

*IDENTIFICATION OF NOVEL
PLASMODIUM VIVAX BLOOD-STAGE
VACCINE TARGETS*



JESSICA BLYTHE HOSTETLER

Lucy Cavendish College
April 2016

This dissertation is submitted for the
degree of Doctor of Philosophy

ABSTRACT

Identification of novel Plasmodium vivax blood-stage vaccine targets

Jessica Blythe Hostetler

A vaccine targeting the illness-inducing blood stage of Plasmodium vivax is hindered by major gaps in our knowledge of P. vivax biology, including critical events during merozoite invasion of erythrocytes. Only a single receptor-ligand interaction is currently known, and natural human immune responses to P. vivax during and after infection, which could provide clues for how to stimulate a protective immune response, have been the subject of only limited study. This lack of understanding of both the molecular details of invasion and the immunological responses during infection correlates with a relatively limited repertoire of potential P. vivax antigens being considered for vaccine development. A comprehensive and systematic approach is needed to advance our understanding of P. vivax invasion and to identify additional vaccine candidates.

I first addressed this knowledge gap by investigating transcription in P. vivax clinical isolates at the schizont stage, just prior to erythrocyte invasion. I purified RNA from schizont-enriched P. vivax isolates obtained directly from Cambodian patients with P. vivax malaria and sequenced the RNA using a strand-specific approach. These RNA-Seq data revealed hundreds of additional genes transcribed at this stage compared to previous microarray studies, and uncovered novel gene transcripts that enabled me to improve and annotate over 300 gene models in the reference genome. Comparisons of RNA-Seq data between clinical isolates revealed that genes related to host invasion were among the most variably-expressed genes. The data provided a basis to prioritize a list of targets consistently expressed during the late blood stages.

I next produced a library of blood-stage P. vivax proteins to use in functional and immuno-epidemiological studies. I selected candidates predicted to localize to the merozoite surface or invasive secretory organelles based on the P. vivax literature, expression during the schizont stage, and homology to P. falciparum vaccine candidates. I successfully expressed 37/39 full-length P. vivax proteins in a mammalian expression system and investigated their function using high-throughput screening methods. I screened all expressed proteins for their ability to bind reticulocytes and normocytes using a flow cytometry-based assay. I screened the complete P. vivax library against an

existing library of 40 erythrocyte surface proteins using an assay designed to detect low-affinity cell surface interactions in order to identify specific erythrocyte receptors. I also performed an intra-P. vivax library screen to identify parasite protein-protein interactions, which established that the interaction between surface proteins P12 and P41 known in P. falciparum is conserved in P. vivax, and detected several novel parasite protein-protein interactions. I then used surface plasmon resonance to biochemically confirm the intra-library interactions.

Finally, I used this protein library to characterize immune responses in P. vivax-exposed individuals. I screened the library using IgG from P. vivax-exposed individuals from Cambodia, the Solomon Islands, and Papua New Guinea (in collaboration), each with distinct transmission dynamics, in one of the first large-scale immunoepidemiological screens of a panel of full-length P. vivax merozoite proteins. Nearly all proteins were immunogenic in all 3 settings, confirming their utility in global immuno-epidemiological studies. We detected age-dependent associations for 12/34 proteins and strong correlations with clinical protection for 3 proteins, including a hypothetical protein of which little is known. Together, these data identified, characterized, and prioritized novel P. vivax vaccine targets for future pre-clinical testing in ex vivo assays.

*To Nick,
my love.*

Thank you for supporting this adventure.

Let's have some more.

*To Oscar,
my son.*

You are magical.

Keep finding the moon.

*To Elsie,
my daughter.*

You are my best surprise.

Keep dancing to your own tune.

DECLARATION

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Acknowledgments, Appendix A, and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or is being concurrently submitted for any such degree, diploma, or other qualification at the University of Cambridge or any other University of similar institution. It does not exceed the prescribed 60,000-word limit (excluding bibliography, figures and appendices) for the Biology Degree Committee.

ACKNOWLEDGMENTS

Gratitude goes first and foremost to my mentors, Julian Rayner and Rick Fairhurst, without whom this work would not have been possible. I have learned more from each of them than I can possibly acknowledge here. This experience has been incredibly rewarding professionally, but even more so personally. It is a time in my life I will never forget, not the least because it overlapped with the adventure that is parenthood. Twice.

A complete list of contributors can be found in Appendix A, but I especially want to thank my thesis committee members, David Dunne and Gavin Wright, for their constant advice and support over the years. Cécile Wright-Crosnier, along with Madushi Wanaguru and Abigail Perrin, provided incredible initial mentorship and training on the HEK293E cell expression system; Cameron Bess, Susana Campino, and Matthew Jones gave generously of their time for training as I got started in the laboratory; Sumana Sharma, William Proto, Josefin Bartholdson, Francis Gallaway, Nicole Muller-Sienerth, and all members of the Wright, Rayner, Bilker and Kwiatkowski laboratories provided enormous support and help during my many UK visits. The Long lab was very supportive throughout the years with time, mentoring and space from Carole Long, Kazutoyo Miura, and Ababacar Diouf in particular. I was more than fortunate to work closely with Camila Franca and Ivo Mueller at the Walter and Eliza Hall Institute for much of the work from Chapter 5. Aaron Neal provided some sanity-saving figure help in the final thesis sprint. My deepest thanks goes to the many field site staff and study participants, without whom much of this work would not have been possible.

I'd like to recognize the scientists who gave me the push to pursue this PhD in the first place, including my Duke University mentors, John Mercer and Louise Roth, and all my many JCVI mentors, but especially John Glass and Bob Strausberg. In addition, I would like to thank my incredibly supportive family and friends. My Team Fairhurst and Malaria Programme lab mates are without equal. I feel so fortunate to have been "at home" in two countries, thanks to many dear friends including Susana Campino and Taane Clark, Georgina and Steve Dunkley, Stefano Mazzucco and Paola Ricciardi, Olivia Cook, Burcu Bronner-Anar, Ellen Bushell, Ana Rita Gomes, and Will Proto who hosted and/or chauffeured me many times while travelling back and forth to the UK. Special thanks to Lia Chappell for chauffeuring this thesis (along with lots of additional help over the years).

I cannot give enough gratitude to the many people who supported and cared for my family while I worked on this thesis, in particular my mother, Suzanne Bartlett, and my fantastic parents-in-law, Michael and Hallie Polt. My urban family repeatedly stepped in to help Nick with a meal or night of caretaking, including Laurel and Vaki Mawema, Chanaki Amaratunga, Brandy St. Laurent, Dave Lemen and Chrissy Monta, Madeleine Lemen, Gray Kimbrough and Holli Jones Kimbrough, Sandie and Hans Chen, Dawn Neuendorffer, Michael and Jenny Kilberg, and many more dear friends and neighbors. Thank you to the Bartlett, Hostetler, and Polt clans, and to my grandparents for their unconditional love. And finally, love and thanks to my siblings and their partners, Stephen and Peyton, Steph and Gabby, Mike, Savannah, Mitch, Cody, Lindsay and Brian, and especially to my parents and stepparents who have always supported me.

As with raising my children, it takes a village to write a thesis.

Thank you to my wonderful village.

TABLE OF CONTENTS

1 INTRODUCTION.....	1
1.1 <i>PLASMODIUM VIVAX</i> : A MAJOR AND NEGLECTED HUMAN PATHOGEN	1
1.1.1 <i>Global burden of malaria historically</i>	1
1.1.2 <i>P. vivax epidemiology</i>	3
1.1.3 <i>Plasmodium life cycle</i>	5
1.1.4 <i>P. vivax pathology</i>	9
1.1.5 <i>Current P. vivax treatment and control measures</i>	10
1.2 <i>P. FALCIPARUM</i> AND <i>P. VIVAX</i> GENOMICS	12
1.2.1 <i>P. falciparum and P. vivax whole genome sequencing</i>	12
1.2.2 <i>P. falciparum and P. vivax transcriptome studies</i>	15
1.3 ERYTHROCYTE INVASION	22
1.3.1 <i>Overview of invasion</i>	22
1.3.2 <i>P. falciparum erythrocyte invasion ligands</i>	26
1.3.3 <i>P. vivax reticulocyte invasion</i>	30
1.4 NATURAL IMMUNITY TO <i>PLASMODIUM</i> INFECTIONS.....	32
1.4.1 <i>Immunity development to P. falciparum</i>	33
1.4.2 <i>Immunity development to P. vivax</i>	35
1.5 APPROACHES TO A <i>PLASMODIUM</i> VACCINE DEVELOPMENT.....	37
1.5.1 <i>Vaccine targets at different stages of the life cycle</i>	37
1.5.2 <i>P. falciparum vaccine development</i>	38
1.5.3 <i>P. vivax blood-stage vaccine candidate: Duffy binding protein</i>	39
1.5.4 <i>Other blood-stage candidates</i>	40
1.5.5 <i>Future P. vivax vaccine development</i>	42
1.6 CHARACTERIZATION OF <i>PLASMODIUM</i> VACCINE ANTIGENS	42
1.6.1 <i>Producing Plasmodium vaccine antigens</i>	42
1.6.2 <i>High-throughput screening of vaccine antigens</i>	44
1.7 SPECIFIC AIMS	46
2 GENERAL METHODS	48
2.1 <i>P. VIVAX</i> SCHIZONT TRANSCRIPTOME SEQUENCING	48
2.1.1 <i>Schizont enrichment of P. falciparum samples for testing</i>	48
2.1.2 <i>Comparison of RNA extraction methods</i>	49

2.1.3	<i>Field isolate collection and enrichment for schizonts</i>	49
2.1.4	<i>RNA extraction of field isolates</i>	50
2.1.5	<i>cDNA synthesis and PCR</i>	51
2.1.6	<i>Strand-specific RNA library production</i>	51
2.2	<i>P. vivax RNA SEQUENCE ANALYSIS</i>	54
2.2.1	<i>Sequence mapping and quality control</i>	54
2.2.2	<i>RNA-Seq expression analysis</i>	56
2.2.3	<i>Expression variability between isolates</i>	57
2.3	<i>PRODUCTION OF RECOMBINANT P. vivax ECTODOMAIN LIBRARY</i>	58
2.3.1	<i>P. vivax candidate selection</i>	58
2.3.2	<i>Subcloning P. vivax recombinant library</i>	59
2.3.3	<i>P. vivax library expression in the HEK293E system</i>	61
2.4	<i>CONFIRMATION AND ASSESSMENT OF PROTEIN EXPRESSION</i>	62
2.4.1	<i>Confirmation of protein expression by ELISA</i>	62
2.4.2	<i>Normalisation of β-lactamase tagged membrane protein ectodomains</i>	63
2.4.3	<i>SDS-PAGE, Western blotting and NativePAGE</i>	63
2.5	<i>HIGH-THROUGHPUT FUNCTIONAL SCREENS</i>	63
2.5.1	<i>Erythrocyte/reticulocyte binding experiments by flow cytometry</i>	63
2.5.2	<i>Protein screens using AVEIXIS</i>	65
2.6	<i>BIOPHYSICAL ANALYSIS OF PROTEIN INTERACTIONS</i>	65
2.6.1	<i>Purification of 6-His-tagged membrane protein ectodomains</i>	65
2.6.2	<i>Surface plasmon resonance (SPR)</i>	66
2.7	<i>P. vivax SEROEPIDEMIOLOGY</i>	66
2.7.1	<i>Seroepidemiology of recombinant P. vivax proteins in Cambodian patients</i>	66
2.7.2	<i>Seroepidemiology of recombinant P. vivax proteins in Solomon Islander patients</i>	70
2.7.3	<i>Seroepidemiology of recombinant P. vivax proteins in Papua New Guinea patients</i>	72
2.8	<i>COMMONLY USED BUFFERS AND SOLUTIONS</i>	75
3	TRANSCRIPTOME PROFILING BEFORE INVASION	76
3.1	<i>INTRODUCTION</i>	76
3.1.1	<i>Benefits of this study</i>	80
3.1.2	<i>Objectives</i>	80

3.2 RESULTS	81
3.2.1 <i>High-quality RNA extracted from P. vivax clinical isolates</i>	81
3.2.2 <i>Library and mapping statistics</i>	83
3.2.3 <i>Assessing DNA contamination</i>	85
3.2.4 <i>Assessing asexual stage time point</i>	89
3.2.5 <i>Assessing gametocyte contamination</i>	92
3.2.6 <i>Using RNA-Seq data to improve the P. vivax reference genome</i>	93
3.2.7 <i>Comparing mapping to P. vivax Sal 1 and P. vivax P01 reference genomes</i> ...	93
3.2.8 <i>Comparing expression data between clinical isolates at genome scale</i>	95
3.2.9 <i>Comparing expression between 4 clinical isolates at the individual gene level</i>	97
3.2.10 <i>Assessing the impact of diversity and mapping on expression data</i>	101
3.3 DISCUSSION.....	102
3.3.1 <i>Limitations and future work</i>	106
3.4 CONCLUSION.....	107
4 P. VIVAX RECOMBINANT PROTEIN LIBRARY	108
4.1 INTRODUCTION.....	108
4.1.1 <i>Benefits of these studies</i>	115
4.1.2 <i>Objectives</i>	115
4.2 RESULTS	116
4.2.1 <i>P. vivax merozoite library candidate selection</i>	116
4.2.2 <i>P. vivax merozoite protein library expression in HEK293E cells</i>	121
4.2.3 <i>Erythrocyte binding experiments by flow cytometry</i>	124
4.2.4 <i>High-throughput protein interaction screens</i>	127
4.2.5 <i>Biophysical analysis with SPR</i>	135
4.3 DISCUSSION.....	141
4.3.1 <i>Limitations and future work</i>	144
4.4 CONCLUSION.....	145
5 IMMUNOEPIDEMIOLOGY OF P. VIVAX PROTEIN LIBRARY	147
5.1 INTRODUCTION.....	147
5.1.1 <i>Benefits of these studies</i>	152
5.1.2 <i>Objectives</i>	153
5.2 RESULTS	153

5.2.1 Antibody responses to <i>P. vivax</i> recombinant proteins in Cambodia	153
5.2.2 Antibody responses to <i>P. vivax</i> recombinant proteins in Solomon Islands and Papua New Guinea	163
5.3 DISCUSSION.....	175
5.3.1 Key findings in Cambodian plasma screens	175
5.3.2 Key findings in SI and PNG populations	177
5.3.3 Limitations and future work.....	179
5.4 CONCLUSION.....	180
6 DISCUSSION	181
6.1 KEY REMAINING CHALLENGES IN <i>P. VIVAX</i> RESEARCH FIELD	181
6.1.1 <i>In vitro</i> culture	181
6.1.2 Hypnozoites.....	182
6.1.3 <i>P. vivax</i> invasion	183
6.1.4 <i>P. vivax</i> infections of Duffy-negative individuals	185
6.1.5 <i>P. vivax</i> vaccine development	186
6.2 <i>P. VIVAX</i> RESEARCH SUMMARY	189
6.3 CONCLUSION.....	192
BIBLIOGRAPHY	194
APPENDIX A: LIST OF CONTRIBUTORS.....	224
APPENDIX B: SUPPLEMENTARY TABLES	227

INDEX OF FIGURES

FIGURE 1.1: <i>P. VIVAX</i> ENDEMICITY IN 2010	5
FIGURE 1.2: <i>P. VIVAX</i> LIFE CYCLE	7
FIGURE 1.3: INTRAERYTHROCYTIC DEVELOPMENT CYCLE (IDC) OVERVIEW FOR <i>P. FALCIPARUM</i>	18
FIGURE 1.4: <i>PLASMODIUM</i> MEROZOITE STRUCTURE.....	23
FIGURE 1.5: OVERVIEW OF <i>P. FALCIPARUM</i> INVASION OF ERYTHROCYTES	25
FIGURE 1.6: AVIDITY-BASED EXTRACELLULAR INTERACTION SCREEN (AVEXIS).....	46
FIGURE 3.1: SHORT TERM <i>EX VIVO</i> CULTURE OF <i>P. VIVAX</i> CLINICAL ISOLATES	78
FIGURE 3.2: RNA-SEQ ILLUMINA LIBRARY CONSTRUCTION	79
FIGURE 3.3: MINOR DNA CONTAMINATION AND/OR INCOMPLETELY SPLICED TRANSCRIPTS IN <i>P. VIVAX</i> RNA EXTRACTIONS	83
FIGURE 3.4: ILLUMINA READ ALIGNMENTS AGAINST THE <i>P. VIVAX</i> , <i>P. FALCIPARUM</i> , AND HUMAN REFERENCE GENOMES	85
FIGURE 3.5: SEQUENCE BREADTH DISTRIBUTIONS ACROSS THE <i>P. VIVAX</i> P01 REFERENCE GENOME IN EXONS, INTRONS, AND OTHER REGIONS	86
FIGURE 3.6: SEQUENCE DEPTH DISTRIBUTIONS ACROSS THE <i>P. VIVAX</i> P01 REFERENCE GENOME IN EXONS, INTRONS, AND OTHER REGIONS	86
FIGURE 3.7: GENE EXPRESSION IN <i>P. VIVAX</i> CLINICAL ISOLATES IS MOST SIMILAR TO THE SCHIZONT STAGE IN <i>P. FALCIPARUM</i>	90
FIGURE 3.8: <i>P. VIVAX</i> CLINICAL ISOLATES CORRELATE WITH THE EARLY SCHIZONT TIME POINT FROM <i>P. VIVAX</i> MICROARRAY DATA.....	91
FIGURE 3.9: <i>P. VIVAX</i> AND <i>P. BERGHEI</i> SCHIZONT RNA-SEQ DATA ARE HIGHLY CORRELATED	92
FIGURE 3.10: EXPRESSION IN <i>P. VIVAX</i> CLINICAL ISOLATES IS HIGHLY CORRELATED.....	94
FIGURE 3.11: FPKM DISTRIBUTIONS ARE SIMILAR BETWEEN <i>P. VIVAX</i> CLINICAL ISOLATES	96
FIGURE 3.12: THREE METHODS FOR RANKING EXPRESSION VARIABILITY IN <i>P. VIVAX</i> CLINICAL ISOLATES	99
FIGURE 3.13: VARIABLE EXPRESSION IN <i>P. VIVAX</i> CLINICAL ISOLATES	101
FIGURE 4.1: GPI-ANCHORED <i>P. VIVAX</i> PROTEINS.....	110
FIGURE 4.2: BAIT AND PREY PROTEIN CONSTRUCTS	112
FIGURE 4.3: HIGH-THROUGHPUT BEAD-BASED INTERACTION SCREENING METHOD	113
FIGURE 4.4: OVERVIEW OF AVEXIS	114
FIGURE 4.5: <i>P. VIVAX</i> RECOMBINANT PROTEIN EXPRESSION DETECTED BY ELISA	122

FIGURE 4.6: WESTERN BLOT ANALYSIS CONFIRMS EXPRESSION OF 34/37 <i>P. vivax</i> RECOMBINANT PROTEINS.	123
FIGURE 4.7: <i>P. vivax</i> RECOMBINANT LIBRARY EXPRESSION AND BEAD SATURATION ASSAY BY ELISA	125
FIGURE 4.8: ERYTHROCYTE AND RETICULOCYTE BINDING TO <i>P. vivax</i> RECOMBINANT PROTEINS	126
FIGURE 4.9: ERYTHROCYTE PREY NORMALIZATION ASSAY.....	129
FIGURE 4.10: AVEXIS BETWEEN <i>P. vivax</i> RECOMBINANT PROTEINS AND ERYTHROCYTE RECEPTOR LIBRARY.....	130
FIGURE 4.11: <i>P. vivax</i> RECOMBINANT PREY NORMALIZATION.....	132
FIGURE 4.12: AVEXIS REVEALS NOVEL INTERACTIONS INVOLVING <i>P. vivax</i> RECOMBINANT PROTEINS	133
FIGURE 4.13: REPLICATED INTRA-LIBRARY INTERACTIONS WITH AVEXIS	134
FIGURE 4.14: SEC FOR <i>P. vivax</i> P12 AND P41 AND MSP7.1.....	136
FIGURE 4.15: P12, P41, AND MSP7.1 MAY EXIST AS HOMODIMERS OR OLIGOMERS.....	137
FIGURE 4.16: QUANTIFICATION OF THE <i>P. vivax</i> P12-P41 INTERACTION AFFINITY BY SURFACE PLASMON RESONANCE	138
FIGURE 4.17: <i>P. vivax</i> P12 AND <i>P. vivax</i> P41 SHOW NO SELF-BINDING BY SURFACE PLASMON RESONANCE.....	139
FIGURE 4.18: SURFACE PLASMON RESONANCE CONFIRMS THE <i>P. vivax</i> MSP3.10-MSP7.1 INTERACTION	140
FIGURE 4.19: SURFACE PLASMON RESONANCE SUPPORTS A WEAK INTERACTION BETWEEN <i>P. vivax</i> P12 AND <i>P. vivax</i> PVX_110945	140
FIGURE 5.1: IMMUNOEPIDEMIOLOGICAL STUDY SITES.....	148
FIGURE 5.2: <i>PLASMODIUM</i> MALARIA CASES IN CAMBODIA FROM 2004 TO 2014.....	149
FIGURE 5.3: MALARIA INCIDENCE IN THE SOLOMON ISLANDS FROM 1969 TO 2011.....	151
FIGURE 5.4: TESTING CAMBODIAN PATIENT PLASMA AGAINST <i>P. vivax</i> RECOMBINANT PROTEINS	154
FIGURE 5.5: CAMBODIAN PATIENT PLASMA IGG REACTIVITY TO FULL-LENGTH <i>P. vivax</i> RECOMBINANT PROTEIN ECTODOMAINS	155
FIGURE 5.6: MULTIPLE <i>P. vivax</i> RECOMBINANT PROTEINS ARE IMMUNOREACTIVE AND CONTAIN CONFORMATIONAL EPITOPES	157
FIGURE 5.7: REACTIVITY IN INDIVIDUAL CAMBODIAN PLASMA SAMPLES	159
FIGURE 5.8: SEROPREVALENCE IN “ACUTE” CAMBODIAN PLASMA SAMPLES	160
FIGURE 5.9: IGG RESPONSES IN ACUTE AND CONVALESCENT CAMBODIAN PATIENT PLASMA SAMPLES (UNPAIRED).....	161

FIGURE 5.10: IGG RESPONSES IN ACUTE AND CONVALESCENT CAMBODIAN PATIENT PLASMA SAMPLES (PAIRED)	162
FIGURE 5.11: <i>P. vivax</i> RECOMBINANT PROTEINS ARE IMMUNOREACTIVE IN SI PATIENT PLASMA SAMPLES.....	166
FIGURE 5.12: SEROREACTIVITY IN SI.....	167
FIGURE 5.13: BREADTH OF ANTIGENS RECOGNIZED IN SI	170
FIGURE 5.14: AGE- AND INFECTION-ASSOCIATED INCREASES IN IGG IN SI	171
FIGURE 5.15: HIGH IGG RESPONSES TO PVX_081550, P12, AND P41 ARE ASSOCIATED WITH REDUCED INCIDENCE OF CLINICAL DISEASE	174
FIGURE 6.1: OVERVIEW OF EACH EXPERIMENTAL CHAPTER, WITH SUMMARIZED AIMS, APPROACHES, AND RESULTS.....	190

INDEX OF TABLES

TABLE 1.1: PRECLINICAL AND CLINICAL <i>PLASMODIUM</i> VACCINE CANDIDATES*	41
TABLE 2.1: EXPRESSION PLASMID BACKBONES	59
TABLE 2.2: PRIMERS	60
TABLE 2.3: PLASMA SAMPLES FROM CAMBODIAN PATIENTS WITH ACUTE VIVAX MALARIA	68
TABLE 2.4: BUFFERS, MEDIA AND SOLUTIONS	75
TABLE 3.1: PATIENT AND SAMPLE PROFILES FOR <i>P. VIVAX</i> CLINICAL ISOLATES	81
TABLE 3.2: RNA EXTRACTION RESULTS	82
TABLE 3.3: RNA EXTRACTION RESULTS AFTER DNA DIGESTION	82
TABLE 3.4: RNA-SEQ MAPPING STATISTICS TO THE <i>P. VIVAX</i> P01 GENOME	84
TABLE 3.5: EXON-TO-INTRON COVERAGE COMPARISON FOR 50 (~1% OF THE GENOME) LOWEST, MIDDLE AND HIGHEST COVERAGE GENES	88
TABLE 3.6: READS MAPPING TO <i>P. VIVAX</i> SAL 1 VS. <i>P. VIVAX</i> P01	93
TABLE 3.7: HOST AND INVASION GENES ENRICHED IN TOP 130 VARIABLY-EXPRESSED GENES FROM <i>P. VIVAX</i> SCHIZONT-STAGE CLINICAL ISOLATES	100
TABLE 4.1: <i>P. VIVAX</i> RECOMBINANT MEROZOITE PROTEINS	116
TABLE 5.1: PILOT CAMBODIAN SEROPOSITIVITY SUMMARY	156
TABLE 5.2: <i>P. VIVAX</i> RECOMBINANT PROTEINS USED IN SI AND PNG SCREENS	164
TABLE 5.3: SEROREACTIVITY IN SI COMPREHENSIVE SCREEN	168
TABLE 5.4: SEROREACTIVITY IN SI AND CAMBODIAN PARASITEMIC PLASMA SAMPLES	168
TABLE 5.5: <i>P</i> VALUES FROM ANOVA FOR SI COMPREHENSIVE SCREEN	172
TABLE 5.6: ASSOCIATION BETWEEN LEVELS OF IGG TO <i>P. VIVAX</i> MEROZOITE PROTEINS AND PROTECTION AGAINST CLINICAL MALARIA IN PNG CHILDREN*	174
SUPPLEMENTARY TABLE A: <i>P. VIVAX</i> SAL 1 REFERENCE ANNOTATION CHANGES BASED ON RNA-SEQ DATA	227
SUPPLEMENTARY TABLE B: VARIABLY-EXPRESSED GENES IN <i>P. VIVAX</i> CLINICAL ISOLATES	236
SUPPLEMENTARY TABLE C: TOP-EXPRESSED GENES UNIQUE TO RNA-SEQ DATA	242
SUPPLEMENTARY TABLE D: <i>P. VIVAX</i> RECOMBINANT PROTEIN ECTODOMAINS	246

LIST OF ABBREVIATIONS

6-cys	Six-cysteine
ACT	Artemisinin combination therapy
AMA1	Apical membrane antigen 1
Amp	Ampicillin
AVEXIS	Avidity-based extracellular interaction screening
BSG	Basigin
BLAST	Basic local alignment search tool
BSA	Bovine serum albumin
Ca ²⁺	Calcium
COMP	Cartilage oligomeric matrix protein
CV	Column volume
DARC	Duffy antigen receptor for chemokines
DBL	Duffy binding-like
DBP	Duffy-binding protein
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DRM	Detergent-resistant membrane
DTT	Dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
EBA	Erythrocyte-binding antigen
EBL	Erythrocyte binding-like
EDTA	Ethylenediaminetetraacetic acid
EGF	Epithelial growth factor
ELISA	Enzyme-linked immunosorbent assay
FBS	Fetal bovine serum
G6PD	Glucose-6-phosphate dehydrogenase
GMEP	Global Malaria Eradication Programme
GPI	Glycosylphosphatidylinositol
HBS	HEPES-buffered saline
HEK293	Human embryonic kidney 293
6-His	Hexa-histidine
HI-FBS	Heat-inactivated fetal bovine serum
HRP	Horseradish peroxidase
IDC	Intraerythrocytic development cycle
IgG	Immunoglobulin G
IRS	Indoor residual spraying
LB	Lysogeny broth/Luria broth
LLIN	Long-lasting insecticide-treated net
LM	Light microscopy

LMVR	Laboratory of Malaria and Vector Research
Mg ²⁺	Magnesium
MolFOB	Molecular force of blood-stage exposure
MSP	Merozoite surface protein
MTRAP	Merozoite thrombospondin-related anonymous protein
MW	Molecular weight
NAI	Naturally-acquired immunity
NCBI	National Center for Biotechnology Information
NEB	New England Biolabs
NHS	National Health Service
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PEI	Polyethylenimine
Pf	<i>Plasmodium falciparum</i>
Pv	<i>Plasmodium vivax</i>
PV	Parasitophorous vacuole
RBL	Reticulocyte binding-like
RCF	Relative centrifugal force
RDT	Rapid diagnostic test
RH	Reticulocyte-binding protein homologue
RNA	Ribonucleic acid
RON	Rhoptry neck protein
RPM	Revolution per minute
RPMI media	Roswell Park Memorial Institute media
RT	Room temperature
RU	Response units
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SNP	single-nucleotide polymorphism
SOC	super optimal broth with catabolite repression
SPR	Surface plasmon resonance
TBE	Tris-borate-EDTA
V	Volts
WHO	World Health Organization
WTSI	Wellcome Trust Sanger Institute

One letter code for amino acids

A	Alanine
C	Cysteine
D	Aspartate
E	Glutamate
F	Phenylalanine

G	Glycine
H	Histidine
I	Isoleucine
K	Lysine
L	Leucine
M	Methionine
N	Asparagine
P	Proline
Q	Glutamine
R	Arginine
S	Serine
T	Threonine
V	Valine
W	Tryptophan
Y	Tyrosine

1 INTRODUCTION

1.1 *Plasmodium vivax*: a major and neglected human pathogen

Plasmodium vivax malaria afflicts millions of people throughout the world each year, particularly in Asia, the Pacific, and Latin America, and continues to be recognized as a neglected tropical disease. The global burden of *P. vivax* costs an estimated US\$1,400,000,000–\$4,000,000,000 per year (Price et al., 2007), yet *P. vivax* research receives only a small fraction of the overall malaria research and development funding, estimated at around 3% of the total malaria research funds from 2006-2009 (PATH, 2011). While there is widespread agreement that an effective vaccine would be a critical component of any global malaria eradication effort, development of a *P. vivax* vaccine has been hindered by a lack of understanding of basic *P. vivax* biology. This thesis aims to begin filling in this significant knowledge gap by identifying novel *P. vivax* vaccine candidates.

1.1.1 Global burden of malaria historically

Malaria is caused by infection from single-celled eukaryotic parasites from the genus *Plasmodium*. A wide variety of vertebrate species from mammals, birds, and reptiles are naturally infected with *Plasmodium* parasites, and there are hundreds of *Plasmodium* species described to date (Levine, 1988). Six species are known to cause human malaria: *P. falciparum*, *P. vivax*, two distinct species of *P. ovale* (Sutherland et al., 2010), *P. malariae*, and *P. knowlesi*, with the latter primarily infecting Southeast Asian macaques, but increasingly detected as a zoonotic source of human infections (Millar and Cox-

Singh, 2015, Singh et al., 2004). However, *P. falciparum* and *P. vivax* cause the vast majority of infections, and will be the focus of this introduction.

The transmission dynamics of *P. falciparum* and *P. vivax* have changed significantly over the last century. The essential role of mosquitoes to the *Plasmodium* life cycle (Figure 1.2) was discovered by Ronald Ross in 1897 and led to significant vector control measures, such as the draining of swamps and the widespread use of insecticides. Cheap and effective drugs, with chloroquine approved for use in 1946, being the exemplar, added a further major weapon in malaria control. Chloroquine, the insecticide DDT, and active case surveillance together underpinned a major world eradication campaign from 1955-1964 by the World Health Organization (WHO) termed the “Global Malaria Eradication Programme” (GMEP), which completely eliminated malaria from some areas (WHO, 1999). Measures such as these have ultimately eliminated malaria in over 50 countries since 1900, shrinking the area for risk of malaria infection by half and saving countless lives (Hay et al., 2004).

While some countries maintained gains and approached elimination, dozens of countries experienced a resurgence in transmission due to a combination of factors (WHO, 1999, Smith et al., 2013). In particular, and as detailed further below, the development of drug resistance in parasites and insecticide resistance in mosquitoes has consistently plagued malaria control efforts worldwide. These factors, coupled with insufficient funding and chronic health infrastructure problems, contributed to the failure of the GMEP, which was discontinued in favor of control programs. Global funding for malaria control declined in the 1970s and 1980s, due in part to the perceived failure of the GMEP and the global economic crises.

The funding landscape has improved significantly over the last 10 years as malaria has, once again, found a place on the global agenda. The Bill and Melinda Gates Foundation committed US\$2.5 billion in basic research and control measures since 2000, and the US President’s Malaria Initiative (PMI) started in 2005 distributes hundreds of millions in control-related funding each year (US\$618 million in 2014). Additional commitments by the Global Fund and World Bank have further contributed to the improvement in funding for both research and control measures, which now nears US\$2 billion annually (Greenwood and Targett, 2011). Malaria mortality declined 32% from 2004-2010 in line with the surge in funding (Murray et al., 2012). A variety of factors contributed to this

decline, including increased use of insecticide-treated bed nets (ITNs), changing transmission dynamics, the use of highly effective treatments such as artemisinin combination therapy (ACT), and improved health infrastructure (Murray et al., 2012). However, recent gains are under threat from the emergence of artemisinin-resistant *P. falciparum* parasites in Southeast Asia and the rise in insecticide resistance in *Plasmodium*-spreading mosquito populations (Greenwood and Targett, 2011). *P. vivax* malaria is also proving difficult to control with initial evidence suggesting that as *P. falciparum* transmission declines, *P. vivax* often assumes a more dominant role in disease incidence (Maude et al., 2014). This is supported by previous eradication campaigns where *P. vivax* transmission persisted long after *P. falciparum* was eliminated or reduced (Yekutieli, 1980). These challenges underline the need for an efficacious vaccine targeting both *P. vivax* and *P. falciparum* to ultimately control this global disease threat.

1.1.2 *P. vivax* epidemiology

At present, more than 200 million cases of malaria occur annually, and up to 45% of the world population is at risk of *Plasmodium* infection (WHO, 2014). *P. falciparum* causes the vast majority of malaria mortality with over 500,000 deaths each year. Over 90% of these deaths occur in Africa, and the majority of deaths occur in children under 5 years of age, although the age spectrum of mortality is changing with changing transmission patterns (WHO, 2014). Malaria deaths disproportionately affect populations in extreme poverty (those living on less than US\$1.25 per day) (WHO, 2014). *P. vivax* malaria, in contrast, has been described as “benign” because it causes relatively few deaths, although a growing body of work is challenging that description due to the large economic and social repercussions of *P. vivax* malaria morbidity, as well as evidence of severe cases (Kochar et al., 2005, Williams et al., 1997, Sharma and Khanduri, 2009, Beg et al., 2002, Kumar et al., 2007, Price et al., 2009). Recent reviews of case reports showed that while *P. vivax* causes far fewer cases of severe disease than *P. falciparum*, the outcomes of severe *P. vivax* and *P. falciparum* cases are similarly poor (Rogerson and Carter, 2008, Tjitra et al., 2008, Nurleila et al., 2012).

P. vivax causes the majority of malaria cases outside Africa; approximately 2.5 billion people are at risk of *P. vivax* infection annually, 80% of whom live in South and Southeast Asia (Gething et al., 2012). Illness caused by *P. vivax* is estimated to affect 80–390 million people each year throughout the tropics, primarily in Southeast Asia and the

Western Pacific (Price et al., 2007, Mendis et al., 2001, Hay et al., 2004). *P. vivax* has a much larger geographical spread than *P. falciparum*, affecting 95 countries across 4 continents. This is partly due to *P. vivax* parasites' ability to survive more readily in seasonal and temperate transmission environments; specifically their dormant liver forms (described below) act as a reservoir, enabling *P. vivax* to relapse into the bloodstream months or even years after the initial infection (Gething et al., 2012, Krotoski et al., 1982). *P. vivax* also has a lower temperature limit in its mosquito host (14.5°C) compared to *P. falciparum* (16°C), contributing to *P. vivax*'s expanded range into more temperate climates (Sinden, 2002, Guerra et al., 2006).

P. vivax maintains stable transmission in areas with much lower endemicity than *P. falciparum*, and *P. vivax* infection prevalence rates rarely exceed 7% (Figure 1.1). This is in stark contrast to *P. falciparum*, which can have prevalence rates as high as 70% across parts of Africa (Gething et al., 2011). It is important to note that most surveys employed microscopy or RDT to determine *P. vivax* malaria incidence, which may significantly underestimate the *P. vivax* burden due to the generally lower parasite densities achieved by this species (Mueller et al., 2009b, Gething et al., 2012). *P. vivax* is noticeably absent throughout most of Africa due to the high rate of the Duffy negative allele, a blood group central to *P. vivax* infection, although recent data from Madagascar and other African countries suggest that the barrier in Duffy-negative individuals is not as absolute as previously thought (Menard et al., 2010, Ngassa Mbenda and Das, 2014, Woldearegai et al., 2013). Nevertheless, selective pressure from *P. vivax* malaria appears to have driven this polymorphism to fixation in the majority of African populations. *P. falciparum* malaria has similarly selected for several other erythrocyte polymorphisms that protect against this disease, such as sickle-cell trait (hemoglobin S, HbS), hemoglobin C (HbC), and α -thalassemia [reviewed in (Taylor et al., 2013)].

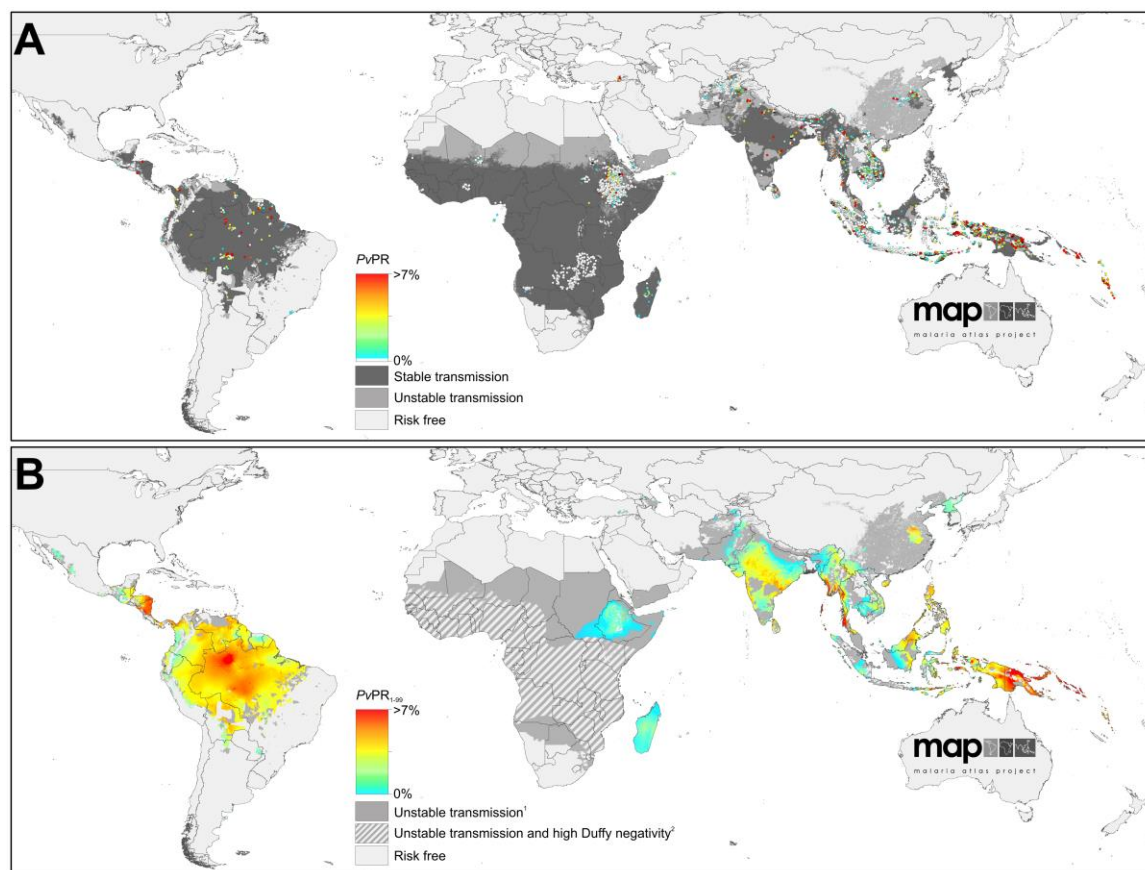


Figure 1.1: *P. vivax* endemicity in 2010

(A) Spatial limits of *P. vivax* malaria based on temperature and aridity data, as well as community surveys (from 1985 to 2010) for annual *P. vivax* parasite incidence (PvAPI). Individual survey data are presented as dots with a color continuum from green to red for a > 0% to > 7% *P. vivax* parasite rate (PvPR*) and white for surveys with zero *P. vivax* cases (see map legend). Transmission areas were defined as stable (dark grey, where PvAPI \geq 0.1 per 1,000 pa), unstable (medium grey, where PvAPI < 0.1 per 1,000 pa), or having no risk (light grey, where PvAPI = 0 per 1,000 pa). (B) Model-based geostatistics for PvPR, for people aged 1-99 years using the color scale from (A). Hatched areas represent areas with Duffy negativity gene frequency exceeding 90%. Reprinted from (Gething et al., 2012) under the Creative Commons Attribution (CC BY) license. *PvPR, the proportion of randomly sampled individuals in a surveyed population with patent parasitemia in their peripheral blood, as detected by microscopy or rapid diagnostic test (RDT)

1.1.3 *Plasmodium* life cycle

Plasmodium species share similar life cycles and develop through several morphologically distinct stages across their vertebrate and mosquito hosts (Figure 1.2) (Mueller et al., 2009a). When an infected female *Anopheles* mosquito takes a blood meal from a human, haploid, motile sporozoites in the mosquito saliva are injected into the skin (Prudencio et al., 2006). The sporozoites burrow through the vascular tissue and

enter the bloodstream about 15 minutes later (Vanderberg and Frevert, 2004). Those parasites that reach the liver and traverse the sinusoidal cell layer to infect hepatocytes, remain largely undetected by the human immune system (Tavares et al., 2013). Once inside hepatocytes, the parasites differentiate and proliferate for 2 days to 3 weeks (depending on the species) into multi-nucleated hepatic schizonts, which each contain many thousands of infectious forms called merozoites packed into vesicles called merosomes (Tarun et al., 2006, Prudencio et al., 2006). The infected hepatocyte detaches, and merosomes begin budding from it into the bloodstream, where they ultimately rupture (Sturm et al., 2006). The infective merozoites are then released, whereupon they attach to and invade erythrocytes (often of specific maturities) in an extremely rapid process that only takes about 1 minute (Weiss et al., 2015). This pattern of multiplication in the liver does not universally occur; in the cases of *P. vivax* and *P. ovale*, for example, some liver-stage parasites convert to dormant forms called hypnozoites, which can remain in situ for months or even years (Krotoski, 1985). These dormant forms are one of the unique challenges of *P. vivax* control, as they enable disease to relapse even in the absence of active transmission.

Upon erythrocyte invasion parasites are described as “ring-stage” (~0-24 hours for *P. falciparum* and *P. vivax*), so called for their appearance when stained with Giemsa and viewed by light microscopy (LM). This is followed by the most metabolically active trophozoite stage, in which the parasite digests host erythrocyte hemoglobin, resulting in crystals of hemozoin in the parasite food vacuole. Giemsa-stained trophozoites appear dark and circular by LM (~24-36 hours for *P. falciparum* and *P. vivax*). The parasites then undergo replication and segmentation into schizonts containing multiple nuclei (~36-48 hours for *P. falciparum* and *P. vivax*). A single schizont can contain 16-32 daughter merozoites in a range of sizes depending on the species. Upon fully maturing, proteases break down the parasitophorous vacuole membrane followed by the host erythrocyte membrane resulting in the daughter cells’ “egress” into the bloodstream (Blackman, 2008, Blackman and Carruthers, 2013), where they invade new erythrocytes (Gilson and Crabb, 2009). This cycle can continue indefinitely unless the infection is treated or cleared by the host’s immune system.

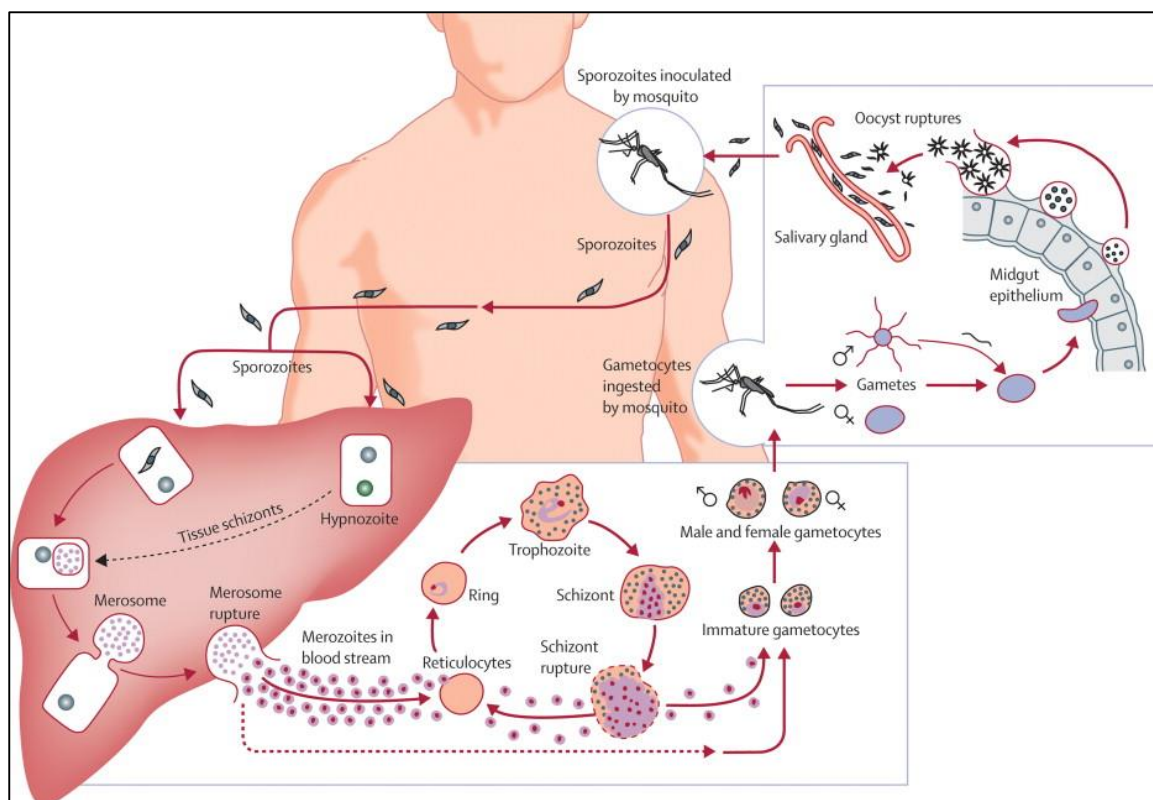


Figure 1.2: *P. vivax* life cycle

P. vivax sporozoites injected during the biting of a female *Anopheles* mosquito, enter the bloodstream and infect hepatocytes, a portion of which lie dormant as hypnozoites. Other parasites mature into schizonts that ultimately rupture and release infective merozoites into the bloodstream. Merozoites attach to and invade reticulocytes, in which they undergo a 48-hour asexual replication cycle to produce daughter merozoites, which subsequently invade new reticulocytes and cause the clinical symptoms of malaria. Some infected cells differentiate into gametocytes, which are ingested by mosquitoes and undergo fertilization in the mosquito midgut. Ookinetes traverse the midgut wall, develop into oocysts, and release sporozoites, which migrate to the salivary glands. Reprinted from (Mueller et al., 2009a), © 2009, with permission from Elsevier.

To complete the life cycle, a small proportion of parasites commit to sexual forms through the expression of transcription factor AP2-G (Sinha et al., 2014, Kafsack et al., 2014). All daughter merozoites from these committed cells will all go on to differentiate into either male (micro-) or female (macro-) gametocytes upon invasion of the next erythrocyte. After 3-6 days of sexual development, mature gametocytes then circulate in the bloodstream, to be ingested by mosquitoes taking a blood meal (Bousema and Drakeley, 2011). Once inside the mosquito midgut, several cues trigger gametocyte ‘activation’ including a drop in temperature to around 18°C and the presence of mosquito-made xanthurenic acid (Billker et al., 1998), resulting in male gametocyte

exflagellation. The motile male and immotile female gametes fuse to form diploid zygotes that mature into ookinetes, a process during which genetic recombination can occur. Ookinetes then traverse the midgut wall and develop into oocysts, which divide into thousands of sporozoites (Rosenberg and Rungsiwongse, 1991). Upon release, the haploid, motile sporozoites migrate through mosquito tissues to the salivary glands, ending a ~2-week process inside mosquitoes, after which the *Plasmodium* parasites can infect new human hosts.

The life cycle is similar for all mammalian *Plasmodium* parasites but with some crucial differences in human-infecting malaria parasites. The period of asexual replication in human blood for *P. vivax*, *P. falciparum*, and *P. ovale* follows a cycle of 48 hours, while this period is 24 hours and 72 hours for *P. knowlesi* and *P. malariae*, respectively. As described above, *P. vivax* and *P. ovale* can cause relapsing malaria through the activation of hypnozoites lying dormant in hepatocytes months to years after initial infection (Krotoski, 1985), which has significant implications for treatment. Both of these relapsing species, as well as initial *P. knowlesi* infections, also display a preference for invading reticulocytes (Mons et al., 1988, Kitchen, 1938, Lim et al., 2013, Collins and Jeffery, 2005). Since reticulocytes account for approximately 1% of the overall erythrocyte count, *P. vivax* achieves much lower parasitemias than *P. falciparum*, which invades erythrocytes of all ages. This host cellular tropism has made *P. vivax* extremely difficult to maintain in continuous *in vitro* culture, as there is no readily available source for reticulocyte-enriched blood; this is a major contributing factor to the relative paucity of our understanding of *P. vivax* basic biology. Low parasitemias also make *P. vivax* more difficult to detect by microscopy and RDTs, which hinders surveillance and control efforts.

P. vivax-infected reticulocytes also display a very different appearance than *P. falciparum*-infected erythrocytes, the former being larger, more deformable, and decorated on their surface with caveolae-vesicle complexes or Schüffner's dots (Aikawa et al., 1975, Barnwell, 1990, Suwanarusk et al., 2004). Deformable *P. vivax*-infected reticulocytes are not cleared by the spleen and thus all developmental stages circulate in the peripheral blood (Suwanarusk et al., 2004). These characteristics contrast with the smaller, rigid, and knobbed morphology of *P. falciparum* trophozoite-infected erythrocytes, which escape splenic clearance by sequestering in microvessels (Luse and Miller, 1971, Cranston et al., 1984). Sequestration is thought to be unique to *P.*

falciparum and along with higher parasitemias contributes to severe disease and mortality in *P. falciparum* infections.

P. vivax gametocytes are frequently present in the blood at the onset of symptoms, in contrast to *P. falciparum*, and consequently these parasite forms can be transmitted to mosquitoes before treatment begins (Douglas et al., 2010). This feature combined with low parasitemias and a potentially large reservoir of asymptomatic carriers further adds to the complexity in interrupting transmission in the many low-transmission *P. vivax* settings.

1.1.4 *P. vivax* pathology

All the symptoms and pathology of malaria are caused by blood-stage parasites, and classically include periodic attacks of fever, chills, and sweats called “paroxysms” (Evans and Wellems, 2002). While periodicity of attacks is common, most patients with uncomplicated *P. vivax* malaria will have a combination of chills, sweats, headaches, nausea, vomiting, body aches, and general malaise (cdc.gov) approximately 7-14 days after a bite from an infectious mosquito that can be mistaken for influenza or other infections in areas where malaria is uncommon. Additional symptoms may include enlarged spleen and/or liver, mild jaundice and increased respiratory rates (cdc.gov).

Severe cases of *P. vivax* malaria have also been reported, with growing agreement that their overall impact has been largely underestimated (Price et al., 2007). Nearly all of the severe complications caused by *P. falciparum* have also been reported for *P. vivax* infections, although coma appears to be rare for *P. vivax*. Severe disease from *P. vivax*-associated anemia, respiratory distress, renal failure, spleen enlargement, and low birth weight in infants have been reported in Pakistan, India, Southeast Asia, and Oceania [reviewed in (Price et al., 2009, White et al., 2014)]. Many aspects of *P. falciparum*-associated severe disease relates to either infected erythrocytes sequestering in organs or high parasitemias. As described above, sequestration of the parasite biomass in organs such as the brain and lungs can cause severe complications, such as coma and pulmonary edema. *P. falciparum* infects erythrocytes of all ages leading to the potential complication of severe anemia as the number of parasites exponentially increases in the blood. The WHO lists parasite counts $>250,000/\mu\text{l}$ as a risk for severe disease in *P. falciparum* infections. In contrast, erythrocytes infected with *P. vivax* parasites do not sequester and parasitemias remain far lower at under 1%, yet still have the potential for severe disease.

A study in Brazil reported admissions to intensive care for *P. vivax* infections with parasite counts $>500/\mu\text{l}$ (Lanca et al., 2012), indicating that a much lower parasite density can lead to severe disease in *P. vivax* compared to *P. falciparum*. The basis of this has not been fully determined, but relapsing or chronic infections with *P. vivax* may play a role (White et al., 2014). Several reviews have found that similar to *P. falciparum*, 10-20% of hospitalizations for *P. vivax* are classified as severe, and 5-15% of severe cases will lead to death (Price et al., 2007, Baird, 2013). Severe disease from *P. vivax* is often confined to populations in lower-resourced, tropical settings with high rates of hypnozoite-based relapse and comorbidities (nutritional, genetic, and immunological disorders), which may make them particularly vulnerable to severe disease (Price et al., 2009). This burden of chronic relapse and its potential contributions to severe pathology underline the need for effective treatments for the dormant liver stage, which is discussed further below.

1.1.5 Current *P. vivax* treatment and control measures

Several aspects of *P. vivax* biology make it challenging for treatment and control measures. As discussed previously, *P. vivax* infections generally reach very low parasitemias (under 1%), and a study in the Solomon Islands demonstrated that traditional measures of detection, such as LM and RDTs, significantly underestimate parasite prevalence in endemic populations (Waltmann et al., 2015). This same study also demonstrated that the vast majority of cases were asymptomatic. These factors make surveillance for *P. vivax* particularly difficult and contribute to the persistence of this pathogen, even in regions where significant progress is made in controlling *P. falciparum* infections.

Vector biting habits also pose challenges for control. The main vector control measures including long-lasting insecticide-treated nets (LLINs) and indoor residual spraying (IRS) aim to prevent biting of human hosts indoors at night and have proven highly effective in reducing the incidence of *P. falciparum* malaria (Goodman et al., 1999). The dominant malaria-transmitting anopheline species in *P. vivax*-endemic Southeast Asia and South America, such as *Anopheles dirus*, *Anopheles minimus*, and *Anopheles maculatus* tend to bite outdoors at dusk, rendering indoor strategies like IRS and LLINs less effective (Trung et al., 2005, Trung et al., 2004). In addition to this, LLIN usage varies significantly in populations and is additionally challenged by increasing mosquito resistance to pyrethroids (Alonso and Tanner, 2013).

Drug treatment challenges pose the final and perhaps greatest threat to *P. vivax* control. The foremost of these problems is the seemingly inevitable development of parasite resistance to frontline treatments. While chloroquine resistance was widespread in *P. falciparum* parasites by the 1980s, it remained the first-line treatment for *P. vivax* through 2010 in all but three malaria-endemic countries (WHO, 2010). The WHO now lists chloroquine as the recommended treatment only in “chloroquine-sensitive areas” (WHO, 2014), and there is ample evidence that chloroquine resistance in *P. vivax* parasites is spreading throughout Southeast Asia, though the mechanism remains unclear (Douglas et al., 2010). This delay in the acquisition of chloroquine resistance in *P. vivax* compared to *P. falciparum* is potentially caused by differences in parasite biology, as *P. vivax* gametocytes present at the time of symptoms while they are delayed by several days during *P. falciparum* infections. Thus, *P. vivax* parasites can be transmitted to mosquitoes before any drug pressure is experienced (Mendis et al., 2001).

Front-line treatment in most *P. vivax*-endemic areas is now ACT for both *P. vivax* and *P. falciparum*. These therapies include 3-day regimens using a short-acting artemisinin derivative as the potent effector of rapid parasite clearance, while a longer-acting partner drug such as mefloquine, lumefantrine, or piperaquine eliminates the remaining parasite burden (WHO, 2015). ACTs have contributed to the 31% reduction in malaria mortality over the last 10 years; however, their continued use is under threat with the development of “slow-clearing” *P. falciparum* parasites in western Cambodia (Dondorp et al., 2009) that were recently shown to be capable of infecting African vectors in laboratory-based experiments (St Laurent et al., 2016). Significant resistance to ACTs in Africa would have devastating consequences for malaria mortality. This adds additional urgency to the need for new treatments and vaccine candidates. ACTs are currently highly effective against *P. vivax* with the exception of artesunate + sulfadoxine-pyrimethamine (SP), due to significant *P. vivax* resistance to SP (WHO, 2015, Young and Burgess, 1959).

P. vivax (along with *P. ovale*) is capable of relapse of disease due to activation of dormant, liver hypnozoites. This stage will make *P. vivax* a much more difficult parasite to eradicate than *P. falciparum*, as the hypnozoites represent a pool of parasites that can re-infect the erythrocytes of individuals that have been previously drug cured. This is readily observed in patients treated by ACTs, where *P. vivax* relapses occur at different intervals depending on the half-lives of the long-acting partner drug used (WHO, 2014). Currently, primaquine is the only drug approved and available to eliminate hypnozoites.

This reliance on a single drug poses a serious risk to public health as parasites in Oceania have already shown resistance to primaquine treatment (Price et al., 2009). Primaquine is also contraindicated in pregnant women and individuals with G6PD deficiency, which affects up to 25% of people in *P. vivax*-endemic areas, as use of primaquine in such patients can cause severe hemolytic anemia (Nkhoma et al., 2009). This problem is complicated by the lack of inexpensive, point-of-care tests for G6PD deficiency, meaning that primaquine is simply not prescribed in many *P. vivax* cases.

There has been some progress at developing alternative drugs for eliminating liver stage parasites. The most developed drug, tafenoquine, is in the same class (8-aminoquinoline) as primaquine and has been shown to prevent relapse in 89% of patients up to 6 months after treatment (Llanos-Cuentas et al., 2014). It is administered as a single dose, a large improvement over the 14-day regimen for primaquine; however, its use is still contraindicated in individuals with G6PD deficiency. This remains a significant limitation in areas where *P. vivax* is most prevalent.

1.2 *P. falciparum* and *P. vivax* genomics

1.2.1 *P. falciparum* and *P. vivax* whole genome sequencing

Several innovations have been truly transformative in the study of *Plasmodium*. One of the first breakthroughs came through the successful adaptation of *P. falciparum* to *in vitro* culture (Trager and Jensen, 1976), providing worldwide access to a continuous supply of parasites for experiments not previously possible. The development impacted nearly all aspects of basic *Plasmodium* research (Trager and Jensen, 1997), and led to the stepwise characterization of a host of individual genes. In contrast, *P. vivax* has remained firmly recalcitrant to reliable *in vitro* culture, in part due to this parasite's preference for invading reticulocytes, and thus it is very difficult to study. Characterizing gene function in *P. vivax* has primarily relied on parasites from primate infections, clinical isolates from *P. vivax*-endemic areas, or both.

Another transformative innovation in *Plasmodium* research has come through the development of the genomics field, beginning with the completion of the entire genome sequence of the *P. falciparum* 3D7 strain in 2002 (Gardner et al., 2002) and the *P. vivax* Sal 1 strain in 2008 (Carlton et al., 2008). The sequencing of both reference genomes

involved multi-year, globally collaborative efforts and relied on the capillary sequencing “Sanger” method, including significant manual genome finishing.

The *P. falciparum* genome produced several important results, and provided a valuable reference for the subsequent *P. vivax* reference genome. The *P. falciparum* 3D7 reference genome led to the annotation (including significant manual annotation) of over 5300 gene models at the time of publication (now near 5400). Manual curation still occurs and is actively maintained by GeneDB (www.genedb.org) and PlasmoDB (www.plasmodb.org). Nuclear genes are organized on 14 chromosomes ranging in size from 0.6 Mb (chromosome 1) to 3.3 Mb (chromosome 14), with a total genome size of ~23 Mb. The majority (53%) of the genome contains protein-coding content. *P. falciparum* also harbors a 6-kb mitochondrial genome and a 35-kb apicoplast genome. The apicoplast, a plastid homologous to chloroplasts (McFadden et al., 1996, Waller and McFadden, 2005), appears to be essential for parasite survival in the asexual blood stage (He et al., 2001) through the biosynthesis of an isoprenoid precursor (Yeh and DeRisi, 2011). The majority of *P. falciparum* genes appeared to be unique to *Plasmodium*; 60% could not be assigned functions, as they were not sufficiently similar to other sequenced and annotated genes in other genomes.

The *P. falciparum* 3D7 reference genome contained three highly variable gene families, including *vars* (variable), *rifs* (repetitive interspersed family) and *stevors* (sub-telomeric variable open reading frame). The 60-member, primarily subtelomeric, *var* gene family is only found in *P. falciparum* and is responsible for encoding *P. falciparum* erythrocyte membrane protein 1 (PfEMP1) (Baruch et al., 1995, Smith et al., 1995, Su et al., 1995). PfEMP1 proteins are trafficked to the surface of infected erythrocytes (Leech et al., 1984), where they mediate adherence to host endothelial receptors (Kyes et al., 2001), enabling parasites to sequester in microvessels and avoid clearance by the spleen. IgG responses directed at PfEMP1 are a primary component of the host protective antibody response in *P. falciparum* infections (Bull et al., 1998). Sequestration appears not to occur in any other human-infecting *Plasmodium* species. The two remaining gene families, *rifs* and *stevors*, are also primarily sub-telomeric with 149 and 28 members, respectively, and are members of the Pir superfamily (Janssen et al., 2004) found in all *Plasmodium* species sequenced to date. Little is known about the proteins produced by these families, *rifs* and *stevors*, but they also appear to be trafficked to the surface of infected erythrocytes and may also be involved in evading the host’s immune system.

The *P. vivax* Sal 1 genome was published several years after the *P. falciparum* 3D7 genome (in 2008) from an isolate from a malaria patient in El Salvador in 1972, and subsequently propagated through *Saimiri boliviensis boliviensis* monkeys. *P. vivax* Sal 1 showed a similar nuclear genome organization to *P. falciparum* 3D7 with ~27 Mb, including over 5400 annotated genes spread over 14 chromosomes, albeit with a much lower AT-content (57.7% vs. 80.6% in *P. falciparum*). Synteny was well conserved and 77% of genes had orthologs with *P. falciparum*, *P. knowlesi*, and *P. yoelii*. No *var* gene equivalents were detected, but there were 346 Pir family representatives, referred to as *virs* in *P. vivax*, primarily in telomeric regions. The genome sequencing did have some limitations, however, with missing or incomplete assemblies in the highly repetitive telomeric regions and around ~4.3 Mb of contigs that could not be linked to specific chromosomes. Gene models were largely predicted through automated gene-model prediction software, and gene names and functions were often inferred from *P. falciparum* annotation. The annotation contained far less manual curation of gene models, making for a less complete data set with missing transcripts and erroneous gene boundaries possible. Nevertheless, it was a huge multi-year achievement and an enormous push forward for the study of *P. vivax* biology.

Genomic sequencing technology has developed rapidly over the last 25 years, beginning with gel- and then capillary-based “Sanger” sequencing technology developed by Fred Sanger (Sanger et al., 1977). This technology produced long (up to 900 bases/read) reads which dominated the reference genome sequencing projects of the late 1990s and early 2000s including most notably, the human genome in 2001 (Lander et al., 2001). Both the *P. falciparum* 3D7 and *P. vivax* Sal 1 reference genomes were sequenced using this expensive sequencing technology. Several sequencing technologies gained prominence as cheaper, higher-throughput options in the 2000s [Illumina/Solexa (Bentley et al., 2008), 454 (Droege and Hill, 2008), SOLiD (Shendure et al., 2005)], with additional new technologies being developed and more commonly used [Pacific Biosciences (Roberts et al., 2013)] or emerging on the horizon (Oxford Nanopore). Sequencing technology today is dominated by Illumina, which generates gigabases (Gb) of data per run (compared to kb/run for capillary sequencing) through medium length, relatively inexpensive reads (up to 150 bases/read). The material required to produce libraries has declined 100-fold from the 5 µg required for a capillary sequencing library to the 50 ng required for an Illumina HiSeq library today (leading to Gb more data). The reductions in sequencing cost and

material requirements and increases in data output have greatly facilitated the development of genome-wide approaches to study *Plasmodium* over the last several years including studies aimed at large-scale population genetics (Manske et al., 2012) and large-scale gene-knockout experiments (Schwach et al., 2015).

P. vivax research has also benefitted from this technological shift and several additional genomes have been fully sequenced since *P. vivax* Sal 1, including monkey-adapted laboratory isolates from India, North Korea, Mauritania, Brazil, and Peru (Neafsey et al., 2012) and field isolates from Madagascar, Cambodia, Peru, and Thailand (Menard et al., 2013, Chan et al., 2012, Auburn et al., 2013, Dharia et al., 2010, Bright et al., 2012). More parasite material was available for the monkey-adapted samples, which therefore generated deeper sequence coverage and the ability for *de novo* assembly and annotation (though assemblies were still more fragmented than the *P. vivax* Sal 1 reference genome). Sequencing from the field isolates was primarily suitable for SNP calling after mapping to the *P. vivax* Sal 1 reference genome. Comparisons between the genomes demonstrated that the genetic diversity in *P. vivax* is twice that of *P. falciparum*, indicative of the long global history of *P. vivax* colonization (Neafsey et al., 2012, Carlton et al., 2013). The data also suggested the “capacity for greater functional variation,” which may add to the challenge of global *P. vivax* malaria elimination (Carlton et al., 2013).

1.2.2 *P. falciparum* and *P. vivax* transcriptome studies

The publication of the *P. falciparum* and *P. vivax* reference genomes laid the foundation for the study of transcription in both organisms. Studies of transcription involve probing for the repertoire of RNA molecules within cells, which may include protein-coding mRNA, non-coding RNA, and regulatory RNA. Transcription studies highlight the dynamic changes within cells, which drive the development of organisms and assist with connecting genes with functions. This is particularly important for *Plasmodium*, as over half the genes have unknown functions. Some early RNA studies connected to reference genome sequencing projects included the single-sided sequencing of cDNA derived from parasite RNA transcripts, called expressed sequence tags (ESTs). These data trained or validated automated gene-model prediction software. However, it represented only a short fragment of a transcript, and was therefore limited in detecting gene boundaries and actual abundance of particular RNA molecules within *Plasmodium* cells. Two major

technologies subsequently dominated genome-wide RNA studies in *Plasmodium*: microarrays and RNA-Seq.

1.2.2.1 Microarrays

Microarray technology enables the abundance of transcripts to be measured simultaneously based on a set of predefined probes often produced from known or predicted transcripts for a given reference genome. The principle behind microarrays involves measuring the light emitted from hybridization events between a set of chemically-labelled DNA probes with known sequences (reference genome sequences, for instance) to a set of unknown DNA or RNA molecules (e.g., RNA from different time points in a life cycle). The proof of principle for DNA microarray technology was established by Fodor et al. from the Affymax Research Institute in 1991 (Fodor et al., 1991). The first use for transcriptome study came in 1995 where binding events between 45 arrayed *Arabidopsis thaliana* cDNAs and fluorescently-labeled *A. thaliana* mRNA were measured. The fluorescence intensity measured was proportional to the hybridization between the cDNA and mRNA, and correlated with the initial mRNA concentration (Schena et al., 1995). This opened the door to assessing the transcription of multiple genes in parallel.

Transcriptional microarray technology was first applied to *Plasmodium* in the early 2000s with the publication of the transcriptional profile of the asexual blood stages for laboratory-adapted *P. falciparum* HB3 (Bozdech et al., 2003). The blood stages cause the symptomatic portion of malaria and were the most tractable for study, given the availability of large volumes of highly synchronous *in vitro* *P. falciparum* HB3 cultures. Bozdech and colleagues measured RNA every hour during blood-stage development using microarrays comprised of long oligonucleotides synthesized based on the *P. falciparum* 3D7 reference genome and spotted onto glass slides. The data showed transcriptional profiles for over 60% of genes and found that the majority of expressed genes (80%) appeared to have single peak expression during the 48-hour cycle (Figure 1.3). The transcription of genes appeared to be a highly-ordered cascade with genes of related function being highly synchronized; for example, genes relating to pyrimidine ribonucleotide synthesis had peak abundance at the same time (18-22 hours post erythrocyte invasion) as genes involved in purine salvage pathways. Similarly, deoxyribonucleotide synthesis reached peak abundance (~32 hours post erythrocyte

invasion) just as DNA replication to produce daughter merozoites is required. Genes relating to erythrocyte invasion, such as AMA1 and EBA175 (discussed further below) reached peak abundance in the schizont stage. This pattern of genes being transcribed maximally when their encoded proteins function was termed the “just in time” model of gene expression. *Plasmodium* appeared to possess this feature much more than other eukaryotes such as yeast (Spellman et al., 1998) or mammalian cells (Whitfield et al., 2002), which exhibit this pattern in only 15% of genes. The profile appeared most similar to developmental processes, such as early development in *Drosophila melanogaster* (Arbeitman et al., 2002). This cyclical transcriptional profile mirroring developmental processes led to the renaming of the asexual blood stages as the intraerythrocytic development cycle or IDC.

A second microarray study of 7 *P. falciparum* IDC time points, as well as sporozoites and gametocyte stages was performed and published in parallel (Le Roch et al., 2003). It confirmed expression in at least 1 stage for 4557 genes (88% of the genome) with up to 5-fold variation in abundance. The work showed single peak abundance for 43% of genes (varying at least 1.5 fold between stages). The ring and trophozoite stages appeared enriched for peak abundance of genes relating to metabolism and cellular processes, such as transcription, translation, and energy metabolism. Genes relating to parasite-host interactions, such as merozoite surface proteins (MSPs) or circumsporozoite protein (CSP), reached peak abundance during schizont and sporozoite stages, respectively. This suggested that parasite transcription is closely linked with the functional requirements of the immediate stage.

A microarray study in 2006 compared the transcriptional profiles of 3 laboratory strains of *P. falciparum*: 3D7, HB3, and DD2 (Llinas et al., 2006). The data showed similar transcriptional profiles for the vast majority of genes in each isolate, suggesting that transcriptional timing is “hard-wired” for nearly all genes. A small group of 69 genes (1.3% of the genome), however, displayed transcriptional shifts of 12 or more hours and was enriched for genes relating to parasite-host interactions and antigenic variation. This pattern was also conserved in newly-adapted field isolates (Mackinnon et al., 2009).

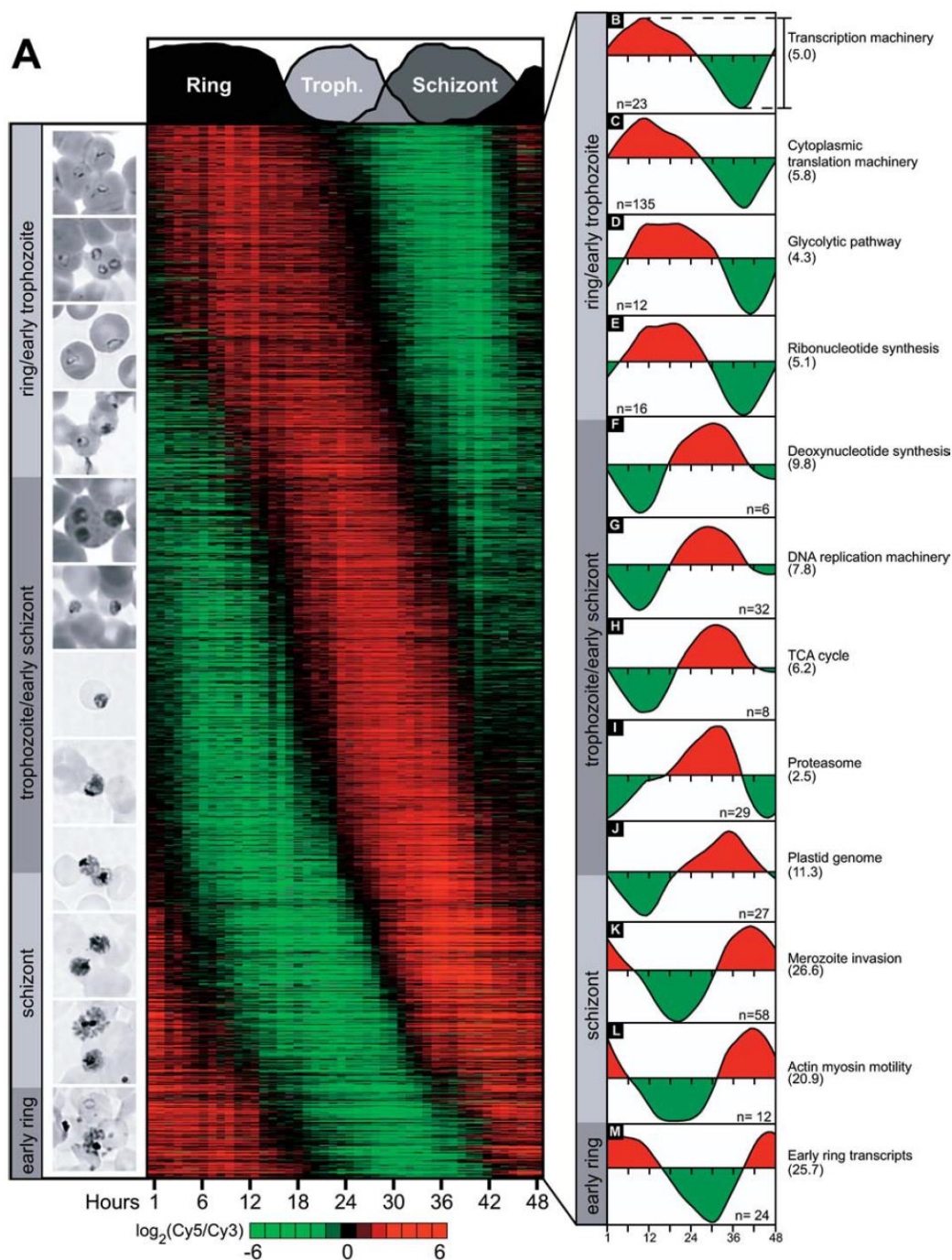


Figure 1.3: Intraerythrocytic development cycle (IDC) overview for *P. falciparum*

(A) Phaseogram for an ordered set of 2712 genes (rows) showing transcription every hour post invasion (columns), with morphological stages shown on the left. (B-M) Average expression profiles for genes with similar biochemical processes and functions at the various stages with the mean peak-to-trough amplitude shown in parentheses. Figure reprinted from (Bozdech et al., 2003) under the Creative Commons Attribution (CC BY) license.

Two *P. vivax* microarray studies have been published to date (Bozdech et al., 2008, Westenberger et al., 2010). The Bozdech *et al.* study focused on the *P. vivax* IDC using 3 highly synchronous patient isolates from Thailand that were cultured *ex vivo* and sampled at 9 time points over 48 hours. The isolates displayed a similar transcriptional cascade to *P. falciparum* with over 3500 genes (70% of the genome) showing peaks in abundance during 1 time point. Transcription profiles were evenly spaced throughout the IDC for syntenic orthologs between *P. falciparum* and *P. vivax*, while non-syntenic orthologs tended to have transcription shifted to the late IDC/schizont-to-ring transition. This implies that genes related to invasion and early establishment of the parasite in new host cells are largely responsible for the key differences between the species. The majority of orthologous, syntenic genes (68%) appeared to have identical expression profiles between *P. vivax* and *P. falciparum* orthologs, and a further 22% (including *msp8*) showed only slight shifts. The remaining 10% appeared to have dramatic shifts between the species (Pearson correlation coefficient between -0.2 and -0.1). This set is likely to contain genes underlying the differences between *P. vivax* and *P. falciparum*; for example, it includes the 2 *P. vivax* orthologs for the knob-associated histidine-rich protein (KAHRP), which have peak expression during the schizont stage in *P. vivax* and the early ring stage in *P. falciparum*. The single-copy KAHRP in *P. falciparum* has been associated with the formation of “knobs” on the surface of infected erythrocytes, which are essential for cytoadherence (Rug et al., 2006). This feature appears to be unique to *P. falciparum*, and KAHRP appears to serve a different purpose in *P. vivax*. In line with the above example, antigenic presentation in general appeared to differ between *P. falciparum* and *P. vivax* with *P. vivax* virs presenting in two waves: just after invasion and later in the schizont stages, when surface antigen transcription in *P. falciparum* is silent.

The Westenberger *et al.* study expanded the transcriptional profile of *P. vivax* to include additional life stages, including sporozoites, gametes, zygotes, and ookinetes, and human asexual blood stages, reporting differential expression in nearly 4326 genes (80% of the genome) (Westenberger et al., 2010). The study included transcriptional data for nearly 200 additional genes not included in the Bozdech *et al.* *P. vivax* microarray study, and established that the general processes for growth, metabolism, and host-parasite interactions are shared between *Plasmodium* species. It also provided a means for prioritizing genes with potential transmission-blocking or pre-erythrocytic stage vaccine development.

The majority of transcriptional studies in *Plasmodium* have focused on the IDC, although there have been some exceptions. Gametocyte stages of *P. falciparum* have been the subject of several studies (Le Roch et al., 2003, Silvestrini et al., 2005, Young et al., 2005). Rodent models have also been used as a model for gametocyte study, including a proteomics study published in 2005 (Khan et al., 2005). Parasite material from outside of the asexual blood-stage is easier to obtain from model rodent malarias, which have also been used to study sporozoite stages (Lasonder et al., 2008, Mikolajczak et al., 2008) and liver stages (Tarun et al., 2008).

In summary, microarrays have proven to be a powerful technology for describing the genome-wide transcriptional profiles of *Plasmodium*. Most studies have focused on the IDC of *P. falciparum* and *P. vivax* due to the limited availability of material from other life stages. These studies establish a highly-ordered periodic transcriptional profile for the majority of *Plasmodium* proteins, and similar transcriptional profiles in laboratory-adapted and clinical isolates. Microarray technology is inherently limited in several ways, however, as transcriptional profiles can only be measured for probes included in the arrays and probe content is contingent on high-quality reference genomes. In the case of *P. vivax*, this made for an incomplete gene list in the Bozdech *et al.* study, as the reference genome annotation was incomplete at the time of the study. Microarrays are also limited in sensitivity, such that genes with low transcription are unlikely to be detected. At the opposite end, highly abundant transcripts will saturate the available probes, thus limiting the ability to describe true abundance. Using pre-designed probes also limits the detection of alternatively-spliced transcripts and 5' and 3' UTR regions.

1.2.2.2 RNA-Seq

A number of the technical limitations of microarrays are addressed by the direct sequencing of cDNA libraries made from RNA, called RNA-Seq. As previously discussed, the technological improvements in sequencing technologies over the last decade have greatly reduced the quantities of starting material needed and the cost for generating whole-genome sequencing data. Transcriptome sequencing in particular benefits from “deep sequencing” and is most commonly generated using Illumina sequencing technology, which currently provides the lowest cost per base. Early RNA-Seq studies focused on model organisms, such as yeast (Wilhelm et al., 2008), mouse (Mortazavi et al., 2008), and human (Pan et al., 2008, Wang et al., 2008), and were

reviewed by (Wang et al., 2009). In addition to producing comparable transcript abundance data with microarrays, additional alternative splicing data were often detected; alternative splicing was estimated to occur in 95% of multi-exon human genes (Pan et al., 2008). Early comparisons to microarray data, for instance by Marioni *et al.*, found RNA-Seq data to be highly reproducible and comparable to microarray data in identifying differential gene expression, with the added benefits of detecting transcripts with low abundance, alternative-splice variants, and novel transcripts (Marioni et al., 2008).

RNA-Seq was first applied to *Plasmodium* in 2010 through the high-throughput sequencing of cDNA made from RNA from a highly synchronous *P. falciparum* 3D7 *in vitro* culture (Otto et al., 2010). The study measured RNA abundance from 7 regularly-spaced time points throughout the 48-hour IDC. The study led to the improvement of over 10% of the ~5400 gene models and identified 121 novel transcripts. It uncovered 84 cases of alternative splicing and confirmed 75% of predicted splice sites. The RNA-Seq data showed good correlation with microarray experiments (Pearson correlation coefficients from 0.7 to 0.8). It also expanded the list of genes exhibiting some expression during the IDC to 4871 genes (~ 90% of the genome), suggesting that the majority of the genome is transcriptionally active during the blood stages. Overall, RNA sequencing proved to validate and expand on microarray results.

At the time of initiating this work, no RNA-Seq data have been published for *P. vivax*. This is primarily due to the inability to culture *P. vivax in vitro*. But recent reductions in the RNA requirements for Illumina libraries have made sequencing from field isolates more feasible. We aimed to address this need using clinical isolates from Cambodia as discussed in Chapter 3. RNA-Seq using *P. vivax* clinical isolates is a particularly pressing need as the initial *P. vivax* microarray data set lacked probes for at least 250 genes now included in the most up-to-date genome annotation. No alternative splicing events are yet known for *P. vivax*. Identifying new blood-stage genes abundant during the schizont stage may expand the list of genes to consider as potential merozoite development or invasion-related genes to prioritize for further functional characterization as vaccine candidates.

1.3 Erythrocyte invasion

1.3.1 Overview of invasion

The process by which infective merozoites invade new host erythrocytes has long been the focus of study and vaccine development for several reasons. Blood-stage parasites cause the symptoms of malaria and vaccines directed at merozoite invasion would therefore halt disease progression. This also represents one of the few points during which free parasite forms are fully exposed to the host immune system, and may therefore represent a “weak link” that both therapeutics and vaccines could target. A full understanding of blood-stage parasites is thus essential for rational prioritization of genes for further consideration as drug or vaccine targets.

Merozoites that egress from schizonts represent one of the smallest eukaryotic cells known, at 1.2 μm in length or less than 1/5 the size of the erythrocytes they invade (Figure 1.4) (Garcia et al., 2008). This tiny form has a polar structure with a pointed apical end and a wider posterior end. The apical end of merozoites contains 3 sets of secretory organelles known as the rhoptries, micronemes, and dense granules, all known to be important for invasion of new erythrocytes. The wider posterior end of the merozoite houses the genetic and metabolic machinery including the nucleus, mitochondrion, and apicoplast (Bannister and Mitchell, 2003, Garcia et al., 2008). These organelles are contained within a double membrane structure called the inner membrane complex (IMC), which is connected to the outer plasma membrane via actin filaments (Farrow et al., 2011). An adhesive coat 15 nm thick covers the outer surface in clumps of narrow protruding bristles (Garcia et al., 2008). The structure of merozoites is supported by 3 cytoskeletal polar rings, 1 of which connects to microtubules linked to the IMC at the posterior end. These structures help to orient the invasive secretory organelles during merozoite assembly.

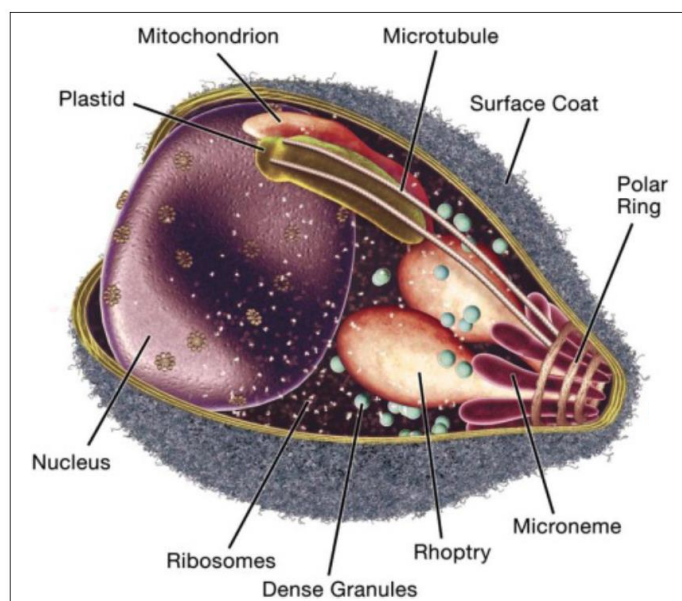


Figure 1.4: *Plasmodium* merozoite structure

Schematic of merozoite structure and organelles, including the genetic material of the posterior (left) end (nucleus, mitochondrion, plastid/apicoplast) and the invasive secretory organelles of the apical (right) end (rhoptries, micronemes, dense granules). Figure reprinted from (Cowman & Crabb, 2006), © 2006, with permission from Elsevier.

Plasmodium invasion of erythrocytes is a multistep process that begins with the initial reversible attachment of a merozoite to an erythrocyte (Figure 1.5). Insight into this process was first described using video microscopy with *P. knowlesi* in 1975 (Dvorak et al., 1975) and confirmed by similar experiments in *P. falciparum* 34 years later (Gilson and Crabb, 2009). This initial merozoite binding can occur at any part of the merozoite surface and is presumed to involve interactions between proteins in the parasite outer coating and the host erythrocyte at somewhat long range (20-30 nm), leading to a weak distortion of the erythrocyte membrane (Cowman and Crabb, 2006, Garcia et al., 2008, Weiss et al., 2016). The merozoite then reorients so that its apical end faces the erythrocyte and the gap between the parasite and erythrocyte narrows (Aikawa et al., 1978, Gilson and Crabb, 2009), leading to a more significant distortion of the erythrocyte membrane (Weiss et al., 2015). Receptor-ligand interactions thought to mediate this and downstream invasion events often involve proteins released from apical organelles, discussed in more detail in the following section. The release of proteins from the apical organelles appears to occur in multiple (potentially calcium or potassium-dependent) stages starting from egress until after the merozoite is engulfed in the erythrocyte (Zuccala et al., 2012, Weiss et al., 2016). An irreversible commitment to invasion occurs,

and a tight junction forms at the point of attachment (Aikawa et al., 1978, Bannister et al., 1975). The tight junction moves from the apical to the posterior end of the merozoite driven by an actin-myosin motor connected to the IMC. The merozoite is engulfed into the erythrocyte or reticulocyte as an internal parasitophorous vacuole (Aikawa et al., 1978, Baum et al., 2006). Surface proteins are shed throughout the invasion process through a calcium-sensitive serine protease, SUB2, discharged from the micronemes, which translocates across the parasite surface at the tight junction (Harris et al., 2005, Withers-Martinez et al., 2012). The adhesive proteins mediating the tight junction are also shed via serine proteases as the merozoite is engulfed and the erythrocyte membrane is resealed (O'Donnell et al., 2006, Garcia et al., 2008). The parasite then begins the process of growth towards either additional asexual replication or gametocyte differentiation.

Both genomic and transcriptional data support that *P. vivax* and *P. falciparum* frequently differ in the genes relating to immune evasion and host-interaction. This underlines the fact that while the general process of merozoite invasion of erythrocytes is conserved for mammalian *Plasmodium* parasites, the ligand-receptor parasite-host interactions mediating invasion tend to be highly parasite-specific.

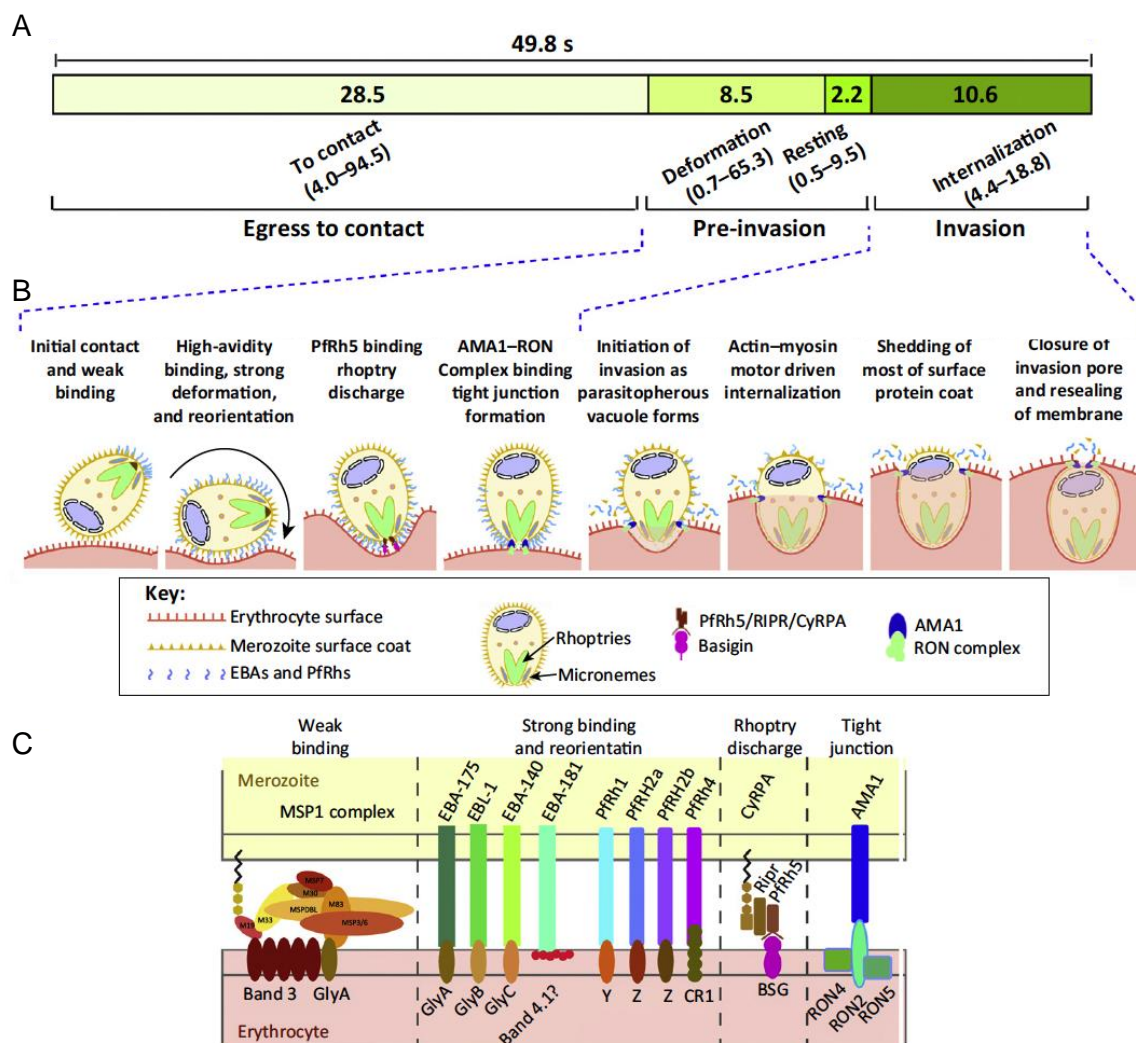


Figure 1.5: Overview of *P. falciparum* invasion of erythrocytes

P. falciparum merozoite egress through invasion is under 1 minute for 77% of *in vitro* invasion events. (A) Bar shows average times (seconds) for invasion stages. Timing for “egress to contact” is likely longer *in vivo* with flowing conditions and higher-density erythrocytes. “Pre-invasion” includes initial contact, significant deformation during merozoite reorientation, and resting, during which formation of tight junction likely occurs. “Internalization” occurs in a tightly-defined window (95% confidence interval of 9.99–11.11 s). (B) Cartoons with general description of events and several *P. falciparum* invasion ligands. (C) Known *P. falciparum* ligands with putative binding partners, including arbitrarily-named Y, Z, and Z for unknown receptors. Abbreviations for erythrocyte receptors: BSG, basigin; CR1, complement receptor 1; GlyA–C, glycoporphins A–C; RON, rhoptry neck protein. Figure reprinted with slight modifications from (Weiss et al., 2016), © 2016, with permission from Elsevier.

