UNIVERSITY OF CALIFORNIA, SAN DIEGO

Battling Drug Resistance: A Tale of Two Pathogens

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Biology

by

Justine Swann

Committee in charge:

Professor Elizabeth Winzeler, Chair
Professor Deborah Spector, Co-Chair
Professor Janelle Ayres
Professor Pieter Dorrestein
Professor Emily Troemel

2016

The Dissertation of Justine Swann is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

Co-Chair

_____

Chair

University of California, San Diego

2016

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SUPPLEMENTAL DATASETS

ACKNOWLEDGEMENTS

I would first like to express my deep appreciation for my thesis advisors, John Young and Elizabeth Winzeler, who have both been instrumental in my development as a scientist. Without their support, this work would not have been possible. I am very grateful for my time in both of their labs. I would also like to thank the members of the Young lab, including John Naughton, Jeff Murry, Arturo Diaz, John Marlett, Mellissa Rodgers, Rose Pilpa, Robyn Kaake and Shannon Siedel, for their guidance and assistance throughout my time at the Salk Institute. I am also indebted to all of the members of the Winzeler lab, especially Stephan Meister, Greg LaMonte, Pamela Orjuela-Sanchez, Victoria Corey, Marissa Hovlid, Jenya Antonova, Christin Reimer, Melaniee Wree, and Erika Flannery for their helpful support and advice during my time as a graduate student in the Winzeler lab. I would also like to thank Paula Maugina and Salley Ganley for support and administrative assistance.

I would like to acknowledge all of the members of my dissertation committee, including Deborah Spector, Emily Troemel, Janelle Ayres, and Pieter Dorrestein. They have all generously taken the time to attend my committee meetings and to offer insightful feedback and constructive suggestions about my research. I would especially like to thank Dr. Spector for providing guidance during my transition from John's lab to Elizabeth's lab.

This work would not be possible without the support of our collaborators. I would like to thank Kevin Olivieri and Manqing Li for helpful

discussion and reagents, and the Salk Gene Transfer, Targeting, and Therapeutics Core (GT3) for lentivirus production. We thank our colleagues from The Genomics Institute of the Novartis Research Foundation, in particular Richard Glynne, Purvi Sanghvi, Annie Mak and Jason Matzen. They provided insight, expertise, and assistance that greatly assisted the research. The Insectary Core Facility at New York University supplied the mosquitos needed for sporozoite dissections described in Chapter 2 and in the Appendix. The assessment of RNA quality and dual RNA-sequencing described in the Appendix was performed by the IGM Core at UCSD. I would especially like to thank Kristen Jepson for her helpful discussion during this process. I would also like to thank Nathan Lewis, Shangzhong Li, and Alyaa M. Abdel-Haleem for performing the dual RNA-sequencing read mapping and differential gene expression analysis described in section A.2 of the Appendix, and for their advice and discussion. We thank Lauren Mack and Ye Zheng for the use of their BD FACS Aria for sorting. I would like to acknowledge Greg Golden for the generation of the Muc13 promoter reporter lentivirus and Pamela Orjuela-Sanchez for the confocal microscopy analysis described in section A.3. The Salk Institute GT3 Viral Vector Core performed tittering of the lentivirus. I thank Emily Troemel for her suggestion to use APEX2-based proximity labeling to characterize the *Plasmodium* liver-stage secretome as described in A.4, and to Tomas Bos for the construction of the APEX2-based lentiviral vectors listed in Table A.5, and to Aaron Reinke for helpful discussion. I am also grateful to

David Gonzales and John Lapek for their collaboration on the APEX2-based proteomics project described in section A.4.

I am grateful to the many sources of funding that have supported my graduate work. The research presented in Chapter 2 was supported by a Salk-Sanofi innovation grant, the National Institutes for Health Cellular and Molecular Training Grant, the Bruce J. Heim Foundation, the Jesse and Caryl Philips Foundation, the James B. Pendleton Charitable Trust, and the NCI Cancer Center Support Grant Award. I would also like to acknowledge the Medicines for Malaria Venture (MMV), the National Science Foundation (NSF), the Welcome Trust, the Bill and Melinda Gates Foundation, and NIH, for funding the research described in Chapter 3. The work in Chapter 4 was completed with generous support from the National Institutes for Health, the Bill and Melinda Gates Foundation, The Medicines for Malaria Venture, and the Novo Nordisk Foundation provided to the Center for Biosustainability at the Technical University of Denmark.

In addition, I would like to acknowledge the UCSD/ Salk Institute Biology PhD Program for being a highly organized and effective training program that has allowed me to develop into a competent biologist. I especially want to thank Thomas Tomp, Cathy Pugh, and Marifel Alforo for their administrative support throughout the years in the program and for their helpful advice.

I am tremendously grateful for the support of my family. My parents, Jeff and Melody Swann, have worked hard to make sure I was given every opportunity throughout my childhood and have been extremely supportive throughout every step of my career. They taught me the importance of commitment and tenacity, both of which were important during my graduate work. I am also very thankful for my identical twin sister and best friend, Lyndsay Swann, who is always there for me when I need her.

I would also like to thank my husband, Dave Albin, for his loving support. I would not have survived graduate school without his constant encouragement, patience, and enthusiasm during all of the ups and downs of my research. He has motivated me to pursue my ambitions and has supported me every step of my journey. I am also thankful for my amazing son, Oliver Albin, for making me the happiest mom there is. He is the light of my life and fills my days with purpose. His toothless smiles were the remedy for any failed experiment.

Chapter 2, in full, has been published in Virology Journal, 2016. Justine Swann, Jeff Murry, John A.T. Young, “Cytosolic sulfotranserase 1A1 regulates HIV-1 reverse transcription in human monocyte-derived macrophages”. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in full, has been published in ACS Infectious Diseases, 2016. Justine Swann, Victoria Doroski, Matthew Abraham, Sang Kim, Kerstin Henson, Maureen Ibanez, David Plouffe, Anne-Marie Zeeman, Clemens H. M.

Kocken, Brice Campo, Case W. McNamara, Elizabeth A. Winzeler, Stephan Meister. “A luciferase-based drug screen identifies MMV compounds with liver stage antimalarial activity.” The dissertation author was the primary investigator and author of this paper.

Chapter 4, in full, has been published in Wiley Interdisciplinary Reviews: Systems Biology and Medicine, 2015. Justine Swann, Neema Jamshidi, Nathan Lewis, Elizabeth Winzeler. “Systems analysis of host-parasite interactions”. The dissertation author was the primary investigator and author of this paper.

The Appendix contains material that may be prepared for publication at a future time. Justine Swann, Neema Jamshidi, Shangzhong Li, Alyaa M. Abdel-Haleem, Nathan Lewis, Pamela Orjuela, Greg Golden, John Lapek, Tomas Bos, David Gonzalez, Elizabeth Winzeler. “Systems biology of *Plasmodium* liver-stage development.”

VITA

2009 Bachelor of Science, University of California, Santa Barbara

2016 Doctor of Philosophy, University of California, San Diego

PUBLICATIONS

1. Maya F. Kotturi, **Justine A. Swann,** Bjoern Peters, Cecilia Lindestam Arlehamn, John Sidney, Ravi V. Kolla, Edward A. James, Rama S. Akondy, Rafi Ahmed, William W. Kwok, Michael J. Buchmeier, and Alessandro Sette. "Human CD8+ and CD4+ T cell memory to lymphocytic choriomeningitis virus infection." *Journal of virology,* 85.22 (2011): 11770-11780.

2. Cecilia S. Lindestam Arlehamn, Anna Gerasimova, Federico Mele, Ryan Henderson, **Justine Swann**, Jason A. Greenbaum, Yohan Kim, John Sidney, Eddie A. James, Randy Taplitz, Denise M. McKinney, William W. Kwok, Howard Grey, Federica Sallusto, Bjoern Peters and Alessandro Sette. "Memory T cells in latent Mycobacterium tuberculosis infection are directed against three antigenic islands and largely contained in a CXCR3+ CCR6+ Th1 subset." *PLoS Pathogens,* 9.1 (2013): e1003130.

3. Jeffrey Murry, Joseph Godoy, Amey Mukim, **Justine Swann**, James Bruce, Paul Ahlquist, Alberto Bosque, Vicente Planelles, Celsa Spina, and John Young. "Sulfonation Pathway Inhibitors Block Reactivation of Latent HIV-1". *Virology Journal,* 471 (2014): 1-12.

4. **Justine Swann,** Neema Jamshidi, Nathan Lewis, Elizabeth Winzeler. "Systems analysis of host-parasite interactions". *Wiley Interdisciplinary Reviews: Systems Biology and Medicine, 7.6 (2015): 381-400.*

5. **Justine Swann**, Victoria Corey, Christina A. Scherer, Eamon Comer, Micah Maetani, Yevgeniya Antonova-Koch, Christin Reimer, Kerstin Gagaring, Maureen Ibanez, David Plouffe, Anne-Marie Zeeman, Clemens H. M. Kocken, Case W. McNamara, Stuart L. Schreiber, Brice

Campo, Elizabeth A. Winzeler, Stephan Meister. “A high-throughput luciferase-based assay for the discovery of therapeutics that prevent malaria.” *ACS Infectious Diseases,* (2016).

6. **Justine Swann**, Jeffrey Murry, John Young. “Cytosolic sulfotranserase 1A1 regulates HIV-1 reverse transcription in human monocyte-derived macrophages.” *Virology Journal*, (2016).

FIELDS OF STUDY

Biology, specialization in microbiology and molecular/ cellular biology

Professor John Young

Professor Elizabeth Winzeler

ABSTRACT OF THE DISSERTATION

Battling Drug Resistance: A Tale of Two Pathogens

by

Justine Swann

Doctor of Philosophy in Biology

University of California, San Diego, 2016

Professor Elizabeth Winzeler, Chair

Professor Deborah Spector, Co-Chair

HIV/AIDS, malaria, and tuberculosis are the world's three deadliest infectious diseases of humans, often referred to as the 'Big Three'. Together, they are responsible for more than 10% of all the deaths worldwide each year. What is perhaps most worrisome is the fact that the current therapies that treat

these conditions are losing their efficacy due to the emergence of antimicrobial drug resistance. Accordingly, research is urgently needed to address the growing problem of drug resistance and to help drive the development of novel therapeutics. In the first chapter of this doctoral dissertation, three strategies to combat drug resistance are discussed: 1. developing therapeutics that target host-derived factors, 2. identifying new antimicrobial inhibitors, and 3. investigating host-pathogen biology using systems analysis. Examples of research utilizing these strategies are discussed in the following chapters, with a particular focus on two of the "Big Three" pathogens, HIV-1 and the malaria parasite, *Plasmodium*. The identification and characterization of a novel host factor that regulates HIV-1 reverse transcription is described in Chapter 2. In Chapter 3, the development of a high-throughput phenotypic assay to identify novel antimalarial drugs is discussed, and in Chapter 4, a broad review of systems biology-based research of host-parasite interactions (with an emphasis on *Plasmodium*) is included. The Appendix includes preliminary data and future directions for systems biology research aimed at understanding *Plasmodium* liver-stage development. Through the combination of these scientific efforts, we will surely strength our position in the ongoing battle against antimicrobial drug resistance.

1. Introduction

1.1. The "Big Three"

The three deadliest infectious diseases of humans include HIV/AIDS, malaria, and tuberculosis. Together, they cause millions of deaths per year and substantial morbidity, socioeconomic decline, and suffering for over a billion more people. Accordingly, these diseases are often referred to as the "Big Three" (Bourzac, 2014). Despite recent funding efforts from many global health organizations such as the Bill and Melinda Gates Foundation, the Welcome Trust, and the Global Fund to Fight AIDS, Tuberculosis, and Malaria, we are still a long way from developing a cure for these deadly infections.

Although HIV/AIDS, malaria, and tuberculosis are caused by very different pathogens (a virus, a parasite, and a bacterium, respectively), these infectious diseases share common characteristics that contribute to their growing lethality: 1.lack of effective vaccines, 2. no natural immunity, and 3. latency and reoccurrence (Bourzac, 2014). Of urgent importance is the increasing report of drug resistance emerging to all of these microbes, threatening to undermine current treatment regimes. For example, close to 10% of HIV-1 patients are now infected with a strain of drug-resistant virus in high-income countries (Wittkop et al., 2011), and data suggests that as access and

distribution of first-line antiretroviral therapies (ART) continues to increase in low-income countries, so will HIV-1 drug resistance (WHO, 2015). Likewise, malaria parasite drug resistance to nearly all of the current antimalarial drugs has been reported worldwide (Cui et al., 2015; Flannery et al., 2013), and infection with multi-drug resistant (MDR) *Mycobacterium tuberculosis* (Mtb) is becoming increasingly more common each year (Keshavjee and Farmer, 2012). New research is therefore needed in order to counteract antimicrobial drug resistance to the “Big Three”.

1.2. Strategies to fight antimicrobial drug resistance

1.2.1. Targeting host factors

An attractive strategy to circumvent antimicrobial drug resistance is to develop therapeutics that target host-derived factors essential for pathogen entry, survival, and replication, rather than targeting the pathogen directly. While rapidly mutating microbes evolve drug resistance at an impressive speed, humans evolve much more slowly and thus developing drugs that target host factors may slow the development of antimicrobial drug resistance. Accordingly, there has been an increasing effort to develop host-directed therapeutics for HIV/AIDS (Arhel and Kirchhoff, 2010), malaria (Prudencio and Mota, 2013), and tuberculosis (Hawn et al., 2015).

A notable example of a successful host factor-directed therapy is the targeting of the HIV-1 co-receptor chemokine receptor 5 (CCR5), which inhibits viral entry. Maraviroc, the first FDA approved HIV-1 drug to target a host factor, is a CCR5 antagonist that was identified by high-throughput screening of a chemical library (Fatkenheuer et al., 2005). Other CCR5-targeted therapies are currently in clinical trials (Gu and Chen, 2014). Moreover, patients with a naturally occurring deletion of CCR5 (Δ32) are highly protected from infection with HIV-1 (Dean et al., 1996; Samson et al., 1996). Host-directed therapies have also been investigated for the treatment of infection with Mtb, including the disruption of the host signaling pathways

needed for replication in macrophages, as well as the inhibition of host factors involved in bacterial uptake and trafficking (Hawn et al., 2015).

Various screening methods have been implemented in recent years in order to identify host factors necessary for microbial infection. These include both forward and reverse chemical genetic and traditional genetic approaches. High-throughput screening using RNA interference (RNAi) has been one of the most commonly used methods for identifying host factors. This is especially true in the HIV-1 field, where many siRNA genome-wide screens have been performed to identify host modulators of HIV-1 replication (Pache et al., 2011). Surprisingly, a meta-analysis of genome-wide HIV-1 host factor screens found limited overlap between screening hits (Bushman et al., 2009). Computational analysis later revealed that the limited overlap between these studies was primarily due to false negatives, rather than false positives, providing further evidence for the utility of RNAi screening approaches (Hao et al., 2013). Other concerns associated with the use of RNAi, such as off-target effects and gene essentiality, can now be addressed using recently developed CRISPR-Cas9 genetic modification systems (Hsu et al., 2014). Smaller scale siRNA screens targeting a more focused set of genes, such as kinases, have also been important for identifying host factors that regulate malaria (Prudencio et al., 2008b) and tuberculosis (Jayaswal et al., 2010) pathogenesis. Likewise, knockdown of candidate host factors identified through other means can help elucidate cellular mechanisms controlling infection.

One of the major criticisms of developing therapies that target host factors is the increased risk for host cellular toxicity. A major advantage of traditional pathogen-directed treatments is that the genetic divergence between host and pathogen limits adverse effects to the host. Therefore, researchers need to critically examine host responses both *in vitro* and *in vivo* when evaluating host-directed therapies. This includes the assessment of cytotoxicity during initial high-throughput screening efforts (Bushman et al., 2009), throughout clinical trials (Gulick et al., 2007) and even after pharmaceutical development (Manfredi et al., 2015). While toxicity may limit the efficacy of many host factor-directed candidate drugs, there have still been many very successful host-directed treatments. A notable example is the CCR5-directed therapies, which, as mentioned above, does not seem to significantly affect patient health (Manfredi et al., 2015).

While in theory host-directed therapies may provide a formidable barrier to the development of drug resistance, it should be noted that this does not mean that microbial drug resistance will not eventually develop to these treatments. In fact, there have been recent reports of pathogens evolving to bypass their dependency on host factors. For example, van der Schaar and colleagues have recently described viral escape mutants that have mutated to overcome the need for host phosphatidylinositol-4-phosphate (PI4P) and are consequently resistant to drugs known to inhibit host PI4KIIIβ kinase (van der Schaar et al., 2012). Resistance to Maraviroc has also been reported in strains

of HIV-1 through mutations in the viral envelope glycoprotein gp120 (Ratcliff et al., 2013). However, it is still very likely that the development of drug resistance to host factors will be a much slower process than to other drugs (Ruiz and Russell, 2012), and more research is therefore needed in order to understand these processes. With all of this in mind, it is generally agreed upon that the best strategy for developing new host-directed treatments is to integrate them into combination-based therapies. For example, it has been recently demonstrated that treatment with Maraviroc improves patient outcomes when used in combination with standard ART therapies (Gulick et al., 2008). Other therapeutics involving host-factor targeting will likely also benefit from a combinatorial approach.

1.2.2. Developing new antimicrobial drugs

Although antimicrobial drug resistance continues to escalate, the development of new antimicrobial agents remains relatively stagnant. The excessive costs of pushing a new drug through the development pipeline ($70 million just for phase III clinical trials), coupled with low profit margins and the risk of future drug resistance, prevents biotech and pharmaceutical companies from actively pursuing the development of anti-infective agents (Cooper and Shlaes, 2011). Despite these challenges, there has been a tremendous effort by the academic community to identify new drugs to fight the “Big Three”. The analysis of high-throughput screening data produced by these research

institutions is the essential first step towards the identification of novel chemical scaffolds with antimicrobial activity.

There are a number of different high-throughput drug screening strategies that have been successfully utilized to identify new antimicrobial agents. These strategies fall into two main categories: 1. phenotypic screens, and 2. target-based screens. Phenotypic screens assess the overall inhibitory activity of a chemical towards the pathogen in question, whereas target-based screens measure the specific activity of a chemical against a purified microbial target *in vitro* (Swinney, 2013) or *in silico* via virtual screening using protein docking programs (Bissantz et al., 2000). The use of target-based virtual screening is an attractive cost-effective option, and has been used to identify potential inhibitors of HIV-1 (Chang et al., 2010), *Plasmodium spp.* (the parasite that causes malaria) (Salzemann et al., 2007; Shah et al., 2015), and Mtb (Cho et al., 2008). While both approaches have their merits, there has been recent concern that target-based screening may not be as effective as originally thought. For example, there is evidence of pharmaceutical industry productivity decline coinciding with the introduction of target-based screening (Sams-Dodd, 2005). This may be coincidental, or may be due to underestimation of the complexity of microbial physiology that occurs when focusing on a single target (Sams-Dodd, 2005). A newer approach, termed 'network pharmacology', seeks to address this issue by exploiting systems-biology to rationally design drugs that act on multiple proteins in a pathway

rather than a single target in isolation (Hopkins, 2008). Even so, there are a limited number of validated targets for some pathogens, such as *Plasmodium*, and thus forward-based screening approaches have the advantage of identifying inhibitors of uncharacterized targets.

There are a number of different types of chemical libraries that can be analyzed, regardless of the screening strategy. To identify novel antimicrobial drugs while maximizing the chemical space, libraries consisting of diversity synthesis-oriented compounds (Tan, 2005) and natural products (Mishra and Tiwari, 2011) have been particularly useful. On the other hand, it is extremely expensive to develop new classes of drugs; therefore, there has been a growing movement toward drug 'repositioning' or 'repurposing', where libraries of FDA-approved drugs can be screened for new uses. This strategy, besides being much more cost-effective, can also save time, as there is generally already a wealth of information about toxicity, dosing, and potential side effects (Savoia, 2015). For example, the drug entecavir, which was originally approved for the treatment of chronic hepatitis B virus (HBV) infection, was found to inhibit HIV-1 at clinically relevant doses (McMahon et al., 2007). Additionally, a number of drugs that were initially developed to inhibit other pathogens have been found to be effective against Mtb, including fluoroquinolones, rifamycins, riminophenazines, oxazolidinones, and β-lactams (Zumla et al., 2013). One of the best examples of successful drug repurposing efforts is for the treatment of malaria, as many already approved

antibiotics are also effective against the parasite, including doxycycline, clindamycin, dapsone, and the sulphonamides (Andrews et al., 2014).

A key objective during the development of new antimicrobials should be to generate compounds less prone to drug resistance. As an exciting example, researchers have recently identified a new antibiotic called teixobactin that seems to resist the development of drug resistance in cultures of *Staphylococcus aureus* and Mtb (Ling et al., 2015). Future research will surely shed light on the molecular mechanisms behind teixobactin's efficacy and the specific properties that block drug resistance. Another strategy that may slow drug resistance is the generation of 'nanoantibitoics', which can stay in the body longer, reduce acute toxicity, and lower development costs (Huh and Kwon, 2011). In addition, designing compounds that target specific stages of the pathogen that are less prone to the development of drug resistance is another desirable option. For example, it has been suggested that drugs targeting the liver-stage of malaria may be less prone to drug resistance (Delves et al., 2012). This is largely due to the fact that the asymptomatic liver-stage represents a 'bottleneck' population of only a few thousand parasites, compared to the billions of blood-stage parasites that are able to evolve resistance much faster.

1.2.3. Using systems biology to understand host-pathogen interactions

Systems biology research has been a valuable weapon in the fight against antimicrobial drug resistance. This relatively new field of biology, based on the generation of whole-cell or whole-organism high-throughput data, took off in the early 90s after the sequencing of the human genome (Venter et al., 2001). The widespread use of the internet facilitated the deposition and distribution of newly generated large datasets (Schneider, 2013). Since then, the post-genomic era has experienced a shift away from traditional 'reductionist' biology research, which analyzes complex biological processes by 'reducing' them into their simpler components, towards systems-analysis, which generates biological understanding through the computational evaluation of system-wide datasets. Accordingly, systems biology has been fundamental to our understanding of the host-pathogen relationships that drive infection (Fruh et al., 2010).

The integration and interpretation of different types of large host-pathogen datasets is critical in order to fully understand infection biology and the development of drug resistance. The major classes of '-omic' methods include genomics, proteomics, and metabolomics, corresponding to the system-wide collection of DNA, protein, and metabolite information, respectively (Joyce and Palsson, 2006). The computational analysis of this 'big data' is fundamental to systems biology (Durmus et al., 2015). Consequently, collaborations between the 'wet' and 'dry' biology labs (bench

science and computer science, respectively) have greatly facilitated host-pathogen systems biology research.

The systems-level analysis of large '–omic' datasets has led to the identification and characterization of novel antimicrobial drug targets and pathways. This is especially important for pathogens with complex organismal biology, such as bacteria and parasites. For example, the *in silico* reconstruction of the Mtb metabolic network from publically available '-omic' host-pathogen data has revealed a list of alternative drug targets to consider when developing new treatments for tuberculosis (Jamshidi and Palsson, 2007). Additionally, systems-based analyses have uncovered important parasite processes that may be targeted by new therapeutics, as reviewed in Chapter 4.

Furthermore, systems-based analyses have been the driving force behind the identification of novel host factors. Functional genomic techniques, including both forward and reverse genetic strategies (such as by chemical mutation or by siRNA screening, respectively), have been especially useful, as discussed in the previous section. The computational analysis of candidate host factors arising from primary screens has revealed complex host-pathogen interaction networks, including pathways critical for pathogen survival and replication (Bushman et al., 2009). These biological processes may include desirable drug targets.

Finally, systems biology has began to uncover the molecular mechanisms that lead to the development of pathogen drug resistance. By artificially selecting for drug resistant pathogens in the lab and whole genome sequencing the resulting mutants (compared to their respective parent strain), researchers are able to pinpoint specific mutations that may confer drug resistance. This strategy has been highly successful for many pathogens, especially *Plasmodium spp.*, which have large, complex, and relatively uncharacterized genomes (Meister et al., 2011; Rottmann et al., 2010; Sidhu et al., 2002). Similarly, whole-genome analysis of drug resistant pathogens from patients *ex vivo* has revealed a number of putative drug resistance genes (Ariey et al., 2014; Dharia et al., 2010).

1.3. Summary

The spread of antimicrobial drug resistance to the 'Big Three' pathogens (HIV, *Plasmosium spp.*, and Mtb) is a serious threat to human health that warrants immediate attention. These three microbes are currently responsible for over 10% of all of the deaths worldwide each year, and if drug resistance continues to increase, this statistic is sure to escalate. We are already experiencing microbial infections that defy all currently approved treatments, therefore new research is urgently needed in order to develop new therapeutics. Several strategies have been particularly successful in the fight against drug resistance, including the characterization of host factors, identification of new antimicrobial drugs, and analysis of host-pathogen systems biology. In this dissertation, specific examples of each of these general approaches will be discussed. Chapter 2 describes the identification and characterization of a novel HIV-1 host factor that was found to regulate an important step during retroviral reverse transcription in monocyte-derived macrophages. In Chapter 3, an improved assay for the high-throughput identification of drugs that inhibit malaria liver-stage development is introduced, and the screening results from several chemical libraries are described. Finally, Chapter 4 includes a review of systems biology research of host-parasite interactions, and the Appendix includes preliminary data and

future directions for several systems biology-based projects that investigate *Plasmodium* liver-stage development.

2. Cytosolic sulfotransferase 1A1 regulates HIV-1 minus-strand DNA elongation in primary human monocyte-derived macrophages

2.1. Abstract

The cellular sulfonation pathway modulates key steps of virus replication. This pathway comprises two main families of sulfonate-conjugating enzymes: Golgi sulfotransferases, which sulfonate proteins, glycoproteins, glycolipids and proteoglycans; and cytosolic sulfotransferases (SULTs), which sulfonate various small molecules including hormones, neurotransmitters, and xenobiotics. Sulfonation controls the functions of numerous cellular factors such as those involved in cell-cell interactions, cell signaling, and small molecule detoxification. We previously showed that the cellular sulfonation pathway regulates HIV-1 gene expression and reactivation from latency. Here we show that a specific cellular sulfotransferase can regulate HIV-1 replication in primary human monocyte-derived macrophages (MDMs) by yet another mechanism, namely reverse transcription.

MDMs were derived from monocytes isolated from donor peripheral blood mononuclear cells (PBMCs) obtained from the San Diego Blood Bank. After one week *in vitro* cell culture under macrophage-polarizing conditions,

MDMs were transfected with sulfotranserase-specific or control siRNAs and infected with HIV-1 or SIV constructs expressing a luciferase reporter. Infection levels were subsequently monitored by luminescence. Western blotting was used to assay siRNA knockdown and viral protein levels, and qPCR was used to measure viral RNA and DNA products.

We demonstrate that the cytosolic sulfotransferase SULT1A1 is highly expressed in primary human MDMs, and through siRNA knockdown experiments, we show that this enzyme promotes infection of MDMs by single cycle VSV-G pseudotyped human HIV-1 and simian immunodeficiency virus vectors and by replication-competent HIV-1. Quantitative PCR analysis revealed that SULT1A1 affects HIV-1 replication in MDMs by modulating the kinetics of minus-strand DNA elongation during reverse transcription.

These studies have identified SULT1A1 as a cellular regulator of HIV-1 reverse transcription in primary human MDMs. The normal substrates of this enzyme are small phenolic-like molecules, raising the possibility that one or more of these substrates may be involved. Targeting SULT1A1 and/or its substrate(s) may offer a novel host-directed strategy to improve HIV-1 therapeutics.

2.2. Introduction

Host cellular machinery is exploited to facilitate all steps of HIV-1 replication. A new paradigm in the treatment of HIV-1 infection is to target these so-called host-derived dependency factors, as exemplified by CCR5 coreceptor antagonists (Fatkenheuer et al., 2005). By contrast to the existing classes of direct-acting antivirals (DAAs) that target viral proteins, host-directed therapies may be less prone to the development of viral drug resistance and offer the potential for broader spectrum therapeutics (Law et al., 2013; Schwegmann and Brombacher, 2008). Therefore, there is currently a great deal of interest in better understanding the roles played by cellular factors and pathways during HIV-1 replication (Brass et al., 2008; Bushman et al., 2009; Konig et al., 2008; Yeung et al., 2009; Zhou et al., 2008).

The cellular sulfonation pathway was first shown to play a key role in regulating HIV-1 infection at the level of cellular entry (Choe et al., 2003; Farzan et al., 1999; Mondor et al., 1998; Roderiquez et al., 1995; Seibert et al., 2002). Previously, we uncovered another role for this pathway, demonstrating that it regulates retroviral transcription (Bruce et al., 2008). In that study, a forward genetic screen implicated two specific bi-functional 3'-phosphoadenosine-5'phosphosulfate (PAPS) synthetase enzymes, PAPSS1 and PAPSS2, as being important for retroviral replication (Bruce et al., 2008). These proteins catalyze two enzymatic steps to generate PAPS, the high-

energy sulfonate-donor used in all cellular sulfonation reactions (Klaassen and Boles, 1997). This specific effect was demonstrated using two different inhibitors of the sulfonation pathway: chlorate, a substrate analog of sulfate that blocks PAPS formation; and guaiacol, a sulfotransferase substrate mimic (Bruce et al., 2008; Hortin et al., 1988). In addition, we have recently shown that treatment with these chemical inhibitors significantly blocks HIV-1 reactivation from latency in a primary human CD4+ T cell model and in established cell lines where latency is maintained by diverse regulatory mechanisms (Murry et al., 2014).

The cellular sulfonation system consists of a family of sulfotransferase enzymes that are responsible for sulfonate transfer within the cell (Chapman et al., 2004). These proteins catalyze the transfer of a sulfuryl group (SO_3^-) from PAPS to a hydroxyl or amino-group on an acceptor molecule. Sulfotransferases are organized into two main sub-families: the Golgi and cytosolic sulfotransferases. The Golgi sulfotransferases are membrane bound enzymes that sulfonate cell surface proteins, carbohydrates, proteoglycans, and glycoproteins, while the cytosolic sulfotransferases are cytoplasmic enzymes that sulfonate endogenous hormones, neurotransmitters, and small metabolites as well as exogenous xenobiotics (Figure 2.2A). Cell surface sulfonation is necessary during normal homeostatic processes such as lymphocyte homing and cellular signaling (Hemmerich and Rosen, 2000; Kawashima et al., 2005; Kramer and Yost, 2003; Nadanaka et al., 2008).

Cytoplasmic sulfonation of metabolites by the cytosolic sulfotransferases (SULTs) generally leads to their inactivation, detoxification, and/or bioactivation (Allali-Hassani et al., 2007; Falany, 1997; Gamage et al., 2006; Glatt, 2000; Hildebrandt et al., 2007). Here we show that one of these cytosolic sulfotransferases, SULT1A1, regulates HIV-1 reverse transcription in MDMs, increasing our knowledge of the roles played by cellular factors in regulating HIV-1 reverse transcription

2.3. Results

2.3.1. Human cytosolic sulfotransferase 1A1 is highly expressed in primary human monocyte-derived macrophages

To determine if any of the 12 known cytosolic SULTs might play a role in regulating HIV-1 infection, we first compared their relative expression levels in the two physiologically relevant cell types, primary human CD4+ T cells and primary monocytic cells. We employed the publically available mRNA expression database Bio-GPS (Wu et al., 2009a). This analysis revealed a strikingly high level of SULT1A1 mRNA in CD14+ monocytes as compared to other cell types including CD4+ T cells (Figure 2.1). Although SULT1A1 was previously found to be important in tissues such as the liver, brain, kidney, and gastrointestinal tract, little or nothing is known about its role in immune cells (Riches et al., 2009; Salman et al., 2009; Teubner et al., 2007). MDM-selective expression was confirmed by immunoblot analysis of protein lysates obtained from primary human CD4+ T cells and MDMs. In agreement with the mRNA expression data, SULT1A1 was highly expressed in MDMs. However, SULT1A1 protein expression was undetectable in resting or activated CD4+ T cells (Figure 2.2B).

2.3.2. SULT1A1 knockdown decreases retroviral infection in MDMs

To test if SULT1A1 plays a role during retroviral infection in MDMs, siRNA knockdown was used to reduce the levels of this protein (Figure 2.3A). Three distinct SULT1A1-targeting siRNAs (#1-3) each reduced SULT1A1 protein levels by 70-80% at 96 hours post-transfection as compared to a control (scrambled sequence) siRNA (Figure 2.3B and Figure 2.3C middle panel). Each of the 3 SULT1A1 siRNAs also reduced the levels of infection seen with a VSV-G-pseudotyped HIV-1 NL43 virus vector as judged by monitoring expression of the firefly luciferase reporter enzyme from the viral genome (Connor et al., 1995). Knockdown of SULT1A1 expression with SULT1A1 siRNAs 1, 2, or 3 resulted in a decrease in HIV-1 reporter gene expression by 50% ($p<0.005$), 76% ($p<0.0005$), and 57% ($p<0.005$), respectively (Figure 2.3C, left panel). None of the siRNAs used significantly impacted cell viability (Figure 2.3C, right panel). The fact that all three SULT1A1-directed siRNAs reduced viral gene expression following infection strongly argues against an indirect off-target effect being responsible for this effect.

To test if SULT1A1 could also influence infection by the related simian immunodeficiency virus, SULT1A1 siRNA-treated and control siRNA-treated MDMs were infected with a VSV-G-pseudotyped SIVagm virus vector encoding firefly luciferase (Figure 2.3D) (Mariani et al., 2003). As for HIV-1, RNAi-mediated knockdown of SULT1A1 directly correlated with a significant

reduction in expression of the virus-encoded reporter enzyme, with an average of 85% inhibition ($p<0.0005$) (Figure 2.3D, left panel).

To further validate the role of SULT1A1 during HIV-1 infection, we next investigated the impact of knocking down expression of this protein on infection of MDMs by replication-competent HIV-1. MDMs that were transfected with control siRNA or with SULT1A1-siRNA#2 were challenged with the replication-competent CCR5-tropic HIV-1 Env+ JM1186 virus encoding Renilla luciferase (Bjorndal et al., 1997; Munch et al., 2007) and luciferase expression was assayed three days post infection (Figure 2.4A). SULT1A1 knockdown was associated with a significant inhibition of virus-encoded luciferase expression in MDMs derived from multiple donors without affecting cell viability (Figure 2.4). Transfection of these cells with SULT1A1 siRNA #2 resulted in an average 72% decrease in HIV-1 reporter expression compared to control ($p<0.0005$) (Figure 2.4, left panel). Taken together, these results demonstrate that SULT1A1 is important for efficient virus gene expression following infection of MDMs with single cycle HIV-1 and SIV vectors and with replication-competent HIV-1.

