of the SET domain-containing methylases for H3K4 methylations, H4K20 methylation, and H3K36 trimethylation, using CRISPR/Cas9 system and studying their effect on *stevor* regulation. The Jiang et al data used for this study showed some discrepancies with my data. And more work still needs to be done.

SUPPLEMENTARY

Gene_ID	new TSS	strand	gene start	gene end
PF3D7_0101800	87202	-	87203	88177
PF3D7_0102100	95049	+	95050	95899
PF3D7_0114600	560473	-	560474	561534
PF3D7_0115400	598464	+	598465	599472
PF3D7_0200400	no data			
PF3D7_0200900	no data			
PF3D7_0201300	66528	+	66529	67545
PF3D7_0221400	860465	+	860466	861504
PF3D7_0222800	899429	+	899430	900445
PF3D7_0300400	52279	-	52280	53273
PF3D7_0324600	1021997	+	1021998	1023008
PF3D7_0400800	66600	-	66601	67573
PF3D7_0401500	89152	+	89153	90184
PF3D7_0402600	159278	-	159279	160314
PF3D7_0425500	1146671	-	1146672	1147697
PF3D7_0500600	39442	-	39443	40488
PF3D7_0617600	738258	-	738259	739265
PF3D7_0631900	1333012	+	1333013	1334035
PF3D7_0700400	36921	-	36922	37927
PF3D7_0700700	45558	+	45559	46607
PF3D7_0732000	1385634	+	1385635	1386626
PF3D7_0832000	1383348	-	1383349	1384377
PF3D7_0832400	1399559	-	1399560	1400583
PF3D7_0832600	1405834	-	1405835	1406999
PF3D7_0832900	1414282	+	1414283	1415293
PF3D7_0900900	55073	-	55074	56081
PF3D7_0901600	77860	+	77861	78874
PF3D7_1000800	57967	+	57968	58980
PF3D7_1040200	1605929	+	1605930	1606953
PF3D7_1100700	56791	+	56792	57840
PF3D7_1149900	2009874	+	2009875	2010902
PF3D7_1254100	2204723	+	2204724	2205714
PF3D7_1254300	2211542	+	2211543	2212554
PF3D7_1254600	2221834	+	2221835	2222848
PF3D7_1300900	62514	-	62515	63547

PF3D7_1372500	2850796	-	2854544	2855568
PF3D7_1372800	2863865	+	2863866	2864869
PF3D7_1400700	24336	-	24337	25320
PF3D7_1479500	3269493	+	3269494	3270496
PF3D7_1479900	3282728	+	3282729	3283752

Supplementary Table 1: *Stevor* TSS from Karmodiya et al. (2015).

The table only shows data for *stevor* genes. The data maps the TSS of *stevor* genes to be at the start of the coding region

Stevor gene	3D730h- Plasmodium.mas5- Signal	3D730h- Plasmodium.mas5- Detection	3J830h- Plasmodium.mas5- Signal	3J830h- Plasmodium.mas5- Detection
PF3D7_0300400	12	A	47	A
PF3D7_0732000	40.2	A	4.4	A
PF3D7_0115400	196.7	P	372.1	P
PF3D7_0201300	17.7	A	3.3	A
PF3D7_0631900	265.7	P	707.3	P
PF3D7_0222800	3.6	A	43.9	A
PF3D7_1400700	83.8	A	169.6	P
PF3D7_0425500	136.8	A	528.7	P
PF3D7_0101800	9.7	A	2.4	A
PF3D7_0402600	38.7	A	3294.8	P
PF3D7_1149900	313.4	A	1152.5	P
PF3D7_0200400	47.6	A	6.7	A
PF3D7_0901600	37.9	A	45.9	A
PF3D7_1254600	2.2	A	63.6	A
PF3D7_1479900	102.2	A	160.3	P
PF3D7_1300900	85.8	A	110.3	P
PF3D7_1479500	7.1	A	965.3	P
PF3D7_0221400	61.4	A	54.5	P
PF3D7_0900900	5.1	A	44.9	A
PF3D7_1254100	20.7	A	69.2	M
PF3D7_1040200	71.1	A	100.6	P
PF3D7_0400800	217.5	P	257.9	P
PF3D7_0617600	1724.6	P	18172	P
PF3D7_0324600	156.7	P	318.3	P

Supplementary Table 2: KO data from Jiang et al (2013) for *stevor* at 30hr post-invasion.

Data only shows what was available from the supplementary information. The letters A, M, and P means Absent, Marginal and Present respectively. This annotation is used to describe the RNA signal 'detection p-value', and is subsequently used to generate a 'detection call'. The paper mentions 18 *stevors* were affected but on the table we see only 11.

Primer annotation		Primer efficiency	R2 value	Slope
0201300 5'3	A	95.867	0.994	-3.425
0201300 5'4	B	82.34	0.89	-3.67
0201300 cod1	C	95.637	0.992	-3.431
0201300 cod2	D	95.764	0.999	-3.428
0201300 cod3	E	96.658	0.997	-3.405
0617600 5'1	A	78.294	0.995	-3.982
0617600 5'2	B	75.469	0.992	-4.095
0617600 5'3	C	87.058	0.999	-3.671
0617600 5'4	D	86.146	0.998	-3.706
0617600 cod1	E	89.526	0.999	-3.601
0617600 cod2	F	83.901	0.997	-3.78
0617600 cod3	G	89.715	0.995	-3.596
0617600 cod4	H	102	0.996	-3.255
0700400 5'1	A	88.862	0.999	-3.621
0700400 5'2	B	93.292	0.995	-3.494
0700400 5'3	C	90.74	0.998	-3.566
0700400 cod1	D	99.66	0.996	-3.33
0700400 cod2	E	105.339	0.998	-3.2
0700400 cod3	F	95.044	0.996	-3.447
0832400 5'1	A	87.735	0.999	-3.656
0832400 5'2	B	97.295	0.998	-3.388
0832400 5'3	C	92.798	0.994	-3.508
0832400 cod1	D	108.9	0.997	-3.126
0832400 cod2	E	96.045	0.998	-3.42
0832400 cod3	F	93.182	0.997	-3.497
0900900 5'1	A	104.084	0.996	-3.228
0900900 5'2	B	97.331	0.996	-3.388
0900900 5'3	C	96.045	0.998	-3.42
0900900 cod1	D	99.383	0.99	-3.337
0900900 cod2	E	100.336	0.998	-3.314
0900900 cod3	F	97.82	0.994	-3.375
1040200 5'1	A	94.701	0.994	-3.456
1040200 5'2	B	107.231	0.989	-3.16

1040200 5'3	C	103.998	0.996	-3.23
1040200 5'4	D	113.194	0.992	-3.042
1040200 cod1	E	97.888	0.99	-3.374
1040200 cod2	F	101.234	0.994	-3.293
1040200 cod3	G	100.517	0.989	-3.31
1040200 cod4	H	103.132	0.996	0.342
1254100 5'1	B	100.949	0.988	-3.299
1254100 5'2	D	99.39	0.996	-3.323
1254100 5'3	E/F	85.095	0.991	-3.74
1254100 5'4	G	98.96	0.964	3.347
1254100 cod1	H	97.561	0.998	-3.382
1254100 cod2	I	92.705	0.998	-3.51
1254100 cod3	J	90.334	1	-3.578
1254100 cod4	K	94.878	0.996	-3.451
1254100hdhfr cod1	H	99.231	0.997	-3.341
1254100hdhfr cod2	I	98.411	0.999	-3.361
1254100hdhfr cod3	J	93.394	0.999	-3.491
1254100hdhfr cod4	K	94.878	0.998	-3.451
1254100hdhfr cod5	L	94.865	0.998	-3.431
1254600 5'2	A	90.358	0.999	-3.577
1254600 5'3	B	95.672	0.998	-3.48
1254600 cod1	C	94.145	0.998	-3.471
1254600 cod2	D	99.976	0.997	-3.322
1254600 cod3	E	102.704	0.99	-3.259
1372500 cod1	A	101.376	0.996	-3.289
1372500 cod2	B	103.661	0.995	-3.237
1372500 cod3	C	89.442	0.996	-3.604

Supplementary Table 3:Primer efficiencies of primers used for ChIP-QPCR studies

Efficiencies were calculated by running primers on 5 fold serial dilution of gDNA.

stevor	slope	efficiency	R2
PF3D7_0101800	-3.395	97.042	0.999
PF3D7_0102100	-3.176	106.472	0.997
PF3D7_0114600	-3.291	101.303	0.999
PF3D7_0115400	-3.36	98.423	0.999
PF3D7_0200400	-3.361	98.415	0.997
PF3D7_0200900	-0.342	99.177	0.997
PF3D7_0201300	-3.378	97.323	0.998
PF3D7_0221400	-3.352	98.764	0.999
PF3D7_0222800	-3.291	101.303	0.999
PF3D7_0300400	-3.359	98.456	0.999
PF3D7_0324600	-3.315	100.273	0.998
PF3D7_0400800	-3.41	96.458	0.999
PF3D7_0401500	-3.411	96.416	0.999
PF3D7_0402600	-3.48	93.812	0.999
PF3D7_0425500	-3.335	99.445	0.999
PF3D7_0500600	-3.236	103.72	0.998
PF3D7_0617600	-3.772	84.118	0.999
PF3D7_0631900	-3.322	99.976	0.999
PF3D7_0732000	-3.462	94.484	0.997
PF3D7_0832900	-3.343	99.14	0.997
PF3D7_0832400	-3.353	98.724	0.999
PF3D7_0832000	-3.351	98.569	0.998
PF3D7_0832600	-3.431	95.622	0.998
PF3D7_0900900	-3.383	97.517	0.998
PF3D7_0700400	-3.3	100.915	0.995
PF3D7_0901600	-3.404	96.701	0.999
PF3D7_1000800	-3.188	105.891	1
PF3D7_1040200	-3.297	101.038	0.998
PF3D7_1100700	-3.397	0.998	96.948
PF3D7_1149900	-3.25	0.995	103.074
PF3D7_1254100	-3.378	0.998	97.692
PF3D7_1254300	-3.503	0.998	92.965
PF3D7_1254600	-3.523	0.998	99.228
PF3D7_0700700	-3.256	0.998	102.822
PF3D7_1372500	-3.489	93.474	0.996
PF3D7_1372800	-3.373	97.896	0.999
PF3D7_1400700	-3.341	99.198	0.998
PF3D7_1479900	-3.36	98.456	0.997
PF3D7_1479500	-3.392	97.149	0.998
PF3D7_1300900	-3.368	98.133	0.998

actin	-3.4	96.842	0.999
hsp80	-3.336	99.416	0.997
trna	-3.429	95.71	0.995

Supplementary Table 4: Primer efficiency of *stevor* genes and genes used as controls.

Efficiencies were calculated by running primers on 5 fold serial dilution of gDNA.

Gene		Blast hit	Blast %	Gaps	Gaps %
PF3D7_1254100	A	232/236	98%	3/236	1%
PF3D7_1254100	B	239/241	99%	2/241	1%
PF3D7_1254100	C	204/204	100%		0%
PF3D7_1254100	D	84/85	99%	1/85	1%
PF3D7_1040200	A	196/197	99%	0/197	0%
PF3D7_1040200	B	221/225	98%	1/225	0%
PF3D7_1040200	C	140/158	89%	4/158	3%
PF3D7_1040200	D	99/99	100%		0%
PF3D7_021300	A	212/213	99%	0/213	0%
PF3D7_021300	B	232/241	96%	3/214	1%
PF3D7_021300	C	215/219	98%	1/219	0%
PF3D7_0700400	A	229/231	99%	0/231	0%
PF3D7_0700402	B	263/265	99%	0/265	0%
PF3D7_0700403	C	235/243	98%	1/243	0%
PF3D7_0900900	A	105/105	100%		0%
PF3D7_0900900	B	233/238	98%	0/238	0%
PF3D7_0900900	C	206/206	100%		0%
PF3D7_0832400	A	290/290	100%		0%
PF3D7_0832400	B	207/209	99%	2/209	1%
PF3D7_0832400	C	272/276	99%	1/275	0%
PF3D7_1254600	A	221/234	94%	1/234	0%
PF3D7_1254600	B	217/226	96%	2/226	1%
PF3D7_1254600	C	164/165	99%	1/165	1%
PF3D7_0617600	A	177/181	98%	1/181	1%
PF3D7_0617600	B	205/206	99%	1/206	0%
PF3D7_0617600	C	205/206	99%	0/206	0%
PF3D7_0617600	D	241/242	99%	1/242	0%

Supplementary Table 5: Sequencing results of 5'UTR PCR products.