1.3.2 *P. falciparum* erythrocyte invasion ligands

Much more is known about the specific interactions required for merozoite invasion in *P. falciparum* parasites (described in detail below) than for *P. vivax* and as a result, studies of invasion in *P. vivax* tend to focus on *P. falciparum* invasion homologs. The molecular details of the *P. falciparum* invasion process are therefore described below in detail, with the known features of *P. vivax* invasion described subsequently.

1.3.2.1 Initial reversible attachment between merozoites and erythrocytes

Plasmodium merozoite attachment to and invasion of erythrocytes is generally mediated by a series of receptor-ligand interactions between erythrocytes and merozoites [recently reviewed in (Weiss et al., 2016)]. Initial reversible binding occurs at the outer coat of the parasite, and over 30 proteins have been identified and localized to the merozoite surface (Cowman and Crabb, 2006, Garcia et al., 2008). Proteins with glycosylphosphatidylinositol (GPI)-anchors are enriched in this set and mostly cluster within detergent-resistant membrane (DRM) regions (Sanders et al., 2005, Cowman and Crabb, 2006). Binding is likely initiated by a complex of proteins including MSP1, the most abundant GPI-anchored surface protein, leading to weak deformation of the erythrocyte membrane in the area as receptors-ligand molecules coalesce in the region of contact (Weiss et al., 2015). MSP1 is processed from a 195-kDa precursor into 4 fragments that remain non-covalently linked prior to invasion. The membrane-bound fragment, MSP1₄₂, is further cleaved by SUB2 during invasion into MSP1₁₉, which remains associated with the surface throughout the invasion process (Blackman et al., 1991, Harris et al., 2005). Blocking the final MSP1-processing event has been shown to block merozoite invasion (Blackman et al., 1994). The MSP1 complex is comprised of the processed MSP1 fragments along with MSP7, MSP6, MSPDBL1, and MSPDBL2 (Stafford et al., 1996, Pachebat et al., 2001, Trucco et al., 2001, Lin et al., 2014)

Additional GPI-anchored proteins are also enriched at the merozoite surface in DRMs, including MSP2, MSP4, MSP5, P12, P92, P38, and P113. A number of these proteins (P12, P92, P38) are members of the six-cysteine (6-cys or Cys₆) protein family, members of which contain 2-7 domains (and occasionally partial domains) with highly-conserved cysteine residues (Templeton and Kaslow, 1999, Williamson, 2003). 6-cys proteins are structurally similar to SAG proteins in *Toxoplasma gondii* (Cowman and Crabb, 2006, Gerloff et al., 2005), with 3 disulfide bonds between the 6 cysteine residues supporting a

conserved structure between family members. 6-cys domain proteins act as ligands with host cell receptors during multiple life stages [i.e., P36 and P36p in sporozoites (Ishino et al., 2005, van Dijk et al., 2005); P230, P48/45 on the surface of sexual gamete stages (van Dijk et al., 2010)], making it highly likely that they serve a similar role in the asexual blood stages.

A number of additional proteins peripherally associate with GPI-anchored merozoite surface proteins during merozoite assembly (Sanders et al., 2005, Cowman and Crabb, 2006, Garcia et al., 2008). This includes the 6-cys protein P41, which lacks a GPI anchor and has been shown to interact with membrane-bound 6-cys protein P12 (Taechalertpaisarn et al., 2012). Members of the SERA family also associate with the merozoite membrane, with SERA5 being the most abundant (Sanders et al., 2005). The proteins contain a central protease domain with an active site cysteine or serine (Hodder et al., 2009). SERA5, while essential in the asexual blood stages, may lack a working protease domain and its function remains unclear (Hodder et al., 2009).

1.3.2.2 Merozoite EBL and RH proteins interact with specific erythrocyte receptors.

After initial attachment, the merozoite reorients, causing significant erythrocyte membrane deformation, so that its apical end interacts more closely with the erythrocyte membrane (Weiss et al., 2015). This requires interactions with proteins that span the plasma membrane and link directly or indirectly to the merozoite cytoskeleton (Weiss et al., 2016). A number of receptor-ligand interactions during this period have been well studied in *P. falciparum*, dominated by 2 protein families: the erythrocyte binding antigens (EBAs and EBLs) and the reticulocyte-binding protein homologues (RHs). Proteins from both groups are secreted from the apical organelles during the invasion process, facilitating it after the initial reversible contact (Harvey et al., 2012, Riglar et al., 2011). The signaling which precipitates the release of the proteins is discussed further below. It is hypothesized that a concentration gradient of EBL and RH proteins concentrated at the apical end of the merozoite may facilitate merozoite reorientation, though the function of the erythrocyte membrane deformation, which is independent of the actin-myosin motor, remains unclear (Farrow et al., 2011, Weiss et al., 2016).

The EBL family contains type-I transmembrane proteins with 2 cysteine-rich regions [Region II (RII) and Region VI (RVI)] in the extracellular region, and include EBA140 (PF3D7_1301600), EBA175 (PF3D7_0731500), EBA181 (PF3D7_0102500), EBL-1

(PF3D7_1371600) and a likely pseudogene with missense mutations, EBA165 (PF3D7_0424300) (Adams et al., 2001, Cowman and Crabb, 2006, Tham et al., 2012, Triglia et al., 2001). The RII contains two tandem Duffy binding-like (DBL) domains, which are homologous to the single DBL domain found in the *P. vivax* DBP (Sim et al., 1994). RH family proteins were first identified as the *P. yoelii* 235-kDa rhoptry protein (Py235) (Holder and Freeman, 1981) followed by the *P. vivax* reticulocyte-binding proteins (PvRBP-1 and 2) 11 years later (Galinski et al., 1992). The superfamily is common to *Plasmodium*, and in *P. falciparum* the group includes 5 members with transmembrane domains [RH1 (PF3D7_0402300), RH2a (PF3D7_1335400), RH2b (PF3D7_1335300), RH4 (PF3D7_0424200)] and a single smaller member lacking a transmembrane domain [RH5 (PF3D7_0424100)] (Rayner et al., 2000, Rayner et al., 2001, Triglia et al., 2001, Taylor et al., 2002, Rodriguez et al., 2008). A sixth member contains missense mutations and is likely encoded by a pseudogene [RH3, PF3D7_1252400] (Taylor et al., 2001).

At least 5 receptor-ligand interactions occur between the *P. falciparum* EBLs or RHs and erythrocytes during invasion (Figure 1.5): Glycophorin A-EBA175, Glycophorin B-EBL-1, Glycophorin C-EBA140, Complement receptor 1-RH4, and Basigin-RH5 (Camus and Hadley, 1985, Sim, 1995, Mayer et al., 2009, Maier et al., 2002, Crosnier et al., 2011). The EBLs and RHs have been described as “alternative pathway” ligands because many of these interactions appear to be functionally redundant; all EBLs and RHs with the exception of RH5 can be disrupted individually without affecting parasite survival (Duraisingh et al., 2003, Baum et al., 2009, Lopaticki et al., 2011). Knockout experiments targeting several members suggest that a minimum number of interactions is essential for survival (Duraisingh et al., 2003). This built-in functional redundancy may aid the parasite in overcoming host receptor polymorphism and phenotypic variation in ligands arising from immune pressure (Harvey et al., 2012, Tham et al., 2012).

The order of and signaling required for microneme and rhoptry protein release is the subject of continued research and debate. Singh, *et al.* found that exposure to low potassium ion concentrations (as in found in blood plasma) led to increased calcium levels in merozoite cytosol and the release of micronemal proteins such as EBA175 and AMA1 to the merozoite surface. Binding of EBA175 to erythrocyte receptor Glycophorin A restored calcium concentrations to basal levels and led to the release of rhoptry proteins CLAG3.1 and RH2b (Singh et al., 2010). The low potassium trigger for invasion was

challenged by Pillai *et al.* who found that asexual growth *in vitro* was not affected by high potassium ion concentrations in the growth media (Pillai *et al.*, 2013). Gao *et al.* showed that monoclonal antibodies against rhoptry protein RH1 blocked invasion by disrupting calcium signalling, and that monoclonal antibodies against EBA175 had no effect on calcium signalling. They further showed that disrupting the calcium signalling via RH1 monoclonal antibodies prevented the release of EBA175 to the merozoite surface (Gao *et al.*, 2013). Together these results suggest that the release of micronemal and rhoptry proteins may not follow a simple two-step process, and additional work is needed to clarify these portions of the invasion process.

Contrary to other receptor-ligand interactions, the basigin (BSG) interaction with RH5 is essential in all isolates studied to date. Recent work supports that the RH5-BSG interaction may be downstream of other EBL and RH receptor-ligand interactions (Weiss *et al.*, 2015). RH5 forms a complex with both RIPR and GPI-anchored CyRPA, and antibodies against any of these 3 proteins block erythrocyte invasion, though neither RIPR nor CyRPA appear to interact with erythrocytes directly (Chen *et al.*, 2011, Reddy *et al.*, 2015). A recent paper, however, contradicts that CyRPA has a GPI-anchor, thus leaving open the question of how the complex is attached to the merozoite surface (Volz *et al.*, 2016). Live imaging of invasion events shows that antibodies blocking the interaction between the RH5 complex and BSG have normal pre-invasion, deformation, and reorientation, indicating the interaction may play a role in the release of rhoptry proteins that are needed to form the tight junction (Weiss *et al.*, 2015).

1.3.2.3 Formation of the tight junction and the molecular motor of invasion

After rhoptry discharge the formation of a tight junction (also called the moving junction) is initiated. This process is thought to hinge on the interaction between the rhoptry neck (RON) proteins and AMA1. The current model suggests that the RON complex is translocated to the surface of erythrocytes (Riglar *et al.*, 2011) where RON2 then binds to AMA1 on the parasite surface to initiate the formation of the tight junction (Bai *et al.*, 2005, Pizarro *et al.*, 2005, Vulliez-Le Normand *et al.*, 2012) and recently reviewed in (Weiss *et al.*, 2016). Micronemal AMA1 binds the an extracellular loop of RON2 in both *P. falciparum* and the related apicomplexan parasite, *Toxoplasma gondii* (Lamarque *et al.*, 2014, Srinivasan *et al.*, 2011, Tonkin *et al.*, 2011, Tyler *et al.*, 2011). This interaction has led to the proposal that the AMA1-RON2 binding mediates the tight junction, which

links either directly or indirectly to the actin-myosin motor connected to the IMC. Recent high resolution imaging work supports that AMA1 is localized to the tight junction, which is at or directly adjacent to the actomyosin force propelling invasion into the host cell (Riglar et al., 2016, Bichet et al., 2014). However, work in both *Plasmodium* and *T. gondii* demonstrates that parasites with knockouts or conditional knockouts of AMA1 were still invasive though with greatly reduced efficiency, suggesting that AMA1 may play a greater role in host cell binding rather than linkage to the tight junction and actin-myosin motor (Bargieri et al., 2013, Giovannini et al., 2011). Component genes of the motor complex itself have been successfully disrupted in *T. gondii* while still maintaining successful invasion (Egarter et al., 2014). This calls into question whether the motor is absolutely required for invasion in other apicomplexans (Bargieri et al., 2014) and supports further, careful study of the potentially versatile invasion process.

Several additional proteins are also linked to invasion. In sporozoites, this process is aided by the membrane-spanning TRAP protein, which contains a thrombospondin type I repeats (TSR) domain and connects extracellular adhesins to the motor. TRAP is linked to the parasite cytoskeleton by aldolase (Buscaglia et al., 2003). In merozoites, it is proposed that MTRAP performs a similar function, as it also contains the TSR domain, is essential to the asexual blood stages, and contains an aldolase-binding cytoplasmic tail (Baum et al., 2006, Morahan et al., 2009). The extracellular region of MTRAP binds to the erythrocyte receptor Semaphorin 7A, although the biological implications for this are not fully understood (Uchime et al., 2012, Bartholdson et al., 2012). Several other TSR domain-containing proteins (PTRAMP, SPATR, TLP) also have peak expression during the asexual blood stage, though their function also remains unclear (Baum et al., 2006, Heiss et al., 2008, Morahan et al., 2009).

1.3.3 *P. vivax* reticulocyte invasion

In stark contrast to *P. falciparum*, current knowledge of *P. vivax* invasion of reticulocytes is extremely limited. Given the lack of continuous *in vitro* culture, it is commonly inferred that much of the *P. vivax* invasion process mirrors that of *P. falciparum*, as many one-to-one orthologs of *P. falciparum* invasion proteins exist. In a few cases, such homologues have been directly shown to have a role in invasion in *P. vivax* or localize to the merozoite surface or rhoptries, including AMA1, MSP1 (Galinski et al., 1992, Galinski and Barnwell, 1996, Cheng and Saul, 1994, Barnwell and Galinski, 1991),

rophtry neck protein 1 (RON1), and 6-cys proteins P12 and P38 (Li et al., 2012, Moreno-Perez et al., 2011, Mongui et al., 2008). In the majority of cases though, their function is simply inferred. It is also worth noting that many of the genes mediating receptor-ligand interactions in *P. falciparum* have no direct *P. vivax* homologs, and in most cases analogous receptor interactions have not yet been found in *P. vivax*. The fact that there are likely to be significant differences between the processes is most obvious in the fact that *P. vivax* displays a preference for invading reticulocytes. The commitment to invading reticulocytes is mediated at a minimum by *P. vivax* reticulocyte-binding proteins 1 and 2 (RBP-1 and RBP-2) (Galinski et al., 1992), although completion of the *P. vivax* genome made it clear that there are several other RBP homologues that may be involved (Carlton et al., 2008). The reticulocyte-specific receptor for these proteins remains unknown.

Only 1 receptor-ligand interaction has been extensively studied in *P. vivax*: that between *P. vivax* Duffy Binding Protein (PvDBP) and the human erythrocyte surface protein Duffy Antigen Receptor for Chemokines (DARC, also referred to as the Duffy antigen or Fy). DBP was first discovered to be the ligand for DARC-dependent invasion through the *P. knowlesi* homolog, PkDBP (Haynes et al., 1988), and subsequently confirmed for *P. vivax* by Wertheimer and Barnwell showing that pre-incubation with purified DARC glycoprotein blocked PvDBP protein binding to erythrocytes (Wertheimer and Barnwell, 1989). Using PkDBP as a probe, PvDBP was cloned in Louis Miller's laboratory in 1991 (Fang et al., 1991), and the binding to DARC was found to depend on the cysteine-rich region II (DBP-RII or DBP-II) (Chitnis and Miller, 1994). DBP interaction with DARC occurs after the initial merozoite attachment and reorientation, and may precipitate the formation of a tight junction between the merozoite and reticulocyte membranes which then moves along the surface of the merozoite (Galinski and Barnwell, 1996). DBP interaction with DARC is a dynamic process where two DBP-II domains initially form a heterotrimer with a single DARC receptor, then recruit a second DARC receptor, forming a heterotetramer (Batchelor et al., 2011, Paing and Tolia, 2014, Malpede and Tolia, 2014, Batchelor et al., 2014).

Duffy negativity, which is highly prevalent in sub-Saharan Africa, was thought to provide complete protection against *P. vivax* infection (Miller et al., 1976). However, the universality of this protection is becoming increasingly disputed, most notably by recent work showing that *P. vivax* infects Duffy-negative individuals in Madagascar (Menard et

al., 2010). Whether this is due to DBP-dependent invasion using a non-Duffy pathway, or represents a truly alternative DBP-independent invasion pathway is not currently known. Genome sequencing of additional *P. vivax* strains has led to the discovery of a DBP paralog (PvEBP) that was missing in the *P. vivax* Sal 1 isolate, which has the hallmarks of an erythrocyte binding protein, including a DBL and cysteine-rich C-terminal domain (Hester et al., 2013). However, the function of the gene remains unknown. The complete sequencing of 1 Malagasy *P. vivax* strain led to the discovery that a duplication of *DBP* was present in over half of Malagasy *P. vivax* infections while the duplication appears to be largely absent elsewhere (Menard et al., 2013). Madagascar has the highest rate of Duffy-negative infections reported thus far, which account for nearly 9% of *P. vivax* infections detected by PCR (Menard et al., 2010), leading to the hypothesis that the *P. vivax* infection of Duffy-negative individuals is connected to *DBP* duplication. But it is currently unknown whether *DBP* duplication is enriched in *P. vivax* infections in Duffy-negative individuals.

1.4 Natural immunity to *Plasmodium* infections

Increasing our knowledge about the biology of *P. vivax* invasion has the potential for identifying new vaccine candidates. This is potentially augmented by understanding the targets of naturally-acquired immunity (NAI) that develops during infections. Humans experiencing repeated infections with *Plasmodium* frequently develop immunity, if not from infection, then from symptoms and the potential complications of severe disease. Understanding the development of antibody-based immunity and the specific antigens targeted, which effectively control infections, may inform and further prioritize candidates for vaccine development.

As reviewed by Doolan et al., one of the first studies of induced immunity against *Plasmodium* came with experimental “malariotherapy” as a treatment for neurosyphilis in 1917 by von Wagner-Jauregg. Patients diagnosed with the disease were intentionally infected with *Plasmodium* parasites of different species leading to the cure of 1 in 3 patients through the fevers associated with the resulting malaria episode. NAI in the adult subjects was observed to occur sometimes after 1 exposure, though often requiring multiple infections. Protection to *P. falciparum* appeared to be acquired more slowly than to *P. vivax*. Immunity also appeared to be species-specific, and protective immunity to at

least *P. vivax* did not persist for long periods after treatment [reviewed in (Doolan et al., 2009)]. Based on the work from Brown and Brown in 1965 on antigenic variation in *P. knowlesi*, a theory of immunity development emerged (Brown and Brown, 1965). Repeated exposures against a wide repertoire of antigenic variants were required to develop protective memory against a highly antigenically variable parasite population.

This cumulative-exposure model (strain-specific immunity acquisition) has been questioned by Doolan et al., however. They reviewed data that age is a more significant factor for acquiring immunity, which is inversely protective depending on whether exposure is acute or chronic. In acute scenarios where non-immune adults and children are exposed to *P. falciparum* for the first time, adults are more susceptible to severe disease than children. In contrast, in settings with chronic exposure, adults will develop immunity and protection from severe disease faster than children. Age-dependent (which may differ from cumulative exposure-dependent) acquisition of immunity appears to be very important in achieving “antiparasite immunity” in *P. falciparum* settings. The fact that adults can develop robust protective and strain-transcending immunity to clinical disease after relatively few episodes (3-4) has potential positive implications for vaccine design, as antigen variation may not represent a barrier to cross-protection against a diverse set of parasite clones (Doolan et al., 2009).

1.4.1 Immunity development to *P. falciparum*

As with most studies in malaria, the research into the development of NAI is much more advanced for *P. falciparum* than for *P. vivax*. Protective immunity to *P. falciparum* develops slowly with repeated exposure and involves the development of a protective IgG response [reviewed in (Langhorne et al., 2008, Doolan et al., 2009)]. In general, protection is defined as the lower risk of clinical disease through the absence of fever (axillary temperature $>37.5^{\circ}\text{C}$) even if parasites are present in the blood. In holoendemic areas across sub-Saharan Africa, populations are often continuously exposed and first develop “antidisease immunity” relatively quickly in young childhood, with significant protection from severe disease complications and death even with higher parasite densities. Subsequently, a more slowly acquired “antiparasite immunity” develops, in which parasite densities are significantly reduced with additional protection from disease. Clearing parasites while limiting host pathology involves a series of pro- and anti-inflammatory or regulatory cytokine responses. An early pro-inflammatory response with

increased tumor necrosis factor (TNF)- α and interferon (IFN)- γ leads to fast clearance of parasites (Kremsner et al., 1995, Walther et al., 2006, D'Ombra et al., 2008) while an increase in regulatory cytokines such as interleukin (IL)-10 and transforming growth factor (TGF)- β is then important for limiting severe disease (Day et al., 1999, Othoro et al., 1999, Perkins et al., 2000).

NAI is not sterilizing, however, and asymptomatic adults with low parasite densities are the norm in high-transmission areas. The most high-risk periods for complications from malaria are from about age 3 months to 5 years as children experience exposure to repeated malaria episodes. Women experiencing first and second pregnancies are also particularly vulnerable due to immunosuppression and the potential exposure to a new *var* gene variant (VAR2CSA) which mediates cytoadherence to chondroitin sulphate A in the placenta (Beeson et al., 2002, Fried et al., 2006).

While several aspects of the human immune response are under active study, the development of IgG antibodies remains the only immune response definitively shown to protect in human studies (Cohen et al., 1961). Protective IgG responses are relevant for vaccine design and there are numerous studies evaluating whether IgG to specific *P. falciparum* proteins correlate with protection. The focus of many studies has been the cytoadherence-mediating PfEMP1, and several studies suggest that the majority of a protective IgG response is directed against this family of proteins and leads to variant-specific protective immunity (Horrocks et al., 2004, Duffy et al., 2001, Kraemer and Smith, 2006). But the development of PfEMP1 as a vaccine candidate is likely limited due to the significant variation and frequent recombination present in the *var* gene repertoire (Claessens et al., 2014). Merozoites also appear to be common targets of NAI, as passive transfer of immune serum following schizogony significantly reduced parasitemia (with apparently little effect on developing or mature trophozoites) (McGregor, 1964). More recently, the search for correlations with protection has been expanded into the evaluation of IgG directed against a panel of *P. falciparum* merozoite proteins in a Kenyan cohort study (Osier et al., 2014). The results concluded that the breadth of antibody response rather than the response to any single antigen was a significant predictor for protection from clinical disease. These findings suggest that a multivalent vaccine with multiple antigen targets may be a winning vaccine strategy rather than a strategy that relies on a single, partially protective antigen.

1.4.2 Immunity development to *P. vivax*

Research into the development of NAI to *P. vivax* is far less advanced, and was recently reviewed in detail (Longley et al., 2016). In contrast to *P. falciparum* rates in Africa, the majority of *P. vivax*-endemic areas experience lower transmission rates. Consequently, populations in low-transmission areas are not repeatedly exposed, and people of all ages experience clinical disease. The exception to this is Papua New Guinea (PNG) where transmission of both *P. falciparum* and *P. vivax* can reach rates as high as hyperendemic regions in Africa. In higher-transmission co-endemic settings such as these, the incidence of *P. vivax* infection peaks at an earlier age than *P. falciparum* (Michon et al., 2007, Maitland et al., 1996, Phimpraphi et al., 2008, Mueller et al., 2009b, Lin et al., 2010). This corroborates reports from experimental infections showing that immunity to *P. vivax* develops more quickly than to *P. falciparum* (Jeffery, 1966, Ciuca, 1934). *P. vivax* infections are frequently polyclonal, and recent work in PNG by the Mueller laboratory suggests this speed of immunity acquisition may be partly due to the greater force of exposure to genetically distinct parasites that occurs in *P. vivax* infections, referred to as the molecular force of blood-stage exposure ($M_{\text{ol}}\text{FOB}$) (Koepfli et al., 2013).

1.4.2.1 Naturally-acquired cellular immunity to *P. vivax*

Naturally-acquired cellular immunity to *P. vivax* infection has been the subject of limited study and frequently with contradictory findings. It has been known for decades that T cells are essential for eliminating *Plasmodium* parasites during liver stages (Hoffman et al., 1989, Weiss et al., 1990, Renia et al., 1993) but whether any cellular responses are directed at dormant *P. vivax* hypnozoites is unknown. The induction of cytokines to *P. vivax* infection has been evaluated most commonly through measuring cytokines in plasma [reviewed in (Longley et al., 2016)], which eliminates the possibility of determining the cellular source of the cytokines. As with *P. falciparum* infection, *P. vivax* infection induces a pro-inflammatory TNF response (Karunaweera et al., 1992), which has been shown to limit infection for *P. falciparum* (Kremsner et al., 1995) but with the dual potential for inducing immunopathology or severe disease (Kwiatkowski et al., 1993, Grau et al., 1989). Multiple subsequent studies have shown the induction of strong pro-inflammatory responses to *P. vivax* infections, such as TNF, IF γ , IL-12, IL-16, IL-1 β and IL-8 [reviewed in (Longley et al., 2016)]. It is largely accepted that *P. vivax* pro-inflammatory cytokine responses are higher at lower parasitemias than during *P. falciparum* infections, thus producing fevers with far fewer parasites (Price et al., 2007).

But the available data is in conflict with several studies finding comparatively higher levels of pro-inflammatory cytokines per parasite during *P. vivax* infections compared to *P. falciparum* infections (Hemmer et al., 2006, Karunaweera et al., 1992), while several other studies find no difference during infection with either species either by measuring cytokines per parasite or comparing cytokine levels for similar parasitemias (Fernandes et al., 2008, Goncalves et al., 2012, Rodrigues-da-Silva et al., 2014). These differing findings may be the result of a variety of factors including subject age, transmission level, location, stage of infection and methods used for measuring cytokines, suggesting the need for standardized assays to resolve the conflicting data (Longley et al., 2016).

The induction of anti-inflammatory or regulatory cytokine IL-10 during *P. vivax* infection has been reported in numerous studies [reviewed in (Longley et al., 2016)], with most studies reporting that levels of IL-10 were higher during *P. vivax* infection than *P. falciparum* infection (Praba-Egge et al., 2003, Fernandes et al., 2008, Goncalves et al., 2012, Yeo et al., 2010). IL-10 has been shown to have a protective effect against experimental cerebral malaria in mice with *P. berghei* infections (Kossodo et al., 1997). The connection between IL-10 levels and severe disease in *P. vivax* is in conflict with one study reporting lower IL-10 levels in severe compared to asymptomatic *P. vivax* cases (Andrade et al., 2010) and several other studies reporting the opposite finding (Goncalves et al., 2012, Mendonca et al., 2013). Additional research is needed to clarify the role of immunoregulatory signals in severe, symptomatic and asymptomatic *P. vivax* infections.

1.4.2.2 Naturally-acquired humoral immunity

Antibody responses to individual *P. vivax* antigens have been evaluated in numerous immunoreactivity studies, but studies evaluating protective associations are few. A recent review of *P. vivax* immunoepidemiological studies found only 3 antigens – MSP1, MSP3.10 (MSP3 α), and MSP9 – were consistently associated with protection (Cutts et al., 2014). DBP has also been a central target in clinical protection studies given its near universal requirement for invasion of reticulocytes, with some exceptions, which have been discussed elsewhere in this introduction, for example (Menard et al., 2010). DBP elicits antibody responses in naturally-infected individuals as first reported by Fraser et al. in 1997, and it was later determined that antibodies directed against the RII domain can block binding to erythrocytes (Fraser et al., 1997, Michon et al., 2000). Antibody levels to

DBP-RII have also been shown to increase with age (Michon et al., 1998, Xainli et al., 2003), which supports the potential development of protective immunity to DBP-RII in natural infections. The general sparseness of data evaluating correlations with protection underline the great need for expanding the list of proteins studied.

1.5 Approaches to a *Plasmodium* vaccine development

Control measures combatting *Plasmodium* have proven effective at reducing malaria deaths over the last decade. But these gains are threatened by the development of artemisinin resistance in *P. falciparum* in Southeast Asia and insecticide resistance in *Anopheles* vectors worldwide. A vaccine targeting *Plasmodium* parasite development is a universally recognized essential component of any global malaria eradication campaign. *P. vivax* vaccine development lags significantly behind *P. falciparum* due to both financial resources and the lack of critical information surrounding *P. vivax* biology.

1.5.1 Vaccine targets at different stages of the life cycle

Vaccine development efforts have focused on several parasite life stages with different goals for blocking parasite development. The majority of efforts have historically been directed at the asexual blood stages for several reasons. This section of the life cycle causes the symptoms of malaria, and blocking merozoite development and invasion would halt the disease progression. It is also the stage most tractable to study, as *P. falciparum in vitro* cultures provide a ready source of parasites. Merozoites are exposed to the immune system briefly after schizont egress and erythrocyte invasion, and IgG responses directed against merozoite proteins have shown correlations with protection. Early experiments suggested that the asexual-blood stages are prime targets of NAI, as serum transfers from immune donor patients led to rapid declines in parasitemia for recipient patients (Boyd, 1939). A vaccine mimicking this natural response may prove to be a useful tool (Greenwood and Targett, 2011).

Targeting the asexual blood stage would not, however, prevent initial sporozoite infection of hepatocytes, and additional vaccine development efforts specifically focus on the pre-erythrocytic-stages. Sporozoites are exposed to the human immune system from bite through to hepatocyte invasion, a process estimated to take at least 15 minutes, and may represent an attractive target for preventing initial infections from taking hold. Evidence

as far back as 1946 from a neurosyphilis patient inoculated with infectious bites and with salivary-gland emulsion demonstrated that pre-erythrocytic forms could be the target of induced immunity (Covell and Nicol, 1951). The two most advanced vaccine candidates for *P. falciparum* target this stage and are discussed in more detail below.

A third strategy in vaccine design is termed “transmission-blocking” and specifically targets the sexual stages. The primary goal in this approach is to prevent parasite fertilization in the mosquito by either targeting surface antigens on gametocyte-infected cells (thus targeting them for degradation) or antigens directly on the surface of sexual forms (gametocytes, gametes, ookinetes). In the latter scenario, mosquitoes would ingest not only gametocytes but also human antibodies blocking critical processes for maturing in mosquitoes. While such a vaccine would have no effect on the progression of illness in the current human host, it would effectively limit the spread of the disease to future hosts. Such an approach may be important as gametocytes in *P. falciparum* can often persist for several weeks and would not be targeted by a blood-stage vaccine administered after symptoms appear. Gametocytes appear almost simultaneously with *P. vivax* asexual forms, potentially indicating their development in liver stages and therefore the possibility of transmission even with protection from a blood-stage vaccine.

Overall, each strategy has some benefits and potential limitations, and it is likely that a combination of candidate antigens targeting multiple stages would provide the greatest chance of success in halting disease and transmission.

1.5.2 *P. falciparum* vaccine development

The majority of *Plasmodium* vaccine efforts have focused on *P. falciparum*, and were historically aimed at the asexual blood stages. The vast majority of initial work focused on the merozoite surface proteins MSP1 and AMA1, but phase II trials have failed to demonstrate significant protection (Geels et al., 2011, Schwartz et al., 2012). These proteins are highly abundant on the merozoite surface and experience significant immune pressure, resulting in a high degree of polymorphism. Antibodies directed against them have not proven to provide cross-protection with other haplotypes, and their usefulness as vaccine candidate may be limited (Schwartz et al., 2012, Hill, 2011, Geels et al., 2011). There are currently 7 merozoite proteins approved for clinical testing: MSP1, MSP2, MSP3, AMA1, EBA175, GLURP, and SERA5 (Schwartz et al., 2012). More recently, antibodies generated against several recombinant merozoite proteins, including *P.*

falciparum RH5, CyRPA, and RIPR have been shown to block invasion, and are all potential vaccine candidates under development (Chen et al., 2011, Reddy et al., 2015). As transmission-blocking vaccine candidates, *P. falciparum* Pfs25, a GPI-anchored surface protein of mosquito-stage zygotes and ookinetes, is the most active candidate (Table 1.1). Rabbits immunized with recombinant forms of Pfs25 and *P. vivax* Pvs25 were shown to block transmission to mosquitoes in membrane-feeding assays (Miura et al., 2007).

There are two pre-erythrocytic *P. falciparum* vaccine candidates in advanced testing, RTS,S and irradiated sporozoites, each of which has drawbacks. RTS,S targets the circumsporozoite protein (CSP) present on the sporozoite surface and has demonstrated only moderate efficacy, with a 36% reduction in malaria episodes but no effect on mortality (RTS, 2015). This represents a useful first step in reducing the number of malaria episodes in highly-endemic regions in Africa, but is likely to have little overall effect in reducing transmission. Improvements in efficacy will be absolutely crucial to making significant changes to the global malaria burden. The second most advanced vaccine involves the intravenous injection of irradiated sporozoites. It has thus far proven to be much more efficacious, with 6/6 participants receiving 5 doses protected by subsequent challenge with *P. falciparum* (Seder et al., 2013). However, requirement for manual dissection of mosquitoes, liquid-nitrogen storage, and multiple intravenous inoculations render the worldwide rollout of this vaccine unlikely (Garcia et al., 2013). While these are incredibly useful first steps, it is likely that additional candidates will be needed to produce a cost-effective, widely distributable, and efficacious vaccine.

1.5.3 *P. vivax* blood-stage vaccine candidate: Duffy binding protein

Very few *P. vivax* candidates are in preclinical or clinical vaccine trials, including only DBP (Phase 1a) and CSP (preclinical) (Table 1.1). The significant role that DBP plays in *P. vivax* invasion of reticulocytes makes it the prime blood-stage vaccine candidate. Previous studies have shown that antibodies against the recombinant DARC-binding DBP region II (DBPII) inhibit merozoite binding and invasion of erythrocytes (Fraser et al., 1997, Michon et al., 2000, Singh et al., 2001, Grimberg et al., 2007). However, the DBPII under consideration is highly polymorphic and under immune pressure (VanBuskirk et al., 2004a, Cole-Tobian et al., 2002), and antibody responses to this domain have been shown to be strain-specific (Ceravolo et al., 2009). Epitopes critical to binding between

DBP and DARC are highly conserved (Batchelor et al., 2011, Singh et al., 2006), suggesting that DBP remains a promising vaccine target. This is further supported by a longitudinal cohort treatment re-infection study in Papua New Guinea showing that naturally-acquired high titer antibodies directed at DBP blocked DBP-DARC binding and were associated with protection from *P. vivax* infection (King et al., 2008). While the diversity of the DBP domain makes the development of a strain-transcending vaccine a complex task, strategies to overcome this challenge are being actively pursued, most notably through the use of a synthetic DBP-based vaccine candidate, termed DEKnull, which lacks an immunodominant variant epitope (Ntumngia et al., 2014, Chen et al., 2015).

1.5.4 Other blood-stage candidates

A limited number of additional proteins have been investigated for their potential as vaccine candidates [reviewed in (Longley et al., 2016)]. These include the *P. vivax* circumsporozoite protein PvCSP (currently in preclinical testing), PvAMA1, PvRBP1, PvRBP2, and PvMSP₁₋₁₉ (Kocken et al., 1999, Galinski et al., 1992, Galinski et al., 2000, Galinski and Barnwell, 1996, Rogers et al., 1999). In the case of a recombinant fragment of RBP1, antibodies directed against the fragment were high but were not protective against infection in an *Aotus* monkey *in vivo* model (Caraballo et al., 2007). More recently, antibody responses to MSP3, MSP9, several TRAg, Virs, and gametocyte antigen 1 (GAM1) were evaluated [reviewed in (Longley et al., 2016)]. A meta-analysis of the available studies by Cutts *et al.* found correlations with protection consistently with MSP3.10 (MSP3 α), MSP9, and MSP1 (Cutts et al., 2014). However, limited progress beyond initial studies has been made for any candidates thus far.

Table 1.1: Preclinical and clinical *Plasmodium* vaccine candidates*

Preclinical	Clinical				
	Translational projects			Vaccine candidates	
	Phase 1a	Phase 2a	Phase 1b	Phase 2b	Phase 3
CSP	ChAd63/MVA ME-TRAP + Matrix M	RTS,S-AS01 fractional dose	R21/Matrix-M1	ChAd63/MVA ME-TRAP	RTS,S-AS01E
<i>P. vivax</i> CSP	PfCelTOS FMP012	PfSPZ-CVac (PfSPZ Challenge + chlor. or + chlor./pyrim.)	AMA1-DiCo	<i>Pf</i> SPZ	
VAR2CSA	R21/AS01B	PfSPZ-CVac (PfSPZ Challenge + chloro.)	P27A		
EBA175/Rh5	PfPEBS	FMP2.1/AS01B	SE36		
RH5.1	ChAd63 RH5 +/- MVA RH5	M3V-D/Ad- PfCA (CSP/AMA1 or CSP/AMA1/SSP2/TRAP)	PRIMVAC (VAR2CSA)	<i>P. falciparum</i>	<i>P. vivax</i>
Pfs 48/45	ChAd63/MVA PvDBP			Pre-erythrocytic	Pre-erythrocytic
	<i>Pfs</i> 25-VLP			Blood-stage	Blood-stage
Pfs25	ChAd63 Pfs25-IMX313/MVA Pfs25-IMX313			Transmission-blocking	
Combo: PE: R21, ME-TRAP E:RH5 TB: Pfs230 and Pfs25	Pfs230D1M-EPA/Alhydrogel and/or Pfs25-EPA/Alhydrogel			Combination	

*Compiled from WHO Rainbow Table of malaria vaccine projects, updated March 2016 (http://www.who.int/immunization/research/development/Rainbow_tables/en/)

1.5.5 Future *P. vivax* vaccine development

DBP remains the most promising *P. vivax* vaccine candidate. However, strong reliance on this single target may prove problematic for successful vaccine design for several reasons. First, as discussed, DBP is highly polymorphic and strain-transcending protection may be difficult to achieve. Second, corollaries in *P. falciparum* vaccine studies targeting single blood-stage antigens, such as AMA1 and MSP1, have failed to generate protective immune responses (Sagara et al., 2009, Ogutu et al., 2009). It is possible that a cumulative immune response to a multivalent vaccine will be needed to produce protective immunity.

Identifying new protein candidates for vaccine trials will be important for ultimate success. The whole genome sequencing and transcriptome studies of both *P. falciparum* and *P. vivax* supplied a long list of unexplored blood-stage candidates. But functions remain unclear for the vast majority of proteins. Considering a panel of candidates in high-throughput functional and immunoepidemiological screens could help to address this by prioritizing candidates for further investment.

1.6 Characterization of *Plasmodium* vaccine antigens

1.6.1 Producing *Plasmodium* vaccine antigens

The current state of vaccine development for *P. vivax* underlines the need for additional proteins to evaluate as vaccine candidates. Given the intractability of *P. vivax* to *in vitro* culture, it is largely impossible to obtain quantities of protein necessary to study and functionally test any desired candidates. New candidate antigens therefore must be expressed and studied as recombinant proteins.

Several systems have been used previously for producing recombinant *Plasmodium* proteins, which are reviewed in (Birkholtz et al., 2008). Most recombinant protein expression aimed at vaccine development has focused on *P. falciparum*, for which protein expression is particularly difficult due to the high AT-content, biased codon-usage, and the presence of repetitive amino acid sequences (Tsuboi et al., 2008). Membrane-bound proteins are often the focus of vaccine research, given their potential roles in invasion. Such proteins are difficult to express in properly-folded and soluble form, and thus most research has focused on producing soluble extracellular ectodomains. The most widely

used and inexpensive system remains *E. coli*, despite frequent issues with solubility that necessitate refolding procedures. The process for obtaining soluble, properly-folded protein fragments in *E. coli* often requires protein-specific protocols, and is therefore limited for high-throughput applications (Birkholtz et al., 2008).

Yeast expression systems (*Saccharomyces cerevisiae*, *Pichia pastoris*) are also frequently used to express vaccine candidates including *P. falciparum* EBA175, AMA1, MSP1, MSP3, and Pfs25 [reviewed in (Birkholtz et al., 2008)]. These systems have very high yield (10-75 mg/l) and overcome many of the solubility issues of the *E. coli* systems by linking expression constructs to an N-terminal yeast pheromone that directs protein secretion into the culture supernatant. As a eukaryotic system producing eukaryotic proteins, folding and post-translational modifications are likely to be much improved over the *E. coli* system. However, this is a drawback in the case of N- and O- linked glycosylation, which do not significantly occur in *Plasmodium*. Disulfide bonding has been shown to be heterogeneous, which may impact functional studies (Stowers et al., 2001). The *S. cerevisiae* system also requires the use of codon-optimization (at least in the case of *P. falciparum* proteins), as certain A-T containing codons are recognized as stop codons, a particular problem in the A-T-rich *P. falciparum* genome.

The baculovirus-infected insect cell system has also been applied for *Plasmodium* vaccine candidate expression, including fragments of *P. falciparum* MSP1, CSP, PfEMP1, EBA175, and SERA [reviewed in (Birkholtz et al., 2008)]. It may have advantages in proper protein folding over yeast expression systems as all of the listed proteins elicited antibody responses, while a yeast-expressed *P. falciparum* MSP1 contained conformational changes that rendered it immunologically inactive (Chang et al., 1992). Several other *P. falciparum* proteins (EBA175, SERA) also proved to be functionally active or were processed into fragments mirroring *in vivo* fragment sizes (Li et al., 2002, Daugherty et al., 1997).

Cell-free expression systems have been on the rise more recently and have several additional advantages over previous expression systems. The most common system involved using ribosomes, translation factors, and post-translational components taken from wheat embryos (Farrokhi et al., 2009). The system is well established for producing *Plasmodium* proteins including the *P. falciparum* vaccine candidates Pfs25, CSP, and AMA-1 (Tsuboi et al., 2008, Crompton et al., 2010, Trieu et al., 2011). The main

advantage of the system is in producing properly-folded proteins in plate format, thus facilitating the use of downstream functional or immunoepidemiological screens. For example, the wheat germ cell-free system was recently used to express a panel of 89 (Chen et al., 2010) or 143 (Lu et al., 2014) *P. vivax* recombinant protein ectodomain fragments; subsequent screening using malaria-exposed patient plasma from Korea established that at least a subset (19% and 28%) of these proteins was highly immunoreactive. The primary drawbacks of the system are that reagents and running costs are high and obtaining extracts is difficult, which limits the usage of this technology in smaller individual laboratories (Farrokhi et al., 2009).

Mammalian expression systems are well characterized and support proper folding of eukaryotic proteins, but were not used significantly to produce recombinant *Plasmodium* proteins due to the low yields achieved in adherent cultures (Birkholtz et al., 2008). This drawback has been largely overcome, however, with the development of liquid cultures leading to protein production on the milligram to gram range (Tom et al., 2008). The most common cell line used for large scale recombinant protein production is the human embryonic kidney 293 (HEK293E) line, which stably expresses the Epstein-Barr virus nuclear antigen 1 (Tom et al., 2008). Prior work in the Wright and Rayner laboratories established the HEK293E expression system for the successful expression of high-quality *P. falciparum* proteins that were functional and immunogenic (Crosnier et al., 2011, Taechalertrpaisarn et al., 2012, Osier et al., 2014).

1.6.2 High-throughput screening of vaccine antigens

Fully characterizing the function of potential vaccine antigens and their immunological underpinnings is an important next step for expanding *P. vivax* vaccine development. Most vaccine research aims at targeting candidates important for host cell interactions, as blocking such interactions halts the invasion of parasites into host cells (i.e., sporozoites into hepatocytes or merozoites into erythrocytes). Interactions between parasites and host cells are often transient and low-affinity, and therefore difficult to detect (Wright, 2009, Bei and Duraisingh, 2012). Interaction and immunoepidemiological screens have traditionally utilized individual assays with individually purified proteins, often with limited ability to scale-up for protein libraries (Wright, 2009, Bei and Duraisingh, 2012). This type of approach often involves a large investment of resources and time focused on

a specific candidate, based on some prior knowledge or predicted homology-based function, which is often the case for *P. vivax*.

This individual approach is frequently used in mammalian expression system, cell-based binding assays to identify *Plasmodium* merozoite proteins and the sub-domains therein that mediate erythrocyte binding. This is reviewed in Birkholtz et al., and the list includes *P. falciparum* EBA175, EBA181, MAEBL, MSP1, AARP, and *P. vivax* DBP [e.g., (Sim et al., 1994, Tolia et al., 2005, VanBuskirk et al., 2004b, Han et al., 2004, Mayer et al., 2009, Mayer et al., 2004, Wickramarachchi et al., 2008)]. These have primarily utilized COS-7, a green monkey kidney cell-line, but have also used CHO-K1 and HeLa cell lines. This approach was used for determining the region of binding of *P. vivax* DBP to erythrocytes. DBP fragments were expressed as fragments embedded in the membrane of COS cells and incubated with erythrocytes. Binding events were visible as clusters of erythrocytes (rosettes) around COS cells. The method was highly effective for determining the required region II domain necessary for binding (VanBuskirk et al., 2004b). These approaches require significant resources and labor devoted to single candidates, however, which is a significant limitation in studying a larger protein library.

An alternative approach to this individual protein strategy is to test a panel of potential proteins in high-throughput interaction screens and immunoepidemiological studies. Evaluating proteins in parallel would prove a less biased approach, as proteins produced on the same platform could be compared in a more systematic way. Indeed, this was the goal for the panels of *P. vivax* protein fragments produced in the wheat germ cell-free system (described in the previous section) most recently used to evaluate general reactivity in Korean isolates (Lu et al., 2014, Chen et al., 2010). Large-scale yeast two-hybrid interaction screens were tested in the mid-2000s to identify potential *P. falciparum* interactions (LaCount et al., 2005), but these studies often used protein fragments with proteins potentially lacking proper conformations, which was a significant limitation.

The protein library approach was developed recently in the Wright and Rayner laboratories to screen a panel of *P. falciparum* proteins against a library of erythrocyte receptors utilizing the HEK293E expression system. The screen uncovered the *P. falciparum* Rh5 interaction with the erythrocyte receptor basigin, which was an essential interaction for invasion in all parasite isolates tested (Crosnier et al., 2011). The primary tool used in the work involved an avidity-based extracellular interaction screen

(AVEXIS) assay which reliably detects highly-transient interactions with half-lives ≤ 0.1 seconds (Figure 1.6)(Bushell et al., 2008); we adapted this assay to screen a *P. vivax* recombinant protein library (discussed in Chapter 4).

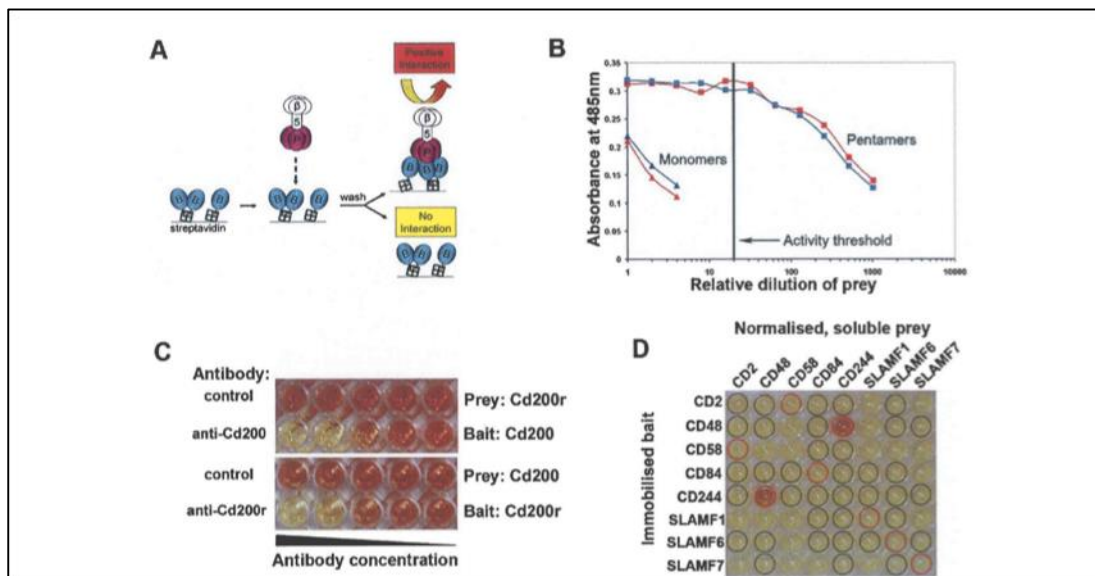


Figure 1.6: AVEXIS-based EXtracellular Interaction Screen (AVEXIS)

(A) Depiction of AVEXIS showing biotinylated bait proteins bound to streptavidin-coated plates. Baits are probed by pentamerized prey proteins tagged with β -lactamase. Interactions are detected by nitrocefin turnover resulting in a color change from yellow to red. (B) Detection is improved using pentamerized preys over monomeric preys. (C) Interactions can be blocked in a dose-dependent manner using inhibitory antibodies. (D) The assay has a low false-positive rate with expected interactions highlighted in red and known non-interactors highlighted in black. Figure reprinted from (Bushell et al., 2008) under the Creative Commons Attribution (CC BY) license.

1.7 Specific Aims

The fact that only a relatively limited repertoire of potential *P. vivax* antigens are currently being considered for vaccine development is due in large part to a lack of understanding of the molecular details of *P. vivax* biology in general, and invasion in particular, and the fact that there have been few large studies of immunological responses generated during *P. vivax* infection, or assessment of whether particular responses correlate with protection. A comprehensive and systematic approach is needed to advance our understanding of *P. vivax* invasion and to identify additional vaccine candidates. My project aimed to enhance our basic understanding of the proteins important for *P. vivax* invasion of erythrocytes, and the development of immunity against those proteins.

My first aim was to study the transcription of *P. vivax* parasites just prior to invasion, which is discussed in Chapter 3. This required the testing and development of RNA extractions for *P. vivax* clinical samples from Cambodia and the bespoke production of strand-specific Illumina RNA-Seq libraries. These data would be used to detect genes with increased transcription in the late blood stages, and hence identify novel vaccine targets. Additionally, it would enable the confirmation or correction of existing gene model predictions in the *P. vivax* reference genome, the prediction of 5' and 3' untranslated regions, and potentially uncover novel gene transcripts and alternative splicing events.

My second aim was to produce a library of blood-stage *P. vivax* proteins to study in functional and immunoepidemiological studies, which is described in Chapter 4. The prioritization of candidates would be enhanced by RNA-Seq abundance data, and would focus on proteins with known or predicted localization to the merozoite surface or invasive secretory organelles, and on homology to *P. falciparum* vaccine candidates. The mammalian expression system, HEK293E cells, has previously been used successfully to express *P. falciparum* proteins and would provide the basis for the development of the library. High-throughput interaction screens developed in the Wright and Rayner laboratories would provide the foundation for investigating protein function.

Next I wanted to utilize the *P. vivax* merozoite protein library to conduct immunoepidemiological studies discussed in Chapter 5. Understanding general protein reactivity and correlation with either exposure or protection from clinical disease can help to prioritize candidates for further functional studies as vaccine candidates. Prior studies have often focused on one or several full-length recombinant proteins, and utilizing a panel of proteins will enable more systematic conclusions to be drawn. Ultimately patient collections from 3 *P. vivax*-endemic countries (Cambodia, Solomon Islands, and Papua New Guinea) were screened to further our understanding of naturally-acquired immunity to *P. vivax*.

This project aims to add significant new knowledge to our understanding of *P. vivax* invasion and the development of natural immune responses to infection. The *P. vivax* recombinant protein library will not only serve as a resource for the scientific community, but also aims to expand the list of potential blood-stage vaccine targets.

2 GENERAL METHODS

Several general methodologies were key to the completion of this thesis. In this chapter, I will describe those methods. Publication note: the methods described in sections 2.3.1, 2.3.3, 2.4.1, 2.4.3, 2.5.2, 2.6.1, 2.6.2, and 2.7.1.3 were slightly modified from a previously published manuscript (Hostetler et al., 2015). I drafted the text described in these sections, which was edited by co-authors prior to publication. I am solely responsible for the work described in sections 2.1-2.7.1, under the supervision of my PhD supervisors, Rick Fairhurst and Julian Rayner, except where noted in the text. The experiments in section 2.7.2 were performed in close collaboration with Camila Franca as specified in the text. Camila Franca performed all experiments in section 2.7.3. The results from sections 2.7.2-2.7.3 were recently published (Franca et al., 2016).

2.1 *P. vivax* schizont transcriptome sequencing

2.1.1 Schizont enrichment of *P. falciparum* samples for testing

To compare RNA extraction methods for isolating RNA from *P. vivax* field samples, mock samples of *P. falciparum* laboratory strains were used. *P. falciparum* cultures (provided by Jennifer Volz) were enriched for schizonts to compare different RNA extraction methods. Five and 10 ml cultures were pelleted (800 RCF, 4°C, 5 min), resuspended in 10 ml of HI-FBS complete media (Table 2.4), pelleted and washed a second time. The pellet was resuspended in HI-FBS complete media to 10% hematocrit, and separated by centrifugation (1500 RCF, 4°C, 20 min, no brakes) on a 60% Percoll® gradient (12.5 ml solution/1 ml erythrocyte pellet at 4°C). The interface containing

schizonts was collected into an RNase-free tube with 20 ml PBS per 1 ml initial pellet. Schizonts were pelleted (800 g, 4°C, 5 min, no brakes), washed with PBS (20 ml/1 ml starting pellet), and pelleted. RNeasy Lysis Buffer (Ambion) was added to isolated schizonts (at 2:1 initial pellet size), and samples were stored at 4°C overnight followed by storage at -20°C until samples were combined for RNA extraction experiments.

2.1.2 Comparison of RNA extraction methods

Schizont-enriched *P. falciparum* parasites in RNeasy Lysis Buffer (Ambion) were combined and divided equally (1.5 ml each) to compare 3 RNA extraction methods. All samples were initially pelleted (8000 RCF, 1 min) and RNeasy Lysis Buffer (Ambion) supernatant was discarded. Samples 1 and 2 were extracted using the RNeasy Plus Mini kit (Qiagen) or the RiboPure Blood kit (Ambion), respectively, according to manufacturer instructions. Sample 3 was extracted using a TRIzol (Invitrogen) extraction protocol modified from Kyes, et al (Kyes et al., 2000). The schizont pellet was lysed by adding TRIzol (Invitrogen, 500 µl incubated at 37°C, 5 min), followed by extraction with chloroform (added 250 µl, mixed well, incubated at RT, 3 min). The sample was centrifuged (8000 rpm, 4°C, 30 min) and the aqueous layer was removed to a new tube, mixed with isopropanol (250 µl), and incubated (4°C, 2 h). RNA was pelleted (14000 rpm, 4°C, 30 min), and the pellet was then washed with 75% ethanol (500 µl), centrifuged (14000 rpm, 4°C, 5 min), and the supernatant was discarded. The sample was air dried (5 min), resuspended in DEPC-H₂O (Ambion), and heated to solubilize the RNA (65°C, 5 min). Extracted samples were all subjected to DNA digestion using the DNA-free kit (Ambion) according to manufacturer instructions and analyzed by Bioanalyzer® (Agilent Technologies, Inc.) using an RNA Nano Chip for RNA quantity and quality.

2.1.3 Field isolate collection and enrichment for schizonts

The collection of *P. vivax* clinical isolates from patients was performed as part of an ongoing NIAID-approved *P. vivax* protocol (ClinicalTrials.gov Identifier: NCT00663546). Chanaki Amaratunga, a staff scientist in the Fairhurst laboratory, determined which samples were best for subjecting to *ex vivo* culture based on the following exclusion criteria: patients who took antimalarials less than 1 month prior to sample collection and parasitemia <0.1%. She performed all sample processing in the field described in this subsection according to a previously published *P. vivax ex vivo* culture protocol (Russell

et al., 2011). *P. vivax* clinical isolates were processed (PV0417-3, PV0563, PV0565, PV0568) from Cambodian patients with malaria. Patient blood samples (16 ml for ages <18 years and 32 ml for >18 years collected by venipuncture into sodium heparin Vacutainers) were centrifuged (2000 rpm for 5 min) and plasma was removed. Samples were diluted with PBS (up to 64 ml, mixed by inverting tube), depleted of white blood cells and platelets (pass over 8 autoclaved, pre-wet CF11 columns packed to the 5.5-ml mark in a 10-ml syringe, and collect flow through), washed twice (centrifuge at 2000 rpm for 5 min, wash with 1x PBS, repeat 1 time), and re-suspend (packed cells at a 10% hematocrit in modified McCoys 5A complete media with 25% AB serum, and culture at 37°C, 5% CO₂) until the parasites matured to schizonts. After maturation, cultures were pelleted (centrifuge at 2000 rpm for 5 min), the supernatant was removed, and cells were resuspended (to 50% hematocrit in 1x PBS). To prevent rosetting, cells were treated with trypsin (7.5 ml of 500 mg/l Trypsin-Versene) and incubated (15 min at 37°C). To stop digestion, samples were diluted (add 2x volume of 1x PBS) and centrifuged (2000 rpm for 5 min). The recovered pellet was incubated with AB serum (6 ml for 5 min at RT), diluted with PBS (up to 30 ml) and separated on a 45% isotonic Percoll® gradient (5 ml of suspension overlaid on six 15-ml tubes containing 5 ml 45% isotonic Percoll® each). The suspensions were centrifuged (1200 g for 15 min) and the fine band of concentrated schizonts on the Percoll® interface was removed, centrifuged (2000 rpm for 5 min), and resuspended (5 ml of 1x PBS). Smears (made by pelleting ~200 µl) were counted with a hemocytometer. The remaining sample was pelleted, mixed with up to 10x RNAlater® (1 ml) and divided into 2 cryovials (500 µl each). Samples were stored at 4°C overnight and transferred to liquid nitrogen*. Samples were shipped to the WTSI on dry ice.

*To note. -20°C storage and transport would have been sufficient and recommended by RNAlater®. Storage in liquid nitrogen appeared to lead to some brown cloudiness (likely lysis), which made pelleting difficult, and multiple fractions were extracted. This did not appear to impact RNA quality and yields were very high.

2.1.4 RNA extraction of field isolates

RNA was isolated from the 4 field isolates (PV0417-3, PV0563, PV0565, PV0568) using the RiboPure Blood kit (Ambion) according to manufacturer instructions and subjected to 2 rounds of DNA digestion using the DNA-free kit (Ambion) according to manufacturer

instructions. Samples were analyzed by Bioanalyzer® (Agilent Technologies, Inc.) using an RNA Nano Chip to test quantity and quality.

2.1.5 cDNA synthesis and PCR

To test for genomic DNA contamination, 1 µl of extracted RNA from parasite isolates (PV0417-3, PV0563, PV0565, PV0568) was used to make cDNA using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems) according to manufacturer instructions. Using both RNA and cDNA samples, a region surrounding an intron in the PvDBP gene PVX_110810 (5'AAACCGCTCTTTATTTGTTCTCC, 3' TTCCTCACTTCTTCTTTCATT) was amplified by PCR. Reaction volumes were as follows: 2.5 µl buffer, 2 µl dNTP mix (10 µM), 0.1 µl Platinum Pfx DNA Polymerase (Invitrogen), 16.9 µl water, 1 µl of each primer (10 µM), 1 µl cDNA or extracted RNA. Thermocycler conditions were as follows: incubation at 95°C for 15 min, 35 cycles consisting of denaturation at 95°C for 40 seconds, annealing at 55°C for 40 seconds, elongation at 68°C for 1 min, followed by a final extension at 65°C for 5 min.

2.1.6 Strand-specific RNA library production

Lia Chappell developed a strand-specific Illumina library protocol for *P. falciparum* and *P. knowlesi* RNA (Chappell, et al, manuscript in preparation), and her protocol was followed with slight modifications for producing strand-specific Illumina libraries for the 4 *P. vivax* RNA samples (PV0417-3, PV0563, PV0565, PV0568). The work in this section used filter tips, RNase-free tubes (0.2 ml), water, and other reagents throughout, and the workspace was regularly cleaned with RNAzap/70% EtOH. Tubes were placed in magnetic racks for any step in which supernatants were removed. Aliquots were taken throughout the process to check for the quantity and quality of material made after each step. I performed all laboratory steps through to the USER treatment of the libraries. Lia Chappell performed the PCR reactions and final clean-up prior to sending samples to the core WTSI sequencing facility.

2.1.6.1 Isolate mRNA using oligo(dT) magnetic beads

Aliquots of 5 µg of total RNA were made for each sample and diluted to a final volume of 50 µl in nuclease-free water. Dynabead Oligo(dT) beads (20 µl) were washed twice (100 µl of 2x RNA Binding Buffer), resuspended (50 µl of 2X RNA Binding Buffer), and combined with the total RNA. RNA was denatured to facilitate mRNA binding to the

beads by heating (65°C for 5 min), cooling (to 4°C), and incubating (RT, 5 min). Beads+mRNA were separated from the solution by placing tubes on the magnetic rack and incubating (RT, 2 min), and the supernatant was stored (-80°C). Beads+mRNA were washed twice (200 µl of Wash Buffer, with thorough pipette mixing 6 times). Beads+mRNA were separated from the solution by placing tubes on the magnetic rack and incubating (RT, 2 min), and the wash buffer was discarded. mRNA was eluted from the beads by removing tubes from the magnetic rack, mixing thoroughly with elution buffer (50 µl) and incubating samples (80°C, 2 min, then hold at 25°C). mRNA isolation was then repeated to further purify the sample. mRNA was bound to beads by adding 2X RNA buffer (50 µl), mixing thoroughly, and incubating (RT, 5 min). Beads+mRNA were separated from the solution by placing tubes on the magnetic rack and incubating (RT, 2 min), and the supernatant was discarded. Beads+mRNA were washed twice (200 µl of Wash Buffer, with thorough pipette mixing 6 times). Beads+mRNA were separated from the solution by placing tubes on the magnetic rack and incubating (RT, 2 min), and the wash buffer was discarded. mRNA was eluted from the beads by removing tubes from the magnetic rack, mixing thoroughly with elution buffer (17 µl) and incubating samples (80°C, 2 min), placing tubes on magnetic rack and collecting the purified mRNA by transferring the supernatant to a clean nuclease-free PCR Tube on ice. Purified mRNA yield and size distribution were assessed using 2 µl on a Bioanalyzer® (Agilent Technologies, Inc.) using a RNA Nano Chip.

2.1.6.2 mRNA fragmentation by Covaris

Samples were diluted (to 120 µl, nuclease-free water), heated (65°C, 5 min) and cooled (on ice, 5 min) to minimize secondary structure. Samples were transferred to Covaris tubes and fragmented using the following settings: Duty cycle 10%, Intensity 5, Cycles per burst 200, Time 60s (more for total RNA), water bath 4-8°C. Fragmentation success was assessed immediately using a 2 µl aliquot on a Bioanalyzer® (Agilent Technologies, Inc.) using a RNA Nano Chip. Fragmented RNA (118 µl) was precipitated by adding the sample to a 1.5-ml eppendorf tube and adding 2.5x the volume of RNase-free 96-100% EtOH (325 µl), 1/10 volume 3M Sodium Acetate (13 µl), glycogen (20 µg) and incubating (-80°C, overnight). Precipitated RNA was centrifuged (14,000 rpm, 4°C, 30 min), with supernatant discarded, washed twice (500 µl, 75% EtOH), and centrifuged (14,000 rpm, 4°C, 5 min). Supernatant was discarded by pipette and pellet was air dried

in a covered box (RT, 2-4 min), resuspended (12 μ l, nuclease-free water), and stored on ice.

2.1.6.3 Reverse transcription/first strand cDNA synthesis and modified second-strand synthesis

Reverse transcription/first strand cDNA synthesis was performed with Superscript II according to manufacturer instructions for step 1 to 6 for a 20- μ l reaction volume using random hexamers and 10 μ l of resuspended fragmented mRNA. Following step 6, the RNA-DNA hybrid was cleaned using 1.8x reaction volume with Agencourt RNAClean XP Spri-beads (Beckman Coulter, Inc.) (36 μ l beads to 20 μ l starting volume). Samples were mixed (vortex mixer), incubated (RT, 5 min), and separated from supernatant by magnetic stand until solution was clear (~5 min). The supernatant was discarded and the beads were washed twice (180 μ l, 80% EtOH, incubated for 30 seconds). The EtOH was discarded and the beads were air-dried (10 min while the tube was on the magnetic stand). The RNA-DNA hybrid was eluted from the beads (25 μ l water, mixed with vortex mixer), and placed on the magnetic stand until the solution was clear. The supernatant (23 μ l) was transferred to a clean tube. Second-strand synthesis incorporated dUTPs instead of dTTPs by mixing together the following: 22.6 μ l cleaned RNA-DNA, 3 μ l Buffer 2 (NEB), 2 μ l of 10 mM dNTP mix (using dUTP instead of dTTP), 0.4 μ l RNase H (Invitrogen), and 2.0 μ l DNA pol I (Invitrogen). Reaction conditions were as follows: 16°C for 2 h, hold at 4°C. Samples were cleaned using Agencourt RNAClean XP Spri-beads (Beckman Coulter, Inc.) and eluted as above but with 39.5 μ l elution buffer.

2.1.6.4 End repair of cDNA libraries

To end repair the cDNA, the following reagents (all from NEB, except cDNA) were mixed together: 37.5 μ l cDNA, 2 μ l dNTP mix, 2.5 μ l T4 DNA polymerase, 0.5 μ l *E. coli* DNA polymerase 1, 2.5 μ l T4 Polynucleotide Kinase, and 5 μ l 10x phosphorylation buffer. The reactions were heated (20°C for 30 min), cleaned using Agencourt RNAClean XP Spri-beads (Beckman Coulter, Inc.), and eluted as above but with 34 μ l elution buffer.

2.1.6.5 dA-Tailing of cDNA libraries

Performed dA-Tailing of cDNA library using the NEBNext DNA modules with the following reaction volumes: 32 μ l purified end-repaired cDNA, 5 μ l NEBNext dA-Tailing Reaction Buffer, 10 μ l 1 mM deoxyadenosine 5'-triphosphate, 3 μ l Klenow fragment. Reactions were incubated (37°C for 30 min), cleaned using Agencourt

RNAClean XP Spri-beads (Beckman Coulter, Inc.), and eluted as above but with 22 µl elution buffer.

2.1.6.6 Ligation of PCR adapters

PCR adapters were added to the libraries using standard Illumina adapters. Reaction volumes were as follows: 20 µl end-repaired, dA-tailed cDNA, 25 µl DNA ligase buffer, 1 µl 10 µM no-PCR IDT P.E., and 4 µl Quick DNA ligase (NEB). Reactions were incubated (25°C for 15 min), cleaned using Agencourt RNAClean XP Spri-beads (Beckman Coulter, Inc.), and eluted as above but with 1x Agencourt RNAClean XP Spri-beads (Beckman Coulter, Inc.) and 22 µl elution buffer. The eluted samples were transferred to clean 1.5-mL LoBind tubes.

2.1.6.7 Digestion of dUTP strand and PCR

The strand containing dUTPs was digested using a standard protocol by mixing together the following: 20 µl cDNA and 2 µl USER enzyme (NEB). Reactions were incubated (37°C for 15 min), heated (95°C for 10 min), and cooled (held at 4°C). The libraries were then amplified by PCR. Reaction volumes were as follows: 20 µl USER treated cDNA, 5 µl water, 25 µl 2x KAPA HIFI HS Master mix, 1 µl of 10 mM PE 1 Illumina primer, and 1 µl of 10 mM Illumina index primer (matching the index of the ligated PCR-free adapters). Reaction conditions were as follows: 95°C for 5 min followed by 4 cycles of 95°C for 20 s, 60°C for 15 s, 72°C for 60, and a final extension of 72°C for 5 min. PCR reactions were then cleaned using Agencourt RNAClean XP Spri-beads (Beckman Coulter, Inc.) and eluted as above but with 1x Agencourt RNAClean XP Spri-beads (Beckman Coulter, Inc.) and 22 µl elution buffer. The eluted samples were transferred to clean 1.5-mL LoBind tubes.

2.2 *P. vivax* RNA sequence analysis

2.2.1 Sequence mapping and quality control

Illumina HiSeq RNA sequences for 4 clinical isolates were mapped to the following reference genomes:

P. vivax Sal 1 (Carlton et al., 2008): http://plasmodb.org/common/downloads/release-9.3/PvivaxSal1/fasta/data/PlasmoDB-9.3_PvivaxSal1_Genome.fasta

P. vivax P01 (Auburn, Manuscript in preparation): <ftp://ftp.sanger.ac.uk/pub/project/pathogens/gff3/CURRENT>

P. falciparum 3D7 (Gardner et al., 2002): http://plasmodb.org/common/downloads/release-9.3/Pfalciparum3D7/fasta/data/PlasmoDB-9.3_Pfalciparum3D7_Genome.fasta

Homo sapiens (Lander et al., 2001): NCBI build V37

TopHat (version 2.0.14) (Kim et al., 2013) was used with the following settings: tophat -g 1 -I 2000. The max intron size was set to 2000 (-I 2000). The '-g 1' option ensured only uniquely mapping reads would be aligned (allows a single alignment to the reference, placed at the top scoring position if more than 1 alignment was found). Samtools (Li et al., 2009, Li, 2011) was used to report the numbers of total reads (samtools flagstat) and reads mapped uniquely (samtools view -c -q 1) to the *P. vivax* Sal 1, *P. vivax* P01, *P. falciparum* 3D7, and *Homo sapiens* reference genomes. Reads mapped to each genome were extracted from the bam files (samtools view -F4 file.bam | cut -f1 | sort | uniq > file_reads.txt) and compared to look for reads mapping to both *P. vivax* and *P. falciparum*, *P. vivax* and *Homo sapiens* (comm -12 file1.txt file2.txt | wc). When comparing the numbers of reads mapping to *P. vivax* Sal 1 and *P. vivax* P01, only the 14 main chromosomes were used (samtools idxstats file.bam | cut -f 1,3).

The *P. vivax* P01 reference genome annotation was divided into 3 categories (Exons, Introns and the remainder or “other”), the process of which is reviewed here: <http://davetang.org/muse/2013/01/18/defining-genomic-regions/>. This involves extracting the regions from the *P. vivax* P01 gtf file into a bed file using unix and Bedtools (2.17.0) (Quinlan and Hall, 2010) as follows:

```
Exons: cat file.gtf | awk 'BEGIN{OFS="\t";} $3=="exon" {print $1,$4-1,$5}' | sortBed | mergeBed -i - > file.exon.beda
```

```
Introns: cat file.gtf | awk 'BEGIN{OFS="\t";} $3=="transcript" {print $1,$4-1,$5}' | sortBed | subtractBed -a stdin -b file.exon.bed > file.intron.beda
```

```
Remaining “other” regions: cat file.gtf | awk 'BEGIN{OFS="\t";} $3=="transcript" {print $1,$4-1,$5}' | sortBed | complementBed -i stdin -g file.txt* > file.other.beda
```

*create tab delimited file.txt listing chromosome ids and sizes

^aApicoplast and mitochondrial records removed from the files before coverage analysis

Coverage histograms were generated for the above genomic region bed files using Bedtools (coverageBed -split -hist -abam file.bam -b file.exon.bed > file.exons.bed.coverage), and aggregate statistics were calculated using Excel spreadsheets or R. Coverage counts for rRNA regions using Bedtools (coverageBed -abam file.bam -b rRNA_locations.bed* > file.rRNA.coverage.counts). *created based on list compiled from the *P. vivax* P01 reference genome annotation (<http://www.genedb.org/Homepage/PvivaxP01>).

2.2.2 RNA-Seq expression analysis

Expression results were generated from the TopHat mapping assemblies using Cufflinks version 2.2.1 (Trapnell et al., 2012, Trapnell et al., 2010) with the following settings: cufflinks -o out.dir -b ref.fasta -u -q -g ref.gtf file.bam. The '-b' with a provided reference fasta file (ref.fasta) uses a bias detection and correction algorithm to improve accuracy of estimated transcript abundance. The '-u' option enables an "initial estimation procedure to more accurately weight reads mapping to multiple locations in the genome," (<http://cole-trapnell-lab.github.io/cufflinks/cufflinks/#transcriptome-assembly-transcriptsgtf>). The '-g' option provides reference genome annotation (file.gtf) to guide the assembly. Using this option, the output file contained both the reference transcripts and any novel transcripts identified. The '-o' directs to the output directory (out.dir) and the '-q' suppresses messages other than warnings/errors. See the following for more details about options: <http://cole-trapnell-lab.github.io/cufflinks/cufflinks/#transcriptome-assembly-transcriptsgtf>

The results of Cufflinks computed a normalized expression level for each gene called **F**ragments **P**er **K**ilobase of transcript per **M**illion mapped reads or FPKMs. The FPKMs for each isolate were combined using unix and a custom perl script (Thomas Otto, Parasite Genomics group, WTSI). The results from the 'genes.fpkm_tracking' files output from Cufflinks were sorted, with gene_id and FPKM columns copied to a new file for each isolate in a new 'Stats' directory from the original 'cuff.SampleX' directories:

```
for x in `ls -d cuff.Sample1 cuff.Sample2 cuff.Sample3 cuff.Sample4` ; do n=$(echo $x | sed 's/cuff.//g' ); echo -e "gene_id\t$n" > Stats/$n.txt; cut -f 1,10 $x/genes.fpkm_tracking | grep -v FPKM | sort >> Stats/$n.txt; done
```

The 4 isolates FPKMs were then combined into a single file (Results.txt in the Stats directory): paste Sample1.txt Sample2.txt Sample3.txt Sample4.txt | cut -f 1,2,4,6,8,10,12 | sed 's/_R/_r/g' > Results.txt

The product descriptions were added to the combined FPKM columns in Results.txt with a perl script and product description file supplied by Thomas Otto (Berriman Parasite Genomics Group, WTSI).

Multiple comparisons of FPKM results from Cufflinks were computed using R or Microsoft Excel. In order to validate that our samples best reflected the schizont stage of the *P. vivax* life cycle, the generated data were compared to published *P. falciparum* RNA-Seq data and microarray data, and *P. vivax* microarray data (Otto et al., 2010, Bozdech et al., 2003, Bozdech et al., 2008). Correlation plots (using the R “corrplot” library) for 1:1 orthologs between the raw FPKM values of the clinical isolates and either the RPKM data from *P. falciparum* RNA-Seq data or the microarray enrichment values for *P. falciparum* and *P. vivax*.

The similarity of expression between the isolates was investigated using R in pair-wise comparisons plotting the log(FPKM) of each sample to log(FPKM) of every other sample. Pearson correlations coefficients were computed for each pair-wise comparison (untransformed data). R (version 3.2.2, <http://cran.r-project.org>) was used to generate histograms and smooth density plots (density function) of the log(FPKM) for each isolate and plotted using the ‘sm’ library.

2.2.3 Expression variability between isolates

Expression variability between isolates was investigated in 3 ways using Microsoft Excel. First the ‘max fold-change’ was computed, by first calculating the mean FPKM from the 4 clinical isolates (PV0563, PV0565, PV0568, PV0417-3) for each gene. The fold-change from the mean for each isolate was calculated (isolate expression divided by the mean), and the maximum fold-change from the mean was recorded. Next, the Coefficient of Variation (c_v) was calculated for each gene as the ratio of standard deviation (σ) of the FPKMs from the 4 isolates to the mean (μ) FPKM:

$$c_v = \frac{\sigma}{\mu}$$

Lastly, the variance-to-mean ratio (VMR) also called index of dispersion (D) was calculated as a way to understand the spread of the expression results. This was computed as a ratio of the variance (σ^2) of the FPKMs from the 4 isolates to the mean (μ) FPKM:

$$D = \frac{\sigma^2}{\mu}$$

The intersection of the top 300 ranked genes for each method was visualized in a Venn diagram.

In order to understand if any types of genes were enriched in the top variably-expressed gene set, the Gene Ontology (GO) terms for the one-to-one homologs with *P. falciparum* were investigated. The list of 77 *P. falciparum* one-to-one homologs was uploaded to panther.db. Analysis type: PANTHER Overrepresentation Test (release 20150430). Annotation version and release date: GO Ontology database, released 2015-08-06. Reported results only from published experimental data.

2.3 Production of recombinant *P. vivax* ectodomain library

2.3.1 *P. vivax* candidate selection

Expression plasmids corresponding to the entire ectodomains of secreted and membrane-embedded *P. vivax* merozoite proteins were designed and constructed essentially as described (Bushell et al., 2008, Crosnier et al., 2013). The entire ectodomain protein sequences were identified by removing the endogenous signal peptide, transmembrane domain, and glycosylphosphatidylinositol (GPI) anchor sequences (if present), and the corresponding nucleic acid sequences were codon-optimized for expression in human cells. N-linked glycosylation sequons were mutated from NXS/T to NXA (where X is any amino acid except proline) to prevent glycosylation when proteins were expressed in human cells. The final constructs were chemically synthesized and sub-cloned into a derivative of the pTT3 expression vector (Durocher et al., 2002), which contains an N-terminal signal peptide, a C-terminal rat CD4 domain 3 and 4 (Cd4d3+d4) tag, and 17 amino acid biotinylatable peptide using flanking NotI and AscI restriction sites (Genart AG, Germany). In the interaction screens described below (section 2.5.2), these biotinylated proteins are referred to as “baits.” All biotinylatable bait expression plasmids are available from the non-profit plasmid repository, Addgene (www.addgene.org).

2.3.2 Subcloning *P. vivax* recombinant library

β -lactamase-tagged “prey”-expressing plasmids were produced by subcloning each ectodomain into a plasmid containing a pentamerization domain conjugated to the β -lactamase tag as described (Crosnier et al., 2013, Bushell et al., 2008). A subset of ectodomains was subcloned into a plasmid containing a six-histidine tag for producing purified proteins. All plasmids with target backbones were provided by Cecile Crosnier-Wright in the Wright laboratory. Expression plasmid backbones summarized in Table 2.1.

Table 2.1: Expression plasmid backbones

Name	Characteristics
Bio	Biotinylatable ‘bait’ pTT3-based expression vector (Durocher et al., 2002) with rat CD4d3+d4 tag (Brown and Barclay, 1994), 17 amino acid substrate for the <i>E. coli</i> biotin ligase BirA (Bushell et al., 2008, Brown et al., 1998)
βlac	β -lactamase-tagged ‘prey’ pTT3-based expression vector (Durocher et al., 2002) with rat CD4d3+d4 tag (Brown and Barclay, 1994), pentamerization domain of the rat cartilaginous oligomeric matrix protein (COMP) (Tomschy et al., 1996) with an ampicillin resistance protein (TEM) β -lactamase (Crosnier et al., 2010, Bushell et al., 2008)
Hexa- His	6-His-tagged pTT3-based expression vector (Durocher et al., 2002) with rat Cd4d3+d4 tag (Brown and Barclay, 1994) (Bushell et al., 2008)

Inserts and target vector backbones were cut using the following reaction components and conditions: 30 μ l plasmid (1 mg/ml), 5 μ l Buffer 4 (10x), 1 μ l BSA (100x), 2 μ l NotI (New England BioLabs, NEB), 4 μ l AscI (NEB), and 8 μ l MilliQ water incubated at 37°C for 3 h or overnight. Cut plasmids were then separated using agarose gel electrophoresis with 1% agarose gels in TAE at 100-150V for 1 h or more to achieve good separation between insert and vector backbone. Inserts or vector backbones were then cut from the gels and extracted using the QIAquick® Gel Extraction Kit (Qiagen) according to manufacturer’s instructions, and DNA concentration was then quantified using a NanoDrop (Thermo Scientific).

Inserts and vector backbones were then ligated in a 3:1 ratio with a target of 60 ng insert to 20 ng vector backbone. The reaction volumes were as follows: 1-2 μ l insert DNA (60 ng), 0.5-1 μ l vector backbone DNA (20 ng), 1 μ l T4 ligase (NEB), 1 μ l T4 ligase buffer (NEB), and MilliQ water up to 10 μ l. The ligations were incubated at RT for 3 h.

Ligations were stored at 4°C for 1-2 days prior to transformation and plating or frozen for use at a later time.

Ligations were then used in transformation and plating. Chemically competent *E. coli* cells (25 µl One Shot® Top10, Invitrogen) were thawed on ice and combined with 3 µl of ligation reaction and incubated on ice for 5-30 min, heated to 42°C for 45 seconds, cooled on ice for 2 min, and then mixed with 100-200 µl of SOC media at RT. The transformed cells were then spread using sterile techniques on LB-agar plates containing 100 µg/ml of ampicillin (Media Team at WTSI or Richard Eastman at LMVR) and grown overnight at 37°C.

A selection of single colonies from positive plates was then checked using polymerase chain reaction (PCR). Single colonies were picked into tubes containing 12.5 µl GoTaq® Green Master Mix (Promega), 1 µl Primer 3610 (100 mM), 1 µl Primer 4006 (100 mM), 10 µl MilliQ water, and run with the following Thermocycler conditions: incubation at 95°C for 10 min, 30 cycles consisting of denaturation at 95°C for 30 seconds, annealing at 60°C for 30 seconds, elongation at 68°C for 1 min, followed by a final extension at 68°C for 10 min. Primers for PCR and subsequent sequencing confirmation summarized in Table 2.2.

Table 2.2: Primers

Primer name	Primer sequence	Location/use
OL3610	GCCACCATGGAGTTTCAGACCCAGGTACT CATGTCCCTGCTGCTCTGCATGTCTGGTGC	5' from insert sequence and Not1 site, in exogenous signal peptide
OL4006	TCCCTGCAGGCTTTCCTCTCCAAGGTTGAG	3' from insert sequence and Asc1 site, in the Cd4d3+d4 tag
OL497	TGAGATCCAGCTGTTGGGGT	5' from insert sequence, in vector backbone
OL498	AGAAGGGGCAGAGATGTCGT	3' from insert sequence, in vector backbone

Half of the total PCR reaction was then separated using agarose gel electrophoresis with 1% agarose gels in TAE run at 100-150V for 30-60 min and checked for size agreement with the known insert size.

Single colonies with inserts of the correct size were then picked into 50 ml of LB-AMP (100 µg/ml) and grown shaking at 37°C overnight. Plasmid DNA was then isolated using PureLink® HiPure Plasmid Filter Maxiprep Kit (Invitrogen) according to manufacturer's instructions. DNA was resuspended in TBE, and the concentration was adjusted to 1 mg/ml using a NanoDrop (Thermo Scientific).

To ensure that the final plasmid contained the expected insert and backbone, all sub-cloned plasmids were subjected to DNA sequencing using the primers summarized in Table 2.2 above, either by sending plasmid and primer aliquots to the in-house WTSI Capillary Sequencing Team, Macrogen, Inc. (USA), or by performing the reactions and running them on the LMVR 96-well capillary sequencing machine, 3730xl DNA Analyzer (Applied Biosystems).

All sequences were then assembled and aligned to the expected insert sequences and vector backbones using SeqMan Pro (version 12, DNASTAR Lasergene software package) or Sequencher (version 5.3, Gene Codes Corporation).

2.3.3 *P. vivax* library expression in the HEK293E system

P. vivax biotinylated bait and β-lactamase-tagged prey recombinant proteins were expressed in HEK293E cells as described (Bushell et al., 2008, Crosnier et al., 2013). HEK293E cells were maintained with shaking in Freestyle™ 293 Expression Medium (Invitrogen, USA) supplemented with G418 (50 mg/l), pen/strep (10000 units/l), and HI-FBS (1%) at 37°C in a 5% CO₂ atmosphere. Bait and prey plasmids were transiently transfected as described (Durocher et al., 2002, Bushell et al., 2008). Cells were split into 50 ml of fresh media (2.5x10⁵ cells/ml), incubated for 24 h, and inoculated with a transfection mix consisting of 25 µl of expression plasmid (at 1 mg/ml), 50 µl of linear PEI (at 1 mg/ml), preincubated for 10 min at RT in 2 ml of non-supplemented Freestyle media or using the 293fectin™ (Invitrogen, USA) according to manufacturer's instructions. Bait proteins were enzymatically biotinylated during synthesis using media supplemented with D-biotin (100 µM) and were co-transfected with a biotin ligase expression plasmid expressing a secreted BirA protein, along with the bait construct in a 1:10 ratio (2.5 µl of a plasmid at 1 mg/ml). After 3-6 days, cultures were centrifuged (4000 rpm for 15 min) and supernatants filtered (0.2-µm filter). Bait proteins were dialyzed in Snakeskin dialysis tubing (10 kDa MWCO, Thermo Scientific) using 5 L of

HBS (5-7 buffer changes over a 48-h period) to remove excess D-biotin. Cultures were stored in 10 mM sodium azide at 4°C until use.

2.4 Confirmation and assessment of protein expression

2.4.1 Confirmation of protein expression by ELISA

Biotinylated bait proteins were immobilized on streptavidin-coated plates (Nunc, USA), and the dilution required to saturate all biotin-binding sites was determined by ELISA as described (Kerr and Wright, 2012). Proteins were serially diluted in HBS/1% BSA for 1 h, and plates were washed 3 times in HBS/0.1% Tween, incubated with mouse anti-rat CD4 antibody (OX68*, 1:1000) in HBS/1% BSA for 1 h, washed 3 times in HBS/0.1% Tween, incubated with mouse anti-rat alkaline phosphatase antibody (1:5000, Sigma, USA) in HBS/1% BSA for 1 h, washed 3 times in HBS/0.1% Tween, and washed once in HBS. Proteins were then incubated with phosphate substrate (1 mg/ml, Sigma) in either diethanolamine buffer (10% diethanolamine, 0.5 mM magnesium chloride, pH 9.2) or coating buffer (0.015 M sodium carbonate, 0.035 M sodium bicarbonate, pH 9.6) and detected by measuring OD at 405 nm. Each protein was concentrated or diluted using Vivaspin® 30-kDa spin columns (Sartorius, Germany) or HBS/1% BSA, respectively. As an approximate guide, the biotin binding sites are saturated at a biotinylated protein concentration of 0.3-0.5 µg/ml (Kerr and Wright, 2012, Osier et al., 2014). Based on this, and as a very rough approximation, protein expression was categorized “high” (likely above 5 µg/ml) for those proteins that could be diluted more than 1:10 and still saturate biotin binding sites, “medium” (approximately 0.5-5 µg/ml) for those proteins using 1:1 through 1:10 dilutions, or “low” (likely under 0.5 µg/ml) for those proteins that showed some signal, but failed to saturate the biotin binding sites over a range of dilutions prior to concentrating. These categories should be treated as a guide only since, due to experimental variation in transient transfections, we observed significant batch-to-batch variation.

*OX68 was isolated from hybridomas by Cecile Crosnier-Wright in the Wright laboratory and aliquoted for regular use.

2.4.2 Normalisation of β -lactamase tagged membrane protein ectodomains

Prey proteins were normalized using the β -lactamase tag activity as a proxy of protein concentration by the rate of nitrocefin, essentially as described (Kerr and Wright, 2012). Proteins were serially diluted (neat, 1: 5, and either 1:10 or 1:50) in HBS/1% BSA, and 20 μ l of each dilution was transferred to a 96-well plate. Nitrocefin solution (Calbiochem) was added to each well (60 μ l at 125 μ g/ml), and absorbance at 485 nm was immediately read every minute for 20 minutes at RT. For each protein dilution, Absorbance at 485 nm was plotted against time and preys were concentrated or diluted as needed to achieve saturation of signal at approximately 10 min (approximating a threshold activity of \sim 2 nmol/min turnover of nitrocefin) using Vivaspin® 20- or 30-kDa spin columns (Sartorius, Germany) or HBS/1% BSA (Kerr and Wright, 2012).

2.4.3 SDS-PAGE, Western blotting and NativePAGE

Western blotting was used to confirm protein sizes. *P. vivax* biotinylated bait proteins were reduced with NuPAGE® Sample Reducing Agent (Invitrogen), heated at 70°C for 10 min, fractionated by SDS-PAGE, and transferred to nitrocellulose membranes. After blocking in HBS/0.1% Tween/2% BSA at 4°C overnight, membranes were incubated with streptavidin horseradish peroxidase (1:2000, Cell Signaling Technology, USA) in HBS/0.1% Tween/2% BSA at RT for 1-2 h, developed using the chemiluminescence substrate Amersham ECL™ Prime (GE Healthcare, USA), and exposed to X-ray film.

In order to determine whether several proteins formed dimers or oligomers, non-reducing conditions were tested. Recombinant 6-His-tagged *P. vivax* P12, P41, and MSP7.1 were fractionated by SDS-PAGE under reducing conditions as above [NuPAGE® Sample Reducing Agent (Invitrogen), heated at 70°C for 10 min] and non-reducing conditions, and stained using SimplyBlue SafeStain (Invitrogen). Samples were also run under native conditions using the NativePAGE Bis-Tris Gel System (Thermo Fisher Scientific), stained with Coomassie Brilliant Blue R-250 (ThermoScientific).

2.5 High-throughput functional screens

2.5.1 Erythrocyte/reticulocyte binding experiments by flow cytometry

P. vivax recombinant protein binding to whole erythrocytes and reticulocytes was investigated using a flow cytometry-based assay developed by Madushi Wanaguru as

described (Crosnier et al., 2013) based on a previously described method (Brown, 2002). The biotinylated proteins were immobilized on streptavidin-coated fluorescent beads (Nile red) and presented to erythrocytes, with binding events detected as a shift in fluorescence intensity by flow cytometry. A bead-saturation ELISA similar to the protocol in section 2.4.1 was first necessary to determine the optimal protein dilution for saturating the biotin binding sites on a uniform number of beads while leaving a minimal amount of unbound protein.

Biotinylated proteins were serially diluted in HBS/1% BSA for 1 h and 100 μ l of each dilution was transferred to a 96-well microtitre plate. Proteins were either pre-incubated for 1 h at 4°C, shaking, either without or with 4 μ l of streptavidin-coated Nile Red fluorescent 0.4–0.6 μ m microbeads (Spherotech Inc.). Prior to use, beads were washed twice and resuspended in HBS/1% BSA with sonication (5 min) after all washes using a bath sonicator at 4°C. The pre-incubated dilution series were then transferred to streptavidin-coated plates (Nunc, USA), where protein not bound to beads would subsequently bind to plates. The ELISA then proceeded as described in section 2.4.1. Protein dilutions pre-incubated with beads indicating saturation of biotin sites on beads, with low remaining quantities of unbound protein (as indicated by low OD values by ELISA) were selected.