2.3.3. SULT1A1 regulates HIV-1 reverse transcription

To determine the step in HIV-1 replication that is influenced by SULT1A1, we monitored the effect of SULT1A1 siRNA #2 on early steps of virus replication at 24 hours post-infection. The VSV-G-pseudotyped HIV-1

vector was used for these studies to avoid multiple cycles of virus infection confounding the interpretation of the results. Quantitative PCR methods were used to monitor viral DNA and RNA products, focusing specifically upon spliced viral RNAs. These RNAs are produced *de novo* in infected cells and thus, are clearly distinguished from the large amounts of input unspliced viral genomic RNA that are present in these cells due to efficient virus uptake. Immunoblotting was used to monitor the expression levels of two independent viral proteins (Vpu and Vif), that are produced from spliced HIV-1 mRNA transcripts.

These studies revealed that SULT1A1 had no effect upon the levels of early reverse transcription products (defined as those generated prior to minus-strand DNA transfer) (Figure 2.5A, left panel). By contrast, knockdown of SULT1A1 was associated with a reduction (58%, $p<0.0005$) in the levels of late reverse transcription products (defined here as those generated subsequent to minus-strand DNA transfer) (Figure 2.5A, right panel). Consistently, SULT1A1 knockdown in MDMs was associated with a 72% reduction in levels of HIV-1 multiply spliced RNA (forward and reverse primers span the first and second exons of Tat/Rev, respectively ($p<0.0005$) (Figure 2.5B). Furthermore, SULT1A1 knockdown was also correlated with a 70% reduction in the levels of both the HIV-1 Vif and Vpu proteins ($p<0.0005$) (Figure 2.5C-D). Taken together, these results indicate that the predominant effect of SULT1A1 on HIV-1 replication is at the level of viral DNA synthesis.

2.3.4. SULT1A1 influences the kinetics of HIV-1 minus-strand DNA elongation

In order to determine the specific stage of HIV-1 reverse transcription that is influenced by SULT1A1, MDMs were transfected with SULT1A1-specific or control siRNA prior to infection with VSV-G pseudotyped NL43-Luc HIV-1 vector and DNA was isolated at different time points post-infection (ranging from 8-24 hours). The relative abundance of different length reverse transcription products was then measured using a series of oligonucleotide primers that were used to amplify different regions of HIV-1 DNA (labeled in order of appearance as follows: ERT, U3-R, 2 KB, 4KB, 8 KB, LRT (Figure 2.6A).

Consistent with our previous result (Figure 2.5), the abundance of early reverse transcription products (ERT and U3-R) was largely the same in both SULT1A1 siRNA transfected and control transfected cell populations (Figure 2.6B all panels). By contrast, there was a marked reduction of DNA products longer than 2 kb in length at 8 hours post-infection in cells deficient in SULT1A1 (Figure 2.6B, left panel). The defect in 2 kb products was less pronounced at 16 hours post-infection and was not evident at 24 hours post-infection (Figure 2.6B). Similarly the defect in the level of 4 kb products was less pronounced at 24 hours as compared with 16 hours post-infection. The defect seen in the levels of longer 8 kb and LRT products was not overcome in the 24-hour time frame. Taken together, these results show that SULT1A1

knockdown is associated with a decreased progression of reverse transcription, a result that is consistent with this enzyme playing an important role during the kinetics of viral minus strand DNA elongation.

2.4. Discussion

This study demonstrates that the cellular sulfotransferase SULT1A1 regulates HIV-1 reverse transcription in primary human monocyte-derived macrophages (MDMs). We showed that SULT1A1 is highly expressed in MDMs and that RNAi-mediated depletion of SULT1A1 in these cells results in a significant reduction in the kinetics of minus-strand DNA elongation during HIV-1 reverse transcription.

The synthesis of HIV-1 proviral DNA from the viral RNA genome is heavily regulated by cellular factors; however, only a small number of these factors have been well characterized to date (Arhel and Kirchhoff, 2010; Friedrich et al., 2011; Goff, 2007; Warren et al., 2009). These proteins include SAMHD1, which blocks reverse transcription in myeloid cells and resting CD4+ T cells by hydrolyzing dNTPs and/or degrading viral RNA (Baldauf et al., 2012; Hrecka et al., 2011; Laguette et al., 2011; Ryoo et al., 2014), and APOBEC3G, which deaminates cytidine to uridine resulting in hypermutation during DNA synthesis (Sheehy et al., 2003; Yu et al., 2003). In addition, other factors have been shown to affect the kinetics of viral DNA synthesis (Konig et al., 2008). Thus, the studies described in this report extend our understanding of the roles of cellular factors in regulating this key step of retroviral replication. This novel role for SULT1A1 in the regulation of retroviral reverse transcription also adds to our growing knowledge about the diverse roles that the

sulfonation pathway plays in the retroviral life cycle. For example, tyrosylprotein sulfotransferases 1 and 2 (TPST1 and TPST2) are Golgi sulfotransferases that sulfonate N-terminal tyrosine residues of CCR5 and enable efficient cellular entry of R5-tropic viruses (Choe et al., 2003; Farzan et al., 1999; Seibert et al., 2002). It has also been demonstrated that HIV-1 binds to sulfonated proteoglycans on cell surfaces (Mondor et al., 1998; Roderiquez et al., 1995). In addition, we previously showed that PAPSS deficiency or treatment of cells with the sulfonation pathway inhibitors chlorate and guaiacol blocked HIV-1 *de novo* gene expression from the viral LTR promoter, and more recently we showed that these inhibitors also block HIV-1 reactivation from latency (Bruce et al., 2008; Murry et al., 2014). Taken together, these observations demonstrate the importance of the sulfonation pathway at multiple steps of HIV-1 replication. It will be important for future studies to determine which sulfotranserase(s) regulate HIV-1 infection and reactivation from latency in CD4+ T cells, as SULT1A1 does not appear to be expressed at the protein level in these cells and control has been demonstrated at level of transcription, not reverse transcription, upon treatment with chlorate and guaiacol.

SULT1A1 is highly polymorphic within the human population, with both genetic polymorphisms and copy number variation conferring different levels of enzymatic activity (Hebbring et al., 2007; Ning et al., 2005; Yu et al., 2010). Moreover, *SULT1A1* variation has been linked to a number of diseases such

as cancer (Nowell et al., 2004; Tang et al., 2003; Wang et al., 2002; Zheng et al., 2003), heart disease (O'Halloran et al., 2009), and inflammatory bowel disease (Imielinski et al., 2009). Consequently, we are now seeking to determine if there is a correlation between *SULT1A1* variability and HIV-1 susceptibility and/or AIDS disease progression.

Further investigation will be aimed at determining if SULT1A1 acts on HIV-1 through a substrate-dependent or -independent mechanism. It is possible that SULT1A1 may act independently of substrate by directly modifying viral proteins (such as reverse transcriptase). If the sulfonation of a specific SULT1A1 substrate is required, on the other hand, then identification of that substrate will be critical for understanding the underlying mechanism involved.

2.5. Conclusions

In summary, we demonstrated that a human cytosolic sulfotransferase, SULT1A1, regulates HIV-1 reverse transcription in human monocyte-derived macrophages (MDMs). We showed that SULT1A1 is highly expressed in primary human monocytes and MDMs. RNAi-knockdown of SULT1A1 in MDMs leads to a substantial decrease in infection by both pseudotyped and replication-competent HIV-1 vectors, as well as by a SIVagm vector. Quantitative PCR analysis revealed that this effect is associated with a defect in minus-strand DNA elongation during HIV-1 reverse transcription. These results support the idea that SULT1A1 is a novel HIV-1 host factor in MDMs, and suggest that targeting SULT1A1 or its substrate may lead to improved HIV-1 therapies.

2.6. Acknowledgements

We thank members of the Young lab for helpful discussion, especially John Naughton. This project was supported by a Salk-Sanofi innovation grant, the NIH Cellular and Molecular Training Grant, the Bruce J. Heim Foundation, the Jesse and Caryl Philips Foundation, the James B. Pendleton Charitable Trust, the NCI Cancer Center Support Grant Award, and for funding and support. We would also like to thank Kevin Olivieri and Manqing Li for helpful discusion and reagents, and the Salk Gene Transfer, Targeting, and Therapeutics Core (GT3) for lentivirus production.

Chapter 2, in full, has been published in Virology Journal, 2016. Justine Swann, Jeff Murry, John A.T. Young, “Cytosolic sulfotranserase 1A1 regulates HIV-1 reverse transcription in human monocyte-derived macrophages”. The dissertation author was the primary investigator and author of this paper.

2.7. Materials and Methods

2.7.1. Reagents

AllStars Negative control and SULT1A1 Flexitube siRNAs were obtained from Qiagen, reconstituted at 20μM in water, and stored at -20°C until use. Cell viability was assayed using Cell Titer-Glo reagent and luciferase activity was measured using Bright-Glo reagent according to the manufacturer's instructions (Promega).

2.1.2. SULT mRNA expression analysis

The expression level for each cytosolic sulfotransferase in CD4+ T cells and CD14+ monocytes was derived from publically available expression data from BioGPS (Wu et al., 2009a), and normalized to the median expression of that sulfotransferase in all tissues tested.

2.7.3. Peripheral blood mononuclear cells

Human donor buffy coats and LRS-WBC (white blood cells isolated in the Leuko-Reduction system via Terumo BCT Trima Automated Collection System) samples were collected from anonymous healthy donors and obtained from the San Diego Blood Bank. Written informed consent for the use of buffy coats and LRS samples for research purposes was obtained from the donors by the San Diego Blood Bank. Samples are routinely tested for the presence of HIV-1 antibodies and all samples used in the study are confirmed

HIV-1 negative. Peripheral blood mononuclear cells (PBMCs) were isolated from these samples by Ficoll density gradient centrifugation. Briefly, buffy coats and LRS samples were diluted in RPMI, layered over Ficoll (Ficoll-Paque Plus, GE Healthcare), and centrifuged at 1850 rpm for 25 minutes without brakes. Cells at the interface were removed and washed twice in RPMI. Platelets were then removed by a final spin at 800 rpm without brakes. The cells were counted and resuspended at a density of $25x10^6$ cells/ml in FBS supplemented with 10% DMSO, transferred to -80°C at least overnight, and then to liquid nitrogen for long-term storage.

2.7.4. CD4+ T cells

PBMCs were thawed in RPMI, washed, and CD4+ T cells were isolated using CD4+ negative selection magnetic bead isolation (EasySep Human CD4+ T cell enrichment kit, StemCell Technologies). Resting CD4+ T cells were lysed directly after separation, and the remaining CD4+ T cells were activated using CD3/CD28 beads for three days (Dynabeads Human T-Activator CD3/CD28, Life Technologies). CD4+ T cells were resuspended at $2x10^6$ cells/100µl 1X LDS lysis buffer (Invitrogen), diluted 1:3, and 10μl lysate was loaded onto a gel 4-12% Bis-Tris gel for immunoblot analysis.

2.7.5. Monocyte-derived macrophages

PBMCs were thawed in RPMI, washed, and CD14+ monocytes were isolated using CD14 positive selection magnetic cell sorting according to the

manufacturer's instructions (Easy Sep Human CD14 Positive Selection kit, StemCell Technologies). CD14+ monocytes were plated at $5x10^6$ cells per 10cm polypropylene petri dish (Fisher) in RPMI 1640 medium (Gibco-BRL) supplemented with 10% Human AB Serum (Sigma), 50 U/ml penicillin, 50 µg/ml streptomycin, 10 mM HEPES, sodium pyruvate, and 20 ng/ml recombinant human M-CSF (Peprotech). Fresh media without cytokine was added after 3-4 days, and by day 7 the MDMs were fully differentiated.

2.7.6. siRNA transfection

MDMs were washed with PBS, treated with Accutase (Stem Pro Accutase Cell Dissociation Reagent, Life Technologies) at 37°C for 30 minutes, and dissociated by gentle pipetting. Cells were transfected with siRNA using the NEON electroporation system (Life Technologies) according to the manufacturer's instructions. Briefly, $1.5x10^5$ MDMs were mixed with 3µl 20µM siRNA in a 10µl final reaction volume, electroporated with 2 pulses of 1500V for 20ms, and then transferred to 500µl pre-warmed RPMI 10% human serum without antibiotics in a 48 well plate and allowed to recover for 4 days. At 4 days post transfection, samples were assayed for protein knockdown by immunoblot. Samples with <60% SULT1A1 knockdown and/or <65% cell viability were not used in the analysis.

2.7.7. Viruses and infection

The HIV-1 NL43-Luc and SIVagm plasmids were obtained from Ned Landau (Connor et al., 1995; Mariani et al., 2003). HIV-1 Env+ JM1186-Rluc plasmid was made by Sumit Chanda's lab by cloning the V3 loop of Gp120 from isolate 92TH014 (Bjorndal et al., 1997) and Renilla luciferase reporter into pBR-NL43-IRES-eGFP (Munch et al., 2007). VSV-G pseudotyped NL43-Luc HIV-1 and Env+ JM1186-Rluc HIV-1 viruses were produced by transient transfection of 293T cells by the Salk viral vector core, and VSV-G-pseudotyped SIVagm-Luc was produced in a similar method as previously described (Konig et al., 2007). HIV-1 and SIVagm viruses were assayed for p24 or p27 content, respectively, by p24 or p27 antigen capture ELISA according to the manufacturer's instructions (Zeptometrix).

For MDM infection experiments, media was removed 96 hours post transfection with siRNA, replaced with 100µl virus (corresponding to 164ng p24 VSV-G pseudotyped NL43-Luc HIV-1, 66.3ng p24 Env+-JM1186-Rluc HIV-1, 234ng p27 VSV-G pseudotyped SIVagm-Luc) diluted in 400µl fresh RPMI 10% human serum, and cells were spinoculated at 1200 x g for 60 minutes. Viral reporter expression and cell viability was assayed 24 hours post infection for VSV-G-pseudotyped viruses, and 3 days post infection for replication competent HIV-1.

2.7.8. Immunoblot and protein quantitation

Cells were lysed in 1X LDS sample buffer (Invitrogen) and stored at -20°C until use. Samples were thawed on ice, incubated at 95°C for 5 minutes, and centrifuged for 3-5 minutes before loading. Precision Plus Protein Dual Color Standards (Bio-Rad) or Novex Sharp Pre-stained standards (Invitrogen) were used as protein standards, and 4-12% Bis-Tris gels (Invitrogen) were used for gel electrophoresis. Protein was transferred onto PVDF membrane (Millipore Immobilon-FL) using BioRad wet transfer apparatus at 100V for 1 hour at 4°C. Membranes were blocked with Casein blocking buffer (Bio-Rad) for at least 1 hour at room temperature, then incubated with primary antibody diluted in 0.1% Casein/0.2X PBS + 0.1% Tween-20 overnight at 4°C. Primary antibodies used in the study include mouse anti-human SULT1A1 mAb (clone 638708, R&D Systems) used at 1:2000 dilution, rabbit anti-HIV-1 Vif (clone A319, NIH AIDS Reagent program) used at 1:500, rabbit anti-HIV-1 Vpu (clone ab81532, Abcam) used at 1:1000 dilution, mouse anti-human Ku-86 mAb (clone B-1, Santa Cruz Biotechnology) used at 1:500 dilution, and anti-human mouse GAPDH (clone 6C5, Abcam) used at 1:5000. Membranes were washed three times with 1X PBS/ 0.1% Tween-20, incubated with Alexa-Fluor 680-conjugated secondary antibody diluted in 0.1% Casein 0.2X PBS 0.1% Tween-20 and 0.01% SDS at room temperature for at least 1 hour. Secondary antibodies used in the study include goat anti-mouse Alexa-Fluor 680 IgG (H+L) (clone A21057, Invitrogen), and donkey anti-rabbit Alexa-Fluor 680 IgG

(H+L) (clone A10043, Invitrogen). Membranes were washed three times with 1X PBS/ 0.1% Tween 20, and imaged using the Licor Odyssey system. Protein expression was quantified using Image Studio software (Licor).

2.7.9. Detection of nucleic acids by real-time quantitative PCR

Total RNA was prepared using Qiazol lysis reagent (Qiagen) and miRNeasy RNA isolation kit (Qiagen) according to the manufacturer's instructions. cDNA was synthesized using Quantitect Reverse Transcription kit (Qiagen) according to the manufacturer's instructions. DNA was isolated from cells using a DNeasy Blood and Tissue Kit (Qiagen) according to the manufacturer's instructions. Quantitation of ERT, LRT, and PBGD was performed using primers and probes diluted in TaqMan universal PCR master mix (Invitrogen). Amplification of other DNA products was monitored using SYBR green fluorescence (Invitrogen). Real-time PCR was performed using the ViiA 7 Real-time PCR system (Life Technologies). Standard curves were generated for all primers and efficiencies were found to be equivalent. Relative expression was calculated using the delta CT method as previously described (Schmittgen and Livak, 2008). All primer and probe sequences used in the study are included in Supplemental Dataset 1.

2.8. Figures

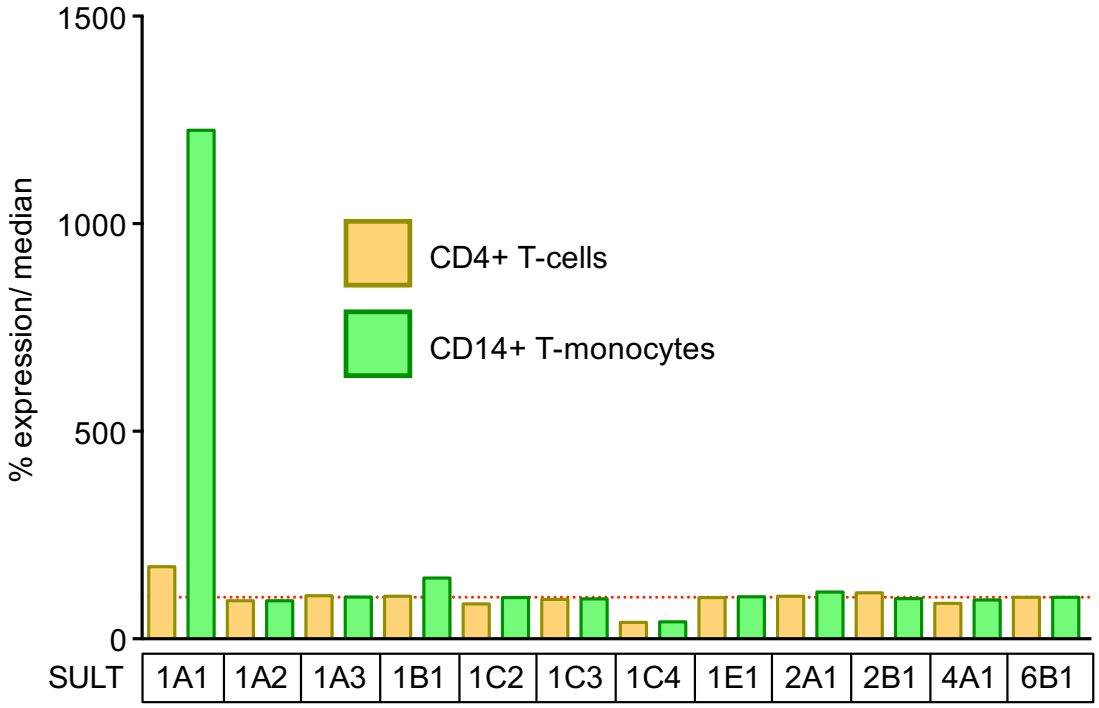


Figure 2.1. SULT1A1 is highly expressed in monocytes. The expression level for each cytosolic sulfotransferase in CD4+ T cells and CD14+ monocytes was derived from publically available expression data from BioGPS and normalized to the median expression of that sulfotransferase in all tissues tested.

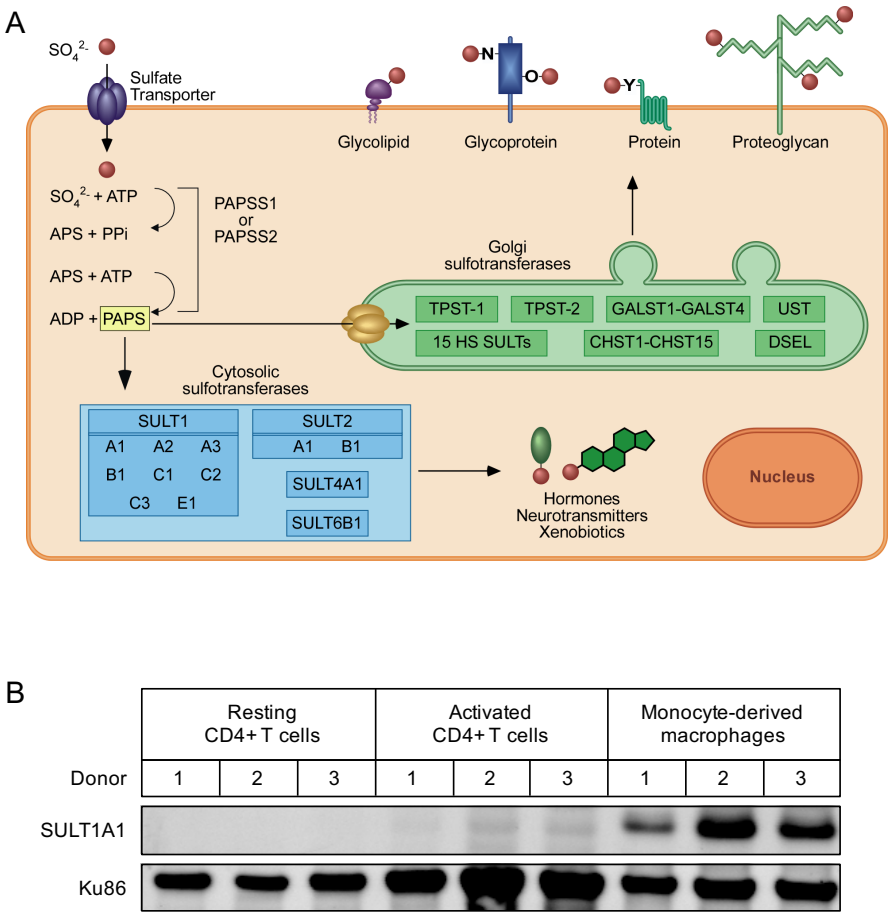


Figure 2.2. SULT1A1 is highly expressed in primary human monocyte-derived macrophages (MDMs). a) The cellular sulfonation pathway. The first step of the cellular sulfonation pathway involves import through a sulfate transporter of a sulfate ion that is then used as a substrate by either 3'-phosphoadenosine-5'phosphosulfate (PAPS) synthetase enzymes PAPSS1 or PAPSS2. These proteins catalyze two enzymatic steps to generate PAPS, the high-energy universal sulfonate-donor from sulfate and two molcules of ATP. PAPS can be transported across the Golgi membrane and used by the Golgi sulfotransferases to generate sulfonated proteins, glycoproteins, glycoproteins, glycolipids, and proteoglycans. Alternatively, PAPS can be used by cytosolic sulfotransferases (SULTS) to sulfonate small molecules such as hormones, neurotransmitters, and xenobiotics. b) Human CD4+ T cells and CD14+ monocytes were isolated from donor PBMCs by magnetic bead isolation. Resting CD4+ T cells were lysed directly after separation, and the remaining CD4+ T cells were activated using CD3/CD28 beads for three days. Monocytes were cultured for 7 days in the presence of 20ng/ml M-CSF, were lysed, subjected to gel electrophoresis, and immunoblotting was performed to detect SULT1A1 or the loading control Ku86 protein.

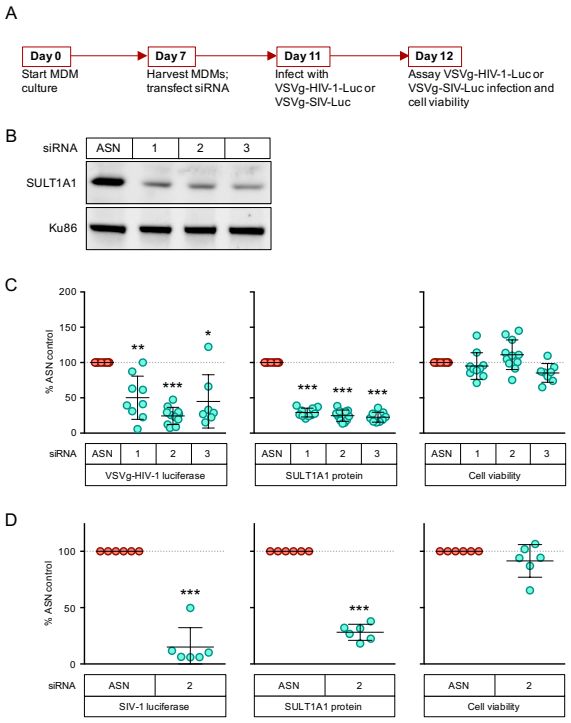


Figure 2.3. SULT1A1 knockdown is associated with decreased viral gene expression following infection of MDMs with VSV-G pseudotyped HIV-1 and SIV vectors. a) Schematic showing experimental timeline. Briefly, CD14+ monocytes were isolated from human donor PBMCs using positive selection with magnetic beads. Monocytes were differentiated into MDMs and on day 7 were electroporated with siRNA and plated at $1.5x10^4$ MDMs per well in a 48 well plate. After 96 hours, protein knockdown was confirmed by immunoblot and cells were infected with 100µl (corresponding to 164ng p24 HIV-1 and 234ng p27 SIV viral vectors) of the indicated virus. Luciferase and cell viability measurements were determined at 24 hours post-infection. b) Representative immunoblot showing SULT1A1 knockdown 96 hours post transfection with SULT1A1 siRNAs (1-3) and AllStars Negative siRNA Control (ASN), Qiagen. SULT1A1 expression is compared to endogenous Ku86 used as a loading control. Results from one representative donor are shown. SULT1A1 expression was generally decreased by 70-80% with siRNA treatment compared to the control ASN siRNA (as shown in Figure 2C, middle panel). c) HIV-1 luciferase reporter expression (left panel), SULT1A1 protein expression (middle panel), and cell viability of MDMs (right panel) were measured 24 hours after infection with the VSV-G pseudotyped NL43-Luc HIV-1 vector. Results shown are from 6 donors assayed twice. All values were compared to ASN control. d) SIV-1 luciferase reporter expression (left panel), SULT1A1 protein expression (middle panel), and cell viability (right panel) for MDMs 24 hours post infection with VSV-G-pseudotyped SIVagm-Luc. Mean and SD shown, *** $p < 0.0005$ ** $p < 0.005$ * $p < 0.05$ one sample t test. Results shown are from 6 donors assayed twice. Samples with <60% SULT1A1 knockdown and/or <65% cell viability were not used in the analysis. All values were compared to ASN control.

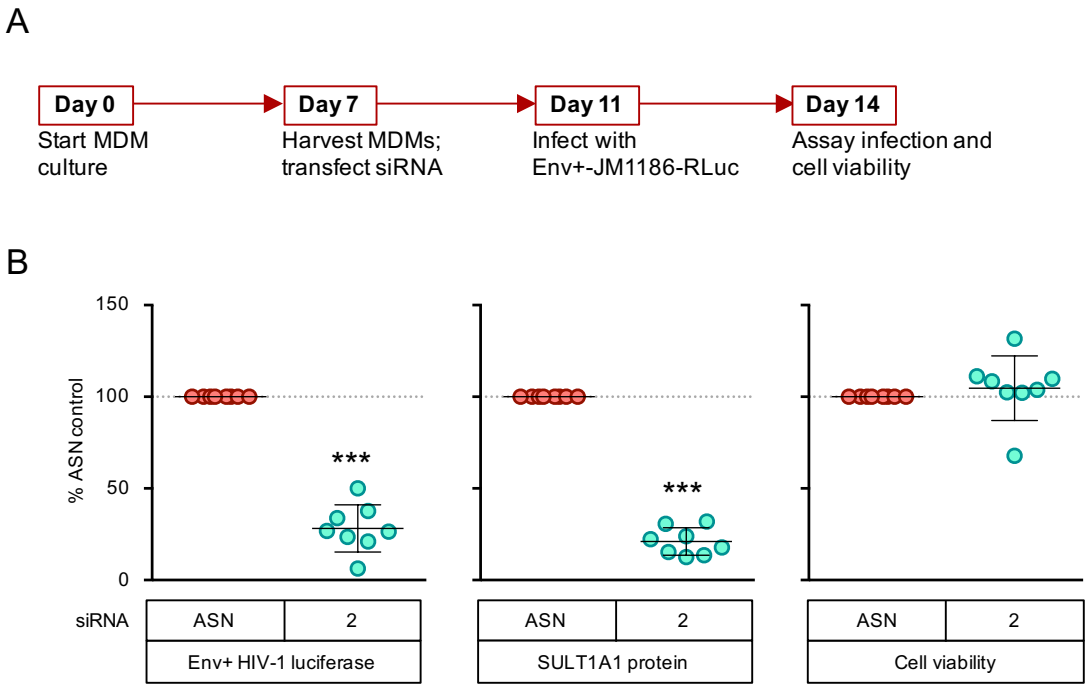


Figure 2.4. SULT1A1 knockdown is associated with decreased viral gene expression following infection of MDMs with a replication-competent HIV-1 vector. a) The same method was used as in Figure 2A, however 100µl (corresponding to 66.3ng p24) HIV-1 Env+ JM1186-Rluc virus was added to the cells at day 11 and luciferase and cell viability was assayed at 72 hours post-infection. b) HIV-1 luciferase reporter expression, SULT1A1 protein expression, and cell viability of MDMs 24 hours post infection with HIV-1 Env+ JM1186-RLuc. Mean and SD shown, *** $p < 0.0005$ ** $p < 0.005$ * $p < 0.05$ one sample t test. Results shown are from 6 donors assayed twice. Samples with <60% SULT1A1 knockdown and/or <65% cell viability were not used in the analysis. All values were compared to ASN control.

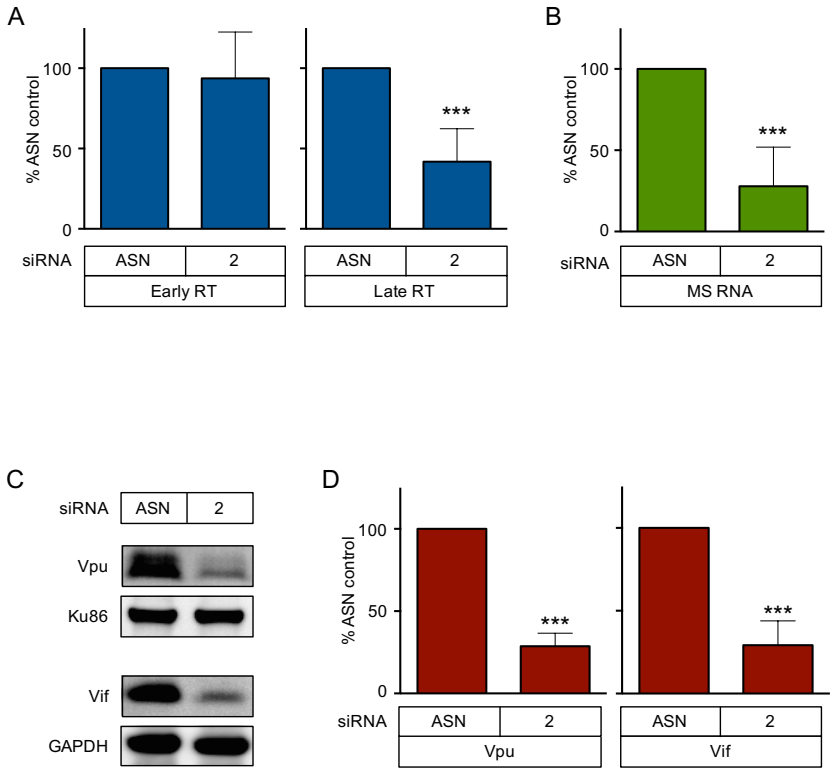


Figure 2.5. SULT1A1 regulates HIV-1 reverse transcription. a) MDMs were treated with control siRNA or SULT1A1-specific siRNA and subesequently challenged with VSV-G pseudotyped NL43-Luc HIV-1 vector. DNA was isolated 24 hours post-infection and qPCR was performed using primers that detect early RT DNA products or late RT DNA products compared to the cellular PBGD gene as an endogenous control. The levels of early and late RT products were normalized to the ASN siRNA control. Results shown are from MDMs derived from 6 donors tested twice. b) MDMs were transfected with SULT1A1 siRNA 2 or ASN siRNA and total cellular RNA was isolated 24 hours post infection with VSV-G pseudotyped NL43-Luc HIV-1 vector. qPCR using primers specific for HIV-1 multiply spliced mRNA (MS RNA, forward and reverse primers span the first and second exons of Tat/Rev, respectively) was performed, and relative MS RNA (normalized to GAPDH) was then normalized to ASN siRNA control. Results shown are from MDMs derived from 6 donors tested twice. c) Representative immunoblot showing HIV-1 Vpu and Vif protein levels compared to endogenous Ku86 or GAPDH loading control, respectively, from protein lysate collected 48 hours post infection with VSV-G pseudotyped NL43-Luc HIV-1 vector pre-treated with either ASN siRNA control or SULT1A1 siRNA 2. d) Quantitative immunoblot analysis using Image Studio software of Vpu and Vif as shown in Figure 4C.
Mean and SD shown, *** $p < 0.0005$ ** $p < 0.005$ * $p < 0.05$ one sample t test. Results shown are from 6 donors assayed twice. Samples with <60% SULT1A1 knockdown and/or <65% cell viability were not used in the analysis. All values were compared to ASN control.