The table shows the blast results for the PCR products of the 5' UTR region used for ChIP-qPCR. I did not calculate any values. The blast function in plasmodb did the calculation.

Sequence 1- Supplementary Figure 1

>PF3D7_1479500_site_1 offset= 297

GGTTTACACCCATAGCAACTATTGTTCTACACATTTATATGTGTATTTGC

>PF3D7_0300400_site_1 offset= 297

GGTTTACACCCATAGCAACTATTGTTCTACACATTTATATGTGTATTTGC

>PF3D7_0732000_site_1 offset= 276

GGTTTACACCCATAACAACACTATTGATCTACACATTTATATGTGTATTTGC

>PF3D7_0700400_site_1 offset= 335

GGTTTACACCCATAACAACACTATTGTTCTACACATTTATATGTGTATTTGA

>PF3D7_1400700_site_1 offset= 343

GGTTTACACCCATAACAACACTATTGTTTATACACATTTATATGTGTATTTGA

>PF3D7_0222800_site_1 offset= 347

GGTTTACCCCCATAACAACACTATTGTTTATACACATTTATATGTGTATTTGA

>PF3D7_1254300_site_1 offset= 276

GGTTTACACCAATAACAACACTATTGATCTACACATTTATATGTGTGTTTGC

>PF3D7_0400800_site_1 offset= 275

GGTTTACACCAATAACAACACTATTGATCTACACATTTATATGTGTGTTTGC

>PF3D7_1254100_site_1 offset= 349

GGTTTACACCCATAACAGCTATTGTTTATACACATTTATATGTGTATTTGA

>PF3D7_0832900_site_1 offset= 334

GGCTTACACCAATAACAACACTATTGTTCTACACATTTATATGTGTATTTGA

>PF3D7_0631900_site_1 offset= 340

GGTTTACACCCATAACAACACTATTGTTTATACACATTTATATGTGTATATGA

>PF3D7_1479900_site_1 offset= 346

GGTTTACACGCATAACAACACTATTGTTCTACACATTAATATGTGTATTTGA

>PF3D7_0200900_site_1 offset= 349

GGTTTACACGCATAACAACACTATTGTTCTACACATTAATATGTGTATTTGA

>PF3D7_0832400_site_1 offset= 288

GGTTTACCCCCATAACAACACTGTTGTTCTACACATTTATATGGATATTCGC

>PF3D7_0200400_site_1 offset= 340

GGTTTACACCCATAACAACACTATTGTTATACACATTTATATGTGTATTTTA

>PF3D7_1300900_site_1 offset= 355

GGTTTATACCCATAACAACACTATTGTTATACACATTTATATGTGTATTTGA

>PF3D7_0900900_site_1 offset= 335

GGTTTATACCCATAACAACACTATTGTTATACACATTTATATGTGTATTTGA

>PF3D7_0324600_site_1 offset= 335

GGTTTATACCCATAACAACACTATTGTTATACACATTTATATGTGTATTTGA

>PF3D7_0401500_site_1 offset= 274

GGTTTACACCAAAAACAACACTATTGATCCACACATTTATATGTGTATTTGC

>PF3D7_0114600_site_1 offset= 301

GGTTTACACTCATAACAACACTATTGATCTACACATTTATATGTGTATATGC

>PF3D7_1254600_site_1 offset= 329

GGTTTACACCCATAACAACACTATTTTTATACACATTTATATGTGTACTTGA

>PF3D7_0901600_site_1 offset= 259

GGTTTACACCAAAAACACTACTATTGATCCACACATTTATATGTGTATTTGC

>PF3D7_1040200_site_1 offset= 351

GGTTTACACCCATAACAATTATTGTTATACACATTTATATATGTATTTGA

>PF3D7_1149900_site_1 offset= 349

GATTTACACCCATAACAACACTATTTTTATACACATTTATATGTGTATTTGA

>PF3D7_0617600_site_1 offset= 303

GGTATATAACCATAACAACACTATTGTTCTACACATTTGTATGTGTATTTGA

>PF3D7_0115400_site_1 offset= 236

GGTTTGTACCTATAAGAATTATTGTTCTACACATTTATATGTGTATTTGC

>PF3D7_0101800_site_1 offset= 341

GGTTTACACCCATAACAACACTATTTTTATACACATTTTATGTGTATTTAA

>PF3D7_1372800_site_1 offset= 331

GGTTTGCACACATAACAACACTATGGTTATATACATTTATATGTGTATTTGA

>PF3D7_0425500_site_1 offset= 253

GTTTTACACCTATAACAACACTATTAATCTACGCATTTATATAGATTTTTTC

>PF3D7_1372500_site_1 offset= 259

GTTTTACACCTATAACAACACTATTAATCTACGCATTTATATAGATTTTTTC

>PF3D7_0402600_site_1 offset= 245

GGCCTACACCTATAACAAGTACTATTCTACACATTTATATAGATTTTTTTT

>PF3D7_0832000_site_1 offset= 235

GTTATACAACCTATAACAACACTATTAATCTACACATGTATATAGATTTTTTC

>PF3D7_1000800_site_1 offset= 275

ATTATACACCCATAACAACACTATTATTAACACATTTATATATATTTCTTA

>PF3D7_0201300_site_1 offset= 259

GAATTAAACCCATAATAACTATTGTTCTACACATTTATTTTTATAGTAAC

Sequence 2- Supplementary Figure 2

>PF3D7_1149900_site_1 offset= 90

ACATAAAATAATAGATACAATAGAGGGTATTTTCTAACATAGAATAACAC

>PF3D7_1372800_site_1 offset= 72

ATATAAAACAATAGATACAATAGAGAGGATTTTCTATCATAGAATAATGC

>PF3D7_0324600_site_1 offset= 75

ACATAAATTAATAGATACAATAGAGAGAATTTTCTGTCATATAATAAGGC

>PF3D7_1479500_site_1 offset= 37

ATATATAATAATAGATACAATAGAGAGTATTTTCTAAAATAGAATAACAC

>PF3D7_0300400_site_1 offset= 37

ATATATAATAATAGATACAATAGAGAGTATTTTCTAAAATAGAATAACAC

>PF3D7_1400700_site_1 offset= 84

ACATAAAATAATAGATACAATAGAGAGAATTTTCTGTCATATAATAAGGC

>PF3D7_0200400_site_1 offset= 82

ACATAAATTAATAGATACAATAGAGAGTATTTTTTAACATAGAATAAAAAC

>PF3D7_0832900_site_1 offset= 75

ACATAAATTAATAGATAAAATAGAGAGGATTTTCTATCATATAATAAGGC

>PF3D7_0832900_site_2 offset= 657

ATATATATTAGAAAACAAAATAGATCATTTTTTGTGAATTAGATTAATAC

>PF3D7_0732000_site_1 offset= 16

ACATAAAATAATAGATACAATAGAGAGCATTTTCTAACACAAAAGAAGGC

>PF3D7_1254300_site_1 offset= 664

ATATATATTAGAAAACAAAATAGATCATTTTTTATGAATTAGATTAATAC

>PF3D7_0901600_site_1 offset= 2

TAACAAAATAATAGATACAATAGAGAGCATTTTCTAACATAGAATAATAC
>PF3D7_0631900_site_1 offset= 664
ATATATATTAGAAAACAAAATAGATCATTTTTTATAAATTAGATTAATAC
>PF3D7_0631900_site_2 offset= 82
TTATAAATTAATAGATACAATAGAGAGAATTTTCTATCATATAATAATGC
>PF3D7_0200400_site_2 offset= 664
ATATATATTAGAAAACAAAAGAGATCATTTTTTGTGAATTAGATTAATAC
>PF3D7_1372800_site_2 offset= 658
ATATATATTAGAAAACAAAATAGATAATTTTTTATGAAATAGATTAATAC
>PF3D7_1149900_site_2 offset= 663
ATATATATTAGAAAACAAAATAGATAATTTTTTATGAATTAGATTAATAC
>PF3D7_1040200_site_1 offset= 92
AAATAAAATAAGAGATACAATAGAGAGTATTTTATAATATAGGATAATGC
>PF3D7_0200900_site_1 offset= 663
ATATATATTAGAAAATAAAACAGATCATTTTTTATAAATTAGATTAATAC
>PF3D7_0700400_site_1 offset= 664
ATATATATTAGAAAACAAAATAGATCATTTTTTATGAATTAGATTAATAA
>PF3D7_0400800_site_1 offset= 663
ATATATATTAGAAAACAAAATAGATCATTTTTTATGAATTAGATTAATAA
>PF3D7_0115400_site_1 offset= 661
TTATGCATTAGAAAACAAAATAGATCATTTTTTATGAATTAGATTAATAC
>PF3D7_1479500_site_2 offset= 664
ATATATATCAGAAAACAAAATAGATCATTTTTTATGAATTAGATTAATAA
>PF3D7_0401500_site_1 offset= 15
ATATAAAACAATAGATATAATAGAGAGCATTTTATAATATAGAATAATGC
>PF3D7_0300400_site_2 offset= 664
ATATATATCAGAAAACAAAATAGATCATTTTTTATGAATTAGATTAATAA
>PF3D7_0114600_site_1 offset= 41
ACATAAAATAATAGATAGAATAGATATTATTTTCTAACCTAGAAGAACAC
>PF3D7_1254600_site_1 offset= 68

ACATAAATTAATAGATACAATAGAGAGAATTTTCTATTATATAATAAGAA
>PF3D7_1300900_site_1 offset= 664

ATATATATTAGAAAACAAAATAGATCATTTTTTTATTAATTAGATTAATAC
>PF3D7_0222800_site_1 offset= 87

TTATAAATTAATAGATACAATAGAGAGAATTTTTTATCATATAATAAGGC
>PF3D7_0832400_site_1 offset= 29

ACATAAAATAATAGATACAATAGAGAGAATTTTCTAACATATAATAAAAT
>PF3D7_0222800_site_2 offset= 662

ATATATATTAGAAAACAAAATAGATCATTTTTTTATTAATTAGATTAGTAC
>PF3D7_0700400_site_2 offset= 77

ACATAAAATAAAAGATACAATAGAGAGAATTTTCTATTGGCGAATAGGTA
>PF3D7_0101800_site_1 offset= 664

ATATATATTATAAAACAAAATAGATCATTTTTTTATAAATTAGATTAATAC
>PF3D7_1040200_site_2 offset= 664

ATATATATTAGAAAACAAAACATATCATTTTTTTATGAATTAGATTAATAC
>PF3D7_0900900_site_1 offset= 664

ATATATATTAGAAAACAAAATATATCATTTTTTTATAAATTAGATTAATAC
>PF3D7_0324600_site_2 offset= 664

ATATATATTAGAAAACAAAATATATCATTTTTTTATAAATTAGATTAATAC
>PF3D7_0900900_site_2 offset= 78

ACATAAACTAATAGATACAATAGAGAGAATTTTCTATCATATAAGGCTTA
>PF3D7_0101800_site_2 offset= 83

ACATAAACTAATAGATACAATAGAGAGAATTTTCTATCATATAAGGCTTA
>PF3D7_0732000_site_2 offset= 663

TTATGCATTAGAAAACAAAATAGATCATTTTTTTGTAAATTAGATTAATAA
>PF3D7_0832400_site_2 offset= 662

TTATACATTATAAAATAAAATAGATCATTTTTTTGTAAATTAGATTAATAC
>PF3D7_1372500_site_1 offset= 5

AAATAACATAATAGATATATTAGAGAGTATATTCTAACCTAGGATAAGGC
>PF3D7_1000800_site_1 offset= 32

AAATAACACAATAGATACATTAGAGAGCATATTGTAACATAGAATAAACC
>PF3D7_0617600_site_1 offset= 43
ACATAAAATAATAGATACAATAGAGAGCATTTTATATGGGTGAATAGATA
>PF3D7_1254600_site_2 offset= 664
AAATATATTAAAGAATAAAATAGATCATTTTTTTATAAATTAGATTAATAA
>PF3D7_1479900_site_1 offset= 86
ATAAAAAATAGTAGATACAATAGAGAGTTATTTCTATTGGTGAATACGTA
>PF3D7_0200900_site_2 offset= 90
ATAAAAAATAGTAGATACAATAGAGAGTTATTTCTATTGGTGAATACGTA
>PF3D7_1400700_site_2 offset= 663
AAATTTATTAAAGAATAAAATAGATCATTTTTTTATAAATTAGATTAATAC
>PF3D7_0114600_site_2 offset= 662
ATATACATTACAAAATAAAATAGATCATTTGATATAAATTAGATTCATAC
>PF3D7_1254300_site_2 offset= 19
ACATAAAATAATAGATACAACAGAGATTATTTTTTAACATGGATAAAATT
>PF3D7_0400800_site_2 offset= 18
ACATAAAATAATAGATACAACAGAGATTATTTTTTAACATGGATAAAATT