For the binding assay, proteins (100 μ l) were then immobilized on 4 μ l streptavidin-coated Nile Red fluorescent 0.4–0.6 μ m microbeads (Spherotech Inc.) (washed and sonicated as above) by incubation with gentle shaking for 1 h at 4°C. Human erythrocytes or hematopoietic stem cell-derived reticulocytes (provided by NHS Blood and Transplant) were washed twice in RPMI and aliquoted into a 96-well flat bottomed plate at 3×10^5 cells/well in 30 μ l of RPMI, creating a monolayer of cells available to the protein-coated beads. Plates containing protein-coated beads were sonicated for 20 min at 4°C and transferred (100 μ l protein-bead mixture resulting in approximately 120 beads/cell) to erythrocyte-containing plates. The erythrocyte-bead mixture was centrifuged (1000 rpm for 20 min) and incubated with gentle shaking for 1 h at 4°C. The cells were analysed using flow cytometry on an LSRII cytometer (BD Biosciences) using the BD FACS Diva software. Nile Red was excited by a blue laser and detected with a 575/26 filter. Voltages for FCS (430 V) and SSC (300 V) and a threshold of 26,100 on FSC were applied when analysing. The results were analysed using Flow Jo v7.5.3 software (Tree Star, Inc.).

2.5.2 Protein screens using AVEXIS

P. vivax biotinylated bait and β -lactamase-tagged prey proteins were screened using AVEXIS (Bushell et al., 2008, Crosnier et al., 2013). For AVEXIS, biotin-binding sites on streptavidin-coated plates (Nunc) were saturated with biotinylated bait proteins for 1 h. Plates were washed 3 times in HBS/0.1% Tween, probed with β -lactamase-tagged prey proteins for 1 h, washed twice with HBS/0.1% Tween, washed once with HBS, and incubated with nitrocefin (60 μ l at 125 μ g/ml; Calbiochem, USA). Positive interactions were indicated by nitrocefin hydrolysis, which was detected by measuring absorbance at 485 nm at 90 min.

2.6 Biophysical analysis of protein interactions

2.6.1 Purification of 6-His-tagged membrane protein ectodomains

The entire ectodomains of P12, P41, and MSP7.1 were sub-cloned into a modified plasmid containing a 6-His tag (Bushell et al., 2008), expressed as above, and purified by immobilized metal ion affinity chromatography using HisTrap HP columns on an AKTA Xpress (GE Healthcare) following the manufacturer's instructions. Columns were pre-equilibrated with binding buffer (20 mM sodium phosphate, 40 mM imidazole, 0.5 M NaCl, pH 7.4) at a flow rate of 1 ml/min. Harvested supernatant (150-500 ml) was supplemented with imidazole (10 mM) and NaCl (100 mM) and passed over the column at a flow rate of 1 ml/min. The column was washed with 15 column volumes (CV) of binding buffer in order to remove non-specific adherents, and the proteins were eluted using 10 CV of elution buffer (20 mM sodium phosphate, 0.4 M imidazole, 0.5 M NaCl, pH 7.4). The eluant was collected in 0.5-ml fractions; fractions containing the highest concentration of purified protein (1-2 total), as estimated by measuring the absorbance at 280 nm, were subsequently subjected to SDS-PAGE and further purification by gel filtration.

In order to remove aggregates, purified proteins were subjected to size-exclusion chromatography (SEC) immediately before use as described (Taechalertpaisarn et al., 2012). Either a SuperdexTM 200 IncreaseTM 10/300 GL column or SuperdexTM 200 TricornTM 10/600 GL column using an AKTA Xpress (GE Healthcare) was pre-equilibrated with 2 CV of running buffer HBS-EP (0.01 M HEPES, 0.15 M NaCl, 3 mM EDTA, 0.005% v/v surfactant P20, pH 7.4 from GE Healthcare) at a flow rate of 0.2

ml/min. Protein samples were then injected, and the column was washed with running buffer at 0.2 ml/min, and 0.5-ml fractions were collected after passage of the void volume of the column. Absorbance at 280 nm and extinction coefficients (calculated using the protein sequence and the ProtParam tool, <http://www.expasy.org/tools/protparam.html>) were used to estimate the protein concentrations of peak fractions. Protein sizes were deduced from the elution volumes of the peak fractions by comparison to a standard curve generated by Josefin Bartholdson using well-defined protein standards from calibration kits (GE Healthcare).

2.6.2 Surface plasmon resonance (SPR)

SPR was used to biochemically confirm all interactions detected by AVEXIS. SPR was performed using a BIAcore T100 instrument (GE Healthcare) at 37°C in HBS-EP running buffer (0.01 M HEPES, 0.15 M NaCl, 3 mM EDTA, 0.005% v/v surfactant P20, pH 7.4 from GE Healthcare). Biotinylated P12, P41, MSP3.10, and PVX_110945 proteins (expressed as in 2.3.3) were immobilized to the sensor chip using the Biotin CAPture Kit (GE Healthcare) in approximately molar equivalents with a negative control reference biotinylated rat CD4d3+4 (tag alone). Increasing concentrations of each soluble purified analyte was injected at low (20 µl/min) flow rates for equilibrium experiments. The chip was regenerated between each injection cycle using regeneration buffer provided by the kit. Biacore evaluation software version 2.0.3 (GE Healthcare) was used to analyze the reference-subtracted sensorgrams. Binding was investigated by plotting the maximum binding response for a range of concentrations (R_{eq}) prior to washing, and plotted as a function of analyte concentration (C) and fitted using the equation $R_{eq}=CR_{max}/(C+K_D)$, where R_{max} is the maximum binding response and K_D is the equilibrium dissociation constant.

2.7 *P. vivax* seroepidemiology

2.7.1 Seroepidemiology of recombinant *P. vivax* proteins in Cambodian patients

2.7.1.1 Collection of Cambodian *P. vivax*-exposed plasma

In Pursat Province, Western Cambodia, Sokunthea Sreng and Seila Suon oversaw the collection and preparation of plasma samples from patients with acute vivax malaria.

Adult patients or the parents of child patients provided written informed consent under a protocol approved by the National Ethics Committee for Health Research in Cambodia and the NIAID Institutional Review Board in the United States (ClinicalTrials.gov Identifier, NCT00663546). Paired samples were also collected including ‘acute’ plasma from clinical presentation and ‘convalescent’ plasma 21-28 days later. Patient plasma samples were used in ELISA experiments described below.

2.7.1.2 ELISA optimization screening biotinylated *P. vivax* recombinant proteins with Cambodian patient plasma

Plasma from Cambodian patients with *P. vivax* infections were used to optimize ELISA conditions using Cambodian patient plasma. Plasma samples were selected based on prior ELISA data showing reactivity using recombinant *P. vivax* MSP1 (data generated and provided by Daria Nikolaeva and Carole Long, LMVR, NIH).

Carole Long and Kazutoyo Miura (LMVR/NIH) advised on the initial ELISA protocols based on the standard ELISA used in the Long Laboratory (Miura et al., 2008). Ababacar Diouf provided extensive initial laboratory training. The ELISA protocol for screening *P. vivax* recombinant proteins with Cambodian *P. vivax*-infected patient plasma was optimised using a subset of 6 *P. vivax* recombinant proteins and the Cd4d3+d4 tag alone expressed as in 2.3.3. To conserve protein supernatant, biotinylated bait proteins were diluted (HBS/1% BSA) to the maximum dilution where all biotin-binding sites on streptavidin-coated plates (Nunc, USA) were saturated. Optimal dilutions were determined by ELISA as described in section 2.4.1, and diluted proteins were immobilized on streptavidin-coated plates (Nunc, USA) for 1 h and washed 3 times (HBS, 0.1% Tween). Proteins were incubated in duplicate against 42 Cambodian *P. vivax* patient plasma samples (diluted 1:200, 1:400; 1:800, and 1:1600) in HBS with 1% BSA, 2 h) and washed 3 times (HBS, 0.1% Tween). Proteins were incubated with alkaline-phosphatase conjugated goat anti-human IgG (KPL, Inc., diluted 1:1000 in HBS with 1% BSA, 2 h), washed 3 times in HBS/0.1% Tween, and washed once in HBS. Proteins were then incubated with phosphate substrate (1 mg/ml, Sigma) in coating buffer (0.015 M sodium carbonate, 0.035 M sodium bicarbonate, pH 9.6) and detected by measuring OD at 405 nm measured at 5, 10, 20 and 25 min.

An OD value of 0.1 was set as a conservative lower limit of detection based on the plate reader used. Samples with OD values under 0.1 were set to 0.1. Differences in population

antibody responses comparing the proteins and the Cd4d3+d4 tag alone were assessed using Mann-Whitney U tests. A reactivity cut-off was set at 2 standard deviations above the mean of 5 American naïve controls after correcting for background responses by subtracting the Cd4d3+d4 values. Analysis and graphing performed with GraphPad Prism (version 6, GraphPad Software, Inc.)

2.7.1.3 Cambodian reactivity screen using a library of biotinylated *P. vivax* recombinant proteins with pooled Cambodian patient plasma

A set of 17 plasma samples was obtained from 14 patients who experienced their third, fourth, or fifth episode of acute vivax malaria within 2 years of their initial episode on our clinical protocol (Table 2.3). Control sera from 5 malaria-naïve individuals were obtained from Interstate Blood Bank (Memphis, TN, USA). Equal volumes of individual plasma or serum samples were pooled for use in ELISA.

Table 2.3: Plasma samples from Cambodian patients with acute vivax malaria

Patient ID	Sex	Age (years) at first episode	Date of episode				
			1	2	3	4	5
PV0005	M	10	20080619	20080723	20080829	20081008	20081114
PV0012	M	10	20080624	20080729	20080906	20081013	20081204
PV0022	M	24	20080627	20080809	20080925	20081212	
PV0092	F	5	20080830	20081005	20081212	20091015	
PV0106	M	8	20080914	20081021	20081201	20090713	20100729
PV0128	M	10	20081009	20100709	20101020		
PV0242	M	22	20090817	20091118	20101217		
PV0340	M	15	20100705	20101116	20101226		
PV0345	M	20	20100710	20100914	20101203		
PV0352	M	39	20100717	20100829	20101015		
PV0359	M	14	20100719	20100902	20101014		
PV0361	M	10	20100720	20100915	20101028		
PV0378	F	8	20100728	20100901	20101013	20101209	
PV0412	F	17	20100814	20100925	20101112	20101226	

Plasma samples were collected from patients in Pursat Province, Western Cambodia, in 2008-2010. Plasma was pooled from 17 malaria episodes (in bold) from 14 patients who presented with their third, fourth, or fifth episode of acute vivax malaria on our clinical protocol. Table provided by Chanaki Amaratunga and modified by Jessica Hostetler. S1 Table from (Hostetler et al., 2015).

ELISAs were used to assess the seroreactivity and conformation of 37 biotinylated *P. vivax* recombinant proteins and the Cd4d3+d4 tag alone (expressed as in section 2.3.3). *P. vivax* biotinylated proteins were denatured at 80°C for 10 min or left untreated, and then immobilized on streptavidin-coated plates (as in section 2.7.1.2). Proteins were incubated in a single assay in triplicate wells with pooled plasma (1:600 or 1:1000) from 14 vivax malaria patients from Cambodia or pooled serum from 5 malaria-naïve individuals from the United States in HBS/1% BSA. The ELISA then proceeded as described in section 2.7.1.2.

2.7.1.4 Cambodian reactivity screen using a 11 biotinylated *P. vivax* recombinant proteins with individual Cambodian patient plasma

In order to better understand IgG reactivity to recombinant *P. vivax* proteins in Cambodian patients, a subset of proteins was screened using ELISA against 48 individual Cambodian patient plasma experiencing their second, third, fourth, or fifth episode of acute vivax malaria within 2 years of their initial episode on our clinical protocol. Cambodian plasma samples (at 1:1000) and 24 sera samples from American malaria-naïve donors (at 1:1000) in HBS/1% BSA were screened in singlicate as described in 2.7.1.2, except that wash steps were performed using a plate washer machine with 4 washes in TBS, 0.1% Tween between each step, rather than HBS, 0.1% Tween.

2.7.1.5 Cambodian acute -convalescent screen using a library of biotinylated *P. vivax* recombinant proteins with Cambodian patient plasma

ELISAs were used to measure IgG responses in ‘acute’ and ‘convalescent’ Cambodian patient plasma to 10 biotinylated *P. vivax* recombinant proteins and the Cd4d3+d4 tag alone (expressed as in section 2.3.3). *P. vivax* biotinylated proteins were immobilized on streptavidin-coated plates (as in section 2.7.1.2). Proteins were incubated in a single assay in duplicate wells with ‘acute’ and ‘convalescent’ plasma (1:600) from 18 vivax malaria patients from Cambodia selected at random. Proteins were incubated individually with serum from 5 malaria-naïve individuals from the United States in HBS/1% BSA. The ELISA then proceeded as described in section 2.7.1.2, with the exception of using a plate washer machine with 4 washes in TBS, 0.1% Tween between each step, rather than HBS, 0.1% Tween.

2.7.1.6 Analysis

Differences in reactivity were tested using Mann-Whitney U tests (unpaired) and Wilcoxon rank-sum tests (paired). Analysis and graphing performed with GraphPad Prism (version 6, GraphPad Software, Inc.)

2.7.2 Seroepidemiology of recombinant *P. vivax* proteins in Solomon Islander patients

2.7.2.1 Cross-sectional study and sample selection

Blood samples were collected in a cross-sectional survey of 3501 individuals aged ≥ 6 months in May 2012 in Ngella, Central Island Province, SI (Waltmann et al., 2015). Samples were collected with ethical approval from the Solomon Islands National Health Research Ethics Committee and the Walter and Eliza Hall Institute, and verbal informed consent was given by all adult participants and parents or guardians of child participants (as approved by Australian and Solomon Islands' IRBs, and documented on each participant's case report form). Survey participants were grouped into a 3x3 factorial design for downstream IgG reactivity screening by ELISA. This included 3 age categories (5-9, 10-19, and 20-99 years) and 3 infection statuses as follows:

1. Not infected [negative by PCR and light microscopy (LM)]
2. PCR + (positive by PCR and negative by LM)
3. PCR + LM + (positive by PCR and positive by LM)

2.7.2.2 Solomon Islands plasma screening of *P. vivax* recombinant proteins using ELISA

Reactivity screening: ELISAs were used to measure IgG response in SI patient plasma to 34 biotinylated *P. vivax* recombinant proteins and the Cd4d3+d4 tag alone (expressed as in section 2.3.3). *P. vivax* biotinylated proteins immobilized on streptavidin-coated plates (as in section 2.7.1.2). Proteins were incubated with individual plasma samples (1:600) from 22 adolescents (10-19 years) and 24 adults (20-50 years) selected at random from the 3 infection categories described above. Proteins were incubated with pooled serum from 5 malaria-naïve individuals from Australia (1:600) in HBS/1% BSA as a negative control and plasma pooled highly-immune PNG adult donors (1:600) in HBS/1% BSA as a positive control. After 2 h of incubation, wells were washed 3 times with HBS/0.1% Tween, incubated with alkaline phosphatase-conjugated goat anti-human IgG (1:1000, KPL Inc., USA) in HBS/1% BSA for 2 h, washed 3 times with HBS/0.1% Tween, and

washed once with HBS. Seroreactivity was detected using phosphate substrate (1 mg/ml, Sigma) in coating buffer (0.015 M sodium carbonate, 0.035 M sodium bicarbonate, pH 9.6) and measuring OD at 405 nm at various times up to 30 min.

Comprehensive screening: ELISAs were repeated as above with an expanded set of plasma samples for 12 highly immunogenic proteins. In order to investigate relationships between infection status, clinical symptoms, and other factors, 144 plasma samples were selected at random from the 3 age categories and 3 infection statuses including the following: 48 children (5-9 years), 48 adolescents (10-19 years), and 48 adults (20-80 years) either without any *Plasmodium* infections (Not Infected), with a current *P. vivax* monoinfection detected by PCR (PCR+), or with a current *P. vivax* monoinfection detected by PCR and light microscopy (PCR+ LM+). Plasma samples were tested in duplicate on separate plates.

2.7.2.3 Statistical analysis

Reactivity screening: An OD value of 0.1 was set as a conservative lower limit of detection based on the plate reader used. Samples with OD values under 0.1 were set to 0.1. Differences in population antibody responses comparing the proteins and the Cd4d3+d4 tag alone, and 3 age categories were assessed using Mann-Whitney U tests. Analysis and graphing performed with GraphPad Prism (version 6, GraphPad Software, Inc.)

Comprehensive screening: An OD value of 0.1 was set as a conservative lower limit of detection based on the plate reader used. Samples with OD values under 0.1 were set to 0.1. OD values from duplicate wells were averaged, the Cd4d3+d4 values were subtracted to correct for background, and $\log_{(10)}$ -transformed. Differences between mean antibody responses between age groups or infection groups were analysed using ANOVA. Positivity cut-offs were set at 2 standard deviations above the mean antibody levels of the negative controls. Negative binomial regression was used to assess differences in the breadth of antibody levels by age and infection group. Analysis and graphing performed using STATA (version 12, StataCorp), GraphPad Prism (version 6, GraphPad Software Inc.), or R (version 3.2.2, <http://cran.r-project.org>).

Multivariate analysis: multivariate ANOVA models were fitted including all variables univariately associated with IgG levels. The best model was determined using backward

elimination using Wald's Chi-square tests for individual variables. Analysis performed by Camila Franca and Ivo Mueller using STATA (version 12, StataCorp).

2.7.3 Seroepidemiology of recombinant *P. vivax* proteins in Papua New Guinea patients

2.7.3.1 Longitudinal cohort study design and sample collection

In order to investigate protein responses that correlated with clinical protection, we screened a subset of proteins using plasma samples collected in a longitudinal cohort from PNG described in detail (Lin et al., 2010). Samples were collected with ethical approval from the Medical Research and Advisory Committee of the Ministry of Health in PNG and the Walter and Eliza Hall Institute, and written informed consent was given by all adult participants and parents or guardians of child participants. The study enrolled 264 children aged 1-3 years near Maprik, East Sepik Province, in March-September, 2006 and followed for 16 months. Blood samples were taken every 2 weeks and passive case detection occurred throughout the study period. Samples that were positive by PCR for *P. vivax* infection were genotyped to determine the number of genetically distinct infections acquired during the follow-up period. This molecular force of blood-stage infections, molFOB has been described in detail (Mueller et al., 2012, Koepfli et al., 2013). Samples from the 230 children who completed follow-up were used in the Luminex screen described below.

2.7.3.2 Papua New Guinea plasma screening of *P. vivax* recombinant proteins using Luminex

A subset of 6 highly immunogenic proteins from the SI 'Comprehensive screening' were selected by screening against plasma samples from the PNG longitudinal cohort. The proteins were prioritized based on reactivity in SI and Cambodia and on expression levels. The entire ectodomains of 6 *P. vivax* proteins (P12, P41, GAMA, CyRPA, ARP, PVX_081550) and the Cd4d3+d4 tag alone were sub-cloned into a modified plasmid containing a 6-His tag (Bushell et al., 2008). Sumana Sharma expressed the proteins as in section 2.3.3, and purified them by immobilized metal-ion affinity chromatography using HisTrap HP columns on an AKTA Xpress (GE Healthcare) following the manufacturer's instructions.

Camila Franca conjugated the purified proteins to Luminex Microplex microspheres (Luminex Corporation) as described (Kellar et al., 2001), with the following

concentrations per 2.5×10^6 beads: P41, $0.5 \mu\text{g/ml}$; PVX_081550, $1.2 \mu\text{g/ml}$; P12, $0.2 \mu\text{g/ml}$; GAMA, $0.015 \mu\text{g/ml}$; ARP, $0.09 \mu\text{g/ml}$; CyRPA, $1.5 \mu\text{g/ml}$; and Cd4, $2 \mu\text{g/ml}$. Coupling efficiency was previously determined by testing for high fluorescence intensity by the reporter fluorochrome using an immune plasma pool from PNG adults with high reactivity to the antigens by ELISA.

Camila Franca performed a Luminex bead array assay as described (Piriou et al., 2009) to measure total IgG response against the proteins in singlicate to 230 PNG patient plasma samples with secondary antibody donkey F(ab')₂ anti-human IgG Fc R-PE (1:100 in PBS) (Jackson ImmunoResearch). Serum from malaria-naïve individuals from Australia was used as negative controls. A dilution series of plasma pooled from highly-immune PNG adult donors was used as a positive control to standardize plate-to-plate variations.

2.7.3.3 Statistical analysis

Camila and Ivo Mueller performed the analyses described below, which were subsequently published (Franca et al., 2016).

To correct plate-to-plate variations, the dilutions of the highly-immune PNG positive control pool were fitted as plate-specific standard curves using a 5-parameter logistic regression model (Giraldo et al., 2002). For each plate, Luminex median fluorescence intensity (MFI) values were interpolated into relative antibody units based on the parameters estimated from the plate's standard curve. Antibody units ranged from 1.95×10^{-5} (i.e., equivalent to 1:51200 dilution of the immune pool) to 0.02 (1:50). To account for the background reactivity to the Cd4-tag, antibody levels were re-scaled by using linear regression to estimate the antibody levels that would be detected if reactivity to the Cd4-tag was zero, as follows:

$$\log(AB_meas) = \log(AB_true) + \beta * \log(Cd4)$$

where AB_meas = measured antibody level, AB_true = true antibody level to a given antigen, and Cd4 = measured antibody level to the Cd4-tag.

Associations between antibodies and age and exposure were assessed using Spearman rank correlation, and differences with infection using 2-tailed unpaired t-test on log₁₀-transformed values. Negative binomial GEE models with exchangeable correlation structure and semi-robust variance estimator (Stanisic

et al., 2013) were used to analyse the relationship between IgG levels and prospective risk of P. vivax episodes (defined as axillary temperature $\geq 37.5^{\circ}\text{C}$ or history of fever in the preceding 48 hours with a current P. vivax parasitemia > 500 parasites/ μL). For this, IgG levels were classified into tertiles and analyses done comparing children with low versus medium and high antibody levels. Children were considered at risk from the first day after the initial blood sample was taken. The molFOB, representing individual differences in exposure, was calculated as the number of new P. vivax clones acquired per year at risk, and square-root transformed for better fit (Koepfli et al., 2013). All GEE models were adjusted for seasonal trends, village of residency, age, and molFOB. In multivariate models that included all antigens univariately associated with protection, the best model was determined by backward elimination using Wald's Chi-square tests for individual variables. To investigate the effect of increasing cumulative IgG levels to the combination of antigens on the risk of P. vivax episodes, we assigned a score of 0, 1, and 2 to low, medium, and high antibody levels, respectively, and then added up the scores to the 6 antigens to generate a breadth score per child. The breadth score was then fitted as a continuous covariate in the GEE model described above. Analyses were performed using STATA version 12 (StataCorp) or R version 3.2.1 (<http://cran.r-project.org>).

2.8 Commonly used buffers and solutions

The several buffers and solutions that were regularly used throughout this thesis work are summarized in Table 2.4 [Jessica Hostetler (JBH), Camila Franca (CF), Susana Campino (SC)].

Table 2.4: Buffers, media and solutions

Name	Recipe	Usage	Produced by
Coating Buffer	0.015 M sodium carbonate, 0.035 M sodium bicarbonate, pH 9.6	ELISAs	Long lab. at NIH, JBH and CF at WEHI
Diethanolamine buffer	10% diethanolamine, 0.5 mM MgCl ₂ , pH 9.2)	ELISAs	Wright Lab. and JBH
Freestyle supplemented media	Freestyle™ 293 Expression Medium (Invitrogen, USA), 50 mg/l G418, 10000 units/l pen/strep, and 1% HI-FBS	HEK293E cell culture	Wright Lab., William Proto, Sumana Sharma, and JBH
HBS	0.14M NaCl, 5mM KCl, 2mM CaCl ₂ , 1mM MgCl ₂ , 10mM HEPES	<i>P. vivax</i> protein dialysis; recipe from Cecile Crosnier-Wright.	10X HBS: WTSI Media Team, JBH at NIH
HI-FBS Complete media	90% incomplete media, 10% HI-FBS, 0.2% sodium bicarbonate	Parasite culturing and schizont enrichment	JBH and SC
Incomplete media	RPMI, 25 mM HEPES, 50 μM hypoxanthine 0.2% gentamycin	Parasite culturing and schizont enrichment	JBH and SC
Modified McCoys media	McCoys media, Ascorbic acid 5ug/L, MgSO ₄ 16ng/L, hypoxanthine 10ng/L, gentamycin 2mL/L (from 10mg/mL stock), 25% heat inactivated AB serum	<i>P. vivax ex vivo</i> parasite culturing	Chanaki Amaratunga
PBS		<i>P. falciparum</i> / <i>P. vivax</i> proteins, western blotting	10X PBS: WTSI Media Team, Invitrogen at NIH

3 TRANSCRIPTOME PROFILING BEFORE INVASION

3.1 Introduction

Previous work aimed at understanding *P. vivax* merozoite invasion of reticulocytes has focused on a narrow set of genes, almost all of which are homologs of *P. falciparum* genes known to be involved in invasion, primarily because *P. vivax* is very difficult to study directly. There is no established *in vitro* culture system for *P. vivax*, despite repeated efforts by multiple groups, meaning that almost all *P. vivax* research relies on samples taken directly from clinical infections. However, *P. vivax* infections are often of low parasitemia due to the parasite's preference for invading reticulocytes, which limits the amount of material obtainable from clinical samples. A search of PubMed finds nearly 50% (311/637) of publications from 1970-2016 related to *P. vivax* merozoites and/or invasion focus on just 3 proteins: DBP, MSP1, and AMA1. These 3 proteins, and several others with shorter publication lists such as the reticulocyte binding proteins (RBPs), are certainly important for invasion, but there are almost certainly other genes encoding for proteins involved in invasion that are not currently being studied at all. Many of these genes may well be *P. vivax*-specific, and so will not be picked up by the approach of focusing only on those genes with clear homologues in *P. falciparum*. To provide a more unbiased assessment of genes that may be important for *P. vivax* merozoite invasion of reticulocytes, I used RNA-Seq to identify genes upregulated during

the schizont stage of the intraerythrocytic development cycle (IDC), when invasive merozoites are developing.

There are several published transcriptome studies for *P. falciparum* (Bozdech et al., 2003, Otto et al., 2010), which used microarray technology and RNA-Seq respectively (described in section 1.2.2 in detail). Current analyses of *P. vivax* transcription rely primarily on 2 microarray experiments from Bozdech *et al.* (Bozdech et al., 2008) and Westenberger *et al.* (Westenberger et al., 2010). The data in these studies provide a strong foundation for understanding transcription during invasion, but each lacks some information. The Bozdech *et al.* microarray experiment lacks complete genome coverage, and so any genes not present in the initial *P. vivax* Sal 1 reference genome annotation published in 2008 will not have a transcriptional profile. The Westenberger *et al.* microarray experiment covers additional unannotated *P. vivax* genome sections; however, it lacks abundance comparisons during the asexual life cycle, which is critical for understanding invasion-related transcription. We chose to use RNA-Seq, rather than microarray technology, because it will enable us to produce unbiased transcript abundance data during the schizont stage that is not constrained by the specific probes used on a microarray chip. RNA-Seq will also allow definition of the boundaries of genes, such that it could uncover novel gene transcripts, alternative splicing events, validate and/or correct current gene models, and predict 5' and 3' untranslated regions. RNA-Seq therefore provides an opportunity to generate a more definitive list of potential invasion associated genes in *P. vivax*.

No RNA-Seq data had yet been published for *P. vivax* at the time of this work, from either clinical isolates or primate models. As *P. vivax* has no reliable long-term *in vitro* culture system, samples could only be obtained from laboratory-controlled primate infections or clinical samples from patients from *P. vivax* endemic areas. Given that we wanted to understand erythrocyte invasion in naturally-occurring human infections, which may differ from erythrocyte invasion in artificially infected primates, samples from human patients were preferable. However, such a study is technically and logistically challenging. Firstly, processing field isolates can be difficult because -80°C storage and transport are not always available. We were aided in this aspect through access to the NIH laboratory field site in Pursat Province, Cambodia (with Rick Fairhurst at LMVR/NIH and Socheat Duong at Cambodian National Center for Parasitology, Entomology, and Malaria Control), which gave us access to *P. vivax* patients and

laboratory facilities to process and ship samples. Secondly, the preference of *P. vivax* for reticulocytes means that parasitemias are regularly under 1%. This leaves little parasite RNA and an overwhelming majority of human RNA (from leukocytes and erythrocytes). We therefore included sample processing steps to remove host leukocytes, and further enrich for the parasite and hence parasite RNA. Finally, *P. vivax* clinical isolates are often asynchronous, with all stages present in a single blood draw. Given that we were most interested in the schizont stage transcriptome for our study, we included an *ex vivo* culturing step to mature the parasites to the schizont stage, followed by Percoll enrichment of schizonts (Figure 3.1 below). A protocol published near the time of our planning stages provided a starting point to use for our study (Russell et al., 2011).

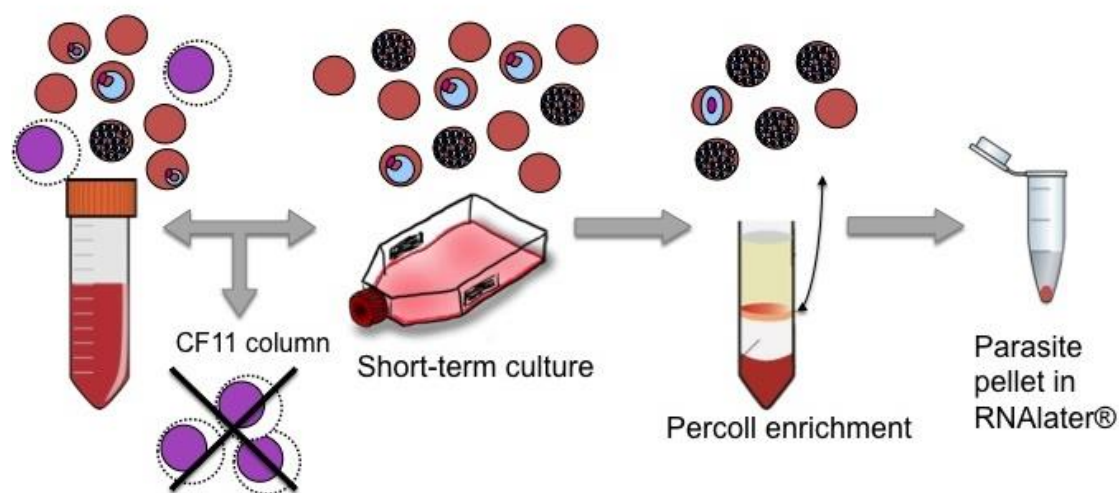


Figure 3.1: Short term *ex vivo* culture of *P. vivax* clinical isolates

The starting material, 16-32 ml of patient blood, contained leukocytes and mixed *P. vivax* life stages. CF11 columns were used to filter leukocytes, which remain in the cellulose matrix as the erythrocytes pass through. Isolates were cultured short-term (11-39 hours) until the majority of parasites were matured to the schizont stage. The culture was enriched for schizonts (with some gametocyte contamination) with a Percoll® gradient. Parasite pellets (up to 200 µl packed cells) were placed in RNAlater® and stored at -20°C. (See section 2.1.3 for greater detail.)

Plasmodium genomes are relatively compact, meaning that the transcribed 5' and 3' UTRs of adjacent genes can overlap. Deconvoluting which gene an RNA-Seq read comes from requires strand-specific Illumina RNA sequencing, but strand-specific RNA-Seq for *Plasmodium* samples was not a standard process during the time of our study. Lia Chappell, a PhD student in the Berriman Laboratory at the WTSI, was actively developing several *Plasmodium* RNA-Seq Illumina library construction protocols at the

start of our study. She provided protocols, training, and guidance for completing the *P. vivax* RNA-Seq libraries before they were sequenced in the WTSI sequencing pipelines (See section 2.1.6 and Figure 3.2 below).

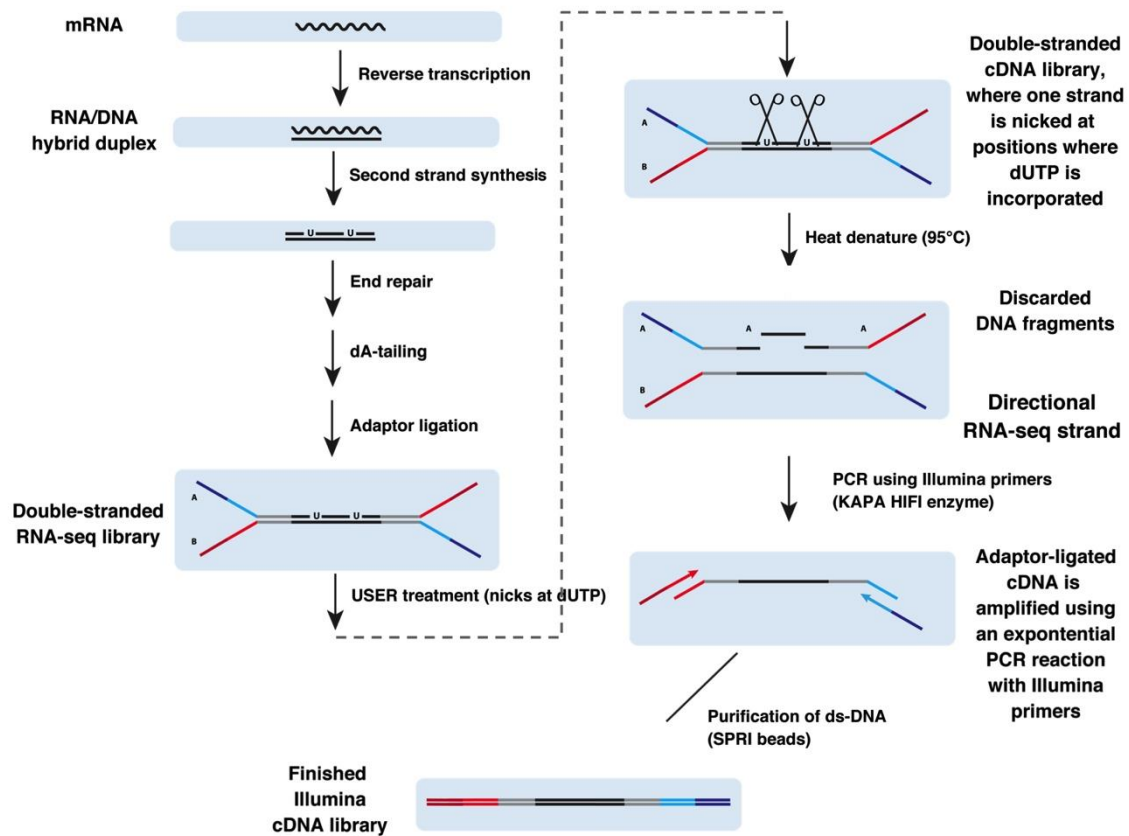


Figure 3.2: RNA-Seq Illumina library construction

Illumina strand-specific library construction protocol developed by Lia Chappell for *Plasmodium* RNA (manuscript in preparation). Messenger RNA (mRNA) transcripts are processed while preserving strand information through the use of second strand synthesis incorporating dUTPs instead of dTTPs. The strand containing dUTPs is ultimately nicked and discarded. An optional final PCR step allows for recovery of samples with concentrations too low for downstream Illumina sequencing (See section 2.1.6 for greater detail). Figure modified from images provided from Lia Chappell.

While obtaining and processing samples was potentially problematic, it was also clear that the downstream data processing and analysis would also be challenging. The *P. vivax* Sal 1 reference genome, published in 2008, represents the *P. vivax* isolate Salvador 1 from an El Salvadoran patient in 1972 that was passed through *Saimiri boliviensis boliviensis* monkeys to generate enough material for whole-genome sequencing (Carlton et al., 2008, Collins et al., 1972). Given its geographical and temporal distance from the

isolates we obtained in Cambodia, there may be significant differences at the genome sequence level. This could substantially hinder read mapping and downstream analysis, particularly in areas of the genome under diversifying selective pressure. Because they are exposed to the adaptive immune system, the merozoite proteins in which we were most interested are particularly likely to suffer from this problem. The MSP3 family serves as a well-known example, where numbers of paralogs differ in different isolates, and large regions are frequently unalignable between isolates (Neafsey et al., 2012, Rice et al., 2014). The *P. vivax* Sal 1 reference genome is also incomplete, particularly at the telomeric ends of chromosomes, and is known to be missing at least 1 gene, EBP or DBL2, which has an erythrocyte-binding domain and may relate to *P. vivax* invasion (Hester et al., 2013). Gene families with high sequence similarities between members, and therefore potentially difficult to assemble properly, are particularly hard to compare. Recently, a new reference genome from a 2014 Indonesian isolate has become available (Auburn et. al, unpublished), containing much more complete telomere assemblies and the most updated gene-model annotation to date. This resource provided a potentially improved assembly to which our sequencing data could be mapped.

3.1.1 Benefits of this study

Transcriptomes of *P. vivax* clinical isolates are not routinely explored, and at the time of this writing, no published RNA-Seq datasets yet exist. Our study aims to benefit the scientific community in several ways. Firstly, the study provides a guide for processing *P. vivax* clinical isolates from collection in the field through Illumina library construction. Secondly, it provides publicly available transcriptome sequencing datasets for *P. vivax* clinical isolates, which will be used for the improvement of reference gene models, as well as evaluating 5' and 3' untranslated regions, novel gene transcripts, and alternative splicing events. Finally, the transcriptomes will inform our understanding of *P. vivax* invasion by identifying genes that are up-regulated in schizonts and enabling inter-isolate comparisons of expression patterns, including the evaluation of multi-gene families known to be important for erythrocyte invasion.

3.1.2 Objectives

- i. To identify genes expressed during the *P. vivax* schizont stage, just prior to invasion of reticulocytes.

- ii. To provide a data resource for building a library of merozoite surface and/or invasion genes, potential asexual stage vaccine candidates, for further study.

3.2 Results

3.2.1 High-quality RNA extracted from *P. vivax* clinical isolates

Chanaki Amaratunga, a staff scientist in the Fairhurst laboratory, selected 4 Cambodian *P. vivax* clinical isolates with high concentrations of ring and/or trophozoite parasite stages, and then cultured the samples *ex vivo* for 11-38.5 hours in the field (Section 2.1.3, Figure 3.1 above), sending parasite pellets in RNeasy® to me at the WTSI (Table 3.1 below). All patients were males aged 14-32 years.

Table 3.1: Patient and sample profiles for *P. vivax* clinical isolates

ID	Sex	Age	Temp (°C)	Parasite density/μl				Volume of blood (mL)	Hours in culture	Parasit- emia %
				Ring	Troph	Schiz.	G'cyte			
PV0563	M	14	38	5544	6099	1504	316	16	38.5	0.3
PV0565	M	32	39	4750	12785	1821	678	32	12	0.4
PV0568	M	28	39	6202	6766	669	387	32	11	0.3
PV0417-3	M	21	38.5	5615	8730	461	384	32	18	0.3

Parasite density counts from starting samples counted from thick smears for rings, trophozoite (Troph.), schizonts (Schiz.), and gametocytes (G'cyte).

I first tested RNA extraction methods to determine which method provided both the highest yield and highest quality. I extracted RNA from a laboratory-adapted *P. falciparum* clone using 3 methods: RNeasy Plus Mini kit (Qiagen), RiboPure Blood kit (Ambion), and a TRIzol (Invitrogen) extraction method adapted from Kyes *et al.* (Kyes *et al.*, 2000) (Table 3.2, below and Sections 2.1.1 and 2.1.2). Overall, the quality of RNA from each method appeared sufficiently high prior to DNA digestion, but the TRIzol method and RiboPure Blood kit performed best in overall yield and quality after 2 rounds of DNA digestion. The RiboPure Blood kit protocol was faster and did not require the use of a fume hood compared to the TRIzol method tested. The TRIzol method, alternatively, was much more economical for large numbers of samples. Given that only 4 samples

were processed, the RiboPure Blood kit was selected for RNA extraction of the *P. vivax* isolates.

Table 3.2: RNA extraction results

Sample	Method	Before DNA digestion		After DNA digestion	
		RNA concentration (ng/μl)*	RNA integrity number (RIN)*	RNA concentration (ng/μl)	RNA integrity number (RIN)
1	RNeasy Plus Mini kit (Qiagen)	134	9.5	73	NA
2	RiboPure Blood kit (Ambion)	302	9.6	210	9.4
3	TRIzol (Invitrogen)	451	9.4	203	9.2

*RNA concentration and RNA integrity number (RIN) computed by Bioanalyzer® (Agilent Technologies, Inc.) using an RNA Nano Chip. RIN is scored 1-10. NA=Not available.

I extracted RNA from the 4 schizont-enriched *P. vivax* isolates, which had been stored in RNAlater® (Ambion) after collection from patients with *P. vivax* malaria and short-term culture carried out in Pursat, Cambodia (ClinicalTrials.gov Identifier: NCT00663546): PV0417-3, PV0563, PV0565, and PV0568. RNA extraction using the RiboPure Blood kit and 2 rounds of DNA digestion (Section 2.1.4) yielded high-quality RNA with concentrations abundant for downstream RNA library construction (Table 3.3 below).

Table 3.3: RNA extraction results after DNA digestion

Patient ID	Blood volume (ml)	Hours in culture	Parasite count (x10 ⁷ / μl)	RNA conc. (ng/μl)	Total volume (μl)	Total RNA (μg)	RIN*
PV0563	16	38.5	2.815	369	80	29.52	9.8
PV0565	32	11	0.387	133	80	10.64	9.7
PV0568	32	18	7.3	408	80	32.64	10
PV0417-3	32	12	2.975	231	80	18.48	9.8

*RNA concentration and RNA Integrity Number (RIN) computed by Bioanalyzer® (Agilent Technologies, Inc.) using an RNA Nano Chip.

To test for contamination with genomic DNA, I used an aliquot of extracted RNA to make cDNA using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems). I then used PCR to amplify a region of the *PvDBP* gene (PVX_110810) containing an intron (Section 2.1.5, Figure 3.3 below). The RNA extracts yielded no

discernible bands to indicate genomic DNA contamination; however, all 4 cDNA samples showed both a bright band corresponding to the product with the intron excised (211 bp) and a faint band corresponding to the expected product if the intron was present (346 bp). This indicated either some genomic DNA might still be present and/or a proportion of RNA was incompletely spliced. While this could raise some problems in interpreting sequencing reads, it was judged that the risk of additional loss in quality from a third round of defrosting and digestion was too great, and no further DNA digestion was performed. Potential DNA contamination in the sequencing output was analysed further below.

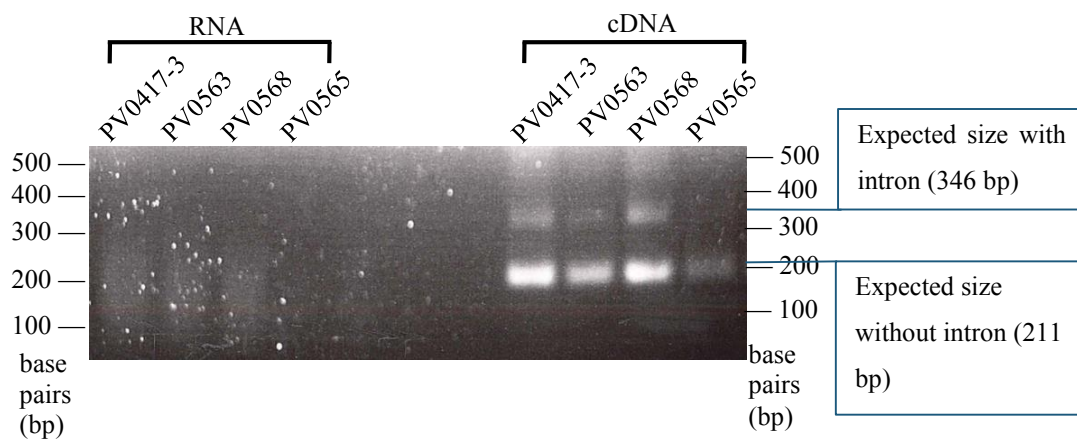


Figure 3.3: Minor DNA contamination and/or incompletely spliced transcripts in *P. vivax* RNA extractions

Polyacrylamide gel showed no discernible bands when testing RNA directly from the 4 Cambodian clinical isolates (PV0417-3, PV0563, PV0568, PV0565) but faint bands were visible after converting RNA to cDNA and amplifying a region of *PvDBP* containing an intron.

3.2.2 Library and mapping statistics

Since the extracted RNA was very high quality (Bioanalyzer RIN > 9.7), I proceeded with the strand-specific Illumina library protocol developed and shared by Lia Chappell with a starting input of 5 µg per sample. The vast majority of the RNA was ribosomal (as with any initial extraction), and therefore, we needed a strategy for its removal. Based on Lia's prior experiments, I removed rRNA through the use of oligo(dT) conjugated to magnetic beads. This step selected for transcripts with poly-A tails, but resulted in the loss of any non-coding RNAs. Future experiments using other methods, such as exonuclease treatment, are planned to study the non-mRNA transcriptome for these

isolates. The yield after the final adapter ligation step necessitated the use of 8 cycles of PCR to reach the DNA quantification required for sequencing on the Illumina HiSeq. The final barcoded strand-specific Illumina RNA sequencing libraries (section 2.1.6) for the 4 *P. vivax* clinical isolates were sequenced together in a single lane on an Illumina HiSeq.

The number of paired-end reads per sample ranged from 55 million to 63 million (Table 3.4). The sequencing and mapping statistics were remarkably similar between the samples, with an average genome coverage ranging from 151x to 186x. The overall amount of ribosomal RNA contamination was very low at 1% per sample, but these rRNA regions corresponded to the regions with the highest coverage overall (maximum coverage, genome-wide in Table 3.4). The samples were also highly pure, with less than 1% human or *P. falciparum* contamination (Figure 3.4).

Table 3.4: RNA-Seq mapping statistics to the *P. vivax* P01 genome

	PV0563	PV0565	PV0568	PV0417-3
Total reads	55720594	52651506	63124660	64316506
Mapped to unique locations^a	47405586	45030514	53779293	55393526
% Mapped to unique locations^a	85	86	85	86
Reads mapped to rRNA^b	833178	656333	606261	818503
% Reads mapped to rRNA^b	1	1	1	1
Fold coverage^b	159	151	180	186
% Genome not covered^b	23	23	23	23
% Genome covered, 1-fold^b	77	77	77	77
% Genome covered, 20-fold^b	59	57	62	61
% Genome covered, 100-fold^b	33	29	34	36
Max. coverage across exon sequences^b	53426	55738	63269	65164
Max. coverage, genome-wide^b	99569	81806	177536	69293

^aReads mapped using TopHat including only reads mapping to unique locations

^bCoverage determined using Bedtools

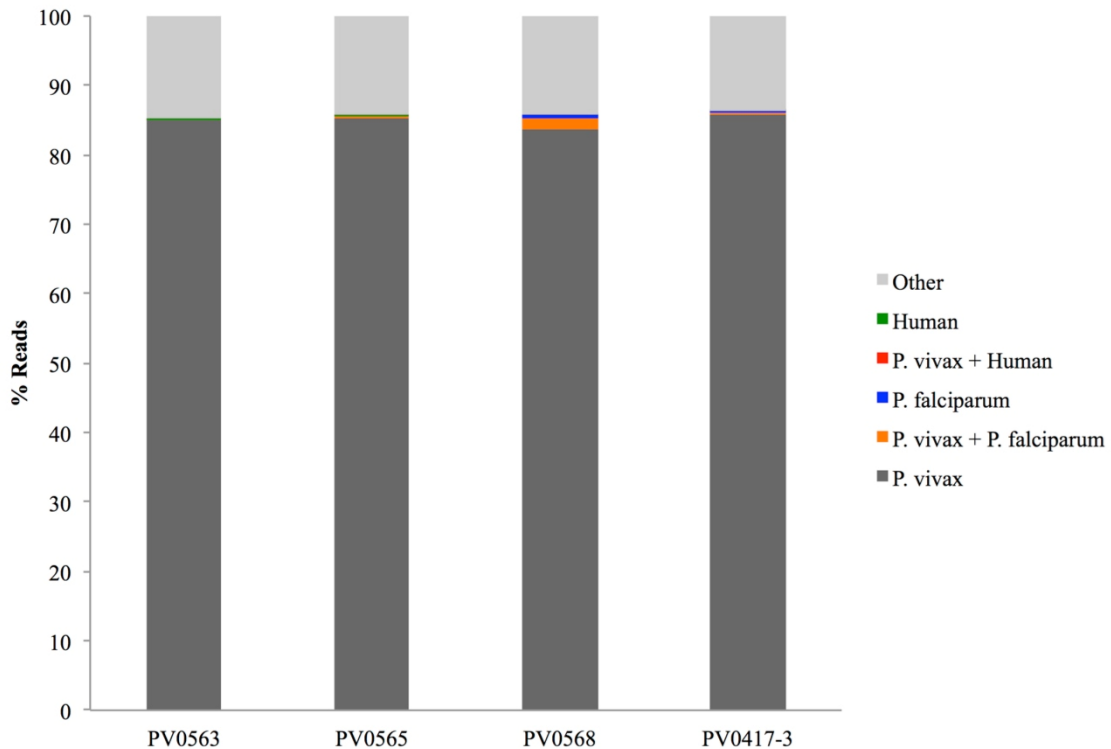


Figure 3.4: Illumina read alignments against the *P. vivax*, *P. falciparum*, and human reference genomes

P. vivax clinical isolate RNA-Seq data contained less than 1% contamination with either human or *P. falciparum* RNA/DNA. A portion of reads, 14-15% per isolate, did not map to *P. vivax* P01, *P. falciparum* 3D7, or human reference genomes.

3.2.3 Assessing DNA contamination

In order to understand the quality of the sequences and assess the amount of genomic DNA contamination, I compared the depths of coverage across 3 types of genomic regions: exons, introns, and intergenic (“other”). The intergenic region includes both true intergenic sequences, but also includes 5’ and 3’ UTR regions, non-coding sequence regions, and rRNA regions. Some of the intergenic region would therefore still be expected to have a significant amount of sequence coverage in RNA-Seq, whereas coverage of introns should be almost completely absent in sequence reads from RNA, so can be used to measure the level of genomic DNA contamination in the samples. Note that this approach to analysing potential DNA contamination is complicated by the presence of incompletely spliced transcripts, which would produce reads from some introns, but is still a useful proxy.

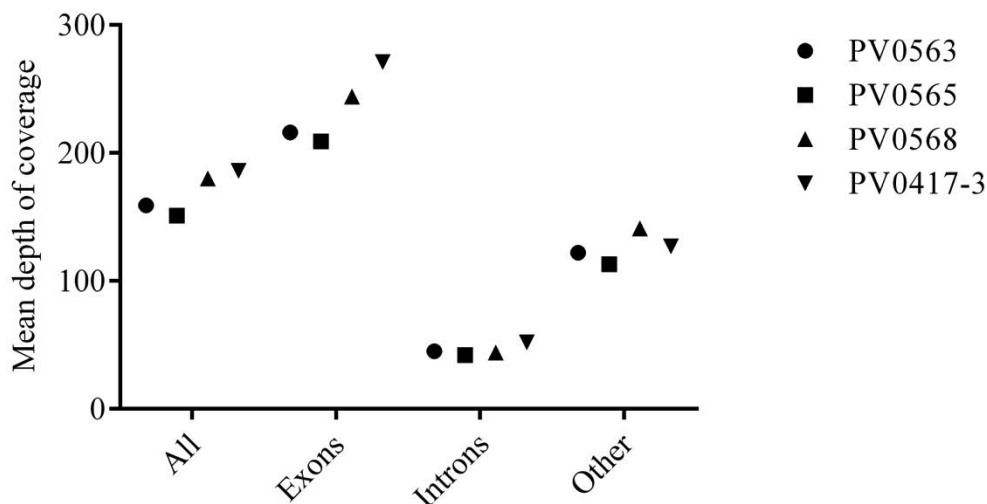


Figure 3.5: Sequence breadth distributions across the *P. vivax* P01 reference genome in exons, introns, and other regions

Mean depth of RNA-Seq coverage for each *P. vivax* clinical isolate in exons, introns, and all remaining regions (“other”). Between isolates, average exon coverage varied from 209- to 271-fold, and average intron coverage varied 42- to 52-fold.

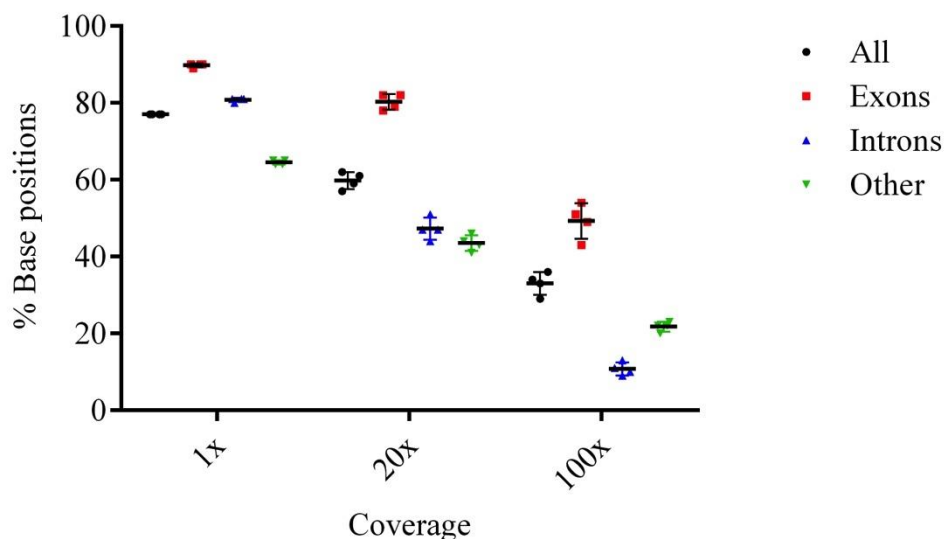


Figure 3.6: Sequence depth distributions across the *P. vivax* P01 reference genome in exons, introns, and other regions

Each isolate is represented by a point, with mean and SD shown. The percentage of bases covered at 1-fold, 20-fold, and 100-fold coverage for exons, introns, and all remaining regions (“other”). Exons consistently had the highest percentage of bases covered at every threshold.

The coverage across exons was about 5 times that of introns on average for each isolate, suggesting that while the majority of reads originated from RNA, some incompletely spliced transcripts and/or genomic DNA may have persisted despite 2 rounds of DNA digestion (Figure 3.5). However, these samples largely represent a single life stage, and will have a highly skewed expression profile, with some genes highly expressed and others not at all. A comparison of the average coverage across introns and exons across the whole genome is therefore not necessarily the most useful approach.

Considering the coverage for each group at low, medium, and high coverage depth cut-offs (1x, 20x, 100x), more exon bases are covered at all the coverage depths (Figure 3.6). At the lowest, 1x threshold, most introns and exon bases are covered (80-90%). At the higher thresholds, however, about 35% more exon bases are consistently covered than intron bases (i.e., at 20x, about 80% of exon bases are covered compared to about 45% of intron bases for each sample; at 100x, about 45% exon bases are covered compared to about 10% of intron bases). This may indicate that genomic DNA contamination is present as 80% of intron bases have some coverage, but the fact that 10% of intron bases have coverage of over 100x (far above the 42-52x average), may indicate that incompletely spliced transcripts play a role in covering introns.

To investigate this more carefully, I compared the average depth of exons and introns for 3 groups: 50 multi-exon genes (~1% of total genome) with the highest, middle, and lowest coverage (setting a minimum coverage depth of 20x, to eliminate any genes with mapping issues rather than true low expression). If genomic DNA represented the primary cause of the coverage of introns, one would expect lower abundance genes to contain a higher ratio of exon-to-intron coverage than higher abundance genes. If intron coverage related primarily to incompletely-spliced transcripts, the exon-to-intron ratio would be similar between the groups. The results suggested that both genomic DNA and incompletely-spliced transcripts contribute to intron coverage (Table 3.5). The exon-to-intron ratio increased from the lowest to highest covered genes, indicative of genomic DNA contamination. However, after the coverage values are normalized by subtracting the average depth of the introns from the lowest covered genes group (a surrogate for the general level of genomic DNA contamination of each sample), the exon-to-intron ratios are much more similar.

Table 3.5: Exon-to-intron coverage comparison for 50 (~1% of the genome) lowest, middle and highest coverage genes

	Average coverage						Exon-to-intron ratio		
	L		M		H		L	M	H
	exon	intron	exon	intron	exon	intron			
PV0563	24	14	174	29	3520	288	1.7	6	12.2
PV0565	22	14	141	28	3518	277	1.6	5	12.7
PV0568	37	20	188	34	3262	267	1.9	5.5	12.2
PV0417.3	37	19	209	35	3746	307	1.9	6	12.2
Normalized									
PV0563	10	0	160	15	3506	274	NA	10.7	12.8
PV0565	8	0	127	14	3504	263	NA	9.1	13.3
PV0568	17	0	168	14	3242	247	NA	12	13.1
PV0417.3	18	0	190	16	3727	288	NA	11.9	12.9

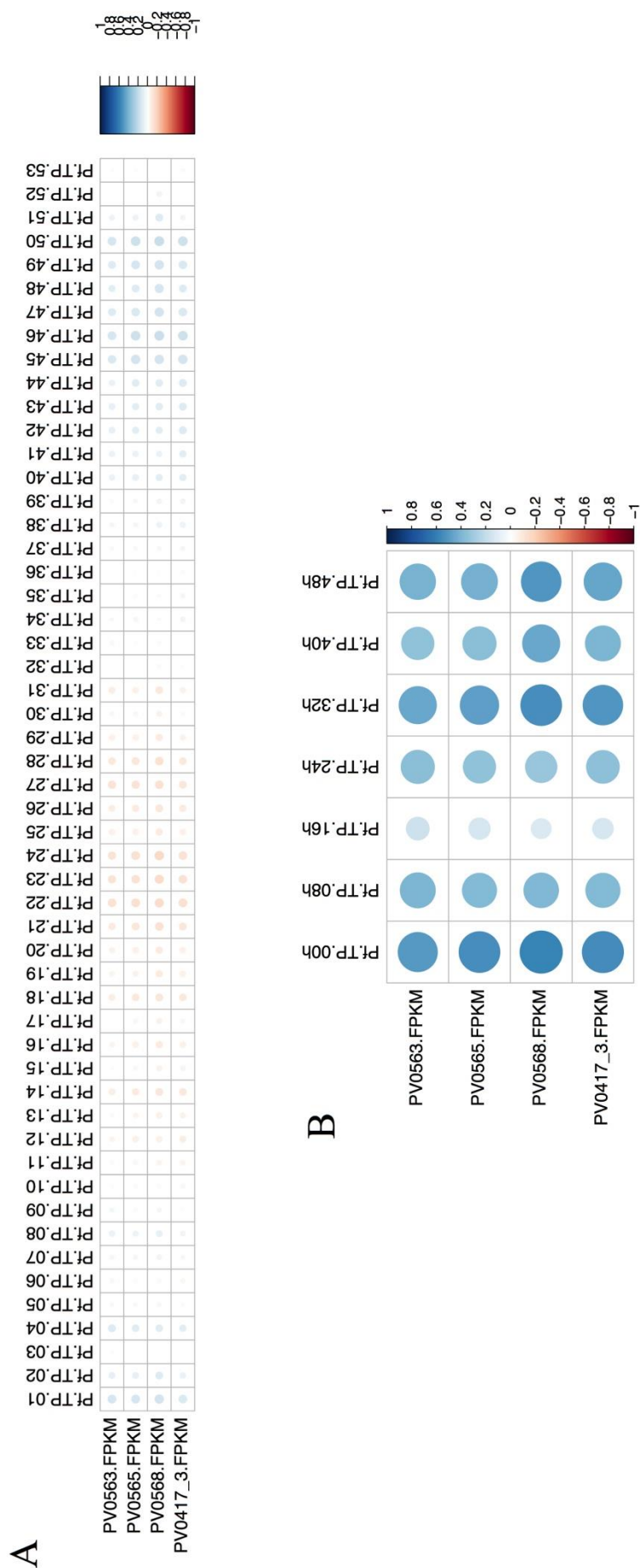
L=lowest, M=Middle, H=Highest; NA=Not applicable

Thus, there appears to be a lower level of genomic DNA contamination contributing 15-20x coverage throughout the genome, or roughly 10% of the average coverage. Above this threshold, coverage in the introns is likely caused by incompletely spliced transcripts. Overall, the isolates appear to have similar levels of both DNA contamination and incompletely spliced transcripts as evidenced by the similar exon-to-intron coverage ratios (1.6-1.9 for genes with lowest coverage; 5-6 for genes with middle coverage; 12.2-12.7 for genes with highest coverage), ensuring that the downstream expression analysis will not be impacted. However, since genomic DNA contamination would mostly impact genes with lower coverage/expression it would be prudent to set a conservative minimum coverage threshold (at least 20x) for including genes in any expression analysis. With the coverage generated in these isolates, this corresponds to a minimum FPKM of 5, and therefore a conservative FPKM lower limit of 10 was set for downstream analysis.

3.2.4 Assessing asexual stage time point

While thin blood smears made after short-term culture and prior to the addition of RNAlater® to the parasite pellets suggested that samples were highly enriched for schizonts, it was also important to test this computationally. In order to assess this, I compared the expression values (FPKMs) for all genes to their one-to-one orthologs in both the published *P. falciparum* microarray dataset (53 time points across the IDC, using enrichment values) and *P. falciparum* RNA-Seq dataset (7 time points across the IDC, using RPKMs) (Figure 3.7). I also compared my data to the published *P. vivax* microarray dataset (9 time points across the IDC, using enrichment values) (Figure 3.8). This work was significantly aided by Lia Chappell who provided an initial file listing one-to-one orthologs between *P. falciparum* 3D7 and *P. vivax* Sal 1 as well as some R scripts.

Pearson correlation coefficients using microarray enrichments values and the clinical isolates' FPKMs showed the highest similarity to the latest time points of the IDC, the schizont stage of the life cycle, for both *P. falciparum* 3D7 and *P. vivax* (Figure 3.7A and Figure 3.8A,C). The correlation to the second *P. vivax* microarray sample SD2 (Figure 3.8B) also showed high similarity to time-point 6, corresponding to the late trophozoite stage. However, this sample correlates less well to both the other 2 *P. vivax* microarray samples and the *P. falciparum* microarray datasets, and may have represented a less synchronous sample (Bozdech et al., 2008). Therefore, this sample is likely a less reliable sample to use for comparisons. The correlations with the *P. falciparum* RNA-Seq time course showed no obvious best match, though with the weakest match to the 16-hour time point corresponding to the early trophozoite (Figure 3.7B). Overall, these comparisons suggested that schizonts are the dominant stage in all 4 clinical samples.



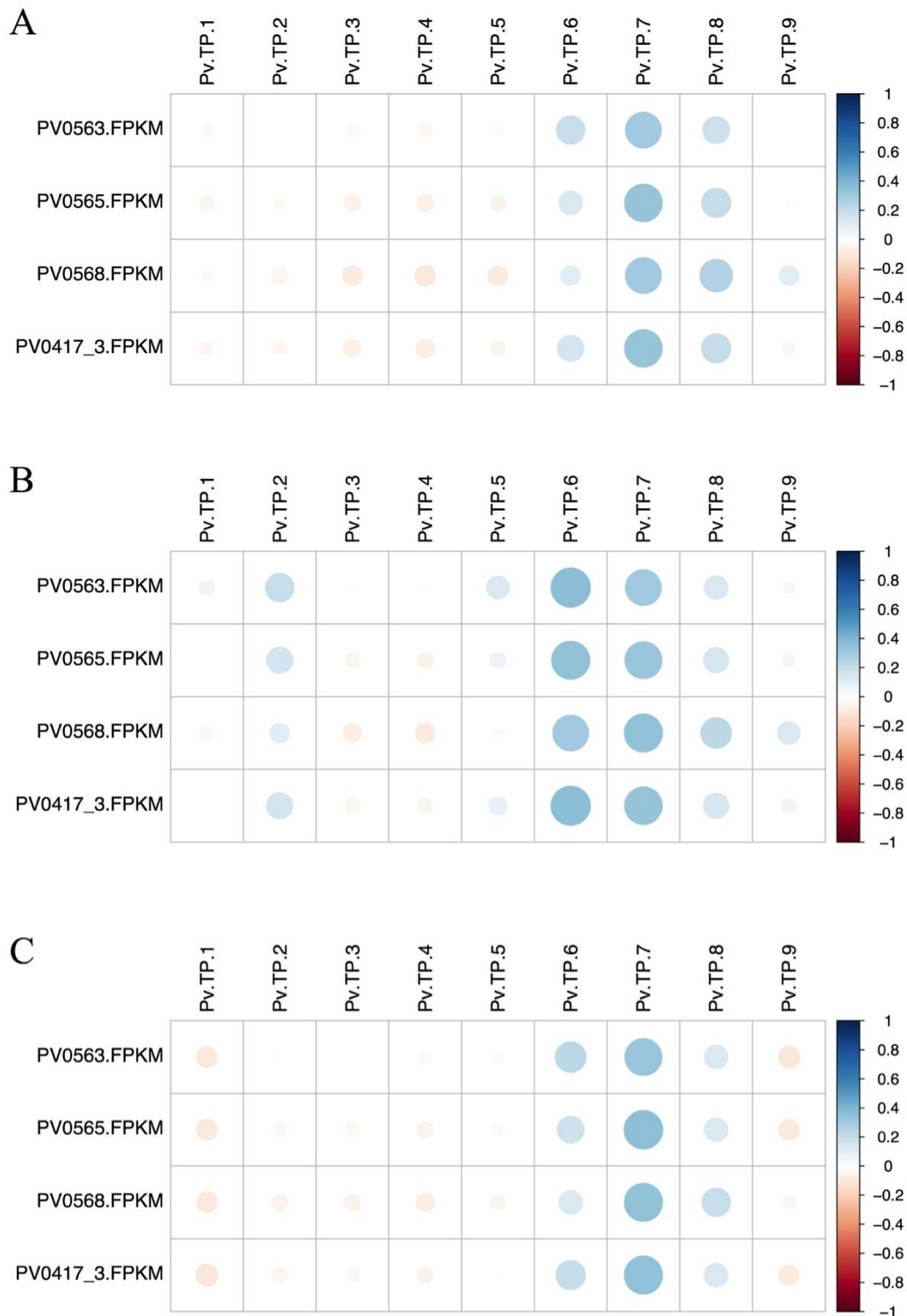


Figure 3.8: *P. vivax* clinical isolates correlate with the early schizont time point from *P. vivax* microarray data

FPKMs from *P. vivax* clinical isolates (PV0563, PV0565, PV0568, PV0417-3) were compared to *P. vivax* microarray data (enrichment values) for 3 isolates (A=SD1, B=SD2, C=SD3) over 9 time points across the 48-hour life cycle using Pearson correlations coefficients. *P. vivax* RNA-Seq data were mapped to *P. vivax* P01 with TopHat, and FPKMs calculated using Cufflinks.

3.2.5 Assessing gametocyte contamination

Thin blood smears made after schizont enrichment established that all samples had some gametocytes present. Significant gametocyte contamination (particularly if very different between samples) could skew the expression analysis, as the gametocyte fraction would diminish the signal of the asexual schizont stage of interest. There are no published RNA-Seq data from *P. vivax* gametocytes, making assessment of the level of gametocyte-specific transcripts challenging. However, RNA-Seq data from several *P. berghei* life stages were available and recently published (Otto et al., 2014). The study contained expression data for 2 biological replicates for the asexual stages (ring, trophozoite, schizont) and gametocytes, and 2 time points for ookinetes. To assess gametocyte contamination, I compared the *P. vivax* isolate's FPKM expression values to all the available *P. berghei* time point FPKMs using Pearson correlation coefficients (Figure 3.9). The *P. vivax* isolates showed the highest correlations with *P. berghei* schizonts, further validating schizont enrichment in the *P. vivax* isolates. All blood-stage comparisons had higher correlations than the gametocyte- and ookinetes-stage comparisons, suggesting that there was no significant gametocyte contamination in the schizont-enriched *P. vivax* samples.

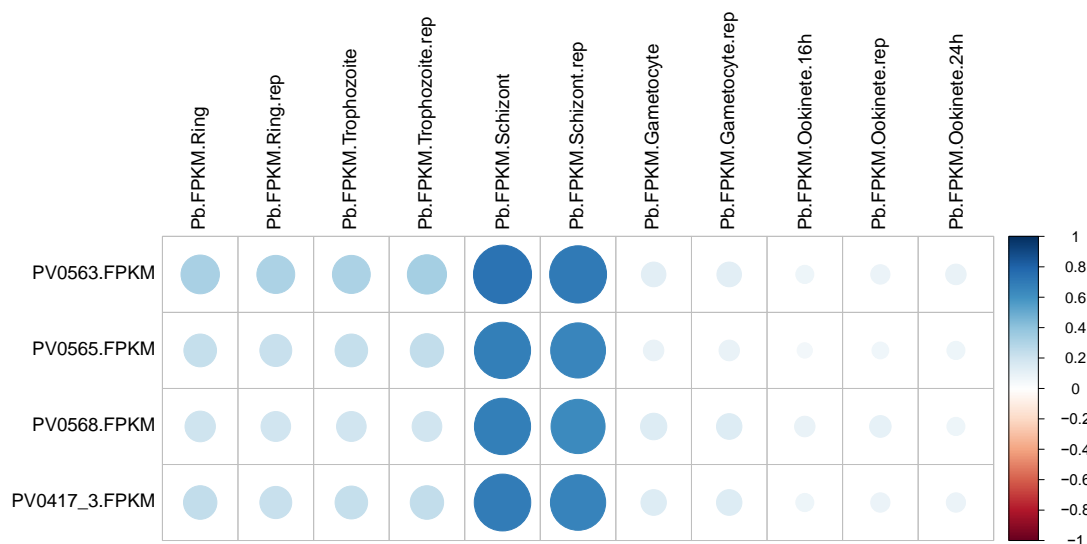


Figure 3.9: *P. vivax* and *P. berghei* schizont RNA-Seq data are highly correlated

FPKMs from *P. vivax* clinical isolates (PV0563, PV0565, PV0568, PV0417-3) were compared to *P. berghei* RNA-Seq data (FPKMs) for 2 replicates (single replicate only for the 24-hour ookinetes stage) over 6 time points during both asexual and sexual stages using Pearson correlation coefficients. *P. vivax* RNA-Seq data were mapped to *P. vivax* P01 with TopHat, and FPKMs calculated using Cufflinks.

3.2.6 Using RNA-Seq data to improve the *P. vivax* reference genome

The *P. vivax* RNA-Seq data was used by Ulrike Böhme in the Parasite Genomics Group at the WTSI to improve 352 gene models in the *P. vivax* Sal 1 reference genome (the only available reference genome at the time of the completion of sequencing). This included detecting 20 novel gene transcripts, merging 7 pairs of genes, splitting 3 genes into 6 separate genes and changing the exons of 350 genes (including adding exons, deleting exons, and changing exon coordinates). No alternative splicing events were detected, however (using the default parameters with Cufflinks tool suite). A detailed list of improvements can be found in Supplementary Table A.

3.2.7 Comparing mapping to *P. vivax* Sal 1 and *P. vivax* P01 reference genomes

Given that 2 reference genomes were available for the analysis phase of this project, I wanted to evaluate the mapping quality to each genome. Running identical assembly parameters, 1% more reads mapped on average to the *P. vivax* Sal 1 reference genome compared to the *P. vivax* P01 reference genome (comparing only the 14 chromosomes) (Table 3.6). Therefore, both reference genomes were fairly similar in terms of overall mapping success. However, the *P. vivax* P01 reference assembly contained more complete chromosomal assemblies (especially in telomere regions) with 24.2 MB assembled compared to *P. vivax* Sal 1 with 22.6 MB assembled and linked to chromosomes. The telomeric regions are highly repetitive and this combined with the unique mapping requirement (using TopHat) may partially explain why slightly fewer reads mapped to *P. vivax* P01 than to *P. vivax* Sal 1. Because we were interested in comparing the largest set of annotated genes, the *P. vivax* P01 genome was used for all inter-isolate expression analysis.

Table 3.6: Reads mapping to *P. vivax* Sal 1 vs. *P. vivax* P01

	Reads mapping to chromosomes			
	PV0563	PV0565	PV0568	PV0417-3
<i>P. vivax</i> Sal 1*	47349527	44926798	54397463	55117488
<i>P. vivax</i> P01*	46963376	44669990	53061483	54933162
Total reads	55720594	52651506	63124660	64316506
% Mapped <i>P. vivax</i> Sal 1	85	85	86	86
% Mapped <i>P. vivax</i> P01	84	85	84	85

*Reads mapped uniquely with TopHat

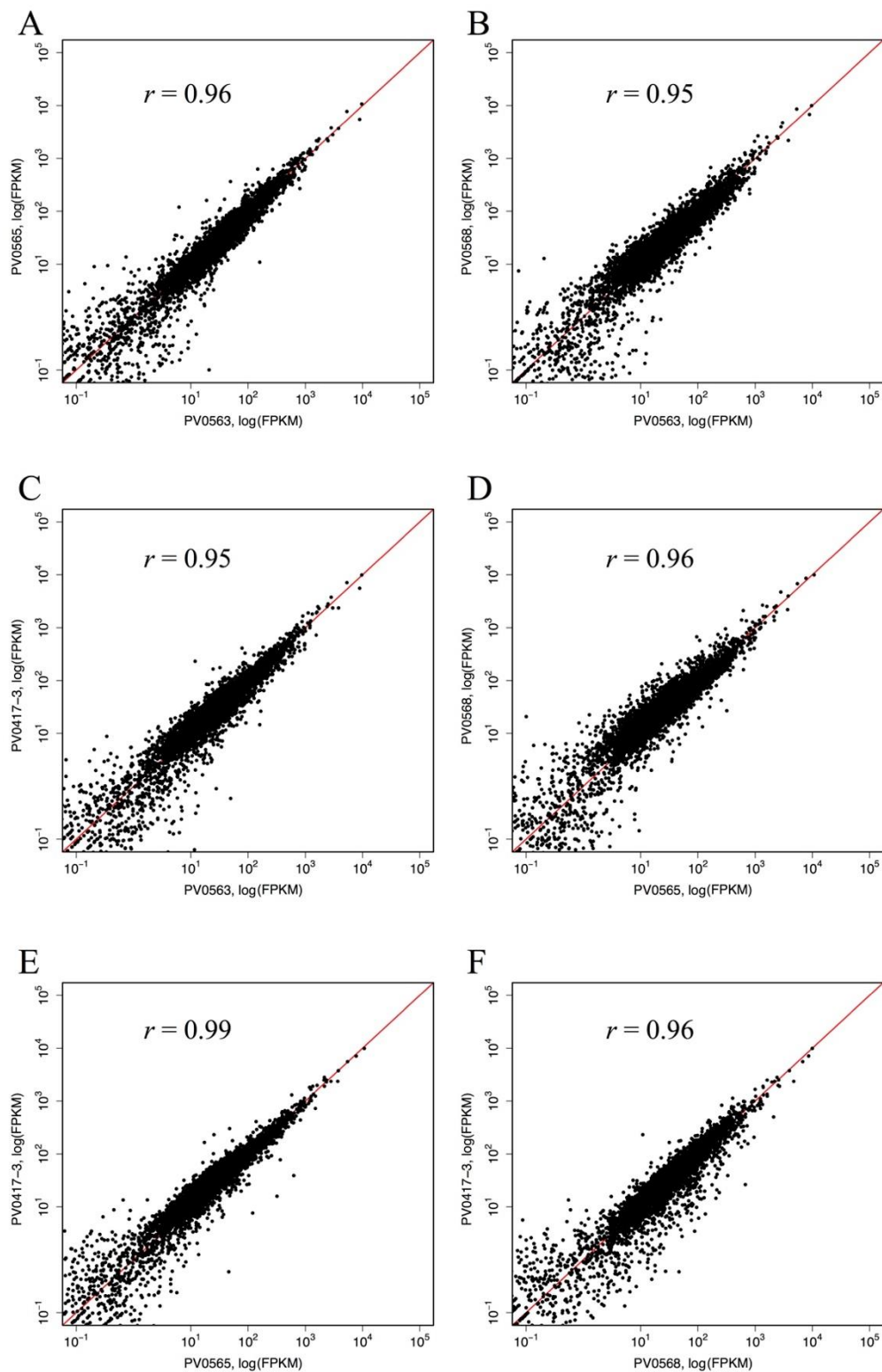


Figure 3.10: Expression in *P. vivax* clinical isolates is highly correlated

(A-F) Pairwise plots of the log(FPKM) for each isolate indicated that the expression profiles between samples were highly correlated with Pearson correlation coefficients (untransformed data) ranging from 0.95 to 0.99. *P. vivax* RNA-Seq data were mapped to *P. vivax* P01 with TopHat, and FPKMs calculated using Cufflinks.

3.2.8 Comparing expression data between clinical isolates at genome scale

In order to understand the how similar (or different) each clinical isolate was to each other, I performed several comparisons. The first was to plot all pair-wise comparisons of the 4 isolates using the log(FPKM) for all genes (Figure 3.10). The Pearson correlation coefficients for each pair (untransformed data) were uniformly high at 0.95 to 0.99, indicating the expression profiles between the isolates were highly similar despite being from separate clinical infections, collected at different times, and subjected to different lengths of *ex vivo* culturing, all of which could potentially impact RNA expression. This high degree of similarity also suggested that the data were not heavily skewed by the gametocyte contamination. This also confirmed that the isolates were matured and collected at very similar time points in the life cycle (despite different lengths of *ex vivo* culture).

Normalising the data to enable clear comparisons was a challenge. While the transcription of both *P. falciparum* and *P. vivax* follows a well-characterized cascade over the 48-hour life cycle with RNA abundance for stage-specific genes peaking and declining in a regular cycle (Bozdech et al., 2003, Bozdech et al., 2008), some transcription for each gene is likely to be detected at every stage with deep sequencing. The ideal way to overcome this problem would have been to sequence control samples that contained all life stages in equal proportion. Such samples would have provided the average RNA abundance for each gene, which could have been used to normalize our data and compare to the enrichment values computed in the *P. falciparum* and *P. vivax* microarray datasets. However, given the difficulty in obtaining even these 4 samples, no such mixed infection control samples could be obtained. This lack of normalization complicates attempts to compare between isolates, and also made it difficult to compare data to the published *Plasmodium* transcriptome datasets. Despite these challenges, some general conclusions could be made, as long as the caveats for comparison are borne in mind.

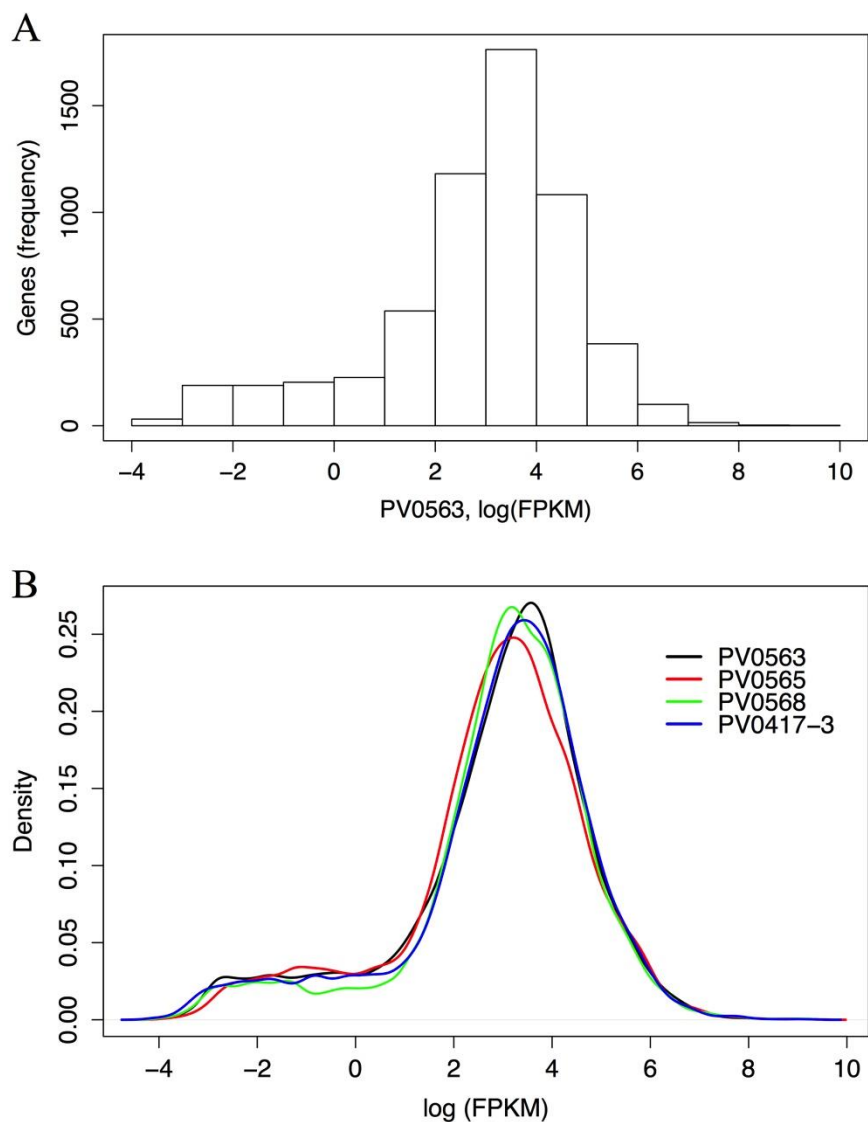


Figure 3.11: FPKM distributions are similar between *P. vivax* clinical isolates

(A) Example histogram with gene frequencies of log(FPKM) for PV0563. (B) Density plots of the log(FPKM) for each of the 4 *P. vivax* clinical isolates (PV0563, PV0565, PV0568, PV0417-3).

The FPKMs for the 4 clinical isolates ranged from 0 to 10,000, with 5290 genes having an FPKM >1. The distribution was very similar in each isolate (Figure 3.11). Over 70% (4704/6672) of the genes had FPKMs ranging from 1 to ~130 ($1^e - 6^e$). More than 25% of genes had FPKMs of 20 to 43 ($3^e - 4^e$). The published *P. vivax* microarray data contained expression data (in at least 1 of the 3 isolates) for 4633 genes. This *P. vivax* RNA-Seq dataset contained expression data for over 450-750* additional genes (*minimum FPKM of 10 for inclusion). Gene naming and annotation changes between

the *P. vivax* microarray dataset, the *P. vivax* Sal 1 dataset, and the *P. vivax* P01 reference make reporting an absolute number difficult. 44% of these genes for which transcription data is available for the first time are annotated as conserved hypothetical proteins, for which little is known. Nearly 100 of the complete list of new genes (including both hypotheticals and named genes) have very high expression (FPKMs over 100, or among the top 12% of expressed genes), including high molecular weight rhostry protein 3, RhopH3, a known invasion-related gene in *P. falciparum* (Sam-Yellowe et al., 1988). The table listing the top-expressed genes unique to the *P. vivax* clinical isolate RNA-Seq data (compared to the published microarray dataset) can be found in Supplementary Table C. These highly expressed genes, for which expression was not previously known, are potentially important for understanding *P. vivax* merozoite invasion of reticulocytes.

3.2.9 Comparing expression between 4 clinical isolates at the individual gene level

While at an overall level the number of expressed genes and FPKM ranges for the 4 clinical isolates was highly similar, I next wanted to evaluate any specific expression differences between the isolates. Standard differential expression analysis compares a set of biological replicates (potentially with technical replicates) to another set of biological replicates after some change in condition (frequently drug pressure) and is commonly performed with tools such as DeSeq or EdgeR (Anders and Huber, 2010, Robinson et al., 2010). In our case we have essentially 4 biological replicates, no technical replicates, and no conditions, as we have no phenotypic categories (e.g., asymptomatic and symptomatic infection) to compare, meaning that applying these approaches is not valid. Looking for differences in our samples is therefore more accurately described as looking at gene expression variability or spread. Before investigating this, I first set a lower FPKM threshold for genes to consider by any method. As discussed above, based on intron coverage indicative of genomic DNA contamination, an initial threshold of 10 FPKMs was set for considering a gene to be expressed. Nearly 64% of genes (4295/6672) have expression over this threshold. However, for this analysis, I was most interested in variable expression among genes that are most relevant in the schizont stage. I therefore considered only the top 50% (3305/6672) of expressed genes, which set a minimum FPKM of 20, in downstream expression variability methods. I then explored variable expression using several different methods.

The first method simply ranked the genes based on how different expression for individual isolates was from the mean expression all 4 isolates, described as the “max fold-change” method. For instance, by comparing all FPKMs to the average FPKM from the group and ranking the genes from highest fold-change to lowest, 56 genes show at least a 2-fold change from the mean in at least 1 isolate. Next I calculated the coefficient of variation (CV) or relative standard deviation (RSD) for the data. This gives a ratio of the standard deviation to the mean for the data, so that we can compare genes with very different means, as is the case with the RNA-Seq data where FPKMs range from 0 to 10000. The third normalized measure of spread is the index of dispersion or variance-to-mean ratio (VMR). This is similar to the first step of the DeSEQ differential expression analysis, which assesses the dispersion (or over-dispersion) of the data for a set of biological replicates, in order to gauge the normal variability or “noisiness” of the biological replicates which will be compared to another condition (Anders and Huber, 2010).

In order to compare how similar the rankings were between the several methods, I arbitrarily compared the top 300 genes (or around 10% of the total included genes) ranked by each method. The ‘max fold-change’ and CV calculations were very similar, as evidenced by the 2 sets overlapping for 256/344 genes (76%) (Figure 3.12). The VMR ranking differed most from the other 2 methods, and overall 130/495 genes (26%) were ranked in the top 300 by all 3 methods. These 130 genes could be considered the most reliable set of variably-expressed genes for further analysis.

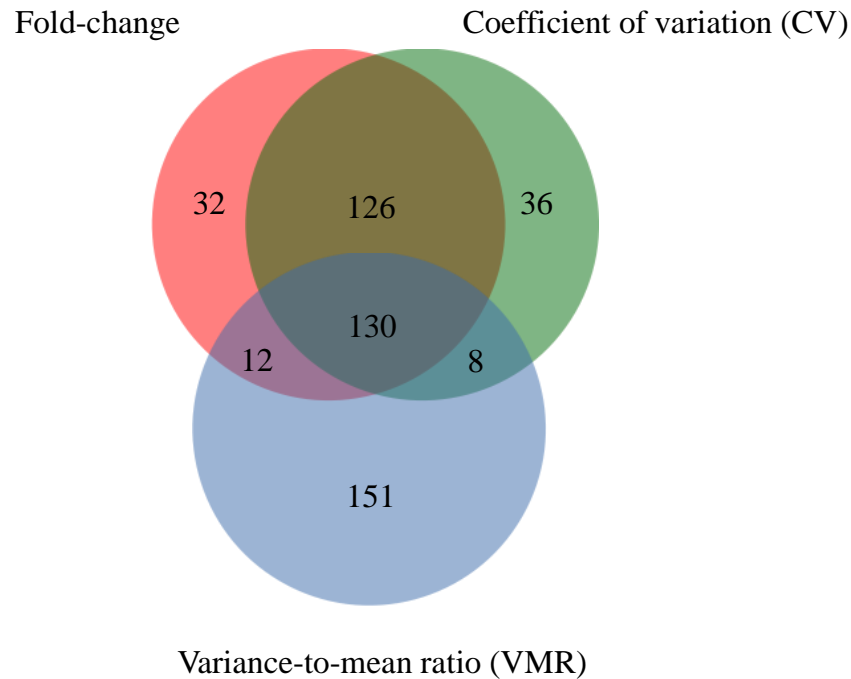


Figure 3.12: Three methods for ranking expression variability in *P. vivax* clinical isolates

Evaluating expression variability between *P. vivax* clinical isolates using fold-change from the mean and 2 normalized measurements of dispersion (coefficient of variation, CV; variance-to-mean ratio, VMR), showed modest agreement, with 26% of genes (130/495) overlapping between the top 300-ranked genes in each approach (setting a minimum average FPKM of 20). The fold-change and CV calculations intersected for 76% of genes (256/344), and the VMR ranking had the least overlap with 31% (142/459) and 30% (138/463) when compared to fold-change and CV, respectively. *P. vivax* RNA-Seq data were mapped to *P. vivax* P01 with TopHat, and FPKMs calculated using Cufflinks.

The 130 genes ranked in the top 300 by each of these 3 methods are summarized in Supplementary Table B. At least 26% of the genes were surface-expressed or invasion-related and/or members of large families, including AMA1, DBP, GAMA, 7 MSP3s, 7 RBPs, and 3 MSP7s. Another 40% (51/129), were either “conserved *Plasmodium* protein, unknown function,” “hypothetical protein,” or “*Plasmodium* exported protein, unknown function,” in keeping with the large number of these genes in the *P. vivax* genome. None of the top 20 differentially expressed *P. berghei* gametocyte genes were present in the top 130 variably-expressed gene set, although 2 variably-expressed genes (PVP01_1258000, gamete egress and sporozoite traversal protein; PVP01_0115300, gamete release protein, putative) did appear to be related to the sexual stage. At least 1 gene from the *P. falciparum* gametocyte expression study (Young et al., 2005), a putative oxidoreductase,

PVP01_1229400 (PF3D7_1325200) was also in the variably-expressed gene set, possibly indicating there were slightly different proportions of gametocytes in the samples. I analysed the list for gene ontology (GO) enrichment using the 77 of these 130 genes that had direct *P. falciparum* 3D7 one-to-one homologs (and therefore for which GO terms were available) and found significant enrichment for invasion and host interacting genes (Table 3.7 below). This is an underestimate of the actual enrichment, as several large *P. vivax* families in the top 130 variably-expressed genes related to the merozoite surface (MSP3 and MSP7) and/or invasion (RBPs) had no direct one-to-one orthologs with *P. falciparum*, and thus had no GO terms.