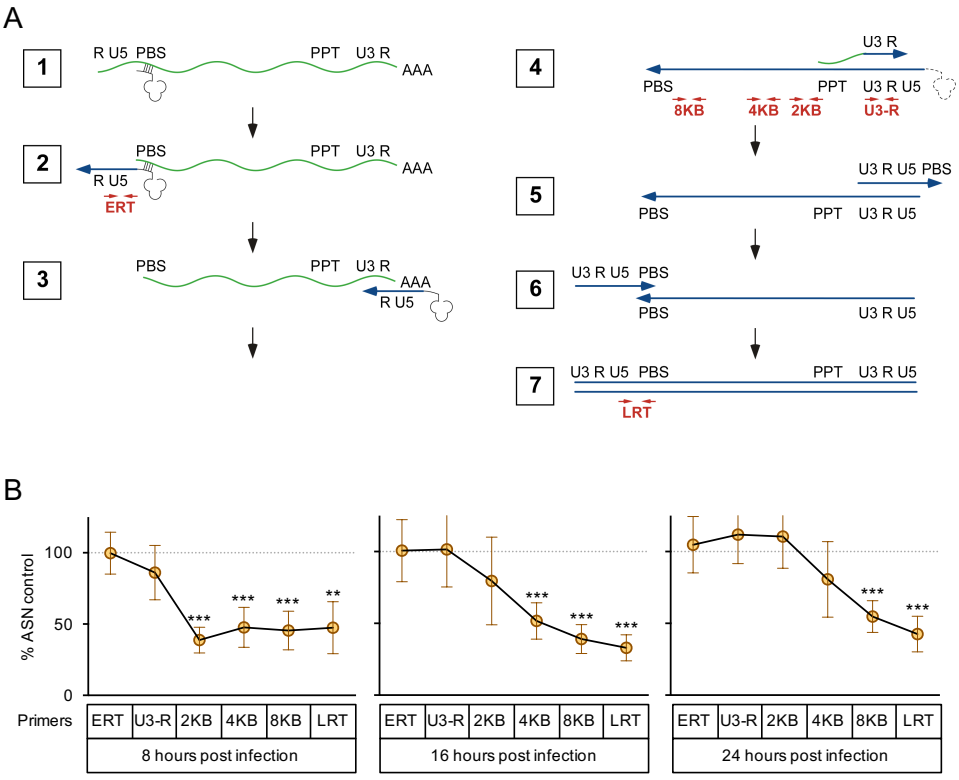


Figure 2.6. SULT1A1 influences the kinetics of minus-strand DNA elongation. a) The step during HIV-1 reverse transcription that SULT1A1 regulates was investigated by quantitative real-time PCR analysis with the indicated primer sets. The stages of reverse transcription are shown: 1. Viral genomic RNA; 2. Minus-strand DNA initiation; 3. Minus-strand DNA transfer; 4. Minus-strand DNA elongation and plus-strand DNA initiation; 5 and 6. Plus-strand DNA transfer; 7. Plus-strand DNA elongation. b) MDMs were treated with control or SULT1A1-specific siRNA and challenged with VSV-G pseudotyped NL43-Luc HIV-1 vector. Total cellular DNA was isolated at 8, 16, or 24 hours post infection. Quantitative real-time PCR was performed to measure the relative abundance of DNA products corresponding to specific steps during the process of reverse transcription using primers indicated in Figure 6A. Results shown are from MDMs derived from 6 donors assayed once. Mean and SD shown, *** $p < 0.0005$ ** $p < 0.005$, one sample t test.

3. A high-throughput luciferase-based assay for the discovery of therapeutics that prevent malaria

3.1. Abstract

In order to identify the most attractive starting points for drugs that can be used to prevent malaria, a diverse chemical space comprising tens of thousands to millions of small molecules may need to be examined. Achieving this throughput necessitates the development of efficient ultra-high-throughput screening methods. Here, we report the development and evaluation of a luciferase-based phenotypic screen of malaria exoerythrocytic-stage parasites optimized for a 1536-well format. This assay uses the exoerythrocytic-stage of the rodent malaria parasite, *Plasmodium berghei,* and a human hepatoma cell line. We use this assay to evaluate several biased and unbiased compound libraries, including two small sets of molecules (400 and 89 compounds, respectively) with known activity against malaria erythrocytic-stage parasites and a set of 9,886 Diversity-Oriented Synthesis (DOS)-derived compounds. Of the compounds screened we obtain hit rates of 12–13% and 0.6% in preselected and naïve libraries, respectively, and identify 52 compounds with exoerythrocytic-stage activity less than 1 μM, and having minimal host cell

toxicity. Our data demonstrate the ability of this method to identify compounds known to have causal prophylactic activity in both human and animal models of malaria, as well as novel compounds, including some exclusively active against parasite exoerythrocytic stages.

3.2. Introduction

Despite being an ancient disease, malaria is still responsible for over a half million deaths and substantial morbidity, poverty, and suffering for hundreds of millions of people each year (Carter and Mendis, 2002; Sachs and Malaney, 2002; Snow et al., 2005). It is a vector-borne disease caused by infection with *Plasmodium* parasites transmitted through the bite of *Anopheles* mosquitos. While eradication campaigns have been successful in most of North America and Europe, malaria continues to devastate developing regions of Asia, Africa, and South America (Sinka et al., 2012). The mortality rates are highest amongst African children, with an estimated one death per minute (WHO). The emergence of resistance to all of the current frontline antimalarial drugs warrants global concern (Flannery et al., 2013). It is therefore critical that new drugs are developed that not only treat disease symptoms, but also contribute towards the elimination and eradication of malaria infections. In order to achieve eradication, new drugs should inhibit multiple developmental stages of the parasite. Following the blood meal of an infected *Anopheles* mosquito, *Plasmodium* sporozoites travel through the bloodstream to reach the liver. The sporozoites traverse multiple cells within the liver before establishing productive invasion within hepatocytes, where they transform into exoerythrocytic-stage exoerythrocytic forms (EEFs) (Prudencio et al., 2006a). Depending on the species, these exoerythrocytic forms enter one of two

developmental pathways: they can form mature exoerythrocytic-stage schizonts, or they can enter a dormant phase called hypnozoites. The determinant factors guiding exoerythrocytic-stage development towards hypnozoite formation in *P. vivax* and *P. ovale* are not understood. Hypnozoites can reinitiate development and give rise to malaria relapses weeks, months or even years after the initial infection (Wells et al., 2010a). Fully developed exoerythrocytic-stage merozoites within schizonts eventually egress from the liver and re-enter the bloodstream (Prudencio et al., 2006a). The continuous replication of asexual bloodstages within red blood cells (RBCs) causes RBC destruction and leads to the characteristic symptoms associated with malaria: anemia, fever and chills (Clark et al., 1997a). A small percentage of these asexual-blood stage parasites will then differentiate into sexual erythrocytic-stage parasites as female and male gametocytes, and the transmission of the sexual blood stage back to the mosquito vector during a subsequent blood meal completes the life cycle (Baker, 2010b).

The majority of the current antimalarials only treat the symptom-causing erythrocytic stages of the parasite (Delves et al., 2012). A few classes, including cytochrome bc1 inhibitors (such as atovaquone) and antifolate drugs (such as pyrimethamine), are active against developing exoerythrocytic forms as well as erythrocytic forms and are able to prevent the establishment of infection (causal prophylactic activity) as well as relieve symptoms (Delves et al., 2012). Antibiotics, such as doxyclycline, clindamycin,

and azithromycin, are also an important class of antimalarial drugs. Doxyclycline is commonly prescribed to travelers to endemic areas and is especially useful for its casual prophylaxis and slow acting blood schizontocidal activity (Tan et al., 2011). The 8-aminoquinolines (such as primaquine and tafenoquine) are a unique class in that they can eliminate hypnozoites as well and can provide a radical cure for *P. vivax* and *P. ovale* (Delves et al., 2012; Wells et al., 2010a). Having new classes of drugs that could be used prophylactically and/or to provide a radical cure would be desirable. Resistance is developing to both napthoquinones and antifolates, and 8-aminoquinolines can be toxic to individuals with glucose-6-phosphate deficiency (Delves et al., 2012; Hyde, 2002). Drugs targeting the exoerythrocytic-stage only would also offer the reduced potential for drug resistance, as there are far fewer parasites at this 'bottleneck' compared to the replicative erythrocytic stages (Delves et al., 2012). Accordingly, the development of an exoerythrocytic-stage specific high-throughput screening assay is necessary in order to identify the next generation of antimalarial drugs.

Although it is possible to create new chemical derivatives of existing drugs with improved therapeutic and resistance profiles, phenotypic screening offers the opportunity to find entirely new classes of small molecules that are active against exoerythrocytic stages of the lifecycle (Ganesan et al., 2012). We have previously reported an immunofluorescence-based malaria

exoerythrocytic-stage assay that we used to screen a library of >4,000 commercially available compounds with erythrocytic-stage activity (Meister et al., 2011). While this platform led to the identification of 275 exoerythrocytic-stage active compounds, the assay is limited to a 384-well throughput and is therefore not suitable for the screening of larger chemical libraries, in part because of the high cost of sporozoites obtained by manual mosquito dissections ($1.00 per well screened). In addition, the requirement for a specialized high-content imaging device limits the accessibility of the assay. Other malaria exoerythrocytic-stage drug screens have utilized *P. berghei* sporozoites that express a luciferase reporter (Pb-Luc) (Derbyshire et al., 2011; Derbyshire et al., 2012; Reader et al., 2015), however, these assays are also limited by a 384-well assay throughput. In this report, we describe the development of a high-throughput luciferase-based assay that can be used to screen chemical libraries in a 1536-well plate format. We demonstrate that the assay is highly sensitive, reproducible, and efficient. As proof of concept, we use this assay to screen the Medicines for Malaria Venture (MMV) Malaria Box for compounds with exoerythrocytic-stage activity (Spangenberg et al., 2013) as well as a larger collection of chemical compounds from the Broad Diversity-Oriented Synthesis Library, a set that includes compounds with and without demonstrated erythrocytic-stage antimalarial activity.

3.3. Results and Discussion

3.3.1. Development of a luciferase-based high-throughput exoerythrocytic-stage assay

In order to develop a high-throughput exoerythrocytic-stage malaria assay capable of screening large libraries of chemical compounds, a number of tests were performed to optimize a 48 hour *in vitro* PbGFP-Luc-SM_{CON} (Janse et al., 2006) infection of HepG2-A16-$CD81^{EGFP}$ hepatocytes (Silvie et al., 2006) (Figure 3.1). This rodent *Plasmodium* strain was previously generated through the integration of a GFP-Luc cassette into the *c-rrna* locus and selecting transgenic *P. berghei* by flow sorting GFP-expressing parasites. For simplicity, we will refer to this strain as Pb-Luc. For these tests, HepG2-A16-$CD81^{EGFP}$ cells were seeded in 1536-well plates 24 hours prior to infection and luciferase bioluminescence measured 48 hours post infection to detect parasite viability. We found the ideal ratio of sporozoites to cells per well to be 1:3, respectively (1×10^3 sporozoites in 5 μl to 3×10^3 cells 5 μl) (Figure 3.2A, Figure 3.1A). At these concentrations, the cells were ideally confluent and the infection rate produced luciferase values that were significantly greater than background values at 48 hours post infection (Figure 3.1A). Furthermore, tests without hepatocytes showed that there was no residual luciferase activity from Pb-Luc sporozoites at 24 hours post infection at 37°C (Figure 3.1B), eliminating the possibility that sporozoites which had

not invaded contribute to the luciferase signal. We also tested different DMSO concentrations (added 18 hours pre-infection) to assess their impact on parasite viability, and found that concentrations up to 0.88% DMSO had an insignificant effect on luciferase activity 48 hours post infection (Figure 3.1C). The final protocol was to add 50 nl of compound in DMSO (resulting in 50 μM compound and 0.5% DMSO concentration in the assay plates) 18 hours pre-infection in the optimized screening assay (Figure 3.2A). An example of the luciferase signal for two replicate plates seeded with a representative small molecule library is shown in Figure 3.2B. Z factor for these plates was between 0.7 and 0.9, an excellent value for a phenotypic screen.

3.3.2. Assay validation through screening of known antimalarial compounds

To validate this assay further, we next examined several compound collections. We first evaluated 50 established antimalarial clinical or tool compounds (Table 3.1) that have been tested in other antimalarial phenotypic assays (Delves et al., 2012). The most active compounds included the electron transport chain inhibitor atovaquone, antifolate pathway inhibitors P218•HCl, pyrimethamine and cycloguanil, the protein biosynthesis inhibitor cycloheximide, and the glutathione reductase inhibitor methylene blue. For these compounds, the exoerythrocytic-stage half maximal inhibitory concentration (IC_{50}) values were similar to the *P. falciparum* erythrocytic-stage

IC_{50} values. Compounds that were less active relative to erythrocytic-stages, on the other hand, included 4-aminoquinolines, amino alcohols, and endoperoxides, which presumably act primarily against hemoglobin degradation, a process that does not occur during hepatic stages. Overall, these results were highly consistent with our previously established high-content imaging (HCI) assay of the hepatic stages against the related rodent parasite, *Plasmodium yoelii* (Delves et al., 2012).

For active compounds, we also compared the IC_{50} values for Pb-Luc generated by the luciferase-based enzymatic assay to values produced by screening Pb-Luc parasites using our previously established (HCI) assay that uses polyclonal *Plasmodium* HSP70 antibody staining at 48 hours as an indicator of infection (Meister et al., 2011) (Figure 3.3). We found that there was a strong correlation between compound activities in both assays, as indicated by an R^2 of 0.83, but the high-throughput luciferase-based assay resulted in IC_{50}s roughly 10× lower (Figure 3.3). Since luciferase is relatively unstable, with a half-life of less than 2 hours (Ignowski and Schaffer, 2004), this may lead to a higher rate of reporter turnover and therefore increased sensitivity to compounds that inhibit parasite growth at later stages compared to that of parasite HSP70 (as measured in the HCI assay). Additionally, non-viable parasites may be stained using the HCI assay, whereas they may not produce luciferase.

3.3.3. Screening the MMV Malaria Box for exoerythrocytic-stage inhibitors

After establishing the high-throughput luciferase-based assay quality, reproducibility, and sensitivity, we sought to test its validity as a platform to screen diverse chemical libraries by screening the MMV Malaria Box (Spangenberg et al., 2013). This open-access set consists of 400 compounds that were selected from a group of ~20,000 antimalarial hits generated from a large-scale erythrocytic-stage screening of >4,000,000 compounds by St. Jude Children's Hospital, Novartis, and GSK (Figure 3.5). The library contains 200 'probe-like' and 200 'drug-like' compounds, selected based on their chemical diversity, erythrocytic-stage antimalarial activity, and commercial availability. The Malaria Box compounds have erythrocytic-stage activity ranging from an IC_{50} of 30 nM to 4 μM (Spangenberg et al., 2013). Compounds originating in the Novartis collection had been previously tested in the *P. yoelii* assay (Meister et al., 2011) but others had not, providing valuable internal controls.

After a first-round screening of the MMV Malaria Box at compound concentrations of 50 μM in duplicate plates, 48 compounds were selected for reconfirmation based on Pb-Luc inhibition of more than 90% and HepG2 cytotoxicity of less than 25% in both assay plates, a hit rate of 12%. These were tested in a 12-point serial dilution dose response beginning at 10 μM. Of those retested, 36 had a Pb-Luc IC_{50} < 10 μM and counter-screening for HepG2 cytotoxicity and luciferase inhibition produced IC_{50} values greater than

10 μM (Figure 3.4, see Supplemental Dataset 2 for full screening results), leading to a confirmation rate of 75%. Furthermore, more than half of these compounds were very active against Pb-Luc exoerythrocytic-stages with IC_{50} values of less than 1 μM (Figure 3.4). Six of the compounds that had been previously tested and confirmed (Meister et al., 2011) were reconfirmed here, but there were 18 compounds that were considered exoerythrocytic-stage active hits in the HCI assay and not in the luciferase-based assay. This is likely due to the higher initial screening concentrations used in the luciferase-based assay (50 μM compared to 10 μM in the HCI assay) leading to increased HepG2 cytotoxicity, as all of these compounds inhibited Pb-Luc activity but were toxic at 50 μM. It should be noted that due to the sensitivity of the luciferase assay and to minimize cytotoxicity, we advise starting with lower initial screening concentrations for future high-throughput screening.

3.3.4. Screening the Broad Diversity-Oriented Synthesis library

To assess performance in an unbiased screening library, two sets of Diversity-Oriented Synthesis (DOS) derived compounds were screened (outlined in Figure 3.6). These diverse compounds combine the stereochemical and skeletal complexity of the entire ensemble of natural products and the efficiency of high-throughput synthesis (Dandapani and Marcaurelle, 2010; Nielsen and Schreiber). This is an attractive validation library because multiple stereoisomers of each structural type are included,

which permits a unique type of structure–activity relationship measurements. The first set included 9,886 compounds selected to represent the structural diversity of the scaffolds from the Broad Institute's 100,000 DOS compound library. This 'informer set' was tested as a naïve library with respect to activity against *Plasmodium*. The second set comprised 89 compounds previously shown to be active in a *P. falciparum* erythrocytic-stage assay ($IC_{50} < 2$ μM) (Nobutaka Kato et al.). Compounds that inhibited the luciferase signal >75% were scored as hits. From the informer set (hit rate = 0.6%), 60 hits and 4 inconclusive (only active in one replicate) were identified and 12 hits were identified from the erythrocytic-stage active set (hit rate = 13.4%). All available compounds (58) plus 25 additional weak actives (60–74% inhibition) from the informer set were tested in dose response in the primary exoerythrocytic-stage assay and also in a SYBR Green erythrocytic-stage assay (Plouffe et al., 2008). All of the compounds from the erythrocytic-stage active set retested at dose with an $IC_{50} < 5$ μM and 10 of these compounds had an $IC_{50} < 1$ μM. In all, 72% (60 compounds) of the naïve informer set hits retested with IC_{50}s < 5 μM. A third of these had an $IC_{50} < 1$ μM (see Supplemetnal File 2 for full screening results). Finally, stereoisomers for selected hits were also examined in dose response. These data (Figure 3.7) showed stereoselective inhibition; for example, only the S,S,S stereoisomer, BRD0326, is active ($IC_{50} = 0.152$ μM) in a set of eight stereoisomers tested.

3.3.5. MMV Malaria Box and Broad Library exoerythrocytic-stage active chemical clustering identifies active scaffolds and important targets

To cross validate these data, we first identified scaffolds that were enriched for compounds with exoerythrocytic-stage activity (relative to erythrocytic-stage or no antimalarial activity) using compound clustering. Given the small size of our initial compound set, we combined the MMV Malaria Box compounds with a GNF library consisting of 4,422 compounds that were previously screened for exoerythrocytic-stage activity in *P. berghei* (Meister et al., 2011). Merging the libraries allowed us to determine if there was any overlap in scaffold hits between the two compound sets. The 4,822 compounds were clustered using a hierarchical clustering method based on substructure similarity. To define scaffold groups, we separated clusters based on a minimum Tanimoto coefficient requirement of 0.65, resulting in 2,335 total clusters that ranged in size from 1–45 compounds. DOS library compounds were also included, but given that the library was designed to eliminate structural redundancy, it was expected that DOS compounds would not contribute significantly to any scaffold clusters.

We identified 15 scaffold series that showed specific enrichment in exoerythrocytic-stage activity with rates higher than expected by chance ($p < 0.001$) (Figure 3.8). The scaffold clustering showed multiple scaffold families already known to be active. One is the diaminotriazine scaffold (1228) similar to the diaminopyrimidine found in antifolate drugs such as pyrimethamine

(probability of enrichment by chance = 4.29e-7). Another is a tetracyclic benzothiazepine scaffold, cluster 1096, that has shown to target the Q_0 site of cytochrome bc1(Dong et al., 2011). Several enriched quinolone compounds (GNF-Pf-2549, GNF-Pf-4577, GNF-Pf-5037; cluster 342) were similar to ELQ300, a possible Q_1 site cytochrome bc1 inhibitor(Capper et al., 2015) with known causal prophylactic activity in mouse models of malaria (Nilsen et al., 2013). In addition, cluster 1096 may contain inhibitors of the electron transport chain (DHOD or cytochrome bc1) as do the three 4-quinolinol scaffolds (clusters 1122, 1613 and 2061). The compounds that were the precursors of the imidazolopiperazine compound in clinical trials, KAF156 (Kuhen et al., 2014), were also found in a cluster of three compounds (GNF-Pf-5069, GNF-Pf-5179, GNF-Pf-5466; cluster 849). These compounds work by an unknown mechanism of action, but resistance is conferred by mutations in the *P. falciparum* cyclic amine resistance locus *(Pfcarl)* (Meister et al., 2011). Additionally, compounds in cluster 2045 have some structural similarity to *P. falciparum* histone methyltransferase inhibitors (Malmquist et al., 2012).

Although most of the over-represented scaffolds have been investigated as starting points of antimalarial drug discovery in recent years, there were also novel notable singleton molecules whose hepatic stage activity had not been previously described. For example, MMV666693 (the most potent compound from the exoerythrocytic-stage screen of the MMV Malaria Box) also strongly inhibits erythrocytic-stage *P. falciparum* with IC_{50}

values reported below 100 nM (data not shown). This compound was previously identified as an allosteric inhibitor of *P. falciparum* Kinesin-5, a microtubule cross-linking enzyme required for cell division (Sinka et al., 2012). It is therefore interesting to speculate that targeting this enzyme may be an efficient means to inhibit parasite replication across multiple developmental stages.

3.3.6. MMV Malaria Box and Broad Library exoerythrocytic-stage active chemical clustering identifies active scaffolds and important targets

Compounds that are active only against exoerythrocytic stages, but not erythrocytic stages, may represent new opportunities for development of drugs for which resistance acquisition may be less of a problem. While 42 of the 63 exoerythrocytic-stage DOS compounds were active in both the exoerythrocytic- and erythrocytic-stage assays, others could be starting points for such drugs. Several of the compounds that were identified in the DOS library unbiased screen are highlighted in Figure 3.7. In particular, cyanoazetidine and a bicyclic azetidine series are shown. Although BRD7539 is active in both the exoerythrocytic- and asexual erythrocytic-stage assays, BRD9781 (its stereoisomer) is only active in the exoerythrocytic-stage assay. Although BRD7539 targets the *Plasmodium* DHODH enzyme (Nobutaka Kato et al.) based on its profile across lifecycle stages, we suspect that BRD9781 has a different target. These results suggest the important role of

stereochemistry in inhibiting different biological targets. BRD47390 appears to result in specific activity in the exoerythrocytic-stage assay, whereas the DOS Phenylalanine–tRNA ligase inhibitor is active against blood- and exoerythrocytic-stages. BRD0326, which has identical stereochemistry to BRD7539 but has differing functionality at the azetidine nitrogen, also has no erythrocytic-stage activity and may have yet another target. Further testing in a DHODH assay (Ross et al., 2014) confirmed that BRD0326 and BRD9781 are inactive (data not shown). BRD0326, BRD9781, and BRD47390 were all retested at dose along with all of their stereoisomers. With the exception of BRD7539, which targets DHODH, these compounds displayed stereoselective activity (Figure 3.7B). Overall these data show that up to a third of the hits in a screen of an unbiased library might have exclusive exoerythrocytic activity, targeting either unique EEF parasite targets or the host factors needed to support parasite replication.

3.3.7. Cross-validation using phenotypic assays to assess time of action during exoerythrocytic-stage development

We sought to further confirm exoerythrocytic-stage activity for the three most potent MMV Malaria Box hits by an orthogonal and complementary assay. This assay utilizes a previously described flow cytometry-based method (Prudencio et al., 2008a; Sinnis et al., 2013) to measure four specific metrics during exoerythrocytic-stage development: 1) sporozoite traversal, 2)

sporozoite invasion, 3) EEF frequency, and 4) EEF development. The assay uses *P. berghei* expressing GFP (Pb-GFP) infection of Huh7.5.1 cells, (Figure 6), a related exoerythrocytic cell line that can also be used to study EEF development (Figure 3.9) and measurements are taken at 2 and 48 hours post infection.

At 2 hours post infection, the number of hepatocytes that have been traversed and invaded is measured. Sporozoite traversal is inferred based on the observation that traversal temporarily ruptures the plasma membrane of hepatocytes, allowing high molecular weight rhodamine-dextran to stain cells that have been traversed but not invaded (Mota et al., 2001). Sporozoite invasion is measured by the percentage of cells expressing GFP and not rhodamine-dextran, as the parasite enters the cells via a moving tight junction, excluding rhodamine-dextran (Mota et al., 2001). Double positive cells, expressing both GFP and rhodamine-dextran, likely represent a population of non-productively infected cells, or cells in the process of being traversed by parasite, and are therefore not used in the measurements of traversal and invasion (Sinnis et al., 2013). Cytochalasin D, a potent inhibitor of actin polymerization, is used as a positive control for traversal and invasion inhibition at 2 hours post infection as it has been previously demonstrated to reduce sporozoite motility (Figure 3.10A) (Sinnis et al., 2013).

At 48 hours post infection, the frequency and development (size) of exoerythrocytic-stage EEFs can be measured. The frequency of EEFs is

determined by the percentage of cells expressing GFP at 48 hours post infection. KDU691, which inhibits *Plasmodium* phosphatidylinositol 4-kinase (PI(4)K), an enzyme that phosphorylates its phosphoinositide substrate to regulate intracellular signaling and trafficking (McNamara et al., 2013), was used as a positive control because it was shown to significantly decrease the number of exoerythrocytic-stage EEFs (Figure 6A). At the same time, the assay measures EEF development by reporting GFP mean fluorescence intensity (MFI), an indicator of EEF size. Here, atovaquone serves as a positive control (Figure 3.10A).

As predicted, all compounds led to changes in cell populations that were detectable by flow cytometry (Figure 3.10B). At 48 hours post infection, all of the compounds tested led to a significant decrease in the EEF size, but not frequency, similar to that of atovaquone (Figure 3.10B). In addition, MMV666693 also appears to have a slight effect on sporozoite traversal of hepatocytes at 2 hours post infection (Figure 3.10B). Somewhat unexpectedly, sporozoite invasion at 2 hours post infection was not affected by MMV666693, and rather, slightly increased. Unlike KDU691, none of these compounds affected EEF frequency. While further investigation is needed to better understand the specific mode of action, these results suggest that these compounds are acting primarily during EEF development. An important caveat of the flow cytometry-based assay is that there may be prolonged

fluorescence well after parasites have lost viability (Corish and Tyler-Smith, 1999).

To provide further validation of EEF development inhibition, MMV666693, MMV007160, and MMV665916, were characterized in an exoerythrocytic-stage time of action assay and compared to compounds with known activity in hepatic stages (Figure 3.11). This included atovaquone, which targets mitochondrial cytochrome bc1 complex and therefore inhibits the parasite's electron transport chain during all developmental stages (Flannery et al., 2013) (Figure 3.11B). Pyrimethamine inhibits dihydrofolate reductase (DHFR) and thus the synthesis of purines and pyrimidines required for DNA synthesis (Peterson et al., 1988). We also included several compounds in development such as DDD107498, which inhibits *Plasmodium* translation elongation factor 2 (eEF2) (Baragana et al., 2015), as well as KDU691 (McNamara et al., 2013). A third compound, GNF179, whose target has not been elucidated but which has potent causal prophylactic activity in mice and resistance mediated by mutations in the aforementioned *Pf*carl gene, was also included (Meister et al., 2011). Compounds were added in 12-point serial dilutions (10 μM to 56 pM) or washed out to be present during specific time frames throughout exoerythrocytic-stage development. Their activity was measured using a modified 384-well version of our high-throughput luciferase assay and bioluminescence recorded at 48 hours post-infection.

The assay supported the flow cytometry data and showed that MMV666693 and MMV007160 inhibit parasite replication during all time-points during exoerythrocytic-stage EEF development, and seem to be most potent when added during trophozoite development 6–24 hours post infection (Figure 3.11B). Unlike the other compounds tested, MMV665916 may need longer incubation *in vitro* in order to achieve optimal activity against exoerythrocytic-stage parasites, as all of the developmental time-points tested resulted in a more than 10-fold IC_{50} change compared to the 2–50 hours post infection control (Figure 3.11B). These data highlight how exoerythrocytic-active compounds may be further classified and how this assay may reveal information about a compound's mechanism of action.

3.4. Conclusions

The assays described here provide a high-throughput approach to identify scaffolds or scaffold families that will have causal prophylactic activity. Although our assay depends on rodent malaria parasites that are not infectious to humans, they have an advantage over human parasites because mosquitos infected with *P. berghei* or *P. yoelii* can be handled and shipped more easily. In addition, the number of sporozoites per mosquito is high (~20,000) enabling higher-throughput methods and evaluation of more starting points. This reduces cost (20 cents per well) as mosquito production and mosquito dissection are very labor intensive. One concern is that activity tests using rodent malaria parasites might not translate into activity against human parasites. In cases where a compound is also active against *P. falciparum* erythrocytic-stages, this is less likely to be a concern. In the small number of cases where compounds are not active in erythrocytic-stages, additional testing using exoerythrocytic-stages of *P. falciparum* or *P. vivax* may be warranted. Compounds of this class could be particularly interesting as leads for drugs that could be used in malaria elimination and eradication campaigns because they could be developed into drugs that could provide long-acting protection and which would not have the same resistance-development liabilities as compounds that act against the billions of erythrocytic-stage parasites that teem in an infected human. There is also the possibility that *P.*

falciparum- or *P. vivax*- specific hits may be lost, however, this is likely a small number of compounds given the screening throughput capacity of the assay.

An unanswered question is whether leads identified with this assay will have radical cure activity. Primaquine and tafenoquine are the only two compounds that can provide radical cures and both behave poorly in cellular assays such as those described here because they depend on host organismal metabolism. These compounds are not even particularly active in assays that involve primary hepatocytes (Bourzac, 2014). Likewise, we have recently tested a number of MMV Malaria Box screening hits in an *ex vivo P. cynomolgi* model of hypnozoite development (Zeeman et al., 2014), however, only one compound, MMV007224, had moderate activity (at 10 μM) against small or large forms in the assay (Table 3.2). Interestingly, this was also the only compound of the set tested with any pharmacokinetic exposure *in vivo* (Keshavjee and Farmer, 2012). This highlights the utility of using hepatoma cells rather than primary hepatocytes for screening purposes, as compounds with less than favorable pharmacokinetic properties can be evaluated for activity. As exoerythrocytic-stage active compounds identified in this report will also likely include inhibitors of *Plasmodium* exoerythrocytic-stage hypnozoites, it will be important to address compound metabolic stability *in vitro* and bioavailability *in vivo* during the development of novel hypnozoite models. This class of compounds will represent an important starting point for

the development of novel treatments capable of providing a malaria radical cure.

3.5. Acknowledgments

We thank our colleagues from The Genomics Institute of the Novartis Research Foundation, in particular Richard Glynne, Purvi Sanghvi, Annie Mak and Jason Matzen. They provided insight, expertise, and assistance that greatly assisted the research. The mosquitos were supplied by the Insectary Core Facility at New York University School of Medicine (http://microbiology-parasitology.med.nyu.edu/research/parasitology/insectary-core-facility-and-parasite-culture). Infection of mosquitos was performed by bite of mice infected with the same parasite in blood stage. This procedure was carried out in strict accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. The protocol was approved by the Institutional Animal Care and Use Committee of New York University School of Medicine (Protocol Number 150816), which is fully accredited by the Association For Assessment and Accreditation Of Laboratory Animal Care International (AAALAC).

The authors acknowledge the Medicines for Malaria Venture (MMV) for access to the MMV Malaria Box. SM, JS, CR, and YA was supported by MMV. VC was supported by the Bill and Melinda Gates Foundation (OPP104040). EAW was funded by the Bill and Melinda Gates Foundation (OPP1054480), MMV, and NIH (R01AI090141 and R01AI103058). MM was supported by the National Science Foundation (NSF) GRPF (DGE1144152), NK, EC, CS and

SLS was supported by the Bill and Melinda Gates Foundation (OPP1032518). SLS is an investigator at the Howard Hughes Medical Institute. CM, KG, and MI were funded by the Welcome Trust (WT078285). AMZ and CHMK were funded by the Wellcome Trust (WT078285) and received additional funding from MMV. The authors would like to acknowledge Omar Vandal for assistance with grant management, and Paula Maguina for administrative assistance.

Chapter 3, in full, has been published in ACS Infectious Diseases, 2016. Justine Swann, Victoria Doroski, Matthew Abraham, Sang Kim, Kerstin Henson, Maureen Ibanez, David Plouffe, Anne-Marie Zeeman, Clemens H. M. Kocken, Brice Campo, Case W. McNamara, Elizabeth A. Winzeler, Stephan Meister. “A luciferase-based drug screen identifies MMV compounds with liver stage antimalarial activity.” The dissertation author was the primary investigator and author of this paper.

3.6. Materials and Methods

3.6.1. Compound libraries

MMV Validation set of antimalarials. This collection of 50 known antimalarial powders were obtained from the MMV and are all commercially available and active primarily against erythrocytic-stage malaria parasites.

MMV Malaria Box. This open source compound library is comprised of 400 diverse open source compounds with proven antimalarial activity. 200 of these compounds are described by MMV as 'drug-like' and 200 as 'probe-like' compounds (Spangenberg et al., 2013). They have been distilled down from ~20,000 hits generated from a screening campaign of 4 million compounds from the libraries of St. Jude's Children's Hospital, Novartis, and GSK. All compounds are commercially available and the library is also available for free from MMV as long as the resulting data are published and placed in the public domain.

Broad Diversity-Oriented Synthesis Library. Two sets of compounds were tested from the Broad Diversity-Oriented Synthesis (DOS) library of 100,000 compounds. The first set included 9,886 compounds selected to represent the structural diversity of all of the scaffolds. This 'informer' set was used as a naïve library to screen for compounds with unknown activity against *Plasmodium*. The second set was comprised of 89 compounds previously

shown to be active in a *P. falciparum* erythrocytic-stage assay (IC_{50} < 2 μM, (Nobutaka Kato et al.)).

3.6.2. Parasites

P. berghei-ANKA-GFP-Luc-SM_{CON} (Pb-Luc)(da Cruz et al., 2012) and *P. berghei-GFP* (Pb-GFP) (McNamara et al., 2013) sporozoites were obtained by dissection of infected *Anopheles stephensi* mosquito salivary glands. Dissected salivary glands were homogenized in a glass tissue grinder and filtered twice through nylon cell strainers (20 μm pore size, Millipore SCNY00020) and counted using a Neubauer hemocytometer. The sporozoites were kept on ice until needed. Both Pb-Luc and Pb-GFP infected *A. stephensi* mosquitos were obtained from the Insectary Core Facility at New York University.

3.6.3. Cell lines

HepG2-A16-$CD81^{EGFP}$(Silvie et al., 2006) cells stably transformed to express a GFP-CD81 fusion protein were cultured at 37 °C and 5% CO_2 in DMEM (Invitrogen, Carlsbad, USA) supplemented with 10% FCS, 0.29 mg/ml glutamine, 100 units penicillin and 100 μg/ml streptomycin. Huh7.5.1 cells were cultured at 37 °C and 5% CO_2 in DMEM (Invitrogen, Carlsbad, USA) supplemented with 10% FCS (Corning cat# 35-011-CV), 200 U/ml penicillin, 200 μg/ml streptomycin (Invitrogen cat# 15140-122), 10 mM Hepes (Invitrogen cat# 15630-080), 1× Glutamax (Invitrogen cat # 35050-061), 1x non-esential

amino acids (Invitrogen). During infection, cell media was supplemented with 50 μg/ml gentamycin and 50 μg/ml neomycin. After infection, the antimycotic 5-fluorocytosine at a final concentration of 50 μg/ml was added to the media.

3.6.4. High-content imaging

The high-content imaging experiments were performed as previously described (Meister et al., 2011). Briefly, we used the HepG2-A16-CD81^{EGFP} host cells and either the *P. yoelii* or Pb-Luc rodent malaria parasites. We seeded the cells in 384-well plates and infected them with a ratio of 2:1 (cells:sporozoites). After staining the parasites with a polyclonal mouse anti-*Py*HSP70 antibody, these data were acquired (image analysis) with the Perkin Elmer Opera. Average parasite size per well served as the primary readout for compound effectiveness.