Sequence 3-Supplementary Figure 3

>PF3D7_1372800_site_1 offset= 588

TGATGAAACTATAAGGTTATTATATCATATAAAATTACAACAAAATTTCCGC

>PF3D7_1479900_site_1 offset= 594

TGATGAAACTATAATGTTATTATATCATATAAAATTACAACAAAATTTCCGC

>PF3D7_1479500_site_1 offset= 595

TGATGAAACTATAATGTTATTATATCATATAAAATTACAACAAAATTTCCGC

>PF3D7_1300900_site_1 offset= 595

TGATGAAACTATAATGTTATTATATCATATAAAATTACAACAAAATTTCCGC

>PF3D7_1254300_site_1 offset= 594

TGATGAAACTATAATGTTATTATATCATATAAAATTACAACAAAATTTCCGC

>PF3D7_1254100_site_1 offset= 594

TGATGAAACTATAATGTTATTATATCATATAAAATTACAACAAAATTTCCGC

>PF3D7_0400800_site_1 offset= 593

TGATGAAACTATAATGTTATTATATCATATAAAATTACAACAAAATTTCCGC

>PF3D7_0300400_site_1 offset= 595

TGATGAAACTATAATGTTATTATATCATATAAAATTACAACAAAATTTCCGC

>PF3D7_1254600_site_1 offset= 594

TGATGAAACTATAATGTTATTATATCACAAAATTACAACAAAATTTCCGC

>PF3D7_0900900_site_1 offset= 594

TGATGAAATTATAAGGTTATTATATCATATAAAATTACAACAAAATTTCCGC

>PF3D7_0101800_site_1 offset= 594
TGATGAAATTATAAGGTTATTATATCATATAAAATTACAACAAAATTTCCGC

>PF3D7_0324600_site_1 offset= 594
TGATGAAATTATAAGGTTATTATATCATATAAAATTACAACAAAATTTCCGC

>PF3D7_0700400_site_1 offset= 592
TGATGAAACTATAATGTTATTATATCATATAAAATTACAACAAAATTTCCGC

>PF3D7_1149900_site_1 offset= 593
TGATGAAACTATAATGTTATTATATCATATAAAATTACAACAAAATTTCCGT

>PF3D7_0631900_site_1 offset= 594
TGATGAAACTATAATGTTATTATATAATAAAATTACAACAAAATTTCCGC

>PF3D7_1400700_site_1 offset= 593
CGATGAAACTATAATGTTCTTATATAATAAAATTACAACAAAATTTCCGC

>PF3D7_0901600_site_1 offset= 594
TGATGAAAATATAATGTTATTATATCATCAAATTATAAGAAAATTTCCGC

>PF3D7_0222800_site_1 offset= 592
TGATGAAGTTATAATGTTATTATATCATATAAAATTACAACAAAATTTCCGT

>PF3D7_0832900_site_1 offset= 587
TGATGAAACTATAATATTATTATATCATATAAGTTACAACAAAATTTTCGC

>PF3D7_0200900_site_1 offset= 593
CGATGAAACTATAATGTTCTTATATAATAAAATTACAACAAAATTTCTGC

>PF3D7_0200400_site_1 offset= 594
TGATGAAACTATAATATTATTATATCATATAAGTTACAACAAAATTTACGC

>PF3D7_1040200_site_1 offset= 594
TGATGAAACTATAATTTTATTATATTATAAAATTACAACAAAATTTCCGC

>PF3D7_0401500_site_1 offset= 582
TGATGAAAATATAATGTTATTATATCATTAATTATAAGAAAATTTCCGT

>PF3D7_0732000_site_1 offset= 593
TGACGAAAATATTATGTTATTATATCATGAAAATATAATAAAATTTCCGC

>PF3D7_0115400_site_1 offset= 591
TGACGAAAATATTATGTTATTATATCATGAAAATATAATAAAATTTCCGC

>PF3D7_0201300_site_1 offset= 533

TGACGAAACTATAAGGTTATTATATCATGAAAATATAAGAATATTTTCTC

>PF3D7_0617600_site_1 offset= 593

TAATTGAAATATAATGTTATTATATCATAAAAATATAAGAAAAATTCCTT

>PF3D7_0114600_site_1 offset= 592

GATGAAAAAATAATGTTCATATATCCTAATATTATAAGAACATTTCCGC

>PF3D7_0832400_site_1 offset= 590

TGATGAATATATAATTCTATTATATCATAATATTATAAAAAAAAATATCC

PUBLICATIONS AND CONFERENCES

Huang X, Huang S, Ong LC, Lim JC-S, Hurst RJM, Mushunje AT, Matsudaira PT, Han J, Preiser PR. (2015). “Differential remodeling associated with different levels of parasite virulence controls disease outcome in malaria parasite infections.”*mSphere* 1(1):e00018-15. doi:10.1128/mSphere.00018-15.

Mushunje, Annals T and Preiser, P.P. Epigenetic regulation of transcription of the stevor multigene family of Plasmodium falciparum. Poster presentation. EMBO Conference: Chromatin and Epigenetics – 2015

Mushunje, Annals T and Preiser, P.P. Epigenetic regulation of transcription of the stevor multigene family of Plasmodium falciparum. Poster presentation. EMBL Conference: BioMalPar XI: Biology and Pathology of the Malaria Parasite - 2015

Mushunje, Annals T and Preiser, P.P. Epigenetic regulation of transcription of the stevor multigene family of Plasmodium falciparum. Oral presentation: Courage Fund- Infectious disease Conference - 2015

REFERENCES

- Abdel-Latif, M.S. et al., 2003. Antibodies to Plasmodium falciparum Rifin Proteins Are Associated with Rapid Parasite Clearance and Asymptomatic Infections. *Infection and Immunity*, 71(11), pp.6229–6233.
- Abdel-Latif, M.S. et al., 2002. Recognition of variant rifin antigens by human antibodies induced during natural Plasmodium falciparum infections. *Infection and Immunity*, 70(12), pp.7013–7021.
- Albrecht, L. et al., 2011. var gene transcription and PfEMP1 expression in the rosetting and cytoadhesive Plasmodium falciparum clone FCR3S1.2. *Malaria Journal*, 10(1), p.17.
- Aley, S.B., Sherwood, J.A. & Howard, R.J., 1984. Knob-positive and knob-negative Plasmodium falciparum differ in expression of a strain-specific malarial antigen on the surface of infected erythrocytes. *The Journal of Experimental Medicine*, 160(November), pp.1585–1590.
- Allfrey VG, Faulkner R, M.A., 1964. Acetylation and Methylation of Histones and Their Possible Role in the Regulation of RNA Synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, 51(1938), pp.786–94.
- Amaratunga, C. et al., 2012. Artemisinin-resistant Plasmodium falciparum in Pursat province, western Cambodia: A parasite clearance rate study. *The Lancet Infectious Diseases*, 12(11), pp.851–858.
- Amit-Avraham, I. et al., 2015. Antisense long noncoding RNAs regulate var gene activation in the malaria parasite Plasmodium falciparum. *Proceedings of the National Academy of Sciences of the United States of America*, 112(9), pp.E982-91.
- An, Y., Wu, W. & Lv, A., 2010. A PCR-after-ligation method for cloning of multiple DNA inserts. *Analytical Biochemistry*, 402(2), pp.203–205.
- Andrews, K.T. et al., 2012. Comparative gene expression profiling of P. falciparum malaria parasites exposed to three different histone deacetylase inhibitors G. Snounou, ed. *PLoS ONE*, 7(2), p.e31847.
- Araki, Y. et al., 2009. Genome-wide Analysis of Histone Methylation Reveals Chromatin State-Based Regulation of Gene Transcription and Function of Memory CD8+ T Cells. *Immunity*, 30(6), pp.912–925.
- Aravind, L. et al., 2003. Plasmodium Biology: Genomic Gleanings. *Cell*, 115(7), pp.771–785.
- Bachmann, A. et al., 2015. A comparative study of the localization and membrane topology of members of the RIFIN, STEVOR and PfMC-2TM protein families in Plasmodium falciparum-infected erythrocytes. *Malaria Journal*, 14(1), p.274.
- Bachmann, A. et al., 2016. Mosquito Passage Dramatically Changes var Gene Expression in Controlled Human Plasmodium falciparum Infections R. Dzikowski, ed. *PLoS Pathogens*, 12(4), p.e1005538.
- Balaji, S. et al., 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(13), pp.3994–4006.

- Bannister, A.J. & Kouzarides, T., 1996. The CBP co-activator is a histone acetyltransferase. *Nature*, 384(6610), pp.641–3.
- Barragan, A. et al., 2000. The Duffy-binding-like domain 1 of *Plasmodium falciparum* erythrocyte membrane protein 1 (PfEMP1) is a heparan sulfate ligand that requires 12 mers for binding. *Blood*, 95(11), pp.3594–9.
- Bártfai, R. et al., 2010. H2A.Z demarcates intergenic regions of the *Plasmodium falciparum* epigenome that are dynamically marked by H3K9ac and H3K4me3 K. Kim, ed. *PLoS Pathogens*, 6(12), p.e1001223.
- Baruch, D.I. et al., 1999. CD36 peptides that block cytoadherence define the CD36 binding region for *Plasmodium falciparum*-infected erythrocytes. *Blood*, 94(6), pp.2121–2127.
- Baruch, D.I. et al., 1995. Cloning the *P. falciparum* gene encoding PfEMP1, a malarial variant antigen and adherence receptor on the surface of parasitized human erythrocytes. *Cell*, 82(1), pp.77–87.
- Baruch, D.I. et al., 1996. *Plasmodium falciparum* erythrocyte membrane protein 1 is a parasitized erythrocyte receptor for adherence to CD36, thrombospondin, and intercellular adhesion molecule 1. *Proceedings of the National Academy of Sciences of the United States of America*, 93(8), pp.3497–502.
- Baxevanis, A.D. & Landsman, D., 1996. Histone sequence database: A compilation of highly conserved nucleoprotein sequences. *Nucleic Acids Research*, 24(1), pp.245–247.
- Beeson, J.G. & Duffy, P.E., 2005. The Immunology and Pathogenesis of Malaria During Pregnancy. In *Immunology and Immunopathogenesis of Malaria*. Berlin/Heidelberg: Springer-Verlag, pp. 187–227.
- Beeson, J.G., Elliott, S.R. & Hommel, M., 2008. Polymorphic and Conserved Targets of Antibodies against *Plasmodium falciparum* during Pregnancy. *The Journal of Infectious Diseases*, 197(9), pp.1350–1351.
- Bernabeu, M. et al., 2012. Functional analysis of *Plasmodium vivax* VIR proteins reveals different subcellular localizations and cytoadherence to the ICAM-1 endothelial receptor. *Cellular Microbiology*, 14(3), pp.386–400.
- Bernstein, B.E. et al., 2005. Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell*, 120(2), pp.169–181.
- Bernstein, B.E. et al., 2002. Methylation of histone H3 Lys 4 in coding regions of active genes. *Proceedings of the National Academy of Sciences of the United States of America*, 99(13), pp.8695–8700.
- Bevington, S. & Boyes, J., 2013. Transcription-coupled eviction of histones H2A/H2B governs V(D)J recombination. *The EMBO Journal*, 32(10), pp.1381–92.
- Bian, C. et al., 2011. Sgf29 binds histone H3K4me2/3 and is required for SAGA complex recruitment and histone H3 acetylation. *The EMBO Journal*, 30(14), pp.2829–2842.
- Bischoff, E. & Vaquero, C., 2010. In silico and biological survey of transcription-associated proteins implicated in the transcriptional machinery during the erythrocytic development of *Plasmodium falciparum*. *BMC Genomics*, 11(1), p.34.
- Blythe, J.E. et al., 2008. *Plasmodium falciparum* STEVOR proteins are highly expressed in