Table 3.7: Host and invasion genes enriched in top 130 variably-expressed genes from *P. vivax* schizont-stage clinical isolates

GO molecular function	Pf. ref (5159)	Var set (77)	Exp'd	+/-	Fold Enrich.	P value
intramolecular oxidoreductase activity (GO:0016860)	2	2	0.03	+	> 5	2.89E-02
host cell surface binding (GO:0046812)	19	5	0.28	+	> 5	7.19E-04
protein binding (GO:0005515)	146	12	2.18	+	> 5	1.17E-04
binding (GO:0005488)	199	15	2.97	+	> 5	1.52E-05
molecular_function (GO:0003674)	391	18	5.84	+	3.08	9.50E-04
Unclassified (UNCLASSIFIED)	4768	60	71.16	-	0.84	0.00E+00
GO biological process						
interaction with host (GO:0051701)	47	5	0.7	+	> 5	4.75E-02
biological_process (GO:0008150)	313	13	4.67	+	2.78	4.67E-02
Unclassified (UNCLASSIFIED)	4846	65	72.33	-	0.9	0.00E+00
GO cellular component						
microneme (GO:0020009)	20	6	0.3	+	> 5	4.26E-05
pellicle (GO:0020039)	18	5	0.27	+	> 5	5.55E-04
inner membrane complex (GO:0070258)	18	5	0.27	+	> 5	5.55E-04
apical complex (GO:0020007)	57	10	0.85	+	> 5	1.02E-06
apical part of cell (GO:0045177)	59	10	0.88	+	> 5	1.40E-06
membrane (GO:0016020)	89	8	1.33	+	> 5	3.84E-03
cell (GO:0005623)	560	20	8.36	+	2.39	1.07E-02
cell part (GO:0044464)	559	19	8.34	+	2.28	3.07E-02
cellular_component (GO:0005575)	733	23	10.94	+	2.1	2.10E-02
Unclassified (UNCLASSIFIED)	4426	55	66.06	-	0.83	0.00E+00

Gene Ontology (GO) ids from experimental evidence published in scientific literature. Pf. Ref = *P. falciparum* 3D7. Var set (77) = one-to-one orthologs between *P. falciparum* and *P. vivax* P01 for a set of most variably-expressed genes during the schizont stage for 4 clinical isolates. Exp'd=Expected fraction. Fold enrich=Fold enrichment. *P* values <0.05.

3.2.10 Assessing the impact of diversity and mapping on expression data

A number of the genes identified as variably expressed were members of highly-diverse large gene families, including 7 members of the MSP3 family. MSP3 family members are known to be under heavy diversifying selection (Neafsey et al., 2012, Rice et al., 2014). It was highly possible that such diversity meant that the “expression variability” was due to mapping differences between the isolates rather than actual expression differences.

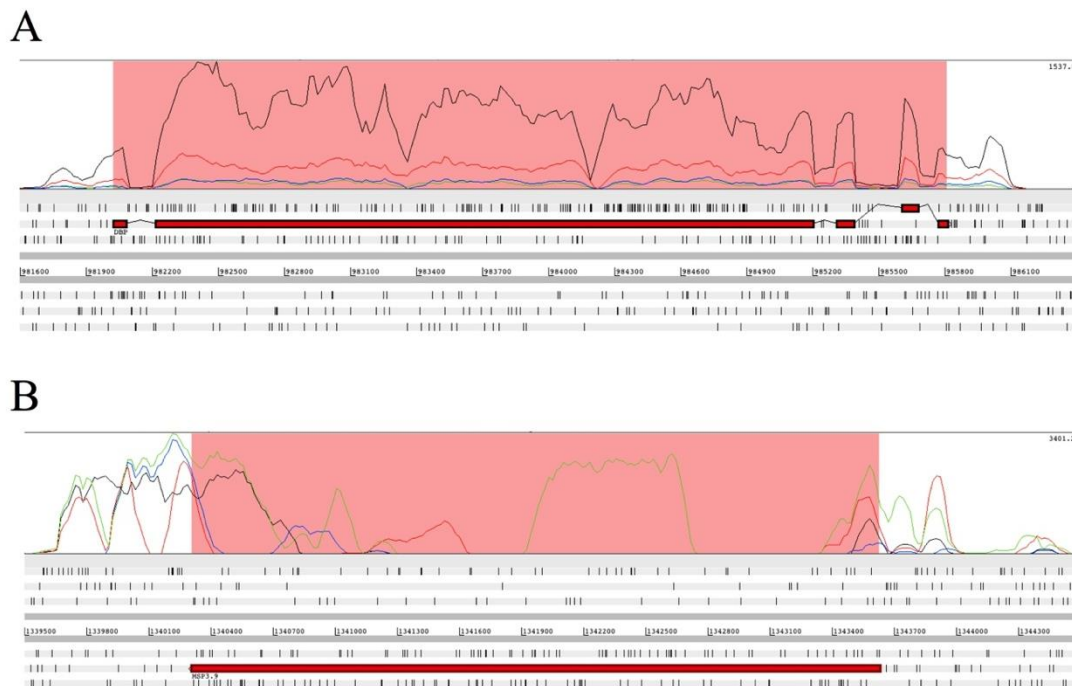


Figure 3.13: Variable expression in *P. vivax* clinical isolates

(A,B) Artemis screenshots showing 2 examples of genes ranked as variably-expressed. RNA-Seq read coverage spans the top half of each panel, with each isolate represented by a different colored line (PV0563=black, PV0565=red, PV0568=blue, PV0417-3=green). The bottom half of each panel shows the 3 forward reading frames (above the chromosomal coordinates) and the 3 reverse reading frames (below the chromosomal coordinates), with stop codons denoted by black vertical bars and gene exons in large red boxes. (A) PVPVP01_0623800, the Duffy Binding Protein (DBP), has mostly even mapping coverage throughout the gene for all 4 isolates, while PV0563 (black) has significantly higher coverage. (B) PVP01_1031200, MSP3.9, has large regions with no coverage in one or more isolates indicating either repetitive regions to which reads could not be uniquely mapped and/or divergence from the *P. vivax* P01 reference.

I reviewed the mapping results for each of the top 130 variably-expressed genes visually using Artemis (Carver et al., 2012, Rutherford et al., 2000) and found that 87% (113/130) contained coverage across the entire gene in all 4 samples. Only 4% (5/130) contained a drop in coverage shared by all 4 samples, and 9% (12/130) had large regions lacking coverage in 1 or more of the isolates, which is indicative of a region of divergence (example shown in Figure 3.13B and noted in Supplementary Table B). Relatively few of the variably-expressed genes can therefore be ascribed to mapping issues, and almost all the mapping issues were from the MSP3 and MSP7 gene families. One RBP showed large gaps in alignment, whereas the other 6 variably-expressed RBPs showed coverage throughout. None of the genes with potential mapping issues had one-to-one orthologs in *P. falciparum*, indicating that our results for GO term enrichment were still valid.

3.3 Discussion

In this chapter, I described the progress toward the goal of studying transcription just prior to *P. vivax* merozoite invasion of erythrocytes with the generation of RNA-Seq data from the schizont stage of 4 Cambodian *P. vivax* clinical isolates. This relied on obtaining a large quantity of high-quality RNA from *P. vivax* samples enriched for schizonts. Starting levels of parasites in all 4 patients were very low at 0.3-0.4%, suggesting that our relatively large blood draws were important for success. The RNA yields after extraction and 2 rounds of DNA digestion were very high, ranging from 10 to 33 μg . At the time the Illumina library protocol required 5 μg of starting RNA, and our high RNA yields meant that multiple attempts could be made from each sample, if needed. The starting material required for library construction has declined 10-fold since we made our libraries, which will make future studies possible with much smaller volumes of patient blood (a limiting factor for many field site protocols). This also raises the possibility of collecting multiple time points from *ex vivo* cultured samples, enabling us to collect transcriptional data throughout the IDC for Cambodian isolates.

RNA-Seq data quantity and quality was similarly high for the 4 isolates. Our schizont-stage sequencing is much deeper than the *P. falciparum* RNA-Seq analogous time point, with over 5 times as many reads mapped in our study compared to the *P. falciparum* RNA-Seq IDC study [Table S2, (Otto et al., 2010)]. The samples contained very little *P. falciparum* or human contamination (under 1% each), validating that these were *P. vivax*

mono-infections and that the leucocyte depletions by CF11 columns were highly effective. It also indicated that our enrichment for schizont-infected erythrocytes by Percoll® largely removed uninfected erythrocytes, which are known to contain high quantities of RNA (Kabanova et al., 2009). The mRNA selection using oligo(dT) magnetic beads was also very effective at removing ribosomal RNA, which accounted for ~1% of the overall data. Thus, our overall laboratory processes were very effective for isolating high quality RNA that was nearly free of contamination and targeted mRNA almost completely. This process will serve as a useful guide for future field studies both within and outside our group.

The average coverage over exons was consistently about 5 times that of introns (209x-271x for exons compared to 42x-52x for introns). The coverage of introns may represent sequencing of incompletely spliced transcripts and/or the presence of contaminating genomic DNA. I investigated this further by comparing the exon and intron coverage for 1% of genes with the highest, middle, and lowest expression. The results indicated that genomic DNA contamination accounted for 15-20x of the overall coverage in each isolate, but that coverage of introns above this threshold was primarily due to incomplete splicing of transcripts. Based on the sequencing depth of each isolate, I set a conservative lower limit of 10 FPKMs for inclusion as an “expressed” gene in downstream analyses. Overall, the similarity in the ratio between exon and intron coverage across all samples suggested that our inter-isolate comparisons were not likely to be affected, even without setting a conservative FPKM threshold.

Comparisons to the published microarray IDC studies for both *P. falciparum* and *P. vivax* confirmed that our samples correlated best to the early schizont stage of the IDC. While the comparison between our samples and the *P. falciparum* RNA-Seq data did not produce a best-matching time point, this may be due to technical issues rather than biological differences. It is noted by the authors that the percentage of uniquely mapping reads was lower at the schizont stage than for any other stage, potentially indicating a loss of transcripts due to high A-T content and low complexity sequence, and no Pearson correlation coefficient comparing the *P. falciparum* microarray and RNA-Seq is reported for the likely best, 40-hour time point (Otto et al., 2010). The *P. falciparum* experiment also used a combination of exonuclease and specific oligos for depletion of rRNAs compared to our oligo(dT) selection of mRNA. It is possible that a bias in sequencing at the schizont stage combined with the differing rRNA depletion methods have reduced the

correlations between the 2 RNA-Seq datasets. In contrast, the comparison with *P. berghei* RNA-Seq for several asexual and sexual stages represented the strongest evidence (with highest Pearson correlation coefficients) that the *P. vivax* clinical isolates represent the schizont stage of the IDC with very little gametocyte contamination.

The RNA-Seq data mapped very similarly to the 2 available reference genomes, *P. vivax* Sal 1 and *P. vivax* P01. On average, 1% more reads mapped to the *P. vivax* Sal 1 genome, which contains about 1.6 fewer Mb of chromosome compared to *P. vivax* P01. One therefore might have expected more data to map to the genome with a longer assembly. This loss in mapping percentage may relate to the stringent mapping parameters I used (only unique placements allowed), as the newest reference is greatly expanded in the repetitive telomeric regions. A more detailed sequence comparison, looking at SNPs between each isolate and each reference, for example, would further define which reference genome is best suited to the Cambodian RNA-Seq dataset. Despite the slightly lower overall mapping, given that the *P. vivax* P01 genome contained over 1000 additionally annotated genes, it was deemed the best reference for subsequent inter-isolate comparisons.

The overall expression of the 4 isolates was highly correlated; pairwise isolate comparisons computed Pearson correlation coefficients ranging from 0.95 to 0.99. This is striking as the samples infected hosts of different ages with potentially different immune statuses, were cultured *ex vivo* for different lengths of time, and underwent a series of laboratory processes from RNA extraction through Illumina library construction, all of which might have affected the resulting comparisons. This overall high similarity in expression values between samples suggests that the lab processing of the samples was very similar and introduced limited technical noise. Critically, the samples were cultured *ex vivo* to the nearly identical stage, early schizonts, before Percoll® enrichment. These high correlations also provided additional support that intron coverage and gametocyte contamination did not greatly skew expression results. It also supports that the transcription during the schizont stage appears to be very similar among different circulating isolates in Cambodia.

I searched for genes with the most dispersed or variable expression using 3 methods and considered in detail those genes that ranked in the top 300 most variably-expressed genes of all 3 methods. Such a cut-off was arbitrary, but was useful to evaluate whether the

genes with the highest variability appeared to be enriched in any way. Assessing GO enrichment using GO terms based on *P. falciparum* one-to-one orthologs, showed a greater than 5-fold enrichment with functions relating to intramolecular oxidoreductase activity, host cell surface binding, protein binding and cellular locations relating to the microneme, pellicle, inner membrane complex, apical complex or apical part of cell, and membrane; almost all of these GO terms relate to invasion in some way. The most variably-expressed genes appeared to be enriched for host interacting genes and locate to the apical end or in invasion-related organelles of merozoites, which might indicate that the isolates were responding to the host environment through the modulation of expression of genes involved in parasite invasion of reticulocytes. Given that several of these variably-expressed genes were members of multi-gene families, such as the RBPs, modulating expression of family members might enable a parasite to evade the immune system and/or improve invasion efficiency. This requires much more study and consideration, however, as it is also likely such genes are overrepresented in general in the schizont stage compared to the entire genome, and may just reflect the genes most likely to be in the schizont transcriptome. Performing a GO enrichment analysis using a curated set of genes combining those expressed throughout the IDC and with peak schizont-stage expression will help to clarify this. One might also hypothesize that the expression variability relates to the efficiency with which each isolate produced merozoites, such that a higher merozoite to schizont ratio might appear as variable expression. Given that the expression for the vast majority of the genes were highly similar, and that there was no pattern of expression ranking (i.e. no single isolate with highest expression throughout the “variably-expressed” list), this seems unlikely.

The vast majority of genes identified as variably-expressed had reads mapping for the entire length of the gene in all 4 isolates, including all of the genes included in the GO enrichment analysis, establishing that differences in expression were not due to mapping artefacts. However, all the variably-expressed members of the MSP3 and MSP7 gene families, as well as a single RBP gene, contained large gaps in alignment for at least 1 underlying isolate indicating either the region was repetitive and reads could not be uniquely mapped or that the reference was highly divergent from the clinical isolates in these areas. *De novo* assemblies will be needed to further investigate expression for these gene families.

3.3.1 Limitations and future work

The biggest limitation of our dataset was having access to only a single life stage of the IDC. In order to understand the full relevance of transcription in our samples, we need a baseline level of transcription for each gene with which to compare the schizont-stage results. This can likely be addressed computationally to some degree requesting the raw intensity values from the *P. vivax* microarray dataset (Bozdech et al., 2008), but would ideally be done through RNA sequencing of a sample with a relatively even mixture of IDC stages at a minimum. The best comparison would be made through completing a study of the complete IDC in Cambodian field isolates. The recent significant reductions in the amount of material needed for Illumina library construction make this highly feasible in future studies.

Shortly before the submission of this dissertation, an RNA-Seq study describing the IDC of *P. vivax* was published (Zhu et al., 2016). The study produced non-stranded RNA-Seq data for 2 Southeast Asian *P. vivax* clinical isolates over 9 time points of the IDC, from the same samples published in the original *P. vivax* IDC microarray study (Bozdech et al., 2008). The study was able to expand on the microarray conclusions to comment on the IDC transcription using the most current *P. vivax* Sal 1 reference annotation as well as non-coding RNAs, long UTRs, and more. These published *P. vivax* RNA-Seq data will be a valuable resource to which the data described in this chapter can be compared. It can be utilized for normalizing the observed expression to best understand which genes are enriched during the schizont stage in naturally-circulating Cambodian clones. As we prepare these data for publication, we will compare and contrast expression between the 2 Thai isolates in the paper and the 4 Cambodian clinical isolates in our dataset. There are also significant features of the Cambodian analysis, which are not captured in the published Thai study. First, we compare mapping and expression data using both *P. vivax* Sal 1 and the new *P. vivax* P01 reference genomes, the latter of which contains over 1000 new annotated genes and was not used in the Thai study. Secondly, our RNA libraries are strand-specific, allowing identification of 5' and 3' UTRs and both sense and antisense transcripts. These findings will therefore add significantly to our understanding of *P. vivax* transcription, even if these data will not now be the first *P. vivax* RNA-Seq data published.

3.4 Conclusion

To characterize gene expression in the schizont stage of *P. vivax* parasites, which contain invasive merozoites and are therefore the most likely source of new blood-stage vaccine targets, we sequenced the transcriptome of 4 *P. vivax* clinical isolates from Pursat Province, Cambodia. We tested 3 RNA extraction methods prior to extracting RNA from merozoite-containing schizonts that were purified using a Percoll® gradient and stored in RNAlater® under field conditions. Strand-specific libraries using only 8 cycles of PCR were generated for Illumina sequencing and mapped to the *P. vivax* Sal 1 and *P. vivax* P01 reference genomes. RNA-Seq data from the clinical isolates correlated most closely with schizont stages from published microarray data from *P. vivax* and *P. falciparum*. This study produced genome-wide unbiased transcript abundance data and enabled the correction of more than 300 gene models. The data showed that expression between the clinical isolates was highly correlated, and the few genes with variable expression were enriched for merozoite surface/invasion genes and or members of large gene families, potentially pointing to differential transcription of genes in response to the host environment. Overall, the production of *P. vivax* schizont stage RNA-Seq data provided a valuable resource to consult for the next stage of my project, the creation of a *P. vivax* merozoite protein library for functional studies.

4 *P. VIVAX* RECOMBINANT PROTEIN LIBRARY

Publication note: portions of the introduction, results (4.2.1, 4.2.2, 4.2.4.2, 4.2.5), and discussion were slightly modified from a published manuscript (Hostetler et al., 2015). I drafted the text in these sections, which was edited by co-authors prior to publication. I am solely responsible for the work described in this chapter under the supervision of my PhD supervisors, Rick Fairhurst and Julian Rayner, except where noted in the text.

4.1 Introduction

A vaccine targeting *P. vivax* will be an essential component of any comprehensive malaria elimination program, but relatively few *P. vivax* antigens have been investigated as vaccine candidates [e.g., (Perera et al., 1998, Malkin et al., 2005, Valderrama-Aguirre et al., 2005, Castellanos et al., 2007, Devi et al., 2007, Wu et al., 2008, Herrera et al., 2011, Mizutani et al., 2014, Yadava et al., 2014)]. This is at least partly due to the difficulty in studying *P. vivax*, which lacks a reliable *in vitro* culture system, and consequently there is limited knowledge about the molecular mechanisms of *P. vivax* merozoite invasion and antimalarial immunity. Reticulocyte invasion represents an essential, completely extracellular part of the asexual blood stage and would make a sensible target for prioritizing new vaccine candidates. The understanding of the interactions required for erythrocyte invasion in *P. vivax* lags significantly behind that of *P. falciparum*. Currently, 5 ligand-receptor interactions are known in *P. falciparum*

(Figure 1.5) including Glycophorin A-EBA175, Glycophorin B-EBL-1, Glycophorin C-EBA140, Complement receptor 1-RH4, and Basigin-RH5 (Camus and Hadley, 1985, Sim, 1995, Mayer et al., 2009, Maier et al., 2002, Crosnier et al., 2011). This is in stark contrast to only 1 known ligand-receptor interaction in *P. vivax*, that between *P. vivax* Duffy Binding Protein (DBP) and the Duffy Antigen Receptor for Chemokines (DARC, also referred to as the Duffy antigen or Fy).

Due to its central role in *P. vivax* invasion, DBP remains a primary target for vaccine development and the only *P. vivax* candidate currently in clinical trials (Table 1.1). DARC binding is mediated by a cysteine-rich region II domain (DBP-II or DBP-RII), which is polymorphic and under immune pressure (VanBuskirk et al., 2004a, Cole-Tobian et al., 2002, Batchelor et al., 2014, Batchelor et al., 2011), leading to strain-specific antibody responses (Ceravolo et al., 2009). Thus, while DBP remains a high-priority target, these complicating factors and previous experience with monovalent *P. falciparum* blood-stage vaccines, which have not thus far provided protection, suggest that a multivalent vaccine that generates cumulative immune responses should be explored. For such a multi-target *P. vivax* vaccine to be generated, a comprehensive and systematic approach is needed to better understand *P. vivax* invasion and identify additional vaccine candidates.

To address this knowledge gap, we initiated a large-scale study of *P. vivax* merozoite proteins that are potentially involved in reticulocyte binding and invasion. Several publications facilitated the assembly of a *P. vivax* merozoite protein library. The primary resource was the 2008 publication of the *P. vivax* Sal 1 reference genome (Carlton et al., 2008), which provided gene model annotation for over 5000 genes and the identification of over 3700 *P. vivax* and *P. falciparum* ortholog pairs. This publication identified 30 proteins with predicted GPI-anchors (Figure 4.1), several of which are known to localize to the merozoite surface and/or invasive secretory organelles in *P. falciparum* orthologs. A subset of proteins with predicted GPI-anchors are part of the 6-cys gene family, members of which are leading vaccine candidates in other stages of the life cycle, for instance, P48/45 and P230 (transmission-blocking) and P36 and P36p (pre-erythrocytic) (Ishino et al., 2005, van Dijk et al., 2005, van Dijk et al., 2010). At least two 6-cys proteins are expressed in each life stage, including P12, P12p, P41, and P38 during the asexual blood stages. *P. vivax* apparently has an additional 6-cys member not found in *P. falciparum* (PVX_001015) (Figure 4.1).

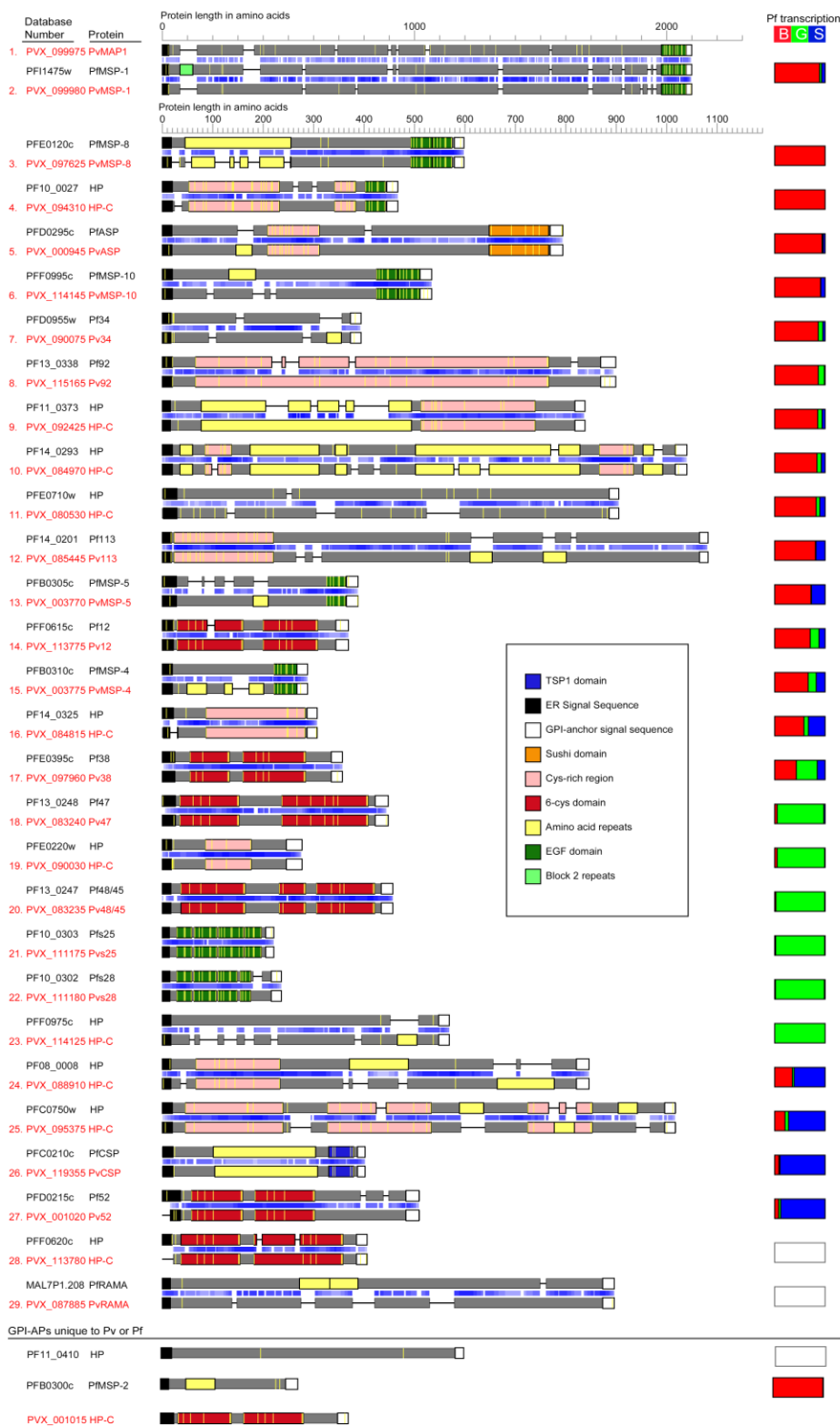


Figure 4.1: GPI-anchored *P. vivax* proteins

Alignments between *P. falciparum* (black gene IDs) and *P. vivax* (red gene IDs) orthologs with predicted GPI anchors, with amino acid similarity proportional to the intensity of blue shading. Depictions include predicted domains (colored boxes), cysteine residues (yellow bars), and microarray expression data for *P. falciparum* (rectangles to the right). HP: hypothetical; HP-C: hypothetical conserved. Reprinted by permission from Macmillan Publishers Ltd: Nature (Carlton et al., 2008) © 2008.

In addition to the reference genome, there are 2 published microarray datasets for *P. vivax* (Bozdech et al., 2008, Westenberger et al., 2010). The Westenberger *et al.* dataset contained average expression data for the asexual blood stages, while the Bozdech *et al.* study spanned 9 time points over the intraerythrocytic development cycle (IDC) of *P. vivax*. These studies, combined with the expression data we generated for 4 schizont-enriched Cambodian clinical isolates (presented in Chapter 3), provide the expression data needed to identify and prioritize candidates with potential involvement in merozoite invasion (i.e., with peak expression in the late blood-stages).

Multiple systems have now been used to express panels of *Plasmodium* proteins (Mehlin et al., 2006, Vedadi et al., 2007, Doolan et al., 2008, Tsuboi et al., 2008, Crompton et al., 2010, Trieu et al., 2011, Chen et al., 2010, Crosnier et al., 2013), and all have their strengths and weaknesses. The wheat germ extract and *E. coli in vitro* protein expression systems are scalable, but lack the context of a complete eukaryotic secretory system including some post-translational modifications, though the lack of *N*-linked glycosylation is helpful in the case of *Plasmodium*, where it is thought to occur rarely or not at all (contrary to most eukaryotes) (Gowda and Davidson, 1999). The mammalian human embryonic kidney, HEK293E, system is more medium-throughput, but is well suited to large proteins, may better represent native post-translational modifications, and has already demonstrated its utility in *Plasmodium* immunoepidemiological and functional studies (Crosnier et al., 2011, Osier et al., 2014, Taechalerpaisarn et al., 2012). This system was applied to uncover the *P. falciparum* RH5 interaction with the erythrocyte receptor basigin, an essential interaction for invasion in all parasite isolates tested (Crosnier et al., 2011), and a library of over 40 *P. falciparum* proteins has now been successfully expressed (Crosnier et al., 2013). Based on these initial successful studies in *P. falciparum* protein expression, we selected the mammalian HEK293E system for large-scale *P. vivax* recombinant protein expression.

The Cell Surface Signalling Laboratory led by Gavin Wright at WTSI was actively developing the methods for expressing *Plasmodium* proteins in the HEK293E expression system with several publications describing protocols for expression and high-throughput screening (Bushell et al., 2008, Crosnier et al., 2013, Kerr and Wright, 2012). Support and training by members of the Wright laboratory were essential to developing the *P. vivax* recombinant protein library. Full-length protein ectodomains, often synthesized by an external company (e.g., GeneArtAG, Germany) were inserted into derivatives of the

pTT3 expression vector (Durocher et al., 2002) including biotinylated ‘bait’ constructs and/or pentamerized ‘prey’ constructs (Figure 4.2). The biotinylation amino acid recognition sequence in bait protein constructs allows the proteins to be enzymatically biotinylated when transiently co-transfected with a plasmid containing the BirA enzyme in expression media containing free biotin. This enables bait proteins to be bound to streptavidin-coated surfaces (i.e., plates or beads) for high-throughput screens without the need for initial purification, which renders large interaction screens more feasible. Both baits and prey constructs contain C-terminal rat Cd4d3+d4 domains, which serve as tags to check for protein expression, as polyclonal antibodies against most *Plasmodium* proteins are not available for such purposes, whereas commercial antibodies to this domain are readily available. The C-terminal cartilage oligomeric matrix protein (COMP) in the β -lactamase enzyme-tagged prey constructs enable ectodomains to be pentamerized, which has been shown to increase the avidity with potential interacting partners in downstream screens (Bushell et al., 2008).

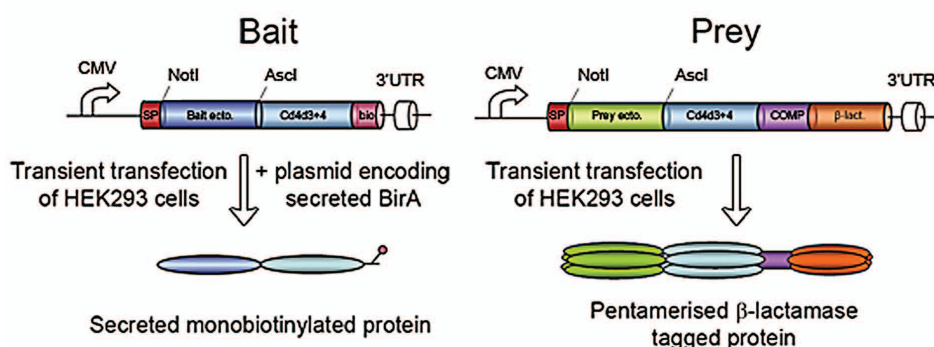


Figure 4.2: Bait and prey protein constructs

Schematic representations of constructs for expression in the mammalian HEK293E cell expression system. Both construct ectodomain (ecto.) regions contain flanking Not1 (5') and Asc1 (3') cut sites and a C-terminal rat Cd4d3+d4 terminal tag (~25 kDa). Bait proteins contain a C-terminal biotinylation recognition sequence that leads to enzymatic mono-biotinylation when transiently co-transfected with free biotin and a BirA enzyme-expressing plasmid. Prey proteins contain a C-terminal pentamerization sequence from the cartilage oligomeric matrix protein (COMP) and the β -lactamase enzyme leading to the pentamerization of prey ectodomains. Figure modified from (Kerr and Wright, 2012).

Once expressed, the *P. vivax* recombinant protein library can be utilized to uncover potential invasion ligands through erythrocyte binding experiments. Such experiments can be performed using several methods. One approach is to employ a cell-based binding assay. Most commonly, this involves transfecting COS7 cells to express membrane-bound protein fragments of interest, followed by incubation with erythrocytes. Binding is detected through counting the clusters (“rosettes”) of erythrocytes surrounding the COS7 cells. This method was used to confirm that region II of *P. vivax* DBP and *P. falciparum* EBA175 was sufficient for binding to erythrocytes (Chitnis and Miller, 1994, Sim et al., 1994). However, this method involves individual experiments per protein and manual counting, equating to significant labor and time to evaluate several dozen proteins in a larger scale library. A flow cytometry-based assay published by Tran et al. in 2005 detected the DBP-RII interaction with erythrocytes using a His-tagged recombinant DBP-RII and a fluorescently-labeled anti-His antibody (Tran et al., 2005). This would enable high-throughput screening in plate formats, but requires His-purified proteins at high concentrations (~50 µg/ml), also a significant limitation for screening a protein library.

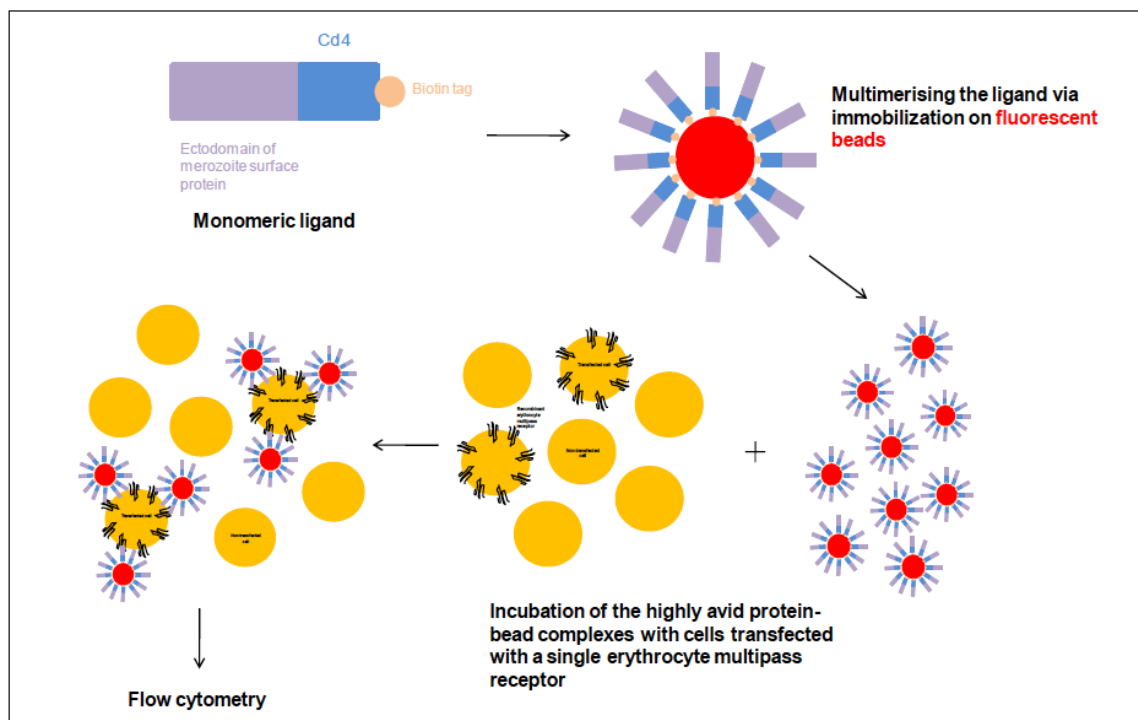


Figure 4.3: High-throughput bead-based interaction screening method

Plate-based interaction screening method developed in Gavin Wright’s laboratory. Mono-biotinylated recombinant proteins are bound to streptavidin-coated beads and incubated with cells. Binding events are detected via a fluorescence shift by flow cytometry. Figure provided by Madushi Wanaguru.

Madushi Wanaguru, in Gavin Wright's laboratory, optimized a flow-cytometry based method for detecting binding based on previous methods in the laboratory (Wright et al., 2000). This method involved binding biotinylated recombinant proteins to streptavidin-coated fluorescent beads and incubating them with cells expressing multi-pass proteins in order to identify specific receptor-ligand interactions (Figure 4.3). The recombinant proteins were multimerized on the beads, potentially increasing the avidity of interactions with erythrocyte receptors, which would prolong the usually transient cell-surface interactions and facilitate their detection. Early testing showed significant fluorescence shifts when incubating erythrocytes with recombinant *P. falciparum* EBA175 and EBA140, which was ultimately published (Crosnier et al., 2013). This high-throughput approach for testing erythrocyte binding would be useful to evaluate a *P. vivax* recombinant protein library in a 96-well format.

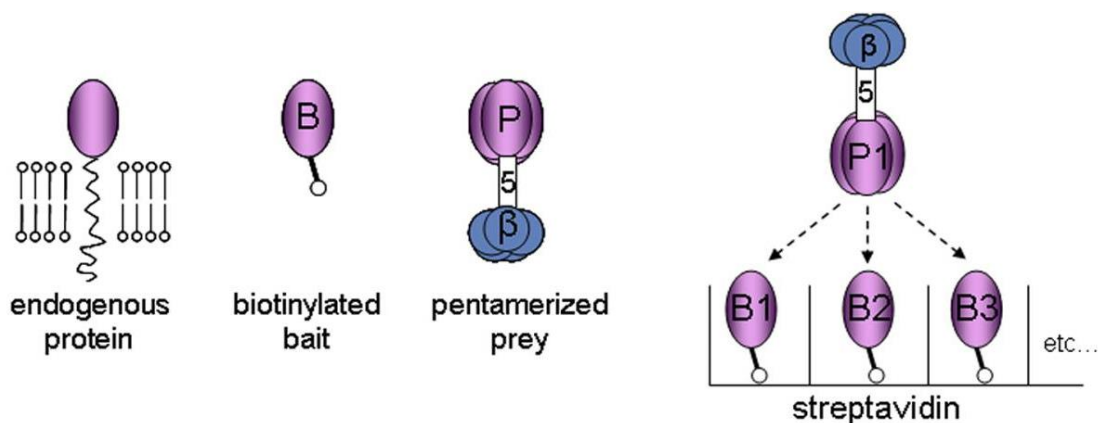


Figure 4.4: Overview of AVEXIS

Depiction of AVEXIS showing biotinylated bait proteins bound to streptavidin-coated plates. Baits are probed by pentamerized prey proteins tagged with β -lactamase. Interactions are detected by nitrocefin turnover resulting in a color change from yellow to red. This figure was originally published in (Martin et al., 2010), © the American Society for Biochemistry and Molecular Biology.

The Wright and Rayner laboratories had also developed a library of over 40 proteins abundant on the surface of mature erythrocytes, which led to the detection of the essential *P. falciparum* RH5 interaction with the erythrocyte receptor basigin (Crosnier et al., 2011). This library was comprised of single-pass erythrocyte surface proteins, expressed in soluble form as full-length ectodomain fragments (Crosnier et al., 2011). The largest surface-exposed ectodomain of several multipass membrane proteins, were also included. We hypothesized that several uncharacterized merozoite-erythrocyte protein interactions

are involved in *P. vivax* recognition and invasion of reticulocytes, and planned to perform large-scale interaction screens to test for specific binding partners to the recombinant *P. vivax* merozoite protein library. Gavin Wright's laboratory had developed a large-scale screening method for detecting specific receptor-ligand interactions called AVidity-based EXtracellular Interaction Screen or AVEXIS (Figure 4.4). The assay reliably detects highly-transient interactions with half-lives ≤ 0.1 seconds (Bushell et al., 2008). This transiency is a common feature of interactions between cell surface proteins (van der Merwe et al., 1994, Dustin and Springer, 1991), and it is plausible that initial *P. vivax* merozoite-reticulocyte interactions will be similarly transient. We plan to use this assay to study *P. vivax* erythrocyte invasion by screening the library of *P. vivax* merozoite surface and invasion antigens against the existing erythrocyte surface protein library. This method can additionally be used to screen for intra-library parasite protein-protein interactions, which may be essential to merozoite function and invasion.

4.1.1 Benefits of these studies

Research into *P. vivax* invasion of reticulocytes lags significantly behind *P. falciparum*, partly due to the lack of an *in vitro* culture system and the difficulty in expressing recombinant *Plasmodium* proteins. Our studies aim to address this need and benefit the scientific community in several ways. First, we aim to assemble a library of *P. vivax* antigens known or predicted to localize to the merozoite surface and/or be involved in erythrocyte invasion. Second, we plan to screen this library in large protein-protein interaction screens and erythrocyte-binding assays to determine protein function and potentially prioritize proteins for further development as vaccine candidates. Lastly, we plan to make this protein library publicly available to facilitate future functional and immunoepidemiological studies.

4.1.2 Objectives

- i. To generate a library of *P. vivax* merozoite proteins and express them in a mammalian expression system
- ii. To investigate protein library function in high-throughput binding and interaction screens

4.2 Results

4.2.1 *P. vivax* merozoite library candidate selection

I incorporated the RNA-Seq data from Chapter 3, existing *P. vivax* microarray data (Westenberger et al., 2010, Bozdech et al., 2008), and homology comparisons to *P. falciparum* to identify candidate merozoite proteins to evaluate in downstream protein interaction screens. This search concentrated on proteins known or predicted to be exported to the merozoite surface, micronemes, or rhoptries (often based on homology with *P. falciparum* proteins) and/or whose transcripts are most abundant during schizogony. A library of 39 *P. vivax* merozoite proteins was identified, and constructs were designed and synthesized for expression in HEK293E cells (GeneArt AG, Germany, see section 2.3.1 for greater detail). These 39 proteins were subdivided into 4 groups Table 4.1.

Table 4.1: *P. vivax* recombinant merozoite proteins

Group	Accession #	Name	Product	Size (kDa)	Pf3D7 homologs /ortholog group (OG)	Subcell. location	GPI anchor (y/n)	Expr level
MSPs ¹	PVX_099980	MSP1	merozoite surface protein 1	215	PF3D7_0930300	Merozoite surface ⁸	y	low
	PVX_097670	MSP3.1 MSP3A ² MSP3 γ	merozoite surface protein 3	114	None ⁶		n	high
	PVX_097680	MSP3.3 MSP3C ² MSP3 β	merozoite surface protein 3	133	None ⁶	Merozoite surface ⁹	n	high
	PVX_097685	MSP3.4 MSP3D1	merozoite surface protein 3	140	OG5_126854		n	high
	PVX_097720	MSP3.10 MSP3H ² MSP3 α	merozoite surface protein 3	113	None ⁶	Merozoite surface ⁹	n	med
	PVX_003775	MSP4 ²	merozoite surface protein 4, putative	45	PF3D7_0207000	Merozoite surface ¹⁰	y	med
	PVX_003770	MSP5 ²	merozoite surface protein 5	62	PF3D7_0206900	Microneme and or/Apical ¹⁰	y	high

	PVX_082700	MSP7.1 ³	merozoite surface protein 7 (MSP7)	70	None		n	low
	PVX_082675	MSP7.6 ³	merozoite surface protein 7 (MSP7)	72	PF3D7_1334300		n	med
	PVX_082655	MSP7.9 ³	merozoite surface protein 7 (MSP7), putative	65	OG5_151950	Merozoite surface ^{7, 11}	n	low
	PVX_114145	MSP10 ²	merozoite surface protein 10, putative	72	PF3D7_0620400	Merozoite surface ^{12, 13}	y	high
6-cys proteins	PVX_113775	P12	6-cysteine protein	61	PF3D7_0612700	Merozoite surface ¹³ , Rhoptry ¹⁴	y	high
	PVX_113780	P12p ²	6-cysteine protein	68	PF3D7_0612800		n	med
	PVX_097960	P38 ²	6-cysteine protein	60	PF3D7_0508000	Merozoite surface and/or Apical ^{7, 15}	y	low
	PVX_000995	P41	6-cysteine protein	67	PF3D7_0404900	Merozoite surface ¹⁶	n	med
	PVX_115165	P92	6-cysteine protein	118	PF3D7_1364100	Merozoite surface ^{7, 15}	y	low
Mero. proteins not in other families	PVX_001725	RON12	rhoptry neck protein 12, putative	54	PF3D7_1017100	Rhoptry ^{7, 17}	n	med
	PVX_088910	GAMA	GPI-anchored micronemal antigen, putative	103	PF3D7_0828800	Merozoite surface ^{7, 18}	y	high
	PVX_090075	Pv34 ²	apical merozoite protein ²	61	PF3D7_0419700	Rhoptry ^{7, 19}	y	high
	PVX_090210	ARP	asparagine-rich protein	54	PF3D7_0423400	Rhoptry ^{7, 20}	n	high
	PVX_090240	CyRPA	cysteine-rich protective antigen, putative	65	PF3D7_0423800	Apical ^{7, 21}	y ²⁵	med
	PVX_095055	RIPR	Rh5 interacting protein,	144	PF3D7_0323400	Microneme and merozoite	n	very low

Identification of novel *Plasmodium vivax* blood-stage vaccine targets

			putative			surface ^{7, 22}		
	PVX_098712	RhopH3	high molecular weight rhoptry protein 3, putative	125	PF3D7_0905400	Rhoptry	n	low
	PVX_110810	DBP	Duffy receptor precursor	135	OG5_148188	Microneme ^{10, 23}	n	low
		DBP-RII	Duffy binding protein region II	64				low
	PVX_111290	MTRAP	merozoite TRAP-like protein, putative	58	PF3D7_1028700	Microneme ^{7, 24}	n	high
	PVX_121885	CLAG ⁴	cytoadherence linked asexual protein, CLAG, putative	159	OG5_138272	Rhoptry ⁷	n	very low
	PVX_081550		StAR-related lipid transfer protein, putative	80	PF3D7_0104200		n	low
	PVX_084815		conserved <i>Plasmodium</i> protein, unknown function ²	53	PF3D7_1434400		y	very low
	PVX_084970		conserved <i>Plasmodium</i> protein, unknown function ²	121	PF3D7_1431400		y	low
	PVX_116775		conserved <i>Plasmodium</i> protein, unknown function ²	60	PF3D7_1321900		n	low
No known <i>Pf</i> 3D7 homolog	PVX_101590	RBP2-like ⁴	reticulocyte-binding protein 2 (RBP2), like	97	None		n	med
	PVX_001015		6-cysteine protein, putative ⁵	62	None		y	low
	PVX_110945		hypothetical protein	61	None		n	low
	PVX_		conserved	66	None		n	low

110950	<i>Plasmodium</i> protein, unknown function ²					
PVX_ 110960	hypothetical protein	92	None		n	low
PVX_ 110965	conserved <i>Plasmodium</i> protein, unknown function ²	69	None		n	low

Proteins and their Plasmodb.org accession numbers are classified into 4 groups, based on whether they are members of the MSP family, the 6-cysteine family, neither family and have a known *P. falciparum* homolog, or neither family and have no known *P. falciparum* homolog. Common names and products are listed based on existing *P. vivax* annotation. The expected size includes a ~25-kDa C-terminal rat Cd4d3+d4 tag. *P. falciparum* accession numbers are listed when only 1 homolog exists and OrthoMCL clusters are listed when multiple homologs exist. Subcellular locations are based on existing *P. vivax* literature or predictions based on existing *P. falciparum* literature, where noted. The presence of a glycosylphosphatidylinositol (GPI) anchor sequence is indicated (Carlton et al., 2008). Expression levels are based on an estimated 0.3-0.5 µg/ml required to saturate biotin binding sites on plates (Osier et al., 2014, Kerr and Wright, 2012). Levels are listed as a guide, as significant batch-to-batch variability was observed. Groups include “high” (>5 µg/ml), “med” for medium (0.5-5 µg/ml), and “low” (<0.5 µg/ml) expressors, as defined in section 2.3.3. Accession links to plasmodb.org, amino acid lengths and boundaries of full-length ectodomains are listed in Supplementary Table D.

¹MSP9 and MSP1P failed to be sub-cloned or expressed, respectively

²Based on genedb.org annotation

³Based on MSP7 family naming in (Kadekoppala and Holder, 2010)

⁴Based on *P. vivax* product description

⁵6-cysteine protein upstream of P52 and P36 in *P. vivax* but not *P. falciparum*

⁶*P. vivax* and *P. falciparum* MSP3s have the same numeric designation, but do not appear to be orthologs (Rice et al., 2014) or fall into the same OrthoMCL clusters

⁷Based on localization in *P. falciparum*

⁸(del Portillo et al., 1988, del Portillo et al., 1991)

⁹(Jiang et al., 2013)

¹⁰(Black et al., 2002)

¹¹(Kadekoppala and Holder, 2010, Kadekoppala et al., 2010)

¹²(Perez-Leal et al., 2005, Giraldo et al., 2009)

¹³(Moreno-Perez et al., 2013)

¹⁴(Li et al., 2012)

¹⁵(Sanders et al., 2005)

¹⁶Cheng et al., 2013

¹⁷(Knuepfer et al., 2014)

¹⁸(Arumugam et al., 2011)

¹⁹(Proellocks et al., 2007)

²⁰(Wickramarachchi et al., 2008)

²¹(Dreyer et al., 2012)

²²(Chen et al., 2011)

²³(Adams et al., 1990, Black et al., 2002)

²⁴(Baum et al., 2006)

²⁵(Reddy et al., 2015)

Merozoite surface proteins. Merozoite surface proteins (MSPs) are primarily exposed on the plasma membrane of the invasive merozoite, and several *P. falciparum* MSPs have been tested as vaccine candidates. We selected a total of 13 MSPs, 8 of which have known subcellular localizations in *P. vivax* (del Portillo et al., 1988, Jiang et al., 2013, Black et al., 2002, Perez-Leal et al., 2005, Moreno-Perez et al., 2013, del Portillo et al., 1991, Giraldo et al., 2009, Oliveira-Ferreira et al., 2004, Cheng et al., 2014) that are similar to those of their *P. falciparum* homologs – except MSP5, which has a micronemal and/or apical distribution in *P. vivax* but a merozoite surface distribution in *P. falciparum* (Black et al., 2002). GPI-anchor sequences were predicted for MSP5 (Carlton et al., 2008) and both predicted and supported by merozoite surface localization experiments in *P. vivax* for MSP1, MSP4, and MSP10 (del Portillo et al., 1988, del Portillo et al., 1991, Black et al., 2002, Perez-Leal et al., 2005, Moreno-Perez et al., 2013, Giraldo et al., 2009, Carlton et al., 2008). GPI anchors were experimentally confirmed in *P. falciparum* homologs for MSP1, MSP4, and MSP5 (Sanders et al., 2005). *P. vivax* has 12 members of the MSP3 family, which are not clear homologues of the *P. falciparum* MSP3 family but may serve similar functions (Rice et al., 2014); 4 representative MSP3s were selected for expression in this library. Similarly 3 MSP7s were selected to represent the 11 MSP7 family members, which Kadekoppala and Holder (Kadekoppala and Holder, 2010) have named sequentially in the order they occur on chromosome 12, beginning with PVX_082700 as MSP7.1 and ending with PVX_082650 as MSP7.11. Despite repeated attempts, we were unable to express the MSP1 paralog, MSP1P, or to sub-clone MSP9 into the expression vector.

6-cysteine proteins. 6-cysteine proteins are defined by a characteristic arrangement of cysteine residues and are expressed at multiple stages of the *P. falciparum* life cycle; several members are considered stage-specific *P. falciparum* vaccine candidates. We selected 5 of 13 *P. vivax* 6-cysteine proteins, 4 of which are most homologous to the *P. falciparum* proteins known to be involved in blood-stage parasite development (Sanders

et al., 2005). Transcriptional data suggest that P12p may also be involved in the blood-stage development of *P. berghei* parasites (van Dijk et al., 2010). GPI anchors were predicted for 3 proteins (P12, P38, and P92) (Carlton et al., 2008), all of which were experimentally confirmed in *P. falciparum* orthologs (Sanders et al., 2005). The predicted GPI anchor in P12 is supported by a merozoite surface and/or rhoptry localization in *P. vivax* (Moreno-Perez et al., 2013, Li et al., 2012).

Merozoite proteins not in other families. Twenty other proteins, which do not belong to the aforementioned *P. vivax* MSP or 6-cysteine protein families, were selected because they are known or suspected to localize to the merozoite surface, micronemes, or rhoptries. Localization to these regions is predicted based on *P. falciparum* homologs (Knuepfer et al., 2014, Arumugam et al., 2011, Proellocks et al., 2007, Wickramarachchi et al., 2008, Dreyer et al., 2012, Chen et al., 2011, Baum et al., 2006) and several have predicted GPI anchors (GAMA, Pv34, PVX_084815, PVX_084970, and PVX_001015) (Carlton et al., 2008). *P. falciparum* CyRPA was also recently determined to have a GPI anchor (Reddy et al., 2015). These 20 proteins fall into 2 groups, with 14 (including PvDBP) having a *P. falciparum* homolog or ortholog, and 6 having neither. The latter 6 proteins include a reticulocyte binding protein (RBP2-like, PVX_101590) that is much smaller than other RBPs, which are typically > 250 kDa and therefore unlikely to express well in the HEK293E cell system; a hypothetical 6-cysteine protein (PVX_001015) that is located upstream from the 6-cysteine proteins P52 and P36 at a microsyntenic breakpoint on *P. vivax* chromosome 3, and has a predicted GPI-anchor (Carlton et al., 2008); and 4 hypothetical proteins (Frech and Chen, 2011) that are syntenic to an MSP-encoding region in *P. falciparum* but have no *P. falciparum* orthologs. Little is known about the hypothetical proteins in our library, except that their transcripts are most abundant in early-to-late schizonts [(Bozdech et al., 2008), Chapter 3 RNA-Seq data] and 1 of them (PVX_110960) contains a merozoite SPAM domain (Pfam accession number, PF07133).

4.2.2 *P. vivax* merozoite protein library expression in HEK293E cells

The expression of 95% (37/39) of the biotinylated full-length protein ectodomains was confirmed by an ELISA that detected the rat Cd4d3+d4 tag (Figure 4.5); 34 of these proteins were also detected by western blot analysis (Figure 4.6). The sizes of expressed proteins ranged from 45 to 215 kDa (which includes the ~25-kDa rat Cd4d3+d4 tag), and generally conform to predicted sizes (Table 4.1).

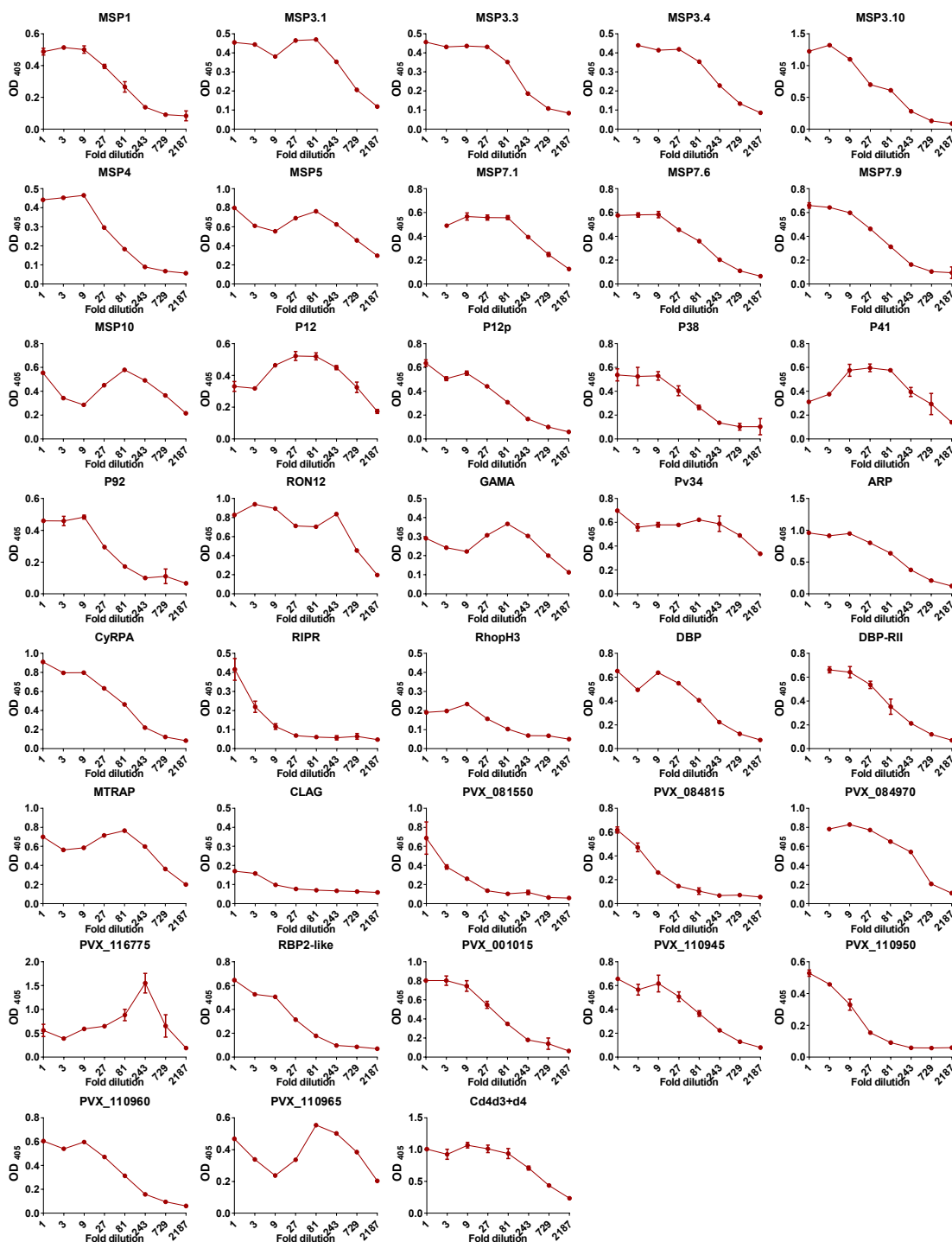


Figure 4.5: *P. vivax* recombinant protein expression detected by ELISA

Three-fold serial dilutions of 37 mono-biotinylated *P. vivax* recombinant proteins were bound to streptavidin-coated plates, and C-terminal Cd4d3+d4 tags were detected using OX68 antibody (primary) and alkaline phosphatase-conjugated anti-mouse antibody (secondary), with alkaline phosphatase activity measured as an increase in optical density (OD) at 405 nm. Mean \pm standard deviation; n=3.

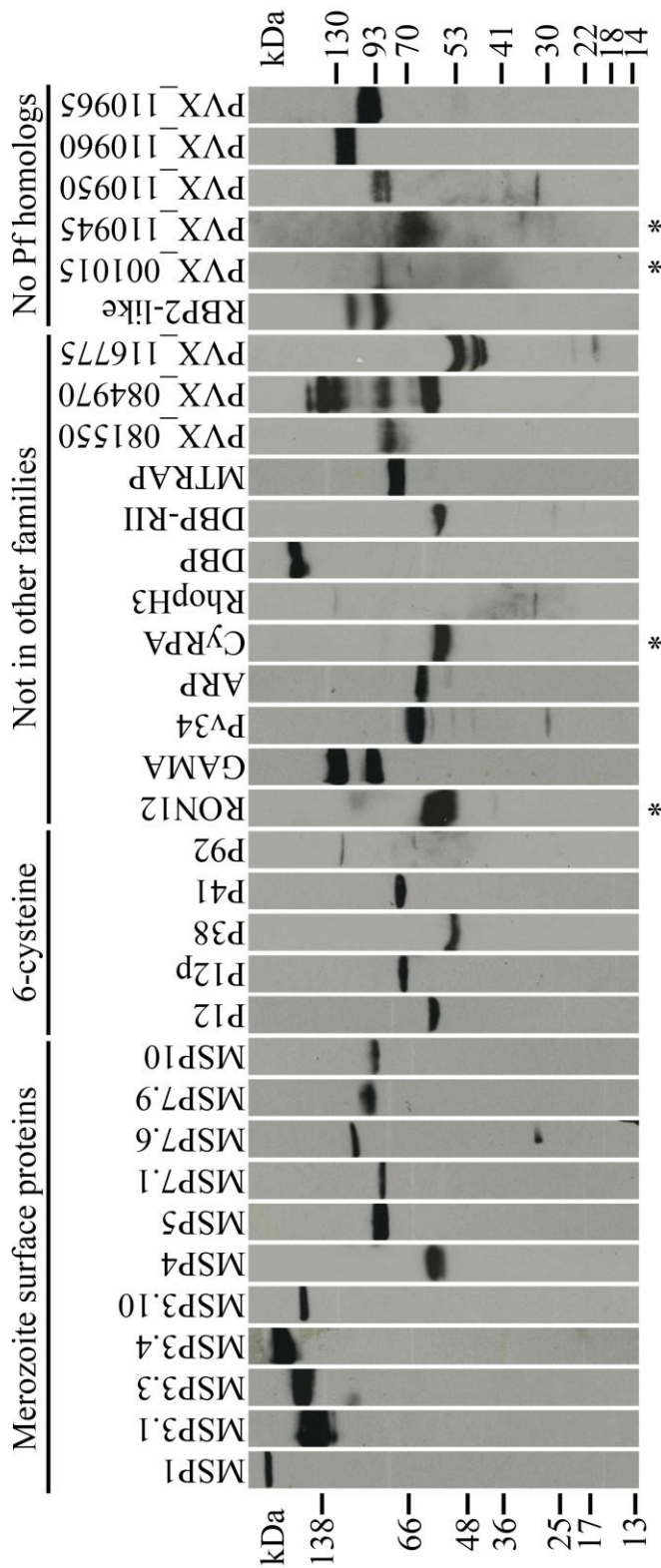


Figure 4.6: Western blot analysis confirms expression of 34/37 *P. vivax* recombinant proteins. Biotinylated proteins were resolved by SDS-PAGE under reducing conditions, blotted, and probed using streptavidin-HRP. All proteins contain a ~25-kDa rat Cd4d3+4 tag. (*) indicates proteins that were run with the right ladder; all others were run with the left or both ladders. Reprinted from (Hostetler, 2015) under the Creative Commons Attribution (CC BY) license.

Several proteins (e.g., GAMA, RBP2-like) showed multiple bands, suggesting they were proteolytically processed, as was seen for some recombinant proteins in an analogous *P. falciparum* library (Crosnier et al., 2013). Protein expression varied widely: 44% (17/39) of the *P. vivax* bait proteins expressed well at over 0.5 µg/ml (see “high” and “medium” in section 2.3.3), and 51% (20/39) of proteins had low expression (<0.5 µg/ml). Most could be effectively concentrated and used in downstream screens. Proteins CLAG, RIPR, and PVX_084815 showed very low expression and were detected by ELISA only.

4.2.3 Erythrocyte binding experiments by flow cytometry

I tested the *P. vivax* recombinant protein library in 96-well format for binding to erythrocytes using a method optimised by Madushi Wanaguru (Wright Laboratory), who provided significant training. The method successfully detected binding of recombinant EBA175 and EBA140 to erythrocytes (Crosnier et al., 2013). The method involved binding recombinant biotinylated *Plasmodium* proteins to streptavidin-coated fluorescent beads (Nile red) and subsequently incubating them with erythrocytes, with binding indicated by a fluorescence shift by flow cytometry. I expressed the complete library as mono-biotinylated proteins in HEK293E cells as described in 2.3.3, and 34/37 proteins showed some expression by dilution series ELISA detection of the Cd4d3+d4 tag (Figure 4.7). Proteins RIPR, RhopH3, and CLAG had no detectable expression while 9 additional proteins had an immediately linear slope in the dilution series, indicating the proteins did not saturate the streptavidin binding sites on plates and were considered to have low protein levels (<0.5 µg/ml after concentrating). The proteins with low expression included MSP1, RON12, PVX_081550, PVX_084815, PVX_084970, PVX_116775, PVX_001015, PVX_110950, and PVX_110960. 34 proteins were multimerized by direct attachment to streptavidin-coated Nile red beads based on the initial ELISA results (excluding CLAG, RIPR, RhopH3, PVX_081550, PVX_110950). I determined the minimum amount of protein needed for saturating the beads by pre-incubating dilutions of protein with uniform quantities of Nile red beads, and assessing the remaining unbound protein by ELISA (Figure 4.7). This was important as any free protein (not bound to beads) in the assay would potentially inhibit binding of the protein-coated beads and reduce the ability to detect erythrocyte binding. Similarly, too little initial protein would result in beads which were not saturated, thereby decreasing the avidity of binding to erythrocytes and again reducing the ability to detect binding.

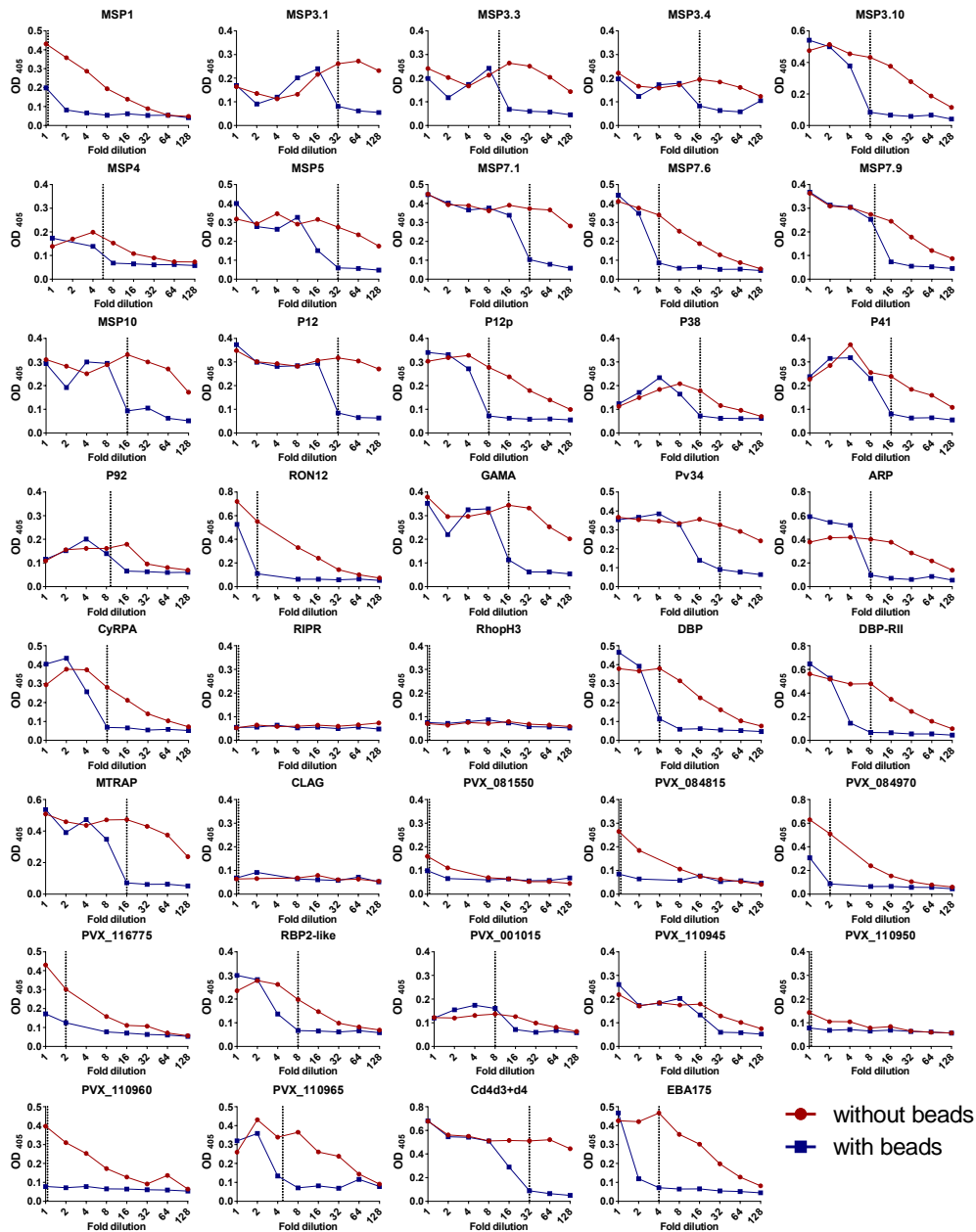


Figure 4.7: *P. vivax* recombinant library expression and bead saturation assay by ELISA

Two-fold serial dilutions of 37 mono-biotinylated *P. vivax* recombinant proteins with (blue line) or without (red line) pre-incubation with streptavidin-coated Nile red beads were bound to streptavidin-coated plates, and C-terminal Cd4d3+d4 tags were detected using OX68 antibody (primary) and alkaline phosphatase-conjugated anti-mouse antibody (secondary), with alkaline phosphatase activity measured as an increase in optical density (OD) at 405 nm. Dotted lines represent the dilution needed for complete saturation of beads with minimal unbound protein. Any unbound protein after pre-incubation with beads, was subsequently bound to the streptavidin-coated plates and resulted in higher OD values, while dilutions where all protein was bound to beads resulted in low OD values (blue lines).

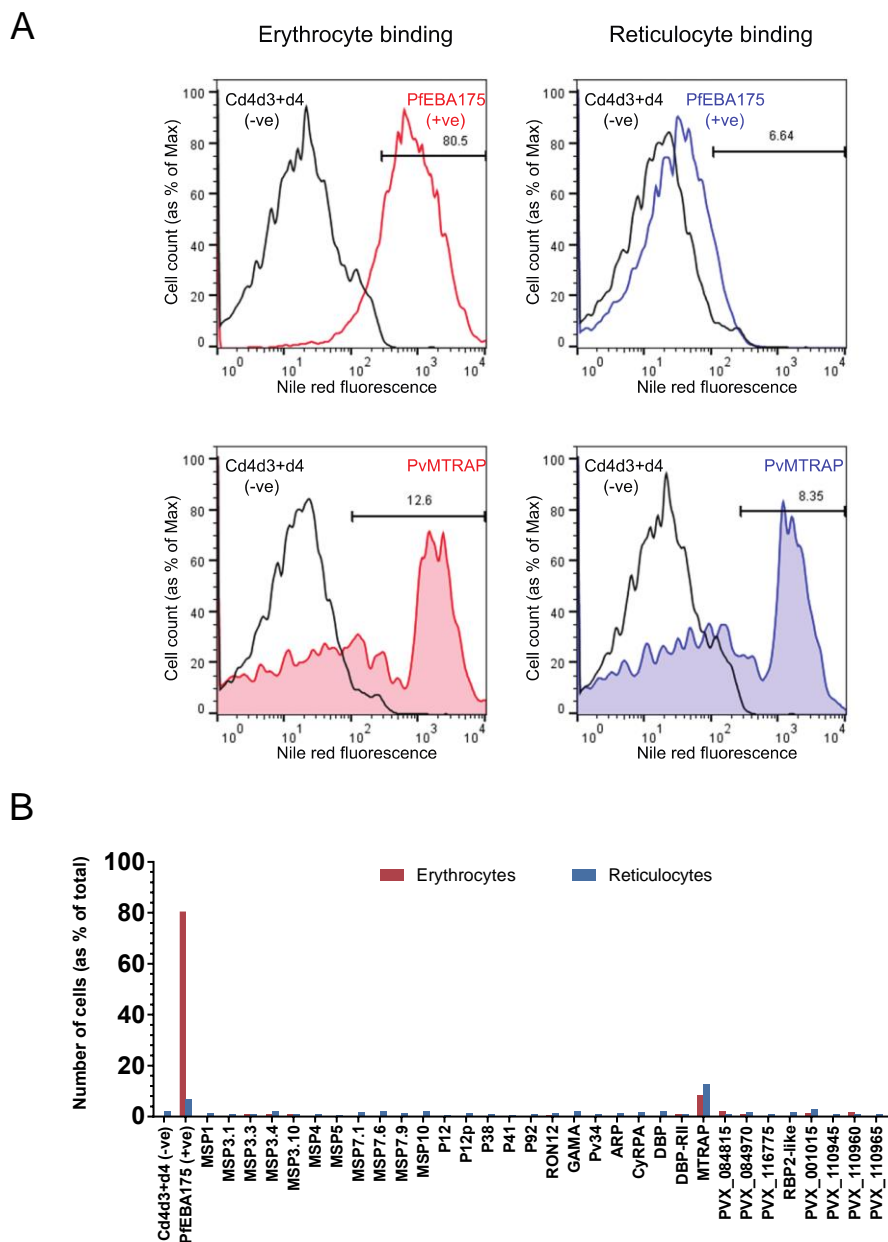


Figure 4.8: Erythrocyte and reticulocyte binding to *P. vivax* recombinant proteins

Recombinant *P. vivax* proteins (34) were multimerized on fluorescent beads (streptavidin-coated, Nile red) and incubated with either erythrocytes (left panel) or hematopoietic stem cell-derived reticulocytes (right panel). (A) Histograms showing fluorescence intensities (at the Nile red emission wavelength) and the number of erythrocytes (as a percentage of total) associated with beads with the gate for binding set based on the negative control, Cd4d3+d4 (-ve) (black). Positive control, *P. falciparum* EBA175 showed 80.5% of erythrocytes (red) and 6.6% of reticulocytes (blue) associated with beads. *P. vivax* MTRAP (lower panel, shaded) displayed some binding with 12.6% of erythrocytes (red) and 8.35% of reticulocytes (blue) associated with beads (B) number of erythrocytes (red) or reticulocytes (blue) (as a percentage of total) associated with beads for 34 *P. vivax* recombinant proteins.

Protein-saturated Nile red beads were then incubated with either erythrocytes or hematopoietic stem cell-derived reticulocytes (provided by NHS Blood and Transplant), and binding was measured using flow cytometry (Figure 4.8). A fluorescence intensity threshold was set based on Cd4d3+d4-coated beads (negative control) in order to calculate the number of erythrocytes (percentage of the total) bound to each protein. *P. falciparum* EBA175-coated beads served as a positive control and 80.5% of erythrocytes and 6.6% of reticulocytes associated with beads. The percentage of erythrocytes or reticulocytes associating with beads was very low (under 2%) for all *P. vivax* recombinant proteins tested except for MTRAP which showed 12.6% of erythrocytes and 8.35% of reticulocytes associated with beads (Figure 4.8B). *P. falciparum* MTRAP has been found to bind with erythrocyte receptor, Semaphorin-7a, so binding may be predicted for the orthologous *P. vivax* MTRAP (Bartholdson et al., 2012). The fluorescence profile however was atypical (i.e. compared to PfEBA175) with a bimodal distribution rather than a unimodal distribution. A repeat of the binding experiment for *P. vivax* MTRAP including re-expression and binding in triplicate failed in the final steps due to plate handling issues and was not able to be repeated within the timeline of this project. Future experiments will be needed to replicate the MTRAP binding results and rule out non-specific binding or bead aggregation. Overall, the assay appeared to distinguish binding in the case of positive control *P. falciparum* EBA175, which binds to glycophorin A, a highly abundant erythrocyte surface protein (10^6 copies/cell), but failed to detect any binding with receptors with low abundance (i.e., no binding was detected in the case of DBP, which bind to lower abundance surface receptor, DARC (10^4 copies/cell) (Anstee, 1990).

4.2.4 High-throughput protein interaction screens

4.2.4.1 AVEXIS between *P. vivax* recombinant proteins and an erythrocyte receptor library

The sensitivity of the bead-based erythrocyte-binding assay was limited in detecting interactions with lower abundance erythrocyte surface receptors. Thus, it was possible that binding between members of the recombinant protein library and erythrocytes would be below the limit of detection of the assay. The Wright and Rayner laboratories had recently developed a library of recombinant erythrocyte prey proteins to facilitate the screening for specific receptor-ligand interactions. This led to the discovery of the

essential *P. falciparum* RH5 interaction with basigin using AVEXIS technology (Crosnier et al., 2011). I performed a similar experiment screening the *P. vivax* recombinant protein library as baits and the erythrocyte recombinant protein library as preys using AVEXIS (Figure 4.4).

AVEXIS consists of testing for interactions between “bait” proteins, which are biotinylated and captured by the streptavidin-coated wells of a 96-well plate, and “prey” proteins, which are enzymatically tagged and contain a pentamerization domain to increase interaction avidity. Thus, I first expressed the recombinant *P. vivax* protein library as mono-biotinylated bait proteins (as described in 2.3.3). Individual protein expression was evaluated by a dilution series ELISA detecting the Cd4d3+d4 tag prior to performing AVEXIS as described in 2.4.1 (Figure 4.5). Nearly 95% (37/39) of the library showed some expression; the remaining 2 proteins either consistently showed no expression (MSP1P) or could not be subcloned in the bait expression construct (MSP9). Seven proteins showed an immediately linear slope in the dilution series, indicating the proteins did not saturate the streptavidin binding sites on plates and were considered to have low protein levels (<0.5 µg/ml after concentrating) (noted with asterisks on Figure 4.10).

Recombinant erythrocyte receptors (37) were expressed (by both Sumana Sharma and myself) as preys similarly to bait proteins as described in section 2.3.3. Prey protein activity was evaluated through the use of a β-lactamase normalization assay described in section 2.4.2. Over 80% (30/37) of preys showed optimal activity, defined as complete turnover of 60 µl at 125 µg/ml nitrocefin by β-lactamase within 20 minutes (Kerr and Wright, 2012, Bushell et al., 2008). The remaining 7 prey proteins (noted with asterisks on Figure 4.10) failed to completely turnover nitrocefin within 20 minutes and were considered to have activity below the threshold of the assay. All 37 bait and 39 prey proteins were included in the AVEXIS screen.

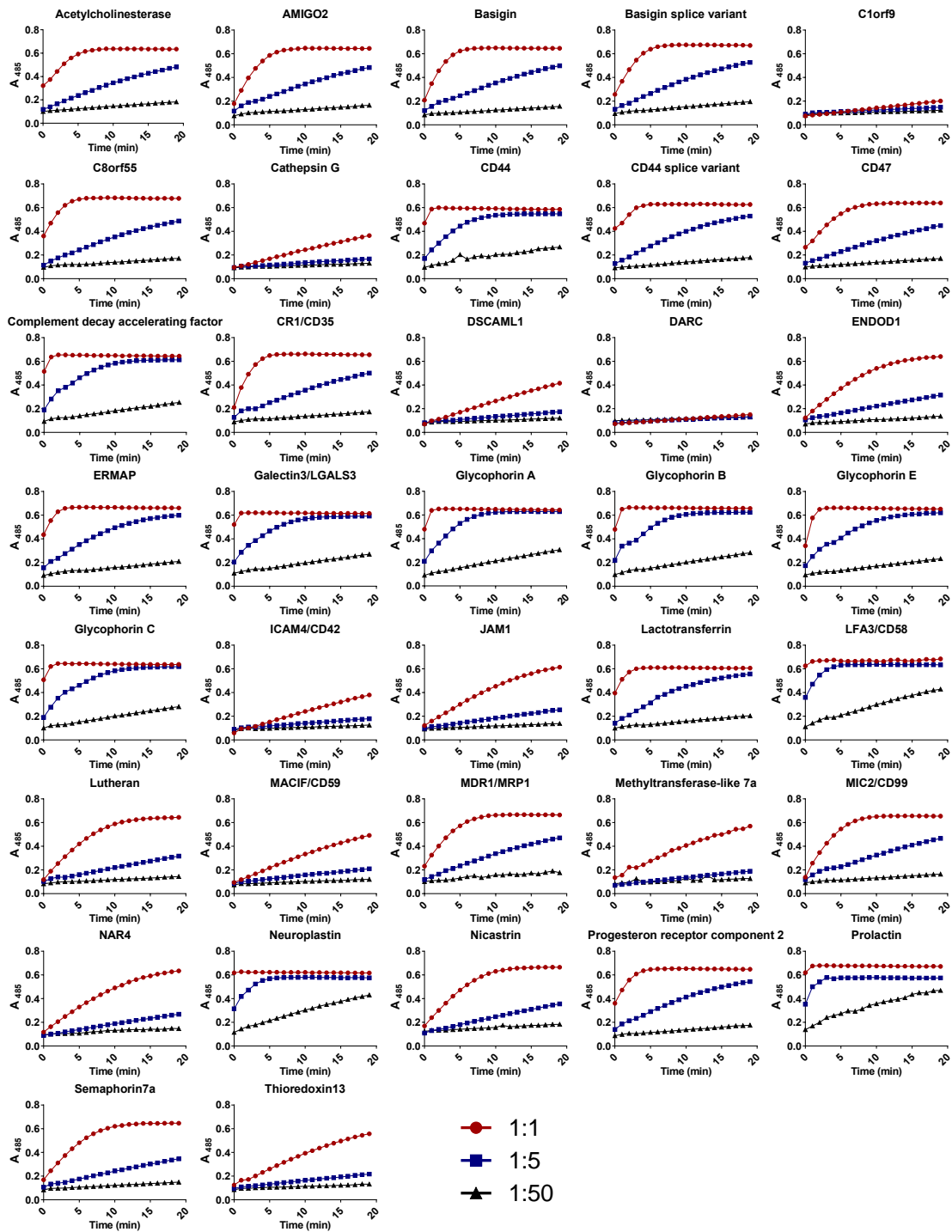


Figure 4.9: Erythrocyte prey normalization assay

Recombinant erythrocyte prey activity was assessed by incubating dilutions of prey proteins (1:1, 1:5, 1:50) with nitrocefin, with β -lactamase turnover of nitrocefin resulting in a color change from yellow to red and increasing absorbance (A) at 485 nm. Optimal activity defined as saturation of signal within 20 minutes.

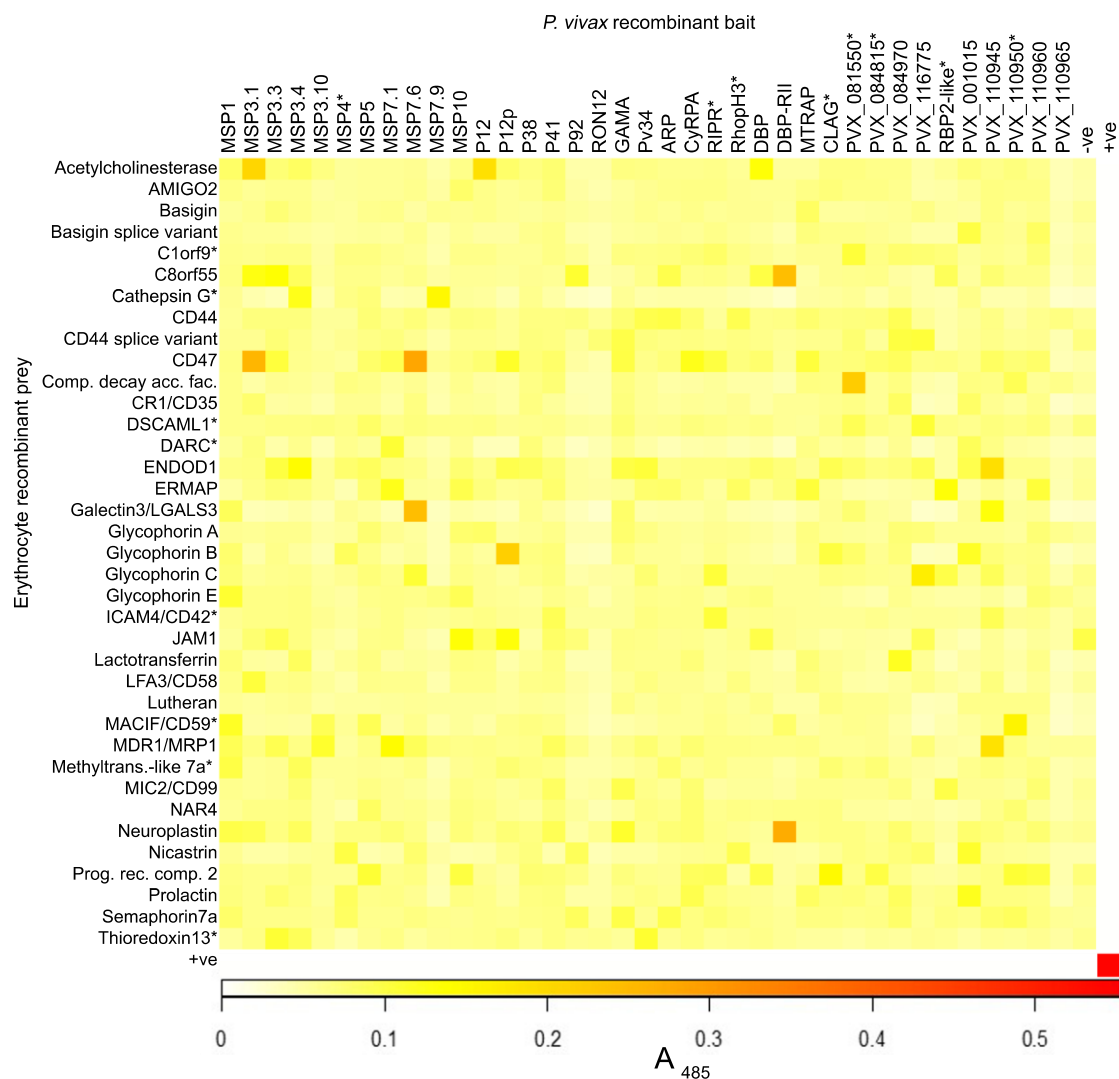


Figure 4.10: AVEXIS between *P. vivax* recombinant proteins and erythrocyte receptor library

Heat map of AVEXIS between *P. vivax* merozoite proteins as baits and erythrocyte receptor proteins as preys, with the intensity of absorbance (A) values at 485 nm after 90 min (values for Neuroplastin-DBP-RII and MSP7.6-CD47 after 18 h). No signal saturations indicating no clear interactions were detected. (*) indicates baits with low protein levels (<0.5 µg/ml after concentrating) and preys with activity below the threshold required by the assay. Positive control (+ve) is the *P. falciparum* P12-P41 interaction. Negative controls (-ve) are rat Cd4d3+d4 tag.

The AVEXIS between the *P. vivax* recombinant protein library and the erythrocyte receptor library detected no clear interactions (Figure 4.10). There were 7 wells with absorbance between 0.22 and 0.29, which was above the general background level, with mean absorbance for the assay 0.07, but well below the absorbance of the positive control (0.57), which is generally observed for validated interactions. These elevated readings

often occurred for the same prey or bait (i.e., CD47 prey and MSP7.6, DBP-RII baits), likely indicating an elevated level of non-specific interactions in these recombinant proteins. A second screen was performed using a complete set of re-expressed proteins, which also failed to detect any interactions (no signals saturated). In order to further characterize the function of proteins in the library in a high-throughput manner, I moved forward to investigating protein-protein interactions between proteins within the *P. vivax* merozoite library.

4.2.4.2 AVEXIS detects predicted and novel interactions between *P. vivax* recombinant proteins

Reflecting the general lack of knowledge about *P. vivax* biology, most of the recombinant proteins in our library have no known function, making it impossible to establish whether they recapitulate the function of native proteins. In some cases, however, protein-protein interactions between members of the library were either known or predicted based on homology to *P. falciparum*. I therefore performed AVEXIS in order to identify predicted and novel *P. vivax* protein-protein interactions. Of the 37 constructs that expressed successfully in the bait vector (Figure 4.5), 34 were successfully sub-cloned and expressed in the prey vector (Figure 4.11), though several had low expression or activity (indicated by asterisks in Figure 4.12). AVEXIS was then performed using all 37 bait and 34 prey *P. vivax* proteins.

This intra-library AVEXIS (Figure 4.12) identified 3 *P. vivax* protein-protein interactions in the bait-prey orientation: P12-P41, P12-PVX_110945, and MSP3.10-MSP7.1. The P12-P41 interaction was predicted, as the *P. falciparum* homologs of these proteins are known to form a heterodimer, an interaction that has been validated by both AVEXIS and immunoprecipitation from parasite interactions (Taechalertrpaisarn et al., 2012). The P12-P41 interaction was also detected in the reciprocal prey-bait orientation (Figure 4.12). While interaction between *P. vivax* P12 and P41 is predicted based on the function of the *P. falciparum* homologues, this is the first time it has been confirmed, implying that the ability of these proteins to interact is essential for their functional activity and predates the evolutionary divergence of *P. falciparum* and *P. vivax*.

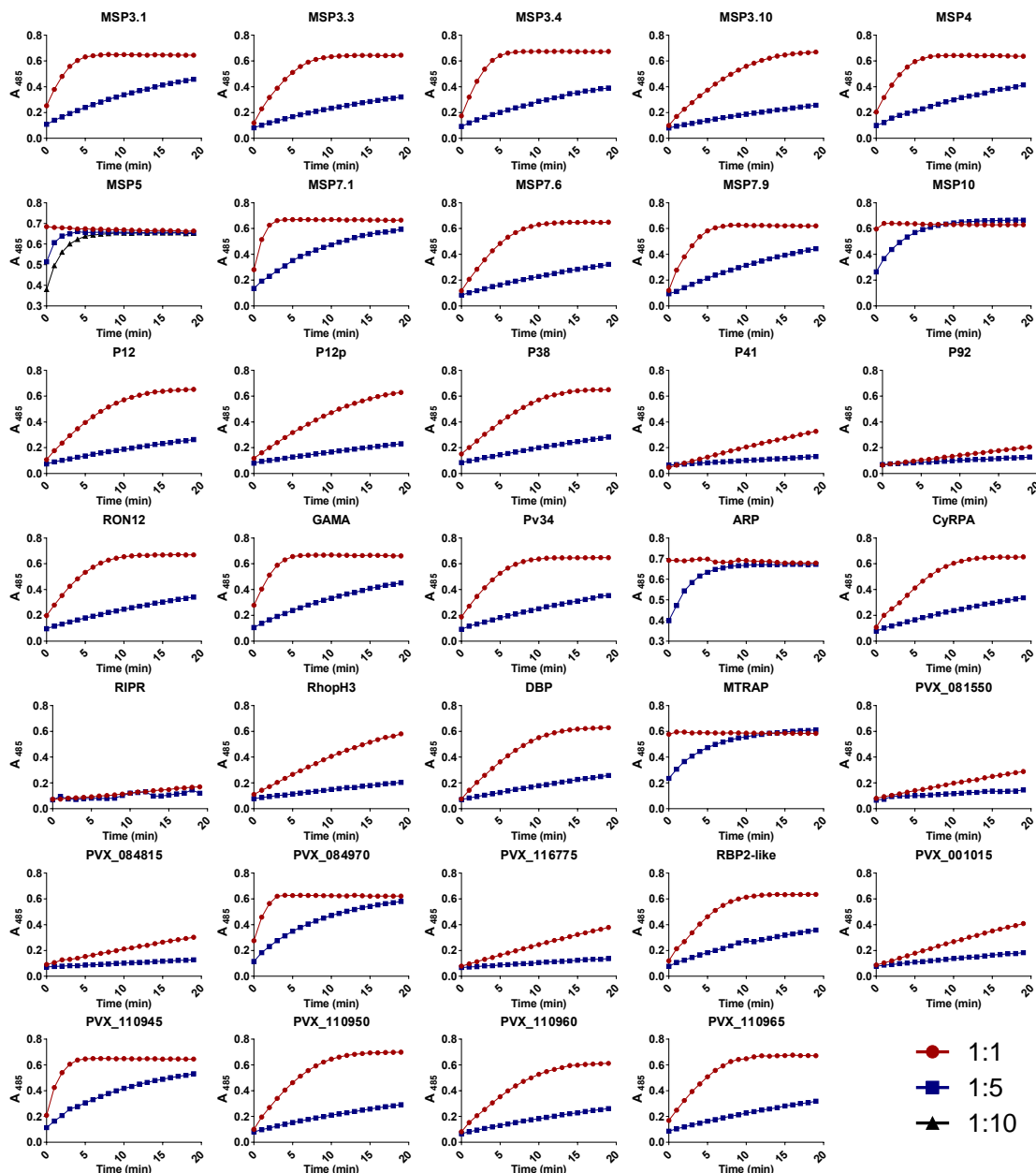


Figure 4.11: *P. vivax* recombinant prey normalization

Recombinant *P. vivax* prey activity was assessed by incubating dilutions of prey proteins (1:1, 1:5, 1:10) with nitrocefin, with β -lactamase turnover of nitrocefin resulting in a color change from yellow to red and increasing absorbance (A) at 485 nm. Optimal activity defined as saturation of signal within 20 min.