3.6.5. Luciferase-based high-throughput screening

Sporozoite infection. For the Pb-Luchigh-throughput screen, we utilized *P. berghei* since its higher infection rates of immortal human hepatocyte cell lines are more conducive to high-throughput screening than the infection rates of human malaria parasites. *P. berghei* is able to infect human hepatocarcinoma HepG2 cells expressing the tetraspanin CD81 receptor (Silvie et al., 2006). 3×10^3. HepG2-A16-CD81^{EGFP} cells in 5 µl medium (2×10^5 cells/ml, 5% FBS, 5×x Pen/Strep/Glu) were seeded into 1536-well, white, solid-bottom plates (Ref# 789173-F, Greiner Bio-One) 20-26 hours prior to the

actual infection. 18 hours prior to infection, 50 nl of compound in DMSO (0.5% final DMSO concentration per well) were transferred with a PinTool (GNF Systems) into the assay plates (10 µM final concentration). Atovaquone (12-point serial dilution starting at 10 µM) and 0.5% DMSO were used as positive and negative controls, respectively. Pb-Luc sporozoites were freshly dissected and their concentration adjusted to 200 sporozoites per µl. Penicillin and streptomycin were added at 5x concentration for a final 5x concentration in the well. The increased antibiotic concentration did not interfere with parasite or HepG2-A16-CD81^{EGFP} growth. The HepG2-A16-CD81^{EGFP} cells were then infected with 1×10^3 sporozoites per well (5 µl) using a single tip Bottle Valve liquid handler (GNF), and the plates were centrifuged for 3 minutes at room temperature and at 330x g (Eppendorf 5810 R centrifuge) on lowest acceleration and brake setting. The plates were then incubated at 37°C for 48 hours in 5% CO_2 with high humidity to minimize media evaporation and edge effect.

Bioluminescence quantification of exoerythrocytic forms (EEFs). After incubation, the parasite EEF growth was quantified by bioluminescence measurement. Media was removed by spinning the inverted plates at 150x *g* for 30 seconds. 2 µl BrightGlo reagent (Promega) were dispensed with the MicroFloliquid handler (BioTek). Immediately after addition of the luminescence reagent, the plates were vortexed at median intensity setting for 10 seconds and read by an EnVision Multilabel Plate Reader (PerkinElmer).

IC_{50} values were obtained using measured bioluminescence intensity and a non-linear variable slope four parameter regression curve fitting model in Prism 6 (GraphPad Software Inc).

Bioluminescence quantification of HepG2 cytotoxicity. After incubation, the HepG2 cytotoxicity was assessed by removing the media through an inverted spin of the plates at 150× *g* for 30 seconds and addition of 2 µl CellTiterGlo reagent (Promega diluted 1:2 with deionized water) per well using the MicroFloliquid handler (BioTek). Immediately after addition of the luminescence reagent, the plates were vortexed for 10 seconds and read with an EnVision Multilabel Reader (PerkinElmer).

Bioluminescence quantification of compound luciferase inhibition. After a 3 hour incubation period, 2 µl BrightGlo (Promega) were added to the wells with the MicroFlo liquid handler (BioTek). Immediately after addition, the plates were read by an EnVision Multilabel Reader (PerkinElmer).

Bioinformatic analysis of EEF inhibition, HepG2 cytotoxicity and luciferase inhibition. For the first screening round, the luminescence reads from each 1536-well plate were analyzed separately. Briefly, a lack of inhibition was defined as the average DMSO readings (64 wells) minus the baseline inhibition readings. For both, the first and second round of screening, the baseline for the EEF inhibition was defined as the average of the five highest atovaquone concentrations (10 wells), the baseline for the HepG2 cytotoxicity was defined as the average of the highest puromycin

concentrations (8 wells), and the baseline of the luciferase inhibition assay was defined as the average of 48 wells with 500 µM resveratrol. For the first round of screening, we determined the inhibition percentage relative to the normalized well concentrations for each compound. This analysis was repeated for every 1536-well plate. For the second reconfirmation round of screening, IC_{50} values were obtained using the average normalized bioluminescence intensity of 4 wells per concentration and plate (96 wells in total for each compound) and a non-linear variable slope four parameter regression curve fitting model in Prism 6 (GraphPad Software Inc).

3.6.6. Culturing asexual erythrocytic-stage (AES) parasites

P. falciparum parasites were cultured in complete medium containing 5% hematocrit in a low-oxygen atmosphere composed of 1% oxygen, 3% carbon dioxide and 96% nitrogen at 37°C. Complete medium consists of RPMI medium 1640 (with L-glutamine, without phenol red, Thermo Fisher Scientific) supplemented with 4.3% heat-inactivated O^+ human serum, 0.2% AlbuMAX II lipid-rich BSA, 0.014 mg/ml hypoxanthine, 3.4 mM NaOH, 38.4 mM Hepes, 0.2% glucose, 0.2% sodium bicarbonate and 0.05 mg/ml gentamicin.

3.6.7. Asexual erythrocytic-stage screening in a 1536-well plate format

Pathogenic asexual erythrocytic-stage parasites were screened using a modified fluorescence-based proliferation assay described previously (Delves et al., 2012). Briefly, *P. falciparum* 3D7 parasites were cultured until a

parasitemia of 3-6% was reached. The level of erythrocytic-stage parasitemia was determined by microscopic inspection of Giemsa-stained blood smears for the presence of parasites. A parasite suspension with 0.3% parasitemia and 4% hematocrit was prepared in screening medium consisting of RPMI medium 1640 (with L-glutamine, without phenol red) supplemented with 0.4% AlbuMAX II lipid-rich BSA, 0.014 mg/ml hypoxanthine, 3.4 mM NaOH, 38.4 mM Hepes, 0.2% glucose, 0.2% sodium bicarbonate and 0.05 mg/ml gentamicin. The parasite culture was gassed with 1% oxygen, 3% carbon dioxide and 96% nitrogen and stored at 37°C until used. Next, 3 μl of screening medium were dispensed into 1,536-well, black, clear-bottom plates (Ref# 789092-F, Greiner Bio-One, Kremsmünster, Austria) using the MultiFlo^{TM}Microplate dispenser (BioTek Instruments, VT, USA). For determining the IC_{50} values, 50 nl of compounds dissolved in DMSO (12-point serial dilutions (1:3) starting at 10 mM) were transferred into the assay plates (62.5 μM final drug concentration, 0.625% final DMSO concentration) using the Biomek® FXP Laboratory Automation Workstation (Beckman Coulter, CA, USA) with a PinTool (V&P Scientific, CA, USA). Artemisinin and DMSO were included as background and baseline controls, respectively. Next, 5 μl of prepared parasite suspension were dispensed into the 1,536-well plates resulting in a final parasitemia of 0.3% and a final hematocrit concentration of 2.5% (MultiFlo^{TM}Microplate dispenser). The assay plates were transferred into Ziploc^{TM} bags and gassed with a gas mixture of 1% oxygen, 3% carbon

dioxide and 96% nitrogen. After a 72 hour incubation at 37°C, 2 µl of detection solution consisting of 10x SYBR Green I (Thermo Fisher Scientific) in lysis buffer (20 mMTris/HCl, 5 mM EDTA, 0.16% Saponin, 1.6% Triton X-100) were added to the plates (MultiFlo™Microplate dispenser) and incubated for 24 hours at room temperature in the dark. After 24 hours, fluorescence signals were measured at 530 nm with a 485 nm excitation from the bottom using the 2104 EnVision®Multilabel Reader (PerkinElmer, MA, USA). After subtracting the signal of the highest concentration of artemisinin (background) from all output values and normalizing the values to the average DMSO signal, the IC_{50} values were calculated by using a non-linear variable slope regression curve-fitting model in GraphPad Prism (GraphPad Software Inc.).

3.6.8. Computational compound clustering

To evaluate clustering of exoerythrocytic-stage hits and enrichment of compound groups, the 400 compound MMV set was co-clustered with a set of > 4,000 compounds that had previously been evaluated in a *P. yoelii* high-content imaging assay (Meister et al., 2011). Briefly, SMILESL (simplified molecular-input line-entry system) strings were loaded into R, and the maximum common substructure Tanimoto coefficient (MCS-TC) was calculated using the fmcsR package (Srivastava et al., 1999). The compounds were then subsequently hierarchically clustered with the hclust package, using the ward.D2 agglomeration method pair clusters. To create compound bins,

the tree branches were separated at a maximum pair-wise distance of 0.4. Hypergeometric mean statistical tests were applied to each compound bin to identify sets where exoerythrocytic-stage activity was enriched.

3.6.9. Flow cytometry assay

Exoerythrocytic-stage Pb-GFP traversal, invasion, and schizont development were measured using a previously established flow cytometry-based method (Sinnis et al., 2013). Briefly, 24 hours prior to infection, $1.75x10^5$ Huh7.5.1 cells were seeded in 24-well plates in 1 ml of DMEM hepatocyte culture medium for the traversal and invasion assay, as well as for the quantitation of EEF size and frequency. Pb-GFP sporozoites were freshly isolated from infected *Anopheles stephensi* mosquitos as described above, and $3.5x10^4$ or $7.0x10^4$ sporozoites were added to the cells for the traversal/ invasion assay or EEF quantitation assay, respectively, and incubated for 2 hours. Rhodamine-dextran was added to the wells at a final concentration of 1 mg/ml for the traversal and invasion assay. The cells were washed after the 2 hour infection, and assayed using flow cytometry for Rhodamine-dextran and GFP signal (traversal and invasion, respectively), or incubated for 48 hours and assayed by flow cytometry for GFP frequency and MFI (EEF frequency and size, respectively). Data were analyzed using the FlowJo Software.

3.6.10. Pb-Luc time of action assay

For the Pb-Luc time-course assay, we seeded 1×10^4 Huh7.5.1 cells in 30 µl hepatocyte culture media per well in a 384-well plate (Greiner Bio) 24 hours before infection. Pb-Luc sporozoites were freshly dissected from infected *Anopheles stephensi* mosquitos and $5x10^3$ sporozoites in 30 µl were added to each well. The plates were centrifuged for 5 minutes at 330××x *g* and incubated for 2 hours at 37°C and 5% CO_2. After incubation, media was removed and 50 µl fresh culture media was added. 12-point serial dilutions of compound in DMSO were added and removed from the plates at the indicated time-points post infection. At 48 hours post infection, media was removed from the plates and 20 µl BrightGlo Reagent (Promega) was added to each well. Luciferase light units were measured by bioluminescence using an EnVision Multilabel Plate Reader (PerkinElmer).

3.6.11. *P. cynomolgi* liver assay

The *P. cynomolgi* assay was performed as previously reported (Zeeman et al., 2014). All Rhesus macaques (*Macaca mulatta*) used in this study were bred in captivity for research purposes, and were housed at the Biomedical Primate Research Centre (BPRC; AAALAC-certified institute) facilities under compliance with the Dutch law on animal experiments, European directive 2010/63/EU and with the 'Standard for Humane Care and Use of Laboratory Animals by Foreign institutions' identification number

A5539-01, provided by the Department of Health and Human Services of the US National Institutes of Health. The local independent ethical committee first approved all protocols.

3.7. Figures

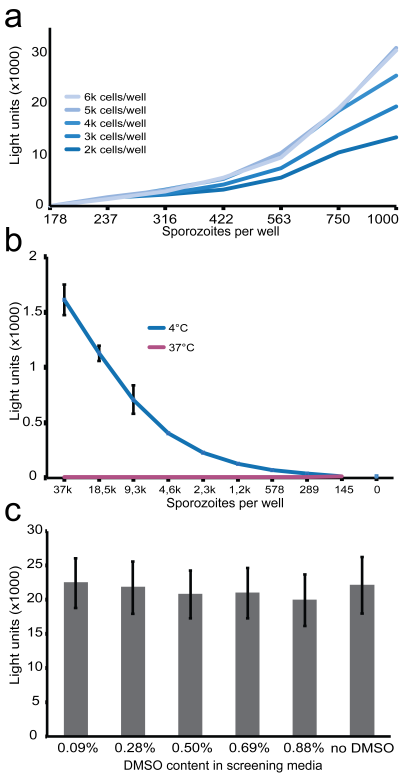


Figure 3.1. Luciferase-based high-throughput screening assay optimization. a) Optimization of Pb-Luc sporozoite to host HepG2-A16-CD81^{EGFP} cell ratio. Bioluminescence light output was measured for varying numbers of sporozoites in relation to cells per well. Higher cell numbers resulted in diminishing increases in light output. A ratio of 750 sporozoites to 3,000 cells was determined to be optimal for our assay. b) Luciferase light signal from sporozoites after 24 hours at 4°C and 37°C. In order to determine the contribution of extracellular sporozoites to the background signal intensity, we incubated sporozoites in cell culture media for 24 hours. After 24 hours at 37°C, there was negligible luminescent signal from even the highest concentration of sporozoites tested (37,000 per well), and thus extracellular sporozoites do not significantly contribute to background signal in our assay. c) Effect of different DMSO concentrations in the screening media on the malaria EEFs. The final DMSO concentration in the assay is 0.5%. DMSO concentrations up to 0.88% do not significantly alter EEF development d) Light output of QuantiLum Recombinant Luciferase in 1536-well assay conditions. We chose 0.02 µl per well to replicate the typical light outputfrom a Pb-Luc infected 1536 well. This was used to measure direct luciferase inhibition by the tested compound.

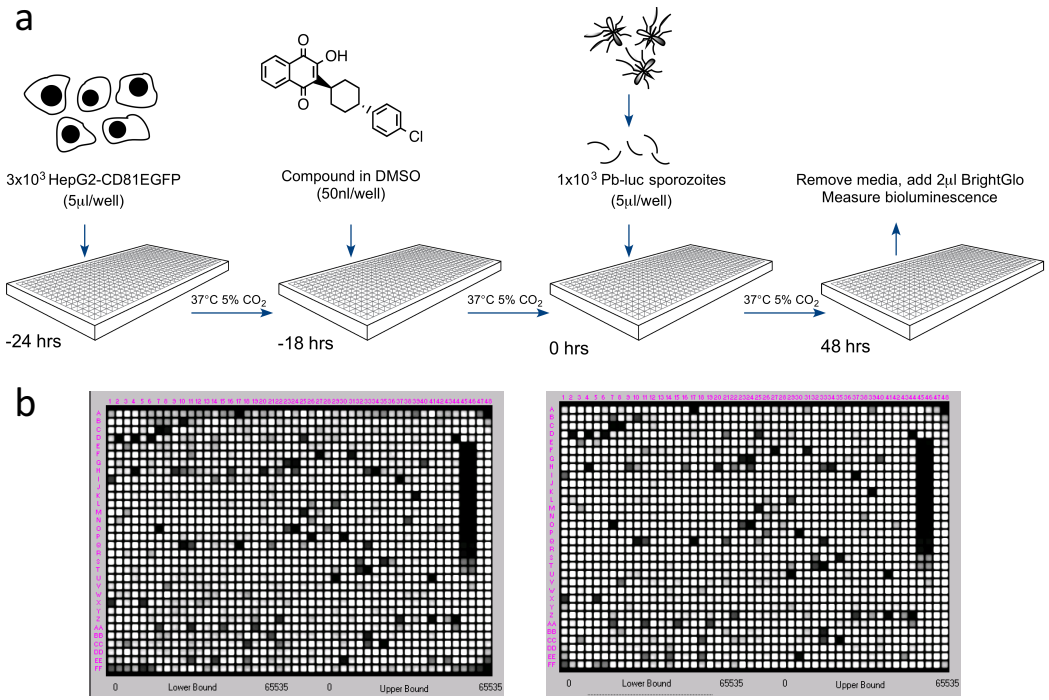


Figure 3.2. A luciferase-based high-throughput screening assay to identify malaria exoerythrocytic-stage inhibitors. a) Assay workflow. 24 hours prior to infection, 3x10^3 HepG2-A16-CD81^{EGFP} cells in 5 µl media were added to wells in a 1536-well assay plate. One to four hours later, 50 nl of compound dissolved in DMSO were added to the wells. At the time of infection, Pb-Luc sporozoites were freshly prepared from infected *A. stephensi* mosquitos and diluted to a concentration of 1x10^3 in 5 µl media per well. After 48 hours, Pb-Luc growth within hepatocytes was measured by bioluminescence.**.** b) As a proof of concept, we screened two plates containing 2,816 natural compounds (GNF) in replicate. One set of replicates is shown here. The average Z factor for these plates was 0.82.

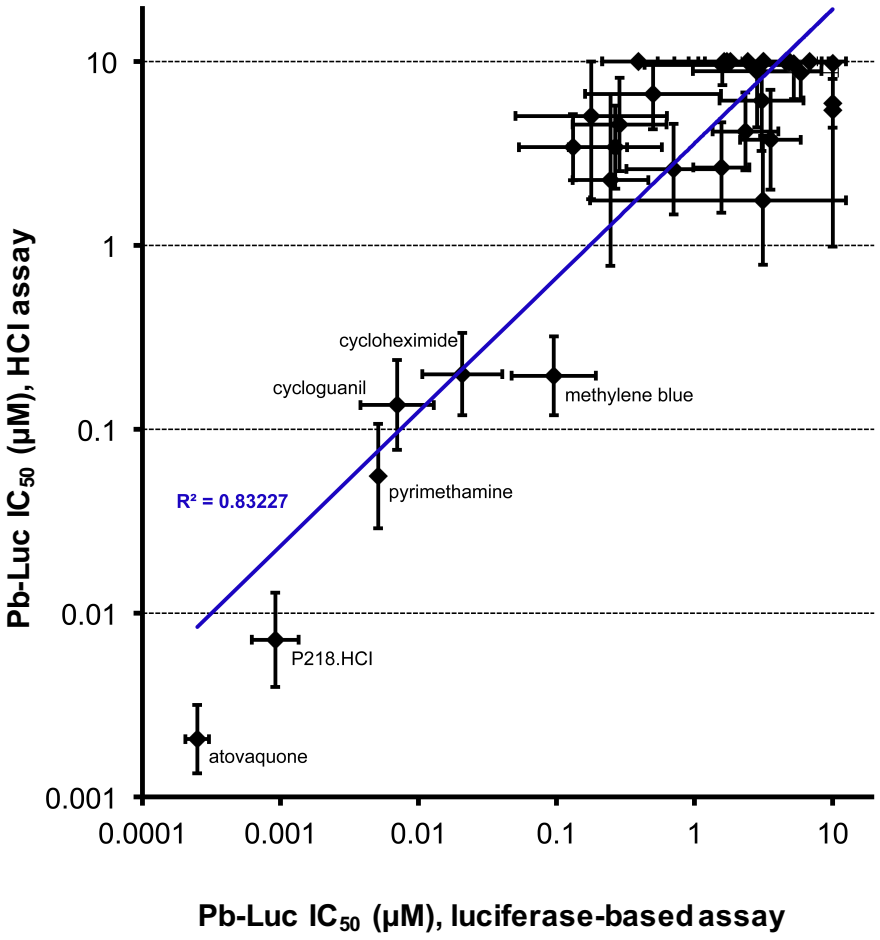


Figure 3.3. 1536-well luciferase-based screening assay is higher-throughput and more sensitive than former 384-well HCI assay. Pb-Luc IC_{50} values for the MMV validation set of antimalarials screened by the luciferase-based 1536-well assay and by the 384-well high-content imaging (HCI) assay are compared. Each data point represents a single antimalarial compound. The most active compounds in both assays are labeled. The assays generate Pb-Luc IC_{50} values that correlate very well with eachother, as demonstrated by an R^2 value of 0.83, however, the luciferase-based assay resulted in IC_{50}s roughly 10× lower than in the HCI assay. Artemether and sulfadoxine were classified as inactive by HCI, however led to parasite IC_{50} values of roughly 1 μM and 100 nM, respectively, by luciferase screening. Figure provided by Stephan Meister.

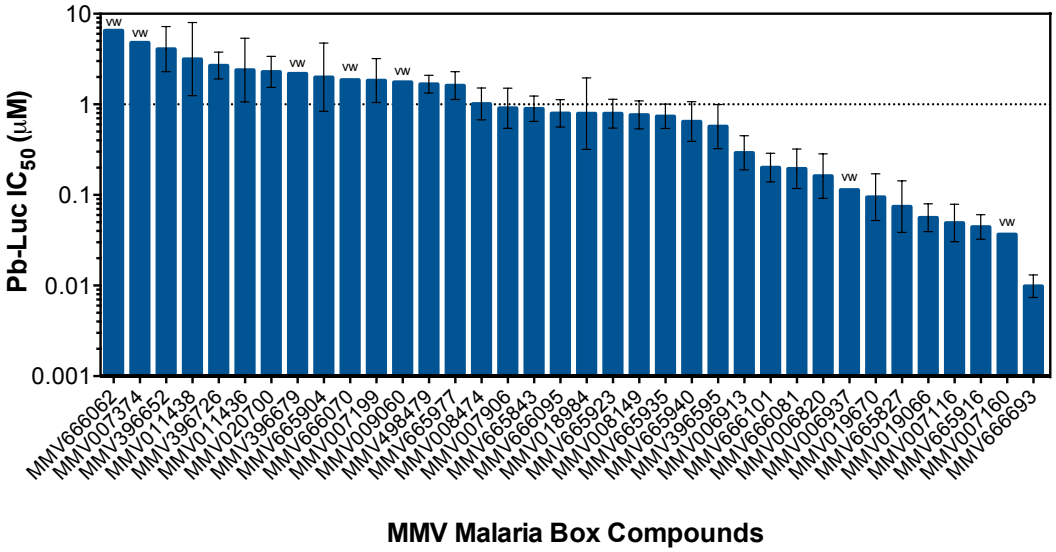


Figure 3.4. MMV Malaria Box compounds are identified as potent malaria exoerythrocytic-stage inhibitors. The 36 exoerythrocytic-stage active MMV Malaria Box hits with a Pb-Luc IC_{50} of less than 10 μM are shown with their respective IC_{50} values. Hits were also selected to demonstrate a HepG2 IC_{50} and a luciferace C_{50} of more than 10 μM. A line at 1 μM highlights that almost half of the exoerythrocytic-stage active hits display nanomolar potency. Error bars represent the 95% confidence interval. vw= very wide confidence interval

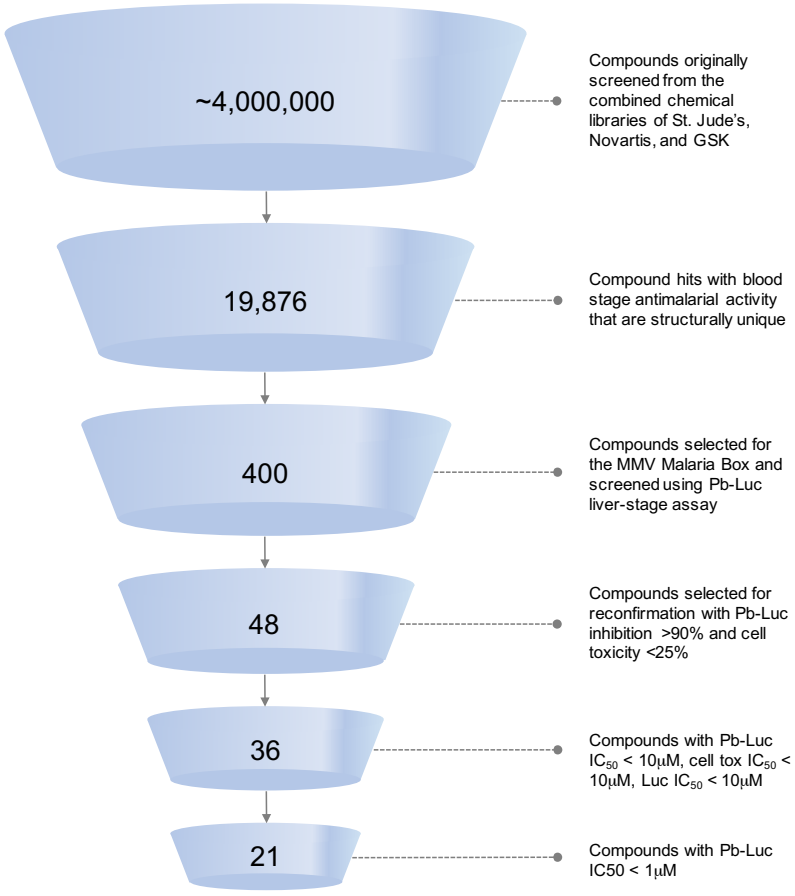


Figure 3.5. The selection process for the MMV Malaria Box compounds and exoerythrocytic-stage active hits. Diagram illustrating the flow of the selection process for the inclusion of MMV Malaria Box compounds and the resulting malaria exoerythrocytic-stage active screening hits. The selection process begins with the creation of the MMV Malaria Box, a collection of 400 diverse compounds with erythrocytic-stage antimalarial activity derived from the combined screening of roughly 4 million compounds. Our high-throughput exoerythrocytic-stage screen identified 36 compounds with exoerythrocytic-stage activity of less than 10 μM, and 21 compounds with exoerythrocytic-stage activity of less than 1 μM.

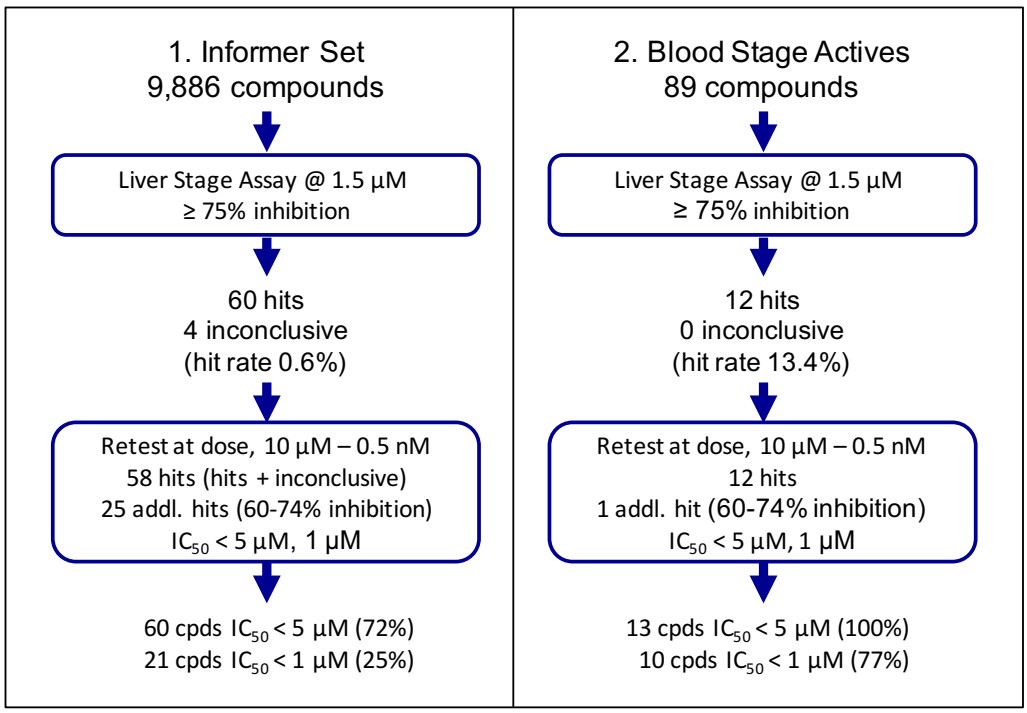
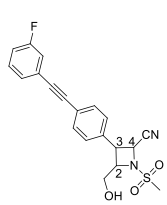


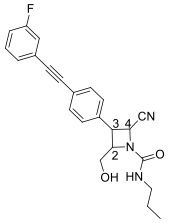
Figure 3.6. Overview of the Broad Diversity-Oriented Synthesis Library Screen. Screening pipeline for naïve informer set compounds and pre-selected erythrocytic-stage compounds. The hit rate for the naïve compounds set was 0.6%, while the hit rate for the pre-selected compounds was 13.4%.

a



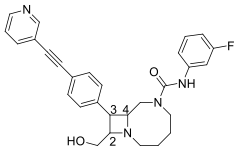
BRD0326
Stereochemistry (C2C3C4)
Pb-Luc IC_{50} (μM)

SRR 42	**SSR** 20	**SRS** 17	**SSS** 0.152
RRR 18	**RSR** 16	**RRS** 19	**RSS** 8



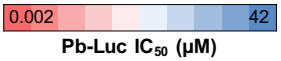
BRD9781/7539 series
Stereochemistry (C2C3C4)
Pb-Luc IC_{50} (μM)

SRR 0.039	**SSR** 19	**SRS** 12	**SSS** 0.002
RRR 1.6	**RSR** 10	**RRS** 10	**RSS** 0.36



BRD47390
Stereochemistry (C2C3C4)
Pb-Luc IC_{50} (μM)

SRR 5.7	**SSR** 0.229	**SRS** 1	**SSS** 3
RRR 2	**RSR** 10	**RRS** 2	**RSS** 3



b

	BRD0326	BRD9781	BRD7539	BRD47390
Stereochemistry ($C_2C_3C_4$)	SSS	SRR	SSS	SSR
Exoeyrthrocytic-stage Pb-Luc IC_{50} (μM)	0.3 (0.15)	0.08 (0.04)	0.01 (0.002)	0.36 (0.23)
Erythrocytic-stage (Pf Dd2) IC_{50} (μM)	> 10	8.547	0.01	2.58
HepG2 cells CC_{50} (μM)	> 10	>10	> 10	> 10

Figure 3.7. DOS compounds exhibit stereoselective inhibition of Pb-Luc exoerythrocytic-stage parasite growth. a) Representative compounds and activity profiles with activity against *P. berghei* in HepG2 cells. Stereocenters (C_n) are listed below the corresponding chemical structure. Pb-Luc exoerythrocytic-stage activity was measured for each of the 8 possible stereoisomers (SRR, SSR, SRS, SSS, RRR, RSR, RRS, and RSS) of each compound. Three of the eight possible stereoisomers of BRD9781 have exoerythrocytic-stage activity, two with potent activity (SRR and SSS, IC_{50} <0.1 μM), and another with moderate activity (RSS, IC_{50} < 1 μM). B. One stereoisomer of BRD0326 (SSS) is active (IC_{50} < 1 μM). C. Two stereoisomers of BRD47390 have significant exoerythrocytic-stage activity (SSR, IC_{50} < 0.1 μM; SRS, IC_{50} < 0.1 μM). b) Compounds were tested in dose in the *P. berghei*/HepG2 assay, a Dd2 erythrocytic stage assay, and in a mammalian cell cytotoxicity assay. Compounds from three scaffold libraries are shown. Compounds were tested twice in the exoerythrocytic-stage assay; values from the second assay are shown in parentheses.

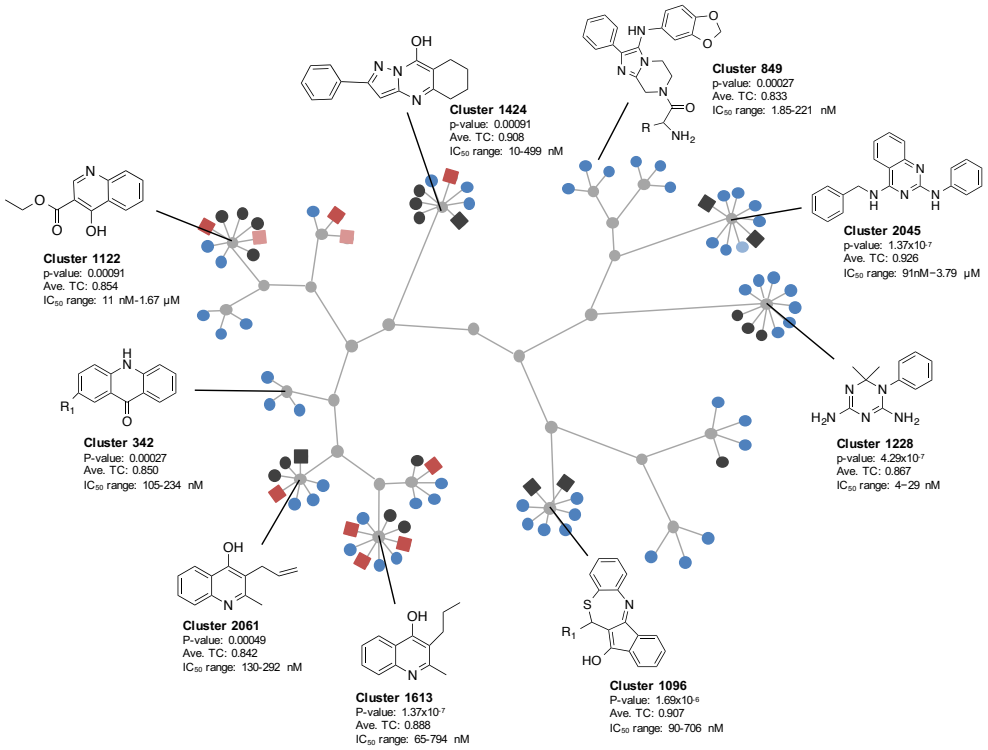


Figure 3.8. Exoerythrocytic-stage active MMV compounds display unique chemical scaffold clustering. Compounds of the GNF and MMV Malaria Box were clustered by their substructure similarity, binning sets based on main common substructure (Tanimoto average compound similarity ≥0.85). Out of 2335 cluster sets, 15 were significantly enriched for exoerythrocytic-stage active compounds (depicted above). GNF Malaria Box compounds are shown as circle nodes, and MMV Malaria Box compounds as square nodes. Active compounds are indicated in red and blue, for MMV and GNF compounds, respectively. Dark red/blue signifies IC_{50} < 1 µM, while light red/blue signifies IC_{50} < 10 µM. Inactive compounds are shown in black, and base scaffolds are shown in grey. It is important to note that of the 16 MMV compounds represented in the scaffold clustering, 13 of them were structurally identical to compounds also screened by the Novartis library using high-content imaging (TC=1).