- patient isolates and located in the surface membranes of infected red blood cells and the apical tips of merozoites. *Infection and Immunity*, 76(7), pp.3329–3336.
- Borst, P. et al., 1998. Control of VSG gene expression sites in *Trypanosoma brucei*. *Molecular and Biochemical Parasitology*, 91(1), pp.67–76.
- Borst, P. et al., 1997. Mechanisms of antigenic variation in African trypanosomes. *Behring Institute Mitteilungen*, March(99), pp.1–15.
- Bozdech, Z. et al., 2003. The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS Biology*, 1(1), p.e5.
- Brancucci, N.M.B. et al., 2014. Heterochromatin protein 1 secures survival and transmission of malaria parasites. *Cell Host and Microbe*, 16(2), pp.165–176.
- Brancucci, N.M.B. et al., 2012. Identification of a cis-acting DNA-protein interaction implicated in singular var gene choice in *Plasmodium falciparum*. *Cellular Microbiology*, 14(12), pp.1836–1848.
- Breman, J.G., 2012. Resistance to artemisinin-based combination therapy. *The Lancet Infectious Diseases*, 12(11), pp.820–822.
- Bright, A.T. & Winzeler, E.A., 2011. Noncoding RNA, antigenic variation, and the virulence genes of *Plasmodium falciparum*. *BMC Biology*, 9(1), p.50.
- Broadbent, K.M. et al., 2011. A global transcriptional analysis of *Plasmodium falciparum* malaria reveals a novel family of telomere-associated lncRNAs. *Genome Biology*, 12(6), p.R56.
- Brolin, K.J.M. et al., 2009. Simultaneous transcription of duplicated var2csa gene copies in individual *Plasmodium falciparum* parasites. *Genome Biology*, 10(10), p.R117.
- Brugat, T. et al., 2014. Sequestration and histopathology in *Plasmodium chabaudi* malaria are influenced by the immune response in an organ-specific manner. *Cellular Microbiology*, 16(5), pp.687–700.
- Bull, P.C. et al., 1998. Parasite antigens on the infected red cell surface are targets for naturally acquired immunity to malaria. *Nature Medicine*, 4(3), pp.358–60.
- Bull, P.C. & Marsh, K., 2002. The role of antibodies to *Plasmodium falciparum*-infected-erythrocyte surface antigens in naturally acquired immunity to malaria. *Trends in Microbiology*, 10(2), pp.55–58.
- Bunnik, E.M. et al., 2014. DNA-encoded nucleosome occupancy is associated with transcription levels in the human malaria parasite *Plasmodium falciparum*. *BMC Genomics*, 15(1), p.347.
- Burns, L.G. & Peterson, C.L., 1997. The yeast SWI-SNF complex facilitates binding of a transcriptional activator to nucleosomal sites in vivo. *Molecular and Cellular Biology*, 17(8), pp.4811–9.
- Cabral, F.J., Fotoran, W.L. & Wunderlich, G., 2012. Dynamic activation and repression of the *Plasmodium falciparum* rif gene family and their relation to chromatin modification. *PLoS ONE*, 7(1), p.e29881.
- Cabral, F.J. & Wunderlich, G., 2009. Transcriptional memory and switching in the *Plasmodium falciparum* rif gene family. *Molecular and Biochemical Parasitology*, 168(2), pp.186–190.

- Calderwood, M.S. et al., 2003. Plasmodium falciparum var Genes Are Regulated by Two Regions with Separate Promoters, One Upstream of the Coding Region and a Second within the Intron. *Journal of Biological Chemistry*, 278(36), pp.34125–34132.
- Callebaut, I. et al., 2005. Prediction of the general transcription factors associated with RNA polymerase II in Plasmodium falciparum: conserved features and differences relative to other eukaryotes. *BMC Genomics*, 6(1), p.100.
- Carlson, M. & Laurent, B.C., 1994. The SNF/SWI family of global transcriptional activators. *Current Opinion in Cell Biology*, 6(3), pp.396–402.
- Carlton, J., Silva, J. & Hall, N., 2005. The genome of model malaria parasites, and comparative genomics. *Current Issues in Molecular Biology*, 7(1), pp.23–38.
- Carlton, J.M. et al., 2002. Genome sequence and comparative analysis of the model rodent malaria parasite Plasmodium yoelii yoelii. *Nature*, 419(6906), pp.512–519.
- Carrozza, M.J. et al., 2005. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell*, 123(4), pp.581–592.
- Carrozza, M.J. et al., 2003. The diverse functions of histone acetyltransferase complexes. *Trends in Genetics*, 19(6), pp.321–329.
- Chaal, B.K. et al., 2010. Histone deacetylases play a major role in the transcriptional regulation of the Plasmodium falciparum life cycle. *PLoS Pathogens*, 6(1), p.e1000737.
- Chan, J.-A., Fowkes, F.J.I. & Beeson, J.G., 2014. Surface antigens of Plasmodium falciparum-infected erythrocytes as immune targets and malaria vaccine candidates. *Cellular and Molecular life sciences : CMLSolecular life sciences : CMLS*, 71(19), pp.3633–57.
- Chen, Q. et al., 1998. Developmental selection of var gene expression in Plasmodium falciparum. *Nature*, 394(6691), pp.392–395.
- Chen, Q. et al., 1998. Identification of Plasmodium falciparum erythrocyte membrane protein 1 (PfEMP1) as the rosetting ligand of the malaria parasite P. falciparum. *The Journal of Experimental Medicine*, 187(1), pp.15–23.
- Chen, Q., Schlichtherle, M. & Wahlgren, M., 2000. Molecular aspects of severe malaria. *Clinical Microbiology Reviews*, 13(3), pp.439–450.
- Chen, Y. et al., 2011. Prediction of RNA Polymerase II recruitment, elongation, and stalling from histone modification data. *BMC Genomics*, 12(1), p.544.
- Cheng, Q. et al., 1998. stevor and rif are Plasmodium falciparum multicopy gene families which potentially encode variant antigens. *Molecular and Biochemical Parasitology*, 97(1–2), pp.161–176.
- Choi, S.W., Keyes, M.K. & Horrocks, P., 2006. LC/ESI-MS demonstrates the absence of 5-methyl-2'-deoxycytosine in Plasmodium falciparum genomic DNA. *Molecular and Biochemical Parasitology*, 150(2), pp.350–352.
- Chookajorn, T. et al., 2007. Epigenetic memory at malaria virulence genes. *Proceedings of the National Academy of Sciences of the United States of America*, 104(3), pp.899–902.
- Chookajorn, T., Ponsuwanna, P. & Cui, L., 2008. Mutually exclusive var gene expression in the malaria parasite: multiple layers of regulation. *Trends in Parasitology*, 24(10), pp.455–461.

- Claessens, A. et al., 2014. Generation of Antigenic Diversity in *Plasmodium falciparum* by Structured Rearrangement of Var Genes During Mitosis K. Deitsch, ed. *PLoS Genetics*, 10(12), p.e1004812.
- Clapier, C.R. & Cairns, B.R., 2009. The biology of chromatin remodeling complexes. *Annual Review of Biochemistry*, 78, pp.273–304.
- Coleman, B.I. & Duraisingh, M.T., 2008. Transcriptional control and gene silencing in *Plasmodium falciparum*. *Cellular Microbiology*, 10(10), pp.1935–1946.
- Corredor, V. et al., 2004. A SICAvr switching event in *Plasmodium knowlesi* is associated with the DNA rearrangement of conserved 3' non-coding sequences. *Molecular and Biochemical Parasitology*, 138(1), pp.37–49.
- Cortés, A. et al., 2007. Epigenetic silencing of *Plasmodium falciparum* genes linked to erythrocyte invasion. *PLoS Pathogens*, 3(8), pp.1023–1035.
- Côté, J. et al., 1994. Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SWI/SNF complex. *Science*, 265(5168), pp.53–60.
- Coutte, L. et al., 2009. Detailed analysis of sequence changes occurring during vlsE antigenic variation in the mouse model of *Borrelia burgdorferi* infection. *PLoS Pathogens*, 5(2), p.e1000293.
- Cowman, A.F., Berry, D. & Baum, J., 2012. The cellular and molecular basis for malaria parasite invasion of the human red blood cell. *Journal of Cell Biology*, 198(6), pp.961–971.
- Cox-Singh, J. et al., 2008. *Plasmodium knowlesi* malaria in humans is widely distributed and potentially life threatening. *Clinical Infectious Diseases*, 46(2), pp.165–171.
- Cross, G.A.M., Wirtz, L.E. & Navarro, M., 1998. Regulation of vsg expression site transcription and switching in *Trypanosoma brucei*. *Molecular and Biochemical Parasitology*, 91(1), pp.77–91.
- Cui, L. et al., 2008. Histone acetyltransferase inhibitor anacardic acid causes changes in global gene expression during in vitro *Plasmodium falciparum* development. *Eukaryotic Cell*, 7(7), pp.1200–1210.
- Cui, L. et al., 2007. PfGCN5-mediated histone H3 acetylation plays a key role in gene expression in *Plasmodium falciparum*. *Eukaryotic Cell*, 6(7), pp.1219–1227.
- Cui, L. & Miao, J., 2010. Chromatin-Mediated epigenetic regulation in the malaria parasite *Plasmodium falciparum*. *Eukaryotic Cell*, 9(8), pp.1138–1149.
- Cunningham, D. et al., 2009. Rapid changes in transcription profiles of the *Plasmodium yoelii* yir multigene family in clonal populations: Lack of epigenetic memory? *PLoS ONE*, 4(1), p.e4285.
- Cunningham, D. et al., 2010. *The pir multigene family of Plasmodium: Antigenic variation and beyond*,
- Deindl, S. et al., 2013. ISWI remodelers slide nucleosomes with coordinated multi-base-pair entry steps and single-base-pair exit steps. *Cell*, 152(3), pp.442–452.
- Deitsch, K.W., 2005. Malaria virulence genes: Controlling expression through chromatin modification. *Cell*, 121(1), pp.1–2.