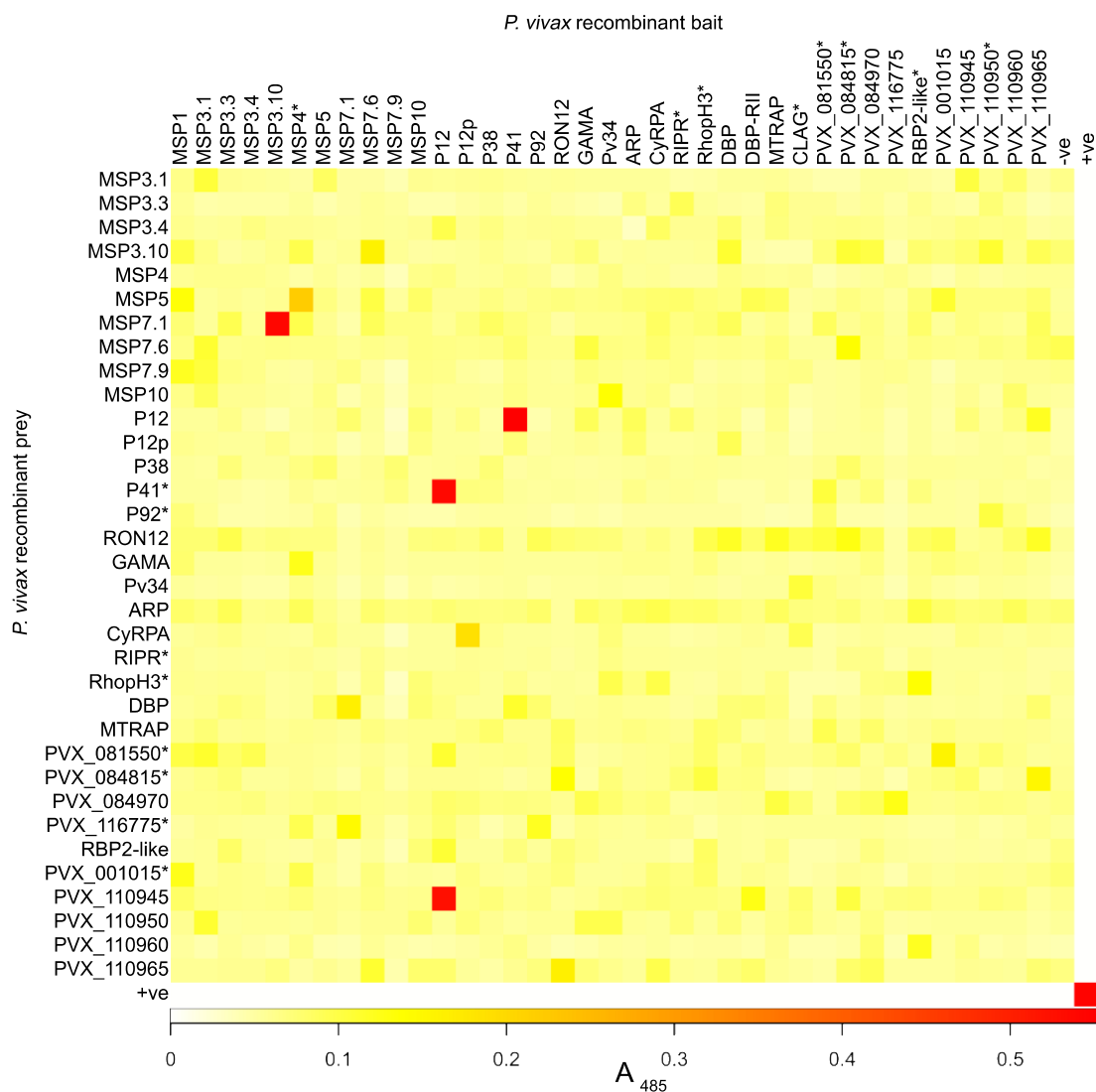


Figure 4.12: AVEXIS reveals novel interactions involving *P. vivax* recombinant proteins

Heat map of the initial *P. vivax* intra-library AVEXIS, with the intensity of absorbance (A) values at 485 nm and positive putative interactions in red: P12-P41 (bait-prey and prey-bait orientations), P12-PVX_110945 (bait-prey orientation), and MSP3.10-MSP7.1 (bait-prey orientation). (*) indicates baits with low protein levels (<0.5 µg/ml after concentrating) and preys with activity below the threshold required by the assay. Positive control (+ve) is the *P. falciparum* P12-P41 interaction. Negative controls (-ve) are rat Cd4d3+d4 tag. Figure reprinted from (Hostetler et al., 2015) under the Creative Commons Attribution (CC BY) license.

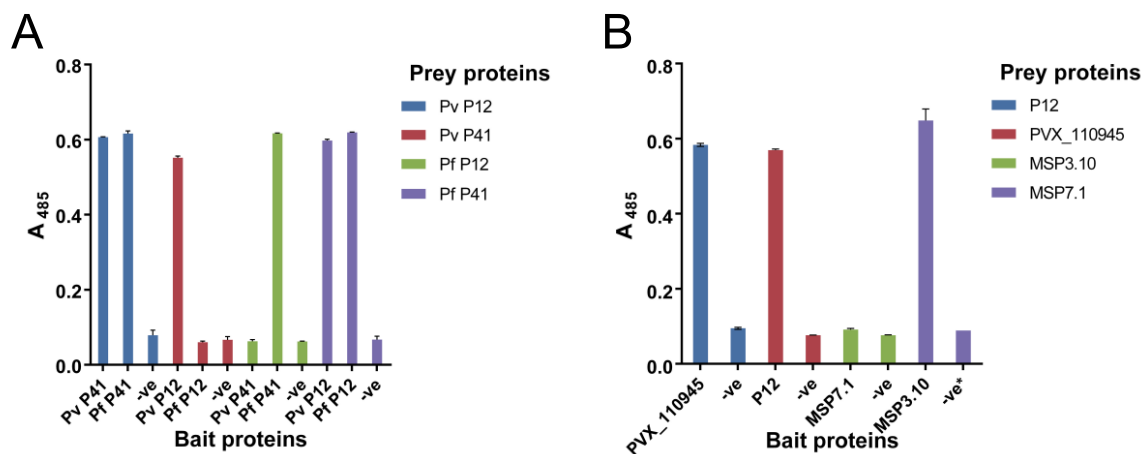


Figure 4.13: Replicated intra-library interactions with AVEXIS

Bar chart shows mean absorbance (A) at 485 nm with range; n=2 (*) indicates n=1. Replicated *P. vivax* intra-library AVEXIS using re-expressed bait proteins. (A) P12-P41 interaction within and between *P. vivax* (Pv) and *P. falciparum* (Pf) proteins by AVEXIS, which detected an interaction between Pv P12-Pv P41 and Pv P12-Pf P41 in both bait and prey orientations. (B) P12-PVX_110945 interaction confirmed in both orientations, and the MSP3.10-MSP7.1 interaction confirmed in only the bait-prey orientation. Positive control (+ve) is the *P. falciparum* P12-P41 interaction. Negative control (-ve) is the rat Cd4d3+d4 tag. Figure reprinted from (Hostetler et al., 2015) under the Creative Commons Attribution (CC BY) license.

To test whether specific amino acid components of the interaction have been conserved over the long evolutionary time frame since the *P. falciparum* and *P. vivax* divergence, I tested the ability of *P. vivax* and *P. falciparum* P12 and P41 to interact with each other. *P. vivax* P12 and *P. falciparum* P41 interacted, suggesting that conserved amino acid contacts do exist (Figure 4.13A), but *P. falciparum* P12 and *P. vivax* P41 did not interact. Investigating the sequences and structures of these proteins in more detail may improve our understanding of their interactions.

While the P12-P41 interaction was predicted, 2 novel interactions, P12-PVX_110945 and MSP3.10-MSP7.1, were also identified during the intra-library screen. Re-expression of PVX_110945 and MSP7.1 as baits instead of preys confirmed the PVX_110945-P12 interaction, but the MSP3.10-MSP7.1 interaction was not detected in the reciprocal prey-bait orientation (Figure 4.13B). The inability to recapitulate the MSP3.10-MSP7.1 interaction does not necessarily indicate that this interaction is biologically insignificant. Protein-protein interactions that are orientation-dependent have been reproducibly detected in other studies (Bushell et al., 2008, Martin et al., 2010, Sollner and Wright,

2009), and in this case may indicate a loss of activity when MSP3.10 is pentamerized in the prey vector.

4.2.5 Biophysical analysis with SPR

4.2.5.1 Size-exclusion chromatographic purification of P12, P41, and MSP7.1

In order to confirm these *P. vivax* protein-protein interactions using surface plasmon resonance (SPR), I subcloned the entire ectodomains of P12, P41, and MSP7.1 into a modified plasmid containing a 6-His tag (Bushell et al., 2008), expressed them (described in 2.3.3), and purified the proteins by immobilized metal-ion affinity chromatography (described in 2.6.1). Purified proteins were subjected to size-exclusion chromatography (SEC) immediately before use, as described in (Taechalertpaisarn et al., 2012) and section 2.6.1. In SEC, recombinant purified *P. vivax* P12 and P41 appeared to elute at apparent molecular masses of 112 and 111 kDa, which are greater than their expected sizes of 60 kDa and 66 kDa, respectively (Figure 4.14A,B). Since the apparent masses are about 2 times the expected sizes, this finding may indicate that these proteins form homodimers in solution, which is supported by gel data obtained under native conditions (Figure 4.15). *P. falciparum* P12 and P41 also eluted at higher-than-expected masses of 90 kDa and 105 kDa, respectively (Taechalertpaisarn et al., 2012). No P12-P12 or P41-P41 self-binding was observed by SPR (Figure 4.17) or AVEXIS (Figure 4.12, Figure 4.13), indicating that the proteins, whether monomers or dimers, are unable to self-associate into higher-order structures.

The recombinant purified *P. vivax* MSP7.1 eluted as a main peak with evidence of higher molecular mass forms by SEC. The main peak eluted at a much larger-than-expected molecular mass of 69 kDa, suggesting oligomerization or aggregation of the protein in solution (Figure 4.14C), which is additionally supported by gel data obtained under native conditions (Figure 4.15). Despite little sequence conservation between MSP7 family members for *P. vivax* and *P. falciparum* (Kadekoppala and Holder, 2010), this is also seen in *P. falciparum* MSP7 purifications, though with additional smaller forms present (Perrin et al., 2015). Other *Plasmodium* surface proteins, such as *P. falciparum* MSP2 and MSP3, are also known to form higher-order structures (Adda et al., 2009, Zhang et al., 2012, Gondeau et al., 2009, Imam et al., 2014).

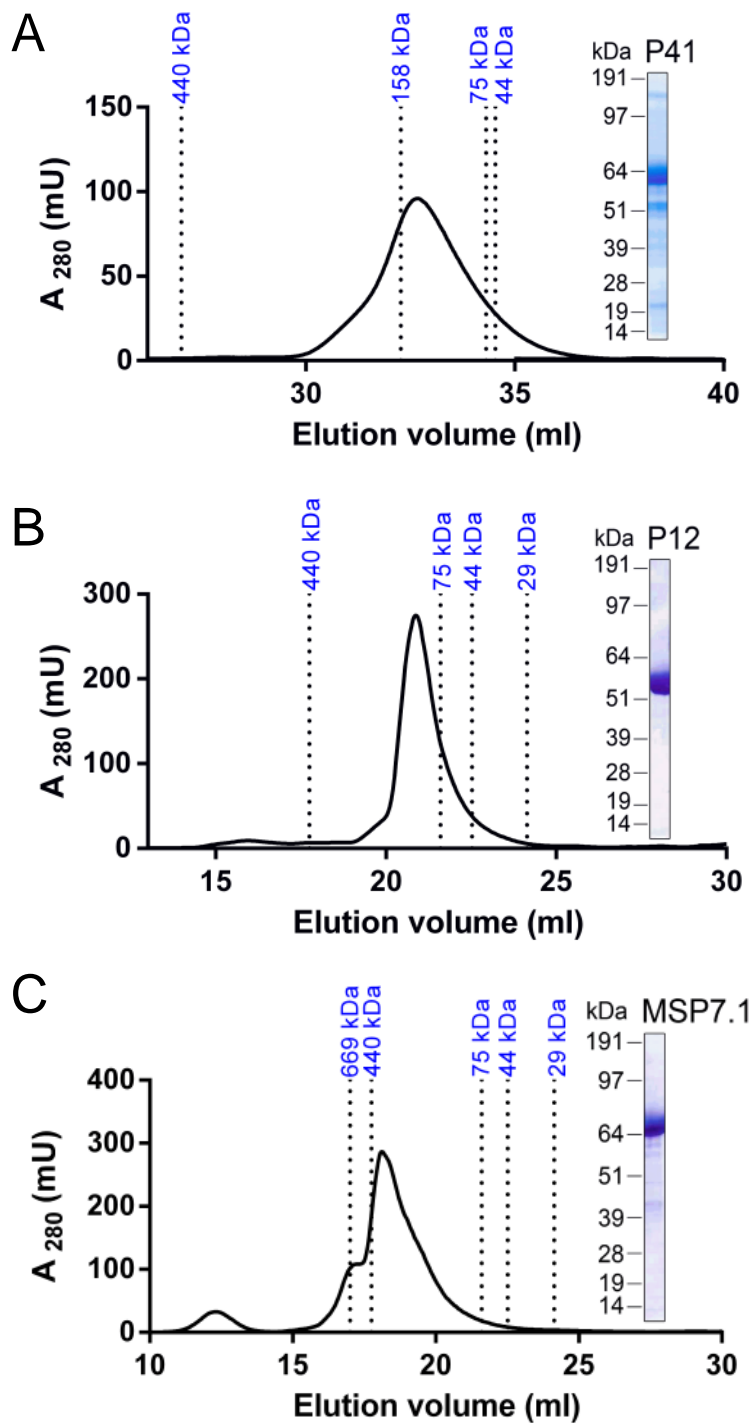


Figure 4.14: SEC for *P. vivax* P12 and P41 and MSP7.1

Recombinant 6-His-tagged *P. vivax* P12 (A) and P41 (B), each eluted as a monodisperse peak after SEC, which was resolved as a single band of the expected size by SDS-PAGE (insets). Recombinant 6-His-tagged *P. vivax* MSP7.1 eluted as a main peak with a small shoulder at higher-than-expected masses after SEC, likely due to oligomerization, and with a main band of the expected size by SDS-PAGE (inset). Figure adapted from (Hostetler et al., 2015) under the Creative Commons Attribution (CC BY) license.

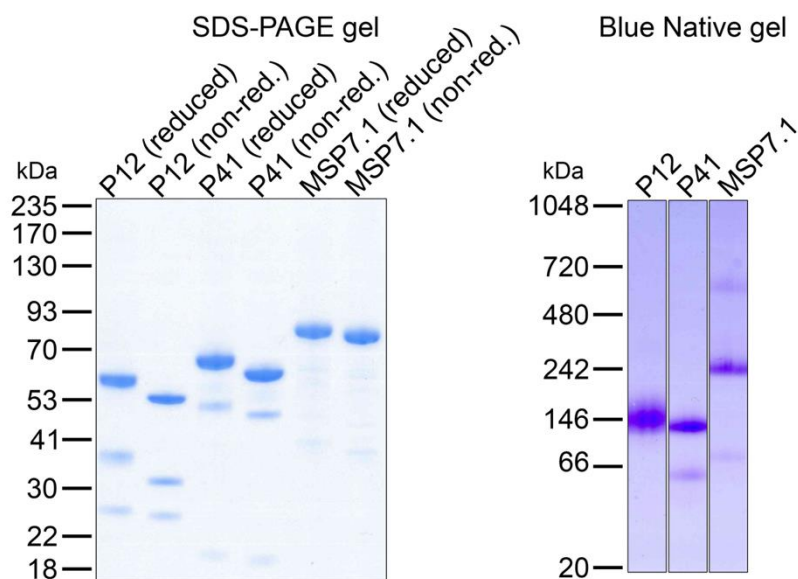


Figure 4.15: P12, P41, and MSP7.1 may exist as homodimers or oligomers

Recombinant 6-His-tagged *P. vivax* P12, P41, and MSP7.1 were fractionated by SDS-PAGE under reducing conditions and non-reducing conditions, and stained. An increase in molecular mass upon reduction suggested that the proteins contain disulphide bonds, as expected. No homodimers or oligomers are observed by SDS-PAGE. Samples run under native conditions support the presence of homodimers for P12 and P41 with bands near 146 kDa. Monomers for these 2 proteins correspond to 60 and 66 kDa, respectively. MSP7.1 shows a faint band near the expected monomer size of 69 kDa and a dominant band near 242 kDa, suggesting that this protein potentially forms an oligomer of 3 or 4 monomers. Figure reprinted from (Hostetler et al., 2015) under the Creative Commons Attribution (CC BY) license.

4.2.5.2 SPR confirms *P. vivax* P12-P41 interaction

SPR was used to validate all interactions and determine the biophysical binding parameters for the P12-P41 interaction. Equilibrium binding experiments between *P. vivax* P12-P41 showed clear evidence of saturation, thus demonstrating the interaction's specificity (Figure 4.16). However, due to limited amounts of protein, several of the lower concentrations of analyte did not achieve equilibrium in both bait-prey orientations with the consequence that the calculated equilibrium dissociation constants (K_D) are likely to be slightly overestimated; that is, the interaction has a higher affinity. When *P. vivax* P12 was used as the purified 6-His-tagged analyte and *P. vivax* P41 as the immobilized biotinylated ligand, the K_D was 120 ± 10 nM, and when the proteins were used in the reverse orientation the K_D was 77 ± 6 nM. The *P. vivax* P12-P41 interaction

therefore has an affinity that is at least 3 times higher than the *P. falciparum* P12-P41 interaction, with the *P. vivax* P12-P41 K_D of <100 nM being much lower than the *P. falciparum* P12-P41 K_D of 310 nM (Taechalertrpaisarn et al., 2012). Binding between *P. falciparum* and *P. vivax* homologues that had been detected using AVEXIS was confirmed by SPR using *P. vivax* P12 as the purified 6-His-tagged analyte and *P. falciparum* P41 as the immobilized biotinylated ligand (Figure 4.16C). The K_D of 31 ± 10 nM for this interaction also suggests that it has a much higher affinity than the *P. falciparum* P12-P41 interaction.

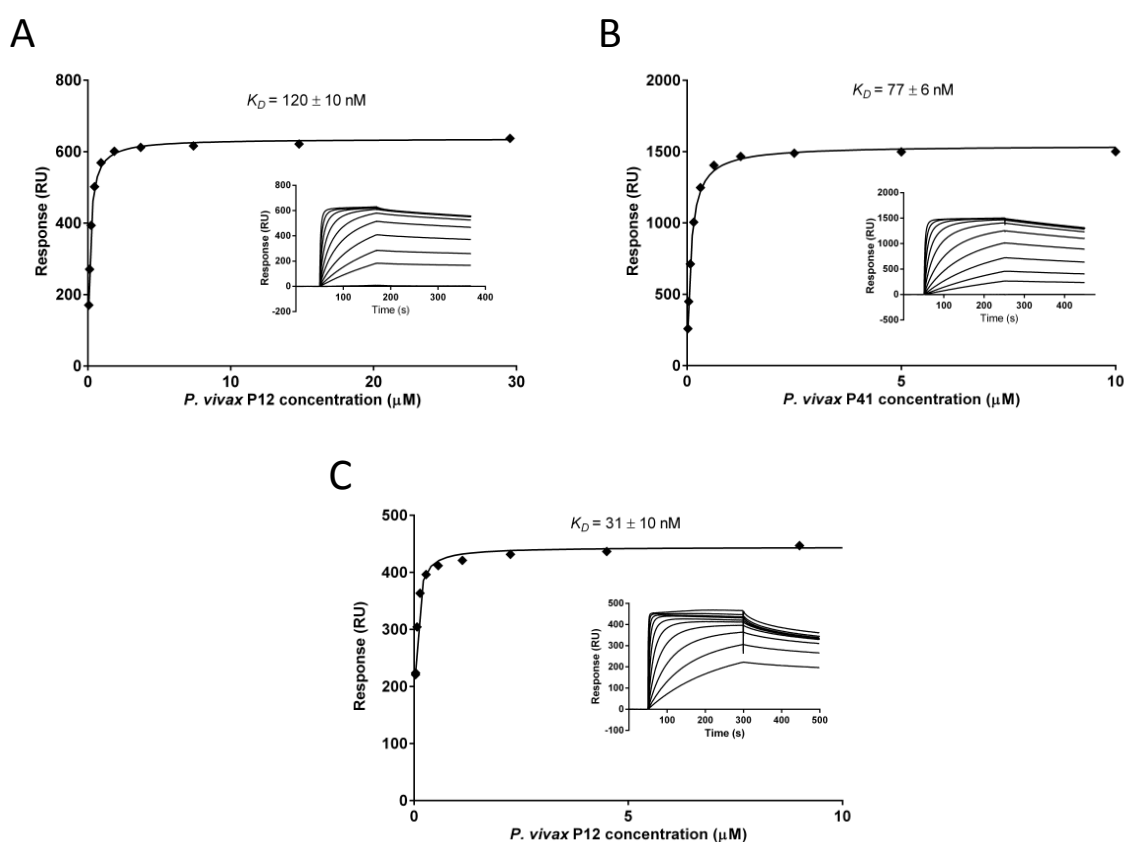


Figure 4.16: Quantification of the *P. vivax* P12-P41 interaction affinity by surface plasmon resonance

Increasing concentrations of analyte protein were injected over immobilized biotinylated ligand protein. Reference-subtracted binding data were plotted as a binding curve and the equilibrium dissociation constant was calculated using $R_{eq} = CR_{max}/(C + K_D)$. Experiments included analyte-ligand combinations *P. vivax* P12-P41 (A), *P. vivax* P41-P12 (B), and *P. vivax* P12-*P. falciparum* P41 (C). Lower concentrations failed to reach equilibrium, which resulted in an overestimated K_D . SEC, size-exclusion chromatography. Figure adapted from (Hostetler et al., 2015) under the Creative Commons Attribution (CC BY) license.

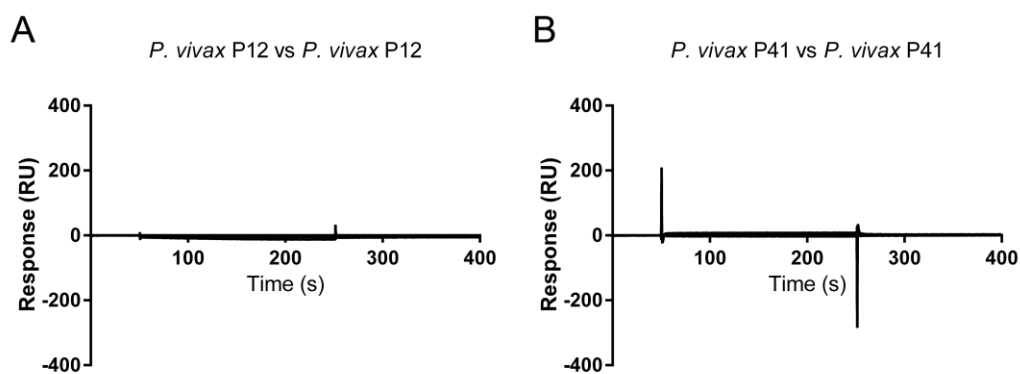


Figure 4.17: *P. vivax* P12 and *P. vivax* P41 show no self-binding by surface plasmon resonance

Increasing concentrations of *P. vivax* P12 (A) or P41 (B) were injected over immobilized biotinylated *P. vivax* P12 (A) or P41 (B), with no interactions observed. Figure reprinted from (Hostetler et al., 2015) under the Creative Commons Attribution (CC BY) license.

4.2.5.3 SPR for novel interactions

SPR was also used to study the *P. vivax* MSP3.10-MSP7.1 interaction. Equilibrium binding experiments using *P. vivax* MSP3.10 as ligand and *P. vivax* MSP7.1 as analyte showed a relatively high binding affinity (Figure 4.18), but a K_D could not be calculated since the binding did not reach equilibrium at any of the concentrations tested. In addition, the binding did not fit a 1:1 model (Figure 4.18, shown in red), most likely due to oligomerization of the purified *P. vivax* MSP7.1. Such complex binding behavior was also observed between P-selectin and *P. falciparum* MSP7, which also oligomerize (Perrin et al., 2015).

SPR was also used to explore the P12-PVX_110945 interaction identified by AVEXIS. This confirmed an extremely weak interaction between recombinant purified *P. vivax* P12 as analyte and biotinylated PVX_110945 as ligand (Figure 4.19). None of the concentrations used reached equilibrium, which prevented the calculation of an equilibrium dissociation constant (K_D). Further functional studies will be needed to determine whether the interaction is biologically relevant.

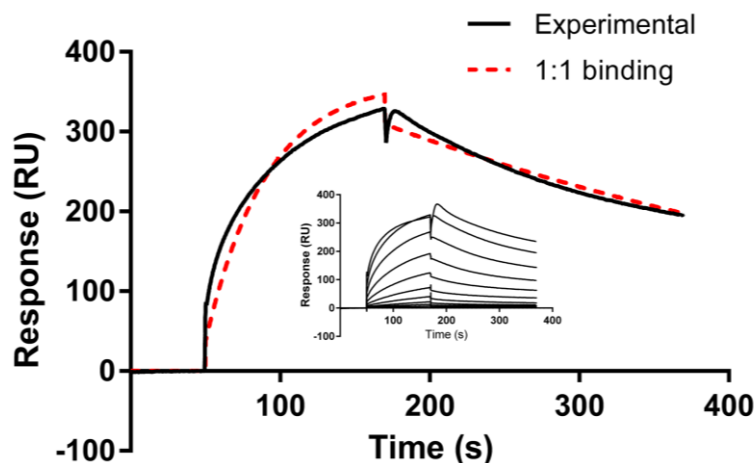


Figure 4.18: Surface plasmon resonance confirms the *P. vivax* MSP3.10-MSP7.1 interaction

Increasing concentrations of *P. vivax* MSP7.1 were injected over immobilized biotinylated *P. vivax* MSP3.10. Relatively high-affinity binding was observed, although none of the concentrations used reached equilibrium (inset). The binding did not fit a 1:1 model (red dashed line). The increase in response units at the start of the dissociation phase at the higher analyte concentrations of *P. vivax* MSP7.1 is likely an artefactual buffer effect. SEC, size-exclusion chromatography. Figure reprinted from (Hostetler et al., 2015) under the Creative Commons Attribution (CC BY) license.

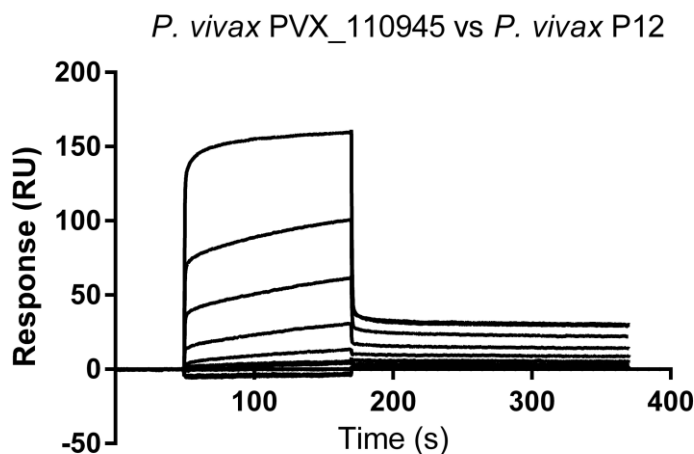


Figure 4.19: Surface plasmon resonance supports a weak interaction between *P. vivax* P12 and *P. vivax* PVX_110945

Increasing concentrations of *P. vivax* P12 were injected over immobilized biotinylated *P. vivax* PVX_110945 with weak binding observed. None of the concentrations used reached equilibrium, which prevented the calculation of an equilibrium dissociation constant (K_D). Figure reprinted from (Hostetler et al., 2015) under the Creative Commons Attribution (CC BY) license.

4.3 Discussion

In this Chapter, I presented the production and characterization of a recombinant *P. vivax* merozoite protein library. Research into *P. vivax* biology generally lags behind *P. falciparum*, in large part due to the lack of a robust *in vitro* culture system for *P. vivax*. This complicates investigation of *P. vivax* erythrocyte invasion and thus identification of new *P. vivax* blood-stage vaccine candidates. Multiple studies have investigated single or few *P. vivax* merozoite proteins, but a comprehensive library of full-length ectodomains of merozoite surface, microneme, and rhoptry proteins for functional studies has not yet been assembled. To build such a library, we selected 39 proteins that are known or predicted to localize to the surface or apical organelles of merozoites based on published *P. vivax* and *P. falciparum* studies, and published microarray data showing gene upregulation during *P. vivax* schizogony. I was able to express 37 of these proteins as recombinant ectodomains, including several members of the MSP and 6-cysteine protein families, additional invasion-related and/or GPI-anchored proteins, as well as several proteins with no *P. falciparum* homologs. This success in expressing 95% of *P. vivax* merozoite proteins at levels useable for biochemical studies is comparable to similar efforts at expressing *P. falciparum* merozoite proteins (Crosnier et al., 2013), and demonstrates the broad utility of the human HEK293E cell expression system in producing high-yield, high-quality proteins for *Plasmodium* research.

Protein expression levels were evaluated for patterns that predicted success. From previous experience using the HEK293E system, proteins larger than 250 kDa are frequently not expressed or expressed at low levels. However, size alone was not a useful predictor of expression levels in this library, as I sometimes observed high and low expression levels for large and small proteins, respectively [e.g., high expression for MSP3.4 (140 kDa) and low expression for PVX_084815 (53 kDa)]. The effects of amino acid composition and predicted protein folding on expression are not yet clear, but useful predictors of expression levels may become apparent as we expand our *Plasmodium* plasmid library.

In order to maximize our chance for successful expression, I used codon-optimized, protein ectodomains with mutated *N*-linked glycosylation sites. In this work I did not seek to systematically study the factors that aid *Plasmodium* expression, seeking instead a “one size fits all” approach to maximize the number of proteins expressed, but minimize the

time and resources invested in optimizing expression. However, prior work in the Wright laboratory suggests that codon optimization is necessary for successful expression of some but not all *Plasmodium* proteins in mammalian expression systems. Previous studies have often used native sequences [reviewed in (Birkholtz et al., 2008)], even though other studies have noted that optimization of codon usage can aid expression (Kocken et al., 2002, Russell et al., 2005). As another example, the Wright and Rayner laboratories previously found that codon optimization alone had little effect on the expression of PfRH5, while the inclusion of an exogenous signal peptide and the mutation of *N*-linked glycosylation sites significantly increased its expression (Crosnier et al., 2013). In addition, the AT content of the *P. vivax* genome (50-60%) is much closer to that of humans (59%) than to that of *P. falciparum* (80%). This significantly impacts codon usage between the 2 *Plasmodium* species [as noted in (Chen and Cheng, 1999)], and potentially means that *P. vivax* proteins would be more easily expressed without codon optimization than *P. falciparum* proteins. If expression constructs are to be generated synthetically, as they were in this study, then codon optimization appears to be a useful step, but not a panacea, in the expression of *P. vivax* proteins.

A flow cytometry-based erythrocyte-binding assay was used to investigate *P. vivax* protein library binding to both erythrocytes and hematopoietic stem cell-derived reticulocytes. The assay allowed for the complete library to be screened in 96-well format using small quantities of protein culture supernatant (i.e., 100 μ l per binding replicate), which is a significant improvement over assays requiring His-purification or manual counting of rosettes. The results demonstrated significant shifts in fluorescence indicative of binding for *P. falciparum* EBA175. *P. vivax* MTRAP also appeared to show binding with less dramatic shifts in fluorescence. However, the atypical distribution may indicate bead aggregation and follow-up experiments are needed to fully explore this potential interaction. Overall, the assay appeared to distinguish binding in the case of positive control *P. falciparum* EBA175, which binds to glycophorin A, a highly abundant erythrocyte surface protein (10^6 copies/cell), but failed to detect any novel interactions for proteins in the *P. vivax* protein library. This could reflect biological reality (i.e., none of these proteins bind erythrocytes), but it could also suggest methodological issues, specifically that the binding detection method has relatively low sensitivity. Supporting this theory is the fact that no binding was detected in the case of DBP, which binds to a relatively low abundance surface receptor, DARC (10^4 copies/cell). More sensitive

binding assays need to be developed to systematically interrogate this and other *Plasmodium* merozoite protein libraries.

The AVEXIS interaction screen between the *P. vivax* recombinant protein library and specific erythrocyte receptors also failed to detect any clear interactions, including the expected interaction between DBP and the N-terminal sequence of DARC. Multipass membrane proteins, such as DARC, are difficult to express as properly-folded amphiphilic fragments, and the recombinant prey protein included in the library represented the largest surface-exposed ectodomain, which included the expected binding region with DBP-RII. However, this tail alone may not be enough to provide binding in the context of an AVEXIS assay, which has not been applied to this interaction before. Again, more sensitive methods may need to be developed to use the library to identify new interactions.

The intra-library AVEXIS assay detected 3 interactions: P12-P41, MSP3.10-MSP7.1, and P12-PVX_110945. This result confirmed for the first time that the known P12-P41 interaction in *P. falciparum* (Taechalertpaisarn et al., 2012) is conserved in the evolutionarily distant but related parasite, *P. vivax*. Biophysical measurements using SPR showed that the *P. vivax* P12-P41 interaction appears to be at least 3 times stronger than the *P. falciparum* P12-P41 interaction. Gene knockout experiments have shown that *P. falciparum* P12 is not essential for parasite invasion or growth *in vitro*, and antibodies against P12 do not significantly block erythrocyte invasion (Taechalertpaisarn et al., 2012). The difference in protein-protein interaction affinity between the species indicates that there could be functional differences between them. Interestingly, an interaction between *P. vivax* P12 and *P. falciparum* P41 was also detected, suggesting the existence of conserved binding sites, which may be attractive targets for vaccines against both *P. falciparum* and *P. vivax*. The reverse interaction between *P. falciparum* P12 and *P. vivax* P41 (in both bait-prey and prey-bait orientations) was not detected, which may assist in mapping the P12-P41 binding site within *Plasmodium* species.

Two novel interactions were also detected with AVEXIS and investigated by SPR; both putative interactions were validated by SPR and/or reciprocation of prey-bait and bait-prey interactions in AVEXIS. The failure of AVEXIS to detect the *P. vivax* MSP3.10-MSP7.1 interaction in the prey-bait orientation may indicate that artificially pentamerized MSP3.10 prey interferes with the formation of an oligomeric structure necessary for

binding. While *P. falciparum* MSP3.1 contains a conserved C-terminus important for oligomerization (Burgess et al., 2005), *P. vivax* MSP3s are not clear homologs and lack this feature (Rice et al., 2014, Jiang et al., 2013). Their potential for forming higher order structures is unknown. SPR showed that the *P. vivax* MSP3.10-MSP7.1 interaction has a relatively high avidity, potentially increased due to oligomerization of MSP7.1. Members of the *P. vivax* MSP3 family are known to peripherally associate with the merozoite surface (Jiang et al., 2013) and *P. vivax* MSP7s are predicted to have a similar location, based on their *P. falciparum* orthologs (Kadekoppala and Holder, 2010, Kadekoppala et al., 2010). MSP3 has no known binding partners, and *P. falciparum* MSP7 is known to form a complex with MSP1 (Kadekoppala and Holder, 2010). An interaction between MSP7 and MSP1, while perhaps predicted based on *P. falciparum* data, has not yet been established for *P. vivax*. In my study, the AVEXIS screen also did not detect such an interaction although only 3 of the 11 *P. vivax* MSP7 family members were included in our library, meaning that 1 of the members not expressed to date could interact with *P. vivax* MSP1. SPR data suggest that the *P. vivax* P12-PVX_110945 interaction is extremely weak. PVX_110945 is a hypothetical protein with no known function, but its transcription and genomic location give it a possible function in merozoite development, invasion of erythrocytes, or both. Additional experiments to co-localize these 2 interacting pairs in parasite isolates may help validate and shed light on the possible biological relevance of these interactions.

4.3.1 Limitations and future work

The *P. vivax* recombinant library AVEXIS screen against erythrocyte receptors failed to detect any interactions, including the known PvDBP-DARC interaction. For several reasons, however, these results do not conclusively indicate that none of these proteins interact. AVEXIS is optimized to have a very low false-positive rate with 2/8 expected interactions detected and 0/37 false positives in validation experiments described in (Bushell et al., 2008). This focus on eliminating false positives may also reduce the number of real positives. In addition, several baits and preys used in the screens were expressed at levels below the optimal concentrations for detecting interactions. Future studies exploring additional erythrocyte receptors and/or focusing on individual proteins in the library will be useful to fully evaluate the proteins' potential for erythrocyte binding.

The *P. vivax* P12-P41 interaction appears to have a much higher affinity than the analogous *P. falciparum* P12-P41 interaction. The difference in protein-protein interaction affinity between species indicates that there could be functional differences between them; therefore, we plan to investigate the potential function of the *P. vivax* P12-P41 interaction in merozoite invasion of reticulocytes and the possible invasion-blocking effects of immune IgG specific for P12 or P41 in Cambodian *P. vivax* isolates *ex vivo*.

We believe that this *P. vivax* recombinant library and expression approach will significantly improve our ability to study the biology of *P. vivax* erythrocyte invasion and the natural development of *P. vivax* immunity. This approach has already been successfully applied to both fronts in *P. falciparum* research (Crosnier et al., 2013, Crosnier et al., 2011, Bartholdson et al., 2012, Osier et al., 2014). Not only will this *P. vivax* library enable new screens for parasite-host interactions, it can also contribute to protein structure and immunoepidemiological studies. Structural studies can greatly increase our understanding of the function of these proteins. Such studies can also enhance the possibility of defining protective versus decoy conformational epitopes, which may have a tremendous impact on their ultimate potential as vaccine candidates (Chen et al., 2013). Future immunoepidemiological studies that use a panel of proteins should enable more systematic comparisons between proteins, in contrast to previous studies that have used only one or several proteins. The *P. vivax* library of expression plasmids is now freely available through the non-profit plasmid repository, Addgene (www.addgene.org), facilitating community efforts to identify, validate, and develop promising vaccines for *P. vivax* malaria. Finally, systematic generation of antibodies against library proteins will generate reagents that can be used in *ex vivo* invasion assays to test for invasion-blocking potential. Identification and prioritisation of new *P. vivax* vaccine antigens is therefore significantly enhanced by the work described here, even if specific receptors were not identified in this work.

4.4 Conclusion

We selected 39 *P. vivax* proteins that are predicted to localize to the merozoite surface or invasive secretory organelles, some of which show homology to *P. falciparum* vaccine candidates. Of these, I was able to express 37 full-length protein ectodomains in a mammalian expression system, which has been previously used to express *P. falciparum*

invasion ligands such as PfRH5. Using a method specifically designed to detect low-affinity, extracellular protein-protein interactions, I confirmed a predicted interaction between *P. vivax* 6-cysteine proteins P12 and P41, suggesting that the proteins are natively-folded and functional. This screen also identified 2 novel protein-protein interactions, between P12 and PVX_110945 and between MSP3.10 and MSP7.1, the latter of which was confirmed by surface plasmon resonance. As well as identifying new interactions for further biological studies, this library will be useful in identifying *P. vivax* proteins with vaccine potential, and studying *P. vivax* malaria pathogenesis and immunoepidemiology.

5 IMMUNOEPIDEMIOLOGY OF *P. VIVAX* PROTEIN LIBRARY

Generation of the recombinant protein library described in Chapter 4 gave us a unique opportunity to perform *P. vivax* immunoepidemiology studies in several *P. vivax*-endemic areas. In this chapter, I will describe those studies. Publication note: the results described in section 5.2.1.2 and portions of the introduction and discussion were slightly modified from a published manuscript (Hostetler et al., 2015). I drafted the text in these sections, which was edited by co-authors prior to publication. I am solely responsible for the work described in section 5.2.1, under the supervision of my PhD supervisors, Rick Fairhurst and Julian Rayner, except where noted in the text. The results in sections 5.2.2.1 and 5.2.2.2 were generated in close collaboration with Camila Franca and Ivo Mueller. Camila Franca performed all experiments in section 5.2.2.4. The results in 5.2.2 were recently published (Franca et al., 2016).

5.1 Introduction

The screening of our *P. vivax* recombinant proteins in several functional studies (Chapter 4) revealed several predicted and novel parasite protein-protein interactions, but did not investigate the potential significance of these proteins as vaccine candidates. Identifying naturally-acquired, clinically-protective associations with any of the proteins, while not confirming a causal relationship, could prioritize them for further functional studies and increase the pool of potential *P. vivax* vaccine candidates. Evaluating our panel of *P.*

vivax recombinant proteins, representing full-length ectodomains, in a series of immunoepidemiological studies would be a useful step in addressing this. In this chapter, I will outline studies in 3 *P. vivax*-endemic countries: Cambodia, the Solomon Islands (SI), and Papua New Guinea (PNG).



Figure 5.1: Immunoepidemiological study sites

Immunoepidemiological studies were designed using patient plasma collected in 3 *P. vivax*-endemic countries: Cambodia, Papua New Guinea, and the Solomon Islands. Map created at <https://www.amcharts.com/>.

Natural human immune responses to *P. vivax* during and after infection have been the subject of only limited study. While a few small-scale studies have produced full-length *P. vivax* recombinant proteins, even fewer have investigated whether immune IgG from *P. vivax*-exposed individuals recognize these proteins (Barbedo et al., 2007, Fowkes et al., 2012, Fraser et al., 1997, Garg et al., 2008, Lima-Junior et al., 2011, Lima-Junior et al., 2012, Michon et al., 1998, Oliveira et al., 2006, Rodrigues et al., 2005, Souza-Silva et al., 2010, Wickramarachchi et al., 2006, Woodberry et al., 2008, Xainli et al., 2003, Yildiz Zeyrek et al., 2011, Zeeshan et al., 2013, Ceravolo et al., 2009). The only 2 large-scale immunoreactivity screens did not exclusively use full-length proteins, and may therefore have missed critical epitopes (Lu et al., 2014, Chen et al., 2010).

It would first be important to evaluate the general IgG reactivity to the *P. vivax* recombinant protein library in clinical samples from *P. vivax*-infected patients. The NIH laboratory field site in Pursat Province, Cambodia (with Rick Fairhurst, LMVR/NIH and Socheat Duong, Cambodian National Center for Parasitology, Entomology, and Malaria

Control), had several ongoing *P. vivax* studies that could provide access to hundreds of plasma samples from *P. vivax*-infected patients. Cambodia is endemic for both *P. falciparum* and *P. vivax*. A recent publication by Maude et al. (2014) found that *P. falciparum* malaria transmission has greatly declined in recent years, with an estimated 81% decline in annual cases from 2009 to 2013 (Figure 5.2). This coincided with several intensified control measures including a scale-up of village malaria workers and insecticide-treated bed nets. In contrast, the number of *P. vivax* cases increased 490% from 2008 to 2011 and then declined 50% from 2011 to 2013. The annual parasite incidence (API) in Cambodia in 2013 was 4.6/1000, with *P. vivax* accounting for 67% of malaria cases (Maude et al., 2014). This finding underscores the challenge of eliminating *P. vivax*, which may become the predominant *Plasmodium* species disease burden in many areas of declining *P. falciparum* infections (Yekutieli, 1980).

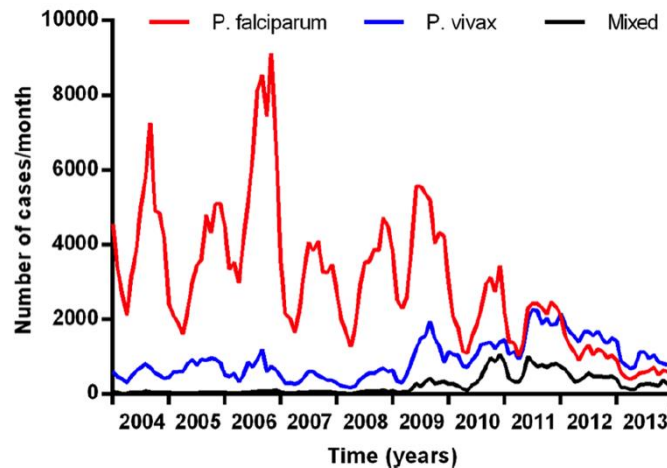


Figure 5.2: *Plasmodium* malaria cases in Cambodia from 2004 to 2014

Monthly malaria cases in Cambodia. *P. falciparum* cases declined while *P. vivax* cases increased from 2009 to 2013. Figure reprinted from (Maude et al., 2014) under the Creative Commons Attribution (CC BY) license.

Any IgG responses to the *P. vivax* recombinant protein library in Cambodian patient samples would suggest that the proteins are properly folded and further validate the use of the mammalian HEK293E system for expressing high-quality *P. vivax* recombinant proteins. Since the protein sequences were only based on the available *P. vivax* reference genome at the time (*P. vivax* Sal 1, originally from an El Salvadoran isolate), it was unknown whether IgG to proteins expressed by naturally-circulating Cambodian isolates would also react to our panel of proteins. The *msp3* genes, for example, are highly divergent among the *P. vivax* isolates sequenced to date (Neafsey et al., 2012, Rice et al.,

2014). The genetic divergence between *P. vivax* Sal 1 and naturally-circulating Cambodian isolates might lead to strain-specific IgG responses, thus limiting their potential usefulness in future immunoepidemiological studies and as vaccine candidates. We hypothesized that some reactivity would be detected, however, as 2 previous immunoreactivity screens, using *P. vivax* Sal 1-based proteins expressed in the wheat germ cell-free system and Korean isolates, detected several highly-immunogenic proteins (Lu et al., 2014, Chen et al., 2010).

The Cambodian patient samples would also facilitate the study of the development of the immune response following infection. In 2011, several hundred pairs of plasma samples were collected from Cambodian patients at the time of clinical presentation with *P. vivax* malaria (“acute”) and 21-28 days later (“convalescent”). Evaluating whether IgG responses to *P. vivax* antigens are boosted or decline during the convalescent period may clarify the early phase of *P. vivax* immunity development. Proteins with a declining IgG response in convalescent plasma may also serve as useful markers for recent past exposure.

Evaluating IgG reactivity to our protein panel in symptomatic patients would be a useful first step for identifying targets of natural immunity. Evaluating IgG reactivity in a cross-sectional study of both infected and uninfected subjects would enable us to investigate whether IgG responses correlate with exposure. Screening a panel of proteins would facilitate more systematic comparisons between proteins, in contrast to previous studies that have used only 1 or several proteins. Understanding exposure is particularly important for malaria control, as *P. vivax* infections are often low-density, asymptomatic, or both, making surveillance challenging (Waltmann et al., 2015). As indicators of recent exposure, IgG responses may serve as a marker for circulating *P. vivax* malaria even when parasite-positive samples are uncommon. This could be very useful in distributing resources in areas nearing *P. vivax* elimination. A cross-sectional survey conducted by Ivo Mueller and colleagues in SI in 2012 provided an ideal patient population to investigate age-stratified associations between IgG responses in subjects with and without *P. vivax* infections (Waltmann et al., 2015).

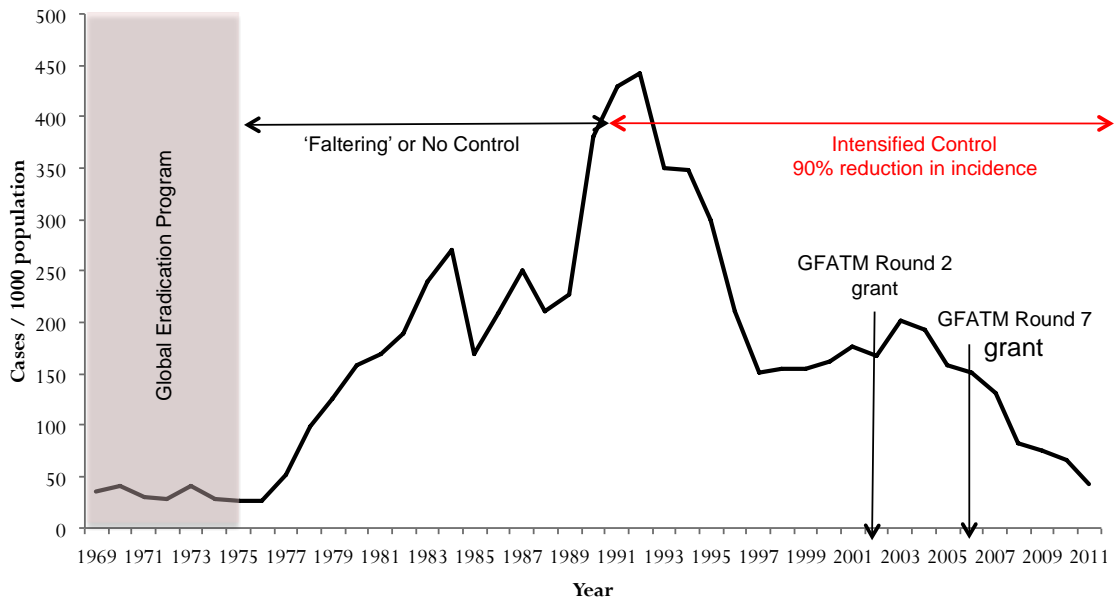


Figure 5.3: Malaria incidence in the Solomon Islands from 1969 to 2011

Malaria cases per 1000 people per year in SI. Malaria neared eradication during the Global Eradication Program in the early 1970's, followed by resurgence through 1992. Intensified control measures, including those funded by 2 Global Fund to Fight AIDS, Tuberculosis, and Malaria (GFATM) grants, reduced incidence by 90% by 2011. Image adapted from Andreea Waltmann; data from National Vector Borne Disease Control Program.

P. vivax malaria in SI is moderately endemic and but has a history of high transmission (Figure 5.3) (Pacific Malaria Initiative Survey Group on behalf of the Ministries of Health of and Solomon, 2010). Malaria incidence has dropped over a period of renewed control measures introduced in the early 1990's and Global Fund to Fight AIDS, Tuberculosis, and Malaria grants awarded in the 2000's. The 2012 cross-sectional survey enrolled 3501 residents of all ages in the Central Islands Province and assessed the prevalence of *P. falciparum*, *P. vivax*, *P. ovale*, and *P. malariae*. In contrast to Cambodia, where both *P. falciparum* and *P. vivax* are present relatively equally (Maude et al., 2014), *P. falciparum* infections in SI were nearly eliminated, with only 5 qPCR-detected cases, all of which were co-infections with *P. vivax*. *P. vivax*, however, maintained a prevalence of 13.4% overall. The vast majority of *P. vivax* infections were submicroscopic (72.9%) and/or afebrile (84.5%), and 23.5% contained gametocytes (Waltmann et al., 2015). Profiling IgG responses in this population would clarify the effects of age and infection status in an area of declining *P. vivax* transmission.

We next wanted to investigate whether any IgG responses to our proteins correlate with protection against clinical *P. vivax* malaria. There are few studies evaluating protective IgG responses, and a review of immunoepidemiological studies of *P. vivax* malaria found that only 3 antigens [MSP1, MSP3.10 (MSP3 α), and MSP9] were consistently associated with protection (Cutts et al., 2014), underscoring the need for a much broader set of antigens for study. The breadth of antibody response was shown to be important for protection from severe malaria in a Kenyan cohort study in 2014 using a panel of *P. falciparum* merozoite proteins expressed in the same system (Osier et al., 2014). Whether this observation holds true for *P. vivax* is unknown, as no studies evaluating protection from disease using a large panel of *P. vivax* proteins has been published. A longitudinal cohort study in PNG, led by Ivo Mueller, provided an ideal study population in which to address this research gap (Lin et al., 2010).

Four of the 5 *Plasmodium* species causing human malaria (*P. falciparum*, *P. vivax*, *P. malariae*, and *P. ovale*) are endemic to PNG. While Cambodia has low transmission and SI has moderate transmission, PNG has one of the highest *P. vivax* transmission rates in the world, with a 25% parasitemia prevalence recorded in 1995 (Genton et al., 1995). The cohort enrolled 264 PNG children aged 1-3 years, and followed them for 16 months. The median age of study participants was 1.7 years (IQR 1.3-2.5) and the PCR-detected prevalence of *P. vivax* infection at the beginning of the study was 55%. Studying naturally-acquired immunity (NAI) to *P. vivax* in very young children is essential since this immunity seems to develop faster than immunity to *P. falciparum* (Jeffery, 1966, Ciuca, 1934). This is supported by numerous studies in co-endemic areas that recorded the incidence of *P. vivax* peaking earlier than that of *P. falciparum* (Michon et al., 2007, Maitland et al., 1996, Phimpraphi et al., 2008, Mueller et al., 2009b, Lin et al., 2010).

5.1.1 Benefits of these studies

The development of NAI during *P. vivax* infection has been the subject of limited study, partially due to the difficulty in producing full-length *Plasmodium* proteins. The *P. vivax* recombinant protein library facilitates the systematic comparison of responses to a panel of proteins produced on the same platform. Our studies with Cambodian patient plasma aim to characterize the IgG reactivity to the proteins as a useful first step to prioritize them for further study. Our studies evaluating IgG responses from population studies in SI and PNG aim to address whether any of these responses is correlated with age,

infection status, and protection from disease. The results from each of these studies will give additional information about the potential utility of the protein library in global *P. vivax* immunoepidemiological studies and prioritize potential vaccine candidates for testing in functional studies.

5.1.2 Objectives

- i. To evaluate IgG reactivity from *P. vivax*-exposed individuals to the library of *P. vivax* recombinant proteins
- ii. To examine whether IgG responses correlate with *P. vivax* exposure or protection from clinical *P. vivax* malaria.

5.2 Results

5.2.1 Antibody responses to *P. vivax* recombinant proteins in Cambodia

5.2.1.1 Pilot reactivity screen using Cambodian patient plasma and a subset of *P. vivax* recombinant proteins

To test whether the *P. vivax* recombinant proteins were immunoreactive in Cambodian patient plasma and what plasma concentration was best to use, I serially diluted (1:200, 1:400, 1:800, 1:1600) plasma samples from 5 *P. vivax* malaria-exposed Cambodian individuals against 3 proteins (P12, P41, and MSP10). Negative controls were a pooled serum sample from American malaria-naïve individuals and the Cd4d3+d4 tag present in all proteins. Cambodian plasma pairs were selected based on prior ELISA data showing reactivity to recombinant *P. vivax* MSP1 (produced by Daria Nikolaeva in the Long laboratory), in order to maximize our chances of observing reactivity. The ELISA data suggested that all our proteins might be immunoreactive, with P12 showing the highest reactivity (Figure 5.4). A single plasma sample, CAM29, showed 5 times higher reactivity to the Cd4d3+d3 negative control than the other plasma samples. Higher dilutions reduced this response to that of negative controls, but correspondingly reduced the overall response to the recombinant proteins. A 1:600 dilution seemed optimal for observing IgG responses for the largest number of proteins versus plasma samples, while minimizing non-specific responses to the negative controls (the CD4d3+d4 tag alone and malaria-naïve control sera). This plasma dilution was then used in subsequent ELISAs.

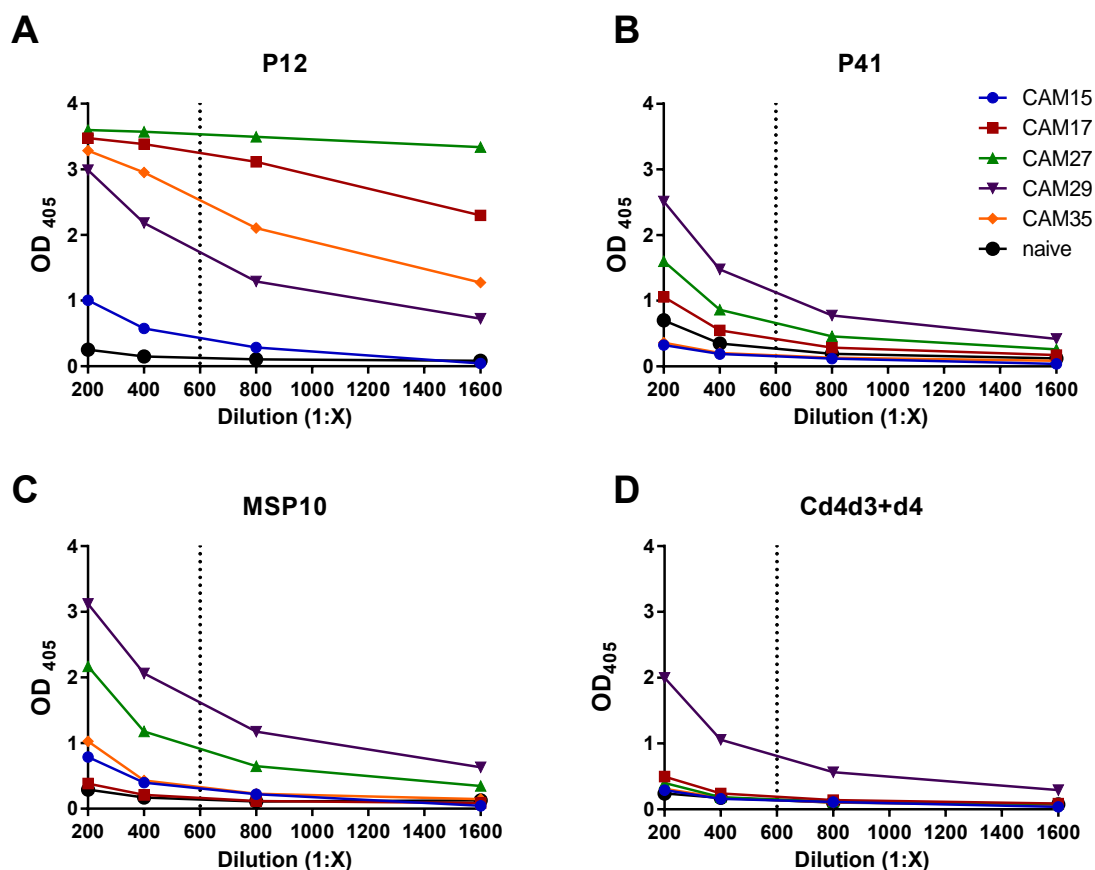


Figure 5.4: Testing Cambodian patient plasma against *P. vivax* recombinant proteins

(A-D) Serially diluted (1:200, 1:400, 1:800, 1:1600) plasma samples from 5 *P. vivax*-infected Cambodian patients (CAM15, CAM17, CAM27, CAM29, CAM35) and a pooled sera sample from American *P. vivax*-naïve individuals (negative control) were screened against 3 *P. vivax* recombinant proteins (P12, P41, MSP10) and a negative control protein (Cd4d3+d4) using ELISA with optical density (OD) measured at 405 nm. Plasma dilution 1:600 (dotted lines) minimized reactivity to negative controls while enabling detection of IgG reactivity to *P. vivax* proteins.

To expand the IgG reactivity screening, I used ELISAs to measure IgG responses in 42 *P. vivax*-infected Cambodian patient plasma samples to 6 full-length *P. vivax* recombinant protein ectodomains [P12, P12p, P41, MSP7 (PVX_082675), MSP10, Pv34]. I included 6 American malaria-naïve sera samples and Cd4+d3+d4 as negative controls. All antigens showed reactivity above Cd4d3+d4 when comparing the mean OD values for all tested samples; however, only those differences for P12, Pv34, P41, and MSP10 were statistically significant (Figure 5.5) (Mann-Whitney U test, $P < 0.0001-0.04$).

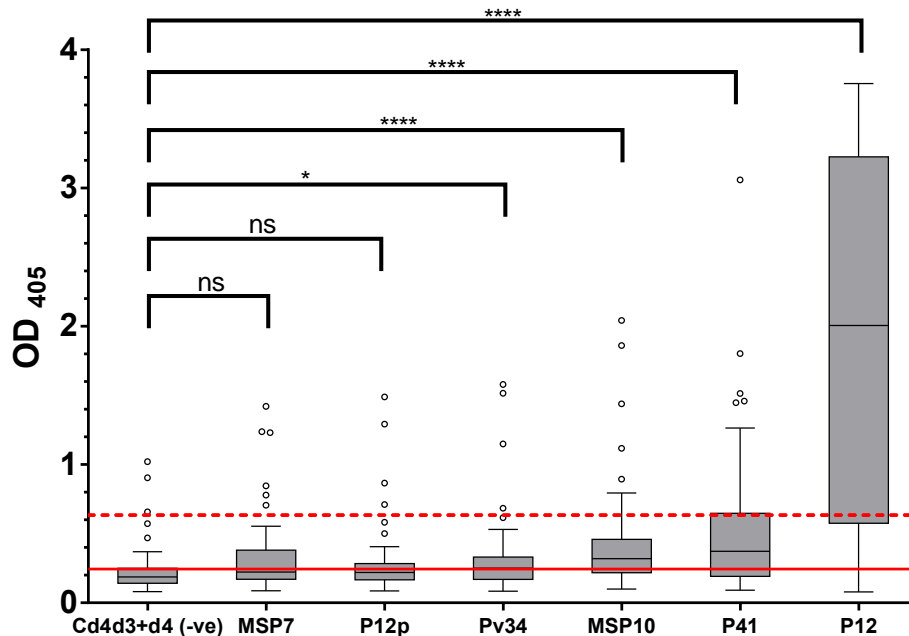


Figure 5.5: Cambodian patient plasma IgG reactivity to full-length *P. vivax* recombinant protein ectodomains

42 Cambodian *P. vivax*-infected plasma samples and 6 American *P. vivax*-naïve control sera were tested against 6 proteins (P12, P12p, Pv34, P41, MSP7.6, MSP10) and the rat CD4d3+d4 tag present in all proteins by ELISA; optical density (OD) was measured at 405 nm. Boxplots show median OD (horizontal bar), IQR (boxes), range (whiskers), and outliers (open circles). IgG responses to the rat Cd4d3+d4 tag alone are shown as mean (solid red line) and mean + 2 standard deviations (dashed red line). P12, Pv34, P41, and MSP10 showed significantly higher population responses than Cd4d3+d4 (*). ns=Not significant. (Mann-Whitney U test, $P < 0.0001-0.02$; P values were deemed significant if < 0.05).

Cd4d3+d4 had higher reactivity against Cambodian *P. vivax*-infected patient plasma compared to American *P. vivax*-naïve control sera, with 62% of patient samples showing reactivity greater than 2 standard deviations above the mean reactivity of the naïve controls. This reactivity to Cd4d3+d4 was not observed in screens using a panel of *P. falciparum* recombinant proteins produced in the same expression system against a cohort of Kenyan samples (Osier et al., 2014). This increased reactivity could be more visible in this ELISA because a 1:600 plasma dilution was used instead of the 1:1000 plasma dilution that was used in the *P. falciparum* ELISAs. The Cambodian plasma samples were selected based on reactivity to recombinant MSP1 (data provided by Daria Nikolaeva and Carole Long, LMVR/NIH). This may have biased the plasma set toward samples that were generally more reactive, even non-specifically; downstream ELISAs included samples based on patient history or chosen randomly (as described) in order to

minimize this potential bias. The Cd4d3+d4 responses were also subtracted to correct for background when setting a reactivity cut-off value.

When considering individual responses, at least some patient samples reacted to all antigens. The reactivity cut-off value was set at 2 standard deviations above the mean of the naïve sera controls after subtracting the Cd4d3+d4 response from each sample. Reactivity to P12 was highest in both magnitude and breadth. There was significant variation in reactivity with P41 and P12 reactive in 43% and 98% of samples, respectively (Table 5.1).

Table 5.1: Pilot Cambodian seropositivity summary

	<i>P. vivax</i> recombinant protein					
	P12	MSP10	Pv34	MSP7.6	P41	P12p
Reactive plasma samples* n (%)	41(98)	29(69)	24(57)	24(57)	18(43)	21(50)

*Reactivity cut-off value set at 2 SD above mean of 6 American malaria-naïve controls after correcting for background responses by subtracting the Cd4d3+d4 values.

5.2.1.2 *P. vivax* recombinant proteins are immunoreactive and contain conformational epitopes.

To test whether the library of biotinylated *P. vivax* recombinant protein ectodomains were immunoreactive, and to establish whether they contained conformational epitopes, I screened them by ELISA against diluted (1:600) pooled plasma from 14 Cambodian patients with acute vivax malaria and pooled control sera from 5 American malaria-naïve individuals (Figure 5.6). Exposed IgG reacted more strongly than naïve IgG to all *P. vivax* proteins, and reacted only weakly to the rat Cd4d3+4 tag present in all expressed proteins. Five proteins (MSP5, P12, GAMA, CyRPA, PVX_081550) were particularly reactive; thus, ELISAs for these proteins were repeated using more-diluted (1:1000) plasma pools (Figure 5.6). Of the 34 proteins tested, 27 showed at least a 2-fold change in seroreactivity between the naïve IgG versus the exposed IgG for each protein, with 7 proteins (P92, PVX_084815, PVX_116775, RBP2-like, PVX_001015, PVX_110960, and PVX_110965) showing a lesser change. The majority of proteins showed at least a 3-fold change (MSP1, MSP3.1, MSP3.3, MSP3.10, MSP4, MSP5, MSP7.1, MSP10, P12, P12p, P38, P41, GAMA, ARP, CyRPA, DBP, DBP-RII, PVX_081550).

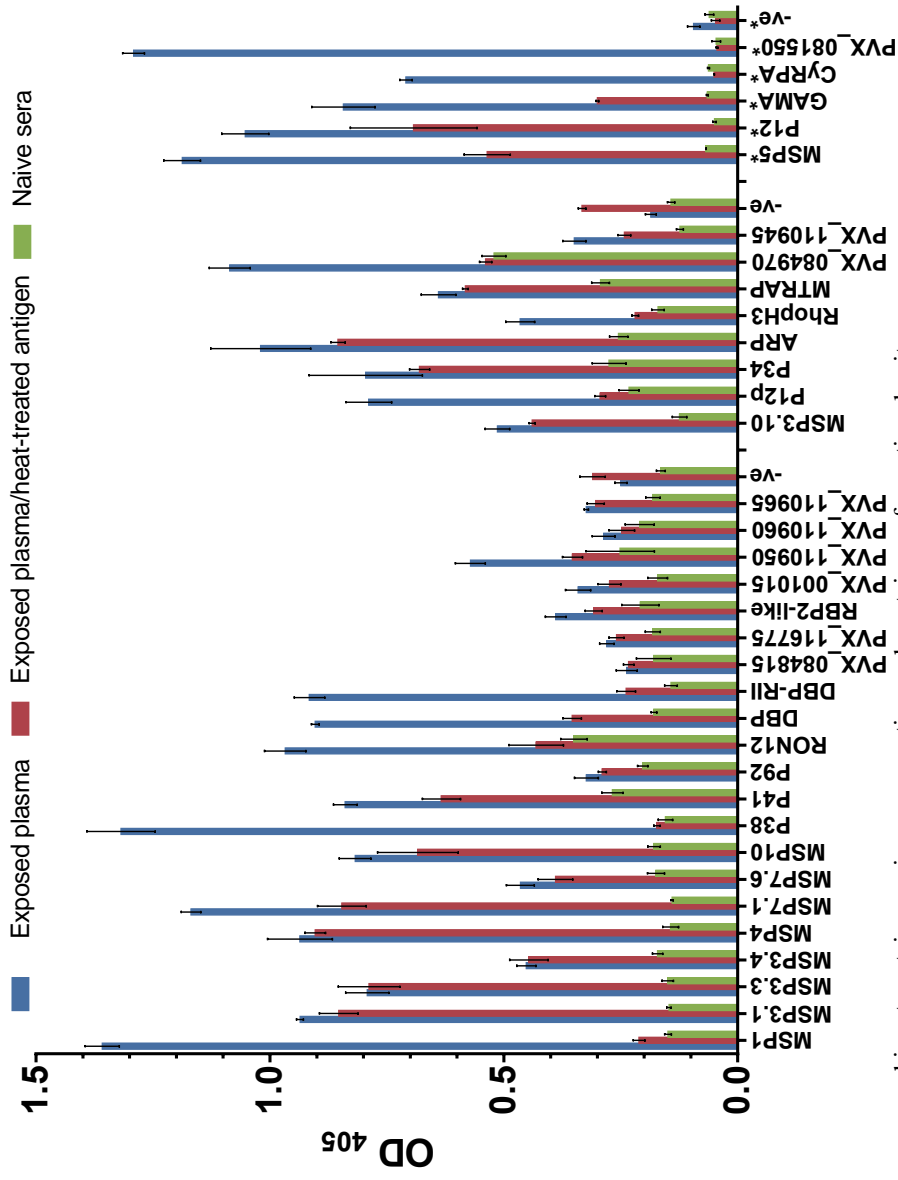


Figure 5.6: Multiple *P. vivax* recombinant proteins are immunoreactive and contain conformational epitopes. The immunoreactivity of 34 biotinylated *P. vivax* recombinant proteins was assessed using diluted (1:600) plasma pools from 14 Cambodian vivax malaria patients (blue bars) and 5 American malaria-naïve individuals (green bars). The immunoreactivity of heat-treated proteins was assessed in parallel using the Cambodian plasma pool (red bars); reduced responses indicate the presence of heat-labile conformational epitopes. The immunoreactivity of highly reactive proteins (*) was assessed using more-diluted (1:1000) plasma pools. Optical density (OD) at 405 nm was measured at various times, but only the mean value nearest to 1.0 for each antigen is shown. Negative control (-ve) was rat Cd4d3-rd4 tag. Bar charts show mean \pm SD; n=3. Figure reprinted from (Hostetler et al., 2015) under the Creative Commons Attribution (CC BY) license.

To test whether IgG responses were directed at conformational epitopes, I heat-treated all 34 proteins and screened them for seroreactivity in parallel with untreated proteins. Of the 34 proteins tested, 18 (MSP1, MSP7.1, P38, P41, RON12, DBP, DBP-RII, RBP2-like, PVX_110950, P12p, RhopH3, PVX_084970, PVX_110945, MSP5, P12, GAMA, CyRPA, PVX_081550) showed at least a 20% decrease in seroreactivity when heat treated (Figure 5.6), indicating they contained conformation-sensitive epitopes, and suggesting that the recombinant proteins were properly-folded. Twelve of these proteins (MSP1, P38, RON12, DBP, DBP-RII, P12p, RhopH3, PVX_084970, MSP5, GAMA, CyRPA, PVX_081550) showed at least a 50% reduction in seroreactivity when heat-treated (Figure 5.6). Naïve IgG showed appreciable reactivity to PVX_084970; screening this protein against individual plasma samples may resolve whether it nonspecifically reacts to IgG from all or only some of the 5 serum donors.

5.2.1.3 IgG response in *P. vivax* Cambodian acute and convalescent plasma

A study conducted in 2011 at the NIH Cambodian field site collected several hundred pairs of “acute” plasma samples when patients presented to the clinic with symptomatic *P. vivax* malaria and “convalescent” samples 21-28 days later. We were interested in understanding whether any IgG boosting occurred during this period, which might serve as a signal for the development of NAI. I first needed to create a standard pool of highly reactive plasma to include on each plate to account for plate-to-plate variations. I started with a subset of 11 available proteins (MSP1, MSP5, MSP10, P12, P41, ARP, CyRPA, DBP, GAMA, RBP2-like, PVX_081550) and screened them by ELISA to 48 Cambodian plasma samples (1:1000) and 24 American malaria-naïve adult plasma samples (1:1000). All proteins showed significantly higher reactivity compared to the Cd4d3+d4 tag (Mann-Whitney U tests, $P < 0.0001$), and reactivity varied for each protein (Figure 5.7). Responses to Cd4d3+d4 were significantly lower than in the ELISA optimization screens. This is likely due to the use of a 1:1000 plasma dilution, which was selected based on the Cambodian reactivity screen using pooled plasma.

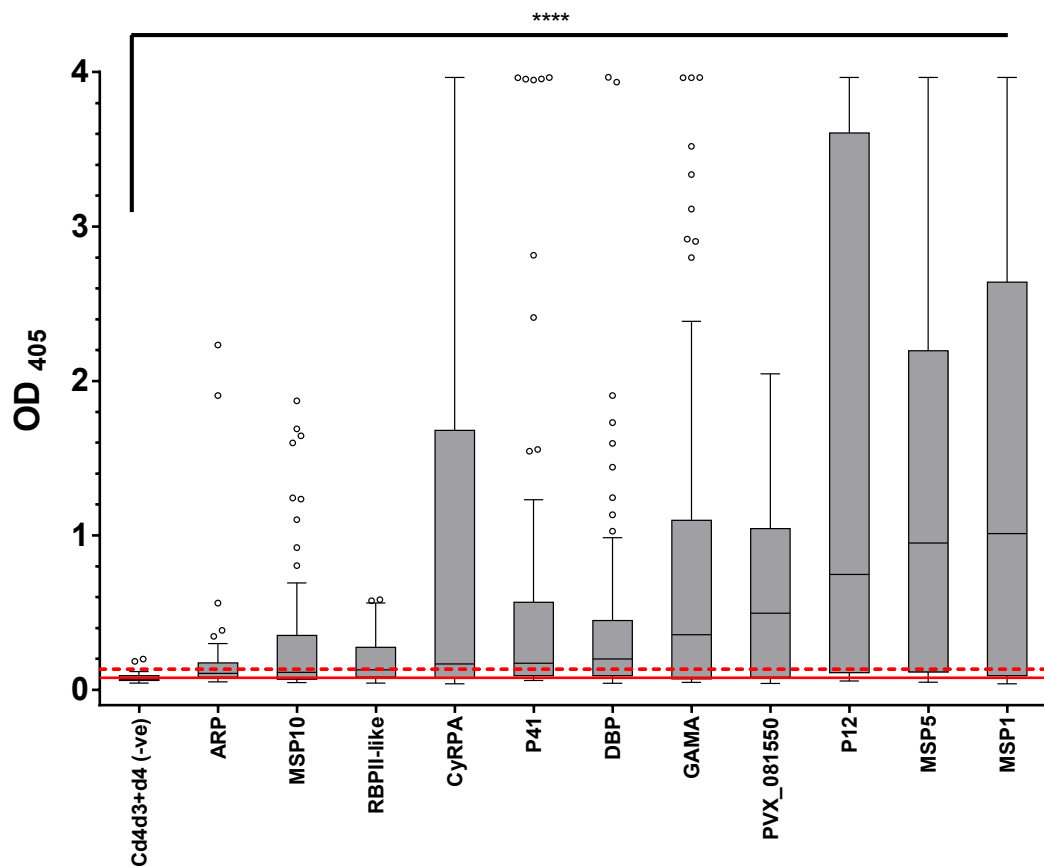


Figure 5.7: Reactivity in individual Cambodian plasma samples

48 Cambodian *P. vivax*-infected plasma samples and 24 American *P. vivax*-naïve controls were screened against 11 proteins (MSP1, MSP5, MSP10, P12, P41, ARP, CyRPA, DBP, GAMA, RBP2-like, PVX_081550) and rat CD4d3+d4 present in all proteins by ELISA. Optical densities (OD) at 405 nm are shown. Boxplots show median OD (horizontal bar), IQR (box) range (whiskers), and outliers (open circles). IgG response to Cd4d3+d4 tag alone are shown with mean (solid red line) and mean + 2 SD (dashed red line). All proteins showed significantly higher population responses than Cd4d3+d4 (*) (Mann-Whitney U tests, $P < 0.0001$; P values were deemed significant if < 0.05).

I calculated reactivity cut-off values for each protein as 2 SD above the mean of the American malaria-naïve adult plasma samples after correcting for background by subtracting the Cd4d3+d4 values. The reactivity ranged from 35% in ARP and MSP10 to 100% for MSP1 and GAMA. Over 90% of samples reacted to 5 proteins (MSP1, GAMA, MSP5, PVX_081550, P12). The results showed a similar relationship between P12, P41, and MSP10 compared to the initial ELISA optimization experiments, with P12 the most

reactive of these 3 proteins. These seropositivity results can subsequently be compared to those obtained at other study sites.

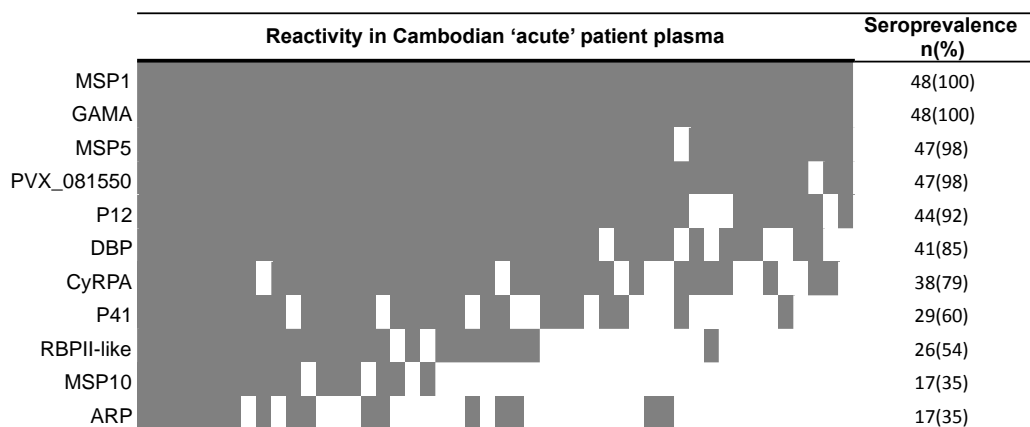


Figure 5.8: Seroprevalence in “acute” Cambodian plasma samples

Heatmap showing reactivity in 48 acute Cambodian plasma samples (columns) for 11 *P. vivax* recombinant proteins (rows; MSP1, MSP5, MSP10, P12, P41, ARP, CyRPA, DBP, GAMA, RBP2-like, PVX_081550). Gray boxes represent optical density (OD) values 2 SD above the mean of the American malaria-naïve adult plasma samples after correcting for background by subtracting the Cd4d3+d4 values.

From these results, I created a standard pool of high responders for each protein to control for plate-to-plate variation in subsequent ELISAs. I next screened 10 *P. vivax* recombinant proteins (MSP1, MSP5, MSP10, P12, P41, ARP, CyRPA, DBP, GAMA, PVX_081550) in duplicate wells against 18 pairs of acute and convalescent plasma samples. I had insufficient amounts of RBP2-like and therefore could not test this protein. All of the proteins showed mean reactivity significantly higher than the Cd4d3+d4 tag alone (Mann-Whitney *U* test, $P < 0.0001$). The mean reactivity in convalescent plasma was 1.9 times higher in ARP, but was similar for other proteins, with acute mean OD values less than 25% higher or lower than convalescent mean OD values. There were no significant differences between acute and convalescent groups when comparing the unpaired data (Mann-Whitney *U* tests, $P = 0.35-0.95$) (Figure 5.9). There were also no significant differences when analysing the plasma time points as pairs (Wilcoxon signed-rank test, $P = 0.08-0.97$) (Figure 5.10). This result could indicate that more samples may be required to detect small changes in IgG in the month following infection. Alternatively, a 3-4 week separation in plasma collection time points may be too narrow to observe changes immediately following infection and recovery.

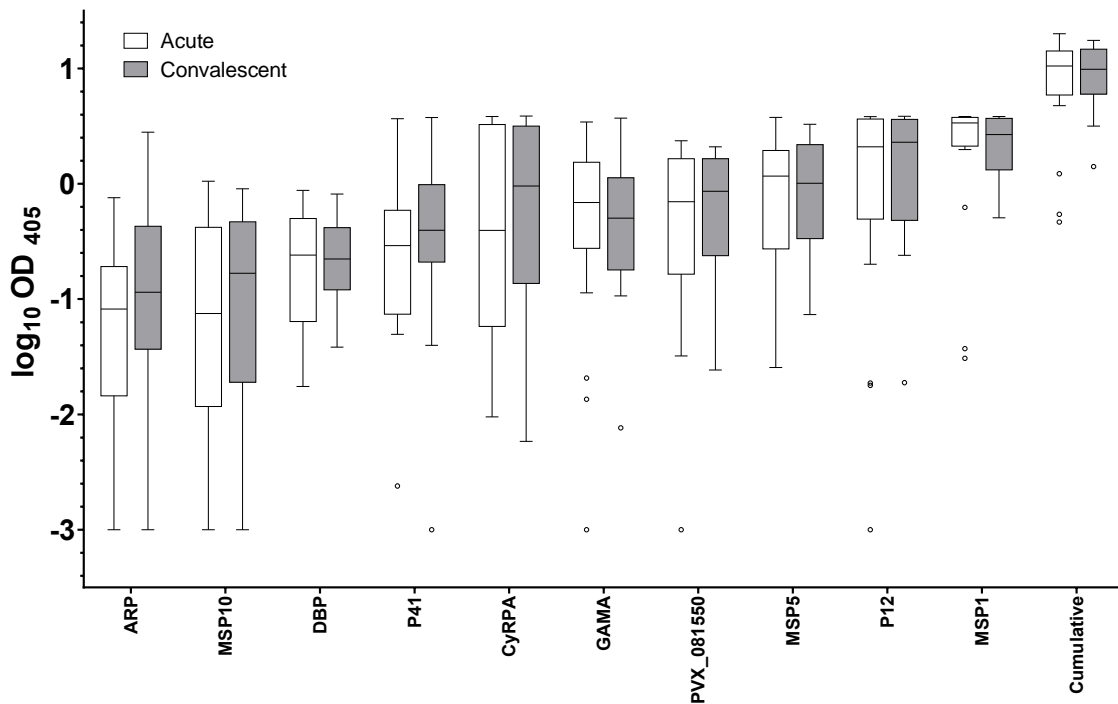


Figure 5.9: IgG responses in acute and convalescent Cambodian patient plasma samples (unpaired)

18 Cambodian paired “acute” and “convalescent” plasma samples were screened against 10 *P. vivax* recombinant proteins (ARP, MSP10, DBP, P41, CyRPA, GAMA, PVX_081550, MSP5, P12, MSP1) by ELISA. Log optical densities (OD) at 405 nm are shown. Boxplots show median (horizontal bar), IQR (box), range (whiskers), and outliers (open circles). No significant differences between plasma time points were detected (Mann-Whitney *U* tests, $P=0.35-0.95$).

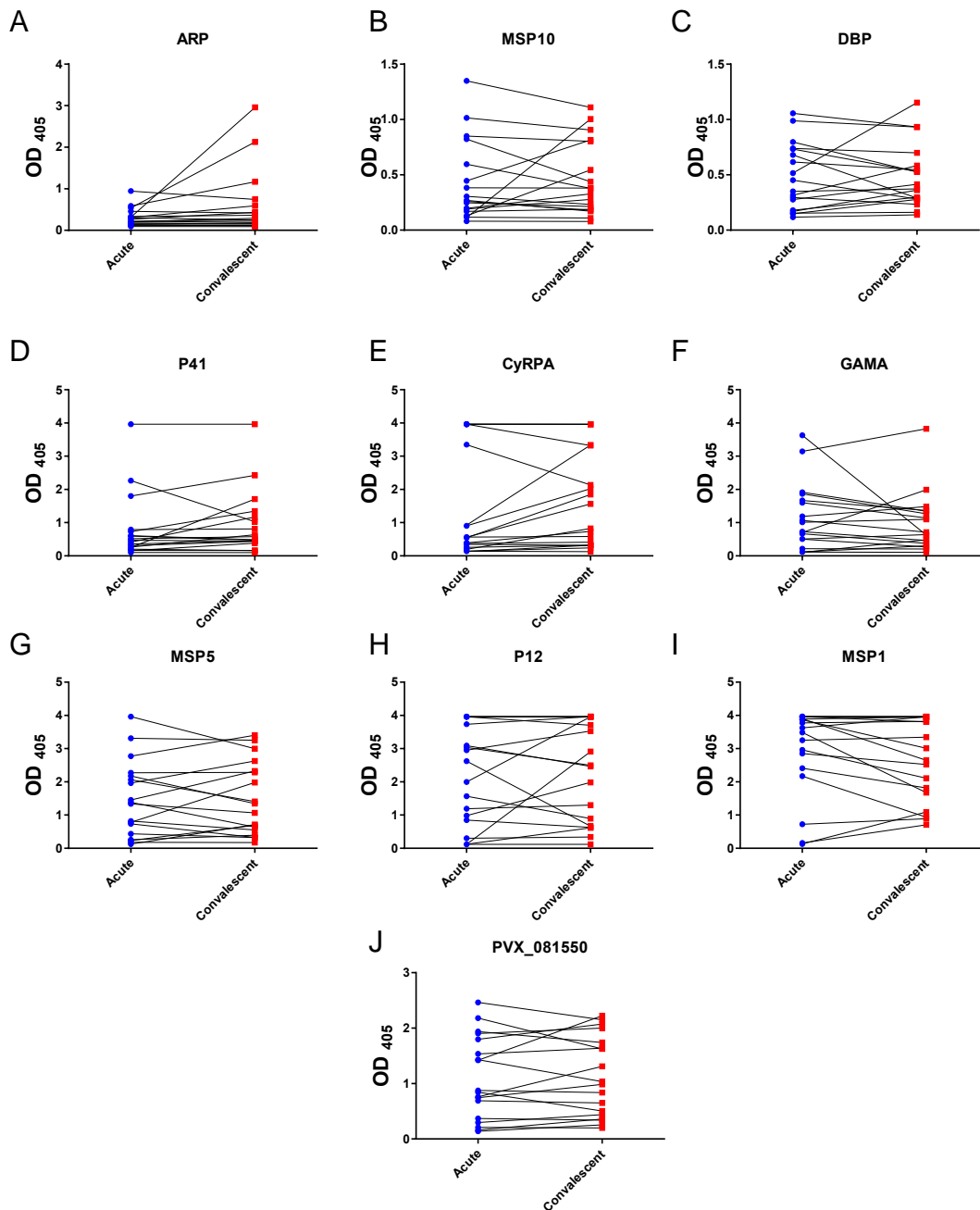


Figure 5.10: IgG responses in acute and convalescent Cambodian patient plasma samples (paired)

ELISA results measuring OD at 405 nm for 18 paired “acute” and “convalescent” plasma samples against 10 *P. vivax* recombinant proteins (ARP, MSP10, DBP, P41, CyRPA, GAMA, PVX_081550, MSP5, P12, MSP1) are shown. The lines connect plasma time points. No significant differences were detected (Wilcoxon signed-rank test, $P=0.08-0.97$).

5.2.2 Antibody responses to *P. vivax* recombinant proteins in Solomon Islands and Papua New Guinea

The Cambodian plasma studies suggested the majority of our antigen library was immunoreactive, but these studies were not designed to investigate correlations with exposure or protection. Examining these correlations required cross-sectional or longitudinal cohort study designs, respectively. We addressed these questions in collaboration with Ivo Mueller's laboratory at WEHI. The Mueller laboratory conducted a collaborative cross-sectional survey in SI in 2011 (Waltmann et al., 2015) and a longitudinal cohort study in PNG in 2006 (Lin et al., 2010). These studies were ideal for pursuing a collaborative set of projects to examine antibody responses comprehensively to our library of *P. vivax* recombinant proteins. This culminated in 3 studies described below, with the relative laboratory contributions of each group noted:

SI reactivity screening: screened 34 recombinant *P. vivax* proteins against 48 plasma samples stratified by 2 age groups. Proteins were expressed by Jessica Hostetler (Rayner and Fairhurst laboratories), and ELISAs were performed by Jessica Hostetler and Camila Franca (Mueller laboratory).

SI comprehensive screening: expanded screening against a subset 12 proteins against 144 plasma samples stratified by 3 age groups and 3 infection statuses. Proteins were expressed by Jessica Hostetler, and ELISAs were performed by Jessica Hostetler and Camila Franca.

PNG cohort screening: selected 6 proteins from our SI screening results to study in high-throughput Luminex assays. Proteins were selected by Jessica Hostetler and Julian Rayner, expressed and purified by Sumana Sharma (Wright and Rayner laboratories). Camila Franca performed all Luminex plasma screening assays, analysed the results, and produced the figures.

The proteins used in each study are summarized in Table 5.2.

Table 5.2: *P. vivax* recombinant proteins used in SI and PNG screens

Group	Accession number	Name	Product	Screen		
				SI R *	SI C *	PNG **
Merozoite surface proteins (MSPs)	PVX_099980	MSP1	merozoite surface protein 1	X	X	
	PVX_097680	MSP3.3, MSP3 β	merozoite surface protein 3	X	X	
	PVX_097720	MSP3.10, MSP3 α	merozoite surface protein 3	X	X	
	PVX_097685	MSP3.4	merozoite surface protein 3	X		
	PVX_003775	MSP4	merozoite surface protein 4, putative	X		
	PVX_003770	MSP5	merozoite surface protein 5	X	X	
	PVX_082700	MSP7.1	merozoite surface protein 7 (MSP7)	X		
	PVX_082675	MSP7.6	merozoite surface protein 7 (MSP7)	X	X	
	PVX_082655	MSP7.9	merozoite surface protein 7 (MSP7), putative	X		
	PVX_114145	MSP10	merozoite surface protein 10, putative	X	X	
6-cysteine proteins	PVX_113775	P12	6-cysteine protein	X	X	X
	PVX_113780	P12p	6-cysteine protein (P12p)	X		
	PVX_097960	P38	6-cysteine protein (P38)	X		
	PVX_000995	P41	6-cysteine protein	X	X	X
	PVX_115165	P92	6-cysteine protein	X		
Merozoite proteins not in other families	PVX_001725	RON12	rhostry neck protein 12, putative	X		
	PVX_088910	GAMA	GPI-anchored micronemal antigen, putative	X	X	X
	PVX_090075	Pv34, PV2	apical merozoite protein (Pf34)	X		
	PVX_090210	ARP	asparagine-rich protein	X	X	X
	PVX_090240	CyRPA	cysteine-rich protective antigen	X	X	X
	PVX_095055	RIPR	Rh5 interacting protein, putative	X	X	
	PVX_098712	RhopH3	high molecular weight rhostry protein 3, putative	X		
	PVX_110810	DBP DBP-RII	Duffy receptor precursor Duffy binding protein reg. II	X X		

	PVX_111290	MTRAP	merozoite TRAP-like protein, putative	X	
	PVX_081550		StAR-related lipid transfer protein, putative	X	X
	PVX_084815		conserved <i>Plasmodium</i> membrane protein, unknown function	X	
	PVX_084970		hypothetical protein, conserved	X	
	PVX_116775		hypothetical protein, conserved	X	
No known <i>P. falciparum</i> 3D7 homologs	PVX_101590	RBP2-like	reticulocyte-binding protein 2 (RBP2), like	X	
	PVX_001015		Pf52-like protein, putative	X	
	PVX_110950		hypothetical protein	X	
	PVX_110960		hypothetical protein	X	
	PVX_110965		hypothetical protein	X	

Abbreviations: SI-R=Solomon Islands reactivity screen; SI-C=Solomon Islands comprehensive screen; PNG=Papua New Guinea screen

* Biotinylated protein

** 6-His-tagged (purified) protein

5.2.2.1 Solomon Islands reactivity screening

I expressed 34 biotinylated *P. vivax* recombinant proteins (as in section 2.3.3), summarized in Table 5.2, and Camila Franca and I screened them against 48 plasma samples (1:500, based on data not shown from pilot sera screening) from a cross-sectional cohort study in SI. Plasma samples from infected and uninfected individuals were selected at random from adolescent and adult age-groups (ages 10-19 and 20-99 years) as they had a higher lifetime exposure to *P. vivax* and were more likely to have acquired immunity. Over 85% (29/34) of the proteins showed mean reactivity significantly higher than the Cd4d3+d4 tag alone (Figure 5.11) (Mann-Whitney *U* tests, $P < 0.0001-0.018$). These initial results confirmed that the protein library was largely immunogenic in a second *P. vivax* endemic area. Reactivity was highly variable with a 12-fold difference in mean response from P92 to CyRPA. Mean responses were similar in adolescents and adults except for 7/34 proteins: PVX_110960, PVX_110965, MTRAP, ARP, GAMA, PVX_081550, and MSP1 ($P < 0.0001-0.042$).