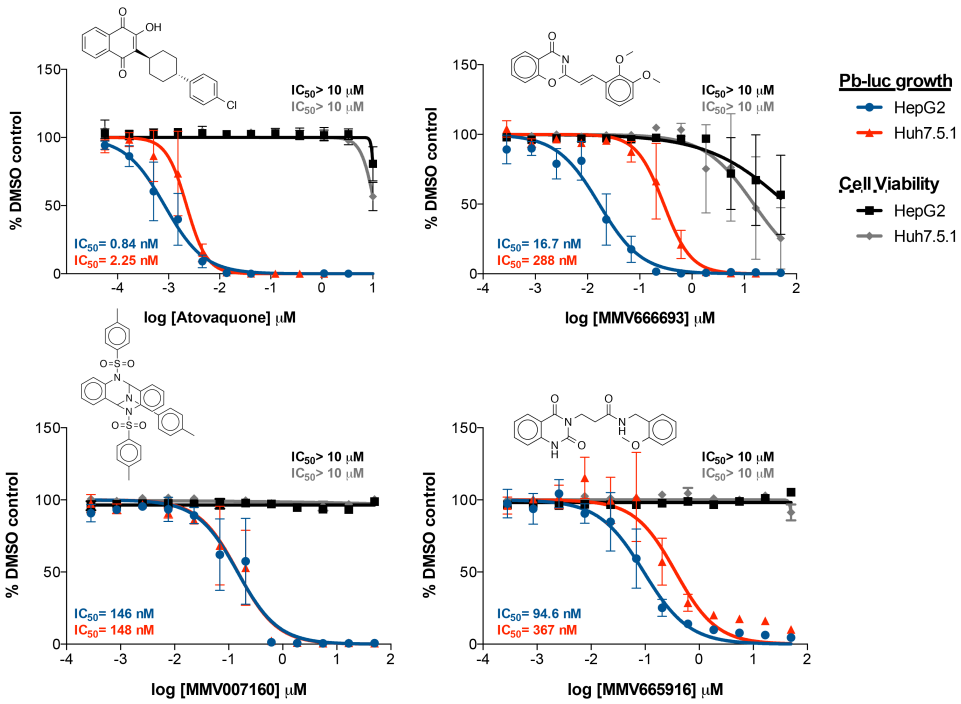


Figure 3.9. Contribution of different host cell lines to compound potency. Serial dilutions of MMV666693, MMV007160, and MMV665916 were performed to calculate Pb-Luc growth in both HepG2 and Huh7.5.1 cell lines using the high-throughput luciferase-based assay. Cell viability was determined by monitoring HepG2 growth using a bioluminescent assay (Cell Titer Glo, Promega). Chemical structures for selected compounds are displayed.

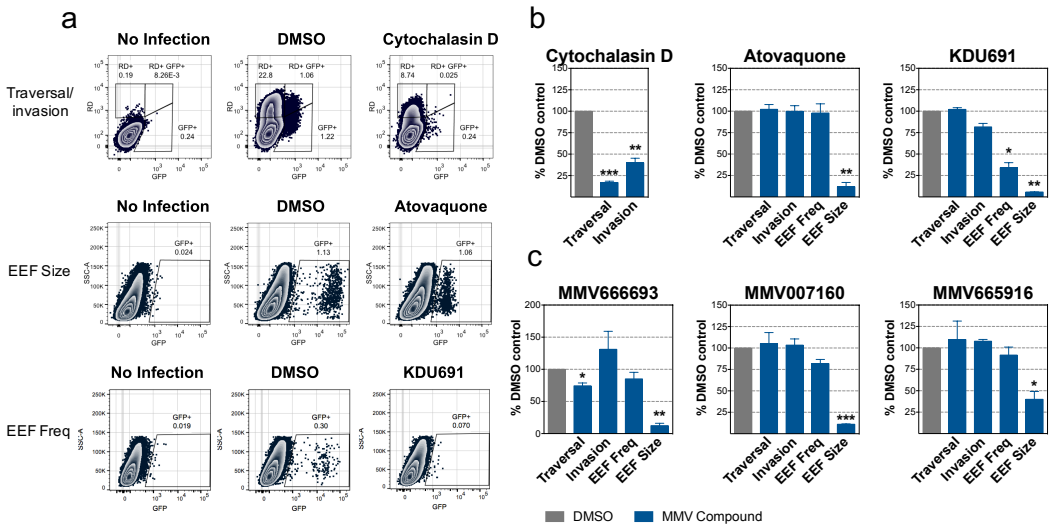
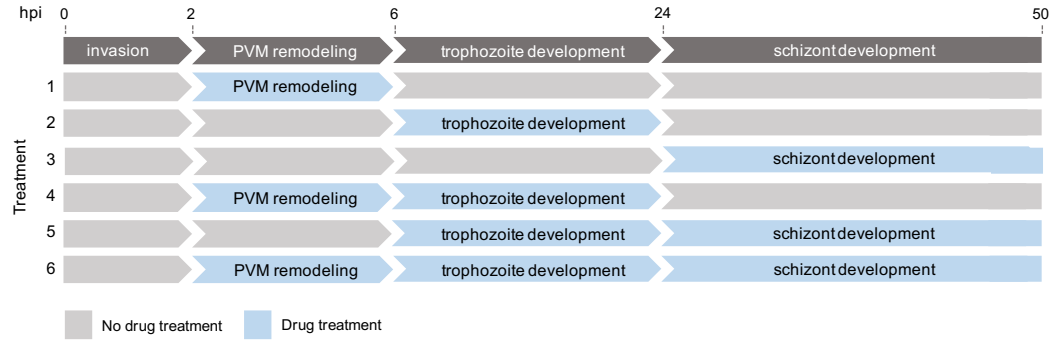


Figure 3.10. Validation of exoerythrocytic-stage activity using an established flow cytometry-based assay. a) Flow cytometry plots measuring traversal and invasion (at 2 hours post infection), and EEF frequency and development (at 48 hours post infection) of exoerythrocytic-stage malaria parasites in Huh7.5.1 cells as previously described for three of the most potent compounds. Cytochalasin D was used as a positive control for traversal and invasion, KDU691 was used as a positive control for EEF frequency, and atovaquone was used as a positive control for EEF frequency and development. While traversal was measured by the percentage of Rhodamine-dextran single-positive cells, invasion was measured by the percentage of Pb-GFP single-positive cells at 2 hours post infection. At 48 hours post infection, EEF frequency was measured by the percentage of Pb-GFP-positive cells, and EEF development was measured by the relative mean fluorescence intensity (MFI). Representative flow cytometry plots are shown. Atovaquone was tested at 1uM (due to slight cytotoxicity at 10μM in Huh7.5.1 cells, (Supplemental Figure 4)), and Cytochalasin D and KDU691 was tested at 10μM. b) The mean and SEM are shown graphically from the traversal/ invasion, EEF frequency, and EEF size control experiments shown in Figure 6a above (cytochalasin D, atovaquone, and KDU691, respectively). Values are normalized to the DMSO control. c) Exoerythrocytic-stage traversal, invasion, EEF frequency, and EEF development are shown for MMV666693, MMV007160, and MMV665916. Mean and SEM from three replicate experiments are shown. MMV66693 was tested at 1 μM (due to slight cytotoxicity at 10 μM in Huh7.5.1 cells, (Figure 3.9), and MMV007160 and MMV665916 were tested at 10 μM.

a



b

Treatment	Established Antimalarials: Atovaquone IC50 (nM)	Atovaquone Fold IC50	Pyrimethamine IC50 (nM)	Pyrimethamine Fold IC50	Antimalarials in Development: GNF179 IC50 (nM)	GNF179 Fold IC50	KDU691 IC50 (nM)	KDU691 Fold IC50	DDD107498 IC50 (nM)	DDD107498 Fold IC50	MMV Malaria Box Liver-Stage Hits: MMV666693 IC50 (nM)	MMV666693 Fold IC50	MMV007160 IC50 (nM)	MMV007160 Fold IC50	MMV665916 IC50 (nM)	MMV665916 Fold IC50
1	34.0	37.6	122.5	6.5	23.3	2.9	656.9	2.8	2.2	3.3	172	14.4	791.5	4.4	5729	15.5
2	2.8	3.1	21.0	1.1	45.1	5.7	265.5	1.1	1.6	2.4	20	1.7	285.4	1.6	5782	15.6
3	7.9	8.8	251.8	13.3	24.2	3.0	249.8	1.0	0.8	1.2	297	24.9	2876	16.1	8162	22.0
4	1.6	1.8	19.7	1.0	8.0	1.0	163.1	0.7	1.1	1.6	15	1.3	177.4	1.0	4290	11.6
5	1.1	1.2	18.8	1.0	15.8	2.0	154.8	0.6	0.9	1.3	16	1.4	198	1.1	785.9	2.1
6	0.9	1.0	18.9	1.0	8.0	1.0	238.7	1.0	0.7	1.0	12	1.0	178.9	1.0	370.2	1.0

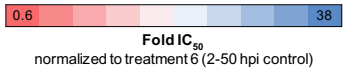


Figure 3.11. Exoerythrocytic-stage active compounds display unique potencies during exoerythrocytic-stage EEF development. a) Diagram illustrating the major stages of malaria parasite exoerythrocytic-stage development, including invasion, parasitophorous vacuole membrane (PVM) remodeling, trophozoite development, and EEF schizont development. During the first four hours after sporozoite invasion, the parasite dramatically remodels its parasitophorous vacuole membrane by degrading host cell-derived proteins and at the same time inserting its own parasite-derived proteins(Prudencio et al., 2006a). During the next 18 hours, the sporozoites transform from their elongated motile form to round, non-motile, and metabolically active trophozoites. The trophozoites undergo impressive nuclear replication starting at around 24 hours post infection, displaying one of the fastest replication rates known to eukaryotic organisms to develop into mature EEFs (Prudencio et al., 2006a). Drug treatments 1-6, corresponding to compound incubation during the exoerythrocytic developmental stages indicated, are shown. b) Pb-Luc IC_{50} data for established antimalarial compounds (atovaquone and pyrimethamine), antimalarials in development (GNF179, KDU691, DDD107498), and the three MMV Malaria Box compounds (MMV666693, MMV007160, and MMV665916) added during Pb-Luc exoerythrocytic-stage development in a modified 384-well luciferase-based assay (discussed in Methods) are shown. Likewise, the Pb-Luc IC_{50} fold changes normalized to the 2-50 hour drug-treated controls are shown and colored based on the indicated heat map.

3.7. Tables

Table 3.1. IC_{50} data for validation set of 50 antimalarials

Compound Class	Common Name	*P. berghei* (EEF[a])		*P. falciparum* (AES[b])	
		IC_{50} (nM)	95% CI (nM)	IC_{50} (nM)	95% CI (nM)
4-aminoquinolines	***Naphthoquine****	***292***	***115 to 741***	***11.4***	***7.6 to 17.2***
	AQ-13	1936	218 to 17220	8.7	ND
	Amodiaquine	4126	571 to 2980	>62500	ND
	Pyronaridine	>10000	ND	9.6	ND
	Piperaquine	>10000	ND	72	ND
	Hydroxychloroquine	>10000	ND	13.9	ND
	Chloroquine	>10000	ND	2.1	1.7 to 2.7
8-aminoquinolines	***Primaquine****	***773***	***207 to 2883***	***695***	***398 to 1212***
	NPC-1161B	9504	1075 to 84020	1320	855 to 2040
	Pamaquine	>10000	ND	>10000	8435 to 36010
	Tafenoquine	>10000	ND	2818	1992 to 3986
Amino alcohols	Halofantrine	2225	360 to 1375	1.5	1.2 to 1.9
	Mefloquine (+ RS)	4638	905 to 23770	16.6	13.1 to 21.1
	Mefloquine (Racemic)	7836	1419 to 43260	<0.4	ND
	Quinidine	8437	108 to 660	2.9	2.3 to 3.8
	Quinine sulfate dihydrate	>10000	ND	12.3	4.2 to 35.8
	Lumefantrine	>10000	ND	7.3	5.9 to 9.1
Antibiotics	***Trimethoprim****	***122***	***39 to 385***	***5640***	***3396 to 9367***
	Thiostrepton	4976	899 to 2756	1908	950 to 3833
	Doxycyclin	>100000	ND	8822	5336 to 14590
	Cis-Mirincamycin	>100000	ND	7560	4457 to 12820
	Trans-Mirincamycin	>100000	ND	>10000	10250 to 36050
	Fosmidomycin	>100000	ND	>10000	5784 to 18320
	Clindamycin	>100000	ND	198	157 to 251
	Tetracycline	>100000	ND	>10000	13340 to 33130
	Azithromycin	>100000	ND	257	199 to 331
Antifolates	***P218.HCl****	***0.32***	***0.18 to 0.55***	***0.4***	***0.3 to 0.4***
	Pyrimethamine*	***3.7***	***1.5 to 9.0***	***19.9***	***12.1 to 32.7***
	Cycloguanil*	***6.7***	***2.9 to 16***	***3.1***	***2.4 to 3.9***
	Dapsone*	***133***	***15 to 1179***	***>62500***	***ND***
	Chlorproguanil HCL	1038	236 to 4562	4716	2697 to 8248
	Proguanil HCL	>10000	ND	6239	3426 to 11360
Endoperoxidases	***Artemisone****	***92***	***19 to 437***	***0,7***	***0.6to 0.8***
	Artemisinin*	***218***	***29 to 1619***	***4.7***	***3.6 to 6.1***
	Artesunate*	***293***	***94 to 916***	***0.9***	***0.6 to 1.3***
	OZ277*	***524***	***92 to 2990***	***2.3***	***1.9 to 2.7***
	Artemether*	***909***	***204 to 4059***	***<0.4***	***ND***
	Artenimol	2462	822 to 7374	<0.4	ND
	OZ439	>10000	ND	2.8	2.5 to 3.2
Others	***Atovaquone****	***0.18***	***0.11 to 0.31***	***<0.4***	***ND***
	Cycloheximide*	***23***	***9.1 to 57***	***1.8***	***1.2 to 2.7***
	Methylene Blue*	***92***	***34 to 246***	***7984***	***1308 to 48740***
	Pentamidine*	***378***	***116 to 1228***	***0.5***	***0.2 to 1.7***
	Deferoxamine	>10000	ND	>10000	11830 to 33830
	Dehydroepiandroster-one sulphate	>10000	ND	>62500	ND
	N-acetyl-D-penicillamine	>10000	ND	>62500	ND
	Riboflavin	>10000	ND	>62500	ND
Sulfonamides	***Sulfadiazine****	***16***	***2.0 to 126***	>62500	***ND***
	Sulfadoxine*	***74***	***103 to 538***	>62500	***ND***
	Sulfamethoxazole*	***216***	***75 to 625***	***>10000***	***ND***

* Antimalarial compounds with exoerythrocytic-stage activity (Pb-Luc IC_{50} < 1000 nM)
EEF, exoerythrocytic form; AES, asexual erthrocytic stage

Table 3.2. Anti-*Plasmodium cynomolgi* exoerythrocytic-stage activity in primary hepatocytes for selected MMV Malaria Box Compounds.

Compound	*P. cynomolgi*, primary hepatocytes (% inhibition normalized to untreated control)					
	Small 0.1 μM	Small 1 μM	Small 10 μM	Large 0.1 μM	Large 1 μM	Large 10 μM
KAI407	-1.07	92.53	100.00	22.98	94.15	100.00
Primaquine	4.09	18.86	24.02	0.84	2.58	12.70
MMV006962	-8.08	11.32	17.09	34.17	24.68	9.71
MMV007160	17.44	-24.73	-11.21	16.97	13.65	0.05
MMV007224	30.48	20.32	66.74	22.75	43.35	95.65
MMV019066	10.32	11.21	3.20	3.22	-1.85	5.11
MMV665807	17.09	33.95	24.02	6.49	10.35	11.96
MMV665916	20.46	-13.35	-15.12	5.59	2.11	-13.23
MMV666101	4.16	19.40	10.85	7.30	10.68	16.79
MMV666693	29.36	16.37	30.07	22.51	-2.95	15.39

4. Systems Analysis of Host-Parasite Interactions

4.1 Abstract

Parasitic diseases caused by protozoan pathogens lead to hundreds of thousands of deaths per year in addition to substantial suffering and socioeconomic decline for millions of people worldwide. The lack of effective vaccines coupled with the widespread emergence of drug-resistant parasites necessitates that the research community take an active role in understanding host-parasite infection biology in order to develop improved therapeutics. Recent advances in next-generation sequencing and the rapid development of publicly accessible genomic databases for many human pathogens have facilitated the application of systems biology to the study of host-parasite interactions. Over the past decade, these technologies have led to the discovery of many important biological processes governing parasitic disease. The integration and interpretation of high-throughput –omic data will undoubtedly generate extraordinary insight into host-parasite interaction networks essential to navigate the intricacies of these complex systems. As systems analysis continues to build the foundation for our understanding of host-parasite biology, this will provide the framework necessary to drive drug

discovery research forward and accelerate the development of new antiparasitic therapies.

4.2 Introduction

Protozoan parasites infect over a half billion people worldwide, and continue to play a significant role in shaping global mortality and morbidity rates despite decades of research (Sacks and Sher, 2002). Some important human diseases caused by these pathogens include malaria, leishmaniasis, African sleeping sickness, toxoplasmosis, Chagas disease, and amoebiasis. The lack of government funding for many of these classically 'neglected' pathogens, the recent emergence of antiparasitic drug resistance, and the absence of licensed vaccines warrant global concern. In addition, host-parasite research is impeded by specific technical and resource limitations. New cost-effective, high-throughput strategies are therefore necessary to circumvent these obstacles and to develop novel therapeutics.

The post-genomic era has generated unparalleled opportunities for creating and integrating systems biology data (i.e. organism- or cellular-scale data produced through a number of –omic, or system-wide, technologies). This holistic approach is in direct contrast to conventional reductionist methods that 'reduce' systems into smaller, more tractable units. Systems-based methods are particularly useful to study complex biological relationships that are: 1. open, with constant information exchange and a net flow of resources, and 2. stochastic, with spatial, temporal, and population heterogeneity (Beiting and Roos, 2011). Host-parasite systems embody all of these defining

characteristics. -Omic technologies are also much more efficient and economical when comparing the cumulative time, labor, and cost per gene to traditional reductionist strategies. Not surprisingly, these methods have been critical for improving our understanding of host-parasite relationships and accelerating anti-parasitic drug discovery (Sakata and Winzeler, 2007; Winzeler, 2006). In this review, we discuss the current state of host-parasite systems biology research. This includes the various obstacles faced by parasite researchers, the advancements and feasibility of several genome-wide technologies, and the key research areas benefiting from such approaches. We aim to emphasize major advances from the past few years as well as the specific hypotheses and gaps emerging from these studies.

4.3 Unique challenges of the host-parasite Interface

4.3.1. Complex lifecycles within multiple hosts

Parasites have evolved elegant strategies to survive and replicate within their hosts. One strategy includes constantly changing their cellular state in order to progress through their lifecycle, while simultaneously evading recognition by the host immune system (Sacks and Sher, 2002). The vast number of developmental stages, combined with distinct tissue tropisms, increases the complexity of host-parasite interactions (Table 4.1). In this review, we focus on the following species due to their global impact on human health and influence in the research community: 1. the Apicomplexans *Toxoplasma gondii,* which causes toxoplasmosis, *Plasmodium* spp., which cause malaria, and *Crypotosporidium* spp., which cause the diarrheal disease cryptosporidosis; 2. the Kinetoplastids *Trypanosoma brucei,* which causes African sleeping sickness, *Trypanosoma cruzi*, which causes Chagas disease, and the *Leishmania* parasites, which cause both cutaneous and visceral leishmaniasis; 3. the Diplomonad *Giardia lamblia*, which causes the intestinal disease giardiasis; and 4. the Amoebozoa *Entamoeba histolytica*, which causes amoebic dysentery (Table 4.1). Many of these parasites, such as the Apicocomplexans and the Kinetoplastids, are vector-borne, intracellular pathogens that complete their lifecycle within multiple hosts. One exception is *T. brucei*, which carries out its lifecycle extracellularly. Others, such as

Cryptosporidium, *Entamoeba*, and *Giardia*, can develop into infectious, resistant cysts that survive outside of their hosts and are generally spread via the fecal-oral route.

Due to the important differences in each lifecycle stage, researchers must consider these unique developmental niches as separate systems when studying host-parasite interactions. This is especially important for systems-based analysis, as parasites display periodic stage-dependent gene expression (Hall et al., 2005). Accordingly, even slight asynchrony within parasite samples can result in inaccurate gene expression measurements, severely limiting statistical power. This is particularly challenging when analyzing clinical samples *ex vivo*, as parasite populations are rarely homogeneous. Therefore, researchers often utilize specialized techniques in order to synchronize parasites in culture, isolate specific cellular stages from mixed culture, or computationally remove stochastic noise (Lemieux et al., 2009). While this experimental isolation of developmental stages will aid in the understanding of stage-specific host-parasite interactions, it will also be important for future systems-based studies to integrate this knowledge into a multi-stage model more representative of physiological mixed parasite populations.

4.3.2. Challenging *in vitro* culture

While it is certainly possible to utilize systems-based approaches for *in vivo* and *ex vivo* study, the establishment of *in vitro* methods is particularly useful for many high-throughput applications. The complex nature of each parasite's lifecycle often requires multiple *in vitro* culture systems in order to study all of the developmental stages. While some parasite stages are easily propagated in culture, others are not (Visvesvara and Garcia, 2002; Widmer et al., 2000). For example, blood-stage *Plasmodium falciparum* parasites can be maintained almost indefinitely in culture if supplied with fresh erythrocytes; however, sporozoites are generally freshly isolated from the salivary glands of infected mosquitos when studying liver stage infection. In order to study hypnozoites (the clinically dormant hepatic stage of *P. vivax* and *P. ovale*), researchers rely on technically challenging, time-intensive assays only available in locations where the species are accessible (Mueller et al., 2009). The absence of methods to isolate developmentally synchronized cysts presents a major hurdle for the study of encystation and excystation by enteric parasites such as *E. histolytica* (Ehrenkaufer et al., 2013) and *C. parvum* (Karanis and Aldeyarbi, 2011). Furthermore, low *in vitro* infection rates for many protozoan pathogens often lead to insufficient material for systems-based analysis.

4.3.3. Large uncharacterized genomes

Not only are parasite genomes generally larger and more complex than their prokaryotic counterparts, but their functional characterization and annotation is severely limited by both lack of genetic tools and resources (Winzeler, 2006). Fully sequenced and annotated genomes greatly strengthen many areas of systems-biology research; this includes determination of coding and non-coding reading frames, alternative splice variants, and the assignment of gene functions. Although many important protozoan parasite genomes have been sequenced, an overwhelming percentage of genes are still assigned 'hypothetical' functions, as illustrated in Figure 4.1. This lack of functional characterization is correlated with the relative magnitude of the research community (Figure 4.1), as well as other factors such as the genetic intractability of certain species. The 24-Mb *P. falciparum* genome, for example, is extremely AT-rich (80.6%) and therefore traditional genetic approaches are particularly challenging (Winzeler, 2006). Not surprisingly, about 40% of the genome is still uncharacterized. For other parasites such as *T. gondii* and *L. major*, this number is even higher, with more than half of annotated genes assigned "hypothetical" functions (Figure 4.1).

Moreover, the lack of functional characterization of parasite genomes makes the interpretation of large datasets difficult. This is especially true when building system-wide networks based on gene ontology, as differentially expressed genes with unknown function are often excluded, which may lead to

an inaccurate representation of the system. Since approximately half of parasite genes fall into this “hypothetical” category, caution must be taken to express the degree of uncertainty when clustering datasets into biological processes. Accordingly, a current focus of parasite biology is to assign global gene function, and thus genome-wide technologies rooted in systems biology are essential.

4.4. Recent advances in systems-based approaches to host-parasite research

4.4.1. Application of –omic technologies

Systems biology utilizes multiple platforms in order to survey global cellular processes. These include the classic -omic technologies, namely transcriptomics, proteomics, and metabolomics. The recent whole-genome sequencing of many important human parasite genomes has led to significant progress in the development of these approaches to the study of parasitic disease. In this section, we will summarize these strategies and their application to host-parasite interactions.

Transcriptomics has been fundamental in shaping our current understanding of parasite infection biology. Probe-dependent cDNA microarrays have been a historically useful tool for gene validation and discovery, as well as determining differential transcript expression in parasites (Duncan, 2004; Gobert et al., 2005). In *T. gondii*, for example, microarray analysis led to the identification of developmentally regulated genes that clustered into distinct processes such as immune avoidance and sugar metabolism (Cleary et al., 2002). Microarray data has also revealed that *P. falciparum* genes in specific pathways are co-regulated (Young et al., 2005), and that many genes that are co-transcribed share common regulatory elements (Young et al., 2008). cDNA microarray chips for a number of

protozoan pathogens are now commercially available, and recently, the first *C. parvum* specific microarray (Zhang et al., 2012) was developed and made accessible to the research community. Microarrays have also provided global insights into the host response to parasitic infections (Blader et al., 2001; Liehl et al., 2014). Cross-hybridization severely limits the scope of probe-dependent techniques, however, as high background reduces the dynamic range of the assay and parasite and host transcriptomes must be analyzed separately (Westermann et al., 2012a). Probe-independent, tag-based methods such as serial or cap analysis of gene expression (SAGE or CAGE, respectively) can provide a more quantitative picture of the transcriptome, and have been particularly useful for gene expression analysis in eukaryotic pathogens, as reviewed elsewhere (Kronstad, 2006).

All of the aforementioned techniques are restricted by the inability to detect specific mRNA isoforms, unannotated non-coding RNAs, and precise splice junctions. Recent advances in next generation sequencing platforms have allowed for deep sequencing of RNA, known as RNA-Seq (Wang et al., 2009). This approach provides quantitative full-genome coverage, is extremely sensitive, and can identify RNA species and alternative splicing events that are undetectable by microarray or tag-based analyses. Recent advances in strand-specific RNA-Seq (Levin et al., 2010; Li et al., 2013; Siegel et al., 2014) have also revealed widespread transcription of natural antisense transcripts (NATs) in many eukaryotic parasites. So far, evidence for NATs

has been found in species such as *P. falciparum, T. gondii, T. brucei, Leishmania spp., and G. lamblia* (Militello et al., 2008). RNA-Seq also enables the simultaneous sequencing of both host and parasite transcriptomes (Westermann et al., 2012a), allowing for unprecedented insights into host-pathogen interactions. In a study by Pittman et. al. (Pittman et al., 2014), *in vivo* dual RNA-Seq analysis of *T. gondii* and its murine host revealed significant influence of both the host environment on parasite gene expression, as well as parasite development on host transcription. Simultaneous sequencing of both human and *P. falciparum* RNA isolated from peripheral blood from 116 malaria patients (Yamagishi et al., 2014) also provided important insights into host-parasite interactions, including the identification of host and pathogen genes that correlate with clinical disease severity. While this paired analysis provides a much more powerful approach than single RNA-Seq, the vast majority of –omic datasets currently survey either the host or the pathogen during infection. As dual RNA-Seq increases in both resolution and cost-effectiveness, it will no doubt continue to provide novel molecular insights into host-parasite transcriptomics.

Due to innovations in existing technologies and the development of new methodologies, the field of proteomics has made significant progress in surveying the complex repertoire of proteins that define host-parasite systems. Mass spectroscopy (MS) (Yates et al., 2009), which measures the mass-to-charge ratio and abundance of ions, has been by far the most widely used

method for proteomic analysis. Prior to the advent of genome sequencing, intact proteins had to be directly analyzed by MS through technically challenging and low-throughput 'top-down' procedures. The post-genomic era has significantly benefited from the implementation of 'bottom-up' approaches that instead utilize enzymatic or chemical fragmentation of proteins (Wastling et al., 2012). The protein sequence is then inferred by mapping of the MS fragmentation spectra to databases built from annotated genomic information. MS-based approaches have been fundamental in the assembly of the whole-cell proteomes of protozoan parasites during multiple life cycle stages, as reviewed elsewhere (Wastling et al., 2012). Despite the increasing number of proteomic datasets publicly available for these organisms, there remains a massive deficit in experimentally validated proteome coverage (i.e. the percentage of protein-coding genes that have evidence for protein expression) (Table 4.3). Moreover, this insufficiency is highly variable among parasites. While more well-studied protozoan parasites such as *P. falciparum, T. gondii, and T. brucei* have more than 50% proteomic coverage, others, such as *G. intestinalis* and *E. histolytica,* have less than 30% coverage, and for *L. major*, less than 5% of the predicted proteome is experimentally validated (Table 4.3). MS-based methods have also been instrumental in profiling the host proteome in response to parasitic infections. For example, Nelson et. al.(Nelson et al., 2008) utilized 2D electrophoresis, difference gel electrophoresis (DIGE), and MS to profile the host proteomic response to infection with *T. gondii*, and

analysis of the resulting dataset suggested extensive global reprogramming of host metabolic pathways. In a truly integrative study (Foth et al., 2011) of both parasite and host cell transcriptomic and proteomic data during the intra-erythrocytic developmental cycle of *P. falciparum*, 24 human proteins were identified in significant quantities within the parasite. Interestingly, these host proteins, like many parasite proteins found in the same study, displayed distinct abundance profiles throughout parasite development.

The current field of parasite proteomics is moving towards more sensitive and specific methods. While it is critical that we continue to map and annotate both host and parasite proteomes during infection, it is also important that we directly measure differential protein expression in order to better understand host-parasite biology. This general approach, commonly referred to as 'quantitative proteomics' (Cox and Mann, 2011), includes relative quantification methods such as isobaric tagging for relative and absolute quantification (iTRAQ) and stable isotope labeling by amino acids in cell culture (SILAC), and label-free methods such as spectral-counting. Quantitative proteomics has been particularly useful in mapping the phosphoproteomes of many parasites (Treeck et al., 2011; Tsigankov et al., 2013; Urbaniak et al., 2013), as well as phosphorylated host proteins in response to parasitic infection (Wu et al., 2009b). These studies have revealed that reversible protein phosphorylation, mediated by protein kinases and phosphatases, is an important regulator of many aspects of host-parasite

biology. In addition to quantitative proteomics, there has been an increasing interest in mapping the proteomes of subcellular organelles, called 'organellar proteomics' (Andersen and Mann, 2006). Organelle isolation prior to proteomic analysis is commonly achieved by cellular fractionation or specific labeling and purification methods. This type of proteomic analysis has enhanced our understanding of sub-cellular protein localization for many protozoan parasites, such as the nuclear proteome for *P. falciparum* (Oehring et al., 2012) and the mitochondrial outer membrane proteome for *T. brucei*(Niemann et al., 2013).

The systems-based application of metabolomics, or the global survey of small molecules (<1kD), has provided significant insight into the metabolic processes governing host-parasite infection biology over the last decade, and has been expertly reviewed elsewhere(Kafsack and Llinas, 2010). Since the majority of antiparasitic drugs target enzymes involved in parasite metabolism, mapping host-parasite metabolomes will be critical for the development of novel therapeutics. The study of parasite metabolism has historically relied on the use of low-throughput radioactive labeling or enzymatic-based assays. Today, high-throughput MS-based technologies as well as Nuclear Magnetic Resonance (NMR) spectroscopy are the major tools used by researchers investigating metabolomes (Bollard et al., 2005; Want et al., 2007). Pioneering studies have utilized these technologies to survey the metabolomes for many parasite life-cycle processes, including *Entamoeba*

cyst formation (Jeelani et al., 2012), *Leishmania* promastigote development (Silva et al., 2011), *Toxoplasma* tachyzoite replication (MacRae et al., 2012), and *Plasmodium* intraerthrocytic progression (Olszewski et al., 2009). Metabolic labeling coupled with MS is an effective strategy for measuring metabolic pathway flux. For example, Ke et. al.(Ke et al., 2015) utilized ^{13}-C labeling of *P. falciparum* genetic knockout lines that have deletions in mitochondiral tricarboxylic acid (TCA) cycle enzymes to show that mitochondrial metabolism is surprisingly flexible throughout the parasite life cycle. Additionally, profiling the host metabolome (Daubener et al., 2001; Itoe et al., 2014) has generated valuable information as to how parasites scavenge host resources and how the host alters its own metabolism to fight infection. Host metabolomic studies also have clinical importance, as this information has been utilized in the identification of diagnostic biomarkers of protozoan infections (Li et al., 2008; Ng Hublin et al., 2013; Wang et al., 2008). Albeit the application of metabolomics to the study of host-parasite interactions is relatively recent, significant progress has already been made towards understanding the dynamic metabolic networks that regulate parasitic infections.

4.4.2. Integrating and interpreting large datasets

Advances in systems-based technologies have required the development of mathematical methods and computational tools to integrate

and interpret multiple data types. This reliance will continue to grow as the size and number of datasets continue to exponentially increase. The computational approaches employed in systems biology span various mathematical disciplines. Here we focus on a few examples where the analysis tools have proven useful. Efforts in data integration can be broadly grouped into the following approaches: 1) data organization and network construction, 2) network analyses, and 3) simulation and modeling.

Given the size and number of large –omic screens, organization of the resulting datasets into databases is critical to facilitate subsequent integration and analysis. Publicly available databases for eukaryotic parasites, as summarized in Table 4.2, provide the information required to populate the 1-D annotation of the organisms, that is, the descriptive content summarizing each measured biological molecule. Examples of this include functional annotation or transcript levels for an individual gene. As exemplified in the summary plot of datasets uploaded to EuPathDB (Aurrecoechea et al., 2010) (Figure 2), there is generally much greater availability of transcriptomic data than proteomic data for protozoan pathogens. In turn, there is much greater availability of proteomic data than metabolomic data. This trend is in part due to technological advancements in the instrumentation and the degree of high-throughput profiling that is feasible for each type of measurement. Public metabolomic databases are also relatively scarce. Metabolights (Haug et al., 2013) is so far the only cross-species, open-access metabolomic database

available, and there is currently no information submitted for protozoan parasites. However, as the value of metabolomic data is recognized (Kafsack and Llinas, 2010), and the number of studies profiling host-parasite metabolomics increases, we anticipate an increase in open-access metabolomic databases. In contrast to the 1-D annotation of organisms, the 2-D annotation includes defining interactions between biological molecules (Palsson, 2004). For example, protein-protein interactions can be detailed as complexes or signaling pathways, and protein-metabolite interactions can be described as metabolic reactions occurring within the organism. The 2-D annotation provides a platform in which different measurements are integrated and subsequently analyzed in order to gain meaningful insight into the capabilities and functions of biological organisms (Palsson, 2004).

Once the components of a system are defined, they need to be merged into a format amenable to the desired analysis style. The type of network that is constructed is dependent on the experimental data available. For example, protein-protein interaction networks can be constructed from MS spectra generated after yeast 2-hybrid screens or MS spectra generated after co-immunoprecipitation (Young et al., 2008). Metabolic network connectivity can be determined by individual metabolites that are linked together via enzymatic (and transport) reactions. Groups of reactions subsequently form pathways. Once such information is collected, the metabolic reactions can be linked together into pathways (Feist et al., 2009). Although advancements in

network construction continue to be made with each new data set, we must also be cognizant that we may never completely characterize all components of an organism, nor comprehensively detail all interactions. Thus, analysis methods will need to be tolerant of the incomplete data sets (Date and Stoeckert, 2006).

Once data are organized and biological networks are constructed, many different methods can be deployed to leverage this information for valuable analyses. The connectivity of biological networks can be studied using methods from statistics and graph theory to systematically characterize relationships between different components (e.g. distance measures, connectivity, etc) in a network and how the different elements within a network are organized. For example, 2,846 protein-protein interactions were elucidated for 1,312 proteins in *P. falciparum* using a high-throughput yeast two-hybrid assay (Date and Stoeckert, 2006). When the resulting network was analyzed using gene co-expression data and ontology information, putative annotation for hypothetical proteins was feasible, and alternative biological functions were suggested for some annotated genes. Suthram et. al.(Suthram et al., 2005) further analyzed this network using a network alignment approach called PathBLAST (Kelley et al., 2004), an algorithm that identifies conserved pathways between organisms by identifying conserved proteins and then testing for conserved interactions. Through this, the authors found that few protein interactions were conserved between *Plasmodium* and several model

organisms, thus demonstrating that the patterns of protein interaction in *Plasmodium* are quite distinct.