- Deitsch, K.W., Moxon, E.R. & Wellems, T.E., 1997. Shared themes of antigenic variation and virulence in bacterial, protozoal, and fungal infections. *Microbiology and Molecular Biology Reviews*, 61(3), pp.281–293.
- Deitsch, K.W., Del Pinal, A. & Wellems, T.E., 1999. Intra-cluster recombination and var transcription switches in the antigenic variation of *Plasmodium falciparum*. *Molecular and Biochemical Parasitology*, 101(1–2), pp.107–116.
- Van Dijk, M.R. et al., 1997. Replication, expression and segregation of plasmid-borne DNA in genetically transformed malaria parasites. *Molecular and Biochemical Parasitology*, 86(2), pp.155–162.
- Dondorp, A.M. et al., 2009. Artemisinin Resistance in. *New England Journal of Medicine*, 361(30), pp.455–467.
- Doolan, D.L., Dobaño, C. & Baird, J.K., 2009. Acquired immunity to Malaria. *Clinical Microbiology Reviews*, 22(1), pp.13–36.
- Dresser, A.R., Hardy, P.O. & Chaconas, G., 2009. Investigation of the genes involved in antigenic switching at the *vlsE* locus in *Borrelia burgdorferi*: An essential role for the RuvAB branch migrase. *PLoS Pathogens*, 5(12), p.e1000680.
- Duffy, M.F. et al., 2009. Ectopic Recombination of a Malaria var Gene during Mitosis Associated with an Altered var Switch Rate. *Journal of Molecular Biology*, 389(3), pp.453–469.
- Duffy, M.F. et al., 2014. Epigenetic regulation of the *Plasmodium falciparum* genome. *Briefings in Functional Genomics*, 13(3), pp.203–216.
- Duffy, M.F. et al., 2012. The role of chromatin in *Plasmodium* gene expression. *Cellular Microbiology*, 14(6), pp.819–828.
- Duffy, M.F. et al., 2002. Transcription of multiple var genes by individual, trophozoite-stage *Plasmodium falciparum* cells expressing a chondroitin sulphate A binding phenotype. *Molecular Microbiology*, 43(5), pp.1285–1293.
- Duraisingh, M.T. et al., 2005. Heterochromatin silencing and locus repositioning linked to regulation of virulence genes in *Plasmodium falciparum*. *Cell*, 121(1), pp.13–24.
- Duraisingh, M.T., Triglia, T. & Cowman, A.F., 2002. Negative selection of *Plasmodium falciparum* reveals targeted gene deletion by double crossover recombination. *International Journal for Parasitology*, 32(1), pp.81–89.
- Dzikowski, R. et al., 2007. Mechanisms underlying the mutually exclusive expression of virulence genes by malaria parasites. *EMBO Reports*, 8(10), pp.959–965.
- Dzikowski, R., Frank, M. & Deitsch, K., 2006. Mutually exclusive expression of virulence genes by malaria parasites is regulated independently of antigen production. *PLoS Pathogens*, 2(3), pp.0184–0194.
- Ebbinghaus, P. & Krücken, J., 2011. Characterization and tissue-specific expression patterns of the *Plasmodium chabaudi* *cir* multigene family. *Malaria Journal*, 10(1), p.272.
- Engwerda, C.R., Beattie, L. & Amante, F.H., 2005. The importance of the spleen in malaria. *Trends in Parasitology*, 21(2), pp.75–80.
- Epp, C. et al., 2009. Chromatin associated sense and antisense noncoding RNAs are transcribed

- from the var gene family of virulence genes of the malaria parasite *Plasmodium falciparum*. *RNA*, 15(1), pp.116–27.
- Evertts, A.G. et al., 2013. H4K20 methylation regulates quiescence and chromatin compaction. *Molecular Biology of the Cell*, 24(19), pp.3025–37.
- Fan, Q., An, L. & Cui, L., 2004. PfADA2, a *Plasmodium falciparum* homologue of the transcriptional coactivator ADA2 and its in vivo association with the histone acetyltransferase PfGCN5. *Gene*, 336(2), pp.251–261.
- Fernandez, V. et al., 1998. Multiple adhesive phenotypes linked to rosetting binding of erythrocytes in *Plasmodium falciparum* malaria. *Infection and Immunity*, 66(6), pp.2969–2975.
- Fernandez, V. et al., 1999. Small, clonally variant antigens expressed on the surface of the *Plasmodium falciparum*-infected erythrocyte are encoded by the rif gene family and are the target of human immune responses. *The Journal of Experimental Medicine*, 190(10), pp.1393–404.
- Figtree, M. et al., 2010. *Plasmodium knowlesi* in human, Indonesian Borneo. *Emerging Infectious Diseases*, 16(4), pp.672–674.
- Fischer, K. et al., 1997. Expression of var genes located within polymorphic subtelomeric domains of *Plasmodium falciparum* chromosomes. *Molecular and Cellular Biology*, 17(7), pp.3679–86.
- Fischer, K. et al., 2003. Ten families of variant genes encoded in subtelomeric regions of multiple chromosomes of *Plasmodium chabaudi*, a malaria species that undergoes antigenic variation in the laboratory mouse. *Molecular Microbiology*, 48(5), pp.1209–1223.
- Flanagan, J.F. et al., 2005. Double chromodomains cooperate to recognize the methylated histone H3 tail. *Nature*, 438(7071), pp.1181–5.
- Flueck, C. et al., 2010. A major role for the *Plasmodium falciparum* ApiAP2 protein PfSIP2 in chromosome end biology. *PLoS Pathogens*, 6(2), p.e1000784.
- Flueck, C. et al., 2009. *Plasmodium falciparum* heterochromatin protein 1 marks genomic loci linked to phenotypic variation of exported virulence factors. *PLoS Pathogens*, 5(9).
- Fonager, J. et al., 2012. Reduced CD36-dependent tissue sequestration of *Plasmodium*-infected erythrocytes is detrimental to malaria parasite growth in vivo. *Journal of Experimental Medicine*, 209(1), pp.93–107.
- Frank, M. et al., 2008. Frequent recombination events generate diversity within the multi-copy variant antigen gene families of *Plasmodium falciparum*. *International Journal for Parasitology*, 38(10), pp.1099–1109.
- Franke-Fayard, B. et al., 2005. Murine malaria parasite sequestration: CD36 is the major receptor, but cerebral pathology is unlinked to sequestration. *Proceedings of the National Academy of Sciences of the United States of America*, 102(32), pp.11468–11473.
- Freitas, L.H. et al., 2005. Telomeric heterochromatin propagation and histone acetylation control mutually exclusive expression of antigenic variation genes in malaria parasites. *Cell*, 121(1), pp.25–36.
- Freitas-Junior, L.H. et al., 2000. Frequent ectopic recombination of virulence factor genes in

- telomeric chromosome clusters of *P. falciparum*. *Nature*, 407(6807), pp.1018–1022.
- Fukuda, H. et al., 2006. Simple histone acetylation plays a complex role in the regulation of gene expression. *Briefings in Functional Genomics and Proteomics*, 5(3), pp.190–208.
- Gannoun-Zaki, L. et al., 2005. A silenced *Plasmodium falciparum* var promoter can be activated in vivo through spontaneous deletion of a silencing element in the intron. *Eukaryotic Cell*, 4(2), pp.490–492.
- Garcia, B.A. et al., 2007. Organismal differences in post-translational modifications in histones H3 and H4. *Journal of Biological Chemistry*, 282(10), pp.7641–7655.
- Gardner, M.J. et al., 2013. Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature*, 419(6906), pp.1–14.
- Gissot, M. et al., 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(1), pp.29–42.
- Goel, S. et al., 2015. RIFINs are adhesins implicated in severe *Plasmodium falciparum* malaria. *Nature Medicine*, 21(4), pp.314–317.
- Grant, P.A. et al., 1997. Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: Characterization of an ada complex and the saga (spt/ada) complex. *Genes and Development*, 11(13), pp.1640–1650.
- Groth, A. et al., 2007. Chromatin Challenges during DNA Replication and Repair. *Cell*, 128(4), pp.721–733.
- Gupta, A.P. et al., 2013. Dynamic Epigenetic Regulation of Gene Expression during the Life Cycle of Malaria Parasite *Plasmodium falciparum*. *PLoS Pathogens*, 9(2), p.e1000737.
- Haber, J.E., 1998. A locus control region regulates yeast recombination. *Trends in Genetics*, 14(8), pp.317–321.
- Hall, N. et al., 2005. A Comprehensive Survey of the *Plasmodium* Life Cycle by Genomic , Transcriptomic ... *Science*, 307(5706), pp.82–86.
- Hall, N. & Carlton, J., 2005. Comparative genomics of malaria parasites. *Current Opinion in Genetics and Development*, 15(6), pp.609–613.
- Han, M. et al., 1988. Depletion of histone H4 and nucleosomes activates the PHO5 gene in *Saccharomyces cerevisiae*. *The EMBO Journal*, 7(7), pp.2221–8.
- Han, M. & Grunstein, M., 1988. Nucleosome loss activates yeast downstream promoters in vivo. *Cell*, 55(6), pp.1137–1145.
- Hebbes, T.R. et al., 1994. Core histone hyperacetylation co-maps with generalized DNase I sensitivity in the chicken beta-globin chromosomal domain. *The EMBO Journal*, 13(8), pp.1823–30.
- Hernandez-Rivas, R. et al., 1997. Expressed var genes are found in *Plasmodium falciparum* subtelomeric regions. *Molecular and Cellular Biology*, 17(2), pp.604–11.
- Hoeijmakers, W.A.M. et al., 2013. H2A.Z/H2B.Z double-variant nucleosomes inhabit the AT-rich promoter regions of the *Plasmodium falciparum* genome. *Molecular Microbiology*, 87(5), pp.1061–1073.

- Hong, L. et al., 1993. Studies of the DNA binding properties of histone H4 amino terminus. Thermal denaturation studies reveal that acetylation markedly reduces the binding constant of the H4 “tail” to DNA. *Journal of Biological Chemistry*, 268(1), pp.305–314.
- Hossain, M.J. et al., 2010. Plasmodium falciparum Tudor Staphylococcal Nuclease interacting proteins suggest its role in nuclear as well as splicing processes. *Gene*, 468(1–2), pp.48–57.
- Howitt, C.A. et al., 2009. Clonally variant gene families in Plasmodium falciparum share a common activation factor. *Molecular Microbiology*, 73(6), pp.1171–1185.
- Huang, Y., 2002. Transcriptional silencing in Saccharomyces cerevisiae and Schizosaccharomyces pombe. *Nucleic Acids Research*, 30(7), pp.1465–1482.
- Issar, N. et al., 2009. Differential sub-nuclear localisation of repressive and activating histone methyl modifications in P. falciparum. *Microbes and Infection*, 11(3), pp.403–407.
- Iyer, L.M. et al., 2008. Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes. *International Journal for Parasitology*, 38(1), pp.1–31.
- Janssen, C.S. et al., 2002. A large gene family for putative variant antigens shared by human and rodent malaria parasites. *Proceedings. Biological Sciences*, 269(1489), pp.431–6.
- Janssen, C.S. et al., 2004. Plasmodium interspersed repeats: The major multigene superfamily of malaria parasites. *Nucleic Acids Research*, 32(19), pp.5712–5720.
- Jensen, A.T.R. et al., 2004. Plasmodium falciparum associated with severe childhood malaria preferentially expresses PfEMP1 encoded by group A var genes. *Journal of Experimental Medicine*, 199(9), pp.1179–1190.
- Jiang, L. et al., 2013. PfSETVs methylation of histone H3K36 represses virulence genes in Plasmodium falciparum. *Nature*, 499(7457), pp.223–7.
- Joannin, N. et al., 2011. RSpred, a set of Hidden Markov Models to detect and classify the RIFIN and STEVOR proteins of Plasmodium falciparum. *BMC Genomics*, 12(1), p.119.
- Joannin, N. et al., 2008. Sub-grouping and sub-functionalization of the RIFIN multi-copy protein family. *BMC Genomics*, 9(January), p.19.
- Joergensen, L. et al., 2010. Surface co-expression of two different PfEMP1 antigens on single plasmodium falciparum-infected erythrocytes facilitates binding to ICAM1 and PECAM1. *PLoS Pathogens*, 6(9).
- Jones, M.K. & Good, M.F., 2006. Malaria parasites up close. *Nature Medicine*, 12(2), pp.170–171.
- Joshi, M.B. et al., 1999. Molecular cloning and nuclear localization of a histone deacetylase homologue in Plasmodium falciparum. *Molecular and Biochemical Parasitology*, 99(1), pp.11–19.
- Kaestli, M. et al., 2006. Virulence of malaria is associated with differential expression of Plasmodium falciparum var gene subgroups in a case-control study. *The Journal of infectious diseases*, 193(11), pp.1567–1574.
- Kafsack, B.F.C. et al., 2014. A transcriptional switch underlies commitment to sexual development in malaria parasites. *Nature*, 507(7491), pp.248–52.