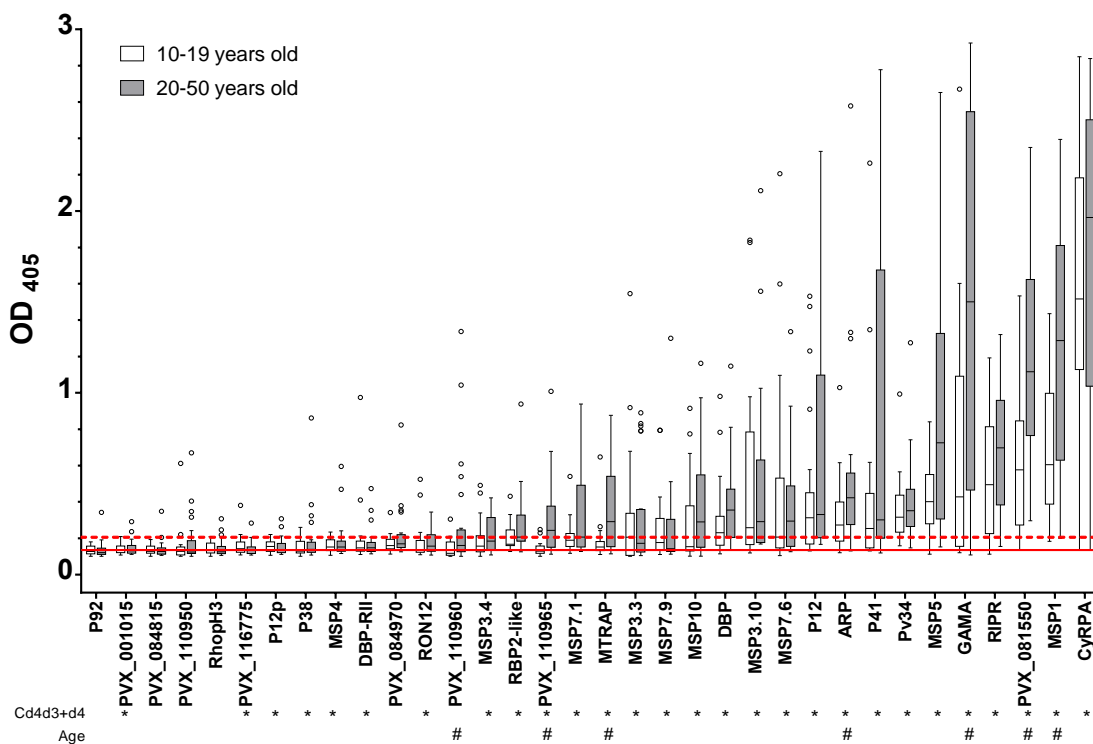


Figure 5.11: *P. vivax* recombinant proteins are immunoreactive in SI patient plasma samples

The total IgG response to 34 biotinylated *P. vivax* recombinant proteins was assessed using diluted (1:500) plasma from Solomon Islander plasma from adolescents (10-19 years, white boxes, n=22) and adults (20-50 years, gray boxes, n=24). Optical densities (OD) at 405 nm are shown. Boxplots show median OD (horizontal bar), IQR (box), range (whiskers), and outliers (open circles). IgG response to rat Cd4d3+d4 tag alone is shown as mean (solid red line) and mean + 2 standard deviations (dashed red line). 29/34 proteins showed significantly higher population responses than Cd4d3+d4 (*). 7/34 proteins showed significant differences between age group responses (#) (Mann-Whitney U tests, $P < 0.0001-0.042$; P values were deemed significant if < 0.05).

5.2.2.2 Solomon Islands comprehensive screening

The protein reactivity varied widely in the SI reactivity screen, and we wanted to expand the screening for a subset of proteins. This was limited based on reagents, supplies, and time to screening 12 of the most highly immunogenic proteins (MSP1, MSP3.3, MSP3.10, MSP5, MSP7.6, MSP10, P12, P41, ARP, CyRPA, GAMA, RIPR) against 144 sera samples (including the 48 included in the reactivity screen) selected randomly for a 3x3 factorial design stratifying 3 age groups (5-9, 10-19, 20-80 years) and by 3 infection groups:

- Negative by PCR and light microscopy (LM) = Not infected
- Positive by PCR and negative by LM = PCR+
- Positive by PCR and positive by LM = PCR+ LM+

We first evaluated reactivity by defining positivity as responses more than 2 SD above the mean of the Australian malaria-naïve adult plasma samples after correcting for background by subtracting the Cd4d3+d4 values (Figure 5.12, Table 5.3).

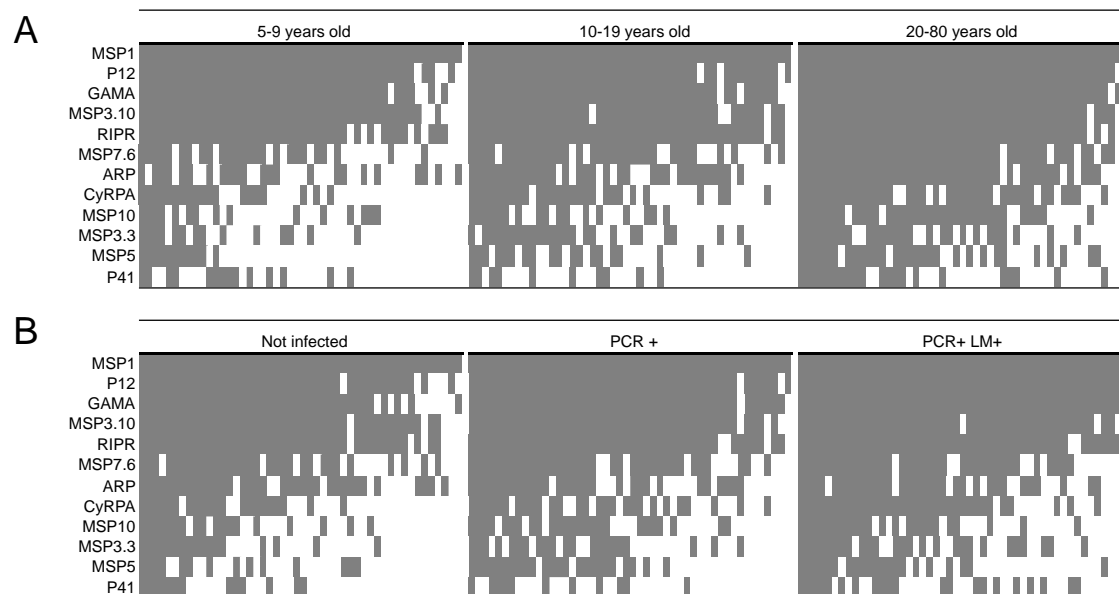


Figure 5.12: Seroreactivity in SI

Heatmaps of seropositivity, with columns representing individual plasma samples and rows representing proteins. Gray boxes represent optical density (OD) values 2 SD above the mean of the Australian malaria-naïve adult plasma samples after correcting for background by subtracting the Cd4d3+d4 values. (A) Samples grouped by age categories (5-9 years, n=48; 10-19 years, n=48; and 20-80 years, n=48). (B) Samples grouped by infection status: Not infected (n=48), samples positive by PCR (PCR+, n=48), and samples positive by both PCR and light microscopy (PCR+ LM+, n=48).

Table 5.3: Seroreactivity in SI comprehensive screen

	Age (years), n (%)			Infection status, n (%)			Total
	5-9	10-19	20-80	NI	PCR+	PCR+ LM +	
MSP1	48(100)	48(100)	48(100)	48(100)	48(100)	48(100)	144(100)
P12	44(92)	45(94)	47(98)	42(88)	46(96)	48(100)	136(94)
GAMA	42(88)	44(92)	47(98)	39(81)	46(96)	48(100)	133(92)
MSP3.10	43(90)	43(90)	46(96)	43(90)	44(92)	45(94)	132(92)
RIPR	41(85)	46(96)	45(94)	42(88)	44(92)	46(96)	132(92)
MSP7.6	25(52)	33(69)	43(90)	31(65)	35(73)	35(73)	101(70)
ARP	23(48)	32(67)	41(85)	35(73)	29(60)	32(67)	96(67)
CyRPA	19(40)	27(56)	36(75)	21(44)	27(56)	34(71)	82(57)
MSP10	15(31)	17(35)	34(71)	18(38)	25(52)	23(48)	66(46)
MSP3.3	14(29)	24(50)	26(54)	17(35)	21(44)	26(54)	64(44)
MSP5	11(23)	19(40)	29(60)	14(29)	24(50)	21(44)	59(41)
P41	14(29)	10(21)	21(44)	13(27)	11(23)	21(44)	45(31)

Table 5.4: Seroreactivity in SI and Cambodian parasitemic plasma samples

	SI (PCR+ LM +)	Cambodia (acute)
MSP1	48(100)	48(100)
P12	48(100)	44(92)
GAMA	48(100)	48(100)
PVX_081550		47(98)
MSP3.10	45(94)	
RIPR	46(96)	
DBP		41(85)
MSP7.6	35(73)	
ARP	32(67)	17(35)
CyRPA	34(71)	38(79)
MSP10	23(48)	17(35)
MSP3.3	26(54)	
RBPII-like		26(54)
MSP5	21(44)	47(98)
P41	21(44)	29(60)

Similar to Cambodia, the overall seropositivity varied widely from 31% for P41 to 100% for MSP1 (Table 5.3). Five of the 12 proteins were recognized in over 90% of samples (MSP1, P12, GAMA, MSP3.10, RIPR). The seropositivity values in the Cambodian survey of acute plasma (positive by LM) and the higher-density infections in SI (PCR+ LM+) were similar (within 16%) for 6/8 proteins in common (MSP1, P12, GAMA, CyRPA, MSP10, P41) (Table 5.4). ARP and MSP5 seropositivity varied more widely with a 32% and 54% difference between the datasets, respectively. The PCR+ LM+ group was primarily asymptomatic in contrast to the acutely febrile Cambodian group, which may contribute to the observed differences.

We next compared the age and infection groups by the breadth of response in each plasma sample. Histograms and density plots for the number of antigens recognized by each age and infection category show a shift in distribution toward greater numbers of antigens recognized in the older age groups and higher parasite densities (Figure 5.13), with mean numbers of antigens recognized as follows: 7.06 in children (5-9 years), 8.08 in adolescents (10-19 years), and 9.65 in adults (20-80 years). The shift between the children and adults was statistically significant (negative binomial regression, $P < 0.001$). The shift between the different infection groups was less pronounced with mean numbers of antigens recognized as follows: 7.56 in uninfected individuals (Not infected), 8.33 in individuals positive by PCR only (PCR+), and 8.90 in individuals positive by both PCR and light microscopy (PCR+ LM+). The shift in numbers of antigens recognized in uninfected and higher-density infection (PCR+ LM+) samples was also statistically significant (negative binomial regression, $P = 0.005$). These results indicated that the breadth of response increased both with age and infection density; shifts were widest between age groups than between infection groups.

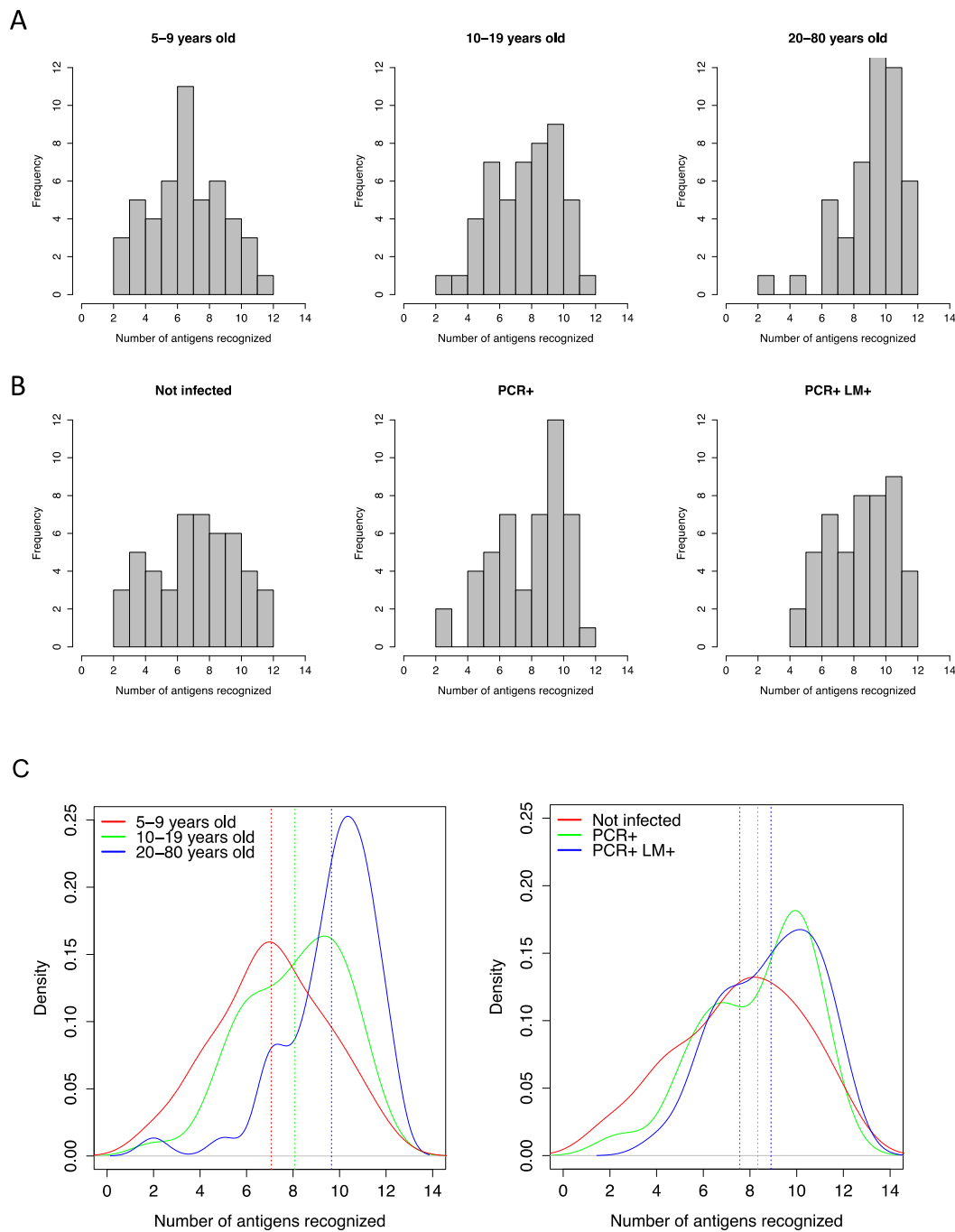


Figure 5.13: Breadth of antigens recognized in SI

(A-B) Histograms of the frequency of number of antigens recognized in SI plasma samples by age group (A) and infection status (B). (C) Density plots of number of antigens recognized by age group (left panel) and infection status (right panel), with mean numbers of antigens recognized by each group denoted by dotted lines. Significant shifts between children (red line, left panel) and adults (blue line, left panel) ($P < 0.001$) and uninfected (red line, right panel) and PCR+ LM+ (blue line, right panel) ($P = 0.005$) were observed. PCR+ = positive by PCR; PCR+LM+ = positive by both PCR and light microscopy. (Negative binomial regression, P values were deemed significant if < 0.05).

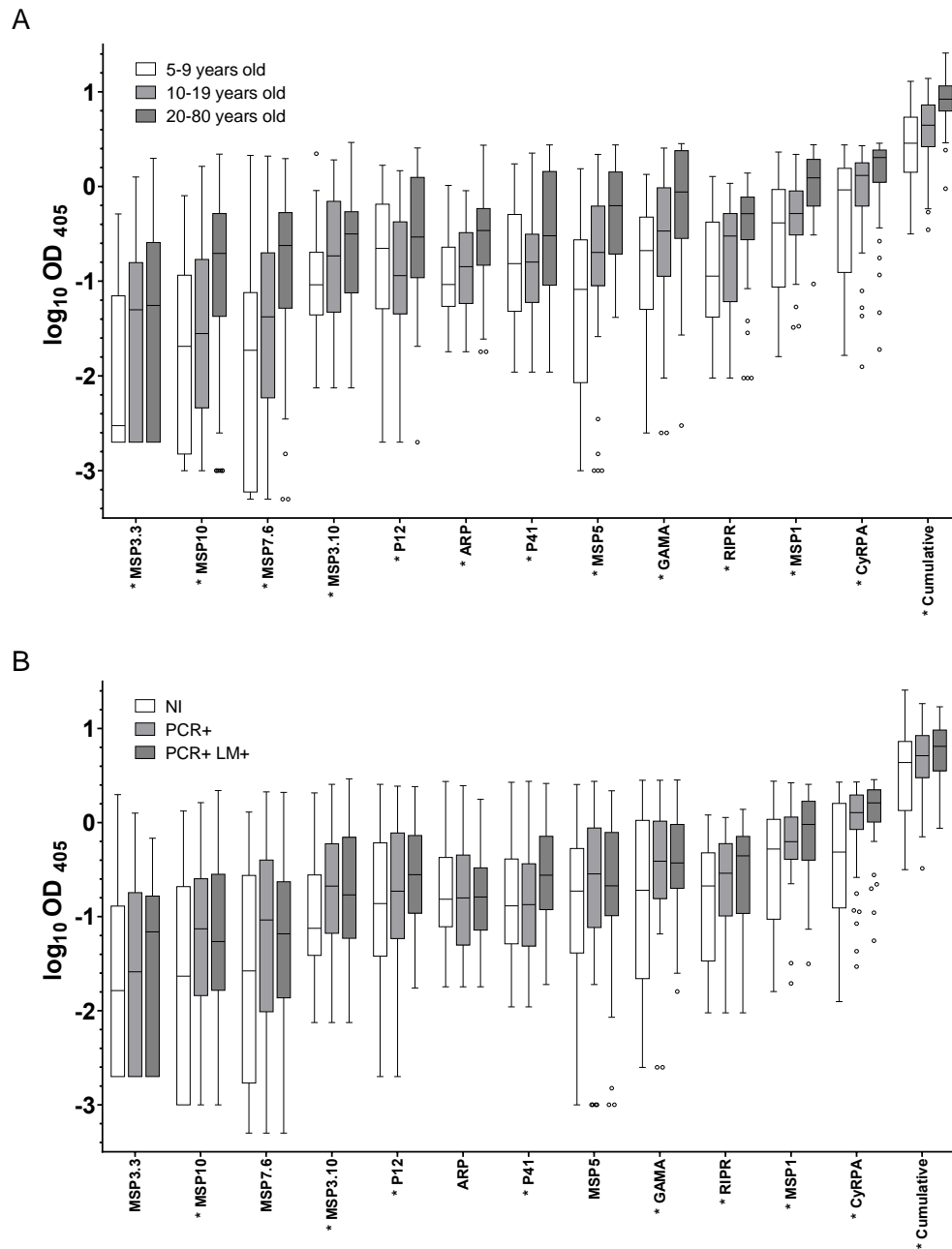


Figure 5.14: Age- and infection-associated increases in IgG in SI

SI plasma samples from 3 age categories (5-9 years, n=48; 10-20 years, n=48; 20-80 years, n=48) with 3 infection statuses (Not infected, NI, n=48; PCR positive, PCR+, n=48; PCR and light microscopy positive, PCR+LM+, n=48) and Australian malaria-naïve sera were screened against 12 *P. vivax* recombinant proteins (MSP3.3, MSP10, MSP7.6, MSP3.10, P12, ARP, P41, MSP5, GAMA, RIPR, MSP1, CyRPA) by ELISA. Log optical densities (OD) at 405 nm are shown. Boxplots show median (horizontal bar), IQR (box), range (whiskers), and outliers (open circles). (A) Age-associated increases in IgG response between children (5-9 years) and adults (20-80 years) were observed for all proteins (*, ANOVA, $P < 0.001-0.027$). (B) Infection-associated increases in IgG responses were observed for a subset of proteins (*, ANOVA $P < 0.001-0.039$). P values were deemed significant if < 0.05 .

We compared the individual responses between the children and adults (Figure 5.14A). Increases in IgG response were strongly associated with age for the cumulative response (combining the IgG response for all proteins) as well as for each individual protein (ANOVA, $P < 0.001$ to 0.027). All P values are reported in Table 5.5.

Table 5.5: P values from ANOVA for SI comprehensive screen

	Children (5-9 years)	Not infected	
	Adults (20-80 years)	PCR+	PCR+ LM+
MSP3.3	0.027	0.581	0.309
MSP10	0.001	0.116	0.039
MSP7.6	<0.001	0.108	0.141
MSP3.10	0.006	0.013	0.032
P12	0.009	0.228	0.020
ARP	<0.001	0.698	0.613
P41	0.012	0.483	0.035
MSP5	<0.001	0.111	0.094
GAMA	<0.001	0.015	0.015
RIPR	0.001	0.175	0.019
MSP1	<0.001	0.344	0.020
CyRPA	<0.001	0.022	<0.001
Cumulative response	<0.001	0.002	<0.001

When comparing the samples by infection status, the cumulative IgG response also increased from non-infected individuals to both low-density infections (PCR+) ($P=0.002$) and high-density infections (PCR+LM+) ($P < 0.001$) (Figure 5.14B, Table 5.5). At the individual protein level, however, only a subset of proteins showed significant associations. In comparing the non-infected group to the low-density PCR+ group, 3 proteins showed significant differences: CyRPA ($P=0.022$), GAMA ($P=0.015$), and MSP3.10 ($P=0.013$). Four additional proteins showed significant increases in antibody levels between non-infected and higher-density infections (PCR+LM+): CyRPA ($P < 0.001$), GAMA ($P = 0.015$), MSP3.10 ($P = 0.032$), P12 ($P=0.020$), P41 ($P=0.035$), MSP1 ($P=0.020$), MSP10 ($P=0.039$), and RIPR ($P=0.019$). Overall, the *P. vivax* recombinant proteins tested are markers of cumulative exposure, showing significant increases in IgG responses with age. Several proteins are also markers of current infection, showing increased IgG responses in individuals with detectable parasites.

5.2.2.3 SI Comprehensive screen multivariate analysis

The SI sample set was analysed by Camila Franca and Ivo Mueller in a multivariate analysis including all available variables to determine if additional associations with antibody responses could be made. Additional variables included region of collection, clinical symptoms, socioeconomic indicators, usage of ITNs, etc. Full details are published in S2 Table in (Franca et al., 2016):

<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4868274/bin/pntd.0004639.s004.xlsx>.

Age- and infection-associated increases in IgG responses remained significant (multivariate ANOVA, $P < 0.001-0.031$), for all comparisons except for 2 individual protein responses within the infection status comparisons; MSP3.10 IgG increases in low-density infections were no longer significant and MSP5 IgG increases in high-density infections were only significantly higher in adults ($P = 0.022$). ITN usage in previous years was the only other variable with significant associations; ITN usage associated with reduced IgG levels for 3 proteins: RIPR ($P = 0.030$), MSP1 ($P = 0.024$), and MSP3.3 ($P = 0.015$). This is potentially due to decreased exposure in these households, which may indicate that IgG responses to these proteins declines in the absence of continued exposure.

5.2.2.4 Correlations with protection in PNG longitudinal cohort study

Our studies in SI demonstrated the utility of screening a panel of proteins, and we wanted to expand the study to investigate whether IgG responses to the proteins correlated with protection from clinical disease. A longitudinal cohort study in PNG led by Ivo Mueller (Lin et al., 2010) provided an ideal sample collection to address this question. Based on the results from the SI plasma screens, we prioritized 6 proteins to be screened in a high-throughput Luminex screening assay. Camila Franca conjugated His-purified *P. vivax* recombinant proteins (P12, P41, CyRPA, ARP, GAMA, PVX_081550) to Luminex beads in order to screen 230 PNG plasma samples, the collaborative results of which were published along with the SI results (Franca et al., 2016). Enrolled children, aged 1-3 years, provided blood samples every 2 weeks for 16 months, and any detected parasites were genotyped to estimate the number of genetically-distinct *P. vivax* infections acquired over time (termed the molecular force of blood-stage infection, molFOB).

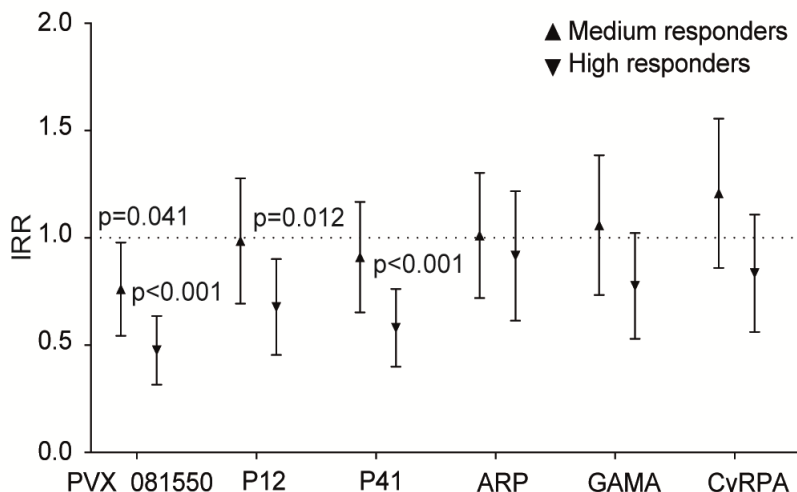


Figure 5.15: High IgG responses to PVX_081550, P12, and P41 are associated with reduced incidence of clinical disease

Incidence rate ratios and 95% confidence intervals, adjusted for exposure (molFOB), age, season, and village of residency are shown. Clinical malaria was defined as axillary temperature $\geq 37.5^{\circ}\text{C}$ or history of fever in the preceding 48 hours with a current *P. vivax* parasitemia >500 parasites/ μL . (Negative binomial GEE models). Figure produced by Camila Franca and reprinted from (Franca et al., 2016) under the Creative Commons Attribution (CC BY) license.

Table 5.6: Association between levels of IgG to *P. vivax* merozoite proteins and protection against clinical malaria in PNG children*

Antigen	uIRR	95%CI		P value	aIRR*	95%CI		P value
PVX_081550 M	0.76	0.54	1.05	0.10	0.74	0.55	0.99	0.041
PVX_081550 H	0.41	0.29	0.60	<0.001	0.46	0.33	0.64	<0.001
ARP M	0.93	0.66	1.32	0.68	0.98	0.73	1.32	0.91
ARP H	1.00	0.69	1.46	0.98	0.88	0.63	1.23	0.47
GAMA M	1.12	0.80	1.57	0.51	1.03	0.75	1.40	0.87
GAMA H	0.82	0.55	1.23	0.34	0.75	0.54	1.04	0.08
P41 M	0.96	0.68	1.36	0.83	0.89	0.67	1.18	0.41
P41 H	0.63	0.43	0.93	0.019	0.56	0.41	0.77	<0.001
P12 M	1.05	0.75	1.47	0.79	0.96	0.71	1.29	0.77
P12 H	0.69	0.47	1.02	0.06	0.65	0.47	0.91	0.012
CyRPA M	1.06	0.76	1.46	0.74	1.17	0.88	1.57	0.28
CyRPA H	0.88	0.59	1.30	0.52	0.81	0.58	1.12	0.20

* With parasite density $>500/\mu\text{L}$ of blood; table by Camila Franca and reprinted from (Franca et al., 2016) under the Creative Commons Attribution (CC BY) license.

The study produced several important results. The data show age-associated increases in IgG response for ARP, CyRPA, and PVX_081550 ($r=0.15-0.25$; $P=0.001-0.027$), indicating that these increases were generalizable to 2 different populations living in areas of moderate and high transmission. Infection was associated with increased IgG levels to CyRPA ($P<0.001$), P12 ($P<0.001$), P41 ($P=0.001$), and PVX_081550 ($P=0.001$), which was also observed in SI (except for PVX_081550, which was not included in the SI comprehensive screen).

Children in the study experienced an average of 1.11 malaria episodes during the follow-up period with *P. vivax* >500 parasites/ μl /year at risk. After adjusting for confounders, high IgG responses to 2 proteins, P12 and P41, and both medium and high IgG responses to PVX_081550 were strongly associated with a reduced incidence of clinical disease (IRR 0.46-0.74; $P<0.001-0.041$) (Figure 5.15, Table 5.6). IgG responses to P12, P41, and PVX_081550 were significantly correlated ($r=0.34-0.66$; $P<0.001$). This may indicate that IgG responses against the 3 proteins are co-acquired. After performing a multivariate analysis, high levels to PVX_081550 remained significantly associated with protection from clinical disease (IRR_H 0.54; $P=0.001$). This result indicates PVX_081550 may be a target of NAI, a marker of a person's immune status, or both.

5.3 Discussion

In this chapter, I presented results demonstrating that the *P. vivax* recombinant protein library is immunoreactive and properly folded. I also presented immunoepidemiological studies from 3 *P. vivax*-endemic areas (Cambodia, SI, and PNG) aimed at furthering our understanding of immune responses during and after *P. vivax* infection.

5.3.1 Key findings in Cambodian plasma screens

Pilot ELISAs that screened 42 individual plasma samples from *P. vivax*-infected patients against 6 proteins suggested that our proteins were immunoreactive, though with varying breadth and levels; P12 showed the greatest initial response with 98% of samples reactive above 2 SD above the mean of American *P. vivax*-naïve controls.

However, limited conclusions could be drawn from this, as the plasma set used was pre-selected based on reactivity to recombinant *P. vivax* MSP1 (produced in the same mammalian expression system, but with a different plasmid backbone), which may not

represent the immunoreactivity of the *P. vivax*-exposed population. This may also have selected for plasma samples that were also non-specifically reactive, as several plasma showed reactivity to Cd4d3+d4 tag more than 5x higher than the reactivity in the American *P. vivax*-naïve controls. Overall, the pilot results established that our proteins were immunoreactive and that a 1:600 plasma dilution was optimal for screening plasma samples against the full *P. vivax* recombinant protein library.

In order to fully explore whether our library proteins were properly folded, contained conformational epitopes, and were targets for naturally-acquired humoral immunity, I screened 34 recombinant *P. vivax* proteins against pooled plasma from 14 Cambodian patients with acute vivax malaria. Of the 34 proteins screened, 27 showed at least a 2-fold change in IgG reactivity between naïve sera and *P. vivax*-exposed plasma samples. Our results aligned well with a *P. vivax* seroreactivity screen in Korean patients (Chen et al., 2010), which detected IgG reactivity to 18 full-length proteins and protein fragments, all produced in the wheat germ cell-free system. Of these 18 proteins, 7 (MSP1, MSP3.3, MSP10, P12, P41, ARP, PVX_081550) were represented in our library and all 7 showed at least a 3-fold change in IgG reactivity between naïve sera and *P. vivax*-exposed plasma samples in our Cambodian screen. This establishes that these antigens are common targets of immunoreactivity across different transmission regions.

Conformational epitopes were present in 18/34 antigens, as indicated by >20% reduction in IgG reactivity after heat treatment. IgG reactivity to 12 of these 18 antigens was predominantly conformation-specific, as indicated by >50% reduction in IgG reactivity after heat treatment. These findings suggest that our library proteins are properly folded. The tertiary structure of MSP3 family proteins is known to be recalcitrant to heat denaturation, so the limited change in immunoreactivity following heat treatment of these antigens does not necessarily indicate a lack of tertiary conformation. In addition to MSP3 proteins, several other proteins also show an absent or weak heat-sensitive response, including those with higher overall responses, such as MSP4 and MSP10, and several with weak overall responses, such as P92 and PVX_110965. The lack of heat-sensitive responses may be due to 1 of 3 factors, each of which might be applicable for a given protein. First, the protein may not denature or remain denatured under the heat treatment we used (80°C for 10 min), which is likely the case for MSP3, which is known to be highly stable. Follow-up experiments testing other techniques, such as chemical denaturation or more extreme heating conditions, could clarify this point. Second, most of

the primary antibody response may target linear epitopes within the protein that are not affected by denaturation. At least 2 of the proteins where responses were not affected by heat treatment, MSP4 and ARP, contain regions of low complexity that could fit with such a model. Third, the lack of a change in response could indicate an issue with protein quality. Although all of these proteins are visible by Western blotting (Figure 4.6), P92 and PVX_001015 are very faint. Additional experiments are needed to fully explore these possibilities, though the fact that the majority of antigens displayed a heat-sensitive response suggests that most library proteins are properly folded. As well as suggesting that the full-length ectodomains contain folded epitopes, these data also indicate that humans naturally acquire IgG responses to multiple *P. vivax* proteins, thus supporting their further exploration as candidate vaccine antigens.

The Cambodian reactivity screen using individual plasma against a subset of 11 proteins demonstrated that reactivity to the *P. vivax* proteins varied widely from 35% for ARP and MSP10 to 100% for MSP1 and GAMA. This variation is similarly observed in *P. falciparum* where IgG responses to merozoite antigens can vary substantially between individuals (Osier et al., 2014). P12 was reactive in 92% of Cambodian samples; *P. falciparum* P12 has shown seroreactivity in 96% of 286 Kenyan individuals (Osier et al., 2014). IgG responses to P12 may be a useful marker of infection with *P. vivax*, *P. falciparum*, or both in broader epidemiological investigations. Recombinant *P. vivax* MSP1 and GAMA were recognized in 100% of Cambodian samples, but recombinant P12 (61 kDa) is smaller than GAMA (103 kDa) and MSP1 (215kDa), has high expression in the HEK293E cell system and is stable for longer periods at 4°C, potentially making it easier to use in future sero-surveillance studies.

5.3.2 Key findings in SI and PNG populations

The SI reactivity screening demonstrated that the *P. vivax* recombinant protein library was largely immunogenic in a second *P. vivax*-endemic country. Over 75% of proteins were recognized in each country with 29/34 proteins recognized in SI compared to 27/34 in Cambodia. This represents the largest number to date of immunogenic responses to a *P. vivax* recombinant protein library, and supports the utility of the *P. vivax* recombinant library in future worldwide immunoepidemiological studies.

We performed an expanded screen for 12 highly immunogenic proteins identified in the SI reactivity screen (MSP3.3, MSP10, MSP7.6, MSP3.10, P12, ARP, P41, MSP5,

GAMA, RIPR, MSP1, CyRPA). Several proteins were immunogenic in over 90% of samples (MSP1, P12, GAMA, MSP3.10, RIPR). Reactivity in over 90% of samples was similarly found for MSP1, GAMA, and P12 in the Cambodian reactivity screen (section 5.2.1.3), indicating any of these 3 proteins may be useful as markers of *P. vivax* exposure. The comprehensive screen also confirmed that IgG increased both with age (12/12 proteins tested) and current infection (7/12 proteins tested). IgG increased more strongly with age than with infection, indicating that cumulative past exposure to *P. vivax* malaria was a stronger driver of immune response than current infection. These results, along with the fact that low-density asymptomatic parasitemias are common in SI indicate that long-lasting stable antibody responses persist even in an area of rapidly declining transmission.

The Cambodian and SI screens enabled us to prioritize 6 proteins (P12, P41, PVX_081550, ARP, GAMA, CyRPA) for screening against a longitudinal cohort of young children in a high *P. vivax* transmission setting in PNG. The PNG study found correlations with protection from disease for 3 proteins (P12, P41, PVX_081550), with a similar reduction in risk of disease as what was found for high antibody titers to *P. vivax* MSP3.10 and MSP9 (Stanisic et al., 2013). As with *P. falciparum* P12 and P41, *P. vivax* P12 and P41 form a heterodimer, though with a much higher affinity than with *P. falciparum* (discussed in detail in Chapter 4). The function of the complex, however, is not known. Antibodies directed individually against *P. falciparum* P12 and P41 did not inhibit parasite invasion or growth *in vitro* (Taechalertpaisarn et al., 2012). Functional work in *P. vivax* field isolates will be needed to establish whether the same is true for this species. *P. falciparum* P41 was associated with clinical protection in Kenya and PNG (Richards et al., 2013, Osier et al., 2014).

The strongest association with protection was for PVX_081550, for which relatively little is known. It was recently characterized as a StAR-related lipid transfer protein in *P. falciparum* (van Ooij et al., 2013). The protein localizes to the parasitophorous vacuole (PV) with some additional localization evidence in the apical region of merozoites. van Ooij et al. hypothesize the protein is involved with PV formation in newly-invaded erythrocytes (van Ooij et al., 2013). The protein was highly immunogenic in the Cambodian and SI screens, and this finding is also supported by a *P. falciparum* screen (Fan et al., 2013). Whether IgG responses directed against the protein actually inhibit invasion or are a by-product of exposure to the immune system after schizont rupture are

unknown. In the latter case, responses to the protein would serve as a useful marker of an individual's immune status. Both the *P. falciparum* and *P. vivax* orthologs appear to be polymorphic with nonsynonymous/synonymous SNP ratios of 1.9-2.3 (plasmodb.org), but whether this indicates any selective pressure is also not known. Follow-up studies involving both species are needed to determine the protein's function and potential as a vaccine candidate.

5.3.3 Limitations and future work

No differences were detected between acute and convalescent plasma samples, but additional experiments are needed to determine if this result is conclusive. A screening of a subset of proteins in more samples will clarify if small IgG changes are occurring in this period, and plasma pairs with the widest separation in dates (28 days) could be prioritized in order to maximize the possibility of observing changes.

The reactivity cut-offs for the SI comprehensive screening were set at 2SD above the mean of multiple readings from a pool of *P. vivax*-naïve Australian sera. While multiple readings were useful for obtaining an accurate measurement of the pool, the standard practice is to measure responses to individual control sera when calculating a cut-off. This potentially lowers the reactivity threshold for the assay, as individual sera may have had more variable responses. The fact that the Cambodian reactivity data, which included 24 individual malaria-naïve US sera samples, shows similar (within 16%) reactivity for most antigens (MSP5 and ARP as the exceptions), suggests that the threshold set using pooled sera was not a significant underestimation of background reactivity in this case. Ideally, all future screenings will include individual *P. vivax*-naïve sera controls.

Strong age-associated IgG increases in SI were detected even as *P. vivax* malaria continues to decline in this study site. It would be useful to follow up the 2011 cross-sectional study in several years to determine whether IgG responses continue to persist without continued exposure, or with greatly reduced exposure. Also, the screening of recombinant *P. vivax* proteins in the SI patient plasma used nearly the full library, with 34/39 proteins screened. The PNG plasma screen, however, included only 6 proteins, as the Luminex screening method required purified proteins. Ideally, this screen would have included additional proteins, but we were limited by both manpower and shared laboratory facilities to producing purified proteins in sufficient quantities for only a

subset. Correlations with protection were found in this study, and repeating the screen with additional purified proteins may expand this list.

The immunoepidemiological screens supported high immunogenicity for several proteins and protection from disease for 3 proteins in particular. From these data, we can prioritize a list of proteins for further functional studies. Such studies would include generating polyclonal antibodies against a subset of the recombinant *P. vivax* proteins for use both in immunofluorescence microscopy studies to establish protein locations within the merozoite (i.e., surface, apical, microneme, etc.), and *P. vivax ex vivo* reticulocyte invasion assays (Russell et al., 2011) to test whether they are able to block invasion. Antibodies could be tested both individually and in combination; such studies are planned to proceed later in 2016.

5.4 Conclusion

To characterize naturally-acquired immune responses in *P. vivax* malaria, we screened the *P. vivax* recombinant protein library against plasma collections from 3 *P. vivax*-endemic countries: Cambodia (low transmission), SI (moderate transmission), and PNG (high transmission). The results from Cambodia indicated that nearly all of the recombinant *P. vivax* full-length ectodomains are immunogenic, over half contain conformational epitopes, and humans naturally acquire IgG responses to multiple *P. vivax* proteins. SI screens supported nearly the entire library as being immunogenic in a second transmission setting, and found strong age-associated and less pronounced infection-associated increases in IgG, indicating that past-cumulative exposure is a stronger driver of IgG responses compared to recent exposure. The PNG screens found that responses to 3 proteins correlated with protection from disease, including a hypothetical protein for which little is known. The overall results support the utility of using a panel of proteins to make systematic comparisons to prioritize candidates for vaccine studies, and identify clear candidates for subsequent functional study.

6 DISCUSSION

6.1 Key remaining challenges in *P. vivax* research field

P. vivax malaria remains a risk for over a quarter of the world's population and a significant global morbidity burden despite the gains in controlling malaria achieved over the last decade. Research into *P. vivax* lags significantly behind that of its deadlier cousin, *P. falciparum*, not least in the area of vaccine research and development. A number of studies suggest that the burden of *P. vivax* is significantly underestimated and a comprehensive strategy to combat both species will be required to achieve the ambitious goal of a malaria-free world.

6.1.1 *In vitro* culture

Despite extensive efforts by many groups, *P. vivax* still lacks a reliable and widely usable *in vitro* blood stage culture system [reviewed in (Noulin et al., 2013)]. This remains the largest barrier to the study of fundamental *P. vivax* biology and the search for better vaccine candidates, at least for erythrocytic stages. While this challenge was not the focus of this work, it remains the continued focus of several other groups using either reticulocytes enriched from cord blood, whole blood or derived from hematopoietic stem-cells (Roobsoong et al., 2015) (Panichakul et al., 2007). In the absence of a robust and widely useable culture system for *P. vivax* itself, there is promising potential for harnessing the more closely-related *P. knowlesi*, which was recently adapted for *in vitro*

culture in human RBCs (Moon et al., 2013, Lim et al., 2013) and for which early transfection work shows positive results. It may be possible to study *P. vivax* genes *in vitro* through allelic replacement into the *P. knowlesi* parasites. *P. knowlesi* shares 4732/5188 (91%) one-to-one orthologous genes with *P. vivax* [(Pain et al., 2008) and OrthoMCL], and the two species are much more phylogenetically closely related to each other than to *P. falciparum* (Carlton et al., 2008). Like *P. vivax*, *P. knowlesi* uses DARC as a receptor for human RBC invasion (Horuk et al., 1993, Singh et al., 2002). It also appears to prefer invading younger RBCs in *in vitro* cultures, though the adaptation to continuous culture has resulted in invasion of a wider range of erythrocyte ages (Lim et al., 2013, Moon et al., 2013, Gruring et al., 2014). This adaptation to erythrocytes of all ages may explain the dangerously high parasitemias observed in some human infections (Lim et al., 2013). Thus there is potential for *P. knowlesi* to serve as an *in vitro* model for *P. vivax* asexual stages, though with potentially divergent and/or adaptable invasion pathways between the species in addition to a 24-hour replication cycle compared to the 48-hour replication cycle in *P. vivax*.

6.1.2 Hypnozoites

P. vivax transmission remains stubbornly entrenched in areas even as prevalence rates of *P. falciparum* decline. This persistence has been linked to a variety of factors including the early appearance and high infectivity of *P. vivax* gametocytes and the biting habits of common *P. vivax* vectors (Mueller et al., 2009a). However, the presence of dormant hypnozoites which leads to relapsing *P. vivax* infections is likely a major cause of continuing transmission (White and Imwong, 2012) with some estimates suggesting that over 50% of infections result from emerging hypnozoites (Betuela et al., 2012). This stage represents a major and so far relatively unexplored roadblock to *P. vivax* elimination. There is only one currently approved drug, primaquine, to remove or kill hypnozoites; its 14-day treatment regimen and contraindication for pregnant women and/or those with G6PD deficiency make it inconvenient at best and unsafe at worst in many of the populations where *P. vivax* is common. A new drug aiming to replace primaquine, tafenoquine, is of the same drug class, will reduce the treatment regimen to a single dose and is currently in Phase III clinical trials (Llanos-Cuentas et al., 2014). If tafenoquine makes it to the clinic, the reduced dose will be a significant improvement for patient compliance, but the drug is still contraindicated for patients with G6PD

deficiency. Lacking a cost-effective, point-of-care test for G6PD deficiency means that neither drug is likely to be used in many transmission regions, such as Southeast Asia where G6PD deficiency is common (Nkhoma et al., 2009). The cellular, molecular and biochemical features of hypnozoites, such as how dormancy is maintained, how infected cells remain silent to the immune system, and the cues which induce activation of hypnozoites are all mysteries. A much more detailed understanding of the hypnozoite life-stage is essential to allow the targeted development of drugs aimed at radical cure of *P. vivax*. Recent work with human-liver chimeric mice showed successful infection with *P. falciparum* sporozoites and subsequent maturation in infected hepatocytes (Sacci et al., 2006, Vaughan et al., 2012). The recent adaptation of this model for *P. vivax* may give a significant boost to the study of hypnozoites, radical cure drug development and vaccines aimed at the pre-erythrocytic stages (Mikolajczak et al., 2015).

6.1.3 *P. vivax* invasion

As discussed throughout this work, the understanding of the interactions required for *P. vivax* invasion lags significantly behind that of *P. falciparum*. While one interaction is known, that between *P. vivax* DBP and human erythrocyte receptor DARC, the human receptor for reticulocyte tropism and any other interactions involved in the invasion cascade are still unknown. The discovery of such ligand-receptor interactions would most certainly widen the field of vaccine candidates to consider for *P. vivax*, for which PvDBP remains the strongest asexual-stage candidate.

The RBPs, specifically PvRBP1b and PvRBP2c, have been shown to have reticulocyte-binding capabilities but their binding partners have yet to be elucidated (Galinski et al., 1992). Whether *P. vivax* RBPs perform similarly redundant roles to the *P. falciparum* EBAs and RHs is also unknown but would not be surprising from the 9 to 10-member group (Carlton et al., 2008). The crystal structure of the erythrocyte binding domain from PvRBP2a (the first RBP structure reported) displays structural similarity to PfRH5, suggesting that the main structure for erythrocyte binding is conserved across species (Gruszczyk et al., 2016). However, the erythrocyte-binding domain was highly polymorphic among field isolates and other surface properties were distinct between PfRH5 and PvRBP2a suggesting that these differences are specific for different ligand-receptor interactions between species (Gruszczyk et al., 2016).

The PfrH5-basigin interaction in *P. falciparum* is of great interest as it appears to represent an essential interaction required to complete erythrocyte invasion, and PfrH5 is under active development as a vaccine candidate (Crosnier et al., 2011, Reddy et al., 2014, Douglas et al., 2015). PfrH5 has been found to form a complex with PfrIPR and PfcyRPA, both of which have one-to-one orthologs in *P. vivax* (Reddy et al., 2015). This leads to the question of whether such a complex also exists in *P. vivax*, and if so, which protein is performing the function of PfrH5. It is interesting to note that PfrH5 and PfrIPR appear to have lower (40-50%) seropositivity rates than most other merozoite surface/invasion proteins (>75%) potentially indicating they are exposed to the immune system for a relatively limited time, or are fundamentally less immunogenic (Richards et al., 2013). This is in contrast to PvCyRPA and PvRIPR, which were among the most immunogenic antigens in several of the screens performed in this work, with 57-79% and 92-96% of patient samples recognizing these antigens respectively (Table 5.3, Table 5.4). This may indicate more exposure to the immune system during *P. vivax* infection, thus suggesting a potential different function of these proteins in the *P. vivax* invasion process.

6.1.3.1 New methods for detecting *P. vivax* erythrocyte interactions

A major goal of this work was aimed at expanding the understanding of proteins involved in *P. vivax* reticulocyte invasion. At the start of this work, the AVEXIS assay had recently uncovered the PfrH5-basigin interaction (Crosnier et al., 2011), and appeared to have great potential for uncovering ligand-receptor interactions in *P. vivax*. A definite strength of AVEXIS in the detection of cell-surface interactions is that the assay produces very few false positives, which is useful for limiting the wasting of resources for downstream functional work. However, it is also known to produce false negatives potentially due to ectodomain misfolding or incorrect glycosylation, and the fact that it can only be applied to surface proteins with extensive ectodomains, meaning some classes such as multi-transmembrane proteins, are omitted from AVEXIS screens.

Several strategies can be used to overcome these limitations. One previously applied strategy is pull-down experiments, where receptors are enriched by immunoprecipitating their binding partner, although this approach, with its multiple wash steps, is ill suited for potentially low-affinity cell-surface interactions. Newer high throughput protein interaction screening strategies are under active development in the Wright and Rayner laboratories as well as other external groups. One promising new approach involves

binding recombinant proteins to cancer cell lines, which often up-regulate a number of surface proteins, including erythroid specific ones. Once binding to a cell line has been identified for a given ligand, CRISPR is then used to create a knock-out library of cells within that line. Sorting by binding and flow cytometry identifies cells that have lost the ability to bind the *Plasmodium* ligand, and gRNA sequencing from the sorted cells identifies which genes have been targeted in the cells which do not bind. This approach has already recapitulated the RH5-basigin interaction (S. Sharma, data not yet published) and may lead to the uncovering of additional parasite ligand-receptor binding events. An alternative approach is under way to express hundreds of receptors in arrays on slides to then screen with recombinant proteins thus overcoming the constraint of AVExis to only include single ectodomains in screening. This approach was recently applied to identify endothelial protein C receptor (EPCR) as the receptor for several PfEMP1 subtypes associated with severe childhood *P. falciparum* malaria (Turner et al., 2013). These two examples demonstrate that high-throughput protein interaction screening methods are under active development and will hopefully bridge the “invasion” knowledge gap in the *P. vivax* research field as well as continuing to expand the known ligand-receptor list in *P. falciparum*, all of which may lead to stronger vaccine candidates than are currently in the development pipeline.

6.1.4 *P. vivax* infections of Duffy-negative individuals

Individuals with the Duffy-negative phenotype were thought to be completely protected from infection with *P. vivax* due to the essential interaction between *P. vivax* DBP and DARC on human erythrocytes. Numerous publications over the last ten years have challenged this assumption with reports of *P. vivax* isolates which are capable of infecting Duffy-negative people throughout Africa (Menard et al., 2010, Mendes et al., 2011, Ngassa Mbenda and Das, 2014, Ryan et al., 2006, Woldearegai et al., 2013). It is not known whether these isolates represent a new and emerging adaptation or a minor but always present capability that had previously been overlooked. If the former is true, the phenomenon may have huge implications for control throughout Africa, where *P. vivax* became absent with the spread of the Duffy-negative phenotype. Understanding the genetic and/or transcriptional profile of such parasites through DNA and RNA sequencing will be essential to addressing these questions. While clinical isolates are still necessary for such studies, improvements in next-generation sequencing technology,

make using patient samples (even with very low parasite densities) more feasible. Studies, like ours, aimed at better understanding the transcription prior to invasion and the similarities and differences driving variation between *P. vivax* isolates, particularly those leading to severe disease, are likely to be on the rise over the next few years. We plan several long-term collaborations for analysing the genomes and transcription of field isolates from Ethiopia, where recent *P. vivax* infection in Duffy-negative individuals has been detected (Woldearegai et al., 2013).

6.1.5 *P. vivax* vaccine development

As discussed throughout this work *P. vivax* is both difficult to study and frequently underestimated in terms of global morbidity, with numerous global reports of severe disease (Price et al., 2007). Both factors have led to the skewed research funding vs the more deadly *P. falciparum* and a relative paucity of active vaccine candidates. A recent review of leading *P. vivax* vaccine candidates shows that while pre-erythrocytic stages, asexual and transmission-blocking vaccine candidates are all under active development, no candidate has progressed further than the Phase I/IIa clinical trial for a PvCSP-based candidate which provided no sterile protection [reviewed in (Mueller et al., 2015)]. The leading asexual candidate in pre-clinical development phase remains *P. vivax* DBP. DBP contains conserved erythrocyte-binding epitopes and has demonstrated antibody-associated protection from clinical disease, both of which support its continued development as a vaccine target. However, the reliance on a single antigen may have drawbacks including limited efficacy (as has been shown in the single-antigen PfCSP-based vaccine) and the potential for being ineffective against parasites which infect Duffy-negative individuals, although as noted above it remains to be seen whether *P. vivax* infection of Duffy-negative individuals are mediated through alternative invasion pathways or perhaps still utilize DBP binding to alternative receptors. Overall, as suggested by Mueller *et al.*, a widely effective *P. vivax* vaccine may require the targeting of multiple critical *P. vivax* antigens from both asexual, sexual and pre-erythrocytic stages (Mueller et al., 2015). The initial RNA-Seq data generated in this work hinted at variable expression for a number of host-interaction genes, including the RBPs, and expanding the number of samples will enable us to establish whether this pattern is the norm. This may provide insight into strategies of the parasite for evading the host-immune response and potentially impact vaccine design, as genes with variable

expression may represent some functional redundancy and should possibly be avoided as vaccine candidates.

6.1.5.1 High-throughput immunoepidemiological studies

Given the challenges of *P. vivax* lab-based research, immunoepidemiological studies remain a key tool for identifying targets of naturally-acquired immunity (NAI) and hence prioritization of vaccine candidates. Screening individual proteins against individual sera samples by ELISA is a reliable and well-understood assay and has formed the basis of most immunoepidemiological studies. However, the process remains labor intensive, which frequently limits the scale of proteins or sera samples for testing. Most seroreactivity studies use purified proteins for screening, adding an additional limitation. We overcame this limitation by producing biotinylated proteins, which could effectively be purified from complex mixtures by coating on streptavidin plates prior to screening with sera samples, and this enabled the simultaneous screening of a panel of 34 recombinant *P. vivax* proteins against plasma samples from Cambodia and SI. This method was also used successfully for studying the development of NAI to *P. falciparum* (Osier et al., 2014). We expect the use of this library of *P. vivax* recombinant proteins, without the need for prior rigorous purification, will continue to be a benefit for drawing systematic conclusions in immunoepidemiological studies in other *P. vivax*-endemic settings outside of Southeast Asia and Oceania, such as South America and India, where *P. vivax* malaria remains a significant public health concern.

The screening technology itself is also changing, and immunoepidemiological studies using protein arrays are beginning to increase in number and usefulness. Protocols for array production are in development in the Wright laboratory, with spotting/arraying of individually expressed proteins on glass slides allowing for systematic screening of many proteins in parallel and the potential for long term, stable storage (in glycerol). This may enable the repeated screening of identical sets of proteins. Currently for *P. vivax*, 3 protein array studies have been used for assessing reactivity (Chen et al., 2010, Lu et al., 2014, Molina et al., 2012). An *E. coli*-expression based array was also recently used to study reactivity in both symptomatic and asymptomatic *P. vivax* infections, showing that symptomatic children carried fewer antibodies (Finney et al., 2014).

Another higher throughput alternative is the Luminex assay, which while requiring purified proteins, enables the screening of hundreds of sera samples in a flow cytometry-

based assay. Because individual sera samples can be screened with several proteins simultaneously, the assay requires less serum volume and has the potential for extending the number of studies accomplishable with finite and precious serum collections. In the context of this work, increasing the number of purified proteins screened by Luminex in the PNG longitudinal cohort study would provide further insight into the development of NAI in a high-transmission *P. vivax* setting, and potentially expand the list of antigens to consider as vaccine candidates. It would also help to further address the hypothesis that the breadth of response is a crucial element of protection from clinical *P. vivax* disease as has been found to be the case for immunity to *P. falciparum* (Osier et al., 2014), and was indicated by screening of only 6 *P. vivax* antigens in this work (Franca et al., 2016).

Cross-protection (or lack thereof) between species will also be important to evaluate in a more systematic way; this can be facilitated by both the *P. vivax* and *P. falciparum* recombinant protein libraries assembled in this and prior work in the Wright and Rayner laboratories. Studies in neurosyphilis patients and the lack of cross-protection between *P. vivax* and *P. falciparum* in PNG suggest that while the development of immunity can be strain transcending, it does not appear to be species transcending (Doolan et al., 2009, Franca et al., 2016). However, more systematic screening using both *P. falciparum* and *P. vivax* merozoite antigens would be useful for fully evaluating this. Any antigens correlating with cross-protection would be prime candidates as vaccine targets, as any mass-administered vaccine must ultimately include efficacious components for both *P. falciparum* and *P. vivax*. Otherwise, we run the risk of reducing *P. falciparum* only to have *P. vivax* continue to persist in causing millions of cases per year and significant global morbidity.

6.1.5.2 Functional validation of *P. vivax* vaccine candidates

The functional and immunoepidemiological studies in Chapters 4 and 5 identified several parasite protein-protein interactions, several highly immunogenic proteins and IgG responses to 3 proteins that correlated with protection. However, the full potential of these candidates as vaccine targets will require additional functional studies. Planning for this work is already underway, and we have generated polyclonal antibodies against several recombinant *P. vivax* proteins, including those associated with protection. We will first confirm that antibodies to the recombinant merozoite proteins react to the corresponding native *P. vivax* merozoite proteins through western/immunoblotting

techniques. Immunofluorescence microscopy will be used to confirm that antibodies bind to expected locations in the merozoite (i.e., surface, apical, microneme, etc.), using *P. vivax*-infected reticulocytes collected from volunteers in the field. This will provide evidence as to whether interaction pairs detected *in vitro* (by AVEXIS and SPR) potentially interact *in vivo* by co-localizing in the parasite.

We also plan to make use of a published *P. vivax* invasion-blocking assay (Russell et al., 2011) to test whether antibodies directed toward the *P. vivax* merozoite antigens significantly inhibit *P. vivax* invasion of reticulocytes. The polyclonal antibodies can be used to potentially disrupt the merozoite protein-protein interactions detected and/or any association with as yet undetected reticulocyte receptors. We will additionally use the candidate recombinant proteins to compete for reticulocyte receptors involved in merozoite invasion. We will attempt to block invasion by targeting single proteins and combinations, as efficient inhibition of *P. vivax* invasion may require a multi-target approach.

6.2 *P. vivax* research summary

This work was specifically aimed at addressing the remaining knowledge gap in understanding the proteins involved in the *P. vivax* merozoite invasion of erythrocytes through the use of transcriptomics, recombinant protein expression and immunoepidemiological studies (Figure 6.1). The ultimate goal of the project was to develop a prioritized list of *P. vivax* proteins for future study as vaccine candidates.

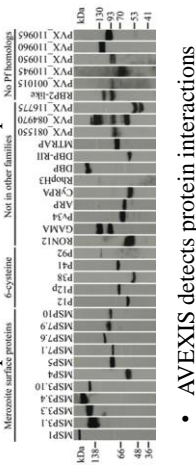
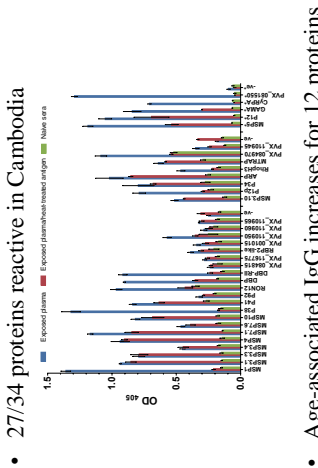
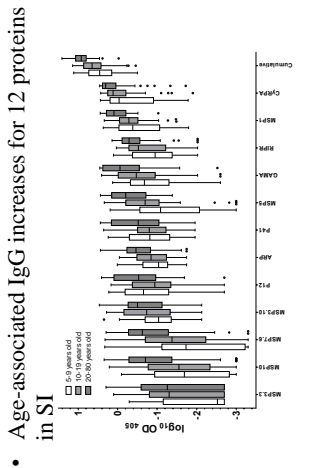
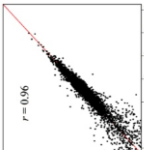
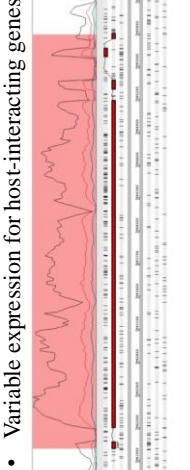
Chapter 3 <i>P. vivax</i> transcriptomics	Chapter 4 <i>P. vivax</i> protein expression	Chapter 5 <i>P. vivax</i> immunoepidemiology
<p>Aim: Study transcription in schizont-stage <i>P. vivax</i> parasites</p> <p>Approaches:</p> <ul style="list-style-type: none"> • Short term <i>ex-vivo</i> culture of <i>P. vivax</i> clinical isolates • Use RNA-seq to produce unbiased transcript abundance data <p>Results:</p> <ul style="list-style-type: none"> • Extracted high quality and quantity RNA from 4 schizont –enriched <i>P. vivax</i> clinical isolates • Generated over 150x average coverage using strand-specific Illumina Hi-Seq libraries • Corrected over 300 gene models, uncovered 20 novel transcripts and expression data for over 400 additional genes 	<p>Aim: Produce a library of <i>P. vivax</i> merozoite proteins and investigate function</p> <p>Approaches:</p> <ul style="list-style-type: none"> • Combine RNA-seq, microarray data and homology with <i>P. falciparum</i> • Utilize mammalian expression system • Assess erythrocyte binding by flow cytometry • High-throughput interaction screens (AVEXIS) <p>Results:</p> <ul style="list-style-type: none"> • Detected parasite protein-protein interactions: predicted (P12-P41), novel (MSP7.1-MSP3.10, P12-PVX_110945) • Confirmed interactions using SPR • Expressed 37/39 candidate proteins  <p>AVEXIS detects protein interactions</p>	<p>Aim: Investigate <i>P. vivax</i> merozoite protein immunoreactivity</p> <p>Approaches:</p> <ul style="list-style-type: none"> • Assess IgG responses from patient plasma to recombinant <i>P. vivax</i> proteins using ELISA and Luminex <p>Results:</p> <ul style="list-style-type: none"> • <i>P. vivax</i> recombinant proteins reactive in three <i>P. vivax</i>-endemic settings • IgG responses to P12, P41, PVX_081550 correlated with protection in PNG • 27/34 proteins reactive in Cambodia  <p>Age-associated IgG increases for 12 proteins in SI</p> 
<p>• Schizont-stage expression</p> <p>• Expression highly correlated</p>  <p>• Variable expression for host-interacting genes</p> 		

Figure 6.1: Overview of each experimental chapter, with summarized aims, approaches, and results.

At the initiation of this project in late 2011, the *P. vivax* Sal 1 reference genome had recently been published (Carlton et al., 2008), as well as the first large-scale studies of transcription using microarray technology (Bozdech et al., 2008, Westenberger et al., 2010). While the first *P. falciparum* RNA-Seq study had been published (Otto et al., 2010), the prospects for similar *P. vivax* RNA-Seq studies were limited, due to the difficulty in obtaining RNA from clinical isolates of sufficiently high quality and quantity. There were few vaccine candidates under consideration partially due to the limited availability of recombinant *P. vivax* proteins. Most blood-stage protein studies relied on one or several protein fragments often produced in different expression systems (with the exception of a single panel of *P. vivax* protein fragments produced in the wheat-germ cell-free system). This significantly limited any ability to make systematic comparisons between proteins (for instance for immunoepidemiological studies), and no studies of panel of proteins for cross-sectional or longitudinal cohorts of *P. vivax*-endemic populations existed at the time.

This project aimed to broaden the potential field of vaccine candidates in a more comprehensive way. This first involved determining the largest pool of potential candidates through identification of proteins upregulated in the schizont stage of asexual blood-stage parasites. The Bozdech *et al.* microarray dataset provided a solid basis for this but lacked a complete set of probes, and we endeavoured to expand on this by producing RNA-Seq data using *P. vivax* clinical isolates from Cambodia. This involved the testing of RNA extraction methods and the bespoke production of Illumina strand-specific libraries, which led to over 150x average coverage for 4 isolates. The data provided schizont-stage transcription data (including transcription data for over 400 additional genes not included in the original microarray dataset), enabled the correction of hundreds of gene models, as well as uncovering novel transcripts. It also opened a window into potential differences between clinical isolates. Overall, it provided a basis from which to begin to build a list of proteins with potential involvement in *P. vivax* merozoite invasion of erythrocytes.

The next phase of the project involved the assembling of a list of *P. vivax* candidates for protein expression and downstream functional and immunoepidemiological screening. Utilizing the available RNA-Seq and published microarray data as well as homology comparisons with *P. falciparum* invasion antigens, we developed a list of 39 *P. vivax* candidates, 37 of which were successfully expressed in the mammalian HEK293E

system. I screened this protein panel in several large-scale functional assays including flow cytometry-based erythrocyte binding assays and 2 AVEXIS assays, which identified both predicted and novel parasite protein interactions. The erythrocyte-binding assays confirmed the binding of *P. falciparum* EBA175 to both erythrocytes and hematopoietic stem cell-derived reticulocytes and potentially the binding of *P. vivax* MTRAP to erythrocytes and reticulocytes, though with the need for additional confirmatory experiments. AVEXIS and subsequent SPR experiments confirmed for the first time that the *P. falciparum* P12-P41 interaction is conserved in *P. vivax* and has a much higher binding affinity. AVEXIS and SPR also supported 2 novel interactions between *P. vivax* MSP3.10 (MSP3 α) and MSP7.1, and between *P. vivax* P12 and PVX_110945, which open the door to further functional studies with field isolates.

The protein library enabled the study of antibody responses directed against the proteins from 3 *P. vivax*-endemic countries (Cambodia, SI, PNG), enabling one of the first screens of a panel of *P. vivax* proteins in reactivity, cross-sectional and longitudinal cohort studies. The vast majority of the protein library was immunogenic in all 3 settings. Several proteins were reactive in over 90% of patient samples in both Cambodia and the SI including MSP1, GAMA, and P12. A factorial screening of 144 patient samples in SI found stronger age-associated increases for IgG responses to 12 proteins compared to infection-associated increases, potentially pointing to the acquisition of long-lived, stable antibody responses to many *P. vivax* antigens. Responses to 3 proteins were additionally found to be associated with protection from clinical disease including P12, P41, and PVX_081550, the latter of which is a hypothetical protein for which little is known.

Overall, the project largely accomplished the primary goal of expanding the number of *P. vivax* antigens for further study as vaccine candidates and sets the stage for a number of follow-up studies in the years to come.

6.3 Conclusion

P. vivax research continues to lag significantly behind *P. falciparum*. However, steady improvements, often facilitated by changing technologies, are making several studies possible that were not feasible when this project began. One of the biggest improvements has been in the field of genomics as reductions in the quantity of DNA needed for making Illumina libraries had made the sequencing of field isolates feasible. A single *P. vivax*

isolate (the reference genome *P. vivax* Sal 1) had been sequenced at the start of this work; in the intervening years next generation sequencing technology has enabled the sequencing and analysis of hundreds of *P. vivax* isolates (Hupaló et al., 2016, Pearson et al., 2016). Genomics research for *P. vivax* may have caught up with *P. falciparum*, but *P. vivax* transcriptome studies, particularly those using RNA-Seq [for which only a single recent study exists (Zhu et al., 2016)], are still far behind.

While this project did not attempt to address the challenges of *P. vivax in vitro* blood stage culture, we did aim to expand the study of the proteins potentially involved in *P. vivax* merozoite invasion of reticulocytes. The availability of proteins to study in functional and immunoepidemiological assays remains a significant barrier for many groups, and we hope this will be somewhat eased by the deposition of our complete library of *P. vivax* vectors at the non-profit plasmid repository, Addgene.org. Several other groups are also assembling either full-length or subdomain *P. vivax* protein libraries (Finney et al., 2014, Lu et al., 2014), primarily for use in immunoreactivity studies. Collectively these protein panels will be important for enabling systematic comparisons between many proteins in the search for markers of exposure and IgG response profiles which correlate with protection, as these remain a potentially critical method for prioritizing candidates to develop as vaccine targets.

In conclusion, the work presented in this thesis identified several predicted and novel parasite protein-protein interactions as well as IgG responses directed against proteins, which were stably acquired with age and/or correlated with protection. All of these proteins are potential *P. vivax* vaccine candidates that we plan to investigate in further functional studies.

BIBLIOGRAPHY

- ADAMS, J. H., BLAIR, P. L., KANEKO, O. & PETERSON, D. S. 2001. An expanding ebl family of *Plasmodium falciparum*. *Trends Parasitol*, 17, 297-9.
- ADAMS, J. H., HUDSON, D. E., TORII, M., WARD, G. E., WELLEMS, T. E., AIKAWA, M. & MILLER, L. H. 1990. The Duffy receptor family of *Plasmodium knowlesi* is located within the micronemes of invasive malaria merozoites. *Cell*, 63, 141-53.
- ADDA, C. G., MURPHY, V. J., SUNDE, M., WADDINGTON, L. J., SCHLOEGEL, J., TALBO, G. H., VINGAS, K., KIENZLE, V., MASCIANTONIO, R., HOWLETT, G. J., HODDER, A. N., FOLEY, M. & ANDERS, R. F. 2009. *Plasmodium falciparum* merozoite surface protein 2 is unstructured and forms amyloid-like fibrils. *Mol Biochem Parasitol*, 166, 159-71.
- AIKAWA, M., MILLER, L. H., JOHNSON, J. & RABBEGE, J. 1978. Erythrocyte entry by malarial parasites. A moving junction between erythrocyte and parasite. *J Cell Biol*, 77, 72-82.
- AIKAWA, M., MILLER, L. H. & RABBEGE, J. 1975. Caveola-vesicle complexes in the plasmalemma of erythrocytes infected by *Plasmodium vivax* and *P cynomolgi*. Unique structures related to Schuffner's dots. *Am J Pathol*, 79, 285-300.
- ALONSO, P. L. & TANNER, M. 2013. Public health challenges and prospects for malaria control and elimination. *Nat Med*, 19, 150-5.
- ANDERS, S. & HUBER, W. 2010. Differential expression analysis for sequence count data. *Genome Biol*, 11, R106.
- ANDRADE, B. B., REIS-FILHO, A., SOUZA-NETO, S. M., CLARENCIO, J., CAMARGO, L. M., BARRAL, A. & BARRAL-NETTO, M. 2010. Severe *Plasmodium vivax* malaria exhibits marked inflammatory imbalance. *Malar J*, 9, 13.
- ANSTEE, D. J. 1990. The nature and abundance of human red cell surface glycoproteins. *J Immunogenet*, 17, 219-25.
- ARBEITMAN, M. N., FURLONG, E. E., IMAM, F., JOHNSON, E., NULL, B. H., BAKER, B. S., KRASNOW, M. A., SCOTT, M. P., DAVIS, R. W. & WHITE, K. P. 2002. Gene expression during the life cycle of *Drosophila melanogaster*. *Science*, 297, 2270-5.
- ARUMUGAM, T. U., TAKEO, S., YAMASAKI, T., THONKUKIATKUL, A., MIURA, K., OTSUKI, H., ZHOU, H., LONG, C. A., SATTABONGKOT, J., THOMPSON, J., WILSON, D. W., BEESON, J. G., HEALER, J., CRABB, B. S., COWMAN, A. F., TORII, M. & TSUBOI, T. 2011. Discovery of GAMA, a *Plasmodium falciparum* merozoite micronemal protein, as a novel blood-stage vaccine candidate antigen. *Infection and immunity*.

- AUBURN, S., MARFURT, J., MASLEN, G., CAMPINO, S., RUANO RUBIO, V., MANSKE, M., MACHUNTER, B., KENANGALEM, E., NOVIYANTI, R., TRIANTY, L., SEBAYANG, B., WIRJANATA, G., SRIPRAWAT, K., ALCOCK, D., MACINNIS, B., MIOTTO, O., CLARK, T. G., RUSSELL, B., ANSTEY, N. M., NOSTEN, F., KWIATKOWSKI, D. P. & PRICE, R. N. 2013. Effective preparation of Plasmodium vivax field isolates for high-throughput whole genome sequencing. *PLoS One*, 8, e53160.
- BAI, T., BECKER, M., GUPTA, A., STRIKE, P., MURPHY, V. J., ANDERS, R. F. & BATCHELOR, A. H. 2005. Structure of AMA1 from Plasmodium falciparum reveals a clustering of polymorphisms that surround a conserved hydrophobic pocket. *Proc Natl Acad Sci U S A*, 102, 12736-41.
- BAIRD, J. K. 2013. Evidence and implications of mortality associated with acute Plasmodium vivax malaria. *Clin Microbiol Rev*, 26, 36-57.
- BANNISTER, L. & MITCHELL, G. 2003. The ins, outs and roundabouts of malaria. *Trends Parasitol*, 19, 209-13.
- BANNISTER, L. H., BUTCHER, G. A., DENNIS, E. D. & MITCHELL, G. H. 1975. Structure and invasive behaviour of Plasmodium knowlesi merozoites in vitro. *Parasitology*, 71, 483-91.
- BARBEDO, M. B., RICCI, R., JIMENEZ, M. C., CUNHA, M. G., YAZDANI, S. S., CHITNIS, C. E., RODRIGUES, M. M. & SOARES, I. S. 2007. Comparative recognition by human IgG antibodies of recombinant proteins representing three asexual erythrocytic stage vaccine candidates of Plasmodium vivax. *Mem Inst Oswaldo Cruz*, 102, 335-9.
- BARGIERI, D., LAGAL, V., ANDENMATTEN, N., TARDIEUX, I., MEISSNER, M. & MENARD, R. 2014. Host cell invasion by apicomplexan parasites: the junction conundrum. *PLoS Pathog*, 10, e1004273.
- BARGIERI, D. Y., ANDENMATTEN, N., LAGAL, V., THIBERGE, S., WHITELAW, J. A., TARDIEUX, I., MEISSNER, M. & MENARD, R. 2013. Apical membrane antigen 1 mediates apicomplexan parasite attachment but is dispensable for host cell invasion. *Nat Commun*, 4, 2552.
- BARNWELL, J. W. 1990. Vesicle-mediated transport of membrane and proteins in malaria-infected erythrocytes. *Blood Cells*, 16, 379-395.
- BARNWELL, J. W. & GALINSKI, M. R. 1991. The adhesion of malaria merozoite proteins to erythrocytes: a reflection of function? *Research in Immunology*, 142, 666-672.
- BARTHOLDSON, S. J., BUSTAMANTE, L. Y., CROSNIER, C., JOHNSON, S., LEA, S., RAYNER, J. C. & WRIGHT, G. J. 2012. Semaphorin-7A is an erythrocyte receptor for P. falciparum merozoite-specific TRAP homolog, MTRAP. *PLoS Pathog*, 8, e1003031.
- BARUCH, D. I., PASLOSKE, B. L., SINGH, H. B., BI, X., MA, X. C., FELDMAN, M., TARASCHI, T. F. & HOWARD, R. J. 1995. Cloning the P. falciparum gene encoding PfEMP1, a malarial variant antigen and adherence receptor on the surface of parasitized human erythrocytes. *Cell*, 82, 77-87.
- BATCHELOR, J. D., MALPEDE, B. M., OMATTAGE, N. S., DEKOSTER, G. T., HENZLER-WILDMAN, K. A. & TOLIA, N. H. 2014. Red blood cell invasion by Plasmodium vivax: structural basis for DBP engagement of DARC. *PLoS Pathog*, 10, e1003869.
- BATCHELOR, J. D., ZAHM, J. A. & TOLIA, N. H. 2011. Dimerization of Plasmodium vivax DBP is induced upon receptor binding and drives recognition of DARC. *Nat Struct Mol Biol*, 18, 908-14.
- BAUM, J., CHEN, L., HEALER, J., LOPATICKI, S., BOYLE, M., TRIGLIA, T., EHLGEN, F., RALPH, S. A., BEESON, J. G. & COWMAN, A. F. 2009. Reticulocyte-binding protein homologue 5 - an essential adhesin involved in invasion of human erythrocytes by Plasmodium falciparum. *Int J Parasitol*, 39, 371-80.
- BAUM, J., RICHARD, D., HEALER, J., RUG, M., KRNAJSKI, Z., GILBERGER, T. W., GREEN, J. L., HOLDER, A. A. & COWMAN, A. F. 2006. A conserved molecular motor drives cell invasion and gliding motility across malaria life cycle stages and other apicomplexan parasites. *J Biol Chem*, 281, 5197-208.

- BEESON, J. G., AMIN, N., KANJALA, M. & ROGERSON, S. J. 2002. Selective accumulation of mature asexual stages of *Plasmodium falciparum*-infected erythrocytes in the placenta. *Infect Immun*, 70, 5412-5.
- BEG, M. A., KHAN, R., BAIG, S. M., GULZAR, Z., HUSSAIN, R. & SMEGO JR., R. A. 2002. Cerebral involvement in benign tertian malaria. *The American journal of tropical medicine and hygiene*, 67, 230-232.
- BEI, A. K. & DURAISINGH, M. T. 2012. Functional analysis of erythrocyte determinants of *Plasmodium* infection. *Int J Parasitol*, 42, 575-82.
- BENTLEY, D. R., BALASUBRAMANIAN, S., SWERDLOW, H. P., SMITH, G. P., MILTON, J., BROWN, C. G., HALL, K. P., EVERS, D. J., BARNES, C. L., BIGNELL, H. R., BOUTELL, J. M., BRYANT, J., CARTER, R. J., KEIRA CHEETHAM, R., COX, A. J., ELLIS, D. J., FLATBUSH, M. R., GORMLEY, N. A., HUMPHRAY, S. J., IRVING, L. J., KARBELASHVILI, M. S., KIRK, S. M., LI, H., LIU, X., MAISINGER, K. S., MURRAY, L. J., OBRADOVIC, B., OST, T., PARKINSON, M. L., PRATT, M. R., RASOLONJATOVO, I. M., REED, M. T., RIGATTI, R., RODIGHIERO, C., ROSS, M. T., SABOT, A., SANKAR, S. V., SCALLY, A., SCHROTH, G. P., SMITH, M. E., SMITH, V. P., SPIRIDOU, A., TORRANCE, P. E., TZONEV, S. S., VERMAAS, E. H., WALTER, K., WU, X., ZHANG, L., ALAM, M. D., ANASTASI, C., ANIEBO, I. C., BAILEY, D. M., BANCARZ, I. R., BANERJEE, S., BARBOUR, S. G., BAYBAYAN, P. A., BENOIT, V. A., BENSON, K. F., BEVIS, C., BLACK, P. J., BOODHUN, A., BRENNAN, J. S., BRIDGHAM, J. A., BROWN, R. C., BROWN, A. A., BUERMANN, D. H., BUNDU, A. A., BURROWS, J. C., CARTER, N. P., CASTILLO, N., CHIARA, E. C. M., CHANG, S., NEIL COOLEY, R., CRAKE, N. R., DADA, O. O., DIAKOUMAKOS, K. D., DOMINGUEZ-FERNANDEZ, B., EARNSHAW, D. J., EGBUJOR, U. C., ELMORE, D. W., ETCHIN, S. S., EWAN, M. R., FEDURCO, M., FRASER, L. J., FUENTES FAJARDO, K. V., SCOTT FUREY, W., GEORGE, D., GIETZEN, K. J., GODDARD, C. P., GOLDA, G. S., GRANIERI, P. A., GREEN, D. E., GUSTAFSON, D. L., HANSEN, N. F., HARNISH, K., HAUDENSCHILD, C. D., HEYER, N. I., HIMS, M. M., HO, J. T., HORGAN, A. M., et al. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, 456, 53-9.
- BETUELA, I., ROSANAS-URGELL, A., KINIBORO, B., STANISIC, D. I., SAMOL, L., DE LAZZARI, E., DEL PORTILLO, H. A., SIBA, P., ALONSO, P. L., BASSAT, Q. & MUELLER, I. 2012. Relapses contribute significantly to the risk of *Plasmodium vivax* infection and disease in Papua New Guinean children 1-5 years of age. *J Infect Dis*, 206, 1771-80.
- BICHET, M., JOLY, C., HENNI, A. H., GUILBERT, T., XEMARD, M., TAFANI, V., LAGAL, V., CHARRAS, G. & TARDIEUX, I. 2014. The toxoplasma-host cell junction is anchored to the cell cortex to sustain parasite invasive force. *BMC Biol*, 12, 773.
- BILLKER, O., LINDO, V., PANICO, M., ETIENNE, A. E., PAXTON, T., DELL, A., ROGERS, M., SINDEN, R. E. & MORRIS, H. R. 1998. Identification of xanthurenic acid as the putative inducer of malaria development in the mosquito. *Nature*, 392, 289-92.
- BIRKHOLTZ, L. M., BLATCH, G., COETZER, T. L., HOPPE, H. C., HUMAN, E., MORRIS, E. J., NGCETE, Z., OLDFIELD, L., ROTH, R., SHONHAI, A., STEPHENS, L. & LOUW, A. I. 2008. Heterologous expression of plasmodial proteins for structural studies and functional annotation. *Malar J*, 7, 197.
- BLACK, C. G., BARNWELL, J. W., HUBER, C. S., GALINSKI, M. R. & COPPEL, R. L. 2002. The *Plasmodium vivax* homologues of merozoite surface proteins 4 and 5 from *Plasmodium falciparum* are expressed at different locations in the merozoite. *Mol Biochem Parasitol*, 120, 215-224.
- BLACKMAN, M. J. 2008. Malarial proteases and host cell egress: an 'emerging' cascade. *Cell Microbiol*, 10, 1925-34.
- BLACKMAN, M. J. & CARRUTHERS, V. B. 2013. Recent insights into apicomplexan parasite egress provide new views to a kill. *Curr Opin Microbiol*, 16, 459-64.

- BLACKMAN, M. J., LING, I. T., NICHOLLS, S. C. & HOLDER, A. A. 1991. Proteolytic processing of the *Plasmodium falciparum* merozoite surface protein-1 produces a membrane-bound fragment containing two epidermal growth factor-like domains. *Mol Biochem Parasitol*, 49, 29-33.
- BLACKMAN, M. J., SCOTT-FINNIGAN, T. J., SHAI, S. & HOLDER, A. A. 1994. Antibodies inhibit the protease-mediated processing of a malaria merozoite surface protein. *J Exp Med*, 180, 389-93.
- BOUSEMA, T. & DRAKELEY, C. 2011. Epidemiology and infectivity of *Plasmodium falciparum* and *Plasmodium vivax* gametocytes in relation to malaria control and elimination. *Clin Microbiol Rev*, 24, 377-410.
- BOYD, M. F. M., C. B. 1939. Further Observations on the Duration of Immunity to the Homologous Strain of *Plasmodium vivax*. *Am. J. Trop. Med. Hyg.*, 19, 63-67.
- BOZDECH, Z., LLINAS, M., PULLIAM, B. L., WONG, E. D., ZHU, J. & DERISI, J. L. 2003. The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS Biol*, 1, E5.
- BOZDECH, Z., MOK, S., HU, G., IMWONG, M., JAIDEE, A., RUSSELL, B., GINSBURG, H., NOSTEN, F., DAY, N. P., WHITE, N. J., CARLTON, J. M. & PREISER, P. R. 2008. The transcriptome of *Plasmodium vivax* reveals divergence and diversity of transcriptional regulation in malaria parasites. *Proc Natl Acad Sci U S A*, 105, 16290-16295.
- BRIGHT, A. T., TEWHEY, R., ABELES, S., CHUQUIYAURI, R., LLANOS-CUENTAS, A., FERREIRA, M. U., SCHORK, N. J., VINETZ, J. M. & WINZELER, E. A. 2012. Whole genome sequencing analysis of *Plasmodium vivax* using whole genome capture. *BMC Genomics*, 13, 262.
- BROWN, K. N. & BROWN, I. N. 1965. Immunity to malaria: antigenic variation in chronic infections of *Plasmodium knowlesi*. *Nature*, 208, 1286-8.
- BROWN, M. H. 2002. Detection of low-affinity ligand-receptor interactions at the cell surface with fluorescent microspheres. *Curr Protoc Immunol*, Chapter 18, Unit 18 2.
- BROWN, M. H. & BARCLAY, A. N. 1994. Expression of immunoglobulin and scavenger receptor superfamily domains as chimeric proteins with domains 3 and 4 of CD4 for ligand analysis. *Protein Eng*, 7, 515-21.
- BROWN, M. H., BOLES, K., VAN DER MERWE, P. A., KUMAR, V., MATHEW, P. A. & BARCLAY, A. N. 1998. 2B4, the natural killer and T cell immunoglobulin superfamily surface protein, is a ligand for CD48. *J Exp Med*, 188, 2083-90.
- BULL, P. C., LOWE, B. S., KORTOK, M., MOLYNEUX, C. S., NEWBOLD, C. I. & MARSH, K. 1998. Parasite antigens on the infected red cell surface are targets for naturally acquired immunity to malaria. *Nat Med*, 4, 358-60.
- BURGESS, B. R., SCHUCK, P. & GARBOCZI, D. N. 2005. Dissection of merozoite surface protein 3, a representative of a family of *Plasmodium falciparum* surface proteins, reveals an oligomeric and highly elongated molecule. *J Biol Chem*, 280, 37236-45.
- BUSCAGLIA, C. A., COPPENS, I., HOL, W. G. & NUSSENZWEIG, V. 2003. Sites of interaction between aldolase and thrombospondin-related anonymous protein in plasmodium. *Mol Biol Cell*, 14, 4947-57.
- BUSHELL, K. M., SOLLNER, C., SCHUSTER-BOECKLER, B., BATEMAN, A. & WRIGHT, G. J. 2008. Large-scale screening for novel low-affinity extracellular protein interactions. *Genome Res*, 18, 622-630.
- CAMUS, D. & HADLEY, T. J. 1985. A *Plasmodium falciparum* antigen that binds to host erythrocytes and merozoites. *Science*, 230, 553-6.
- CARABALLO, J. R., DELGADO, G., RODRIGUEZ, R. & PATARROYO, M. A. 2007. The antigenicity of a *Plasmodium vivax* reticulocyte binding protein-1 (PvRBP1) recombinant fragment in humans and its immunogenicity and protection studies in Aotus monkeys. *Vaccine*, 25, 3713-3721.

- CARLTON, J. M., ADAMS, J. H., SILVA, J. C., BIDWELL, S. L., LORENZI, H., CALER, E., CRABTREE, J., ANGIUOLI, S. V., MERINO, E. F., AMEDEO, P., CHENG, Q., COULSON, R. M. R., CRABB, B. S., DEL PORTILLO, H. A., ESSIEN, K., FELDBLYUM, T. V., FERNANDEZ-BECERRA, C., GILSON, P. R., GUEYE, A. H., GUO, X., KANG'A, S., KOOIJ, T. W. A., KORSINCZKY, M., MEYER, E. V. S., NENE, V., PAULSEN, I., WHITE, O., RALPH, S. A., REN, Q., SARGEANT, T. J., SALZBERG, S. L., STOECKERT, C. J., SULLIVAN, S. A., YAMAMOTO, M. M., HOFFMAN, S. L., WORTMAN, J. R., GARDNER, M. J., GALINSKI, M. R., BARNWELL, J. W. & FRASER-LIGGETT, C. M. 2008. Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*. *Nature*, 455, 757-763.
- CARLTON, J. M., DAS, A. & ESCALANTE, A. A. 2013. Genomics, population genetics and evolutionary history of *Plasmodium vivax*. *Adv Parasitol*, 81, 203-22.
- CARVER, T., HARRIS, S. R., BERRIMAN, M., PARKHILL, J. & MCQUILLAN, J. A. 2012. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics*, 28, 464-9.
- CASTELLANOS, A., AREVALO-HERRERA, M., RESTREPO, N., GULLOSO, L., CORRADIN, G. & HERRERA, S. 2007. *Plasmodium vivax* thrombospondin related adhesion protein: immunogenicity and protective efficacy in rodents and Aotus monkeys. *Mem Inst Oswaldo Cruz*, 102, 411-6.
- CERAVOLO, I. P., SANCHEZ, B. A., SOUSA, T. N., GUERRA, B. M., SOARES, I. S., BRAGA, E. M., MCHENRY, A. M., ADAMS, J. H., BRITO, C. F. & CARVALHO, L. H. 2009. Naturally acquired inhibitory antibodies to *Plasmodium vivax* Duffy binding protein are short-lived and allele-specific following a single malaria infection. *Clin Exp Immunol*, 156, 502-510.
- CHAN, E. R., MENARD, D., DAVID, P. H., RATSIMBASOA, A., KIM, S., CHIM, P., DO, C., WITKOWSKI, B., MERCEREAU-PUJALON, O., ZIMMERMAN, P. A. & SERRE, D. 2012. Whole genome sequencing of field isolates provides robust characterization of genetic diversity in *Plasmodium vivax*. *PLoS Negl Trop Dis*, 6, e1811.
- CHANG, S. P., GIBSON, H. L., LEE-NG, C. T., BARR, P. J. & HUI, G. S. 1992. A carboxyl-terminal fragment of *Plasmodium falciparum* gp195 expressed by a recombinant baculovirus induces antibodies that completely inhibit parasite growth. *J Immunol*, 149, 548-55.
- CHEN, E., PAING, M. M., SALINAS, N., SIM, B. K. & TOLIA, N. H. 2013. Structural and functional basis for inhibition of erythrocyte invasion by antibodies that target *Plasmodium falciparum* EBA-175. *PLoS Pathog*, 9, e1003390.
- CHEN, E., SALINAS, N. D., NTUMNGIA, F. B., ADAMS, J. H. & TOLIA, N. H. 2015. Structural analysis of the synthetic Duffy Binding Protein (DBP) antigen DEKnull relevant for *Plasmodium vivax* malaria vaccine design. *PLoS Negl Trop Dis*, 9, e0003644.
- CHEN, J. H., JUNG, J. W., WANG, Y., HA, K. S., LU, F., LIM, C. S., TAKEO, S., TSUBOI, T. & HAN, E. T. 2010. Immunoproteomics profiling of blood stage *Plasmodium vivax* infection by high-throughput screening assays. *J Proteome Res*, 9, 6479-89.
- CHEN, L., LOPATICKI, S., RIGLAR, D. T., DEKIWADIA, C., UBOLDI, A. D., THAM, W. H., O'NEILL, M. T., RICHARD, D., BAUM, J., RALPH, S. A. & COWMAN, A. F. 2011. An EGF-like protein forms a complex with PfRh5 and is required for invasion of human erythrocytes by *Plasmodium falciparum*. *PLoS pathogens*, 7, e1002199.
- CHEN, N. & CHENG, Q. 1999. Codon usage in *Plasmodium vivax* nuclear genes. *Int J Parasitol*, 29, 445-9.
- CHENG, Q. & SAUL, A. 1994. Sequence analysis of the apical membrane antigen I (AMA-1) of *Plasmodium vivax*. *Mol Biochem Parasitol*, 65, 183-187.
- CHENG, Y., SHIN, E. H., LU, F., WANG, B., CHOE, J., TSUBOI, T. & HAN, E. T. 2014. Antigenicity studies in humans and immunogenicity studies in mice: an MSP1P subdomain as a candidate for malaria vaccine development. *Microbes Infect*, 16, 419-28.