Beyond the analysis of network organization, various modeling approaches are being employed to simulate host-pathogen interaction pathways across spatial and temporal scales. Different types of calculations can be performed, depending on the modeling approach that is used. Cell-scale modeling approaches are proving valuable for integrating and analyzing the increasingly large volumes of -omic data. One notable example is constraint-based modeling (Militello et al., 2008), which uses metabolic network reconstructions (Feist et al., 2009). In this approach, all metabolic reactions in an organism are linked together and represented in a specific mathematical format that enables the calculation of network characteristics as well as simulation of different metabolic network flux states. As illustrated in Figure 4.3, every gene in the network reconstruction includes gene-protein-reaction associations. These connections describe how transcripts are related to the proteins they encode, as well as their corresponding enzymatic reactions. Network reconstructions thus provide the foundation for the hierarchical data integration of biological models. Since the relationships between the components in these networks are defined by logical relationships, it is then relatively trivial to integrate multiple types of different datasets from the host and pathogen and analyze them simultaneously.

Constraint-based modeling has a growing number of methodologies (Lewis et al., 2012b) enabling one to make diverse predictions including metabolic pathway usage, gene essentiality, and potential drug targets. Analysis of network reconstructions for *Leishmania major* have demonstrated utility in predicting minimal requisite media conditions for growth (Chavali et al., 2008) and further for simulating lethality versus non-lethality responses to drug treatments. Network reconstructions contain the requisite pathways for biomass synthesis (i.e. growth), thus, *in silico* simulations can be carried out in a systematic fashion in which each gene is 'deleted' and the ability of the cell to grow can then be tested. Similarly, *in silico* experiments can be performed in which the response to a particular drug is tested. This is achieved by inhibiting or eliminating the activity of an enzymatic target of a drug within the network and then interrogating the capabilities of the *in silico* organism (e.g. testing the ability to generate biomass, carry flux through particular pathways, etc.) (Cleary et al., 2002; Haug et al., 2013; Kim et al., 2011). In a study by Chavali et al (Chavali et al., 2012), the authors elucidated potential combinatorial drug treatment regimens that would inhibit growth. Similar algorithms could be employed to study other parasites and complement drug-screening efforts. Constraint-based reconstructions and analyses have also provided insight into growth conditions and drug sensitivity for other important human parasites (Plata et al., 2010; Raghunathan et al., 2010; Roberts et al., 2009).

In addition to the analysis of individual parasites, models are being developed that include host cell pathways. Specifically, based on host and parasite genome annotation, computational models can be reconstructed for both the host and parasite. Then computational simulations can elucidate how the pathways of the two different organisms influence each other. For example, the host-pathogen model analysis carried out by Bordbar et al (Bordbar et al., 2010) was the first integrated, simulation-capable host-pathogen metabolic network reconstruction, in which two genome-scale network reconstructions (a human host cell and a pathogen) were functionally integrated. Host pathogen interactions in different infectious states were characterized through analysis of transcriptomic data, which revealed differences in flux states of the pathogen (*Mycobacterium tuberculosis*) in latent versus pulmonary versus meningeal tuberculosis. These differing metabolic states are the result of the different tissues as well as the different types of interactions between the pathogen and host cell. Further, such differences may suggest different treatment strategies, depending on the site of the infection. This methodology is likely to be critical in the analysis and interpretation of protozoan pathogen data. The development of host-parasite models represents a new avenue of application and much needed development. With the continued expansion and completion of 1-D annotation of human parasites (Table 4.2), the concurrent and increasingly active 2-D annotation of these pathogens will lead to an improved understanding of host-

parasite interactions, and in the process, yield meaningful predictions and new hypotheses to test and validate in the wet lab.

4.4.3. Systems biology to help mitigate the challenges associated with host-parasite research

In addition to expanding the arsenal of tools available to researchers, advancements in systems-biology based methods have helped address some of the most challenging obstacles associated with host-parasite research. As summarized earlier, some of these challenges include the complex, stochastic nature of parasitic lifecycles, the lack of effective methods for culturing certain parasite developmental stages, and the overwhelming percentage of uncharacterized parasite genes. In this section, we will highlight some of the most recent systems-biology based studies that have aimed to overcome these barriers.

Although a number of methods have been developed to synchronize or isolate specific developmental parasite stages *in vitro*, mixed parasite populations are often unavoidable *in vivo,* especially when analyzing patient samples. This is problematic when examining system-wide expression data, as different stages have been shown to display distinct expression profiles. Clinical surveillance of patients harboring transmissible parasite stages, such as the *Plasmodium* sexual gametocyte stage, is critical for transmission reduction (Alonso et al., 2011); however, this is difficult, as gametocytes

comprise only a very small fraction of all blood-stage *Plasmodium* parasites during infection. To address this issue for mixed populations of *P. falciparum*, researchers have developed a statistical method (Lemieux et al., 2009) that estimates the relative contributions of cell cycle and developmental stage variation to the overall stochasticity of gene expression data. The method was based on both transcriptomic and microscopy analysis, and when applied to a published dataset of *in vivo* patient samples, they found that the previously reported variation in gene expression was directly correlated with a changing proportion of sexual stage parasites. In addition, a recent study (Joice et al., 2013) utilized computational analysis of transcriptomic data in order to develop a novel qRT-PCR-based method that can estimate the amount of both asexual and sexual stages in patient samples. This strategy relied on the selection and validation of a small panel of developmentally associated transcriptional markers, a procedure deeply rooted in systems biology.

In vitro culture of parasites throughout their life cycle is a valuable technique to more easily study host-parasite biology. However, this is technically challenging for many pathogen developmental stages. A notable example is the stage conversion that occurs during *E. histolytica* encystation, or the development from pathogenic trophozoites into transmissible cysts, as this process cannot be currently reproduced *in vitro*. In *E. invadens,* a related *Entamoeba* species that infects reptiles, stage conversion can be induced *in vitro*. Several groups have recently mapped *E. invadens* encystation on a

system-wide scale in order to better understand the biological processes controlling cyst formation, and in the process they provide insight into the development of *in vitro* culture methods to induce *E. histolytica* encystation (Ehrenkaufer et al., 2013; Jeelani et al., 2012). These studies led to the sequencing and assembly of the *E. invadens* genome and global characterization of both transcriptomic and metabolomic changes during encystation. Interestingly, RNA-Seq analysis (Ehrenkaufer et al., 2013) revealed that phospholipase D, an enzyme involved in lipid second messenger signaling, is required for efficient *E. invadens* stage conversion *in vitro*. In addition, MS-based metabolomics (Jeelani et al., 2012) revealed that despite an overall decrease in energy generation, there is an increase in the levels of certain biogenic amines as well as γ-aminobutyric acid (GABA) during encystation. While it is still unclear how the biological processes revealed by these data specifically contribute to *Entamoeba* stage conversion, these studies provide important insight into pathways that may be targeted to induce *E. histolytica in vitro*. Additionally, the metabolic enzymes controlling these processes may be suitable targets for the development of transmission-blocking drugs.

As previously emphasized, the majority of protozoan parasite genomes are only half annotated, with around 50% of genes assigned a hypothetical or unknown function (Figure 4.1). Since our understanding of host-parasite interactions requires knowledge of both host and parasite gene function,

incomplete gene annotation significantly stifles progress in this field. This severely limits our basic understanding of parasite biology and stunts our progress towards improved antiparasitic therapies, as drug discovery research benefits from the functional annotation of parasite genes. There has been an increased effort in recent years to apply high-throughput phenotypic screening and chemical genomics to identify novel parasite drug targets; however, often the genes targeted by promising compounds are uncharacterized (Winzeler, 2006). Accordingly, significant effort has been made to apply systems biology-based methods in order to assign global gene function. Bioinformatic analysis using comparative genomics is a widely used strategy for predicting the function of uncharacterized proteins (Pellegrini et al., 1999). This method relies on the evolutionary conservation of proteins with similar function. While comparative genomics has been important in the assignment of putative gene function for many parasite genes (Rout et al., 2015; Silber and Pereira, 2012), this analysis alone is not sufficient for the characterization of whole parasite genomes. There are notable examples where structurally similar proteins have divergent functions, and likewise, where proteins that have similar functions have divergent sequences (Saghatelian and Cravatt, 2005). Additionally, there are a number of parasite proteins that do not have orthologs with known function, and therefore traditional comparative genomics would not be applicable. For example, *C. parvum* is particularly divergent, with only 4% of all the predicted open reading frames (ORFs) initially assigned putative

functions based on sequence homology (Abrahamsen et al., 2004). Classical systems-based profiling of parasite transcriptomes, proteomes, and metabolomes has helped to build biological context for a number of these uncharacterized proteins (Bozdech et al., 2003; Bringaud et al., 2015; Butler et al., 2014; Le Roch et al., 2003). However, experimental evidence linking genotype to phenotype is still required in order to adequately characterize protein function.

In model systems, protein functional characterization has been largely achieved by phenotypic screens of genetically manipulated organisms. This -omics strategy, called functional genomics, includes forward genetic approaches, which identify the genetic basis for phenotype, and reverse genetic approaches, which identify the phenotypic consequence of genetic alteration. Unfortunately, these functional genetic strategies are challenging for many important human parasite systems, as genome manipulation is technically difficult. Despite the challenges, a number of recent genome-wide screening strategies have been successfully executed. Transposon mutagenesis using the *piggyBac* transposable system has been particularly useful as a forward genetic strategy in *Plasmodium* species (Fonager et al., 2011; Ikadai et al., 2013). Additionally, improvements in forward genetic methods for chemical mutagenesis have facilitated functional genetics in organisms such as *T. gondii* (Farrell et al., 2014). Reverse genetic strategies, including both gain of function and loss of function genetic screens, have also

come a long way in recent years. Genome-wide overexpression screens have been a valuable platform for characterizing protozoan parasite gene function. These studies have identified genes involved in phosphatidylinositol signaling as well as phagocytosis for the human protozoan pathogen *E. histolytica* (King et al., 2012; Koushik et al., 2014). While overexpression screens are useful because they can provide biological insight while avoiding the problem of genetic redundancy, loss of function screens can directly assess the phenotypic consequence of repressing endogenous gene expression, which in many cases is physiologically preferable. Moreover, the ease of gene knockdown technologies such as RNA interference (RNAi) has facilitated high throughput screening. While a number of protozoan pathogens such as *P. falciparum* lack the cellular machinery necessary for RNAi, others, like *T. brucei*, have a functional RNAi pathway amenable to reverse genetics (Kolev et al., 2011). A number of recent genome-wide RNAi screens have been carried out in *T. brucei* (Jones et al., 2014; Monnerat et al., 2009; Mony et al., 2014), and these studies have led to the identification of many parasite genes controlling important biological processes such as cell cycle progression, differentiation, and quorum sensing. Alternative methods to regulate gene expression in the genetically intractable *Plasmodium* parasite species are highly sought after. Significant progress has been made to this end, with the recent development of reverse genetic technologies including a tetracycline-repressible transactivator system (Pino et al., 2012), a glmS ribozyme-based

post-transcriptional knockdown system (Prommana et al., 2013), and an inducible TetR-aptamer system (Goldfless et al., 2014). Very recently, a genome-scale library consisting of bar-coded genetic modification vectors was developed as a reverse genetic screening resource for *P. berghei* (Gomes et al., 2015). Application of the genome-editing CRISPR-Cas9 system to the study of malaria parasites has also been successful (Ghorbal et al., 2014; Wagner et al., 2014; Zhang et al., 2014), and as this technology further develops, it will certainly improve our understanding of many hypothetical parasite proteins and thus host-parasite biology as a whole.

4.5. Systems analysis has advanced our understanding of key aspects of host-parasite biology

An enormous amount of information has been produced from the generation of host-parasite systems-biology datasets within the last decade. The proper integration and interpretation of this 'big data' is critical in order to link experimental findings to useful biological knowledge. Due to the system-wide nature of these datasets, a vast number of important and interesting conclusions can usually be drawn from any given systems-based study, although researchers often choose to pursue only a limited number of noteworthy findings. Interestingly, many systems-based publications have followed up on data that enhance our understanding of specific sub-fields of host-parasite biology. Although this review will not attempt to encompass all of these findings, we will review some of the leading concepts arising from the analysis of recent genome-wide datasets.

4.5.1. Regulation of parasite gene expression

Understanding how parasites regulate gene expression throughout their lifecycle within a host is necessary in order to fully appreciate the scope of host-parasite interactions. For example, parasites can actively interfere with host cell translation in order to hijack the cellular resources required for their own gene expression as well as suppress immune responses, as reviewed elsewhere (Jaramillo et al., 2011; Mohr and Sonenberg, 2012). The system-

wide investigation of parasite gene expression is also vital in understanding the coordinated set of events underlying important host-parasite interactions. The up- or down-regulation of parasite proteins during specific developmental stages is dependent on the host cellular environment and needs to be carefully controlled to ensure parasite survival.

Genome-wide approaches have been particularly important in the elucidation of the regulatory mechanisms governing parasite gene expression in recent years. Although transcriptomics has emerged as a powerful systems-based approach, it must be emphasized that the quantitation of mRNA abundance is often an imperfect indicator of global gene expression. Indeed, for both prokaryotic and eukaryotic organisms, it has been demonstrated that mRNA levels correlate with protein expression for only 50-70% of genes[91], and for protozoan parasites, this number may be even lower (Lahav et al., 2011). Systems-based approaches have been especially useful for characterizing the dynamic control of parasite gene expression in recent years. These studies have revealed that precise control of gene expression is essential to drive the dramatic transformation that takes place as parasites cycle through developmental stages, and that the apparent lack of tight transcriptional regulation is remedied by extensive post-transcriptional mechanisms.

In particular, translational delay, a process in which protein expression is actively suspended for expressed mRNA transcripts, is a common strategy employed by many protozoan parasites. Translational delay may be a

particularly advantageous strategy for parasites, as they must quickly adapt to new environments and undergo developmental switching in order to survive; storing transcripts necessary for such adaptations allows for rapid changes in gene expression by circumventing the time needed for transcription. Genome-wide next generation sequencing of both steady state mRNA as well as polysome-associated transcripts during the asexual erythrocytic stage in *P. falciparum* (Bunnik et al., 2013) revealed widespread translational repression across the genome during different stages of the parasite lifecycle. Surprisingly, more than 30% of parasite genes were found to be associated with translational delay. Many of the repressed genes appeared to be regulated by cell cycle stage and they clustered into discrete biological processes. For example, many genes associated with early stage processes, such as nutrient acquisition and erythrocyte remodeling, were transcribed during the trophozoite or schizont stages, and were only actively translated immediately following merozoite invasion. Another genome-wide ribosomal profiling study of *P. falciparum* blood stages (Caro et al., 2014) provides additional support for a model whereby transcription of important merozoite genes occurs during the previous stage and are translationally upregulated during invasion. Translational delay has also been demonstrated during the sexual gametocyte stage by temporary storage of specific transcripts in P-bodies (Mair et al., 2006). Unlike the majority of other eukaryotic organisms, *Trypanosomes* transcribe almost all of their genes as large polycistronic

clusters, and thus lack transcriptional control for most genes. Despite the absence of regulation at the level of transcription, transcriptomic surveys (Jensen et al., 2009) have revealed extensive variation in mRNA abundance across developmental stages, suggesting widespread post-transcriptional control. Furthermore, the comparison of proteomic expression using SILAC and MS to transcriptomic datasets suggests that like *Plasmodium*, mRNA abundance does not predict protein expression for at least 30% of the *T. brucei* genome (Gunasekera et al., 2012). The integration of data surveying global protein expression, polysome-associated transcript abundance, and total mRNA during these stages revealed extensive translational repression during the time when *T. brucei* prepares for transmission (Capewell et al., 2013).

4.5.2. Parasite utilization of host resources

While eukaryotic pathogens are often able to synthesize a number of nutrients required for growth *de novo*, it is often more advantageous to conserve the energy required for biosynthesis and to instead hijack host-derived resources. This is especially true for the acquisition of host lipids, as protozoan parasites must quickly assemble a large amount of new membrane during replication within host cells. In Apicomplexan parasites such as *P. falciparum* and *T. gondii*, fatty acids are taken up from the host and converted into triacylglycerides, where they are then stored in lipid bodies (Mazumdar

and Striepen, 2007). Recently, a system-wide survey of the *Plasmodium* lipidome during liver stage infection (Itoe et al., 2014) revealed a significant enrichment in fatty acids important for membrane biogenesis, including phosphatidylcholine. Upon further investigation, it was found that the parasite actively acquires host-derived phospatidylcholine and that this process is essential for parasite survival within hepatocytes (Itoe et al., 2014). It has also been shown that *Leishmania* parasites, while unable to synthesize sphingomyelin themselves, are able to hydrolyze host sphingomyelin in order to produce essential metabolites (Zhang et al., 2009). A comparative genomics study (Zhang et al., 2009) identified a parasite enzyme, LaISCL, which is responsible for the degradation of host-derived sphingomyelin, and showed that this process is necessary for the proliferation of *L. major* parasites within their mammalian hosts. More recently, the same group showed(Pillai et al., 2012) that this enzyme is also responsible for sphingomyelin turnover in *L. amazonensis*, although in this species, the role of shingomyelin degradation in promoting virulence is quite different.

Many protozoan parasites live an intracellular auxotrophic lifestyle, actively acquiring metabolites from their nutrient-rich host in order to survive. For instance, blood stage *Plasmodium* parasites have lost the ability to biosynthesize purine rings or amino acids, and therefore scavenge host nucleotides to synthesize DNA and catabolize host hemoglobin to generate amino acids (Gardner et al., 2002; Olszewski et al., 2009). Recent system-

wide metabolomic studies have been instrumental in profiling the complex exchange of nutrients between parasites and their hosts. A comprehensive MS-based approach (Olszewski et al., 2009) revealed significant modulation of host metabolites during blood stage *Plasmodium* development. The authors found that host arginine depletion was particularly extensive, suggesting that this may contribute to human malarial hypoargininemia and progression to cerebral malaria. Another Apicocomplexan parasite, *T. gondii*, relies on host nutrients, such as carbon, in order to proliferate within host cell vacuoles. In a combined metabolomic and stable isotope labeling approach, a recent study (MacRae et al., 2012) mapped the carbon metabolism pathway for *T. gondii* tachyzoites. This systems-based analysis revealed that active catabolism of host glucose and glutamine through an oxidative tricarboxylic acid (TCA) cycle is essential for parasite replication. Through these and similar systems-biology based surveys, it is becoming clear that protozoan parasites have evolved complex strategies to both usurp and exploit host resources.

4.5.3. Host immune response to parasitic infection

In order to fully appreciate the complexity of host-parasite interactions, the host immune response must be considered. It is well established that while most protozoan infections are self-limiting in immunocompetent hosts, however, immunocompromised individuals can develop severe and often life-threatening disease, suggesting that an effective immune response is

essential for regulating parasitic disease. Many -omic-based strategies have contributed to our current knowledge of how the innate and adaptive immune systems resist parasitic infection, and in many cases, exacerbate disease. In particular, recent transcriptomic analyses of host-parasite systems have implicated the host innate Toll-like receptor (TLR) and interferon (IFN) -mediated proinflammatory pathways in the regulation of disease progression. Microarray analysis of malaria patient samples (Franklin et al., 2009) demonstrated an up-regulation of TLR signaling genes that had sites for IFN-inducible transcription factors. Upon subsequent analysis of *Plasmodium*-infected rodents (Franklin et al., 2009), it was revealed that TLR9 and MyD88 are critical to initiate the cytokine responses leading to acute malaria *in vivo*. Another transcriptomic analysis of patient responses (Sharma et al., 2011) further confirmed the enhancement of IFN-stimulated genes (ISGs) upon infection with malaria parasites, and interestingly, the same study determined that TLR9-independent sensing of AT-rich *Plasmodium* DNA induces type-I IFNs. In a dual RNA-Seq approach (Pittman et al., 2014), a recent report mapped host and pathogen transcriptomes during acute and chronic infection with *T. gondii*. Analysis of the differentially expressed transcripts revealed that many of the acute infection specific-genes included ISGs such as guanylate binding proteins. Chronic infection specific-transcripts were shown to comprise a unique set of immune genes, including those important for antigen recognition and presentation. Thus, these systems-level analyses

indicate that innate sensing of protozoan pathogens is important for the induction of proinflammatory responses aimed at controlling infection.

Parasitic disease is an evolutionary arms race; as our immune systems attempt to fight off infection, pathogens quickly respond by adapting to and subverting these attacks, often through elegant biological maneuvers. Multiple –omic-based surveys have contributed to our knowledge of how protozoan parasites actively manipulate the host immune response in order to avoid detection. Over a decade of systems-biology research has shown that *T. gondii* downregulates the innate immune response by multiple mechanisms. This includes preventing host nuclear translocation of proinflammatory transcription factors such as nuclear factor kappa β (NF-κβ) and signal transducer and activator of transcription 1 (STAT1α), as well as upregulating anti-inflammatory pathways such as those involving the suppressor of cytokine signaling (SOCS) proteins (Hunter and Sibley, 2012). A notable systems-based study (Saeij et al., 2007) utilized transcriptomics and pathway analysis to show that *Toxoplasma* actively regulates host immune responses, and through forward genetics, discovered a parasite rhoptry kinase, ROP16, that is secreted into the host cytoplasm to interfere with STAT signaling. Additionally, *Plasmodium* parasites also secrete virulence factors that specifically block host innate immune signaling. During liver stage development, *Plasmodium* circumsporozoite protein (CSP) is exported and localized to the host cell nucleus where it interferes with the nuclear translocation of NF-κβ, and

microarray analysis confirms that at least 40 NF-κβ responsive genes are downregulated with CSP expression (Singh et al., 2007). Likewise, in the blood stages of the parasite, a high-throughput protein interaction screen (Waisberg et al., 2012) found that *Plasmodium* merozoite surface protein 1 (MSP1) specifically binds to the human proinflammatory cytokine S100P, and that this interaction blocked activation of the host NF-κβ -mediated innate immune response. Through these and other genome-wide investigations, it is clear that while the host innate immune system is essential in controlling parasitic infection, parasites have evolved complex strategies to effectively dampen these responses.

4.6. Conclusions

Parasitic disease research has significantly benefited from systems analyses. Host-parasite systems are complex, with stochasticity across and within developmental stages, are often technically challenging to model experimentally, and are built upon incompletely characterized genomic foundations. Despite the challenges, recent improvements in systems-level technologies have facilitated the generation of 'big data' to model host-pathogen interactions. These analyses have improved our current knowledge of the basic biology driving parasitic infection, and have also yielded novel tools to facilitate further research. Many -omic surveys have been conducted, and global expression data for important human protozoan parasites are now publically accessible through several pathogen databases. Furthermore, algorithms for the integration and interpretation of genomic, transcriptomic, and metabolomic data have elucidated novel insights and hypotheses into host-parasite interactions. In particular, systems biology approaches have shed light on how parasites utilize post-transcriptional gene regulation to quickly adapt to changing host environments, hijack host-derived resources to establish intracellular replication, and neutralize host immune responses to escape host proinflammatory attacks.

Each new -omics survey comes with the promise to 'solve biology' and serve as a singular framework for biological understanding. Inevitably, this

fails, not because of overestimation of the utility of a particular measurement, but rather, the failure to recognize the need for multiple data types and for analysis to be carried out in an integrated, cohesive manner. Significant insights into host-parasite biology have been made with systems biology, but technical challenges still limit the application of systems approaches to parasite systems, leading to an uneven distribution of genome-wide datasets across protozoan species and developmental stages. While a number of genomic and transcriptomic datasets have been generated for these pathogens, functional annotation is still absent for approximately half of all parasite genes, and proteomic coverage is severely lacking. Moreover, there is a complete absence of publically accessible metabolomic databases for protozoan pathogens.

Moving forward, the field of host-parasite biology would greatly benefit from overcoming key deficiencies in systems biology research. Necessary advances include the optimization of parasite culturing methods, the development of functional genetic approaches (e.g., through the CRISPR-Cas9 system), and computational models of host-parasite interactions. These tools will enable the generation of more genome-wide datasets for functional characterization of parasite genes and provide tools for the analysis of these data. Thus, there are many opportunities for researchers to leverage systems biology to further a field that is far from saturated. There is no doubt that the increasing efficacy of systems-based approaches will continue to improve our

current understanding of host-parasite interactions, and accordingly, the treatment of parasitic disease.

4.7. Acknowledgments

We would like to thank members of the Winzeler lab for critical review and discussion of the manuscript. This work was completed with generous support from the National Institutes for Health, the Bill and Melinda Gates Foundation, The Medicines for Malaria Venture, and the Novo Nordisk Foundation provided to the Center for Biosustainability at the Technical University of Denmark.

Chapter 4, in full, has been published in Wiley Interdisciplinary Reviews: Systems Biology and Medicine, 2015. Justine Swann, Neema Jamshidi, Nathan Lewis, Elizabeth Winzeler. "Systems analysis of host-parasite interactions". The dissertation author was the primary investigator and author of this paper.

4.8. Figures

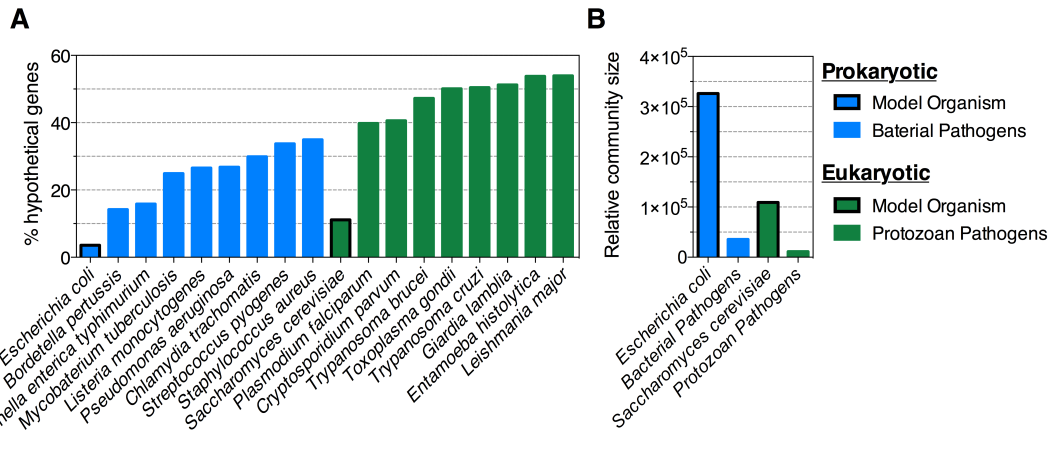


Figure 4.1. Percentage of 'hypothetical' genes and relative community size for important unicellular human pathogens and their model organisms. a) The percentage of 'hypothetical' genes for selected prokaryotic and eukaryotic pathogens compared to their relevant model organism, *E. coli* and *S. cerevisiae*, respectively. Percentages for each species were calculated from the number of genes including 'hypothetical', 'unknown', or 'uncharacterized' in the gene description compared to the total number of pathogen genes from the NCBI database for model organisms and bacterial pathogens, and from the corresponding EuPathDB databases for protozoan pathogens. b) The relative community size for model organisms, and the mean relative community size for the bacterial and protozoan pathogens listed in A, based on the number of results generated from a Pubmed (http://www.ncbi.nlm.nih.gov/pubmed) search of the species name.

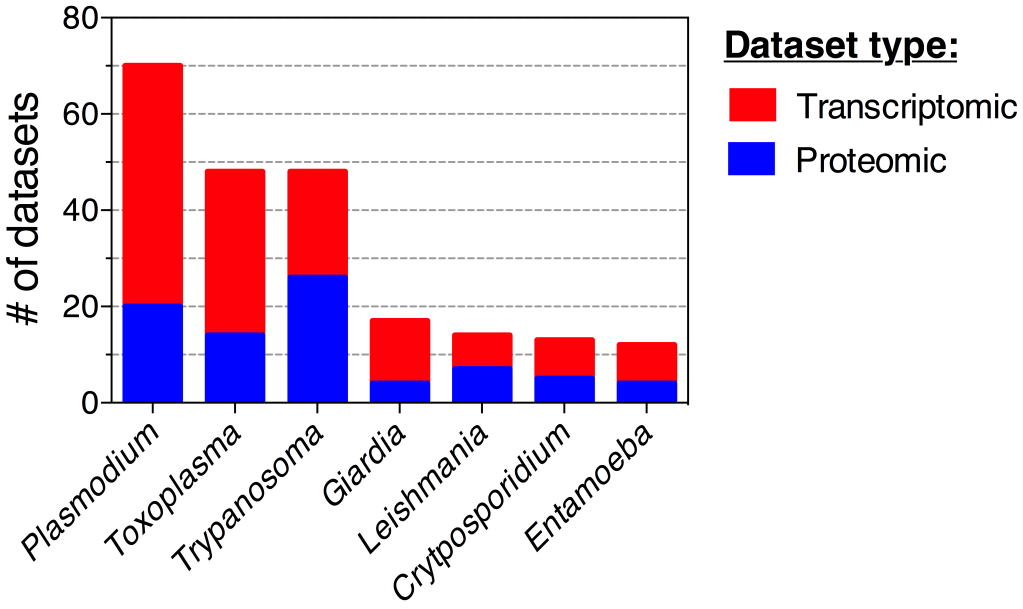


Figure 4.2. Distribution of transcriptomic and proteomic datasets uploaded to EuPathDB for selected Protozoan parasites. The number of transcriptomic and proteomic datasets submitted to the EuPathDB(Aurrecoechea et al., 2010) family of databases (see Table 2) for each Protozoan parasite genus. The total number of datasets is plotted for each parasite group, with the proportion of transcriptomic datasets (colored in red) and proteomic datasets (colored in blue) displayed within each bar graph.

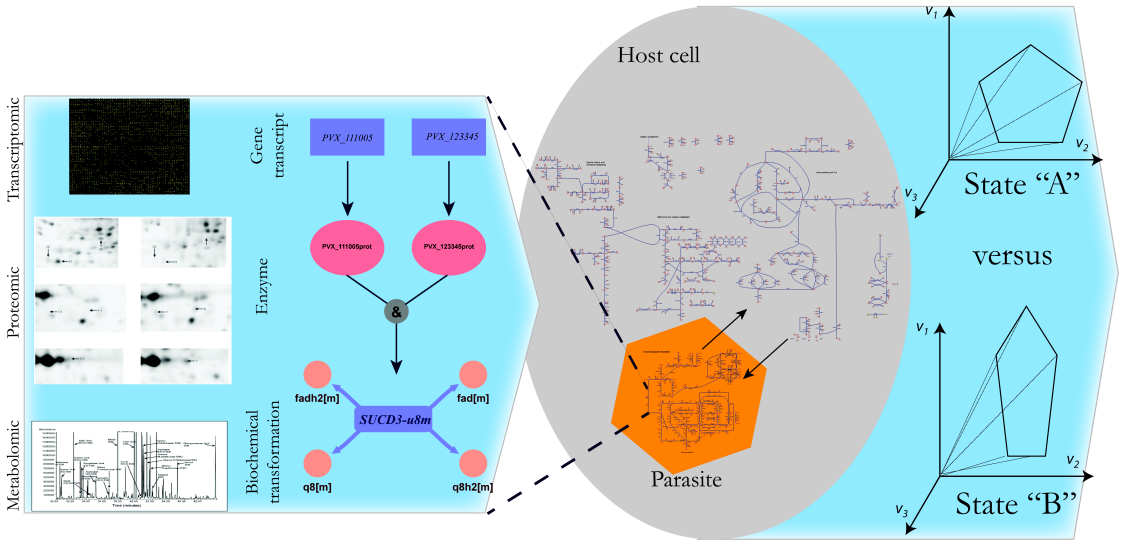


Figure 4.3. Organization, integration, and analysis of -omic datasets in metabolic network reconstructions used in constraint-based modelling. Moving from left to right in the figure, various -omic data types (transcriptomic, proteomic, metabolomic) are mapped onto the different components of the model. This includes the genes, enzymes, or small metabolites within a network for every reaction in the reconstruction (for which such data is available). Host-pathogen models can be constructed by connecting (or infecting) a host cell with the intracellular parasite. Subsequent simulations may characterize differences in the flux states in the non-infected versus infected state of the host cell. A gene-protein-reaction relationship for *Plasmodium* succinate dehydrogenase is highlighted in the figure.

4.9. Tables

Table 4.1. Protozoan parasites that cause human disease.