- Karachentsev, D. et al., 2005. PR-Set7-dependent methylation of histone H4 Lys 20 functions in repression of gene expression and is essential for mitosis. *Genes and Development*, 19(4), pp.431–435.
- Karmodiya, K. et al., 2015. A comprehensive epigenome map of *Plasmodium falciparum* reveals unique mechanisms of transcriptional regulation and identifies H3K36me2 as a global mark of gene suppression. *Epigenetics and Chromatin*, 8(1), p.32.
- Karmodiya, K. et al., 2012. H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells. *BMC genomics*, 13(1), p.424.
- Kaviratne, M. et al., 2002. Small variant STEVOR antigen is uniquely located within Maurer's clefts in *Plasmodium falciparum*-infected red blood cells. *Eukaryotic Cell*, 1(6), pp.926–935.
- Kemp, D.J. et al., 1992. A chromosome 9 deletion in *Plasmodium falciparum* results in loss of cytoadherence. *Memorias do Instituto Oswaldo Cruz*, 87 Suppl 3, pp.85–89.
- Khattab, A. et al., 2008. *Plasmodium falciparum* variant STEVOR antigens are expressed in merozoites and possibly associated with erythrocyte invasion. *Malaria Journal*, 7(1), p.137.
- Khattab, A. & Meri, S., 2011. Exposure of the *Plasmodium falciparum* clonally variant STEVOR proteins on the merozoite surface. *Malaria Journal*, 10(1), p.58.
- Kim, A., Kiefer, C.M. & Dean, A., 2007. Distinctive signatures of histone methylation in transcribed coding and noncoding human beta-globin sequences. *Molecular and Cellular Biology*, 27(4), pp.1271–9.
- Knezetic, J.A. & Luse, D.S., 1986. The presence of nucleosomes on a DNA template prevents initiation by RNA polymerase II in vitro. *Cell*, 45(1), pp.95–104.
- Komaki-Yasuda, K. et al., 2013. Identification of a Novel and Unique Transcription Factor in the Intraerythrocytic Stage of *Plasmodium falciparum*. *PLoS ONE*, 8(9), p.e74701.
- Korir, C.C. & Galinski, M.R., 2006. Proteomic studies of *Plasmodium knowlesi* SICA variant antigens demonstrate their relationship with *P. falciparum* EMP1. *Infection, Genetics and Evolution*, 6(1), pp.75–79.
- Kouzarides, T., 2007. Chromatin Modifications and Their Function. *Cell*, 128(4), pp.693–705.
- Kraemer, S.M. et al., 2007. Patterns of gene recombination shape var gene repertoires in *Plasmodium falciparum*: comparisons of geographically diverse isolates. *BMC Genomics*, 8(1), p.45.
- Kraemer, S.M. & Smith, J.D., 2003. Evidence for the importance of genetic structuring to the structural and functional specialization of the *Plasmodium falciparum* var gene family. *Molecular Microbiology*, 50(5), pp.1527–1538.
- Kuo, M.H. et al., 1996. Transcription-linked acetylation by Gcn5p of histones H3 and H4 at specific lysines. *Nature*, 383(6597), pp.269–272.
- Kurdistani, S.K., Tavazoie, S. & Grunstein, M., 2004. Mapping global histone acetylation patterns to gene expression. *Cell*, 117(6), pp.721–733.
- Kyes, S. et al., 2007. *Plasmodium falciparum* var gene expression is developmentally controlled at the level of RNA polymerase II-mediated transcription initiation. *Molecular Microbiology*, 63(4), pp.1237–1247.

- Kyes, S.A. et al., 1999. Rifins: a second family of clonally variant proteins expressed on the surface of red cells infected with *Plasmodium falciparum*. *Proceedings of the National Academy of Sciences of the United States of America*, 96(16), pp.9333–8.
- Kyes, S.A., Kraemer, S.M. & Smith, J.D., 2007. Antigenic variation in *Plasmodium falciparum*: Gene organization and regulation of the var multigene family. *Eukaryotic Cell*, 6(9), pp.1511–1520.
- Kyes, S., Horrocks, P. & Newbold, C., 2001. Antigenic Variation at the Infected Red Cell Surface in Malaria. *Annual Review of Microbiology*, 55(1), pp.673–707.
- Langhorne, J. et al., 2008. Immunity to malaria: more questions than answers. *Nature Immunology*, 9(7), pp.725–732.
- Lavazec, C., Sanyal, S. & Templeton, T.J., 2007a. Expression switching in the stevor and Pfmc-2TM superfamilies in *Plasmodium falciparum*. *Molecular Microbiology*, 64(6), pp.1621–1634.
- Lavazec, C., Sanyal, S. & Templeton, T.J., 2007b. Expression switching in the stevor and Pfmc-2TM superfamilies in *Plasmodium falciparum*. *Molecular Microbiology*, 64(6), pp.1621–1634.
- Lavstsen, T. et al., 2003. Sub-grouping of *Plasmodium falciparum* 3D7 var genes based on sequence analysis of coding and non-coding regions. *Malaria Journal*, 2, p.27.
- Lawton, J. et al., 2012. Characterization and gene expression analysis of the cir multi-gene family of *Plasmodium chabaudi chabaudi* (AS). *BMC Genomics*, 13(1), p.125.
- Laybourn, P.J. & Kadonaga, J.T., 1991. Role of nucleosomal cores and histone H1 in regulation of transcription by RNA polymerase II. *Science*, 254(5029), pp.238–45.
- Lee, D.Y. et al., 1993. A positive role for histone acetylation in transcription factor access to nucleosomal DNA. *Cell*, 72(1), pp.73–84.
- Lee, J.S. & Shilatifard, A., 2007. A site to remember: H3K36 methylation a mark for histone deacetylation. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis*, 618(1–2), pp.130–134.
- Leech, J.H. et al., 1984. Identification of a strain-specific malarial antigen exposed on the surface of *Plasmodium falciparum*-infected erythrocytes. *The Journal of Experimental Medicine*, 159(6), pp.1567–75.
- Lenstra, T.L. et al., 2011. The Specificity and Topology of Chromatin Interaction Pathways in Yeast. *Molecular Cell*, 42(4), pp.536–549.
- Li, B., Carey, M. & Workman, J.L., 2007. The Role of Chromatin during Transcription. *Cell*, 128(4), pp.707–719.
- Li, X.Y. et al., 2008. High-resolution mapping of epigenetic modifications of the rice genome uncovers interplay between DNA methylation, histone methylation, and gene expression. *Plant Cell*, 20(2), pp.259–276.
- Liu, Y. et al., 2007. A Plant Homeodomain in Rag-2 that Binds Hypermethylated Lysine 4 of Histone H3 Is Necessary for Efficient Antigen-Receptor-Gene Rearrangement. *Immunity*, 27(4), pp.561–571.
- Lopez, F.J. et al., 2013. A new computational approach redefines the subtelomeric vir

- superfamily of *Plasmodium vivax*. *BMC Genomics*, 14(1), p.8.
- Lopez-Rubio, J.J. et al., 2007. 5' Flanking Region of Var Genes Nucleate Histone Modification Patterns Linked To Phenotypic Inheritance of Virulence Traits in Malaria Parasites. *Molecular Microbiology*, 66(6), pp.1296–1305.
- Lopez-Rubio, J.J., Mancio-Silva, L. & Scherf, A., 2009. Genome-wide Analysis of Heterochromatin Associates Clonally Variant Gene Regulation with Perinuclear Repressive Centers in Malaria Parasites. *Cell Host and Microbe*, 5(2), pp.179–190.
- Lorch, Y., LaPointe, J.W. & Kornberg, R.D., 1987. Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. *Cell*, 49(2), pp.203–210.
- Lu, X. et al., 2008. The effect of H3K79 dimethylation and H4K20 trimethylation on nucleosome and chromatin structure. *Nature Structural & Molecular Biology*, 15(10), pp.1122–1124.
- Luger, K. et al., 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, 389(6648), pp.251–60.
- Lygerou, Z. et al., 1994. The yeast BDF1 gene encodes a transcription factor involved in the expression of a broad class of genes including snRNAs. *Nucleic Acids Research*, 22(24), pp.5332–5340.
- MacPherson, G.G. et al., 1985. Human cerebral malaria. A quantitative ultrastructural analysis of parasitized erythrocyte sequestration. *The American Journal of Pathology*, 119(3), pp.385–401.
- Maier, A.G. et al., 2008. Exported Proteins Required for Virulence and Rigidity of *Plasmodium falciparum*-Infected Human Erythrocytes. *Cell*, 134(1), pp.48–61.
- Maier, A.G. et al., 2006. Negative selection using yeast cytosine deaminase/uracil phosphoribosyl transferase in *Plasmodium falciparum* for targeted gene deletion by double crossover recombination. *Molecular and Biochemical Parasitology*, 150(1), pp.118–121.
- Maltby, V.E. et al., 2012. Histone H3 lysine 36 methylation targets the Isw1b remodeling complex to chromatin. *Molecular and Cellular Biology*, 32(17), pp.3479–85.
- Mamoun, C. Ben et al., 2001. Co-ordinated programme of gene expression during asexual intraerythrocytic development of the human malaria parasite *Plasmodium falciparum* revealed by microarray analysis. *Molecular Microbiology*, 39(1), pp.26–36.
- Mbengue, A. et al., 2015. A molecular mechanism of artemisinin resistance in *Plasmodium falciparum* malaria. *Nature*, 520(7549), pp.683–687.
- McLean, S.A., Pearson, C.D. & Phillips, R.S., 1982. *Plasmodium chabaudi*: Antigenic variation during recrudescence parasitaemias in mice. *Experimental Parasitology*, 54(3), pp.296–302.
- McRobert, L. et al., 2004. Distinct trafficking and localization of STEVOR proteins in three stages of the *Plasmodium falciparum* life cycle. *Infection and Immunity*, 72(11), pp.6597–6602.
- Merrick, C.J. et al., 2010. The effect of *Plasmodium falciparum* Sir2a histone deacetylase on clonal and longitudinal variation in expression of the var family of virulence genes. *International Journal for Parasitology*, 40(1), pp.35–43.

- Miao, J. et al., 2006. The malaria parasite *Plasmodium falciparum* histones: Organization, expression, and acetylation. *Gene*, 369(1–2), pp.53–65.
- Miao, J. et al., 2010. The MYST family histone acetyltransferase regulates gene expression and cell cycle in malaria parasite *Plasmodium falciparum*. *Molecular Microbiology*, 78(4), pp.883–902.
- Michels, P.A., 1984. Activation of genes for variant surface glycoproteins in trypanosomes. *Oxford Surveys on Eukaryotic Genes*, 1, pp.145–68.
- Miller, L.H. et al., 2002. The pathogenic basis of malaria. *Nature*, 415(6872), pp.673–679.
- Mok, B.W. et al., 2008. Default pathway of var2csa switching and translational repression in *Plasmodium falciparum*. *PLoS ONE*, 3(4), p.e1982.
- Mok, S. et al., 2015. Supp Materials: Population transcriptomics of human malaria parasites reveals the mechanism of artemisinin resistance. *Science*, 347(6220), pp.431–435.
- Mota, M.M. et al., 2000. *Plasmodium chabaudi*-infected erythrocytes adhere to CD36 and bind to microvascular endothelial cells in an organ-specific way. *Infection and Immunity*, 68(7), pp.4135–4144.
- Mukherjee, P. et al., 2008. Structural insights into the *Plasmodium falciparum* histone deacetylase 1 (PfHDAC-1): A novel target for the development of antimalarial therapy. *Bioorganic and Medicinal Chemistry*, 16(9), pp.5254–5265.
- Mwakalinga, S.B. et al., 2012. Expression of a type B RIFIN in *Plasmodium falciparum* merozoites and gametes. *Malaria Journal*, 11(1), p.429.
- Narlikar, G.J., Sundaramoorthy, R. & Owen-Hughes, T., 2013. Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. *Cell*, 154(3), pp.490–503.
- Ng, H.H. et al., 2003. Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Molecular Cell*, 11(3), pp.709–719.
- Niang, M. et al., 2014. STEVOR is a *Plasmodium falciparum* erythrocyte binding protein that mediates merozoite invasion and rosetting. *Cell Host and Microbe*, 16(1), pp.81–93.
- Niang, M., Xue, Y.Y. & Preiser, P.R., 2009. The *Plasmodium falciparum* STEVOR multigene family mediates antigenic variation of the infected erythrocyte. *PLoS Pathogens*, 5(2), p.e1000680.
- Nishioka, K. et al., 2002. PR-Set7 is a nucleosome-specific methyltransferase that modifies lysine 20 of histone H4 and is associated with silent Chromatin. *Molecular Cell*, 9(6), pp.1201–1213.
- Noble, R. et al., 2013. The antigenic switching network of *Plasmodium falciparum* and its implications for the immuno-epidemiology of malaria. *eLife*, 2013(2).
- Noedl, H. et al., 2008. Evidence of Artemisinin-resistant malaria in Western Cambodia. *New England Journal of Medicine*, 359(24), pp.2619–2620.
- Noviyanti, R. et al., 2001. Multiple var gene transcripts are expressed in *Plasmodium falciparum* infected erythrocytes selected for adhesion. *Molecular and Biochemical Parasitology*, 114(2), pp.227–237.