- CHITNIS, C. E. & MILLER, L. H. 1994. Identification of the erythrocyte binding domains of *Plasmodium vivax* and *Plasmodium knowlesi* proteins involved in erythrocyte invasion. *J Exp Med*, 180, 497-506.
- CIUCA, M. B., L.; CHELARESCU-VIERU, M. 1934. Immunity in malaria. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 27, 619-622.
- CLAESSENS, A., HAMILTON, W. L., KEKRE, M., OTTO, T. D., FAIZULLABHOY, A., RAYNER, J. C. & KWIATKOWSKI, D. 2014. Generation of antigenic diversity in *Plasmodium falciparum* by structured rearrangement of Var genes during mitosis. *PLoS Genet*, 10, e1004812.
- COHEN, S., MC, G. I. & CARRINGTON, S. 1961. Gamma-globulin and acquired immunity to human malaria. *Nature*, 192, 733-7.
- COLE-TOBIAN, J. L., CORTES, A., BAISOR, M., KASTENS, W., XAINLI, J., BOCKARIE, M., ADAMS, J. H. & KING, C. L. 2002. Age-acquired immunity to a *Plasmodium vivax* invasion ligand, the duffy binding protein. *J Infect Dis*, 186, 531-539.
- COLLINS, W. E., CONTACOS, P. G., KROTOSKI, W. A. & HOWARD, W. A. 1972. Transmission of four Central American strains of *Plasmodium vivax* from monkey to man. *J Parasitol*, 58, 332-5.
- COLLINS, W. E. & JEFFERY, G. M. 2005. *Plasmodium ovale*: parasite and disease. *Clin Microbiol Rev*, 18, 570-81.
- COVELL, G. & NICOL, W. D. 1951. Clinical, chemotherapeutic and immunological studies on induced malaria. *Br Med Bull*, 8, 51-5.
- COWMAN, A. F. & CRABB, B. S. 2006. Invasion of red blood cells by malaria parasites. *Cell*, 124, 755-766.
- CRANSTON, H. A., BOYLAN, C. W., CARROLL, G. L., SUTERA, S. P., WILLIAMSON, J. R., GLUZMAN, I. Y. & KROGSTAD, D. J. 1984. *Plasmodium falciparum* maturation abolishes physiologic red cell deformability. *Science*, 223, 400-403.
- CROMPTON, P. D., KAYALA, M. A., TRAORE, B., KAYENTAO, K., ONGOIBA, A., WEISS, G. E., MOLINA, D. M., BURK, C. R., WAISBERG, M., JASINSKAS, A., TAN, X., DOUMBO, S., DOUMTABA, D., KONE, Y., NARUM, D. L., LIANG, X., DOUMBO, O. K., MILLER, L. H., DOOLAN, D. L., BALDI, P., FELGNER, P. L. & PIERCE, S. K. 2010. A prospective analysis of the Ab response to *Plasmodium falciparum* before and after a malaria season by protein microarray. *Proc Natl Acad Sci U S A*, 107, 6958-63.
- CROSNIER, C., BUSTAMANTE, L. Y., BARTHOLDSON, S. J., BEI, A. K., THERON, M., UCHIKAWA, M., MBOUP, S., NDIR, O., KWIATKOWSKI, D. P., DURAISINGH, M. T., RAYNER, J. C. & WRIGHT, G. J. 2011. Basigin is a receptor essential for erythrocyte invasion by *Plasmodium falciparum*. *Nature*, 480, 534-537.
- CROSNIER, C., STAUDT, N. & WRIGHT, G. J. 2010. A rapid and scalable method for selecting recombinant mouse monoclonal antibodies. *BMC Biol*, 8, 76.
- CROSNIER, C., WANAGURU, M., MCDADE, B., OSIER, F. H., MARSH, K., RAYNER, J. C. & WRIGHT, G. J. 2013. A library of functional recombinant cell-surface and secreted *P. falciparum* merozoite proteins. *Mol Cell Proteomics*, 12, 3976-86.
- CUTTS, J. C., POWELL, R., AGIUS, P. A., BEESON, J. G., SIMPSON, J. A. & FOWKES, F. J. 2014. Immunological markers of *Plasmodium vivax* exposure and immunity: a systematic review and meta-analysis. *BMC Med*, 12, 150.
- D'OMBRAIN, M. C., ROBINSON, L. J., STANISIC, D. I., TARAICA, J., BERNARD, N., MICHON, P., MUELLER, I. & SCHOFIELD, L. 2008. Association of early interferon-gamma production with immunity to clinical malaria: a longitudinal study among Papua New Guinean children. *Clin Infect Dis*, 47, 1380-7.
- DAUGHERTY, J. R., MURPHY, C. I., DOROS-RICHERT, L. A., BARBOSA, A., KASHALA, L. O., BALLOU, W. R., SNELLINGS, N. J., OCKENHOUSE, C. F. & LANAR, D. E. 1997.

- Baculovirus-mediated expression of *Plasmodium falciparum* erythrocyte binding antigen 175 polypeptides and their recognition by human antibodies. *Infect Immun*, 65, 3631-7.
- DAY, N. P., HIEN, T. T., SCHOLLAARDT, T., LOC, P. P., CHUONG, L. V., CHAU, T. T., MAI, N. T., PHU, N. H., SINH, D. X., WHITE, N. J. & HO, M. 1999. The prognostic and pathophysiologic role of pro- and antiinflammatory cytokines in severe malaria. *J Infect Dis*, 180, 1288-97.
- DEL PORTILLO, H. A., GYSIN, J., MATTEI, D. M., KHOURI, E., UDAGAMA, P. V., MENDIS, K. N. & DAVID, P. H. 1988. *Plasmodium vivax*: cloning and expression of a major blood-stage surface antigen. *Exp Parasitol*, 67, 346-53.
- DEL PORTILLO, H. A., LONGACRE, S., KHOURI, E. & DAVID, P. H. 1991. Primary structure of the merozoite surface antigen 1 of *Plasmodium vivax* reveals sequences conserved between different *Plasmodium* species. *Proceedings of the National Academy of Sciences of the United States of America*, 88, 4030-4034.
- DEVI, Y. S., MUKHERJEE, P., YAZDANI, S. S., SHAKRI, A. R., MAZUMDAR, S., PANDEY, S., CHITNIS, C. E. & CHAUHAN, V. S. 2007. Immunogenicity of *Plasmodium vivax* combination subunit vaccine formulated with human compatible adjuvants in mice. *Vaccine*, 25, 5166-74.
- DHARIA, N. V., BRIGHT, A. T., WESTENBERGER, S. J., BARNES, S. W., BATALOV, S., KUHEN, K., BORBOA, R., FEDERE, G. C., MCCLEAN, C. M., VINETZ, J. M., NEYRA, V., LLANOS-CUENTAS, A., BARNWELL, J. W., WALKER, J. R. & WINZELER, E. A. 2010. Whole-genome sequencing and microarray analysis of ex vivo *Plasmodium vivax* reveal selective pressure on putative drug resistance genes. *Proc Natl Acad Sci U S A*, 107, 20045-50.
- DONDORP, A. M., NOSTEN, F., YI, P., DAS, D., PHYO, A. P., TARNING, J., LWIN, K. M., ARIEY, F., HANPITHAKPONG, W., LEE, S. J., RINGWALD, P., SILAMUT, K., IMWONG, M., CHOTIVANICH, K., LIM, P., HERDMAN, T., AN, S. S., YEUNG, S., SINGHASIVANON, P., DAY, N. P., LINDEGARDH, N., SOCHEAT, D. & WHITE, N. J. 2009. Artemisinin resistance in *Plasmodium falciparum* malaria. *N Engl J Med*, 361, 455-67.
- DOOLAN, D. L., DOBANO, C. & BAIRD, J. K. 2009. Acquired immunity to malaria. *Clin Microbiol Rev*, 22, 13-36, Table of Contents.
- DOOLAN, D. L., MU, Y., UNAL, B., SUNDARESH, S., HIRST, S., VALDEZ, C., RANDALL, A., MOLINA, D., LIANG, X., FREILICH, D. A., OLOO, J. A., BLAIR, P. L., AGUIAR, J. C., BALDI, P., DAVIES, D. H. & FELGNER, P. L. 2008. Profiling humoral immune responses to *P. falciparum* infection with protein microarrays. *Proteomics*, 8, 4680-94.
- DOUGLAS, A. D., BALDEVIANO, G. C., LUCAS, C. M., LUGO-ROMAN, L. A., CROSNIER, C., BARTHOLDSON, S. J., DIOUF, A., MIURA, K., LAMBERT, L. E., VENTOCILLA, J. A., LEIVA, K. P., MILNE, K. H., ILLINGWORTH, J. J., SPENCER, A. J., HJERRILD, K. A., ALANINE, D. G., TURNER, A. V., MOORHEAD, J. T., EDGEL, K. A., WU, Y., LONG, C. A., WRIGHT, G. J., LESCOANO, A. G. & DRAPER, S. J. 2015. A PfPR5-based vaccine is efficacious against heterologous strain blood-stage *Plasmodium falciparum* infection in aotus monkeys. *Cell Host Microbe*, 17, 130-9.
- DOUGLAS, N. M., ANSTEY, N. M., ANGUS, B. J., NOSTEN, F. & PRICE, R. N. 2010. Artemisinin combination therapy for vivax malaria. *The Lancet infectious diseases*, 10, 405-16.
- DREYER, A. M., MATILE, H., PAPASTOGIANNIDIS, P., KAMBER, J., FAVUZZA, P., VOSS, T. S., WITTLIN, S. & PLUSCHKE, G. 2012. Passive immunoprotection of *Plasmodium falciparum*-infected mice designates the CyRPA as candidate malaria vaccine antigen. *J Immunol*, 188, 6225-37.
- DROEGE, M. & HILL, B. 2008. The Genome Sequencer FLX System--longer reads, more applications, straight forward bioinformatics and more complete data sets. *J Biotechnol*, 136, 3-10.
- DUFFY, P. E., CRAIG, A. G. & BARUCH, D. I. 2001. Variant proteins on the surface of malaria-infected erythrocytes--developing vaccines. *Trends Parasitol*, 17, 354-6.

- DURASINGH, M. T., TRIGLIA, T., RALPH, S. A., RAYNER, J. C., BARNWELL, J. W., MCFADDEN, G. I. & COWMAN, A. F. 2003. Phenotypic variation of *Plasmodium falciparum* merozoite proteins directs receptor targeting for invasion of human erythrocytes. *EMBO J*, 22, 1047-57.
- DUROCHER, Y., PERRET, S. & KAMEN, A. 2002. High-level and high-throughput recombinant protein production by transient transfection of suspension-growing human 293-EBNA1 cells. *Nucleic Acids Res*, 30, E9.
- DUSTIN, M. L. & SPRINGER, T. A. 1991. Role of lymphocyte adhesion receptors in transient interactions and cell locomotion. *Annu Rev Immunol*, 9, 27-66.
- DVORAK, J. A., MILLER, L. H., WHITEHOUSE, W. C. & SHIROISHI, T. 1975. Invasion of erythrocytes by malaria merozoites. *Science*, 187, 748-50.
- EGARTER, S., ANDENMATTEN, N., JACKSON, A. J., WHITELAW, J. A., PALL, G., BLACK, J. A., FERGUSON, D. J., TARDIEUX, I., MOGILNER, A. & MEISSNER, M. 2014. The toxoplasma Acto-MyoA motor complex is important but not essential for gliding motility and host cell invasion. *PLoS One*, 9, e91819.
- EVANS, A. G. & WELLEMS, T. E. 2002. Coevolutionary genetics of *Plasmodium malaria* parasites and their human hosts. *Integr Comp Biol*, 42, 401-7.
- FAN, Y. T., WANG, Y., JU, C., ZHANG, T., XU, B., HU, W. & CHEN, J. H. 2013. Systematic analysis of natural antibody responses to *P. falciparum* merozoite antigens by protein arrays. *J Proteomics*, 78, 148-58.
- FANG, X. D., KASLOW, D. C., ADAMS, J. H. & MILLER, L. H. 1991. Cloning of the *Plasmodium vivax* Duffy receptor. *Mol Biochem Parasitol*, 44, 125-32.
- FARROKHI, N., HRMOVA, M., BURTON, R. A. & FINCHER, G. B. 2009. Heterologous and cell free protein expression systems. *Methods Mol Biol*, 513, 175-98.
- FARROW, R. E., GREEN, J., KATSIMITSOULIA, Z., TAYLOR, W. R., HOLDER, A. A. & MOLLOY, J. E. 2011. The mechanism of erythrocyte invasion by the malarial parasite, *Plasmodium falciparum*. *Semin Cell Dev Biol*, 22, 953-60.
- FERNANDES, A. A., CARVALHO, L. J., ZANINI, G. M., VENTURA, A. M., SOUZA, J. M., COTIAS, P. M., SILVA-FILHO, I. L. & DANIEL-RIBEIRO, C. T. 2008. Similar cytokine responses and degrees of anemia in patients with *Plasmodium falciparum* and *Plasmodium vivax* infections in the Brazilian Amazon region. *Clin Vaccine Immunol*, 15, 650-8.
- FINNEY, O. C., DANZIGER, S. A., MOLINA, D. M., VIGNALI, M., TAKAGI, A., JI, M., STANISIC, D. I., SIBA, P. M., LIANG, X., AITCHISON, J. D., MUELLER, I., GARDNER, M. J. & WANG, R. 2014. Predicting antidisease immunity using proteome arrays and sera from children naturally exposed to malaria. *Mol Cell Proteomics*, 13, 2646-60.
- FODOR, S. P., READ, J. L., PIRRUNG, M. C., STRYER, L., LU, A. T. & SOLAS, D. 1991. Light-directed, spatially addressable parallel chemical synthesis. *Science*, 251, 767-73.
- FOWKES, F. J., MCGREADY, R., CROSS, N. J., HOMMEL, M., SIMPSON, J. A., ELLIOTT, S. R., RICHARDS, J. S., LACKOVIC, K., VILADPAI-NGUEN, J., NARUM, D., TSUBOI, T., ANDERS, R. F., NOSTEN, F. & BEESON, J. G. 2012. New insights into acquisition, boosting, and longevity of immunity to malaria in pregnant women. *J Infect Dis*, 206, 1612-21.
- FRANCA, C. T., HOSTETLER, J. B., SHARMA, S., WHITE, M. T., LIN, E., KINIBORO, B., WALTMANN, A., DARCY, A. W., LI WAI SUEN, C. S., SIBA, P., KING, C. L., RAYNER, J. C., FAIRHURST, R. M. & MUELLER, I. 2016. An Antibody Screen of a *Plasmodium vivax* Antigen Library Identifies Novel Merozoite Proteins Associated with Clinical Protection. *PLoS Negl Trop Dis*, 10, e0004639.
- FRASER, T., MICHON, P., BARNWELL, J. W., NOE, A. R., AL-YAMAN, F., KASLOW, D. C. & ADAMS, J. H. 1997. Expression and serologic activity of a soluble recombinant *Plasmodium vivax* Duffy binding protein. *Infect Immun*, 65, 2772-2777.

- FRECH, C. & CHEN, N. 2011. Genome Comparison of Human and Non-Human Malaria Parasites Reveals Species Subset-Specific Genes Potentially Linked to Human Disease. *PLoS Computational Biology*, 7, e1002320.
- FRIED, M., DOMINGO, G. J., GOWDA, C. D., MUTABINGWA, T. K. & DUFFY, P. E. 2006. Plasmodium falciparum: chondroitin sulfate A is the major receptor for adhesion of parasitized erythrocytes in the placenta. *Exp Parasitol*, 113, 36-42.
- GALINSKI, M. R. & BARNWELL, J. W. 1996. Plasmodium vivax: Merozoites, invasion of reticulocytes and considerations for malaria vaccine development. *Parasitol Today*, 12, 20-29.
- GALINSKI, M. R., MEDINA, C. C., INGRAVALLO, P. & BARNWELL, J. W. 1992. A reticulocyte-binding protein complex of Plasmodium vivax merozoites. *Cell*, 69, 1213-1226.
- GALINSKI, M. R., XU, M. & BARNWELL, J. W. 2000. Plasmodium vivax reticulocyte binding protein-2 (PvRBP-2) shares structural features with PvRBP-1 and the Plasmodium yoelii 235 kDa rhoptry protein family. *Mol Biochem Parasitol*, 108, 257-262.
- GAO, X., GUNALAN, K., YAP, S. S. & PREISER, P. R. 2013. Triggers of key calcium signals during erythrocyte invasion by Plasmodium falciparum. *Nat Commun*, 4, 2862.
- GARCIA, C. R., DE AZEVEDO, M. F., WUNDERLICH, G., BUDU, A., YOUNG, J. A. & BANNISTER, L. 2008. Plasmodium in the postgenomic era: new insights into the molecular cell biology of malaria parasites. *Int Rev Cell Mol Biol*, 266, 85-156.
- GARCIA, C. R., MANZI, F., TEDIOSI, F., HOFFMAN, S. L. & JAMES, E. R. 2013. Comparative cost models of a liquid nitrogen vapor phase (LNVP) cold chain-distributed cryopreserved malaria vaccine vs. a conventional vaccine. *Vaccine*, 31, 380-6.
- GARDNER, M. J., HALL, N., FUNG, E., WHITE, O., BERRIMAN, M., HYMAN, R. W., CARLTON, J. M., PAIN, A., NELSON, K. E., BOWMAN, S., PAULSEN, I. T., JAMES, K., EISEN, J. A., RUTHERFORD, K., SALZBERG, S. L., CRAIG, A., KYES, S., CHAN, M. S., NENE, V., SHALLOM, S. J., SUH, B., PETERSON, J., ANGIUOLI, S., PERTEA, M., ALLEN, J., SELENGUT, J., HAFT, D., MATHER, M. W., VAIDYA, A. B., MARTIN, D. M., FAIRLAMB, A. H., FRAUNHOLZ, M. J., ROOS, D. S., RALPH, S. A., MCFADDEN, G. I., CUMMINGS, L. M., SUBRAMANIAN, G. M., MUNGALL, C., VENTER, J. C., CARUCCI, D. J., HOFFMAN, S. L., NEWBOLD, C., DAVIS, R. W., FRASER, C. M. & BARRELL, B. 2002. Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature*, 419, 498-511.
- GARG, S., CHAUHAN, S. S., SINGH, N. & SHARMA, Y. D. 2008. Immunological responses to a 39.8kDa Plasmodium vivax tryptophan-rich antigen (PvTRAg39.8) among humans. *Microbes Infect*, 10, 1097-105.
- GEELS, M. J., IMOUKHUEDE, E. B., IMBAULT, N., VAN SCHOOTEN, H., MCWADE, T., TROYE-BLOMBERG, M., DOBBELAER, R., CRAIG, A. G. & LEROY, O. 2011. European Vaccine Initiative: lessons from developing malaria vaccines. *Expert Rev Vaccines*, 10, 1697-708.
- GENTON, B., AL-YAMAN, F., BECK, H. P., HII, J., MELLOR, S., NARARA, A., GIBSON, N., SMITH, T. & ALPERS, M. P. 1995. The epidemiology of malaria in the Wosera area, East Sepik Province, Papua New Guinea, in preparation for vaccine trials. I. Malariometric indices and immunity. *Ann Trop Med Parasitol*, 89, 359-76.
- GERLOFF, D. L., CREASEY, A., MASLAU, S. & CARTER, R. 2005. Structural models for the protein family characterized by gamete surface protein Pfs230 of Plasmodium falciparum. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 13598-603.
- GETHING, P. W., ELYAZAR, I. R., MOYES, C. L., SMITH, D. L., BATTLE, K. E., GUERRA, C. A., PATIL, A. P., TATEM, A. J., HOWES, R. E., MYERS, M. F., GEORGE, D. B., HORBY, P., WERTHEIM, H. F., PRICE, R. N., MUELLER, I., BAIRD, J. K. & HAY, S. I. 2012. A long neglected world malaria map: Plasmodium vivax endemicity in 2010. *PLoS Negl Trop Dis*, 6, e1814.

- GETHING, P. W., PATIL, A. P., SMITH, D. L., GUERRA, C. A., ELYAZAR, I. R., JOHNSTON, G. L., TATEM, A. J. & HAY, S. I. 2011. A new world malaria map: *Plasmodium falciparum* endemicity in 2010. *Malar J*, 10, 378.
- GILSON, P. R. & CRABB, B. S. 2009. Morphology and kinetics of the three distinct phases of red blood cell invasion by *Plasmodium falciparum* merozoites. *Int J Parasitol*, 39, 91-6.
- GIOVANNINI, D., SPATH, S., LACROIX, C., PERAZZI, A., BARGIERI, D., LAGAL, V., LEBUGLE, C., COMBE, A., THIBERGE, S., BALDACCI, P., TARDIEUX, I. & MENARD, R. 2011. Independent roles of apical membrane antigen 1 and rhoptry neck proteins during host cell invasion by apicomplexa. *Cell Host Microbe*, 10, 591-602.
- GIRALDO, J., VIVAS, N. M., VILA, E. & BADIA, A. 2002. Assessing the (a)symmetry of concentration-effect curves: empirical versus mechanistic models. *Pharmacol Ther*, 95, 21-45.
- GIRALDO, M. A., AREVALO-PINZON, G., ROJAS-CARABALLO, J., MONGUI, A., RODRIGUEZ, R. & PATARROYO, M. A. 2009. Vaccination with recombinant *Plasmodium vivax* MSP-10 formulated in different adjuvants induces strong immunogenicity but no protection. *Vaccine*, 28, 7-13.
- GONCALVES, R. M., SCOPEL, K. K., BASTOS, M. S. & FERREIRA, M. U. 2012. Cytokine balance in human malaria: does *Plasmodium vivax* elicit more inflammatory responses than *Plasmodium falciparum*? *PLoS One*, 7, e44394.
- GONDEAU, C., CORRADIN, G., HEITZ, F., LE PEUCH, C., BALBO, A., SCHUCK, P. & KAJAVA, A. V. 2009. The C-terminal domain of *Plasmodium falciparum* merozoite surface protein 3 self-assembles into alpha-helical coiled coil tetramer. *Mol Biochem Parasitol*, 165, 153-61.
- GOODMAN, C. A., COLEMAN, P. G. & MILLS, A. J. 1999. Cost-effectiveness of malaria control in sub-Saharan Africa. *Lancet*, 354, 378-85.
- GOWDA, D. C. & DAVIDSON, E. A. 1999. Protein glycosylation in the malaria parasite. *Parasitol Today*, 15, 147-52.
- GRAU, G. E., TAYLOR, T. E., MOLYNEUX, M. E., WIRIMA, J. J., VASSALLI, P., HOMMEL, M. & LAMBERT, P. H. 1989. Tumor necrosis factor and disease severity in children with falciparum malaria. *N Engl J Med*, 320, 1586-91.
- GREENWOOD, B. M. & TARGETT, G. A. 2011. Malaria vaccines and the new malaria agenda. *Clin Microbiol Infect*, 17, 1600-7.
- GRIMBERG, B. T., UDOMSANGPETCH, R., XAINLI, J., MCHENRY, A., PANICHAKUL, T., SATTABONGKOT, J., CUI, L., BOCKARIE, M., CHITNIS, C., ADAMS, J., ZIMMERMAN, P. A. & KING, C. L. 2007. *Plasmodium vivax* invasion of human erythrocytes inhibited by antibodies directed against the Duffy binding protein. *PLoS Med*, 4, e337.
- GRURING, C., MOON, R. W., LIM, C., HOLDER, A. A., BLACKMAN, M. J. & DURAISINGH, M. T. 2014. Human red blood cell-adapted *Plasmodium knowlesi* parasites: a new model system for malaria research. *Cell Microbiol*, 16, 612-20.
- GRUSZCZYK, J., LIM, N. T., ARNOTT, A., HE, W. Q., NGUITRAGOOL, W., ROOBSOONG, W., MOK, Y. F., MURPHY, J. M., SMITH, K. R., LEE, S., BAHLO, M., MUELLER, I., BARRY, A. E. & THAM, W. H. 2016. Structurally conserved erythrocyte-binding domain in *Plasmodium* provides a versatile scaffold for alternate receptor engagement. *Proc Natl Acad Sci U S A*, 113, E191-200.
- GUERRA, C. A., SNOW, R. W. & HAY, S. I. 2006. Defining the global spatial limits of malaria transmission in 2005. *Adv Parasitol*, 62, 157-79.
- HAN, H. J., PARK, S. G., KIM, S. H., HWANG, S. Y., HAN, J., TRAIKOFF, J., KHO, W. G. & CHUNG, J. Y. 2004. Epidermal growth factor-like motifs 1 and 2 of *Plasmodium vivax* merozoite surface protein 1 are critical domains in erythrocyte invasion. *Biochemical and biophysical research communications*, 320, 563-570.

- HARRIS, P. K., YEOH, S., DLUZEWSKI, A. R., O'DONNELL, R. A., WITHERS-MARTINEZ, C., HACKETT, F., BANNISTER, L. H., MITCHELL, G. H. & BLACKMAN, M. J. 2005. Molecular identification of a malaria merozoite surface sheddase. *PLoS Pathog*, 1, 241-251.
- HARVEY, K. L., GILSON, P. R. & CRABB, B. S. 2012. A model for the progression of receptor-ligand interactions during erythrocyte invasion by *Plasmodium falciparum*. *International journal for parasitology*, 42, 567-73.
- HAY, S. I., GUERRA, C. A., TATEM, A. J., NOOR, A. M. & SNOW, R. W. 2004. The global distribution and population at risk of malaria: past, present, and future. *Lancet Infectious Diseases*, 4, 327-336.
- HAYNES, J. D., DALTON, J. P., KLOTZ, F. W., MCGINNISS, M. H., HADLEY, T. J., HUDSON, D. E. & MILLER, L. H. 1988. Receptor-like specificity of a *Plasmodium knowlesi* malarial protein that binds to Duffy antigen ligands on erythrocytes. *J Exp Med*, 167, 1873-81.
- HE, C. Y., STRIEPEN, B., PLETCHER, C. H., MURRAY, J. M. & ROOS, D. S. 2001. Targeting and processing of nuclear-encoded apicoplast proteins in plastid segregation mutants of *Toxoplasma gondii*. *J Biol Chem*, 276, 28436-42.
- HEISS, K., NIE, H., KUMAR, S., DALY, T. M., BERGMAN, L. W. & MATUSCHEWSKI, K. 2008. Functional characterization of a redundant *Plasmodium* TRAP family invasin, TRAP-like protein, by aldolase binding and a genetic complementation test. *Eukaryot Cell*, 7, 1062-70.
- HEMMER, C. J., HOLST, F. G., KERN, P., CHIWAKATA, C. B., DIETRICH, M. & REISINGER, E. C. 2006. Stronger host response per parasitized erythrocyte in *Plasmodium vivax* or *ovale* than in *Plasmodium falciparum* malaria. *Trop Med Int Health*, 11, 817-23.
- HERRERA, S., FERNANDEZ, O. L., VERA, O., CARDENAS, W., RAMIREZ, O., PALACIOS, R., CHEN-MOK, M., CORRADIN, G. & AREVALO-HERRERA, M. 2011. Phase I safety and immunogenicity trial of *Plasmodium vivax* CS derived long synthetic peptides adjuvanted with montanide ISA 720 or montanide ISA 51. *Am J Trop Med Hyg*, 84, 12-20.
- HESTER, J., CHAN, E. R., MENARD, D., MERCEREAU-PUJALON, O., BARNWELL, J., ZIMMERMAN, P. A. & SERRE, D. 2013. De novo assembly of a field isolate genome reveals novel *Plasmodium vivax* erythrocyte invasion genes. *PLoS Negl Trop Dis*, 7, e2569.
- HILL, A. V. 2011. Vaccines against malaria. *Philos Trans R Soc Lond B Biol Sci*, 366, 2806-14.
- HODDER, A. N., MALBY, R. L., CLARKE, O. B., FAIRLIE, W. D., COLMAN, P. M., CRABB, B. S. & SMITH, B. J. 2009. Structural insights into the protease-like antigen *Plasmodium falciparum* SERA5 and its noncanonical active-site serine. *J Mol Biol*, 392, 154-65.
- HOFFMAN, S. L., ISENBARGER, D., LONG, G. W., SEDEGAH, M., SZARFMAN, A., WATERS, L., HOLLINGDALE, M. R., VAN DER MEIDE, P. H., FINBLOOM, D. S. & BALLOU, W. R. 1989. Sporozoite vaccine induces genetically restricted T cell elimination of malaria from hepatocytes. *Science*, 244, 1078-81.
- HOLDER, A. A. & FREEMAN, R. R. 1981. Immunization against blood-stage rodent malaria using purified parasite antigens. *Nature*, 294, 361-4.
- HORROCKS, P., PINCHES, R., CHRISTODOULOU, Z., KYES, S. A. & NEWBOLD, C. I. 2004. Variable var transition rates underlie antigenic variation in malaria. *Proc Natl Acad Sci U S A*, 101, 11129-34.
- HORUK, R., CHITNIS, C. E., DARBONNE, W. C., COLBY, T. J., RYBICKI, A., HADLEY, T. J. & MILLER, L. H. 1993. A receptor for the malarial parasite *Plasmodium vivax*: the erythrocyte chemokine receptor. *Science*, 261, 1182-4.
- HOSTETLER, J. B., SHARMA, S., BARTHOLDSON, S. J., WRIGHT, G. J., FAIRHURST, R. M. & RAYNER, J. C. 2015. A Library of *Plasmodium vivax* Recombinant Merozoite Proteins Reveals New Vaccine Candidates and Protein-Protein Interactions. *PLoS Negl Trop Dis*, 9, e0004264.
- HUPALO, D. N., LUO, Z., MELNIKOV, A., SUTTON, P. L., ROGOV, P., ESCALANTE, A., VALLEJO, A. F., HERRERA, S., AREVALO-HERRERA, M., FAN, Q., WANG, Y., CUI, L., LUCAS, C. M., DURAND, S., SANCHEZ, J. F., BALDEVIANO, G. C., LESCANO, A. G., LAMAN, M.,

- BARNADAS, C., BARRY, A., MUELLER, I., KAZURA, J. W., EAPEN, A., KANAGARAJ, D., VALECHA, N., FERREIRA, M. U., ROOBSONG, W., NGUITRAGOOL, W., SATTABONKOT, J., GAMBOA, D., KOSEK, M., VINETZ, J. M., GONZALEZ-CERON, L., BIRREN, B. W., NEAFSEY, D. E. & CARLTON, J. M. 2016. Population genomics studies identify signatures of global dispersal and drug resistance in *Plasmodium vivax*. *Nat Genet*.
- IMAM, M., SINGH, S., KAUSHIK, N. K. & CHAUHAN, V. S. 2014. *Plasmodium falciparum* merozoite surface protein 3: oligomerization, self-assembly, and heme complex formation. *J Biol Chem*, 289, 3856-68.
- ISHINO, T., CHINZEI, Y. & YUDA, M. 2005. Two proteins with 6-cys motifs are required for malarial parasites to commit to infection of the hepatocyte. *Mol Microbiol*, 58, 1264-75.
- JANSSEN, C. S., PHILLIPS, R. S., TURNER, C. M. & BARRETT, M. P. 2004. *Plasmodium* interspersed repeats: the major multigene superfamily of malaria parasites. *Nucleic Acids Res*, 32, 5712-20.
- JEFFERY, G. M. 1966. Epidemiological significance of repeated infections with homologous and heterologous strains and species of *Plasmodium*. *Bull World Health Organ*, 35, 873-82.
- JIANG, J., BARNWELL, J. W., MEYER, E. V. & GALINSKI, M. R. 2013. *Plasmodium vivax* merozoite surface protein-3 (PvMSP3): expression of an 11 member multigene family in blood-stage parasites. *PLoS One*, 8, e63888.
- KABANOVA, S., KLEINBONGARD, P., VOLKMER, J., ANDREE, B., KELM, M. & JAX, T. W. 2009. Gene expression analysis of human red blood cells. *Int J Med Sci*, 6, 156-9.
- KADEKOPPALA, M. & HOLDER, A. A. 2010. Merozoite surface proteins of the malaria parasite: the MSP1 complex and the MSP7 family. *Int J Parasitol*, 40, 1155-61.
- KADEKOPPALA, M., OGUN, S. A., HOWELL, S., GUNARATNE, R. S. & HOLDER, A. A. 2010. Systematic genetic analysis of the *Plasmodium falciparum* MSP7-like family reveals differences in protein expression, location, and importance in asexual growth of the blood-stage parasite. *Eukaryot Cell*, 9, 1064-74.
- KAFSACK, B. F., ROVIRA-GRAELLS, N., CLARK, T. G., BANCELLS, C., CROWLEY, V. M., CAMPINO, S. G., WILLIAMS, A. E., DROUGHT, L. G., KWIATKOWSKI, D. P., BAKER, D. A., CORTES, A. & LLINAS, M. 2014. A transcriptional switch underlies commitment to sexual development in malaria parasites. *Nature*, 507, 248-52.
- KARUNAWEEERA, N. D., GRAU, G. E., GAMAGE, P., CARTER, R. & MENDIS, K. N. 1992. Dynamics of fever and serum levels of tumor necrosis factor are closely associated during clinical paroxysms in *Plasmodium vivax* malaria. *Proc Natl Acad Sci U S A*, 89, 3200-3.
- KELLAR, K. L., KALWAR, R. R., DUBOIS, K. A., CROUSE, D., CHAFIN, W. D. & KANE, B. E. 2001. Multiplexed fluorescent bead-based immunoassays for quantitation of human cytokines in serum and culture supernatants. *Cytometry*, 45, 27-36.
- KERR, J. S. & WRIGHT, G. J. 2012. Avidity-based extracellular interaction screening (AVEXIS) for the scalable detection of low-affinity extracellular receptor-ligand interactions. *J Vis Exp*, e3881.
- KHAN, S. M., FRANKE-FAYARD, B., MAIR, G. R., LASONDER, E., JANSE, C. J., MANN, M. & WATERS, A. P. 2005. Proteome analysis of separated male and female gametocytes reveals novel sex-specific *Plasmodium* biology. *Cell*, 121, 675-87.
- KIM, D., PERTEA, G., TRAPNELL, C., PIMENTEL, H., KELLEY, R. & SALZBERG, S. L. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*, 14, R36.
- KING, C. L., MICHON, P., SHAKRI, A. R., MARCOTTY, A., STANISIC, D., ZIMMERMAN, P. A., COLE-TOBIAN, J. L., MUELLER, I. & CHITNIS, C. E. 2008. Naturally acquired Duffy-binding protein-specific binding inhibitory antibodies confer protection from blood-stage *Plasmodium vivax* infection. *Proc Natl Acad Sci U S A*, 105, 8363-8.
- KITCHEN, S. F. 1938. The infection of reticulocytes by *Plasmodium vivax*. *The American journal of tropical medicine and hygiene*, 18, 347-359.

- KNUEPFER, E., SULEYMAN, O., DLUZEWSKI, A. R., STRASCHIL, U., O'KEEFFE, A. H., OGUN, S. A., GREEN, J. L., GRAINGER, M., TEWARI, R. & HOLDER, A. A. 2014. RON12, a novel Plasmodium-specific rhoptry neck protein important for parasite proliferation. *Cell Microbiol*, 16, 657-72.
- KOCHAR, D. K., SAXENA, V., SINGH, N., KOCHAR, S. K., KUMAR, S. V. & DAS, A. 2005. Plasmodium vivax malaria. *Emerg Infect Dis*, 11, 132-134.
- KOCKEN, C. H., WITHERS-MARTINEZ, C., DUBBELD, M. A., VAN DER WEL, A., HACKETT, F., VALDERRAMA, A., BLACKMAN, M. J. & THOMAS, A. W. 2002. High-level expression of the malaria blood-stage vaccine candidate Plasmodium falciparum apical membrane antigen 1 and induction of antibodies that inhibit erythrocyte invasion. *Infect Immun*, 70, 4471-6.
- KOCKEN, C. H. M., DUBBELD, M. A., VAN DER WEL, A., PRONK, J. T., WATERS, A. P., LANGERMANS, J. A. M. & THOMAS, A. W. 1999. High-level expression of Plasmodium vivax apical membrane antigen 1 (AMA-1) in Pichia pastoris: Strong immunogenicity in Macaca mulatta immunized with P-vivax AMA-1 and adjuvant SBAS2. *Infection and immunity*, 67, 43-49.
- KOEPFLI, C., COLBORN, K. L., KINIBORO, B., LIN, E., SPEED, T. P., SIBA, P. M., FELGER, I. & MUELLER, I. 2013. A high force of plasmodium vivax blood-stage infection drives the rapid acquisition of immunity in papua new guinean children. *PLoS Negl Trop Dis*, 7, e2403.
- KOSSODO, S., MONSO, C., JUILLARD, P., VELU, T., GOLDMAN, M. & GRAU, G. E. 1997. Interleukin-10 modulates susceptibility in experimental cerebral malaria. *Immunology*, 91, 536-40.
- KRAEMER, S. M. & SMITH, J. D. 2006. A family affair: var genes, PfEMP1 binding, and malaria disease. *Curr Opin Microbiol*, 9, 374-80.
- KREMSNER, P. G., WINKLER, S., BRANDTS, C., WILDLING, E., JENNE, L., GRANINGER, W., PRADA, J., BIENZLE, U., JUILLARD, P. & GRAU, G. E. 1995. Prediction of accelerated cure in Plasmodium falciparum malaria by the elevated capacity of tumor necrosis factor production. *Am J Trop Med Hyg*, 53, 532-8.
- KROTOSKI, W. A. 1985. Discovery of the hypnozoite and a new theory of malarial relapse. *Trans R Soc Trop Med Hyg*, 79, 1-11.
- KROTOSKI, W. A., COLLINS, W. E., BRAY, R. S., GARNHAM, P. C., COGSWELL, F. B., GWADZ, R. W., KILLICK-KENDRICK, R., WOLF, R., SINDEN, R., KOONTZ, L. C. & STANFILL, P. S. 1982. Demonstration of hypnozoites in sporozoite-transmitted Plasmodium vivax infection. *The American journal of tropical medicine and hygiene*, 31, 1291-1293.
- KUMAR, S., MELZER, M., DODDS, P., WATSON, J. & ORD, R. 2007. P. vivax malaria complicated by shock and ARDS. *Scand J Infect Dis*, 39, 255-256.
- KWIATKOWSKI, D., MOLYNEUX, M. E., STEPHENS, S., CURTIS, N., KLEIN, N., POINTAIRE, P., SMIT, M., ALLAN, R., BREWSTER, D. R., GRAU, G. E. & ET AL. 1993. Anti-TNF therapy inhibits fever in cerebral malaria. *Q J Med*, 86, 91-8.
- KYES, S., HORROCKS, P. & NEWBOLD, C. 2001. Antigenic variation at the infected red cell surface in malaria. *Annu Rev Microbiol*, 55, 673-707.
- KYES, S., PINCHES, R. & NEWBOLD, C. 2000. A simple RNA analysis method shows var and rif multigene family expression patterns in Plasmodium falciparum. *Mol Biochem Parasitol*, 105, 311-315.
- LACOUNT, D. J., VIGNALI, M., CHETTIER, R., PHANSALKAR, A., BELL, R., HESSELBERTH, J. R., SCHOENFELD, L. W., OTA, I., SAHASRABUDHE, S., KURSCHNER, C., FIELDS, S. & HUGHES, R. E. 2005. A protein interaction network of the malaria parasite Plasmodium falciparum. *Nature*, 438, 103-7.
- LAMARQUE, M. H., ROQUES, M., KONG-HAP, M., TONKIN, M. L., RUGARABAMU, G., MARQ, J. B., PENARETE-VARGAS, D. M., BOULANGER, M. J., SOLDATI-FAVRE, D. & LEBRUN, M. 2014. Plasticity and redundancy among AMA-RON pairs ensure host cell entry of Toxoplasma parasites. *Nat Commun*, 5, 4098.

- LANCA, E. F., MAGALHAES, B. M., VITOR-SILVA, S., SIQUEIRA, A. M., BENZECRY, S. G., ALEXANDRE, M. A., O'BRIEN, C., BASSAT, Q. & LACERDA, M. V. 2012. Risk factors and characterization of Plasmodium vivax-associated admissions to pediatric intensive care units in the Brazilian Amazon. *PLoS One*, 7, e35406.
- LANDER, E. S., LINTON, L. M., BIRREN, B., NUSBAUM, C., ZODY, M. C., BALDWIN, J., DEVON, K., DEWAR, K., DOYLE, M., FITZHUGH, W., FUNKE, R., GAGE, D., HARRIS, K., HEAFORD, A., HOWLAND, J., KANN, L., LEHOCZKY, J., LEVINE, R., MCEWAN, P., MCKERNAN, K., MELDRIM, J., MESIROV, J. P., MIRANDA, C., MORRIS, W., NAYLOR, J., RAYMOND, C., ROSETTI, M., SANTOS, R., SHERIDAN, A., SOUGNEZ, C., STANGETHOMANN, Y., STOJANOVIC, N., SUBRAMANIAN, A., WYMAN, D., ROGERS, J., SULSTON, J., AINSCOUGH, R., BECK, S., BENTLEY, D., BURTON, J., CLEE, C., CARTER, N., COULSON, A., DEADMAN, R., DELOUKAS, P., DUNHAM, A., DUNHAM, I., DURBIN, R., FRENCH, L., GRAFHAM, D., GREGORY, S., HUBBARD, T., HUMPHRAY, S., HUNT, A., JONES, M., LLOYD, C., MCMURRAY, A., MATTHEWS, L., MERCER, S., MILNE, S., MULLIKIN, J. C., MUNGALL, A., PLUMB, R., ROSS, M., SHOWNKEEN, R., SIMS, S., WATERSTON, R. H., WILSON, R. K., HILLIER, L. W., MCPHERSON, J. D., MARRA, M. A., MARDIS, E. R., FULTON, L. A., CHINWALLA, A. T., PEPIN, K. H., GISH, W. R., CHISSOE, S. L., WENDL, M. C., DELEHAUNTY, K. D., MINER, T. L., DELEHAUNTY, A., KRAMER, J. B., COOK, L. L., FULTON, R. S., JOHNSON, D. L., MINX, P. J., CLIFTON, S. W., HAWKINS, T., BRANSCOMB, E., PREDKI, P., RICHARDSON, P., WENNING, S., SLEZAK, T., DOGGETT, N., CHENG, J. F., OLSEN, A., LUCAS, S., ELKIN, C., UBERBACHER, E., FRAZIER, M., et al. 2001. Initial sequencing and analysis of the human genome. *Nature*, 409, 860-921.
- LANGHORNE, J., NDUNGU, F. M., SPONAAS, A.-M. & MARSH, K. 2008. Immunity to malaria: more questions than answers. *Nature immunology*, 9, 725-32.
- LASONDER, E., JANSE, C. J., VAN GEMERT, G. J., MAIR, G. R., VERMUNT, A. M., DOURADINHA, B. G., VAN NOORT, V., HUYNEN, M. A., LUTY, A. J., KROEZE, H., KHAN, S. M., SAUERWEIN, R. W., WATERS, A. P., MANN, M. & STUNNENBERG, H. G. 2008. Proteomic profiling of Plasmodium sporozoite maturation identifies new proteins essential for parasite development and infectivity. *PLoS Pathog*, 4, e1000195.
- LE ROCH, K. G., ZHOU, Y., BLAIR, P. L., GRAINGER, M., MOCH, J. K., HAYNES, J. D., DE LA VEGA, P., HOLDER, A. A., BATALOV, S., CARUCCI, D. J. & WINZELER, E. A. 2003. Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science*, 301, 1503-8.
- LEECH, J. H., BARNWELL, J. W., MILLER, L. H. & HOWARD, R. J. 1984. Identification of a strain-specific malarial antigen exposed on the surface of Plasmodium falciparum-infected erythrocytes. *J Exp Med*, 159, 1567-75.
- LEVINE, N. D. 1988. Progress in taxonomy of the Apicomplexan protozoa. *J Protozool*, 35, 518-20.
- LI, H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, 27, 2987-93.
- LI, H., HANDSAKER, B., WYSOKER, A., FENNEL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G., DURBIN, R. & GENOME PROJECT DATA PROCESSING, S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25, 2078-9.
- LI, J., ITO, D., CHEN, J. H., LU, F., CHENG, Y., WANG, B., HA, K. S., CAO, J., TORII, M., SATTABONGKOT, J., TSUBOI, T. & HAN, E. T. 2012. Pv12, a 6-Cys antigen of Plasmodium vivax, is localized to the merozoite rhoptry. *Parasitology international*, 61, 443-9.
- LI, J., MATSUOKA, H., MITAMURA, T. & HORII, T. 2002. Characterization of proteases involved in the processing of Plasmodium falciparum serine repeat antigen (SERA). *Mol Biochem Parasitol*, 120, 177-86.
- LIM, C., HANSEN, E., DESIMONE, T. M., MORENO, Y., JUNKER, K., BEI, A., BRUGNARA, C., BUCKEE, C. O. & DURAISINGH, M. T. 2013. Expansion of host cellular niche can drive adaptation of a zoonotic malaria parasite to humans. *Nat Commun*, 4, 1638.

- LIMA-JUNIOR, J. C., JIANG, J., RODRIGUES-DA-SILVA, R. N., BANIC, D. M., TRAN, T. M., RIBEIRO, R. Y., MEYER, V. S., DE-SIMONE, S. G., SANTOS, F., MORENO, A., BARNWELL, J. W., GALINSKI, M. R. & OLIVEIRA-FERREIRA, J. 2011. B cell epitope mapping and characterization of naturally acquired antibodies to the *Plasmodium vivax* merozoite surface protein-3alpha (PvMSP-3alpha) in malaria exposed individuals from Brazilian Amazon. *Vaccine*, 29, 1801-11.
- LIMA-JUNIOR, J. C., RODRIGUES-DA-SILVA, R. N., BANIC, D. M., JIANG, J., SINGH, B., FABRICIO-SILVA, G. M., PORTO, L. C., MEYER, E. V., MORENO, A., RODRIGUES, M. M., BARNWELL, J. W., GALINSKI, M. R. & DE OLIVEIRA-FERREIRA, J. 2012. Influence of HLA-DRB1 and HLA-DQB1 alleles on IgG antibody response to the *P. vivax* MSP-1, MSP-3alpha and MSP-9 in individuals from Brazilian endemic area. *PLoS One*, 7, e36419.
- LIN, C. S., UBOLDI, A. D., MARAPANA, D., CZABOTAR, P. E., EPP, C., BUJARD, H., TAYLOR, N. L., PERUGINI, M. A., HODDER, A. N. & COWMAN, A. F. 2014. The merozoite surface protein 1 complex is a platform for binding to human erythrocytes by *Plasmodium falciparum*. *J Biol Chem*, 289, 25655-69.
- LIN, E., KINIBORO, B., GRAY, L., DOBBIE, S., ROBINSON, L., LAUMAEA, A., SCHOPFLIN, S., STANISIC, D., BETUELA, I., BLOOD-ZIKURSH, M., SIBA, P., FELGER, I., SCHOFIELD, L., ZIMMERMAN, P. & MUELLER, I. 2010. Differential patterns of infection and disease with *P. falciparum* and *P. vivax* in young Papua New Guinean children. *PLoS One*, 5, e9047.
- LLANOS-CUENTAS, A., LACERDA, M. V., RUEANGWEERAYUT, R., KRUDSOOD, S., GUPTA, S. K., KOCHAR, S. K., ARTHUR, P., CHUENCHOM, N., MOHRLE, J. J., DUPARC, S., UGWUEGBULAM, C., KLEIM, J. P., CARTER, N., GREEN, J. A. & KELLAM, L. 2014. Tafenoquine plus chloroquine for the treatment and relapse prevention of *Plasmodium vivax* malaria (DETECTIVE): a multicentre, double-blind, randomised, phase 2b dose-selection study. *Lancet*, 383, 1049-58.
- LLINAS, M., BOZDECH, Z., WONG, E. D., ADAI, A. T. & DERISI, J. L. 2006. Comparative whole genome transcriptome analysis of three *Plasmodium falciparum* strains. *Nucleic Acids Res*, 34, 1166-73.
- LONGLEY, R. J., SATTABONGKOT, J. & MUELLER, I. 2016. Insights into the naturally acquired immune response to *Plasmodium vivax* malaria. *Parasitology*, 143, 154-70.
- LOPATICKI, S., MAIER, A. G., THOMPSON, J., WILSON, D. W., THAM, W. H., TRIGLIA, T., GOUT, A., SPEED, T. P., BEESON, J. G., HEALER, J. & COWMAN, A. F. 2011. Reticulocyte and erythrocyte binding-like proteins function cooperatively in invasion of human erythrocytes by malaria parasites. *Infection and immunity*, 79, 1107-17.
- LU, F., LI, J., WANG, B., CHENG, Y., KONG, D. H., CUI, L., HA, K. S., SATTABONGKOT, J., TSUBOI, T. & HAN, E. T. 2014. Profiling the humoral immune responses to *Plasmodium vivax* infection and identification of candidate immunogenic rhoptry-associated membrane antigen (RAMA). *J Proteomics*, 102, 66-82.
- LUSE, S. A. & MILLER, L. H. 1971. *Plasmodium falciparum* malaria. Ultrastructure of parasitized erythrocytes in cardiac vessels. *The American journal of tropical medicine and hygiene*, 20, 655-660.
- MACKINNON, M. J., LI, J., MOK, S., KORTOK, M. M., MARSH, K., PREISER, P. R. & BOZDECH, Z. 2009. Comparative transcriptional and genomic analysis of *Plasmodium falciparum* field isolates. *PLoS Pathog*, 5, e1000644.
- MAIER, A. G., DURAISINGH, M. T., REEDER, J. C., PATEL, S. S., KAZURA, J. W., ZIMMERMAN, P. A. & COWMAN, A. F. 2002. *Plasmodium falciparum* erythrocyte invasion through glycophorin C and selection for Gerbich negativity in human populations. *Nat Med*, 9, 87-92.
- MAITLAND, K., WILLIAMS, T. N., BENNETT, S., NEWBOLD, C. I., PETO, T. E., VIJI, J., TIMOTHY, R., CLEGG, J. B., WEATHERALL, D. J. & BOWDEN, D. K. 1996. The interaction between *Plasmodium falciparum* and *P. vivax* in children on Espiritu Santo island, Vanuatu. *Trans R Soc Trop Med Hyg*, 90, 614-20.

- MALKIN, E. M., DURBIN, A. P., DIEMERT, D. J., SATTABONGKOT, J., WU, Y., MIURA, K., LONG, C. A., LAMBERT, L., MILES, A. P., WANG, J., STOWERS, A., MILLER, L. H. & SAUL, A. 2005. Phase 1 vaccine trial of Pvs25H: a transmission blocking vaccine for *Plasmodium vivax* malaria. *Vaccine*, 23, 3131-8.
- MALPEDE, B. M. & TOLIA, N. H. 2014. Malaria adhesins: structure and function. *Cell Microbiol*, 16, 621-31.
- MANSKE, M., MIOTTO, O., CAMPINO, S., AUBURN, S., ALMAGRO-GARCIA, J., MASLEN, G., O'BRIEN, J., DJIMDE, A., DOUMBO, O., ZONGO, I., OUEDRAOGO, J. B., MICHON, P., MUELLER, I., SIBA, P., NZILA, A., BORRMANN, S., KIARA, S. M., MARSH, K., JIANG, H., SU, X. Z., AMARATUNGA, C., FAIRHURST, R., SOCHEAT, D., NOSTEN, F., IMWONG, M., WHITE, N. J., SANDERS, M., ANASTASI, E., ALCOCK, D., DRURY, E., OYOLA, S., QUAIL, M. A., TURNER, D. J., RUANO-RUBIO, V., JYOTHI, D., AMENGA-ETEGO, L., HUBBART, C., JEFFREYS, A., ROWLANDS, K., SUTHERLAND, C., ROPER, C., MANGANO, V., MODIANO, D., TAN, J. C., FERDIG, M. T., AMAMBUA-NGWA, A., CONWAY, D. J., TAKALA-HARRISON, S., PLOWE, C. V., RAYNER, J. C., ROCKETT, K. A., CLARK, T. G., NEWBOLD, C. I., BERRIMAN, M., MACINNIS, B. & KWIATKOWSKI, D. P. 2012. Analysis of *Plasmodium falciparum* diversity in natural infections by deep sequencing. *Nature*, 487, 375-9.
- MARIONI, J. C., MASON, C. E., MANE, S. M., STEPHENS, M. & GILAD, Y. 2008. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res*, 18, 1509-17.
- MARTIN, S., SOLLNER, C., CHAROENSAWAN, V., ADRYAN, B., THISSE, B., THISSE, C., TEICHMANN, S. & WRIGHT, G. J. 2010. Construction of a large extracellular protein interaction network and its resolution by spatiotemporal expression profiling. *Mol Cell Proteomics*, 9, 2654-65.
- MAUDE, R. J., NGUON, C., LY, P., BUNKEA, T., NGOR, P., CANAVATI DE LA TORRE, S. E., WHITE, N. J., DONDORP, A. M., DAY, N. P., WHITE, L. J. & CHUOR, C. M. 2014. Spatial and temporal epidemiology of clinical malaria in Cambodia 2004-2013. *Malar J*, 13, 385.
- MAYER, D. C., COFIE, J., JIANG, L., HARTL, D. L., TRACY, E., KABAT, J., MENDOZA, L. H. & MILLER, L. H. 2009. Glycophorin B is the erythrocyte receptor of *Plasmodium falciparum* erythrocyte-binding ligand, EBL-1. *Proc Natl Acad Sci U S A*, 106, 5348-52.
- MAYER, D. C., MU, J. B., KANEKO, O., DUAN, J., SU, X. Z. & MILLER, L. H. 2004. Polymorphism in the *Plasmodium falciparum* erythrocyte-binding ligand JESEBL/EBA-181 alters its receptor specificity. *Proc Natl Acad Sci U S A*, 101, 2518-23.
- MCFADDEN, G. I., REITH, M. E., MUNHOLLAND, J. & LANG-UNNASCH, N. 1996. Plastid in human parasites. *Nature*, 381, 482.
- MCGREGOR, I. A. 1964. The Passive Transfer of Human Malarial Immunity. *Am J Trop Med Hyg*, 13, SUPPL 237-9.
- MEHLIN, C., BONI, E., BUCKNER, F. S., ENGEL, L., FEIST, T., GELB, M. H., HAJI, L., KIM, D., LIU, C., MUELLER, N., MYLER, P. J., REDDY, J. T., SAMPSON, J. N., SUBRAMANIAN, E., VAN VOORHIS, W. C., WORTHEY, E., ZUCKER, F. & HOL, W. G. 2006. Heterologous expression of proteins from *Plasmodium falciparum*: results from 1000 genes. *Mol Biochem Parasitol*, 148, 144-60.
- MENARD, D., BARNADAS, C., BOUCHIER, C., HENRY-HALLDIN, C., GRAY, L. R., RATSIMBASOA, A., THONIER, V., CAROD, J. F., DOMARLE, O., COLIN, Y., BERTRAND, O., PICOT, J., KING, C. L., GRIMBERG, B. T., MERCEREAU-PUIJALON, O. & ZIMMERMAN, P. A. 2010. *Plasmodium vivax* clinical malaria is commonly observed in Duffy-negative Malagasy people. *Proc Natl Acad Sci U S A*, 107, 5967-5971.
- MENARD, D., CHAN, E. R., BENEDET, C., RATSIMBASOA, A., KIM, S., CHIM, P., DO, C., WITKOWSKI, B., DURAND, R., THELLIER, M., SEVERINI, C., LEGRAND, E., MUSSET, L., NOUR, B. Y., MERCEREAU-PUIJALON, O., SERRE, D. & ZIMMERMAN, P. A. 2013. Whole genome sequencing of field isolates reveals a common duplication of the Duffy binding protein gene in Malagasy *Plasmodium vivax* strains. *PLoS Negl Trop Dis*, 7, e2489.

- MENDES, C., DIAS, F., FIGUEIREDO, J., MORA, V. G., CANO, J., DE SOUSA, B., DO ROSARIO, V. E., BENITO, A., BERZOSA, P. & AREZ, A. P. 2011. Duffy negative antigen is no longer a barrier to *Plasmodium vivax*--molecular evidences from the African West Coast (Angola and Equatorial Guinea). *PLoS neglected tropical diseases*, 5, e1192.
- MENDIS, K., SINA, B. J., MARCHESINI, P. & CARTER, R. 2001. The neglected burden of *Plasmodium vivax* malaria. *The American journal of tropical medicine and hygiene*, 64, 97-106.
- MENDONCA, V. R., QUEIROZ, A. T., LOPES, F. M., ANDRADE, B. B. & BARRAL-NETTO, M. 2013. Networking the host immune response in *Plasmodium vivax* malaria. *Malar J*, 12, 69.
- MICHON, P., COLE-TOBIAN, J. L., DABOD, E., SCHOEPFLIN, S., IGU, J., SUSAPU, M., TARONGKA, N., ZIMMERMAN, P. A., REEDER, J. C., BEESON, J. G., SCHOFIELD, L., KING, C. L. & MUELLER, I. 2007. The risk of malarial infections and disease in Papua New Guinean children. *Am J Trop Med Hyg*, 76, 997-1008.
- MICHON, P., FRASER, T. & ADAMS, J. H. 2000. Naturally acquired and vaccine-elicited antibodies block erythrocyte cytoadherence of the *Plasmodium vivax* Duffy binding protein. *Infect Immun*, 68, 3164-3171.
- MICHON, P. A., AREVALO-HERRERA, M., FRASER, T., HERRERA, S. & ADAMS, J. H. 1998. Serologic responses to recombinant *Plasmodium vivax* Duffy binding protein in a Colombian village. *Am J Trop Med Hyg*, 59, 597-9.
- MIKOLAJCZAK, S. A., SILVA-RIVERA, H., PENG, X., TARUN, A. S., CAMARGO, N., JACOBS-LORENA, V., DALY, T. M., BERGMAN, L. W., DE LA VEGA, P., WILLIAMS, J., ALY, A. S. & KAPPE, S. H. 2008. Distinct malaria parasite sporozoites reveal transcriptional changes that cause differential tissue infection competence in the mosquito vector and mammalian host. *Mol Cell Biol*, 28, 6196-207.
- MIKOLAJCZAK, S. A., VAUGHAN, A. M., KANGWANRANGSAN, N., ROOBSOONG, W., FISHBAUGHER, M., YIMAMNUAYCHOK, N., REZAKHANI, N., LAKSHMANAN, V., SINGH, N., KAUSHANSKY, A., CAMARGO, N., BALDWIN, M., LINDNER, S. E., ADAMS, J. H., SATTABONGKOT, J. & KAPPE, S. H. 2015. *Plasmodium vivax* liver stage development and hypnozoite persistence in human liver-chimeric mice. *Cell Host Microbe*, 17, 526-35.
- MILLAR, S. B. & COX-SINGH, J. 2015. Human infections with *Plasmodium knowlesi*-zoonotic malaria. *Clin Microbiol Infect*.
- MILLER, L. H., MASON, S. J., CLYDE, D. F. & MCGINNISS, M. H. 1976. The resistance factor to *Plasmodium vivax* in blacks. The Duffy-blood-group genotype, FyFy. *N Engl J Med*, 295, 302-304.
- MIURA, K., KEISTER, D. B., MURATOVA, O. V., SATTABONGKOT, J., LONG, C. A. & SAUL, A. 2007. Transmission-blocking activity induced by malaria vaccine candidates Pfs25/Pvs25 is a direct and predictable function of antibody titer. *Malar J*, 6, 107.
- MIURA, K., ORCUTT, A. C., MURATOVA, O. V., MILLER, L. H., SAUL, A. & LONG, C. A. 2008. Development and characterization of a standardized ELISA including a reference serum on each plate to detect antibodies induced by experimental malaria vaccines. *Vaccine*, 26, 193-200.
- MIZUTANI, M., IYORI, M., BLAGBOROUGH, A. M., FUKUMOTO, S., FUNATSU, T., SINDEN, R. E. & YOSHIDA, S. 2014. Baculovirus-vectored multistage *Plasmodium vivax* vaccine induces both protective and transmission-blocking immunities against transgenic rodent malaria parasites. *Infect Immun*, 82, 4348-57.
- MOLINA, D. M., FINNEY, O. C., AREVALO-HERRERA, M., HERRERA, S., FELGNER, P. L., GARDNER, M. J., LIANG, X. & WANG, R. 2012. *Plasmodium vivax* pre-erythrocytic-stage antigen discovery: exploiting naturally acquired humoral responses. *Am J Trop Med Hyg*, 87, 460-9.
- MONGUI, A., ANGEL, D. I., GUZMAN, C., VANEGAS, M. & PATARROYO, M. A. 2008. Characterisation of the *Plasmodium vivax* Pv38 antigen. *Biochemical and biophysical research communications*, 376, 326-30.

- MONS, B., COLLINS, W. E., SKINNER, J. C., VAN DER STAR, W., CROON, J. J. & VAN DER KAAJ, H. J. 1988. Plasmodium vivax: in vitro growth and reinvasion in red blood cells of Aotus nancymai. *Exp Parasitol*, 66, 183-188.
- MOON, R. W., HALL, J., RANGKUTI, F., HO, Y. S., ALMOND, N., MITCHELL, G. H., PAIN, A., HOLDER, A. A. & BLACKMAN, M. J. 2013. Adaptation of the genetically tractable malaria pathogen Plasmodium knowlesi to continuous culture in human erythrocytes. *Proc Natl Acad Sci U S A*, 110, 531-6.
- MORAHAN, B. J., WANG, L. & COPPEL, R. L. 2009. No TRAP, no invasion. *Trends Parasitol*, 25, 77-84.
- MORENO-PEREZ, D. A., AREIZA-ROJAS, R., FLOREZ-BUITRAGO, X., SILVA, Y., PATARROYO, M. E. & PATARROYO, M. A. 2013. The GPI-anchored 6-Cys protein Pv12 is present in detergent-resistant microdomains of Plasmodium vivax blood stage schizonts. *Protist*, 164, 37-48.
- MORENO-PEREZ, D. A., MONGUI, A., SOLER, L. N., SANCHEZ-LADINO, M. & PATARROYO, M. A. 2011. Identifying and characterizing a member of the RhopH1/Clag family in Plasmodium vivax. *Gene*, 481, 17-23.
- MORTAZAVI, A., WILLIAMS, B. A., MCCUE, K., SCHAEFFER, L. & WOLD, B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*, 5, 621-8.
- MUELLER, I., GALINSKI, M. R., BAIRD, J. K., CARLTON, J. M., KOCHAR, D. K., ALONSO, P. L. & DEL PORTILLO, H. A. 2009a. Key gaps in the knowledge of Plasmodium vivax, a neglected human malaria parasite. *Lancet Infect Dis*, 9, 555-566.
- MUELLER, I., SCHOEPFLIN, S., SMITH, T. A., BENTON, K. L., BRETSCHER, M. T., LIN, E., KINIBORO, B., ZIMMERMAN, P. A., SPEED, T. P., SIBA, P. & FELGER, I. 2012. Force of infection is key to understanding the epidemiology of Plasmodium falciparum malaria in Papua New Guinean children. *Proc Natl Acad Sci U S A*, 109, 10030-5.
- MUELLER, I., SHAKRI, A. R. & CHITNIS, C. E. 2015. Development of vaccines for Plasmodium vivax malaria. *Vaccine*, 33, 7489-95.
- MUELLER, I., WIDMER, S., MICHEL, D., MARAGA, S., MCNAMARA, D. T., KINIBORO, B., SIE, A., SMITH, T. A. & ZIMMERMAN, P. A. 2009b. High sensitivity detection of Plasmodium species reveals positive correlations between infections of different species, shifts in age distribution and reduced local variation in Papua New Guinea. *Malar J*, 8, 41.
- MURRAY, C. J., ROSENFELD, L. C., LIM, S. S., ANDREWS, K. G., FOREMAN, K. J., HARING, D., FULLMAN, N., NAGHAVI, M., LOZANO, R. & LOPEZ, A. D. 2012. Global malaria mortality between 1980 and 2010: a systematic analysis. *Lancet*, 379, 413-31.
- NEAFSEY, D. E., GALINSKY, K., JIANG, R. H. Y., YOUNG, L., SYKES, S. M., SAIF, S., GUJJA, S., GOLDBERG, J. M., YOUNG, S., ZENG, Q., CHAPMAN, S. B., DASH, A. P., ANVIKAR, A. R., SUTTON, P. L., BIRREN, B. W., ESCALANTE, A. A., BARNWELL, J. W. & CARLTON, J. M. 2012. The malaria parasite Plasmodium vivax exhibits greater genetic diversity than Plasmodium falciparum. *Nature Genetics*.
- NGASSA MBENDA, H. G. & DAS, A. 2014. Molecular evidence of Plasmodium vivax mono and mixed malaria parasite infections in Duffy-negative native Cameroonians. *PLoS One*, 9, e103262.
- NKHOMA, E. T., POOLE, C., VANNAPPAGARI, V., HALL, S. A. & BEUTLER, E. 2009. The global prevalence of glucose-6-phosphate dehydrogenase deficiency: a systematic review and meta-analysis. *Blood Cells, Molecules and Diseases*, 42, 267-278.
- NOULIN, F., BORLON, C., VAN DEN ABBEELE, J., D'ALESSANDRO, U. & ERHART, A. 2013. 1912-2012: a century of research on Plasmodium vivax in vitro culture. *Trends Parasitol*, 29, 286-94.
- NTUMNGIA, F. B., BARNES, S. J., MCHENRY, A. M., GEORGE, M. T., SCHLOEGEL, J. & ADAMS, J. H. 2014. Immunogenicity of a synthetic vaccine based on Plasmodium vivax Duffy binding protein region II. *Clin Vaccine Immunol*, 21, 1215-23.

- NURLEILA, S., SYAFRUDDIN, D., ELYAZAR, I. R. & BAIRD, J. K. 2012. Serious and fatal illness associated with falciparum and vivax malaria among patients admitted to hospital at West Sumba in eastern Indonesia. *Am J Trop Med Hyg*, 87, 41-9.
- O'DONNELL, R. A., HACKETT, F., HOWELL, S. A., TREECK, M., STRUCK, N., KRNAJSKI, Z., WITHERS-MARTINEZ, C., GILBERGER, T. W. & BLACKMAN, M. J. 2006. Intramembrane proteolysis mediates shedding of a key adhesin during erythrocyte invasion by the malaria parasite. *J Cell Biol*, 174, 1023-33.
- OGUTU, B. R., APOLLO, O. J., MCKINNEY, D., OKOTH, W., SIANGLA, J., DUBOVSKY, F., TUCKER, K., WAITUMBI, J. N., DIGGS, C., WITTES, J., MALKIN, E., LEACH, A., SOISSON, L. A., MILMAN, J. B., OTIENO, L., HOLLAND, C. A., POLHEMUS, M., REMICH, S. A., OCKENHOUSE, C. F., COHEN, J., BALLOU, W. R., MARTIN, S. K., ANGOV, E., STEWART, V. A., LYON, J. A., HEPPNER, D. G. & WITHERS, M. R. 2009. Blood stage malaria vaccine eliciting high antigen-specific antibody concentrations confers no protection to young children in Western Kenya. *PLoS One*, 4, e4708.
- OLIVEIRA, T. R., FERNANDEZ-BECERRA, C., JIMENEZ, M. C., DEL PORTILLO, H. A. & SOARES, I. S. 2006. Evaluation of the acquired immune responses to *Plasmodium vivax* VIR variant antigens in individuals living in malaria-endemic areas of Brazil. *Malar J*, 5, 83.
- OLIVEIRA-FERREIRA, J., VARGAS-SERRATO, E., BARNWELL, J. W., MORENO, A. & GALINSKI, M. R. 2004. Immunogenicity of *Plasmodium vivax* merozoite surface protein-9 recombinant proteins expressed in *E. coli*. *Vaccine*, 22, 2023-2030.
- OSIER, F. H., MACKINNON, M. J., CROSNIER, C., FEGAN, G., KAMUYU, G., WANAGURU, M., OGADA, E., MCDADE, B., RAYNER, J. C., WRIGHT, G. J. & MARSH, K. 2014. New antigens for a multicomponent blood-stage malaria vaccine. *Sci Transl Med*, 6, 247ra102.
- OTHORO, C., LAL, A. A., NAHLEN, B., KOECH, D., ORAGO, A. S. & UDHAYAKUMAR, V. 1999. A low interleukin-10 tumor necrosis factor-alpha ratio is associated with malaria anemia in children residing in a holoendemic malaria region in western Kenya. *J Infect Dis*, 179, 279-82.
- OTTO, T. D., BOHME, U., JACKSON, A. P., HUNT, M., FRANKE-FAYARD, B., HOEIJMAKERS, W. A., RELIGA, A. A., ROBERTSON, L., SANDERS, M., OGUN, S. A., CUNNINGHAM, D., ERHART, A., BILLKER, O., KHAN, S. M., STUNNENBERG, H. G., LANGHORNE, J., HOLDER, A. A., WATERS, A. P., NEWBOLD, C. I., PAIN, A., BERRIMAN, M. & JANSE, C. J. 2014. A comprehensive evaluation of rodent malaria parasite genomes and gene expression. *BMC Biol*, 12, 86.
- OTTO, T. D., WILINSKI, D., ASSEFA, S., KEANE, T. M., SARRY, L. R., BOHME, U., LEMIEUX, J., BARRELL, B., PAIN, A., BERRIMAN, M., NEWBOLD, C. & LLINAS, M. 2010. New insights into the blood-stage transcriptome of *Plasmodium falciparum* using RNA-Seq. *Mol Microbiol*, 76, 12-24.
- PACHEBAT, J. A., LING, I. T., GRAINGER, M., TRUCCO, C., HOWELL, S., FERNANDEZ-REYES, D., GUNARATNE, R. & HOLDER, A. A. 2001. The 22 kDa component of the protein complex on the surface of *Plasmodium falciparum* merozoites is derived from a larger precursor, merozoite surface protein 7. *Mol Biochem Parasitol*, 117, 83-89.
- PACIFIC MALARIA INITIATIVE SURVEY GROUP ON BEHALF OF THE MINISTRIES OF HEALTH OF, V. & SOLOMON, I. 2010. Malaria on isolated Melanesian islands prior to the initiation of malaria elimination activities. *Malar J*, 9, 218.
- PAIN, A., BOHME, U., BERRY, A. E., MUNGALL, K., FINN, R. D., JACKSON, A. P., MOURIER, T., MISTRY, J., PASINI, E. M., ASLETT, M. A., BALASUBRAMMANIAM, S., BORGHARDT, K., BROOKS, K., CARRET, C., CARVER, T. J., CHEREVACH, I., CHILLINGWORTH, T., CLARK, T. G., GALINSKI, M. R., HALL, N., HARPER, D., HARRIS, D., HAUSER, H., IVENS, A., JANSSEN, C. S., KEANE, T., LARKE, N., LAPP, S., MARTI, M., MOULE, S., MEYER, I. M., ORMOND, D., PETERS, N., SANDERS, M., SANDERS, S., SARGEANT, T. J., SIMMONDS, M., SMITH, F., SQUARES, R., THURSTON, S., TIVEY, A. R., WALKER, D., WHITE, B., ZUIDERWIJK, E., CHURCHER, C., QUAIL, M. A., COWMAN, A. F., TURNER, C. M., RAJANDREAM, M. A., KOCKEN, C. H., THOMAS, A. W., NEWBOLD, C. I.,

- BARRELL, B. G. & BERRIMAN, M. 2008. The genome of the simian and human malaria parasite *Plasmodium knowlesi*. *Nature*, 455, 799-803.
- PAING, M. M. & TOLIA, N. H. 2014. Multimeric assembly of host-pathogen adhesion complexes involved in apicomplexan invasion. *PLoS Pathog*, 10, e1004120.
- PAN, Q., SHAI, O., LEE, L. J., FREY, B. J. & BLENCOWE, B. J. 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet*, 40, 1413-5.
- PANICHAKUL, T., SATTABONGKOT, J., CHOTIVANICH, K., SIRICHAISINTHOP, J., CUI, L. & UDOMSANGPETCH, R. 2007. Production of erythropoietic cells in vitro for continuous culture of *Plasmodium vivax*. *Int J Parasitol*, 37, 1551-1557.
- PATH 2011. Staying the course? Malaria research and development in a time of economic uncertainty. Seattle.
- PEARSON, R. D., AMATO, R., AUBURN, S., MIOTTO, O., ALMAGRO-GARCIA, J., AMARATUNGA, C., SUON, S., MAO, S., NOVIYANTI, R., TRIMARSANTO, H., MARFURT, J., ANSTEY, N. M., WILLIAM, T., BONI, M. F., DOLECEK, C., TRAN, H. T., WHITE, N. J., MICHON, P., SIBA, P., TAVUL, L., HARRISON, G., BARRY, A., MUELLER, I., FERREIRA, M. U., KARUNAWEEERA, N., RANDRIANARIVELOJOSIA, M., GAO, Q., HUBBART, C., HART, L., JEFFERY, B., DRURY, E., MEAD, D., KEKRE, M., CAMPINO, S., MANSKE, M., CORNELIUS, V. J., MACINNIS, B., ROCKETT, K. A., MILES, A., RAYNER, J. C., FAIRHURST, R. M., NOSTEN, F., PRICE, R. N. & KWIATKOWSKI, D. P. 2016. Genomic analysis of local variation and recent evolution in *Plasmodium vivax*. *Nat Genet*.
- PERERA, K. L., HANDUNNETTI, S. M., HOLM, I., LONGACRE, S. & MENDIS, K. 1998. Baculovirus merozoite surface protein 1 C-terminal recombinant antigens are highly protective in a natural primate model for human *Plasmodium vivax* malaria. *Infect Immun*, 66, 1500-6.
- PEREZ-LEAL, O., SIERRA, A. Y., BARRERO, C. A., MONCADA, C., MARTINEZ, P., CORTES, J., LOPEZ, Y., SALAZAR, L. M., HOEBEKE, J. & PATARROYO, M. A. 2005. Identifying and characterising the *Plasmodium falciparum* merozoite surface protein 10 *Plasmodium vivax* homologue. *Biochem Biophys Res Commun*, 331, 1178-84.
- PERKINS, D. J., WEINBERG, J. B. & KREMSNER, P. G. 2000. Reduced interleukin-12 and transforming growth factor-beta1 in severe childhood malaria: relationship of cytokine balance with disease severity. *J Infect Dis*, 182, 988-92.
- PERRIN, A. J., BARTHOLDSON, S. J. & WRIGHT, G. J. 2015. P-selectin is a host receptor for *Plasmodium* MSP7 ligands. *Malar J*, 14, 238.
- PHIMPRAPHI, W., PAUL, R. E., YIMSAMRAN, S., PUANGSA-ART, S., THANYAVANICH, N., MANEEBOONYANG, W., PROMMONGKOL, S., SORNKLOM, S., CHAIMUNGKUN, W., CHAVEZ, I. F., BLANC, H., LOOAREESUWAN, S., SAKUNTABHAI, A. & SINGHASIVANON, P. 2008. Longitudinal study of *Plasmodium falciparum* and *Plasmodium vivax* in a Karen population in Thailand. *Malar J*, 7, 99.
- PILLAI, A. D., ADDO, R., SHARMA, P., NGUITRAGOOL, W., SRINIVASAN, P. & DESAI, S. A. 2013. Malaria parasites tolerate a broad range of ionic environments and do not require host cation remodelling. *Mol Microbiol*, 88, 20-34.
- PIRIOU, E., KIMMEL, R., CHELIMO, K., MIDDELDORP, J. M., ODADA, P. S., PLOUTZ-SNYDER, R., MOORMANN, A. M. & ROCHFORD, R. 2009. Serological evidence for long-term Epstein-Barr virus reactivation in children living in a holoendemic malaria region of Kenya. *J Med Virol*, 81, 1088-93.
- PIZARRO, J. C., VULLIEZ-LE NORMAND, B., CHESNE-SECK, M. L., COLLINS, C. R., WITHERS-MARTINEZ, C., HACKETT, F., BLACKMAN, M. J., FABER, B. W., REMARQUE, E. J., KOCKEN, C. H., THOMAS, A. W. & BENTLEY, G. A. 2005. Crystal structure of the malaria vaccine candidate apical membrane antigen 1. *Science*, 308, 408-11.

- PRABA-EGGE, A. D., MONTENEGRO, S., AREVALO-HERRERA, M., HOPPER, T., HERRERA, S. & JAMES, M. A. 2003. Human cytokine responses to meso-endemic malaria on the Pacific Coast of Colombia. *Ann Trop Med Parasitol*, 97, 327-37.
- PRICE, R. N., DOUGLAS, N. M. & ANSTEY, N. M. 2009. New developments in *Plasmodium vivax* malaria: severe disease and the rise of chloroquine resistance. *Curr Opin Infect Dis*, 22, 430-5.
- PRICE, R. N., TJITRA, E., GUERRA, C. A., YEUNG, S., WHITE, N. J. & ANSTEY, N. M. 2007. Vivax malaria: neglected and not benign. *The American journal of tropical medicine and hygiene*, 77, 79-87.
- PROELLOCKS, N. I., KOVACEVIC, S., FERGUSON, D. J., KATS, L. M., MORAHAN, B. J., BLACK, C. G., WALLER, K. L. & COPPEL, R. L. 2007. *Plasmodium falciparum* Pf34, a novel GPI-anchored rhoptry protein found in detergent-resistant microdomains. *International journal for parasitology*, 37, 1233-41.
- PRUDENCIO, M., RODRIGUEZ, A. & MOTA, M. M. 2006. The silent path to thousands of merozoites: the *Plasmodium* liver stage. *Nat Rev Microbiol*, 4, 849-56.
- QUINLAN, A. R. & HALL, I. M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26, 841-2.
- RAYNER, J. C., GALINSKI, M. R., INGRAVALLO, P. & BARNWELL, J. W. 2000. Two *Plasmodium falciparum* genes express merozoite proteins that are related to *Plasmodium vivax* and *Plasmodium yoelii* adhesive proteins involved in host cell selection and invasion. *Proc Natl Acad Sci U S A*, 97, 9648-9653.
- RAYNER, J. C., VARGAS-SERRATO, E., HUBER, C. S., GALINSKI, M. R. & BARNWELL, J. W. 2001. A *Plasmodium falciparum* homologue of *Plasmodium vivax* reticulocyte binding protein (PvRBP1) defines a trypsin-resistant erythrocyte invasion pathway. *J Exp Med*, 194, 1571-81.
- REDDY, K. S., AMLABU, E., PANDEY, A. K., MITRA, P., CHAUHAN, V. S. & GAUR, D. 2015. Multiprotein complex between the GPI-anchored CyRPA with PfrH5 and PfrRipr is crucial for *Plasmodium falciparum* erythrocyte invasion. *Proc Natl Acad Sci U S A*, 112, 1179-84.
- REDDY, K. S., PANDEY, A. K., SINGH, H., SAHAR, T., EMMANUEL, A., CHITNIS, C. E., CHAUHAN, V. S. & GAUR, D. 2014. Bacterially expressed full-length recombinant *Plasmodium falciparum* RH5 protein binds erythrocytes and elicits potent strain-transcending parasite-neutralizing antibodies. *Infect Immun*, 82, 152-64.
- RENIA, L., GRILLOT, D., MARUSSIG, M., CORRADIN, G., MILTGEN, F., LAMBERT, P. H., MAZIER, D. & DEL GIUDICE, G. 1993. Effector functions of circumsporozoite peptide-primed CD4+ T cell clones against *Plasmodium yoelii* liver stages. *J Immunol*, 150, 1471-8.
- RICE, B. L., ACOSTA, M. M., PACHECO, M. A., CARLTON, J. M., BARNWELL, J. W. & ESCALANTE, A. A. 2014. The origin and diversification of the merozoite surface protein 3 (msp3) multi-gene family in *Plasmodium vivax* and related parasites. *Mol Phylogenet Evol*, 78, 172-84.
- RICHARDS, J. S., ARUMUGAM, T. U., REILING, L., HEALER, J., HODDER, A. N., FOWKES, F. J., CROSS, N., LANGER, C., TAKEO, S., UBOLDI, A. D., THOMPSON, J. K., GILSON, P. R., COPPEL, R. L., SIBA, P. M., KING, C. L., TORII, M., CHITNIS, C. E., NARUM, D. L., MUELLER, I., CRABB, B. S., COWMAN, A. F., TSUBOI, T. & BEESON, J. G. 2013. Identification and prioritization of merozoite antigens as targets of protective human immunity to *Plasmodium falciparum* malaria for vaccine and biomarker development. *J Immunol*, 191, 795-809.
- RIGLAR, D. T., RICHARD, D., WILSON, D. W., BOYLE, M. J., DEKIWADIA, C., TURNBULL, L., ANGRISANO, F., MARAPANA, D. S., ROGERS, K. L., WHITCHURCH, C. B., BEESON, J. G., COWMAN, A. F., RALPH, S. A. & BAUM, J. 2011. Super-resolution dissection of coordinated events during malaria parasite invasion of the human erythrocyte. *Cell Host Microbe*, 9, 9-20.

- RIGLAR, D. T., WHITEHEAD, L., COWMAN, A. F., ROGERS, K. L. & BAUM, J. 2016. Localisation-based imaging of malarial antigens during erythrocyte entry reaffirms a role for AMA1 but not MTRAP in invasion. *J Cell Sci*, 129, 228-42.
- ROBERTS, R. J., CARNEIRO, M. O. & SCHATZ, M. C. 2013. The advantages of SMRT sequencing. *Genome Biol*, 14, 405.
- ROBINSON, M. D., MCCARTHY, D. J. & SMYTH, G. K. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26, 139-40.
- RODRIGUES, M. H., RODRIGUES, K. M., OLIVEIRA, T. R., COMODO, A. N., RODRIGUES, M. M., KOCKEN, C. H., THOMAS, A. W. & SOARES, I. S. 2005. Antibody response of naturally infected individuals to recombinant Plasmodium vivax apical membrane antigen-1. *Int J Parasitol*, 35, 185-92.
- RODRIGUES-DA-SILVA, R. N., LIMA-JUNIOR JDA, C., FONSECA BDE, P., ANTAS, P. R., BALDEZ, A., STORER, F. L., SANTOS, F., BANIC, D. M. & OLIVEIRA-FERREIRA, J. 2014. Alterations in cytokines and haematological parameters during the acute and convalescent phases of Plasmodium falciparum and Plasmodium vivax infections. *Mem Inst Oswaldo Cruz*, 109, 154-62.
- RODRIGUEZ, M., LUSTIGMAN, S., MONTERO, E., OKSOV, Y. & LOBO, C. A. 2008. PfRH5: a novel reticulocyte-binding family homolog of plasmodium falciparum that binds to the erythrocyte, and an investigation of its receptor. *PLoS One*, 3, e3300.
- ROGERS, W. O., GOWDA, K. & HOFFMAN, S. L. 1999. Construction and immunogenicity of DNA vaccine plasmids encoding four Plasmodium vivax candidate vaccine antigens. *Vaccine*, 17, 3136-44.
- ROGERSON, S. J. & CARTER, R. 2008. Severe vivax malaria: newly recognised or rediscovered. *PLoS Med*, 5, e136.
- ROOBSOONG, W., THARINJAROEN, C. S., RACHAPHAEW, N., CHOBSON, P., SCHOFIELD, L., CUI, L., ADAMS, J. H. & SATTABONGKOT, J. 2015. Improvement of culture conditions for long-term in vitro culture of Plasmodium vivax. *Malar J*, 14, 297.
- ROSENBERG, R. & RUNGSIWONGSE, J. 1991. The number of sporozoites produced by individual malaria oocysts. *Am J Trop Med Hyg*, 45, 574-7.
- RTS, S. C. T. P. 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. *Lancet*, 386, 31-45.
- RUG, M., PRESCOTT, S. W., FERNANDEZ, K. M., COOKE, B. M. & COWMAN, A. F. 2006. The role of KAHRP domains in knob formation and cytoadherence of P falciparum-infected human erythrocytes. *Blood*, 108, 370-8.
- RUSSELL, B., SUWANARUSK, R., BORLON, C., COSTA, F. T., CHU, C. S., RIJKEN, M. J., SRIPRAWAT, K., WARTER, L., KOH, E. G., MALLERET, B., COLIN, Y., BERTRAND, O., ADAMS, J. H., D'ALESSANDRO, U., SNOUNOU, G., NOSTEN, F. & RENIA, L. 2011. A reliable ex vivo invasion assay of human reticulocytes by Plasmodium vivax. *Blood*.
- RUSSELL, C., MERCEREAU-PUIJALON, O., LE SCANF, C., STEWARD, M. & ARNOT, D. E. 2005. Further definition of PfEMP-1 DBL-1alpha domains mediating rosetting adhesion of Plasmodium falciparum. *Mol Biochem Parasitol*, 144, 109-13.
- RUTHERFORD, K., PARKHILL, J., CROOK, J., HORSNELL, T., RICE, P., RAJANDREAM, M. A. & BARRELL, B. 2000. Artemis: sequence visualization and annotation. *Bioinformatics*, 16, 944-5.
- RYAN, J. R., STOUTE, J. A., AMON, J., DUNTON, R. F., MTALIB, R., KOROS, J., OWOUR, B., LUCKHART, S., WIRTZ, R. A., BARNWELL, J. W. & ROSENBERG, R. 2006. Evidence for transmission of Plasmodium vivax among a duffy antigen negative population in Western Kenya. *The American journal of tropical medicine and hygiene*, 75, 575-581.

- SACCI, J. B., JR., ALAM, U., DOUGLAS, D., LEWIS, J., TYRRELL, D. L., AZAD, A. F. & KNETEMAN, N. M. 2006. Plasmodium falciparum infection and exoerythrocytic development in mice with chimeric human livers. *Int J Parasitol*, 36, 353-60.
- SAGARA, I., DICKO, A., ELLIS, R. D., FAY, M. P., DIAWARA, S. I., ASSADOU, M. H., SISSOKO, M. S., KONE, M., DIALLO, A. I., SAYE, R., GUINDO, M. A., KANTE, O., NIAMBELE, M. B., MIURA, K., MULLEN, G. E., PIERCE, M., MARTIN, L. B., DOLO, A., DIALLO, D. A., DOUMBO, O. K., MILLER, L. H. & SAUL, A. 2009. A randomized controlled phase 2 trial of the blood stage AMA1-C1/Alhydrogel malaria vaccine in children in Mali. *Vaccine*, 27, 3090-3098.
- SAM-YELLOWE, T. Y., SHIO, H. & PERKINS, M. E. 1988. Secretion of Plasmodium falciparum rhoptry protein into the plasma membrane of host erythrocytes. *J Cell Biol*, 106, 1507-13.
- SANDERS, P. R., GILSON, P. R., CANTIN, G. T., GREENBAUM, D. C., NEBL, T., CARUCCI, D. J., MCCONVILLE, M. J., SCHOFIELD, L., HODDER, A. N., YATES, J. R. & CRABB, B. S. 2005. Distinct protein classes including novel merozoite surface antigens in Raft-like membranes of Plasmodium falciparum. *J Biol Chem*, 280, 40169-76.
- SANGER, F., NICKLEN, S. & COULSON, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*, 74, 5463-7.
- SCHENA, M., SHALON, D., DAVIS, R. W. & BROWN, P. O. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, 270, 467-70.
- SCHWACH, F., BUSHELL, E., GOMES, A. R., ANAR, B., GIRLING, G., HERD, C., RAYNER, J. C. & BILLKER, O. 2015. PlasmoGEM, a database supporting a community resource for large-scale experimental genetics in malaria parasites. *Nucleic Acids Res*, 43, D1176-82.
- SCHWARTZ, L., BROWN, G. V., GENTON, B. & MOORTHY, V. S. 2012. A review of malaria vaccine clinical projects based on the WHO rainbow table. *Malar J*, 11, 11.
- SEDER, R. A., CHANG, L. J., ENAMA, M. E., ZEPHIR, K. L., SARWAR, U. N., GORDON, I. J., HOLMAN, L. A., JAMES, E. R., BILLINGSLEY, P. F., GUNASEKERA, A., RICHMAN, A., CHAKRAVARTY, S., MANOJ, A., VELMURUGAN, S., LI, M., RUBEN, A. J., LI, T., EAPPEN, A. G., STAFFORD, R. E., PLUMMER, S. H., HENDEL, C. S., NOVIK, L., COSTNER, P. J., MENDOZA, F. H., SAUNDERS, J. G., NASON, M. C., RICHARDSON, J. H., MURPHY, J., DAVIDSON, S. A., RICHIE, T. L., SEDEGAH, M., SUTAMIHARDJA, A., FAHLE, G. A., LYKE, K. E., LAURENS, M. B., ROEDERER, M., TEWARI, K., EPSTEIN, J. E., SIM, B. K., LEDGERWOOD, J. E., GRAHAM, B. S., HOFFMAN, S. L. & TEAM, V. R. C. S. 2013. Protection against malaria by intravenous immunization with a nonreplicating sporozoite vaccine. *Science*, 341, 1359-65.
- SHARMA, A. & KHANDURI, U. 2009. How benign is benign tertian malaria? *J Vector Borne Dis*, 46, 141-144.
- SHENDURE, J., PORRECA, G. J., REPPAS, N. B., LIN, X., MCCUTCHEON, J. P., ROSENBAUM, A. M., WANG, M. D., ZHANG, K., MITRA, R. D. & CHURCH, G. M. 2005. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science*, 309, 1728-32.
- SILVESTRI, F., BOZDECH, Z., LANFRANCOTTI, A., DI GIULIO, E., BULTRINI, E., PICCI, L., DERISI, J. L., PIZZI, E. & ALANO, P. 2005. Genome-wide identification of genes upregulated at the onset of gametocytogenesis in Plasmodium falciparum. *Mol Biochem Parasitol*, 143, 100-10.
- SIM, B. K. 1995. EBA-175: an erythrocyte-binding ligand of Plasmodium falciparum. *Parasitol Today*, 11, 213-7.
- SIM, B. K., CHITNIS, C. E., WASNIOWSKA, K., HADLEY, T. J. & MILLER, L. H. 1994. Receptor and ligand domains for invasion of erythrocytes by Plasmodium falciparum. *Science*, 264, 1941-4.
- SINDEN, R. E. G., H. M. 2002. The malaria parasites. In: WARREL, D. A. G., H. M. (ed.) *Essential malariology*. 4th ed. London, United Kingdom: Hodder Arnold.