	Species	Disease	Host(s)	Human Tissue Tropism	Parasite Developmental Stages
Apicocomplexans	*Toxoplasma gondii*	Toxoplasmosis	Domestic cats and humans	Intestine, muscle, neural tissue	Oocysts, tachyzoites, tissue cysts
	Plasmodium spp.	Malaria	Infected female *Anopheles* mosquitos and humans	Hepatocytes, erythrocytes, central nervous system	Sporozoites, liver stages (trophozoites, shizonts, merozoites, hypnozoites in some species), blood stages (erythrocyte ring stages, mature trophozoites, shizonts, merozoites), gametocytes, mosquito stages (zygotes, ookinetes, oocysts)
	Cryptosporidium spp.	Cryptosporidiosis	Humans	Epithelial cells of gastrointestinal or respiratory tract	Oocysts, sporozoites, trophozoites meronts, merozoites, gamonts, micro- and macro- gamonts, zygotes
Kinetoplastids	*Trypanosoma brucei*	African sleeping sickness	Tsetse fly and humans	Bloodstream, lymphatic system, central nervous system	Metacyclic trypomastigotes, bloodstream trypomastigotes, procyclic trypomastigotes, epimastigotes
	Trypanosoma cruzi	Chagas disease	Triatomine bug and humans	A variety of cell types near the site(s) of infection, bloodstream	Metacyclic trypomastigotes, intracellular amastigotes, bloodstream trypomastigotes, epimastigotes
	Leishmania spp.	Leishmaniasis	Sandflies and humans	Mononuclear phagocytes in various tissues	Promastigotes, amastigotes
Diplomonads	*Giardia lamblia*	Giardiasis	Humans	Small intestine, proximal small bowel, colon	Cysts, trophozoites
Amoebozoa	*Entamoeba histolytica*	Amoebic dysentery	Humans	Small intestine, large intestine, liver, brain, lungs	Cysts, trophozoites

Information gathered from Centers for Disease Control (CDC), www.cdc.gov

Table 4.2. Resources and databases for Protozoan parasites

General or parasite species(s)	Database	Description	Web address
General Databases	PHI-based: Pathogen-Host Interactions(Ke et al., 2015)	Expertly curated database of experimentally verified genes from pathogens.	http://www.phi-base.org/
	Pathogen Portal	Integrative repository linking the NIAID Bioinformatics Resource Centers (BRCs) and providing -omics data for eukaryotic pathogens, all bacteria, and all viral families.	http://www.pathogenportal.org/portal/portal/PathPort/Home
	ProtozoaDB(Bozdech et al., 2003)	Gene-based protozoan database with emphasis on distant similarities (HMM-based) and phylogeny-based annotations, including orthology analysis.	http://protozoadb.biowebdb.org/
	HPIDB: Host-Pathogen Interaction Database (Wagner et al., 2014)	Host-pathogen database integrating experimentally derived protein-protein interaction data from various public databases; BLASTP enabled.	http://www.agbase.msstate.edu/hpi/main.html
	PRIDE Archive Proteomics Data Repository(Zhang et al., 2014)	European Bioinformatics Institute repository of mass spectrometry proteomics data.	http://www.ebi.ac.uk/pride/archive/
	EuPathDB (Eukaryotic Pathogen Database Resources) (Mohr and Sonenberg, 2012)	Integrative database of Eukaryotic pathogens housing sequencing data, microarray data, proteomics data, metabolic pathways, and phenotype information.	http://eupathdb.org
	OMIC tools(Jaramillo et al., 2011)	Metadatabase providing a compendium of over 4,400 web-based tools for the analysis of genomic, transcriptomic, proteomic, and metabolomic data.	http://omictools.com/
Crypotosporidium spp.	*CryptoDB(Gardner et al., 2002)*	Part of the EuPathDB family of databases	http://cryptodb.org/cryptodb/

Table 4.2. Resources and databases for Protozoan parasites (continued)

General or parasite species(s)	Database	Description	Web address
Entamoeba histolytica	*Entamoeba histolytica* Assembly and Annotation(Aurrecoechea et al., 2010)	Assembly and annotation of *E. histolytica* with content imported from AmoebaDB	http://protists.ensembl.org/Entamoeba_histolytica/Info/Annotation/
	AmoebaDB(Aurrecoechea et al., 2011)	Part of the EuPathDB family of databases	http://amoebadb.org/amoeba/
Giardia spp.	*GiardiaDB(Aurrecoechea et al., 2009a)*	Part of the EuPathDB family of databases	http://giardiadb.org/giardiadb/
***Leishmania* spp.**	Leishmania major - LeischCyc	Pathway/genome database for Leishmania major based on the BioCyc ontology.	http://biocyc.org/LEISH/organism-summary?object=LEISH
	TriTrypDB(Aslett et al., 2010)	Part of the EuPathDB family of databases, resource for Kinetoplastid species (including *Leishmania* spp.)	http://tritrypdb.org/tritrypdb/
***Plasmodium* spp.**	*PlasmoDB(Aurrecoechea et al., 2009b)*	Part of the EuPathDB family of databases	http://plasmodb.org/plasmo/
	Full-Malaria	Full-length cDNA database of Plasmodium species with web-accessible analysis tool	http://fullmal.hgc.jp/index_ajax.html
Toxoplasma gondii	*ToxoDB(Gajria et al., 2008)*	Part of the EuPathDB family of databases	http://toxodb.org/toxo/
	Trypanosoma brucei on GeneDB(Logan-Klumpler et al., 2012)	Genomic and proteomic database resources for *T. brucei* that is part of the Sanger Institute Pathogen Program, GeneDB project.	http://www.genedb.org/Homepage/Tbruceibrucei927
***Trypanosoma* spp.**	TrypanoCyc(Shameer et al., 2015)	Pathway/genome database for *T. brucei* based on the BioCyc ontology.	http://www.metexplore.fr/trypanocyc/
	TriTrypDB(Aslett et al., 2010)	Part of the EuPathDB family of databases, resource for Kinetoplastid species (including *Trypanosoma* spp.)	http://tritrypdb.org/tritrypdb/

Table 4.3. Experimental proteome coverage for Protozoan parasites

Species	Strain	Total genes	Protein coding genes	Proteomic expression	Proteome coverage
Toxoplasma gondii	GT1	8637	8460	4488	53%
Plasmodium falciparum	3D7	5777	5542	4104	74%
Cryptosporidium parvum	Iowa II	3886	3805	1320	35%
Trypanosoma brucei	TREU927	12094	11567	6632	57%
Trypanosoma cruzi	CL-Brener Esmeraldo-like	10597	10339	3674	36%
Leishmania major	Friedlin	9378	8400	329	4%
Giardia lamblia	Assemblage A Isolate WB	9747	9667	2166	22%
Entamoeba histolytica	HM-1:IMSS	8333	8306	2443	29%

Information gathered from EuPathDB databases, http://eupathdb.org

5. Conclusion

Antimicrobial drug resistance is serious problem that threatens human health on a global scale. The highly mutable nature of microbes coupled with the overuse of antibiotics contributes to the spread of drug resistant pathogens worldwide. Moreover, the lack of interest by the pharmaceutical industry to develop new drugs further complicates the problem. Drug resistance is especially problematic when treating HIV/AIDS, malaria, and tuberculosis. These infectious diseases are often called the 'Big Three' because they lead to the highest mortality rates each year. Drug resistant strains of each of these deadly pathogens have been reported to defy standard treatment regimens. Therefore, new research is needed in order to develop novel therapeutics. In this dissertation, three strategies are described that aim to combat drug resistance: 1. targeting of host factors (discussed in Chapter 2), 2. identification of new antimicrobials (discussed in Chapter 3), and 3. understanding systems-level host-pathogen interactions (discussed in Chapter 4 and in the Appendix). The combination and integration of these research approaches will surely drive the generation of new treatments less prone to the development of antimicrobial drug resistance.

Appendix. Systems biology of *Plasmodium berghei* liver-stage development

A.1. Introduction

Malaria is responsible for over a half million deaths per year, the vast majority among children in Africa (Carter and Mendis, 2002). Furthermore, malaria is a significant contributor to poverty in areas where the disease is endemic. The socioeconomic burden of malaria limits the resources necessary for disease prevention. Thus, while eradication campaigns have been successful in developed regions such as North America and Europe, malaria continues to devastate many third world countries (Sinka et al., 2012).

Malaria is the vector-borne disease caused by infection with *Plasmodium* parasites transmitted through the bite of *Anopheles* mosquitos. *Plasmodium* sporozoites are injected into the skin when an infected mosquito takes a bloodmeal and travel through the bloodstream to reach the liver. The sporozoites traverse multiple cells within the liver before establishing productive invasion within hepatocytes, where they transform into liver-stage forms (Prudencio et al., 2006b). Fully developed liver-stage merozoites within schizonts eventually exit the liver and re-enter the bloodstream. Alternatively, sporozoites may invade hepatocytes and develop into hypnozoites, a dormant parasite form that become activated months to years after initial infection and

are associated with malaria relapse. These forms are specific to certain species of *Plasmodium* such as *P. vivax* and *P. ovale*, and represent a major hurdle for malaria elimination (Wells et al., 2010b). Once in the bloodstream, the continuous replication of asexual blood stages within red blood cells causes their destruction and leads to the fever and chills associated with malarial disease (Clark et al., 1997b). A small percentage of these parasites enter a sexually reproductive cycle as female and male gametocytes, and the transmission of this stage back to the mosquito vector completes the life cycle (Baker, 2010a).

Systems biology is a powerful approach to studying cellular processes on a global scale. Recent advances in 'omic-' methods such as genomics, transcriptomics, proteomics, and metabolomics, have resulted in the generation of large datasets of biological information. The analysis and integration of this 'big data' will be important for understanding biological processes on a cellular and organismal scale. This type of analysis is especially suited for the investigation of parasitic disease biology, as –omic data from both host and pathogen can be incorporated to reveal a complex set of host-pathogen interactions (Swann et al., 2015).

A.2. Dual RNA-sequencing of host and pathogen transcriptomes during *Plasmodium berghei* liver-stage development

Dual RNA-sequencing of host and pathogen is a powerful systems-based method to uncover novel host-pathogen interactions, and is only recently being utilized by researchers (Westermann et al., 2012b). Unlike traditional transcriptomic approaches that analyze RNA reads separately from host and pathogen, a dual-approach maps the mixed reads to each annotated genome within samples of pathogen-infected host cells. This is a powerful method, as host and pathogen transcriptomes can be analyzed simultaneously.

We have taken advantage of this dual-RNA sequencing approach in order to gain insight into *Plasmodium* liver stage development within host hepatocytes. Briefly, Huh7.5.1 hepatocytes were infected *in vitro* with *P. berghei* sporozoites expressing a GFP reporter freshly dissected from infected *Anopheles stephansi* mosquitos, and cells were collected throughout liver-stage development using flow cytometry. This included samples collected at time zero (uninfected hepatocytes and sporozoites before infection), time 24 hours post infection (when the sporozoites have transformed into trophozoites), and time 48-50 hours post infection (when the trophozoites have transformed into liver-stage schizonts). An overview of the experimental

timeline illustrating the collection of samples for RNA-seq analysis is described in Figure A.9.1A. Infected cells were isolated from uninfected cells at each time point using fluorescence-activated cell sorting (FACS) (Figure A.9.1B). Total RNA was isolated from each sample, RNA concentration and quality was assessed, and dual RNA-sequencing was performed (experimental outline illustrated in Figure A.9.2).

Read mapping and differential gene expression analysis were then performed (Supplemental Dataset 4 and Supplemental Dataset 5, respectively). Briefly, total reads were filtered and aligned to the *P. berghei* reference genome (PlasmoDB-13.0 PbANKA) or the human reference genome (human reference genome GRch38) (Figure A.9.2). Principal component analysis (PCA) was performed to assess the variability between samples, and we found the samples to cluster together as expected (Figure A.9.3). *P. berghei* replicates clustered more closely than the Huh7.5.1 replicates, suggesting that the host response may be more variable than parasite gene expression between experimental replicates (Figure A.9.3). Differential analysis of host genes was performed between reads mapping to the human genome of uninfected and infected cells samples at 24 hours and 48 hours post infection (Figure A.9.4A). This pairwise comparison was ideal as it eliminated background host responses generated by exposure to mosquito salivary gland material. DEG was also performed for reads mapping to *P. berghei* between life cycle stages (sporozoite to trophozoite, 24 hour post

infection analysis; and sporozoite to schizont, 48-50 hours post infection analysis) in order to elucidate transcriptomic changes during development (Figure A.9.4B).

In total, 815 and 2,327 host DEG were identified at 24 hours and 48 hours post infection, respectively, with an adjusted p value (padj) <0.01 and log2 transformed fold change (log2FC) >1 (corresponding to upregulated genes) or <-1 (corresponding to downregulated genes) (Figure A.9.5A). At 24 hours post infection, 296 genes were downregulated and 518 genes were upregulated, and at 48 hours post infection, 696 genes were downregulated, and 1,631 genes were upregulated. In general, there were far more statistically significant *P. berghei* DEG at each time point, 3,207 total DEG at 24 hours post infection, and 3788 DEG at 48 hours post infection (Figure A.9.5B). 1,467 genes were downregulated and 1,740 genes upregulated at 24 hours post infection, and 1,672 genes were downregulated and 2,116 genes upregulated at 48 hours post infection. DEG were filtered based on padj <0.001 and log2FC >1 (corresponding to upregulated genes) or <-1 (corresponding to downregulated genes). In general, there were higher numbers of DEG at 48 hours post infection compared to 24 hours post infection for both host and parasite samples, and more genes were upregulated than downregulated. The full dual RNA-seq DEG results are included in Supplemental Dataset 4.

A.3. Validation of dual RNA-sequencing data

In order to validate the dual RNA-seq datasets for both host and parasite transcriptional responses, DEG were analyzed for differentially regulated pathways and processes and compared to the *Plasmodium* liver-stage infection literature. To elucidate host pathways that are differentially regulated upon infection with *P. berghei,* the lists of host DEG at 24 hours and 48 hours post infection (described above) were analyzed using Ingenuity Pathway Analysis (IPA) software (Figure A.9.6). Through this analysis, host pathways previously described as differentially regulated by *Plasmodium* liver-stage infection were identified. This includes mTOR signaling, which was one of the top canonical pathways identified at both 24 hours post infection and 48 hours. The proliferative phosphorylated states of mTOR have been previously shown to be upregulated during infection using a protein lysate microarray, and it has been suggested that this may lead to cellular protection from autophagy (Kaushansky et al., 2013). In the same study, it was shown that *Plasmodium* liver-stage infection induces a host cellular response that is anti-apoptotic (Kaushansky et al., 2013). Accordingly, many anti-apoptotic pathways, including many cancer-signaling pathways, were revealed as differentially regulated in our RNA-seq dataset (Figure A.9.6). Our analysis identified other pathways, such as the acute stress response and EIF signaling

(Figure A.9.6), which have also been reported as differentially regulated by *Plasmodium* liver-stage infection in the literature (Inacio et al., 2015).

Similarly, we analyzed the corresponding parasite DEG at 24 and 48 hour post infection for enrichment corresponding to specific biological processes. To do this, *P. berghei* DEG were analyzed using the gene ontology enrichment application in the *Plasmodium*-specific database PlasmoDB (www.plasmodb.org) (Aurrecoechea et al., 2009b). Tables A.10.1-A.10.4 include the results from the gene onology enrichment analysis for the top parasite DEG at each time point. For the analysis of the 48 hour time points, more stringent cutoffs (log2FC>2.5 and log2FC>4 for downregulated and upregulated DEG, respectively) were used in order to identify enriched processes, as the large number of genes in the DEG lists (1,672 genes downregulated and 2.116 genes upregulated) prevented the statistical enrichment of biological processes. Overall, the analysis yielded results that correspond to up- and down-regulated biological processes that are expected from the literature of *Plasmodium* liver-stage infection. For example, the biological processes corresponding to locomotion and movement within host environments are downregulated at 24 hours post infection, while metabolic processes are upregulated. This is expected, as this time point corresponds to the transformation of sporozoites into trophozoites, and it is known that parasites actively downregulate genes that are needed for motility once they have productively invaded hepatocytes, and instead upregulate genes needed

to increase the cellular metabolism needed for replication (Prudencio et al., 2006b). As further validation of our dataset, we found that genes such TRAP, CSP, UIS3, and UIS4 (well characterized *P. berghei* genes involved in sporozoite traversal and invasion) were significantly downregulated after transformation from the sporozoite to liver-stage forms (Supplemental dataset 4). Likewise, at 48 hours post infection, our results indicate that parasites downregulate genes involved in movement and interaction with the host, and instead upreguate genes involved in intracellular transport and localization. This also makes sense, as liver-stage merozoites within schizonts at 48 hours post infection need to transport and localize newly synthesized cellular components such as DNA, membrane, and organelles, as they undergo asexual replication within hepatocytes (Prudencio et al., 2006b). Accordingly, genes involved in DNA replication, metabolism, and merozoite infectivity are significantly upregulated at 48 hours post infection (Supplemental dataset 4).

A.4. Utilization of dual RNA-sequencing data to build a host-pathogen metabolic network for *Plasmodium* liver-stage development

Dual RNA-sequencing generates powerful datasets that can then be used to investigate host-pathogen interactions using *in silico* modeling approaches, such as constraint-based modeling. Constraint-based modeling has a growing number of methodologies enabling one to make diverse predictions including metabolic pathway usage, gene essentiality, and potential drug targets (Lewis et al., 2012a). In addition to the analysis of individual parasites, models are being developed that include host cell pathways. Specifically, based on host and parasite genome annotation, computational models can be reconstructed for both the host and parasite (Swann et al., 2015). Then computational simulations can elucidate how the pathways of the two different organisms influence each other. Future studies aim to utilize our recently generated dual RNA-seq data to build a host-pathogen metabolic network model of *P. berghei* liver-stage development.

A.5. Generation of a host biomarker reporter cell line to label hepatocytes infected with *Plasmodium vivax*

Plasmodium vivax is the most widespread human malaria species, endemic in many areas of South America, Asia, and Oceania (Gething et al., 2012). While not as lethal as *P. falciparum*, this species causes significant human morbidity and socioeconomic decline for millions of people worldwide. In addition, it is one of only two species of *Plasmodium* that can develop into the dormant liver-stage parasites called hypnozoites, which act as a major barrier to malaria elimination and the development of a radical cure (Wells et al., 2010b). Despite its important impact on human health, *P. vivax* is one of the least understood malaria parasites as it is extremely difficult to study in the lab. It cannot be maintained in cell culture, and current tools and resources to study this species are expensive, time-consuming, and globally-restricted (Wells et al., 2010b). To overcome these challenges, we are working to develop a host biomarker transcriptional reporter hepatocyte line in order to label cells productively infected with liver-stage *P. vivax in vitro* (Figure A.9.7). In this way, we will be able to easily assess and quantitate liver-stage infection with *P. vivax*, which at this moment, is very difficult with the tools available.

To do this, we utilized the gene expression dataset generated from the dual-RNA sequencing of *P. berghei*-infected hepatocytes (Supplemental

Dataset 4) and looked for candidate host DEGs that are significantly upregulated in response to liver-stage infection. From this data, we have identified a candidate host biomarker, Muc13, that is the most significantly overexpressed in *P. berghei* infected hepatocytes 48 hours post infection compared to uninfected cells (Figure A.9.7). Interestingly confocal microscopy analysis revealed apparent co-localization of host Muc13 and *P. vivax* UIS4 (Figure A.9.8). While these results need to be further validated and investigated, they suggest that host Muc13 may be a candidate host biomarker of *Plasmodium* liver-stage infection and may play an interesting role during the development of *Plasmodium* liver-stages.

In order to generate a host biomarker cell line based on these results, we have cloned the Muc13 promoter into the transcriptional response element (TRE) of a lentiviral vector, pGreenFire Lenti-Reporter (System Biosciences), expressing both GFP and luciferase reporters. Lentivirus was made by transfection of 293T cells with the lentiviral vector and pPACK lentiviral packaging vectors (System Biosciences). Titers were determined by transduction in 293T cells by qPCR against universal lentiviral integration sequences.

Future experiments will include the transduction of hepatocyte cell lines at an MOI of <0.5 to generate single integration events. Stable cell lines with successful integration of the lentivirus will be generated using puromycin

selection. The cell lines will be validated for efficacy as a host biomarker for *Plasmodium* infection using a variety of assays, including FACS, immunofluorescence microscopy, and qPCR against Plasmodium 18s rRNA. Through these efforts, we aim to ease the challenges associated with *P. vivax* research.

A.6. APEX2-based proximity labeling and localized proteomics to map the *Plasmodium* liver-stage secretome

Although the word's first malaria vaccine, Mosquirix, was recently licensed in Europe, it is only about 30% effective at preventing disease in recent clinical trials (Wilby et al., 2012). Therefore, new vaccines with improved efficacy should be developed in order to improve the quality of life for millions of people worldwide. The most promising malaria vaccine candidates to date target *Plasmodium* liver stage antigens. This includes Mosquirix, which is directed against the *Plasmodium* circumsporozoite protein (CSP) expressed by sporozoites and liver-stage forms (Gonzalez-Aseguinolaza, 2009). Immunity generated to CSP elicits a strong antibody mediated response in addition to CD4+ T cell response (Leroux-Roels et al., 2014). However, sterile immunity (as demonstrated by the immunization of irradiated sporozoites) has been associated with a robust CD8+ T cell effector response directed against multiple liver stage antigens (Schmidt et al., 2010). The main targets for CD8+ T cells during malaria liver-stage development are antigens derived from proteins exported into the host cell cytoplasm (Riley and Stewart, 2013). Unfortunately, there are currently only a handful of characterized *Plasmodium* proteins that have been described that fall into this category (Ingmundson et al., 2014).

Our proposed research will be the first comprehensive analysis of the malaria liver-stage secretome and therefore has direct implications for improved vaccine development. Briefly, we will isolate Pb-GFP sporozoites from infected *Anopheles* mosquitos and use them to infect immortalized human hepatocytes *in vitro*. After 48 hours, Pb-GFP infected hepatocytes can be isolated from uninfected cells using fluorescence-activated cell sorting (FACS). We will take advantage of a recently described localized proteomics technology from the Ting laboratory at MIT that utilizes an engineered peroxidase enzyme called APEX2 that catalyzes the oxidation of biotin-phenol to covalently label proteins within the direct vicinity of the enzyme (Lam et al., 2015; Rhee et al., 2013). Immunoprecipitation of biotin-labeled proteins coupled with mass spectrometry analysis can be used to effectively map cellular proteomes. A visual overview of this experimental strategy is depicted in Figure A.9.9.

When APEX2 is fused to a cellular localization signal sequence, it can be used to directly label the proteomes of specific cellular compartments with a greater sensitivity and accuracy than traditional organellar proteomic techniques. This is especially important for our proposed research, as it is technically challenging to isolate liver-stage forms away from host material using traditional approaches. In order to specifically map the malaria liver stage secretome, we will utilize this technique to generate a comprehensive inventory of parasite proteins secreted into the host cytoplasm (as well as

other host organelles) of infected hepatocytes isolated using FACS. We are currently in the process of constructing a set of 12 APEX2-based lentiviral vectors that target different cellular compartments, including the following: 1. Nucleolus, 2. Nucleus, 3. Exosomes, 4. Cytoplasm, 5. ER, 6. Golgi, 7. Membrane, 8. Recycling endosomes, 9. Mitochondria, 10. Early endosomes, 11. Late endosomes, and 12. Lysosomes. Table A.10.5. lists these APEX2-based lentiviral constructs with their respective localization sequences.

In summary, this proposed research aims to identify and characterize the *Plasomodium* liver-stage secretome using an innovative localized proteomics approach. In doing so, new parasite proteins with access to the host cell microenvironment may be considered for future vaccine development.

A.7. Acknowledgements

The assessment of RNA quality and dual RNA-sequencing described in section A.2. was performed by the IGM Core at UCSD. We would especially like to thank Kristen Jepson at the IGM Core for her helpful discussion during this process. We thank Lauren Mack and Ye Zheng for allowing us to use their BD FACS Aria cell sorter. We would also like to thank Nathan Lewis, Shangzhong Li, and Alyaa M. Abdel-Haleem for performing the dual RNA-sequencing read mapping and differential gene expression analysis described in section A.2, and for their advice and discussion. We thank Lauren Mack and Ye Zheng for the use of their BD FACS Aria for sorting. We would like to acknowledge Greg Golden for the generation of the Muc13 promoter reporter lentivirus and Pamela Orjuela-Sanchez for the confocal microscopy analysis described in section A.3. The Salk Institute GT3 Viral Vector Core performed titering of the lentivirus. We thank Emily Troemel for her suggestion to use APEX2-based proximity labeling to characterize the *Plasmodium* liver-stage secretome as described in A.4, and to Tomas Bos for the construction of the APEX2-based lentiviral vectors listed in Table A.5, and to Aaron Reinke for helpful discussion. We are also grateful to David Gonzales and John Lapek for their collaboration on the APEX2-based proteomics project described in section A.4.

The Appendix contains material that may be prepared for publication at a future time. Justine Swann, Neema Jamshidi, Shangzhong Li, Alyaa M. Abdel-Haleem, Nathan Lewis, Pamela Orjuela, Greg Golden, John Lapek, Tomas Bos, David Gonzalez, Elizabeth Winzeler. “Systems biology of *Plasmodium* liver-stage development.”

A.8. Materials and Methods

A.8.1. Hepatocyte cell culture

Huh7.5.1 hepatocyte cells were cultured at 37 °C and 5% CO_2 in DMEM (Invitrogen, Carlsbad, USA) supplemented with 10% FCS (Corning cat# 35-011-CV), 200 U/ml penicillin, 200 μg/ml streptomycin (Invitrogen cat# 15140-122), 10 mM Hepes (Invitrogen cat# 15630-080), 1× Glutamax (Invitrogen cat # 35050-061), 1x non-esential amino acids (Invitrogen). During infection, cell media was supplemented with 50μg/ml gentamycin and 50 μg/ml neomycin. After infection (2 hours post infection with sporozoites), the antimycotic 5-fluorocytosine at a final concentration of 50 μg/ml was added to the media described above.

A.8.2. Parasites

P. berghei-GFP (Pb-GFP) (Franke-Fayard et al., 2004) sporozoites were obtained by dissection of infected *Anopheles stephensi* mosquito salivary glands. Dissected salivary glands were homogenized in a glass tissue grinder and filtered twice through nylon cell strainers (20 μm pore size, Millipore SCNY00020) into fresh DMEM without FBS and counted using a Neubauer hemocytometer. The sporozoites were kept on ice until needed. Pb-GFP infected *A. stephensi* mosquitos were obtained from the Insectary Core Facility at New York University.

A.8.3. Preparation of *Plasmodium berghei* liver-stage infection samples and RNA isolation

Huh7.5.1 hepatocytes were infected *in vitro* with *P. berghei* sporozoites expressing a GFP reporter (Franke-Fayard et al., 2004) freshly dissected from infected *A. stephansi* mosquitos (100,000 cells to 30,000 sporozoites per well in 24 well plates). Plates were centrifuged at 330xg for 4 minutes to bring sporozoites closer to cells, and plates were then incubated at 37°C and 5% CO_2 for 2 hours. During this time the sporozoites traverse and invade hepatocytes (Prudencio et al., 2006b). After 2 hours, the cells were washed once with DMEM, then fresh DMEM containing antibiotics/ antimycotics was added, and the cells were returned to the incubator. Cells were collected at time zero (uninfected hepatocytes and sporozoites before infection), time 24 hours post infection (when the sporozoites have transformed into trophozoites), and time 48 hours post infection (when the trophozoites have transformed into liver stage merozoites). To do this, cells were dissociated from plates using TrypLE according to the manufacturer's instructions (ThermoFisher), washed and resuspended in fresh media, and passed through a 40µM cell strainer (Falcon). Uninfected cells were isolated from infected cells by FACS sorting with BD FACS Aria by gating on GFP mean fluorescence intensity (MFI). An average of 10,000 cells were sorted directly into 500µl Qiazol reagent (Qiagen) and total RNA was isolated using a Qiagen miRNEasy kit (Qiagen).

A.8.4. Dual RNA-sequencing

Total RNA was assessed for quantity and quality by an Agilent Tapestation, and RNA libraries were generated using Illumina's TruSeq Stranded Total RNA Sample Prep Kit using at least 100 ng of RNA. RNA libraries were multiplexed and sequenced with 100 basepair (bp) paired single end reads (SR100) to a depth of approximately 15 million reads per sample on an Illumina HiSeq2500. Illumina RTA determined base-calling (bcl2fastq demultiplexes).

A.8.5. Dual RNA-sequencing analysis

Total host-pathogen dual RNA-sequencing reads were filtered and aligned to their respective reference genomes. Tophat.2 was used for the alignment (--library-type fr-unstranded) (Kim et al., 2013). PICARD (http://broadinstitute.github.io/picard/) and SAMtools-0.1.19 (http://samtools.sourceforge.net/) were used for processing the aligned reads. HTSeq (Anders et al., 2015) was used to produce read counts (--stranded=no) To generate *P. berghei* filtered reads, the total reads were first aligned to the human reference genome (GRCh38). Any reads that aligned to this genome were removed, and the remaining reads were then aligned to the reference *P. berghei* –ANKA genome (PlasmoDB13.0-PbergheiANKA). Similarly, host Huh7.5.1 hepatocyte filtered reads were filtered by the alignment to the *P. berghei* genome first, removing these reads, and then aligning the remaining

reads to the human reference genome. Read mapping and alignment statistics are included in Supplemental Dataset 5. All of the raw fastq files as well as differential gene expression analysis have been uploaded to the GEO and SRA databases (http://www.ncbi.nlm.nih.gov/geo/; http://www.ncbi.nlm.nih.gov/sra).

Principal component analysis (PCA) was performed in order to assess the variability of datasets between samples and between experimental replicates. PCA was performed for the filtered reads aligning to the *P. berghei* reference genome and to the human reference genome. The regularized logarithm transformation for raw read counts was used to generate the PCA plots.

Filtered reads aligning to the *P. berghei* and human reference genomes were compared in a pairwise analysis to generate lists of differentially expressed genes (DEG). DESeq2 was used for the differential expression analysis. Three experimental replicates were included for each analysis. Host DEG were generated by comparing infected and uninfected samples at each time point. This analysis is ideal as it removes any artificial changes in host gene expression due to exposure to mosquito salivary gland extracts, as both infected and uninfected cells from each sample are exposed to the same conditions. Further analysis was performed with DEG with a adjusted p value (padj) <0.01 and log2FC>1 or <-1 (for upregulated and downregulated genes,

respectively). *P. berghei* DEG were generated by comparing parasite reads from liver-stage trophozoites (collected at 24 hours post infection) and liver-stage merozoites (collected at 48-50 hours post infection) to sporozoites before infection (collect at zero hours). In this way, gene expression can be monitored as the parasites develop and transform into the different liver stages.

Further analysis of host DEG was performed with the top canonical pathways application using Ingenuity Pathway Analysis (IPA) software (www.ingenuity.com). Functional gene ontology (GO) term for biological processes was performed for *P. berghei* DEG using the open-source *Plasmodium*-specific database PlasmoDB (www.plasmodb.org) (Aurrecoechea et al., 2009b).

A.8.6. Generation of Muc13 promoter reporter lentivirus

The Muc13 promoter was cloned out of the pLightSwitch Prom vector (SwitchGear Genomics, cat#S713988) into the transcriptional response element (TRE) of the lentiviral vector, pGreenFire Lenti-Reporter (System Biosciences, cat# TR010PA-P), expressing both GFP and luciferase reporters. The primers used to amplify the Muc13 promoter out of the pLightSwitch plasmid were as follows: forward, GTAGTAATCGATTTAAGAAGACCCAAATCCAAGCAG; reverse, GTAGTAACTAGTTCCTGGCTACCTTTCGTTTTAACTGTC). The empty

vector was used as a negative control. Lentivirus was made by transfection of 293T cells with the Muc13 promoter pGreenFire lentiviral vector or empty control vector and pPACK lentiviral packaging vectors (System Biosciences, cat# LV500A-1) according to the manufacturers instructions. 293T cell supernatant was collected from cells 28 hours post transfection, centrifuged at 1500rpm for 5 minutes to remove cellular debris, and filtered through a sterile 0.22µm syringe filter (Millipore, cat# SLGVV255F), aliquoted, and stored at -80°C. Titers were determined by transduction of 293T cells with serial dilutions of virus, and qPCR against universal lentiviral integration sequences 24 hours post transduction (titering was performed by the Salk GT3 Viral Vector Core). The pGreenFire Muc13 lentivirus prep was found to have a titer of 9.81 x 10^8 TU/mL, and the pGreenFire Negative Control lentivirus prep was found to have a titer of 3.63 x 10^7 TU/mL

A.8.8. Generation of APEX2 lentiviral constructs

Signal sequences flanked by Gateway attb sites were ordered as gBlocks (Integrated DNA Technologies) and recombined into the pDONR221 Gateway plasmid (LifeTech). Next, the pDONR211 and APEX2-expressing plasmid were recombined to yield fusion proteins targeting the APEX2 reporter gene to different organelles. Clones were sequenced and positive clones were used to flip the signal sequence into a modified pLEX307 (Addgene plasmid # 41392) plasmid containing an APEX2 or APEX2-NES in frame with the signal

sequence. Using this strategy, we generated plasmids targeting 12 different organelles, as described in Table A.7.5.

A.9. Figures

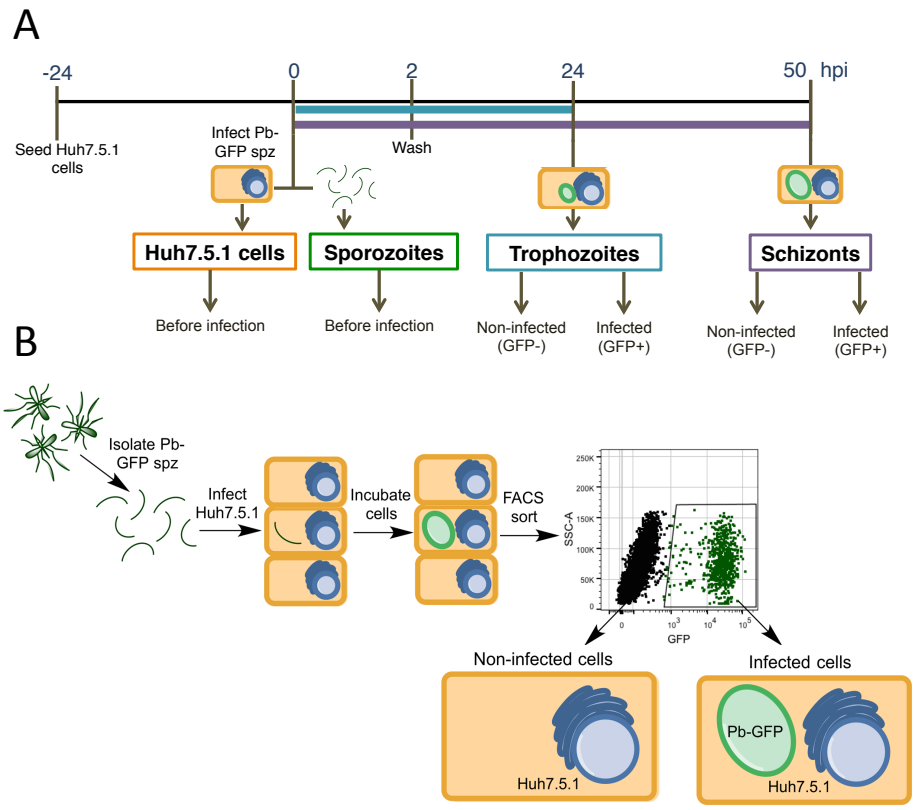


Figure A.9.1. Experimental overview for the collection of *Plasmodium berghei* liver-stage infection samples for dual RNA-sequencing. A) Huh7.5.1 hepatocytes were infected with *Plasmodium berghei* sporozoites expressing recombinant GFP in vitro, and infected and uninfected cells were collected throughout liver-stage development, from 0 to 48-50 hours post infection. At time point zero, before infection, uninfected Huh7.5.1 cells and freshly isolated salivary and sporozoites were collected. After 24 hours post infection, corresponding to a time point in which *P. berghei* has transformed into liver-stage trophozoites, infected hepatocytes were isolated from uninfected hepatocytes by fluorescence-activated cell sorting (FACS). based on *P. berghei* GFP reporter expression (illustrated in B). Similarly, at 50 hours post infection, corresponding to a time point in which *P. berghei* has developed into liver-stage schizonts, B) Overview of FACS sorting technique for the separation of infected and uninfected hepatocytes during the experiment described above in A. Cells are harvested at the indicated time point, and uninfected and infected cells separated using a BS FACS Aria cell sorter based on *P. berghei* reporter GFP expression. A representative example of gating for cells collected at 48-50 hours post infection is shown.