- Nunes-Alves, C., 2015. Parasite physiology: RIFINs promote rosette formation during malaria. *Nature Reviews Microbiology*, 13(5), pp.250–250.
- O'Donnell, R.A. et al., 2002. A genetic screen for improved plasmid segregation reveals a role for Rep20 in the interaction of *Plasmodium falciparum* chromosomes. *EMBO Journal*, 21(5), pp.1231–1239.
- O'Donnell, R. a et al., 2001. An alteration in concatameric structure is associated with efficient segregation of plasmids in transfected *Plasmodium falciparum* parasites. *Nucleic Acids Research*, 29(3), pp.716–24.
- Oda, H. et al., 2009. Monomethylation of histone H4-lysine 20 is involved in chromosome structure and stability and is essential for mouse development. *Molecular and Cellular Biology*, 29(8), pp.2278–2295.
- Ogryzko, V. V et al., 1996. The transcriptional activators p300 and CBP are acetyltransferases. *Cell*, 87, pp.953–959.
- Otto, T.D. et al., 2014. A comprehensive evaluation of rodent malaria parasite genomes and gene expression. *BMC Biology*, 12(1), p.215.
- Pain, A. et al., 2008. The genome of the simian and human malaria parasite *Plasmodium knowlesi*. *Nature*, 455(7214), pp.799–803.
- Parseghian, M.H. & Hamkalo, B. a, 2001. A compendium of the histone H1 family of somatic subtypes: an elusive cast of characters and their characteristics. *Biochemistry and Cell Biology = Biochimie et Biologie Cellulaire*, 79(3), pp.289–304.
- Partnership, C.T., 2011. First Results of Phase 3 Trial of RTS,S/AS01 Malaria Vaccine in African Children. *New England Journal of Medicine*, 365(20), pp.1863–1875.
- Pasternak, N.D. & Dzikowski, R., 2009. PfEMP1: an antigen that plays a key role in the pathogenicity and immune evasion of the malaria parasite *Plasmodium falciparum*. *The International Journal of Biochemistry & Cell Biology*, 41(7), pp.1463–6.
- Pays, E. & Nolan, D.P., 1998. Expression and function of surface proteins in *Trypanosoma brucei*. *Molecular and Biochemical Parasitology*, 91(1), pp.3–36.
- Pérez-Toledo, K. et al., 2009. *Plasmodium falciparum* heterochromatin protein 1 binds to trimethylated histone 3 lysine 9 and is linked to mutually exclusive expression of var genes. *Nucleic Acids Research*, 37(8), pp.2596–2606.
- Petter, M. et al., 2013. H2A.Z and H2B.Z double-variant nucleosomes define intergenic regions and dynamically occupy var gene promoters in the malaria parasite *Plasmodium falciparum*. *Molecular Microbiology*, 87(6), pp.1167–1182.
- Petter, M. et al., 2007. Variant proteins of the *Plasmodium falciparum* RIFIN family show distinct subcellular localization and developmental expression patterns. *Molecular and Biochemical Parasitology*, 156(1), pp.51–61.
- Petter, M., Bonow, I. & Klinkert, M.Q., 2008. Diverse expression patterns of subgroups of the rif multigene family during *Plasmodium falciparum* gametocytogenesis J. G. Beeson, ed. *PLoS ONE*, 3(11), p.e3779.
- Pokholok, D.K. et al., 2005. Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell*, 122(4), pp.517–527.

- Pollack, Y., Kogan, N. & Golenser, J., 1991. Plasmodium falciparum: Evidence for a DNA methylation pattern. *Experimental Parasitology*, 72(4), pp.339–344.
- Pongponratn, E. et al., 1991. Microvascular sequestration of parasitized erythrocytes in human falciparum malaria: A pathological study. *American Journal of Tropical Medicine and Hygiene*, 44(2), pp.168–175.
- Ponts, N. et al., 2013. Genome-wide mapping of DNA methylation in the human malaria parasite plasmodium falciparum. *Cell Host and Microbe*, 14(6), pp.696–706.
- Ponts, N. et al., 2010. Nucleosome landscape and control of transcription in the human malaria parasite. *Genome Research*, 20(2), pp.228–238.
- del Portillo, H.A. et al., 2012. The role of the spleen in malaria. *Cellular Microbiology*, 14(3), pp.343–355.
- del Portillo, H. a et al., 2001. A superfamily of variant genes encoded in the subtelomeric region of Plasmodium vivax. *Nature*, 410(6830), pp.839–842.
- Portugal, S., Pierce, S.K. & Crompton, P.D., 2013. Young Lives Lost as B Cells Falter: What We Are Learning About Antibody Responses in Malaria. *The Journal of Immunology*, 190(7), pp.3039–3046.
- Pray-Grant, M.G. et al., 2005. Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation. *Nature*, 433(January), pp.434–438.
- Quan, T.K. & Hartzog, G.A., 2010. Histone H3K4 and K36 methylation, Chd1 and Rpd3S oppose the functions of Saccharomyces cerevisiae Spt4-Spt5 in transcription. *Genetics*, 184(2), pp.321–334.
- Radeva-Petrova, D. et al., 2014. Drugs for preventing malaria in pregnant women in endemic areas: any drug regimen versus placebo or no treatment. *The Cochrane Database of Systematic Reviews*, 10(10), p.CD000169.
- Raisner, R.M. & Madhani, H.D., 2006. Patterning chromatin: Form and function for H2A.Z variant nucleosomes. *Current Opinion in Genetics and Development*, 16(2), pp.119–124.
- Ralph, S.A., Scheidig-Benatar, C. & Scherf, A., 2005. Antigenic variation in Plasmodium falciparum is associated with movement of var loci between subnuclear locations. *Proceedings of the National Academy of Sciences of the United States of America*, 102(15), pp.5414–9.
- Ralph, S.A. & Scherf, A., 2005. The epigenetic control of antigenic variation in Plasmodium falciparum. *Current Opinion in Microbiology*, 8(4), pp.434–440.
- Recker, M. et al., 2011. Antigenic variation in Plasmodium falciparum malaria involves a highly structured switching pattern K. Kim, ed. *PLoS Pathogens*, 7(3), p.e1001306.
- Reeder, J.C. & Brown, G. V, 1996. Antigenic variation and immune evasion in Plasmodium falciparum malaria. *Immunology and Cell Biology*, 74(6), pp.546–554.
- Rine, J. et al., 1979. A suppressor of mating-type locus mutations in Saccharomyces cerevisiae: Evidence for and identification of cryptic mating-type loci. *Genetics*, 93(4), pp.877–901.
- Le Roch, K.G., 2003. Discovery of Gene Function by Expression Profiling of the Malaria Parasite Life Cycle. *Science*, 301(5639), pp.1503–1508.

- Rottmann, M. et al., 2006. Differential expression of var gene groups is associated with morbidity caused by *Plasmodium falciparum* infection in Tanzanian children. *Infection and Immunity*, 74(7), pp.3904–3911.
- Rovira-Graells, N. et al., 2015. Deciphering the principles that govern mutually exclusive expression of *Plasmodium falciparum* clag3 genes. *Nucleic Acids Research*, 43(17), pp.8243–8257.
- Rovira-Graells, N. et al., 2012. Transcriptional variation in the malaria parasite *Plasmodium falciparum*. *Genome Research*, 22(5), pp.925–938.
- Rowe, A. et al., 1995. *Plasmodium falciparum* rosetting is associated with malaria severity in Kenya. *Infection and Immunity*, 63(6), pp.2323–2326.
- Rowe, J. a et al., 1997. *P. falciparum* rosetting mediated by a parasite-variant erythrocyte membrane protein and complement-receptor 1. *Nature*, 388(6639), pp.292–295.
- RTSS Clinical Trials Partnership, 2015. Efficacy and safety of RTS , S / AS01 malaria vaccine with or without a booster dose in infants and children in Africa : final results of a phase 3 , individually randomised , controlled trial. *The Lancet*, 6736(15), pp.31–45.
- Rubio, J.P., Thompson, J.K. & Cowman, A.F., 1996. The var genes of *Plasmodium falciparum* are located in the subtelomeric region of most chromosomes. *The EMBO Journal*, 15(15), pp.4069–77.
- Sakaguchi, A. & Steward, R., 2007. Aberrant monomethylation of histone H4 lysine 20 activates the DNA damage checkpoint in *Drosophila melanogaster*. *Journal of Cell Biology*, 176(2), pp.155–162.
- Sakamoto, S. et al., 2012. E2A and CBP/p300 act in synergy to promote chromatin accessibility of the immunoglobulin kappa locus. *Journal of Immunology*, 188(11), pp.5547–5560.
- Salanti, A. et al., 2004. Evidence for the involvement of VAR2CSA in pregnancy-associated malaria. *The Journal of Experimental Medicine*, 200(9), pp.1197–1203.
- Salanti, A. et al., 2003. Selective upregulation of a single distinctly structured var gene in chondroitin sulphate A-adhering *Plasmodium falciparum* involved in pregnancy-associated malaria. *Molecular Microbiology*, 49(1), pp.179–191.
- Salcedo-Amaya, A.M. et al., 2009. Dynamic histone H3 epigenome marking during the intraerythrocytic cycle of *Plasmodium falciparum*. *Proceedings of the National Academy of Sciences of the United States of America*, 106(24), pp.9655–60.
- Sam-Yellowe, T.Y. et al., 2004. A *Plasmodium* gene family encoding maurer's cleft membrane proteins: Structural properties and expression profiling. *Genome Research*, 14(6), pp.1052–1059.
- Sanders, S.L. et al., 2004. Methylation of histone H4 lysine 20 controls recruitment of Crb2 to sites of DNA damage. *Cell*, 119(5), pp.603–614.
- Santos-Rosa, H. et al., 2002. Active genes are tri-methylated at K4 of histone H3. *Nature*, 419(6905), pp.407–411.
- Sanyal, S. et al., 2012. *Plasmodium falciparum* STEVOR proteins impact erythrocyte mechanical properties. *Blood*, 119(2), pp.e1–e8.
- Sargeant, T.J. et al., 2006. Lineage-specific expansion of proteins exported to erythrocytes in

- malaria parasites. *Genome Biology*, 7(2), p.R12.
- Sautel, C.F. et al., 2007. SET8-mediated methylations of histone H4 lysine 20 mark silent heterochromatic domains in apicomplexan genomes. *Molecular and Cellular Biology*, 27(16), pp.5711–24.
- Scherf, A. et al., 1998. Antigenic variation in malaria: In situ switching, relaxed and mutually exclusive transcription of var genes during intra-erythrocytic development in *Plasmodium falciparum*. *EMBO Journal*, 17(18), pp.5418–5426.
- Scherf, A., Figueiredo, L.M. & Freitas-Junior, L.H., 2001. Plasmodium telomeres: A pathogen's perspective. *Current Opinion in Microbiology*, 4(4), pp.409–414.
- Schiltz, R.L. et al., 1999. Overlapping but distinct patterns of histone acetylation by the human coactivators p300 and PCAF within nucleosomal substrates. *Journal of Biological Chemistry*, 274(3), pp.1189–1192.
- Schneider, R. et al., 2004. Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. *Nature Cell Biology*, 6(1), pp.73–77.
- Schofield, L. & Grau, G.E., 2005. Immunological processes in malaria pathogenesis. *Nature Reviews. Immunology*, 5(9), pp.722–735.
- Schotta, G. et al., 2004. A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes and Development*, 18(11), pp.1251–1262.
- Schübeler, D. et al., 2004. The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. *Genes and Development*, 18(11), pp.1263–1271.
- Sealy, L. & Chalkley, R., 1978. DNA associated with hyperacetylated histone is preferentially digested by DNase I. *Nucleic Acids Research*, 5(6), pp.1863–1876.
- Sharp, S. et al., 2006. Programmed transcription of the var gene family, but not of stevor, in *Plasmodium falciparum* gametocytes. *Eukaryotic Cell*, 5(8), pp.1206–1214.
- Sherman, I.W., Eda, S. & Winograd, E., 2003. Cytoadherence and sequestration in *Plasmodium falciparum*: Defining the ties that bind. *Microbes and Infection*, 5(10), pp.897–909.
- Shogren-Knaak, M. et al., 2006. Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science*, 311(5762), pp.844–7.
- Sierra-Miranda, M. et al., 2012. Two long non-coding RNAs generated from subtelomeric regions accumulate in a novel perinuclear compartment in *Plasmodium falciparum*. *Molecular and Biochemical Parasitology*, 185(1), pp.36–47.
- Silamut, K. et al., 1999. A quantitative analysis of the microvascular sequestration of malaria parasites in the human brain. *The American Journal of Pathology*, 155(2), pp.395–410.
- De Silva, E.K. et al., 2008. Specific DNA-binding by apicomplexan AP2 transcription factors. *Proceedings of the National Academy of Sciences of the United States of America*, 105(24), pp.8393–8.
- Sims, R.J. et al., 2005. Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains. *Journal of Biological Chemistry*, 280(51), pp.41789–41792.