- SINGH, A. P., PURI, S. K. & CHITNIS, C. E. 2002. Antibodies raised against receptor-binding domain of *Plasmodium knowlesi* Duffy binding protein inhibit erythrocyte invasion. *Mol Biochem Parasitol*, 121, 21-31.
- SINGH, B., KIM SUNG, L., MATUSOP, A., RADHAKRISHNAN, A., SHAMSUL, S. S., COX-SINGH, J., THOMAS, A. & CONWAY, D. J. 2004. A large focus of naturally acquired *Plasmodium knowlesi* infections in human beings. *Lancet*, 363, 1017-24.
- SINGH, S., ALAM, M. M., PAL-BHOWMICK, I., BRZOSTOWSKI, J. A. & CHITNIS, C. E. 2010. Distinct external signals trigger sequential release of apical organelles during erythrocyte invasion by malaria parasites. *PLoS Pathog*, 6, e1000746.
- SINGH, S., PANDEY, K., CHATTOPADHAYAY, R., YAZDANI, S. S., LYNN, A., BHARADWAJ, A., RANJAN, A. & CHITNIS, C. 2001. Biochemical, biophysical, and functional characterization of bacterially expressed and refolded receptor binding domain of *Plasmodium vivax* duffy-binding protein. *J Biol Chem*, 276, 17111-6.
- SINGH, S. K., HORA, R., BELRHALI, H., CHITNIS, C. E. & SHARMA, A. 2006. Structural basis for Duffy recognition by the malaria parasite Duffy-binding-like domain. *Nature*, 439, 741-4.
- SINHA, A., HUGHES, K. R., MODRZYNSKA, K. K., OTTO, T. D., PFANDER, C., DICKENS, N. J., RELIGA, A. A., BUSHHELL, E., GRAHAM, A. L., CAMERON, R., KAFSACK, B. F., WILLIAMS, A. E., LLINAS, M., BERRIMAN, M., BILLKER, O. & WATERS, A. P. 2014. A cascade of DNA-binding proteins for sexual commitment and development in *Plasmodium*. *Nature*, 507, 253-7.
- SMITH, D. L., COHEN, J. M., CHIYAKA, C., JOHNSTON, G., GETHING, P. W., GOSLING, R., BUCKEE, C. O., LAXMINARAYAN, R., HAY, S. I. & TATEM, A. J. 2013. A sticky situation: the unexpected stability of malaria elimination. *Philos Trans R Soc Lond B Biol Sci*, 368, 20120145.
- SMITH, J. D., CHITNIS, C. E., CRAIG, A. G., ROBERTS, D. J., HUDSON-TAYLOR, D. E., PETERSON, D. S., PINCHES, R., NEWBOLD, C. I. & MILLER, L. H. 1995. Switches in expression of *Plasmodium falciparum* var genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. *Cell*, 82, 101-10.
- SOLLNER, C. & WRIGHT, G. J. 2009. A cell surface interaction network of neural leucine-rich repeat receptors. *Genome Biol*, 10, R99.
- SOUZA-SILVA, F. A., DA SILVA-NUNES, M., SANCHEZ, B. A., CERAVOLO, I. P., MALAFRONTI, R. S., BRITO, C. F., FERREIRA, M. U. & CARVALHO, L. H. 2010. Naturally acquired antibodies to *Plasmodium vivax* Duffy binding protein (DBP) in Brazilian Amazon. *Am J Trop Med Hyg*, 82, 185-93.
- SPELLMAN, P. T., SHERLOCK, G., ZHANG, M. Q., IYER, V. R., ANDERS, K., EISEN, M. B., BROWN, P. O., BOTSTEIN, D. & FUTCHER, B. 1998. Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol Biol Cell*, 9, 3273-97.
- SRINIVASAN, P., BEATTY, W. L., DIOUF, A., HERRERA, R., AMBROGGIO, X., MOCH, J. K., TYLER, J. S., NARUM, D. L., PIERCE, S. K., BOOTHROYD, J. C., HAYNES, J. D. & MILLER, L. H. 2011. Binding of *Plasmodium* merozoite proteins RON2 and AMA1 triggers commitment to invasion. *Proc Natl Acad Sci U S A*, 108, 13275-80.
- ST LAURENT, B., MILLER, B., BURTON, T. A., AMARATUNGA, C., MEN, S., SOVANNAROTH, S., FAY, M. P., MIOTTO, O., GWADZ, R. W., ANDERSON, J. M. & FAIRHURST, R. M. 2016. Corrigendum: Artemisinin-resistant *Plasmodium falciparum* clinical isolates can infect diverse mosquito vectors of Southeast Asia and Africa. *Nat Commun*, 7, 10345.
- STAFFORD, W. H., GUNDER, B., HARRIS, A., HEIDRICH, H. G., HOLDER, A. A. & BLACKMAN, M. J. 1996. A 22 kDa protein associated with the *Plasmodium falciparum* merozoite surface protein-1 complex. *Mol Biochem Parasitol*, 80, 159-69.

- STANISIC, D. I., JAVATI, S., KINIBORO, B., LIN, E., JIANG, J., SINGH, B., MEYER, E. V., SIBA, P., KOEPLI, C., FELGER, I., GALINSKI, M. R. & MUELLER, I. 2013. Naturally acquired immune responses to *P. vivax* merozoite surface protein 3 α and merozoite surface protein 9 are associated with reduced risk of *P. vivax* malaria in young Papua New Guinean children. *PLoS Negl Trop Dis*, 7, e2498.
- STOWERS, A. W., ZHANG, Y., SHIMP, R. L. & KASLOW, D. C. 2001. Structural conformers produced during malaria vaccine production in yeast. *Yeast*, 18, 137-50.
- STURM, A., AMINO, R., VAN DE SAND, C., REGEN, T., RETZLAFF, S., RENNINGER, A., KRUEGER, A., POLLOK, J. M., MENARD, R. & HEUSSLER, V. T. 2006. Manipulation of host hepatocytes by the malaria parasite for delivery into liver sinusoids. *Science*, 313, 1287-90.
- SU, X. Z., HEATWOLE, V. M., WERTHEIMER, S. P., GUINET, F., HERRFELDT, J. A., PETERSON, D. S., RAVETCH, J. A. & WELLEMS, T. E. 1995. The large diverse gene family var encodes proteins involved in cytoadherence and antigenic variation of *Plasmodium falciparum*-infected erythrocytes. *Cell*, 82, 89-100.
- SUTHERLAND, C. J., TANOMSING, N., NOLDER, D., OGUIKE, M., JENNISON, C., PUKRITTAYAKAMEE, S., DOLECEK, C., HIEN, T. T., DO ROSARIO, V. E., AREZ, A. P., PINTO, J., MICHON, P., ESCALANTE, A. A., NOSTEN, F., BURKE, M., LEE, R., BLAZE, M., OTTO, T. D., BARNWELL, J. W., PAIN, A., WILLIAMS, J., WHITE, N. J., DAY, N. P., SNOUNOU, G., LOCKHART, P. J., CHIODINI, P. L., IMWONG, M. & POLLEY, S. D. 2010. Two nonrecombining sympatric forms of the human malaria parasite *Plasmodium ovale* occur globally. *J Infect Dis*, 201, 1544-50.
- SUWANARUSK, R., COOKE, B. M., DONDORP, A. M., SILAMUT, K., SATTABONGKOT, J., WHITE, N. J. & UDOMSANGPETCH, R. 2004. The deformability of red blood cells parasitized by *Plasmodium falciparum* and *P. vivax*. *J Infect Dis*, 189, 190-194.
- TAECHALERTPAISARN, T., CROSNIER, C., BARTHOLDSON, S. J., HODDER, A. N., THOMPSON, J., BUSTAMANTE, L. Y., WILSON, D. W., SANDERS, P. R., WRIGHT, G. J., RAYNER, J. C., COWMAN, A. F., GILSON, P. R. & CRABB, B. S. 2012. Biochemical and Functional Analysis of Two *Plasmodium falciparum* Blood-Stage 6-Cys Proteins: P12 and P41. *PLoS one*, 7, e41937.
- TARUN, A. S., BAER, K., DUMPIT, R. F., GRAY, S., LEJARCEGUI, N., FREVERT, U. & KAPPE, S. H. 2006. Quantitative isolation and in vivo imaging of malaria parasite liver stages. *Int J Parasitol*, 36, 1283-93.
- TARUN, A. S., PENG, X., DUMPIT, R. F., OGATA, Y., SILVA-RIVERA, H., CAMARGO, N., DALY, T. M., BERGMAN, L. W. & KAPPE, S. H. 2008. A combined transcriptome and proteome survey of malaria parasite liver stages. *Proc Natl Acad Sci USA*, 105, 305-10.
- TAVARES, J., FORMAGLIO, P., THIBERGE, S., MORDELET, E., VAN ROOIJEN, N., MEDVINSKY, A., MENARD, R. & AMINO, R. 2013. Role of host cell traversal by the malaria sporozoite during liver infection. *J Exp Med*, 210, 905-15.
- TAYLOR, H. M., GRAINGER, M. & HOLDER, A. A. 2002. Variation in the expression of a *Plasmodium falciparum* protein family implicated in erythrocyte invasion. *Infect Immun*, 70, 5779-89.
- TAYLOR, H. M., TRIGLIA, T., THOMPSON, J., SAJID, M., FOWLER, R., WICKHAM, M. E., COWMAN, A. F. & HOLDER, A. A. 2001. *Plasmodium falciparum* homologue of the genes for *Plasmodium vivax* and *Plasmodium yoelii* adhesive proteins, which is transcribed but not translated. *Infect Immun*, 69, 3635-45.
- TAYLOR, S. M., CERAMI, C. & FAIRHURST, R. M. 2013. Hemoglobinopathies: slicing the Gordian knot of *Plasmodium falciparum* malaria pathogenesis. *PLoS Pathog*, 9, e1003327.
- TEMPLETON, T. J. & KASLOW, D. C. 1999. Identification of additional members define a *Plasmodium falciparum* gene superfamily which includes Pfs48/45 and Pfs230. *Mol Biochem Parasitol*, 101, 223-7.
- THAM, W. H., HEALER, J. & COWMAN, A. F. 2012. Erythrocyte and reticulocyte binding-like proteins of *Plasmodium falciparum*. *Trends Parasitol*, 28, 23-30.

- TJITRA, E., ANSTEY, N. M., SUGIARTO, P., WARIKAR, N., KENANGALEM, E., KARYANA, M., LAMPAH, D. A. & PRICE, R. N. 2008. Multidrug-resistant *Plasmodium vivax* associated with severe and fatal malaria: a prospective study in Papua, Indonesia. *PLoS Med*, 5, e128.
- TOLIA, N. H., ENEMARK, E. J., SIM, B. K. & JOSHUA-TOR, L. 2005. Structural basis for the EBA-175 erythrocyte invasion pathway of the malaria parasite *Plasmodium falciparum*. *Cell*, 122, 183-93.
- TOM, R., BISSON, L. & DUROCHER, Y. 2008. Culture of HEK293-EBNA1 Cells for Production of Recombinant Proteins. *CSH Protoc*, 2008, pdb prot4976.
- TOMSCHY, A., FAUSER, C., LANDWEHR, R. & ENGEL, J. 1996. Homophilic adhesion of E-cadherin occurs by a co-operative two-step interaction of N-terminal domains. *EMBO J*, 15, 3507-14.
- TONKIN, M. L., ROQUES, M., LAMARQUE, M. H., PUGNIERE, M., DOUGUET, D., CRAWFORD, J., LEBRUN, M. & BOULANGER, M. J. 2011. Host cell invasion by apicomplexan parasites: insights from the co-structure of AMA1 with a RON2 peptide. *Science*, 333, 463-7.
- TRAGER, W. & JENSEN, J. B. 1976. Human malaria parasites in continuous culture. *Science*, 193, 673-5.
- TRAGER, W. & JENSEN, J. B. 1997. Continuous culture of *Plasmodium falciparum*: its impact on malaria research. *Int J Parasitol*, 27, 989-1006.
- TRAN, T. M., MORENO, A., YAZDANI, S. S., CHITNIS, C. E., BARNWELL, J. W. & GALINSKI, M. R. 2005. Detection of a *Plasmodium vivax* erythrocyte binding protein by flow cytometry. *Cytometry A*, 63, 59-66.
- TRAPNELL, C., ROBERTS, A., GOFF, L., PERTEA, G., KIM, D., KELLEY, D. R., PIMENTEL, H., SALZBERG, S. L., RINN, J. L. & PACHTER, L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc*, 7, 562-78.
- TRAPNELL, C., WILLIAMS, B. A., PERTEA, G., MORTAZAVI, A., KWAN, G., VAN BAREN, M. J., SALZBERG, S. L., WOLD, B. J. & PACHTER, L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*, 28, 511-5.
- TRIEU, A., KAYALA, M. A., BURK, C., MOLINA, D. M., FREILICH, D. A., RICHIE, T. L., BALDI, P., FELGNER, P. L. & DOOLAN, D. L. 2011. Sterile protective immunity to malaria is associated with a panel of novel *P. falciparum* antigens. *Mol Cell Proteomics*, 10, M111 007948.
- TRIGLIA, T., THOMPSON, J., CARUANA, S. R., DELORENZI, M., SPEED, T. & COWMAN, A. F. 2001. Identification of proteins from *Plasmodium falciparum* that are homologous to reticulocyte binding proteins in *Plasmodium vivax*. *Infect Immun*, 69, 1084-1092.
- TRUCCO, C., FERNANDEZ-REYES, D., HOWELL, S., STAFFORD, W. H., SCOTT-FINNIGAN, T. J., GRAINGER, M., OGUN, S. A., TAYLOR, W. R. & HOLDER, A. A. 2001. The merozoite surface protein 6 gene codes for a 36 kDa protein associated with the *Plasmodium falciparum* merozoite surface protein-1 complex. *Mol Biochem Parasitol*, 112, 91-101.
- TRUNG, H. D., BORTEL, W. V., SOCHANATHA, T., KEOKENCHANH, K., BRIET, O. J. & COOSEMANS, M. 2005. Behavioural heterogeneity of *Anopheles* species in ecologically different localities in Southeast Asia: a challenge for vector control. *Trop Med Int Health*, 10, 251-62.
- TRUNG, H. D., VAN BORTEL, W., SOCHANATHA, T., KEOKENCHANH, K., QUANG, N. T., CONG, L. D. & COOSEMANS, M. 2004. Malaria transmission and major malaria vectors in different geographical areas of Southeast Asia. *Trop Med Int Health*, 9, 230-7.
- TSUBOI, T., TAKEO, S., IRIKO, H., JIN, L., TSUCHIMOCHI, M., MATSUDA, S., HAN, E. T., OTSUKI, H., KANEKO, O., SATTABONGKOT, J., UDOMSANGPETCH, R., SAWASAKI, T., TORII, M. & ENDO, Y. 2008. Wheat germ cell-free system-based production of malaria proteins for discovery of novel vaccine candidates. *Infect Immun*, 76, 1702-8.
- TURNER, L., LAVSTSEN, T., BERGER, S. S., WANG, C. W., PETERSEN, J. E., AVRIL, M., BRAZIER, A. J., FREETH, J., JESPERSEN, J. S., NIELSEN, M. A., MAGISTRADO, P.,

- LUSINGU, J., SMITH, J. D., HIGGINS, M. K. & THEANDER, T. G. 2013. Severe malaria is associated with parasite binding to endothelial protein C receptor. *Nature*, 498, 502-5.
- TYLER, J. S., TREECK, M. & BOOTHROYD, J. C. 2011. Focus on the ringleader: the role of AMA1 in apicomplexan invasion and replication. *Trends Parasitol*, 27, 410-20.
- UCHIME, O., HERRERA, R., REITER, K., KOTOVA, S., SHIMP, R. L., JR., MIURA, K., JONES, D., LEBOWITZ, J., AMBROGGIO, X., HURT, D. E., JIN, A. J., LONG, C., MILLER, L. H. & NARUM, D. L. 2012. Analysis of the conformation and function of the Plasmodium falciparum merozoite proteins MTRAP and PTRAMP. *Eukaryot Cell*, 11, 615-25.
- VALDERRAMA-AGUIRRE, A., QUINTERO, G., GOMEZ, A., CASTELLANOS, A., PEREZ, Y., MENDEZ, F., AREVALO-HERRERA, M. & HERRERA, S. 2005. Antigenicity, immunogenicity, and protective efficacy of Plasmodium vivax MSP1 PV2001: a potential malaria vaccine subunit. *Am J Trop Med Hyg*, 73, 16-24.
- VAN DER MERWE, P. A., BARCLAY, A. N., MASON, D. W., DAVIES, E. A., MORGAN, B. P., TONE, M., KRISHNAM, A. K., IANELLI, C. & DAVIS, S. J. 1994. Human cell-adhesion molecule CD2 binds CD58 (LFA-3) with a very low affinity and an extremely fast dissociation rate but does not bind CD48 or CD59. *Biochemistry*, 33, 10149-10160.
- VAN DIJK, M. R., DOURADINHA, B., FRANKE-FAYARD, B., HEUSSLER, V., VAN DOOREN, M. W., VAN SCHAIJK, B., VAN GEMERT, G. J., SAUERWEIN, R. W., MOTA, M. M., WATERS, A. P. & JANSE, C. J. 2005. Genetically attenuated, P36p-deficient malarial sporozoites induce protective immunity and apoptosis of infected liver cells. *Proc Natl Acad Sci U S A*, 102, 12194-9.
- VAN DIJK, M. R., VAN SCHAIJK, B. C., KHAN, S. M., VAN DOOREN, M. W., RAMESAR, J., KACZANOWSKI, S., VAN GEMERT, G. J., KROEZE, H., STUNNENBERG, H. G., ELING, W. M., SAUERWEIN, R. W., WATERS, A. P. & JANSE, C. J. 2010. Three members of the 6-cys protein family of Plasmodium play a role in gamete fertility. *PLoS Pathog*, 6, e1000853.
- VAN OOIJ, C., WITHERS-MARTINEZ, C., RINGEL, A., COCKCROFT, S., HALDAR, K. & BLACKMAN, M. J. 2013. Identification of a Plasmodium falciparum phospholipid transfer protein. *J Biol Chem*, 288, 31971-83.
- VANBUSKIRK, K. M., COLE-TOBIAN, J. L., BAISOR, M., SEVOVA, E. S., BOCKARIE, M., KING, C. L. & ADAMS, J. H. 2004a. Antigenic drift in the ligand domain of Plasmodium vivax duffy binding protein confers resistance to inhibitory antibodies. *J Infect Dis*, 190, 1556-1562.
- VANBUSKIRK, K. M., SEVOVA, E. & ADAMS, J. H. 2004b. Conserved residues in the Plasmodium vivax Duffy-binding protein ligand domain are critical for erythrocyte receptor recognition. *Proc Natl Acad Sci U S A*, 101, 15754-15759.
- VANDERBERG, J. P. & FREVERT, U. 2004. Intravital microscopy demonstrating antibody-mediated immobilisation of Plasmodium berghei sporozoites injected into skin by mosquitoes. *Int J Parasitol*, 34, 991-6.
- VAUGHAN, A. M., MIKOLAJCZAK, S. A., WILSON, E. M., GROMPE, M., KAUSHANSKY, A., CAMARGO, N., BIAL, J., PLOSS, A. & KAPPE, S. H. 2012. Complete Plasmodium falciparum liver-stage development in liver-chimeric mice. *J Clin Invest*, 122, 3618-28.
- VEDADI, M., LEW, J., ARTZ, J., AMANI, M., ZHAO, Y., DONG, A., WASNEY, G. A., GAO, M., HILLS, T., BROKX, S., QIU, W., SHARMA, S., DIASSITI, A., ALAM, Z., MELONE, M., MULICHAK, A., WERNIMONT, A., BRAY, J., LOPPNAU, P., PLOTNIKOVA, O., NEWBERRY, K., SUNDARARAJAN, E., HOUSTON, S., WALKER, J., TEMPEL, W., BOCHKAREV, A., KOZIERADZKI, I., EDWARDS, A., ARROWSMITH, C., ROOS, D., KAIN, K. & HUI, R. 2007. Genome-scale protein expression and structural biology of Plasmodium falciparum and related Apicomplexan organisms. *Mol Biochem Parasitol*, 151, 100-10.
- VOLZ, J. C., YAP, A., SISQUELLA, X., THOMPSON, J. K., LIM, N. T., WHITEHEAD, L. W., CHEN, L., LAMPE, M., THAM, W. H., WILSON, D., NEBL, T., MARAPANA, D., TRIGLIA, T., WONG, W., ROGERS, K. L. & COWMAN, A. F. 2016. Essential Role of the PfRh5/PfRipr/CyRPA Complex during Plasmodium falciparum Invasion of Erythrocytes. *Cell Host Microbe*, 20, 60-71.

- VULLIEZ-LE NORMAND, B., TONKIN, M. L., LAMARQUE, M. H., LANGER, S., HOOS, S., ROQUES, M., SAUL, F. A., FABER, B. W., BENTLEY, G. A., BOULANGER, M. J. & LEBRUN, M. 2012. Structural and functional insights into the malaria parasite moving junction complex. *PLoS Pathog*, 8, e1002755.
- WALLER, R. F. & MCFADDEN, G. I. 2005. The apicoplast: a review of the derived plastid of apicomplexan parasites. *Curr Issues Mol Biol*, 7, 57-79.
- WALTHER, M., WOODRUFF, J., EDELE, F., JEFFRIES, D., TONGREN, J. E., KING, E., ANDREWS, L., BEJON, P., GILBERT, S. C., DE SOUZA, J. B., SINDEN, R., HILL, A. V. & RILEY, E. M. 2006. Innate immune responses to human malaria: heterogeneous cytokine responses to blood-stage *Plasmodium falciparum* correlate with parasitological and clinical outcomes. *J Immunol*, 177, 5736-45.
- WALTMANN, A., DARCY, A. W., HARRIS, I., KOEPFLI, C., LODO, J., VAHI, V., PIZIKI, D., SHANKS, G. D., BARRY, A. E., WHITTAKER, M., KAZURA, J. W. & MUELLER, I. 2015. High Rates of Asymptomatic, Sub-microscopic *Plasmodium vivax* Infection and Disappearing *Plasmodium falciparum* Malaria in an Area of Low Transmission in Solomon Islands. *PLoS Negl Trop Dis*, 9, e0003758.
- WANG, E. T., SANDBERG, R., LUO, S., KHREBTUKOVA, I., ZHANG, L., MAYR, C., KINGSMORE, S. F., SCHROTH, G. P. & BURGE, C. B. 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature*, 456, 470-6.
- WANG, Z., GERSTEIN, M. & SNYDER, M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*, 10, 57-63.
- WEISS, G. E., CRABB, B. S. & GILSON, P. R. 2016. Overlaying Molecular and Temporal Aspects of Malaria Parasite Invasion. *Trends Parasitol*.
- WEISS, G. E., GILSON, P. R., TAECHALERTPAISARN, T., THAM, W. H., DE JONG, N. W., HARVEY, K. L., FOWKES, F. J., BARLOW, P. N., RAYNER, J. C., WRIGHT, G. J., COWMAN, A. F. & CRABB, B. S. 2015. Revealing the sequence and resulting cellular morphology of receptor-ligand interactions during *Plasmodium falciparum* invasion of erythrocytes. *PLoS Pathog*, 11, e1004670.
- WEISS, W. R., MELLOUK, S., HOUGHTEN, R. A., SEDEGAH, M., KUMAR, S., GOOD, M. F., BERZOFKY, J. A., MILLER, L. H. & HOFFMAN, S. L. 1990. Cytotoxic T cells recognize a peptide from the circumsporozoite protein on malaria-infected hepatocytes. *J Exp Med*, 171, 763-73.
- WERTHEIMER, S. P. & BARNWELL, J. W. 1989. *Plasmodium vivax* interaction with the human Duffy blood group glycoprotein: identification of a parasite receptor-like protein. *Exp Parasitol*, 69, 340-350.
- WESTENBERGER, S. J., MCCLEAN, C. M., CHATTOPADHYAY, R., DHARIA, N. V., CARLTON, J. M., BARNWELL, J. W., COLLINS, W. E., HOFFMAN, S. L., ZHOU, Y., VINETZ, J. M. & WINZELER, E. A. 2010. A systems-based analysis of *Plasmodium vivax* lifecycle transcription from human to mosquito. *PLoS Negl Trop Dis*, 4, e653.
- WHITE, N. J. & IMWONG, M. 2012. Relapse. *Adv Parasitol*, 80, 113-50.
- WHITE, N. J., PUKRITTAYAKAMEE, S., HIEN, T. T., FAIZ, M. A., MOKUOLU, O. A. & DONDORP, A. M. 2014. Malaria. *Lancet*, 383, 723-35.
- WHITFIELD, M. L., SHERLOCK, G., SALDANHA, A. J., MURRAY, J. I., BALL, C. A., ALEXANDER, K. E., MATESE, J. C., PEROU, C. M., HURT, M. M., BROWN, P. O. & BOTSTEIN, D. 2002. Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol Biol Cell*, 13, 1977-2000.
- WHO 1999. Rolling Back Malaria. *The World Health Report*.
- WHO 2010. Global report on Antimalarial Drug efficacy and Drug Resistance: 2000-2010.
- WHO 2014. *World Malaria Report 2014*, Geneva, Switzerland, World Health Organisation.

- WHO 2015. *Guidelines for the Treatment of Malaria*. 3rd ed. Geneva.
- WICKRAMARACHCHI, T., DEVI, Y. S., MOHMMED, A. & CHAUHAN, V. S. 2008. Identification and characterization of a novel *Plasmodium falciparum* merozoite apical protein involved in erythrocyte binding and invasion. *PLoS one*, 3, e1732.
- WICKRAMARACHCHI, T., PREMARATNE, P. H., PERERA, K. L., BANDARA, S., KOCKEN, C. H., THOMAS, A. W., HANDUNNETTI, S. M. & UDAGAMA-RANDENIYA, P. V. 2006. Natural human antibody responses to *Plasmodium vivax* apical membrane antigen 1 under low transmission and unstable malaria conditions in Sri Lanka. *Infect Immun*, 74, 798-801.
- WILHELM, B. T., MARGUERAT, S., WATT, S., SCHUBERT, F., WOOD, V., GOODHEAD, I., PENKETT, C. J., ROGERS, J. & BAHLER, J. 2008. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature*, 453, 1239-43.
- WILLIAMS, T. N., MAITLAND, K., PHELPS, L., BENNETT, S., PETO, T. E., VIJI, J., TIMOTHY, R., CLEGG, J. B., WEATHERALL, D. J. & BOWDEN, D. K. 1997. *Plasmodium vivax*: a cause of malnutrition in young children. *QJM*, 90, 751-757.
- WILLIAMSON, K. C. 2003. Pfs230: from malaria transmission-blocking vaccine candidate toward function. *Parasite Immunol*, 25, 351-9.
- WITHERS-MARTINEZ, C., SUAREZ, C., FULLE, S., KHER, S., PENZO, M., EBEJER, J. P., KOUSSIS, K., HACKETT, F., JIRGENSONS, A., FINN, P. & BLACKMAN, M. J. 2012. *Plasmodium* subtilisin-like protease 1 (SUB1): insights into the active-site structure, specificity and function of a pan-malaria drug target. *Int J Parasitol*, 42, 597-612.
- WOLDEAREGAI, T. G., KREMSNER, P. G., KUN, J. F. & MORDMULLER, B. 2013. *Plasmodium vivax* malaria in Duffy-negative individuals from Ethiopia. *Trans R Soc Trop Med Hyg*, 107, 328-31.
- WOODBERRY, T., MINIGO, G., PIERA, K. A., HANLEY, J. C., DE SILVA, H. D., SALWATI, E., KENANGALEM, E., TJITRA, E., COPPEL, R. L., PRICE, R. N., ANSTEY, N. M. & PLEBANSKI, M. 2008. Antibodies to *Plasmodium falciparum* and *Plasmodium vivax* merozoite surface protein 5 in Indonesia: species-specific and cross-reactive responses. *J Infect Dis*, 198, 134-42.
- WRIGHT, G. J. 2009. Signal initiation in biological systems: the properties and detection of transient extracellular protein interactions. *Mol Biosyst*, 5, 1405-12.
- WRIGHT, G. J., PUKLAVEC, M. J., WILLIS, A. C., HOEK, R. M., SEDGWICK, J. D., BROWN, M. H. & BARCLAY, A. N. 2000. Lymphoid/neuronal cell surface OX2 glycoprotein recognizes a novel receptor on macrophages implicated in the control of their function. *Immunity*, 13, 233-42.
- WU, Y., ELLIS, R. D., SHAFFER, D., FONTES, E., MALKIN, E. M., MAHANTY, S., FAY, M. P., NARUM, D., RAUSCH, K., MILES, A. P., AEBIG, J., ORCUTT, A., MURATOVA, O., SONG, G., LAMBERT, L., ZHU, D., MIURA, K., LONG, C., SAUL, A., MILLER, L. H. & DURBIN, A. P. 2008. Phase I trial of malaria transmission blocking vaccine candidates Pfs25 and Pvs25 formulated with montanide ISA 51. *PLoS One*, 3, e2636.
- XAINLI, J., COLE-TOBIAN, J. L., BAISOR, M., KASTENS, W., BOCKARIE, M., YAZDANI, S. S., CHITNIS, C. E., ADAMS, J. H. & KING, C. L. 2003. Epitope-specific humoral immunity to *Plasmodium vivax* Duffy binding protein. *Infect Immun*, 71, 2508-15.
- YADAVA, A., HALL, C. E., SULLIVAN, J. S., NACE, D., WILLIAMS, T., COLLINS, W. E., OCKENHOUSE, C. F. & BARNWELL, J. W. 2014. Protective efficacy of a *Plasmodium vivax* circumsporozoite protein-based vaccine in *Aotus nancymae* is associated with antibodies to the repeat region. *PLoS Negl Trop Dis*, 8, e3268.
- YEH, E. & DERISI, J. L. 2011. Chemical rescue of malaria parasites lacking an apicoplast defines organelle function in blood-stage *Plasmodium falciparum*. *PLoS Biol*, 9, e1001138.
- YEKUTIEL, P. 1980. III The Global Malaria Eradication Campaign. In: MA, K. (ed.) *Eradication of infectious diseases: a critical study*. . Basel, Switzerland: Karger.

- YEO, T. W., LAMPAH, D. A., TJITRA, E., PIERA, K., GITAWATI, R., KENANGALEM, E., PRICE, R. N. & ANSTEY, N. M. 2010. Greater endothelial activation, Weibel-Palade body release and host inflammatory response to *Plasmodium vivax*, compared with *Plasmodium falciparum*: a prospective study in Papua, Indonesia. *J Infect Dis*, 202, 109-12.
- YILDIZ ZEYREK, F., PALACPAC, N., YUKSEL, F., YAGI, M., HONJO, K., FUJITA, Y., ARISUE, N., TAKEO, S., TANABE, K., HORII, T., TSUBOI, T., ISHII, K. J. & COBAN, C. 2011. Serologic markers in relation to parasite exposure history help to estimate transmission dynamics of *Plasmodium vivax*. *PLoS One*, 6, e28126.
- YOUNG, J. A., FIVELMAN, Q. L., BLAIR, P. L., DE LA VEGA, P., LE ROCH, K. G., ZHOU, Y., CARUCCI, D. J., BAKER, D. A. & WINZELER, E. A. 2005. The *Plasmodium falciparum* sexual development transcriptome: a microarray analysis using ontology-based pattern identification. *Mol Biochem Parasitol*, 143, 67-79.
- YOUNG, M. D. & BURGESS, R. W. 1959. Pyrimethamine resistance in *Plasmodium vivax* malaria. *Bull World Health Organ*, 20, 27-36.
- ZEESHAN, M., BORA, H. & SHARMA, Y. D. 2013. Presence of memory T cells and naturally acquired antibodies in *Plasmodium vivax* malaria-exposed individuals against a group of tryptophan-rich antigens with conserved sequences. *J Infect Dis*, 207, 175-85.
- ZHANG, X., ADDA, C. G., LOW, A., ZHANG, J., ZHANG, W., SUN, H., TU, X., ANDERS, R. F. & NORTON, R. S. 2012. Role of the helical structure of the N-terminal region of *Plasmodium falciparum* merozoite surface protein 2 in fibril formation and membrane interaction. *Biochemistry*, 51, 1380-7.
- ZHU, L., MOK, S., IMWONG, M., JAIDEE, A., RUSSELL, B., NOSTEN, F., DAY, N. P., WHITE, N. J., PREISER, P. R. & BOZDECH, Z. 2016. New insights into the *Plasmodium vivax* transcriptome using RNA-Seq. *Sci Rep*, 6, 20498.
- ZUCCALA, E. S., GOUT, A. M., DEKIWADIA, C., MARAPANA, D. S., ANGRISANO, F., TURNBULL, L., RIGLAR, D. T., ROGERS, K. L., WHITCHURCH, C. B., RALPH, S. A., SPEED, T. P. & BAUM, J. 2012. Subcompartmentalisation of proteins in the rhoptries correlates with ordered events of erythrocyte invasion by the blood stage malaria parasite. *PLoS One*, 7, e46160.

APPENDIX A: LIST OF CONTRIBUTORS

The work in this thesis was guided by mentors Julian Rayner and Rick Fairhurst, and made possible through funding from the Laboratory of Malaria and Vector Research (LMVR), National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland, USA, as well as the Malaria Programme, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK. Special thanks goes to Thomas Wellems for the opportunity to join LMVR. The work described in Chapter 5 was done in collaboration with Camila Franca and Ivo Mueller at the Walter and Eliza Hall Institute. Below is a list of contributors by institution.

Centre for Sustainable Development, University of Cambridge, UK

Friedman, Kayla: produced Microsoft Word template modified to produce this thesis

Morgan, Malcolm: produced Microsoft Word template modified to produce this thesis

Laboratory of Malaria and Vector Research, NIH, Rockville, MD

Amaratunga, Chanaki: Performed *ex vivo* culturing to deliver schizont-enriched *P. vivax* clinical isolates in RNAlater®, prepared and shipped Cambodian *P. vivax* patient plasma, provided Table 2.3 which was slightly modified

Bess, Cameron: Provided training on western blots and orientation to the Fairhurst laboratory

Diouf, Ababacar: Provided training for ELISAs, assistance in developing the ELISA used for Cambodian plasma screening

Eastman, Richard: Provided advice and materials for subcloning plasmids, collected colony plates

Long, Carole: Advised in developing the ELISA used for plasma screening

Miura, Kazutoyo: Advised in developing the ELISA used for plasma screening and analyzing results

Nikolaeva, Daria: Provided prioritized Cambodian *P. vivax* patient sera based on MSP1 reactivity

Pursat Regional Hospital, Cambodia

Sreng, Sokunthea: collected and prepared Cambodian patient plasma

Suon, Seila: collected and prepared Cambodian patient plasma

Walter and Eliza Hall Institute, Melbourne, Australia

Franca, Camila: Provided sera samples used in SI sera study, SI ELISAs (performed with Jessica Hostetler) at the WEHI, performed PNG experiments and analyzed results

Mueller, Ivo: Provided laboratory space and input for designing SI/PNG studies, assisted in interpreting results

Wellcome Trust Sanger Institute, Hinxton, UK

Bartholdson, Josefin (Wright laboratory): Extensive Biacore and Akta training for protein purification and SPR experiments, assistance in interpreting SPR results

Bianchi, Enrica (Wright laboratory): Shared HEK293E cell cultures

Böhme, Ulrike (Berriman laboratory): Used RNA-Seq data to correct *P. vivax* gene model annotation

Campino, Susana (Kwiatkowski laboratory): Provided training on parasite culturing and Percoll® enrichment used for RNA-Seq work

Chappell, Lia (Berriman/Rayner Laboratories): Provided training and protocols for *P. vivax* RNA sequencing, completed final PCR step of RNA-Seq library construction, provided files listing 1:1 orthologs between *P. falciparum* 3D7 and *P. vivax* Sal 1, several custom R scripts, and figure images (as attributed in the text)

Drury, Eleanor (Kwiatkowski laboratory): Assisted in dialysis several times during bait protein production

Dundas, Kirsten (Wright laboratory): Shared HEK293E cell cultures

Galaway, Francis (Wright laboratory): significant training and guidance for protein purification

Gomes, Ana Rita (Billker laboratory): Provided miscellaneous help in the lab during UK visits

Jones, Matthew (Rayner laboratory): Provided extensive initial laboratory orientation and training, training on western blots.

Knoeckel, Julia (Wright laboratory): Shared HEK293E cell cultures

Modrzynska, Katarzyna (Billker laboratory): provided *P. berghei* gametocyte expression data, and assisted with shipping *P. vivax* recombinant proteins

Muller-Sienerth, Nicole (Wright Laboratory): Prepared HEK293E cell cultures for several UK trips

Onley, Catherine (Wright Laboratory): Shared HEK293E cell cultures

Otto, Thomas (Berriman laboratory): advice and bioinformatics assistance with RNA-Seq for *P. vivax* clinical samples

Perrin, Abigail (Wright laboratory): subcloned 3 *P. vivax* MSP7 plasmids, provided CD4d3+d4 used as positive control in some AVEKIS assays, training on HEK293E system

Prestwood, Liam (Rayner laboratory): Provided significant logistical help in travelling back and forth to the UK, including securing lab space and shipping samples

Proto, William (Rayner laboratory): Developed some western blots, prepared HEK293E cell cultures for several UK trips

Sanderson, Theodore: Provided thesis advice

Sharma, Sumana (Wright/Rayner laboratories): Expressed several erythrocyte prey proteins, expressed and purified proteins used in PNG study, frequently aided in protein collection, prepared HEK293E cell cultures for several UK trips

Volz, Jennifer (Rayner laboratory): Provided *P. falciparum* cultures used in testing RNA extraction methods

Wanaguru, Madushi (Wright laboratory): Provided training on HEK293E system, ELISAs, and erythrocyte binding assays

Wright, Gavin: Provided mentoring throughout the project, contributed to *P. vivax* construct discussion and ectodomain trimming, invaluable analysis and advice for SPR results

Wright-Crosnier, Cécile (Wright laboratory): provided mentoring throughout the project, Ox68 antibody, aliquots of plasmids used for subcloning, training on HEK293E system, primers for the plasmid backbone, advice for *P. vivax* construct ordering

Zenonos, Zenon (Wright/Rayner Laboratories): Provided partial list of *P. falciparum* invasion homologs, advice on HEK293E protein expression, and provided training in the Wright Laboratory, Biacore CAPchip usage

APPENDIX B: SUPPLEMENTARY TABLES

Supplementary Table A: *P. vivax* Sal 1 reference annotation changes based on RNA-Seq data

gene_id^a	date	type	details
PVX_000015	01/12/2012	exon	added exon 1
PVX_000530	01/12/2012	exon	deleted intron 9*
PVX_000580	29/11/2012	exon	changed coordinates of exon 2, added exon 10
PVX_000604	29/11/2012	exon	changed coordinates of exon 5
PVX_000650	29/04/2013	exon	deleted intron 4
PVX_000660	29/11/2012	exon	changed coordinates of exon 3
PVX_000825	29/11/2012	exon	changed coordinates of exon 1
PVX_000900	29/11/2012	exon	added exon 5, changed coordinates of exon 6
PVX_001100	06/12/2012	exon	extended exon 2, added exon 1
PVX_001105	06/12/2012	exon	extended exon 2, added exon 1
PVX_001615	29/04/2013	exon	added exon 1
PVX_001630	07/12/2012	exon	added intron 1
PVX_001700	02/12/2012	exon	added exon 1, extended exon 2
PVX_001783	05/12/2012	exon	changed coordinates of exon 1
PVX_001840	30/11/2012	exon	changed coordinates of exon 10
PVX_002485	23/12/2012	exon	added exon 4
PVX_002485	29/11/2012	exon	changed coordinates of exon 2
PVX_002507	01/12/2012	exon	added exon 1, extended exon 2
PVX_002745	23/12/2012	exon	added exon 10
PVX_002855	21/04/2013	exon	changed coordinates of exon 1
PVX_002865	29/11/2012	exon	changed coordinates of exon 2
PVX_002930	29/11/2012	exon	changed coordinates of exon 2
PVX_003487	21/04/2013	exon	changed coordinates of exon 2
PVX_003550	08/10/2013	exon	changed coordinates of exon 1*
PVX_003555	01/12/2012	exon	changed coordinates of exon 1
PVX_003580	29/11/2012	exon	added exon 1*
PVX_003685	06/12/2012	exon	added exon 1, 2*
PVX_003750	21/04/2013	exon	changed coordinates of exon 2
PVX_003795	29/11/2012	exon	changed coordinates of exon 2
PVX_003810	29/11/2012	exon	changed coordinates of exon 2
PVX_003825	29/11/2012	exon	changed coordinates of exon 2
PVX_003840	29/11/2012	exon	changed coordinates of exon 2
PVX_003845	23/12/2012	exon	changed coordinates of exon 2
PVX_003915	29/11/2012	exon	changed coordinates of exon 6
PVX_004530	21/04/2013	exon	changed coordinates of exon 1
PVX_004530	30/11/2012	exon	changed coordinates of exon 2

PVX_004537	30/11/2012	exon	added exon 3
PVX_005040	30/11/2012	exon	added exon 2*
PVX_005058	05/12/2012	exon	added intron 2
PVX_079740	01/12/2012	exon	changed coordinates of exon 3
PVX_079785	01/12/2012	exon	changed coordinates of exon 4
PVX_079825	28/12/2012	exon	added exon 10
PVX_079860	28/12/2012	exon	added exon 11
PVX_079860	01/12/2012	exon	changed coordinates of exon 6
PVX_079895	29/12/2012	exon	changed coordinates of exon 1*
PVX_080055	01/12/2012	exon	changed coordinates of exon 2
PVX_080115	29/12/2012	exon	changed coordinates of exon 4
PVX_080125	29/12/2012	exon	changed coordinates of exon 2
PVX_080130	01/12/2012	exon	changed coordinates of exon 2
PVX_080147	08/12/2012	exon	changed coordinates of exon 4*
PVX_080150	29/12/2012	exon	added exon 6
PVX_080150	05/12/2012	exon	changed coordinates of exon 2*
PVX_080255	01/12/2012	exon	changed coordinates of exon 2
PVX_080305	05/12/2012	exon	added exon 7, changed coordinates of exon 6
PVX_080345	01/12/2012	exon	changed coordinates of exon 2
PVX_080475	01/12/2012	exon	changed coordinates of exon 5
PVX_080617	01/12/2012	exon	changed coordinates of exon 3, 8*
PVX_080670	01/12/2012	exon	changed coordinates of exon 2
PVX_080690	01/12/2012	exon	added exon 8
PVX_081250	29/11/2012	exon	changed coordinates of exon 2
PVX_081275	29/11/2012	exon	changed coordinates of exon 1
PVX_081360	29/11/2012	exon	extended exon 1
PVX_081410	29/11/2012	exon	changed coordinates of exon 4, 5*
PVX_081490	29/11/2012	exon	changed coordinates of exon 2
PVX_081605	29/11/2012	exon	changed coordinates of exon 5
PVX_081835	01/12/2012	exon	added exon 1
PVX_081840	23/12/2012	exon	changed coordinates of exon 3
PVX_081847	29/11/2012	exon	changed coordinates of exon 2
PVX_082360	01/12/2012	exon	changed coordinates of exon 2
PVX_082380	01/12/2012	exon	added exon 8
PVX_082830	01/12/2012	exon	changed coordinates of exon 17
PVX_082835	01/12/2012	exon	changed coordinates of exon 1
PVX_083110	22/04/2013	exon	changed coordinates of exon 1
PVX_083335	19/12/2012	exon	added exon 5
PVX_083345	01/12/2012	exon	added exon 9
PVX_083350	22/04/2013	exon	changed coordinates of exon 1
PVX_083370	12/12/2012	exon	added exon 2

PVX_083545	12/12/2012	exon	added exon 1
PVX_083575	22/04/2013	exon	changed coordinates of exon 2*
PVX_084120	02/01/2013	exon	added exon 1*
PVX_084130	03/12/2012	exon	changed coordinates of exon 3
PVX_084440	03/12/2012	exon	changed coordinates of exon 1*
PVX_084515	02/01/2013	exon	changed coordinates of exon 6, added exon 7*
PVX_084525	03/12/2012	exon	changed coordinates of exon 1
PVX_084925	03/12/2012	exon	changed coordinates of exon 10
PVX_084935	03/12/2012	exon	changed coordinates of exon 1
PVX_085005	03/12/2012	exon	changed coordinates of exon 1
PVX_085115	03/12/2012	exon	changed coordinates of exon 1
PVX_085375	05/12/2012	exon	changed coordinates of exon 1*
PVX_085410	02/01/2013	exon	changed coordinates of exon 1
PVX_085415	09/12/2012	exon	added exon 2
PVX_085500	02/01/2013	exon	changed coordinates of exon 9
PVX_085585	03/12/2012	exon	deleted exon 2
PVX_085905	03/12/2012	exon	changed coordinates of exon 1
PVX_085910	02/01/2013	exon	added exon 1
PVX_085977	14/12/2012	exon	added exon 1*
PVX_086030	03/12/2012	exon	changed coordinates of exon 4
PVX_086135	02/01/2013	exon	changed coordinates of exon 7*
PVX_086205	02/01/2013	exon	changed coordinates of exon 18
PVX_086205	03/12/2012	exon	changed coordinates of exon 7, 8*
PVX_086280	09/12/2012	exon	added exon 1, 2, 3*
PVX_086330	03/12/2012	exon	changed coordinates of exon 7
PVX_086860	30/11/2012	exon	changed coordinates of exon 1
PVX_086870	30/11/2012	exon	changed coordinates of exon 1
PVX_086930	30/11/2012	exon	changed coordinates of exon 1
PVX_086940	27/01/2013	exon	added exon 1*
PVX_086945	30/11/2012	exon	changed coordinates of exon 1
PVX_086985	24/12/2012	exon	changed coordinates of exon 2
PVX_087080	07/12/2012	exon	added exon 5, 6, 7*
PVX_087105	05/12/2012	exon	changed coordinates of exon 2*
PVX_087855	06/12/2012	exon	added exon 2, shortened exon 1
PVX_088035	29/11/2012	exon	changed coordinates of exon 2
PVX_088165	22/12/2012	exon	added exon 5
PVX_088205	06/12/2012	exon	added exon 1
PVX_088240	29/11/2012	exon	changed coordinates of exon 9
PVX_088250	06/12/2012	exon	added exon 2
PVX_088805	30/11/2012	exon	changed coordinates of exon 1
PVX_088805	21/04/2013	exon	changed coordinates of exon 3*

PVX_088810	01/12/2012	exon	added exon 1
PVX_088845	03/12/2012	exon	added exon 1, extended exon 2
PVX_088975	05/12/2012	exon	changed coordinates of exon 1*
PVX_089020	10/12/2012	exon	added exon 1
PVX_089360	30/11/2012	exon	extended exon 1
PVX_089410	30/11/2012	exon	changed coordinates of exon 1
PVX_089460	21/04/2013	exon	changed coordinantes of exon 1
PVX_089470	30/11/2012	exon	changed coordinates of exon 1
PVX_089473	23/12/2012	exon	changed coordinates of exon 1*
PVX_089475	23/12/2012	exon	changed coordinates of exon 1*
PVX_089500	30/11/2012	exon	changed coordinates of exon 3, added exon 5
PVX_089592	23/12/2012	exon	changed coordinates of exon 1
PVX_089700	30/11/2012	exon	changed coordinates of exon 3
PVX_089720	05/12/2012	exon	changed coordinates of exon 1*
PVX_089852	23/12/2012	exon	changed coordinates of exon 1*
PVX_089910	30/11/2012	exon	changed coordinates of exon 1
PVX_090270	02/12/2012	exon	extended exon 2, added exon 1
PVX_090330	03/12/2012	exon	deleted exon 3, added exon 1
PVX_090335	30/11/2012	exon	changed coordinates of exon 2
PVX_090940	30/11/2012	exon	changed coordinates of exon 8, added exon 12
PVX_091000	27/12/2012	exon	added exon 3
PVX_091136	30/11/2012	exon	changed coordinates of exon 2
PVX_091140	30/11/2012	exon	changed coordinates of exon 5
PVX_091430	30/11/2012	exon	changed coordinates of exon 10
PVX_091570	30/11/2012	exon	added exon 2, 3, 17*
PVX_091580	30/11/2012	exon	changed coordinates of exon 4
PVX_091610	30/11/2012	exon	changed coordinates of exon 1*
PVX_091660	10/01/2013	exon	added exon 1*
PVX_091750	30/11/2012	exon	changed coordinates of exon 10
PVX_091905	03/01/2013	exon	changed coordinates of exon 6*
PVX_092005	30/11/2012	exon	added exon 2
PVX_092035	07/12/2012	exon	deleted intron 2*
PVX_092135	30/11/2012	exon	changed coordinates of exon 3
PVX_092190	03/01/2013	exon	changed coordinates of exon 4
PVX_092330	07/12/2012	exon	added exon 1
PVX_092360	07/12/2012	exon	changed coordinates of exon 4, added exon 6
PVX_092370	30/11/2012	exon	changed coordinates of exon 1
PVX_092415	10/01/2013	exon	added exon 1*
PVX_092475	22/04/2013	exon	changed coordinates of exon 1
PVX_092490	05/12/2012	exon	changed coordinates of exon 1*
PVX_092670	30/11/2012	exon	changed coordinates of exon 1

PVX_092680	07/12/2012	exon	added exon 2
PVX_092680	30/11/2012	exon	added exon 2*
PVX_092790	30/11/2012	exon	changed coordinates of exon 2*
PVX_092855	05/12/2012	exon	deleted exon 2, changed coordinates of exon 1, 2*
PVX_092925	01/12/2012	exon	changed coordinates of exon 1
PVX_092930	22/04/2013	exon	changed coordinates of exon 2
PVX_093710	29/11/2012	exon	changed coordinates of exon 2*
PVX_093715	29/11/2012	exon	added exon 4, changed coordinates of exon 1, 2*
PVX_094240	30/11/2012	exon	changed coordinates of exon 1
PVX_094240	30/11/2012	exon	changed coordinates of exon 3
PVX_094243	30/11/2012	exon	added exon 3*
PVX_094245	30/11/2012	exon	deleted intron 2
PVX_094250	30/11/2012	exon	changed coordinates of exon 1
PVX_094255	24/12/2012	exon	changed coordinates of exon 1*
PVX_094260	16/12/2012	exon	added exon 2, extended exon 1*
PVX_094425	30/11/2012	exon	changed coordinates of exon 5
PVX_094540	11/12/2012	exon	added exon 8
PVX_094820	30/11/2012	exon	changed coordinates of exon 2
PVX_094890	30/11/2012	exon	changed coordinates of exon 6
PVX_094905	07/01/2013	exon	added exon 5
PVX_095035	09/04/2014	exon	added exon 2
PVX_095095	30/11/2012	exon	changed coordinates of exon 1
PVX_095125	30/11/2012	exon	added exon 3, 10*
PVX_095125	22/04/2013	exon	changed coordinates of exon 6
PVX_095140	30/11/2012	exon	changed coordinates of exon 3
PVX_095165	07/12/2012	exon	added exon 3*
PVX_095315	05/12/2012	exon	changed coordinates of exon 1*
PVX_095405	30/11/2012	exon	changed coordinates of exon 4
PVX_095475	07/12/2012	exon	deleted intron*
PVX_095495	30/11/2012	exon	changed coordinates of exon 1
PVX_095990	29/11/2012	exon	changed coordinates of exon 1
PVX_096010	04/12/2012	exon	changed coordinates of exon 1
PVX_096045	10/12/2012	exon	added exon 1, changed coordinates of exon 2
PVX_096060	06/12/2012	exon	added exon 2
PVX_096065	04/12/2012	exon	changed coordinates of exon 1*
PVX_096195	29/11/2012	exon	changed coordinates of exon 4
PVX_096395	23/12/2012	exon	added exon 1*
PVX_096920	26/04/2013	exon	added exon 1
PVX_096920	29/11/2012	exon	changed coordinates of exon 1*
PVX_096930	10/12/2012	exon	added exon 1, changed coordinates of exon 3
PVX_097530	01/12/2012	exon	added exon 1

PVX_097540	01/12/2012	exon	changed coordinates of exon 2*
PVX_097557	01/12/2012	exon	changed coordinates of exon 1
PVX_097580	01/12/2012	exon	added exon 1
PVX_097585	01/12/2012	exon	added exon 1, extended exon 2
PVX_097787	01/12/2012	exon	changed coordinates of exon 2*
PVX_097890	29/12/2012	exon	added exon 2
PVX_098045	01/12/2012	exon	changed coordinates of exon 5*
PVX_098582	16/12/2012	exon	added exon 1
PVX_098625	07/12/2012	exon	added exon 4*
PVX_098660	21/04/2013	exon	changed coordinates of exon 9
PVX_098690	11/12/2012	exon	added exon 3
PVX_098725	30/11/2012	exon	changed coordinates of exon 9
PVX_098820	30/11/2012	exon	changed coordinates of exon 3
PVX_099015	30/11/2012	exon	changed coordinates of exon 4
PVX_099065	24/12/2012	exon	changed coordinates of exon 1
PVX_099385	30/11/2012	exon	changed coordinates of exon 2
PVX_099470	24/12/2012	exon	added exon 3*
PVX_099475	07/12/2012	exon	added exon 3, 4*
PVX_099530	11/12/2012	exon	added exon 4*
PVX_099635	30/11/2012	exon	changed coordinates of exon 1
PVX_099665	30/11/2012	exon	changed coordinates of exon 1, 2*
PVX_099740	30/11/2012	exon	changed coordinates of exon 7
PVX_099800	24/12/2012	exon	changed coordinates of exon 4
PVX_099915	07/12/2012	exon	added exon 1*
PVX_099970	24/12/2012	exon	changed coordinates of exon 3
PVX_100005	24/12/2012	exon	added exon 3
PVX_100515	03/12/2012	exon	changed coordinates of exon 2
PVX_100595	22/04/2013	exon	changed coordinates of exon 6*
PVX_100630	03/12/2012	exon	changed coordinates of exon 2
PVX_100670	22/04/2013	exon	changed coordinates of exon 1
PVX_101350	14/12/2012	exon	added exon 2
PVX_101442	22/04/2013	exon	changed coordinates of exon 7
PVX_101505	01/12/2012	exon	added exon 1, extended exon 2
PVX_101510	01/12/2012	exon	added exon 1
PVX_101550	22/04/2013	exon	added exon 1
PVX_101560	22/04/2013	exon	added exon 1
PVX_101565	01/12/2012	exon	changed coordinates of exon 4
PVX_101570	01/12/2012	exon	added exon 4
PVX_101575	01/12/2012	exon	added exon 1
PVX_101575	22/04/2013	exon	added exon 1, extended exon 2
PVX_101585	01/12/2012	exon	added exon 1

PVX_101590	22/04/2013	exon	added exon 1
PVX_101620	01/12/2012	exon	changed coordinates of exon 1
PVX_110830	30/11/2012	exon	added exon 1
PVX_110835	30/11/2012	exon	deleted intron 2
PVX_111150	30/11/2012	exon	changed coordinates of exon 2
PVX_111150	24/12/2012	exon	changed coordinates of exon 5
PVX_111270	30/11/2012	exon	added exon 3
PVX_111292	11/12/2012	exon	changed coordinates of exon 1
PVX_111415	30/11/2012	exon	added exon 7, changed coordinates of exon 8*
PVX_111435	24/12/2012	exon	changed coordinates of exon 7
PVX_111440	30/11/2012	exon	changed coordinates of exon 2*
PVX_111535	11/12/2012	exon	added exon 1, 2, 3*
PVX_113220	16/12/2012	exon	added exon 3
PVX_113230	01/12/2012	exon	deleted exon 4
PVX_113260	01/12/2012	exon	added exon 1, changed coordinates of exon 2
PVX_113500	12/12/2012	exon	added exon 2
PVX_113525	01/12/2012	exon	changed coordinates of exon 1
PVX_113731	01/12/2012	exon	added exon 4*
PVX_113945	05/12/2012	exon	changed coordinates of exon 3
PVX_113980	08/12/2012	exon	deleted intron*
PVX_114070	05/12/2012	exon	changed coordinates of exon 8*
PVX_114250	30/12/2012	exon	changed coordinates of exon 1*
PVX_114270	30/12/2012	exon	merged exon 8*
PVX_114346	01/12/2012	exon	changed coordinates of exon 17, added exon 14
PVX_114350	01/12/2012	exon	changed coordinates of exon 2
PVX_114350	05/12/2012	exon	changed coordinates of exon 5*
PVX_114920	05/12/2012	exon	changed coordinates of exon 3*
PVX_114990	01/12/2012	exon	changed coordinates of exon 7
PVX_115125	12/12/2012	exon	added exon 1
PVX_115425	01/12/2012	exon	changed coordinates of exon 1
PVX_115445	12/12/2012	exon	added exon 5
PVX_115470	01/12/2012	exon	added exon 1, extended exon 2
PVX_116525	01/12/2012	exon	changed coordinates of exon 15, added exon 16
PVX_116582	24/01/2013	exon	added exon 1*
PVX_116585	24/01/2013	exon	added exon 2*
PVX_116635	01/12/2012	exon	added intron 1, changed coordinates of exon 2
PVX_116700	24/01/2013	exon	added exon 1*
PVX_116750	12/12/2012	exon	changed coordinates of exon 1*
PVX_116815	01/12/2012	exon	changed coordinates of exon 3*
PVX_117120	01/12/2012	exon	changed coordinates of exon 2
PVX_117335	22/04/2013	exon	changed coordinates of exon 10, 11, 20, 23, added

			exon 19
PVX_117725	13/12/2012	exon	added exon 1
PVX_117915	22/04/2013	exon	added exon 4
PVX_118445	22/04/2013	exon	changed coordinates of exon 4*
PVX_118560	03/12/2012	exon	changed coordinates of exon 14
PVX_118600	03/12/2012	exon	changed coordinates of exon 11
PVX_118690	13/12/2012	exon	deleted exon 1*
PVX_119265	11/12/2012	exon	added exon 2
PVX_119325	05/12/2012	exon	changed coordinates of exon 4*
PVX_119440	30/11/2012	exon	added exon 4
PVX_119500	30/11/2012	exon	changed coordinates of exon 2
PVX_119550	30/11/2012	exon	added exon 2
PVX_119550	19/02/2013	exon	changed coordinates of exon 1
PVX_119595	30/11/2012	exon	changed coordinates of exon 1
PVX_119660	07/12/2012	exon	deleted intron 4
PVX_119725	30/11/2012	exon	changed coordinates of exon 5
PVX_121870	22/04/2013	exon	added exon 1
PVX_121876	01/12/2012	exon	changed coordinates of exon 2
PVX_121895	02/01/2013	exon	added exon 1
PVX_121965	03/12/2012	exon	changed coordinates of exon 2
PVX_122077	14/12/2012	exon	added exon 9
PVX_122105	03/12/2012	exon	changed coordinates of exon 1
PVX_122130	03/12/2012	exon	changed coordinates of exon 1
PVX_122270	03/12/2012	exon	changed coordinates of exon 4
PVX_122635	09/12/2012	exon	added exon 1
PVX_122655	03/12/2012	exon	changed coordinates of exon 2
PVX_122865	03/12/2012	exon	changed coordinates of exon 1
PVX_123283	03/12/2012	exon	changed coordinates of exon 3
PVX_123290	03/12/2012	exon	changed coordinates of exon 2*
PVX_123415	22/04/2013	exon	changed coordinates of exon 1*
PVX_123530	22/04/2013	exon	added exon 5
PVX_123675	09/12/2012	exon	deleted exon 1*
PVX_123780	03/12/2012	exon	changed coordinates of exon 2
PVX_124020	09/12/2012	exon	deleted exon 2*
PVX_124140	05/12/2012	exon	changed coordinates of exon 1, 2, 5*
PVX_124190	03/12/2012	exon	changed coordinates of exon 7
PVX_124700	01/12/2012	exon	changed coordinates of exon 1*
PVX_124705	01/12/2012	exon	changed coordinates of exon 2
PVX_124708	01/12/2012	exon	added intron 1
PVX_001660	07/12/2012	merge	merged PVX_001660 with PVX_001665
PVX_088790	21/04/2013	merge	merged PVX_088790 with PVX_088785

PVX_095470	30/11/2012	merge	merged PVX_095465 with PVX_095470
PVX_097885	29/12/2012	merge	merged PVX_097880 with PVX_097885
PVX_101385	22/07/2014	merge	merged PVX_101380 with PVX_101385*
PVX_123675	14/12/2013	merge	merged PVX_123675 with PVX_123680*
PVX_124130	14/12/2012	merge	PVX_124125 with PVX_124130
PVX_081294	26/11/2014	split	PVX_081295 has been split into PVX_081294*
PVX_081296	26/11/2014	split	PVX_081295 has been split into PVX_081294*
PVX_098784	28/01/2013	split	split PVX_098785 into PVX_098784*
PVX_098786	28/01/2013	split	split PVX_098785 into PVX_098784*
PVX_122218	09/12/2012	split	split PVX_122220 into PVX_122218*
PVX_122222	09/12/2012	split	split PVX_122220 into PVX_122218*
PVX_002952	20/01/2014	new	gene model added*
PVX_079727	22/11/2014	new	new gene model added
PVX_084277	09/12/2012	new	gene model added*
PVX_088775	03/12/2012	new	gene model added
PVX_088797	03/12/2012	new	gene model added
PVX_089863	24/12/2012	new	new gene model added*
PVX_090830	28/04/2013	new	gene model added
PVX_090835	28/04/2013	new	gene model added
PVX_092897	07/12/2012	new	gene model added
PVX_094270	16/12/2012	new	new gene model added
PVX_094303	07/12/2012	new	gene model added
PVX_095452	07/12/2012	new	gene model added
PVX_095997	06/12/2012	new	gene model added
PVX_096007	04/12/2012	new	gene model added
PVX_097583	08/12/2012	new	gene model added*
PVX_099117	24/12/2012	new	gene model added*
PVX_101503	22/04/2013	new	new gene model added
PVX_101592	01/12/2012	new	gene model added
PVX_110832	30/11/2012	new	new gene model added
PVX_110834	30/11/2012	new	gene model added

*Based on homology and RNA-Seq

^aAll annotation changes made by Ulrike Böhme in the Parasite Genomics group at the WTSI. Gene ids listed may only be present in GeneDB.org which has more up-to-date annotation than Plasmodb.org.

Supplementary Table B: Variably-expressed genes in *P. vivax* clinical isolates

<i>P. vivax</i> P01 id	Product	<i>P. falciparum</i> 3D7 (1:1 hom)	Mean expr.	FC	CV	VMR
PVP01_0936800	3-oxo-5-alpha-steroid 4-dehydrogenase, putative	PF3D7_1135900	132	1.6	0.41	22.71
PVP01_1271300	40 kDa heat shock protein, putative	NA	30	2.17	0.81	19.94
PVP01_1318400	50S ribosomal protein L17, putative	PF3D7_1431000	90	1.95	0.72	47.1
PVP01_0111000	60S ribosomal protein L34, putative	PF3D7_0710600	283	1.57	0.44	55.57
PVP01_1430300	acyl-CoA binding protein, putative	NA	1050	2.01	0.7	511.13
PVP01_0409900	acyl-CoA synthetase, putative	NA	56	1.71	0.58	19.12
PVP01_0920400	alternative splicing factor ASF-1, putative	PF3D7_1119800	102	1.73	0.49	24.2
PVP01_0934200	apical membrane antigen 1	PF3D7_1133400	84	2.1	0.79	52.84
PVP01_0521400	armadillo-domain containing rhoptry protein,	PF3D7_0414900	120	1.96	0.69	56.65
PVP01_0703100	bacterial histone-like protein, putative	PF3D7_0904700	63	1.79	0.55	18.93
PVP01_0407500	calcium-dependent protein kinase 1, putative	PF3D7_0217500	87	1.63	0.45	17.99
PVP01_1121200	choline/ethanolaminephosphotransferase,	PF3D7_0628300	170	1.61	0.45	33.71
PVP01_0204400	chromatin assembly factor 1 protein WD40 domain,	PF3D7_0110700	98	1.57	0.41	16.19
PVP01_0106500	conserved Plasmodium protein, unknown function	NA	59	1.81	0.57	18.75
PVP01_0110200	conserved Plasmodium protein, unknown function	PF3D7_0709800	200	1.78	0.52	55.04
PVP01_0117300	conserved Plasmodium protein, unknown function	PF3D7_0802900	69	1.66	0.53	19.24
PVP01_0321100	conserved Plasmodium protein, unknown function	PF3D7_0725400	432	1.68	0.45	87.73
PVP01_0414300	conserved Plasmodium protein, unknown function	PF3D7_0210600	139	1.79	0.56	43.16
PVP01_0421700	conserved Plasmodium protein, unknown function	PF3D7_0203400	128	1.69	0.49	30.12
PVP01_0510600	conserved Plasmodium protein, unknown function	PF3D7_0823700	94	1.75	0.51	24.83
PVP01_0511400	conserved Plasmodium protein, unknown function	PF3D7_0822900	122	1.81	0.59	42.51
PVP01_0526400	conserved Plasmodium protein, unknown function	PF3D7_0417000	163	1.59	0.5	41.32
PVP01_0530400	conserved Plasmodium	NA	284	1.57	0.42	50.07

	protein, unknown function					
PVP01_0602700	conserved Plasmodium protein, unknown function	PF3D7_1017500	188	1.53	0.54	54.33
PVP01_0611800	conserved Plasmodium protein, unknown function	NA	38	1.96	0.68	17.36
PVP01_0617300	conserved Plasmodium protein, unknown function	PF3D7_1032200	37	3.26	1.51	84.24
PVP01_0620300	conserved Plasmodium protein, unknown function	NA	347	1.88	0.61	128.85
PVP01_0709600	conserved Plasmodium protein, unknown function	PF3D7_0911100	37	2.2	0.85	27.01
PVP01_0712200	conserved Plasmodium protein, unknown function	NA	61	2.14	0.83	41.75
PVP01_0724700	conserved Plasmodium protein, unknown function	NA	368	1.61	0.44	70.55
PVP01_0730700	conserved Plasmodium protein, unknown function	PF3D7_0932000	321	1.57	0.44	63.43
PVP01_0733300	conserved Plasmodium protein, unknown function	PF3D7_0934600	53	1.81	0.56	16.72
PVP01_0734300	conserved Plasmodium protein, unknown function	NA	36	1.93	0.68	16.66
PVP01_0811300	conserved Plasmodium protein, unknown function	PF3D7_1011200	63	1.7	0.5	15.57
PVP01_0907700	conserved Plasmodium protein, unknown function	PF3D7_1106900	66	1.66	0.52	17.89
PVP01_1207500	conserved Plasmodium protein, unknown function	PF3D7_1347300	64	1.91	0.63	25.19
PVP01_1227100	conserved Plasmodium protein, unknown function	PF3D7_1327300	55	1.77	0.54	16.34
PVP01_1229300	conserved Plasmodium protein, unknown function	PF3D7_1325300	64	1.67	0.49	15.5
PVP01_1247000	conserved Plasmodium protein, unknown function	PF3D7_1459900	139	1.63	0.44	27.37
PVP01_1304000	conserved Plasmodium protein, unknown function	NA	35	2.19	0.85	25.08
PVP01_1335700	conserved Plasmodium protein, unknown function	PF3D7_1413200	40	1.82	0.62	15.08
PVP01_1344100	conserved Plasmodium protein, unknown function	PF3D7_1404700	33	2.35	0.94	29.22
PVP01_1345600	conserved Plasmodium protein, unknown function	PF3D7_1403200	77	1.64	0.54	22.82
PVP01_1402500	conserved Plasmodium protein, unknown function	NA	30	2.91	1.28	49.9
PVP01_1412100	conserved Plasmodium protein, unknown function	PF3D7_1311100	85	1.71	0.51	21.88
PVP01_1419600	conserved Plasmodium protein, unknown function	PF3D7_1318700	109	1.83	0.59	37.52
PVP01_1425800	conserved Plasmodium protein, unknown function	PF3D7_0814500	23	2.14	0.83	15.67
PVP01_1427200	conserved Plasmodium	PF3D7_0813100	66	1.72	0.48	15.36

Identification of novel *Plasmodium vivax* blood-stage vaccine targets

	protein, unknown function						
PVP01_1433600	conserved Plasmodium protein, unknown function	PF3D7_1214800	277	1.66	0.52	76.36	
PVP01_1436200	conserved Plasmodium protein, unknown function	PF3D7_1217400	50	1.9	0.62	19.18	
PVP01_1460500	conserved Plasmodium protein, unknown function	NA	25	2.37	0.93	21.51	
PVP01_MIT03400	cytochrome b		54	1.71	0.53	14.94	
PVP01_0623800	duffy receptor precursor	PF3D7_1301600	77	2.68	1.16	102.89	
PVP01_1208800	dynactin subunit 2, putative	PF3D7_1346000	23	2.2	0.8	14.72	
PVP01_1258000	gamete egress and sporozoite traversal protein,	PF3D7_1449000	111	1.63	0.54	32.72	
PVP01_0115300	gamete release protein, putative	PF3D7_0805200	41	2.44	0.98	38.68	
PVP01_1244000	glyceraldehyde-3-phosphate dehydrogenase,	PF3D7_1462800	361	1.73	0.55	109.14	
PVP01_0505600	GPI-anchored micronemal antigen, putative	PF3D7_0828800	53	2.17	0.83	36.87	
PVP01_1316500	H/ACA ribonucleoprotein complex subunit 3,	PF3D7_1433000	132	1.6	0.41	22.24	
PVP01_0917400	heat shock protein 101, putative	PF3D7_1116800	80	1.92	0.63	31.88	
PVP01_0423300	hypothetical protein	NA	95	2.49	1.03	99.68	
PVP01_0620700	hypothetical protein	NA	93	2.22	0.85	66.72	
PVP01_0620900	hypothetical protein	NA	123	1.76	0.63	49.58	
PVP01_1031100	hypothetical protein	NA	48	2.22	0.82	32.18	
PVP01_0812500	hypoxanthine-guanine phosphoribosyltransferase,	PF3D7_1012400	123	2.08	0.75	69.78	
PVP01_0803400	inner membrane complex protein 1c, putative	PF3D7_1003600	383	1.95	0.69	182.15	
PVP01_1203100	inner membrane complex protein 1f, putative	PF3D7_1351700	51	1.76	0.56	15.92	
PVP01_1008000	inner membrane complex protein 1g, putative	PF3D7_0525800	181	1.93	0.64	74.97	
PVP01_1221800	isoleucine--tRNA ligase, putative	PF3D7_1332900	72	1.66	0.45	14.93	
PVP01_0926500	kelch protein, putative	PF3D7_1125800	169	1.81	0.55	50.38	
PVP01_1031200*	merozoite surface protein 3	NA	60	2.77	1.18	83.57	
PVP01_1031400*	merozoite surface protein 3	NA	46	2.29	0.87	34.13	
PVP01_1031500*	merozoite surface protein 3	NA	70	3.31	1.54	166.66	
PVP01_1031600*	merozoite surface protein 3	NA	155	2.36	0.93	133.65	
PVP01_1031700*	merozoite surface protein 3	NA	114	2.82	1.25	176.54	
PVP01_0010200*	merozoite surface protein 3, putative	NA	21	1.76	0.84	14.76	
PVP01_0010220*	merozoite surface protein 3, putative	NA	149	2.03	0.73	78.39	
PVP01_0418400	merozoite surface protein 5	NA	74	1.66	0.54	21.68	

PVP01_1219800*	merozoite surface protein 7 (MSP7)	NA	94	2.29	0.9	75.58
PVP01_1220200*	merozoite surface protein 7 (MSP7), putative	NA	287	2.19	0.9	231.74
PVP01_1220300*	merozoite surface protein 7 (MSP7), putative	NA	63	1.76	0.63	24.71
PVP01_0613800	merozoite TRAP-like protein, putative	PF3D7_1028700	39	1.93	0.63	15.5
PVP01_0824100	microneme associated antigen, putative	PF3D7_0316000	45	1.87	0.6	16.37
PVP01_1315900	microsomal signal peptidase 12 kDa subunit,	PF3D7_1433600	98	1.74	0.54	28.44
PVP01_0818500	microtubule and actin binding protein, putative	PF3D7_0321700	88	2.02	0.69	41.97
PVP01_1463500	myosin A tail domain interacting protein,	PF3D7_1246400	154	1.77	0.62	58.63
PVP01_1212200	myosin A, putative	PF3D7_1342600	135	1.89	0.67	60.67
PVP01_1300900	NIMA related kinase 3, putative	PF3D7_1201600	68	1.67	0.48	15.8
PVP01_0925400	nuclear preribosomal assembly protein, putative	PF3D7_1124800	51	1.79	0.56	15.96
PVP01_1105500	nucleoside diphosphate kinase, putative	PF3D7_1366500	42	2	0.7	20.79
PVP01_1229400	oxidoreductase, putative	PF3D7_1325200	197	2.07	0.74	107.47
PVP01_0206100	photosensitized INA-labeled protein 1, PhIL1,	PF3D7_0109000	136	1.77	0.52	36.76
PVP01_0011000	PIR protein	NA	52	3.11	1.41	101.98
PVP01_0102200	PIR protein, pseudogene	NA	305	1.71	0.48	71.45
PVP01_0001440	Plasmodium exported protein (PHIST), unknown	NA	64	1.75	0.57	20.52
PVP01_0119200	Plasmodium exported protein (PHIST), unknown	PF3D7_0801000	189	1.66	0.45	38.01
PVP01_0504000	Plasmodium exported protein (PHIST), unknown	PF3D7_0830600	50	1.81	0.6	17.86
PVP01_0601700	Plasmodium exported protein (PHIST), unknown	PF3D7_1016600	38	2.5	1.03	40.2
PVP01_0623700	Plasmodium exported protein (PHIST), unknown	NA	302	1.6	0.4	48.71
PVP01_0623100	Plasmodium exported protein, unknown function	NA	412	1.94	0.65	172.34
PVP01_0948800	Plasmodium exported protein, unknown function	NA	53	1.57	0.56	16.57
PVP01_0949000	Plasmodium exported protein, unknown function	NA	27	2.22	0.83	19.02
PVP01_1033800	Plasmodium exported protein, unknown function	NA	270	2.13	0.8	170.97
PVP01_1147000	Plasmodium exported protein, unknown function	NA	35	2.24	0.88	27.37
PVP01_1147500	Plasmodium exported protein, unknown function	NA	40	1.93	0.63	15.79

Identification of novel *Plasmodium vivax* blood-stage vaccine targets

PVP01_1201400	Plasmodium exported protein, unknown function	NA	49	2.77	1.2	70.01
PVP01_1271400	Plasmodium exported protein, unknown function	NA	74	1.73	0.5	18.19
PVP01_1402200	Plasmodium exported protein, unknown function	NA	241	2.8	1.22	360.43
PVP01_1234800	pre-mRNA-splicing factor ISY1, putative	PF3D7_1472000	83	2.05	0.73	43.84
PVP01_1457000	protein kinase 2, putative	PF3D7_1238900	46	2.05	0.75	26.16
PVP01_1421400	ras-related protein Rab-11A, putative	PF3D7_1320600	245	1.62	0.41	41.82
PVP01_0701200	reticulocyte binding protein 1a	NA	48	2.27	0.89	37.81
PVP01_1471400	reticulocyte binding protein 2, pseudogene	NA	36	2.17	0.86	27
PVP01_1402400	reticulocyte binding protein 2a	NA	27	2.33	0.94	24.39
PVP01_0800700	reticulocyte binding protein 2b	NA	29	2.29	0.91	24.08
PVP01_0534300*	reticulocyte binding protein 2c	NA	22	2.26	0.89	17.81
PVP01_1469400	reticulocyte binding protein 3, pseudogene	NA	32	2.16	0.81	20.94
PVP01_0700500	reticulocyte binding protein, pseudogene	NA	54	2.38	0.96	50.18
PVP01_0802200	RNA-binding protein, putative	PF3D7_1002400	98	1.61	0.43	17.85
PVP01_1411700	RNA-binding protein, putative	PF3D7_1310700	474	2.01	0.7	234.05
PVP01_1322100	selenoprotein, putative	PF3D7_1427200	102	1.6	0.42	18.01
PVP01_0417400*	serine-repeat antigen 4 (SERA)	NA	52	1.94	0.66	22.65
PVP01_0504400	sporozoite invasion-associated protein 2,	PF3D7_0830300	32	2.42	1.03	33.53
PVP01_1026500	subtilisin-like protease 1	PF3D7_0507500	168	1.69	0.47	36.92
PVP01_1313200	translocon component PTEX150, putative	PF3D7_1436300	47	1.77	0.57	15.13
PVP01_1267100	triosephosphate isomerase, putative	PF3D7_1439900	69	1.65	0.54	19.87
PVP01_0000170	tryptophan-rich antigen	NA	74	2.17	0.87	56.53
PVP01_1401800	tryptophan-rich antigen (Pv-fam-a)	PF3D7_0102700	40	2.61	1.1	47.66
PVP01_0834500	ubiquitin-conjugating enzyme E2, putative	PF3D7_0305700	210	1.58	0.44	40.57

*Potential artefact due to at least 1 isolate contained large gaps in mapping alignment indicating divergence from the reference.

Top 130 genes variably expressed between clinical isolates (PV0563, PV0565, PV0568, PV0417-3) as ranked by three methods for detecting spread in data. FC is the 'maximum fold-change' from the mean FPKM of at least one of the isolates. CV is the coefficient of variation and reflects the ratio of the standard deviation to the mean. VMR is the variance-to-mean ratio and provides a normalized view of the dispersion of the expression of the isolates. Mean expr. = Mean FPKMs of the 4 isolates. FPKMs computed by Cufflinks from assemblies mapped with TopHat to *P. vivax*.

Supplementary Table C: Top-expressed genes unique to RNA-Seq data

<i>P. vivax</i> P01 id	FPKM, <i>P. vivax</i> isolate					Product
	563	565	068	417-3	Mean	
PVP01_1311000	660	900	1277	981	954	polyubiquitin 5, putative
PVP01_0703800	673	1010	1188	833	926	high molecular weight rhoptry protein 3,
PVP01_1265900	909	778	961	760	852	KS1 protein precursor, putative
PVP01_0830600	565	627	338	497	507	60S acidic ribosomal protein P2, putative
PVP01_1260400	450	538	474	489	487	centrin-2, putative
PVP01_0321100	340	325	724	340	432	conserved Plasmodium protein, unknown function
PVP01_0903800	454	330	363	380	382	hypothetical protein
PVP01_1303100	369	339	406	361	369	ubiquitin-conjugating enzyme E2, putative
PVP01_1244000	271	623	160	389	361	glyceraldehyde-3-phosphate dehydrogenase,
PVP01_1034400	313	313	420	308	339	Plasmodium exported protein, unknown function
PVP01_0504300	391	350	285	290	329	conserved Plasmodium protein, unknown function
PVP01_0730700	341	280	503	160	321	conserved Plasmodium protein, unknown function
PVP01_0102200	276	206	522	218	305	PIR protein, pseudogene
PVP01_0607600	292	269	311	340	303	serine/arginine-rich splicing factor 4,
PVP01_1131900	314	296	274	319	301	conserved Plasmodium protein, unknown function
PVP01_1020200	205	275	374	341	299	secreted ookinete protein, putative
PVP01_1312300	254	325	264	330	293	ribonucleoside-diphosphate reductase, large
PVP01_0530400	197	193	447	299	284	conserved Plasmodium protein, unknown function
PVP01_0512900	259	281	290	272	276	conserved Plasmodium protein, unknown function
PVP01_1133400	278	350	225	249	275	RNA and export factor binding protein, putative
PVP01_1033800	262	107	576	135	270	Plasmodium exported protein, unknown function
PVP01_0710400	267	269	227	259	255	inhibitor of cysteine proteases, putative
PVP01_1136000	283	295	178	238	248	conserved Plasmodium protein, unknown function
PVP01_1004500	204	248	257	263	243	conserved Plasmodium protein, unknown function
PVP01_1402200	167	96	674	26	241	Plasmodium exported protein, unknown function

PVP01_0010990	233	207	313	210	241	Plasmodium exported protein, unknown function
PVP01_1336900	175	264	273	214	231	p1/s1 nuclease, putative
PVP01_0816700	209	227	299	188	231	survival motor neuron-like protein, putative
PVP01_1143000	281	237	246	155	230	long chain polyunsaturated fatty acid elongation
PVP01_0530800	202	172	219	320	228	alpha tubulin 2, putative
PVP01_1135900	236	141	229	290	224	conserved Plasmodium protein, unknown function
PVP01_0528600	184	298	156	242	220	CGI-141 protein homolog, putative
PVP01_1247100	245	260	184	184	218	bax inhibitor 1, putative
PVP01_1147200	200	169	289	158	204	Plasmodium exported protein, unknown function
PVP01_1471900	310	183	115	209	204	RAD protein (Pv-fam-e)
PVP01_1439200	209	195	215	194	203	heterochromatin protein 1, putative
PVP01_0110200	356	137	140	168	200	conserved Plasmodium protein, unknown function
PVP01_0730200	201	195	220	150	192	conserved Plasmodium protein, unknown function
PVP01_0916600	156	211	209	174	188	rhoptry neck protein 4
PVP01_1113500	183	254	151	158	186	conserved Plasmodium protein, unknown function
PVP01_1412600	166	175	218	147	176	conserved Plasmodium protein, unknown function
PVP01_0928300	161	156	187	152	164	CRAL/TRIO domain-containing protein, putative
PVP01_0915400	156	157	151	190	163	serine/threonine-protein kinase PRP4K, putative
PVP01_1134500	124	219	144	160	162	conserved Plasmodium protein, unknown function
PVP01_1229000	187	161	78	196	155	mitochondrial fission 1 protein, putative
PVP01_1116400	170	123	161	164	155	serine/threonine protein phosphatase 5,
PVP01_1463500	188	90	271	65	154	myosin A tail domain interacting protein,
PVP01_1125900	131	169	124	183	152	conserved Plasmodium protein, unknown function
PVP01_0825600	143	158	139	165	151	conserved Plasmodium protein, unknown function
PVP01_1216900	132	187	141	142	151	calcium-dependent protein kinase 5, putative
PVP01_0010220	127	48	119	302	149	merozoite surface protein 3, putative
PVP01_1241400	143	113	161	173	148	conserved Plasmodium protein, unknown function
PVP01_0920600	132	144	138	161	144	conserved Plasmodium protein,

						unknown function
PVP01_1206600	136	141	145	153	144	step II splicing factor, putative
PVP01_0618200	153	155	124	129	140	S-adenosylmethionine decarboxylase/ornithine
PVP01_1441600	125	153	110	169	139	phospholipid-transporting ATPase, putative
PVP01_1006200	90	205	133	107	134	conserved Plasmodium protein, unknown function
PVP01_0316300	146	117	143	121	132	conserved Plasmodium protein, unknown function
PVP01_0729900	126	154	100	137	129	selenoprotein, putative
PVP01_0733800	176	117	130	89	128	conserved Plasmodium protein, unknown function
PVP01_0945800	117	168	92	134	128	coatamer subunit gamma, putative
PVP01_1232800	141	156	97	114	127	protein transport protein SFT2, putative
PVP01_1218400	142	128	104	132	126	conserved Plasmodium protein, unknown function
PVP01_0733200	112	126	112	151	125	vacuolar ATP synthase subunit e, putative
PVP01_1022400	98	168	116	117	125	apical rhoptry neck protein, putative
PVP01_1449300	122	114	118	134	122	ras-related protein Rab-2, putative
PVP01_0712800	103	116	147	117	121	conserved Plasmodium protein, unknown function
PVP01_0517800	179	94	87	117	119	protein kinase c inhibitor-like protein,
PVP01_0109000	105	125	118	126	118	cloroquine resistance associated protein Cg8,
PVP01_0923900	105	133	110	120	117	RING zinc finger protein, putative
PVP01_1125000	173	109	87	93	116	anaphase promoting complex subunit, putative
PVP01_0726900	109	128	96	125	115	conserved Plasmodium protein, unknown function
PVP01_1442400	110	105	112	126	113	conserved Plasmodium protein, unknown function
PVP01_0111900	83	137	94	139	113	regulator of chromosome condensation, putative
PVP01_1241800	120	92	117	124	113	conserved Plasmodium protein, unknown function
PVP01_1033600	96	167	100	87	112	skeleton-binding protein 1, putative
PVP01_0823400	114	143	57	133	112	HVA22/TB2/DP1 family protein, putative
PVP01_1230400	92	96	146	111	111	micro-fibrillar-associated protein, putative
PVP01_1111800	87	101	146	110	111	RNA-binding protein, putative
PVP01_1316000	102	86	145	108	110	DNA topoisomerase II, putative
PVP01_0921000	137	78	130	96	110	alpha/beta hydrolase fold domain

PVP01_1138300	108	114	94	123	110	containing transcription elongation factor SPT5, putative
PVP01_0317600	102	125	70	131	107	conserved Plasmodium protein, unknown function
PVP01_0920800	80	157	81	109	107	conserved Plasmodium protein, unknown function
PVP01_0201900	115	90	109	112	106	PIR protein
PVP01_1437400	95	118	91	122	106	arginyl-tRNA synthetase, putative
PVP01_1329200	129	91	93	108	105	conserved Plasmodium protein, unknown function
PVP01_1430000	151	83	90	96	105	protein phosphatase PPM5, putative
PVP01_0603500	107	122	86	102	105	conserved Plasmodium protein, unknown function
PVP01_0505500	85	97	117	113	103	conserved Plasmodium protein, unknown function
PVP01_1212900	103	104	88	115	102	vacuolar ATP synthase subunit d, putative
PVP01_1263900	91	99	105	113	102	conserved Plasmodium protein, unknown function
PVP01_0934600	102	104	81	119	102	RNA (uracil-5-)methyltransferase, putative
PVP01_1212100	137	79	115	73	101	transcription activator, putative
PVP01_0409400	91	105	86	119	100	conserved Plasmodium protein, unknown function
PVP01_1261400	88	105	128	80	100	RNA-binding protein, putative
PVP01_1301500	103	74	121	103	100	mitochondrial phosphate carrier protein,

Genes with mean FPKMs over 100 from the *P. vivax* RNA-Seq data not in the *P. vivax* microarray dataset. It is possible that some microarray probes overlap this data, but naming and annotation inconsistencies between the *P. vivax* microarray dataset and the *P. vivax* P01 reference genome made them difficult to detect. FPKMs of the 4 isolates computed by Cufflinks from assemblies mapped with TopHat to *P. vivax* P01.

Supplementary Table D: *P. vivax* recombinant protein ectodomains

Accession number	Common name	Length (aa)	Expected size (kDa)	Region expressed
<u>PVX_099980</u>	MSP1	1702	215	E20-P1721
<u>PVX_097670</u>	MSP3.1, MSP3A ² , MSP3 γ	825	114	N21-K845
<u>PVX_097680</u>	MSP3.3, MSP3C ² , MSP3 β	997	133	D20-K1016
<u>PVX_097685</u>	MSP3.4, MSP3D1	1091	140	N21-M1111
<u>PVX_097720</u>	MSP3.10, MSP3H ² , MSP3 α	829	113	E24-W852
<u>PVX_003775</u>	MSP4 ²	202	45	A26-S227
<u>PVX_003770</u>	MSP5 ²	346	62	R22-S367
<u>PVX_082700</u>	MSP7.1 ³	399	70	E22-Y420
<u>PVX_082675</u>	MSP7.6 ³	428	72	A26-N453
<u>PVX_082655</u>	MSP7.9 ³	364	65	E24-V387
<u>PVX_114145</u>	MSP10 ²	435	72	A23-S457
<u>PVX_113775</u>	P12	316	61	F24-A339
<u>PVX_113780</u>	P12p ²	395	68	V24-P418
<u>PVX_097960</u>	P38 ²	306	60	K29-G334
<u>PVX_000995</u>	P41	363	67	E22-E384
<u>PVX_115165</u>	P92	833	118	D23-H855
<u>PVX_001725</u>	RON12	274	54	L25-S298
<u>PVX_088910</u>	GAMA	729	103	L21-S749
<u>PVX_090075</u>	Pv34 ²	323	61	N25-S347
<u>PVX_090210</u>	ARP	265	54	K21-P285
<u>PVX_090240</u>	CyRPA	344	65	T23-D366
<u>PVX_095055</u>	RIPR	1054	144	N22-A1075
<u>PVX_098712</u>	RhopH3	866	125	R25-T890
<u>PVX_110810</u>	DBP	986	135	V23-T1008
	DBP-RII	328	64	D194-T521
<u>PVX_111290</u>	MTRAP	292	58	K24-G315
<u>PVX_121885</u>	CLAG ⁴	1138	159	Y25-L1162
<u>PVX_081550</u>		473	80	R23-F495

<u>PVX_084815</u>		249	53	R22-A270 ⁵
<u>PVX_084970</u>		892	121	D24-S915
<u>PVX_116775</u>		314	60	E21-S334
<u>PVX_101590</u>	RBP2-like ⁴	619	97	K22-K640
<u>PVX_001015</u>		330	62	Q28-T357
<u>PVX_110945</u>		316	61	D28-K343
<u>PVX_110950</u>		379	66	K24-L402
<u>PVX_110960</u>		648	92	A23-K670
<u>PVX_110965</u>		421	69	D21-A441

Boundaries of expressed regions are delimited by the N- and C-terminal amino acid residues according to their codon positions along the *P. vivax* Sal 1 reference protein sequence.

¹MSP9 and MSP1P failed to be sub-cloned or expressed, respectively

²Based on genedb.org annotation

³Based on MSP7 family naming in (Kadekoppala and Holder, 2010)

⁴Based on *P. vivax* product description

⁵S270 in unmodified sequence