- Sinha, A. et al., 2014. A cascade of DNA-binding proteins for sexual commitment and development in *Plasmodium*. *Nature*, 507(7491), pp.253–7.
- Smith, J.D. et al., 1995. Switches in expression of *plasmodium falciparum* var genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. *Cell*, 82(1), pp.101–110.
- Smolle, M. et al., 2012. Chromatin remodelers Isw1 and Chd1 maintain chromatin structure during transcription by preventing histone exchange. *Nature Structural and Molecular Biology*, 19(9), pp.884–892.
- Snow, R.W. et al., 1997. Relation between severe malaria morbidity in children and level of *Plasmodium falciparum* transmission in Africa. *Lancet*, 349(9066), pp.1650–1654.
- Snow, R.W. & Marsh, K., 1998. New insights into the epidemiology of malaria relevant for disease control. *British Medical Bulletin*, 54(2), pp.293–309.
- Spence, P.J. et al., 2013. Vector transmission regulates immune control of *Plasmodium* virulence. *Nature*, 498(7453), pp.228–31.
- Srivastava, S. et al., 2014. Histone H3K9 acetylation level modulates gene expression and may affect parasite growth in human malaria parasite *Plasmodium falciparum*. *FEBS Journal*, 281(23), pp.5265–5278.
- Stockdale, C. et al., 2008. Antigenic variation in *Trypanosoma brucei*: joining the DOTs. *PLoS Biology*, 6(7), pp.445–470.
- Stoenner, H.G., Dodd, T. & Larsen, C., 1982. Antigenic variation of *Borrelia hermsii*. *The Journal of Experimental Medicine*, 156(5), pp.1297–311.
- Straimer, J. et al., 2014. K13-propeller mutations confer artemisinin resistance in *Plasmodium falciparum* clinical isolates. *Science*, 2624(1985), pp.428–431.
- Su, X.Z. et al., 1995. The large diverse gene family var encodes proteins involved in cytoadherence and antigenic variation of *Plasmodium falciparum*-infected erythrocytes. *Cell*, 82(1), pp.89–100.
- Sullivan, W.J., Naguleswaran, A. & Angel, S.O., 2006. Histones and histone modifications in protozoan parasites. *Cellular Microbiology*, 8(12), pp.1850–1861.
- Swamy, L., Amulic, B. & Deitsch, K.W., 2011. *Plasmodium falciparum* var gene silencing is determined by cis DNA elements that form stable and heritable interactions. *Eukaryotic Cell*, 10(4), pp.530–539.
- Tamburini, B.A. & Tyler, J.K., 2005. Localized histone acetylation and deacetylation triggered by the homologous recombination pathway of double-strand DNA repair. *Molecular and Cellular Biology*, 25(12), pp.4903–13.
- Tamura, K. et al., 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, 30(12), pp.2725–2729.
- Tan, J. et al., 2016. A LAIR1 insertion generates broadly reactive antibodies against malaria variant antigens. *Nature*, 529(7584), pp.105–109.
- Taylor, H.M., Kyes, S.A. & Newbold, C.I., 2000. Var gene diversity in *Plasmodium falciparum* is generated by frequent recombination events. *Molecular and Biochemical Parasitology*, 110(2), pp.391–397.

- Templeton, T.J. et al., 2004. Comparative analysis of apicomplexa and genomic diversity in eukaryotes. *Genome Research*, 14(9), pp.1686–1695.
- Tham, W.H. et al., 2007. Identification of basic transcriptional elements required for rif gene transcription. *International Journal for Parasitology*, 37(6), pp.605–615.
- The RTS, S.C.T.P., 2012. The Phase 3 Trial of RTS,S/AS01 Malaria Vaccine in African Infants. *New England Journal of Medicine*, 365(24), p.367:2284-2295.
- Thompson, J.K. et al., 1997. The chromosomal organization of the *Plasmodium falciparum* var gene family is conserved. *Molecular and Biochemical Parasitology*, 87(1), pp.49–60.
- Tompa, R. & Madhani, H.D., 2007. Histone H3 lysine 36 methylation antagonizes silencing in *Saccharomyces cerevisiae* independently of the Rpd3S histone deacetylase complex. *Genetics*, 175(2), pp.585–593.
- Tonkin, C.J. et al., 2009. Sir2 paralogues cooperate to regulate virulence genes and antigenic variation in *Plasmodium falciparum*. *PLoS Biology*, 7(4), pp.0771–0788.
- Trelle, M.B. et al., 2009. Global histone analysis by mass spectrometry reveals a high content of acetylated lysine residues in the malaria parasite *Plasmodium falciparum*. *Journal of Proteome Research*, 8(7), pp.3439–3450.
- Trossaert, M. et al., 1991. Cytoadherence of *Plasmodium falciparum* and complications of malaria. *Dakar médical*, 36(2), pp.192–7.
- Turner, B.M., 2007. *Chromatin and Gene Regulation: Molecular Mechanisms in Epigenetics* 1 edition., Wiley-Blackwell.
- Ukaegbu, U.E. et al., 2014. Recruitment of PfSET2 by RNA Polymerase II to Variant Antigen Encoding Loci Contributes to Antigenic Variation in *P. falciparum* W. J. Sullivan, ed. *PLoS Pathogens*, 10(1), p.e1003854.
- Ukaegbu, U.E. & Deitsch, K.W., 2015. The Emerging Role for RNA Polymerase II in Regulating Virulence Gene Expression in Malaria Parasites L. J. Knoll, ed. *PLOS Pathogens*, 11(7), p.e1004926.
- Vázquez-Macías, A. et al., 2002. A distinct 5' flanking var gene region regulates *Plasmodium falciparum* variant erythrocyte surface antigen expression in placental malaria. *Molecular Microbiology*, 45(1), pp.155–167.
- Volz, J. et al., 2010. Potential epigenetic regulatory proteins localise to distinct nuclear sub-compartments in *Plasmodium falciparum*. *International Journal for Parasitology*, 40(1), pp.109–121.
- Volz, J.C. et al., 2012. PfSET10, a *Plasmodium falciparum* Methyltransferase, Maintains the Active var Gene in a Poised State during Parasite Division. *Cell Host and Microbe*, 11(1), pp.7–18.
- Voss, T.S. et al., 2006. A var gene promoter controls allelic exclusion of virulence genes in *Plasmodium falciparum* malaria. *Nature*, 439(7079), pp.1004–1008.
- Voss, T.S. et al., 2007. Alterations in local chromatin environment are involved in silencing and activation of subtelomeric var genes in *Plasmodium falciparum*. *Molecular Microbiology*, 66(1), pp.139–150.
- Voss, T.S. et al., 2000. Genomic distribution and functional characterisation of two distinct and

- conserved *Plasmodium falciparum* var gene 5' flanking sequences. *Molecular and Biochemical Parasitology*, 107(1), pp.103–115.
- Wang, C.W. et al., 2010. Identification of a major rif transcript common to gametocytes and sporozoites of *Plasmodium falciparum*. *Malaria Journal*, 9(1), p.147.
- Wang, P. et al., 2009. Global analysis of H3K4 methylation defines MLL family member targets and points to a role for MLL1-mediated H3K4 methylation in the regulation of transcriptional initiation by RNA polymerase II. *Molecular and Cellular Biology*, 29(22), pp.6074–6085.
- Wang, Z. et al., 2008. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nature Genetics*, 40(7), pp.897–903.
- Ward, C.P. et al., 1999. Analysis of *Plasmodium falciparum* PfEMP-1/var genes suggests that recombination rearranges constrained sequences. *Molecular and Biochemical Parasitology*, 102(1), pp.167–177.
- Warimwe, G.M. et al., 2009. *Plasmodium falciparum* var gene expression is modified by host immunity. *Proceedings of the National Academy of Sciences of the United States of America*, 106(51), pp.21801–21806.
- Waterkeyn, J.G., Crabb, B.S. & Cowman, A.F., 1999. Transfection of the human malaria parasite *Plasmodium falciparum*. *International Journal for Parasitology*, 29(6), pp.945–955.
- Weber, J.L., 1988. Interspersed repetitive DNA from *Plasmodium falciparum*. *Molecular and Biochemical Parasitology*, 29(2–3), pp.117–124.
- Weiss, G.E. et al., 2009. Atypical memory B cells are greatly expanded in individuals living in a malaria-endemic area. *Journal of Immunology*, 183(3), pp.2176–82.
- Welsh, R.M. & Fujinami, R.S., 2007. Pathogenic epitopes, heterologous immunity and vaccine design. *Nature Reviews. Microbiology*, 5(7), pp.555–563.
- Westenberger, S.J. et al., 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(1), p.610.
- WHO, 2015. World Malaria Report 2014 Summary. *WHO*.
- Wikramaratna, P.S. et al., 2013. The antigenic evolution of influenza: drift or thrift? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 368(1614), p.20120200.
- Winter, G. et al., 2005. SURFIN is a polymorphic antigen expressed on *Plasmodium falciparum* merozoites and infected erythrocytes. *The Journal of Experimental Medicine*, 201(11), pp.1853–63.
- Witmer, K. et al., 2012. Analysis of subtelomeric virulence gene families in *Plasmodium falciparum* by comparative transcriptional profiling. *Molecular Microbiology*, 84(2), pp.243–259.
- Wolffe, A.P. & Hayes, J.J., 1999. Chromatin disruption and modification. *Nucleic Acids Research*, 27(3), pp.711–720.
- Xu, C.-R. & Feeney, A.J., 2009. The epigenetic profile of Ig genes is dynamically regulated during B cell differentiation and is modulated by pre-B cell receptor signaling. *Journal of*

- Immunology*, 182(3), pp.1362–9.
- Yam, X.Y. et al., 2016. Characterization of the Plasmodium Interspersed Repeats (PIR) proteins of Plasmodium chabaudi indicates functional diversity. *Scientific Reports*, 6, p.23449.
- Yang, X.J. et al., 1996. A p300/CBP-associated factor that competes with the adenoviral oncoprotein E1A. *Nature*, 382, pp.319–324.
- Yipp, B.G. et al., 2003. Recombinant PfEMP1 peptide inhibits and reverses cytoadherence of clinical Plasmodium falciparum isolates in vivo. *Blood*, 101(1), pp.331–337.
- Zhang, J.R. & Norris, S.J., 1998. Genetic variation of the Borrelia burgdorferi gene vlsE involves cassette-specific, segmental gene conversion. *Infection and Immunity*, 66(8), pp.3698–3704.
- Zhang, Q. et al., 2011. A Critical role of perinuclear filamentous Actin in spatial repositioning and mutually exclusive expression of virulence genes in malaria parasites. *Cell Host and Microbe*, 10(5), pp.451–463.