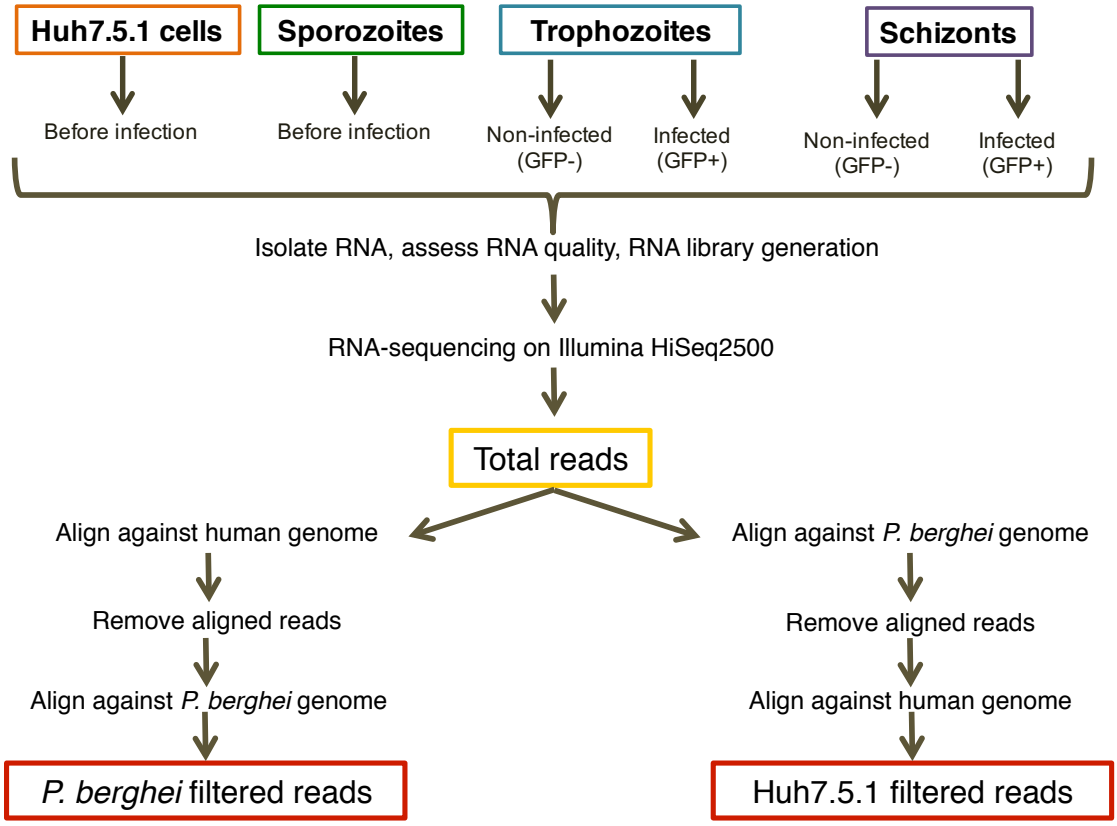


Figure A.9.2. Experimental overview for the dual RNA-sequencing and analysis of samples collected throughout *P. berghei* liver-stage development *in vitro*. Uninfected and infected hepatocytes were isolated throughout *P. berghei*-GFP liver-stage infection *in vitro*, as described above in Figure A.6.1. Total RNA was isolated from samples, and after determination of adequate RNA quality and the generation of corresponding RNA libraries, RNA-sequencing was performed on an Illumina HiSeq2500. Total sequencing reads were then filtered and aligned to their respective genomes. To generate *P. berghei* filtered reads, the total reads were first aligned to the human reference genome. Any reads that aligned to this genome were removed, and the remaining reads were then aligned to the reference *P. berghei* –ANKA genome. Similarly, host Huh7.5.1 hepatocyte filtered reads were filtered by the alignment to the *P. berghei* genome first, removing these reads, and then aligning the remaining reads to the human reference genome.

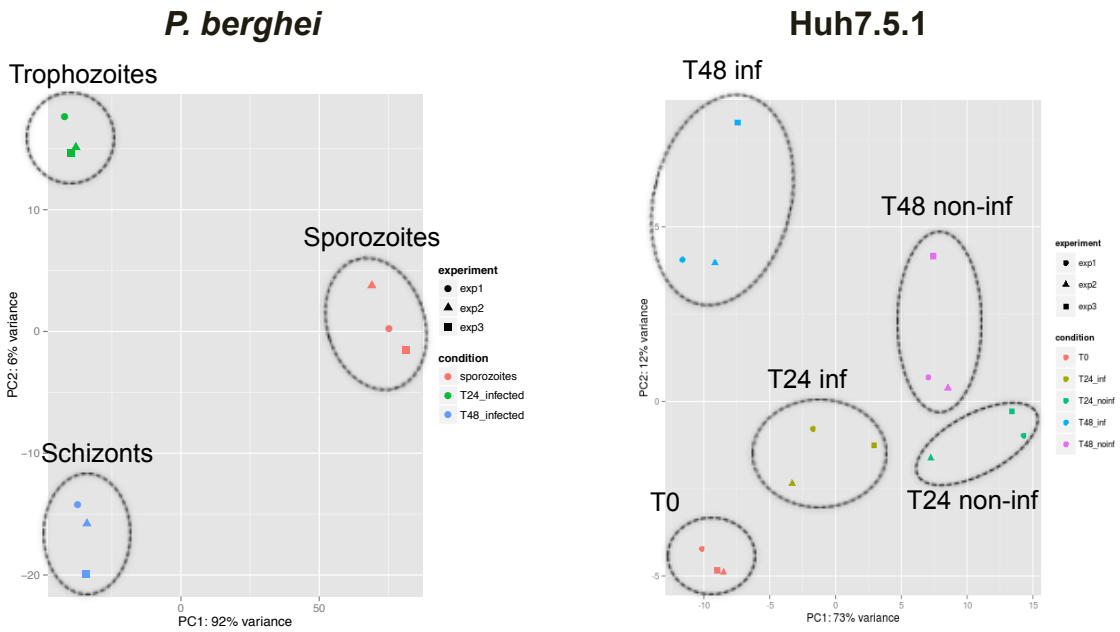


Figure A.9.3. Principal component analysis (PCA) to assess variability of dual RNA-sequencing reads. PCA was performed for the filtered reads aligning to the *P. berghei* reference genome (left panel, *P. berghei*) and to the human reference genome (right panel, Huh7.5.1). Samples are color-coded based on time point (sporozoites, time 0; Huh7.5.1, time zero; trophozoites, infected cells, and uninfected cells time 24 hours post infection; and schizonts, infected cells, and uninfected cells at time 48 hours post infection), and dashed grey circles group these samples together. The shape of the data points (circle, triangle, and square) indicates experimental replicate number (experiments 1, 2, and 3, respectively).

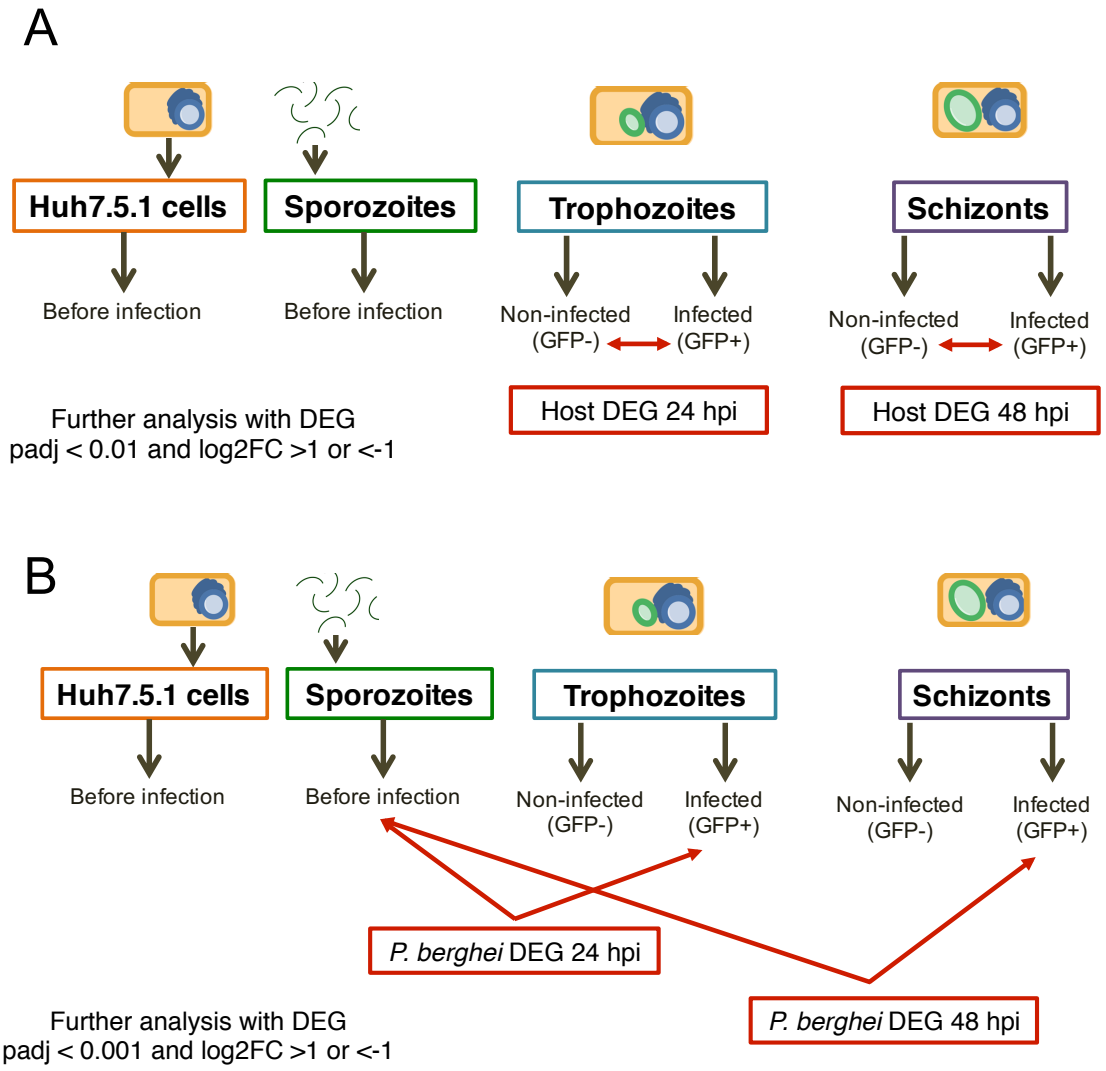


Figure A.9.4. Overview of pairwise comparisons for differential gene expression analysis. Filtered reads aligning to the *P. berghei* and human reference genomes were compared in a pairwise analysis to generate lists of differentially expressed genes (DEG). Three experimental replicates were included for each analysis. A) Host DEG were generated by comparing infected and uninfected samples at each time point. This analysis is ideal as it removes any artificial changes in host gene expression due to mosquito salivary gland extracts, as both infected and uninfected cells from each sample *are exposed to the same conditions. Further analysis was performed with DEG with a* padj <0.01 and log2FC>1 or <-1 (for upregulated and downregulated genes, respectively). B) *P. berghei* DEG were generated by comparing parasite reads from trophozoites (collected at 24 hours post infection) and schizonts (collected at 48-50 hours post infection) to sporozoites before infection (collect at zero hours). In this way, gene expression can be monitored as the parasites develop and transform into the different liver stages.

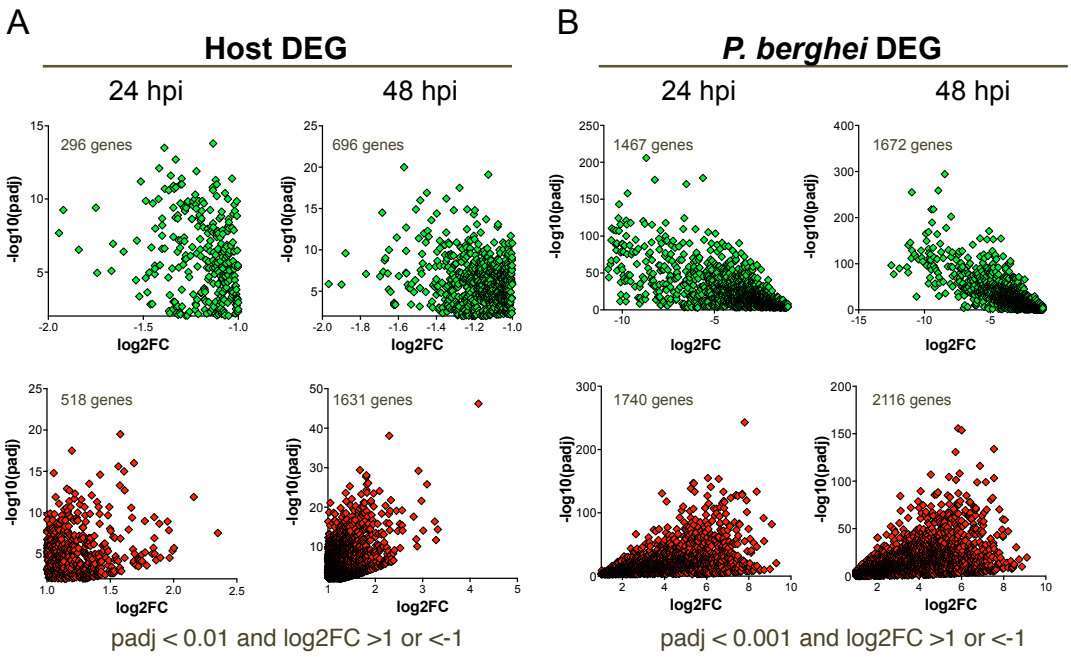
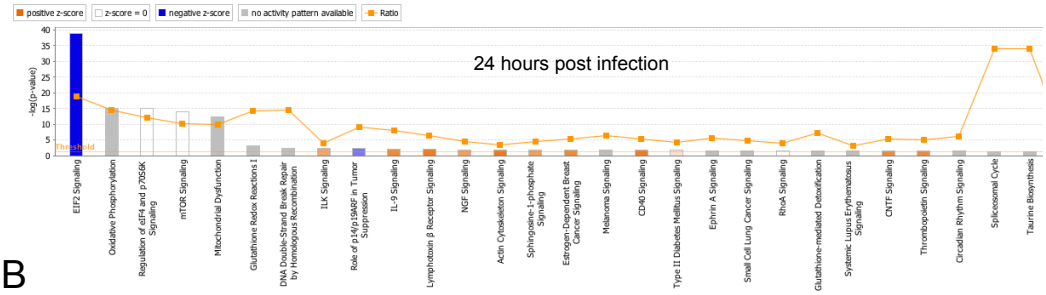


Figure A.9.5. Distribution of host and *P. berghei* differentially expressed genes (DEG). The distribution of host (A) and *P. berghei* (B) DEG is illustrated for each time point. Downregulated genes, with a log2FC<1 are shown in green (top panel) and upregulated genes, with a log2FC>1 are shown in red (bottom panel). All DEG plotted have a padj< 0.01. Each graph shows the transformed p value (-log10(padj)) compared to log2FC in order to visualize the distribution of significantly up- and down-regulated genes. For each analysis, the total number of genes is included in the upper left hand corner. FC= fold change. The full list of host and pathogen DEG for each time point are included in Appendix X.

A



B

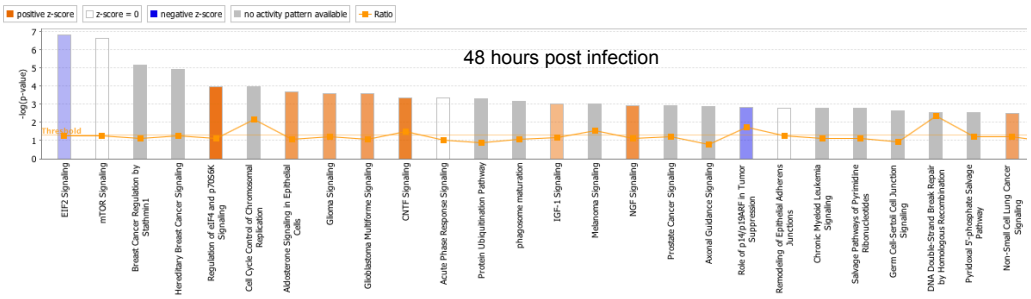


Figure A.9.6. Top canonical pathway analysis for host DEG. The list of host DEG for 24 hours post infection (A) and 48 hours post infection (B) (illustrated in Figure A.6.5 and listed in Appendix X) were analyzed for the top canonical pathways using Ingenuity Pathway Analysis (IPA) Software (http://www.ingenuity.com). Pathways with a positive z-score (indicating a predicted activation of the pathway) are colored in orange, and pathways with a negative z-score (indicating a predicted deactivation of the pathway) are colored in blue. Bars correspond to the –log(p-value) for each pathway, and the relative ratio of DEG to total genes in a pathway is illustrated by the orange line and square data points.

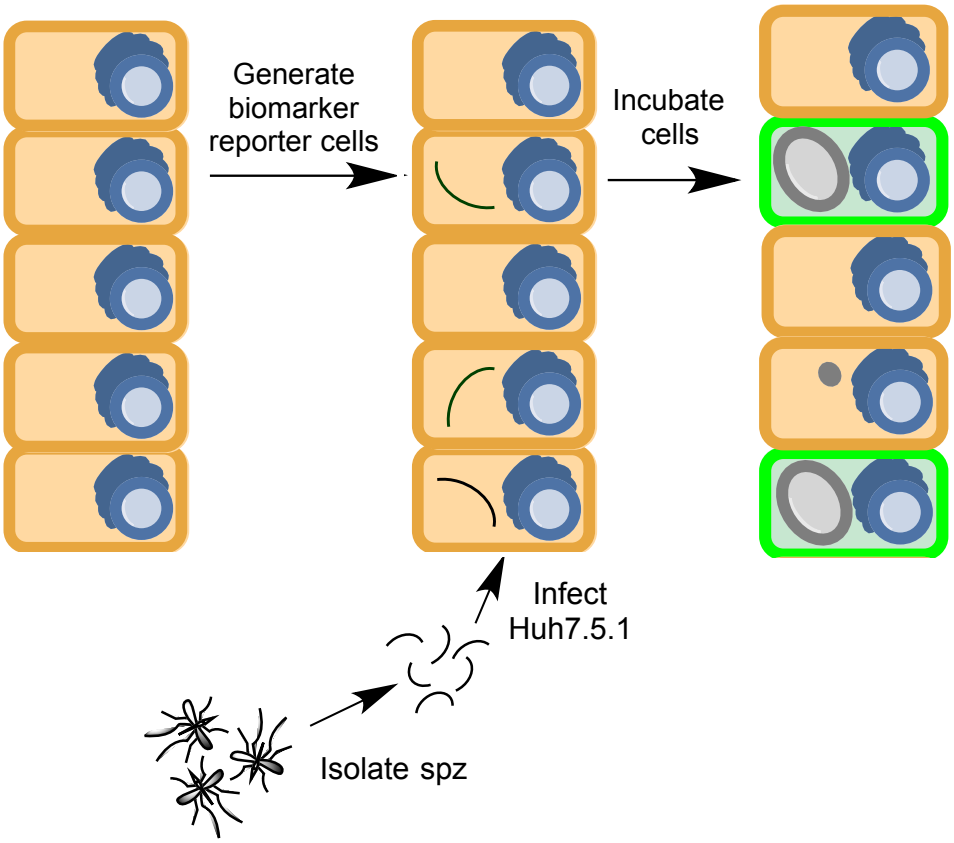


Figure A.9.7. The development of a host biomarker reporter cell line. To overcome the challenges associated with *P. vivax* research, we are developing a host biomarker reporter hepatocyte line in order to label cells productively infected with liver-stage *P. vivax in vitro*. In this way, we will be able to easily assess and quantitate liver-stage infection with *P. vivax*, which at this moment, is very difficult to do. We will use transcriptomic data from *P. berghei* infection to identify and validate potential host biomarkers. A depicted in the illustration, the ideal biomarker cell line will label cells productively infected with *P. vivax* through the production of a reporter protein (such as GFP or luciferase).

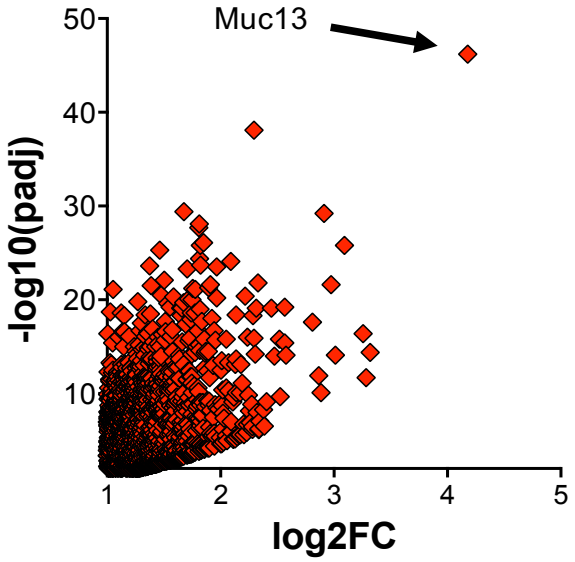


Figure A.9.8. Host Muc13 is significantly upregulated upon infection with *Plasmodium berghei*. A) Host 48 hour post infection DEG alaysis as shown in Figure A.9.5A, with Muc13 labeled. As illustrated, it is the most significantly upregulated host gene after infection with *P. berghei* 48 hours post infection.

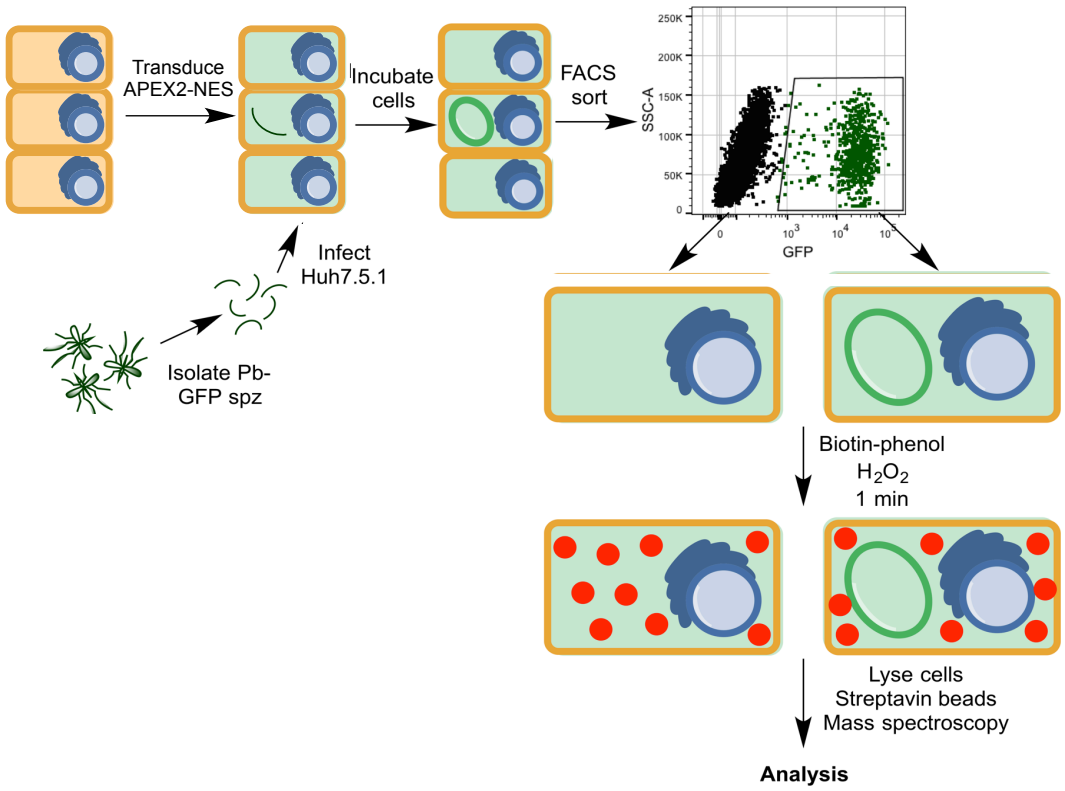


Figure A.9.9. Experimental strategy for APEX2 proximity labeling for mapping the *P. berghei* liver-stage secretome. An overview of the experimental strategy for APEX2-based proximity labeling of the Plasmodium liver-stage secretome is illustrated. In this example, the generation of cells stably expressing APEX-NES is shown, which labels proteins within the cytosol, is shown. Briefly, hepatocytes stably expressing APEX2- localized to a specific cellular compartment (the cytosol in the above example) will be infected with *P. berghei*-GFP, and infected and uninfected cells will be isolated using FACS. Then, APEX2 based proximity labeling will be performed by the incubation of cells with biotin-phenol and H2O2 for 1 minute, and then the cells will be lysed and biotin-labeled proteins will be immunoprecipitated with streptavidin-coated magnetic beads and analyzed via mass spectroscopy.

A.10. Tables

Table A.10.1. GO Term enrichment for *P. berghei* downregulated DEG 24 hours post infection. *P. berghei* 24 hour DEG with a padj <0.001 and log2FC <-1 were analyzed using the gene ontology enrichment application through PlasmoDB (www.plasmodb.org). The table includes GO term ID, name of biological process, background count of genes in each process, number of gene results from the DEG list included in each process, the corresponding percent of background as well as fold enrichment for each GO term, and the odds ratio and p-value for statistical enrichment of DEG for each GO term.

ID	Name	Bgd count	Result count	Pct of bgd	Fold enrichment	Odds ratio	P-value
GO:0040011	locomotion	35	27	77.1	2.81	2.88	7.27E-05
GO:0052126	movement in host environment	27	21	77.8	2.83	2.89	0.00042649
GO:0052192	movement in environment of other organism involved in symbiotic interaction	27	21	77.8	2.83	2.89	0.00042649
GO:0051701	interaction with host	31	21	67.7	2.47	2.51	0.00151533
GO:0044403	symbiosis, encompassing mutualism through parasitism	32	21	65.6	2.39	2.43	0.00201283
GO:0044419	interspecies interaction between organisms	32	21	65.6	2.39	2.43	0.00201283
GO:0044409	entry into host	23	17	73.9	2.69	2.73	0.00219527
GO:0051828	entry into other organism involved in symbiotic interaction	23	17	73.9	2.69	2.73	0.00219527
GO:0051704	multi-organism process	39	23	59	2.15	2.19	0.00344702
GO:0006793	phosphorus metabolic process	120	52	43.3	1.58	1.62	0.00386222
GO:0006796	phosphate-containing compound metabolic process	120	52	43.3	1.58	1.62	0.00386222
GO:0006928	cellular component movement	30	19	63.3	2.31	2.34	0.00432546
GO:0030260	entry into host cell	21	15	71.4	2.6	2.64	0.00495273
GO:0051806	entry into cell of other organism involved in symbiotic interaction	21	15	71.4	2.6	2.64	0.00495273
GO:0007018	microtubule-based movement	19	12	63.2	2.3	2.32	0.02194588
GO:0016310	phosphorylation	90	37	41.1	1.5	1.52	0.02482927
GO:0016311	dephosphorylation	22	13	59.1	2.15	2.17	0.02496525
GO:0006468	protein phosphorylation	79	33	41.8	1.52	1.55	0.02812186
GO:0006470	protein dephosphorylation	19	11	57.9	2.11	2.13	0.04152578
GO:0051674	localization of cell	12	8	66.7	2.43	2.44	0.04748984
GO:0048870	cell motility	12	8	66.7	2.43	2.44	0.04748984

Table A.10.2. GO Term enrichment for *P. berghei* upregulated DEG 24 hours post infection. *P. berghei* 24 hour DEG with a padj <0.001 and log2FC>1 were analyzed using the gene ontology enrichment application through PlasmoDB (www.plasmodb.org). The table includes GO term ID, name of biological process, background count of genes in each process, number of gene results from the DEG list included in each process, the corresponding percent of background as well as fold enrichment for each GO term, and the odds ratio and p-value for statistical enrichment of DEG for each GO term.

ID	Name	Bgd count	Result count	Pct of bgd	Fold enrichm	Odds ratio	P-value
GO:0008152	metabolic process	1153	584	50.7	1.22	1.47	7.92E-08
GO:0044237	cellular metabolic process	974	502	51.5	1.24	1.44	3.88E-07
GO:0044238	primary metabolic process	1000	505	50.5	1.21	1.4	3.04E-06
GO:0010467	gene expression	391	231	59.1	1.42	1.53	3.24E-06
GO:0006412	translation	221	145	65.6	1.57	1.66	6.84E-06
GO:0009987	cellular process	1265	604	47.7	1.15	1.33	5.25E-05
GO:0009058	biosynthetic process	445	246	55.3	1.33	1.42	5.68E-05
GO:0044260	cellular macromolecule metabolic process	759	385	50.7	1.22	1.34	9.91E-05
GO:0044249	cellular biosynthetic process	433	238	55	1.32	1.41	0.00010418
GO:0044267	cellular protein metabolic process	479	257	53.7	1.29	1.38	0.00017555
GO:0008150	biological_process	1602	734	45.8	1.1	1.31	0.00024795
GO:0043170	macromolecule metabolic process	825	409	49.6	1.19	1.3	0.00026582
GO:0019538	protein metabolic process	544	280	51.5	1.23	1.32	0.00071261
GO:0044281	small molecule metabolic process	159	97	61	1.46	1.51	0.00156217
GO:0034645	cellular macromolecule biosynthetic process	330	175	53	1.27	1.32	0.003509
GO:0009059	macromolecule biosynthetic process	331	175	52.9	1.27	1.32	0.00386456
GO:0006396	RNA processing	117	70	59.8	1.44	1.47	0.0094977
GO:0046483	heterocycle metabolic process	86	53	61.6	1.48	1.5	0.01524334
GO:0006807	nitrogen compound metabolic process	430	211	49.1	1.18	1.22	0.01887634
GO:0006163	purine nucleotide metabolic process	37	27	73	1.75	1.77	0.01908982
GO:0006139	nucleobase-containing compound metabolic process	404	199	49.3	1.18	1.22	0.02024803
GO:0044283	small molecule biosynthetic process	30	23	76.7	1.84	1.86	0.02042903
GO:0034641	cellular nitrogen compound metabolic process	427	209	48.9	1.17	1.22	0.02097428
GO:0009141	nucleoside triphosphate metabolic process	31	23	74.2	1.78	1.8	0.02595308
GO:0006414	translational elongation	12	12	100	2.4	2.41	0.02661527
GO:0009259	ribonucleotide metabolic process	35	25	71.4	1.71	1.73	0.02779683
GO:0009117	nucleotide metabolic process	64	40	62.5	1.5	1.52	0.02799304
GO:0006753	nucleoside phosphate metabolic process	64	40	62.5	1.5	1.52	0.02799304
GO:0055086	nucleobase-containing small molecule metabolic process	74	45	60.8	1.46	1.48	0.02803482
GO:0009205	purine ribonucleoside triphosphate metabolic process	30	22	73.3	1.76	1.78	0.03147831
GO:0009144	purine nucleoside triphosphate metabolic process	30	22	73.3	1.76	1.78	0.03147831
GO:0009199	ribonucleoside triphosphate metabolic process	30	22	73.3	1.76	1.78	0.03147831
GO:0009150	purine ribonucleotide metabolic process	34	24	70.6	1.69	1.71	0.03349923
GO:0008380	RNA splicing	34	24	70.6	1.69	1.71	0.03349923
GO:0044248	cellular catabolic process	90	52	57.8	1.39	1.41	0.03571658
GO:0006397	mRNA processing	42	28	66.7	1.6	1.62	0.03624659
GO:0072521	purine-containing compound metabolic process	46	30	65.2	1.56	1.58	0.03710873
GO:0043414	macromolecule methylation	13	12	92.3	2.21	2.23	0.03782327
GO:0009056	catabolic process	114	63	55.3	1.33	1.35	0.04059899
GO:0032259	methylation	19	15	78.9	1.89	1.91	0.04800699

Table A.10.3. GO Term enrichment for *P. berghei* downregulated DEG 48 hours post infection. *P. berghei* 48 hour DEG with a padj <0.001 and log2FC<-2.5 were analyzed using the gene ontology enrichment application through PlasmoDB (www.plasmodb.org). The table includes GO term ID, name of biological process, background count of genes in each process, number of gene results from the DEG list included in each process, the corresponding percent of background as well as fold enrichment for each GO term, and the odds ration p-value for statistical enrichment of DEG for each GO term.

ID	Name	Bgd count	Result count	Pct of bgd	Fold enrichm	Odds ratio	P-value
GO:0051701	interaction with host	31	13	41.9	2.26	2.29	0.0142608
GO:0040011	locomotion	35	14	40	2.15	2.19	0.01516698
GO:0044419	interspecies interaction between organisms	32	13	40.6	2.18	2.22	0.01727919
GO:0044403	symbiosis, encompassing mutualism through parasitism	32	13	40.6	2.18	2.22	0.01727919
GO:0052192	movement in environment of other organism involved in symbiotic interaction	27	11	40.7	2.19	2.22	0.02711617
GO:0052126	movement in host environment	27	11	40.7	2.19	2.22	0.02711617
GO:0051704	multi-organism process	39	14	35.9	1.93	1.96	0.02972251
GO:0051806	entry into cell of other organism involved in symbiotic interaction	21	9	42.9	2.3	2.33	0.03519109
GO:0030260	entry into host cell	21	9	42.9	2.3	2.33	0.03519109

Table A.10.4. GO Term enrichment for *P. berghei* upregulated DEG 48 hours post infection. *P. berghei* 48 hour DEG with a padj <0.001 and log2FC>2 were analyzed using the gene ontology enrichment application through PlasmoDB (www.plasmodb.org). The table includes GO term ID, name of biological process, background count of genes in each process, number of gene results from the DEG list included in each process, the corresponding percent of background as well as fold enrichment for each GO term, and the odds ration p-value for statistical enrichment of DEG for each GO term.

ID	Name	Bgd count	Result count	Pct of bgd	Fold enrichm	Odds ratio	P-value
GO:0051179	localization	240	48	20	1.34	1.39	0.03391307
GO:0055085	transmembrane transport	66	17	25.8	1.73	1.76	0.03450624
GO:0046907	intracellular transport	90	21	23.3	1.56	1.6	0.04474784
GO:0051716	cellular response to stimulus	96	22	22.9	1.54	1.57	0.04680381
GO:0051234	establishment of localization	223	44	19.7	1.32	1.36	0.04884147
GO:0006810	transport	223	44	19.7	1.32	1.36	0.04884147

Table A.10.5. APEX2 constructs to label multiple intracellular organelles. The 12 APEX2-based lentiviral constructs we are currently generating are listed in terms of their organellar targets. In parenthesis, the cellular localization signal that is used for targeting is listed.

1. Nucleolus (NOP10)	7. Membrane (palmitoyl from p63)
2. Nucleus (NLS)	8. Recycling endosomes (Rab11)
3. Exosomes (CD63)	9. Mitochondria (COX8A 1-29)
4. Cytoplasm (NES)	10. Early endosomes (Rab5a)
5. ER (ER retention signal)	11. Late endosomes (Rab7a)
6. Golgi (B4GALT1)	12. Lysosomes (Lamp1)

REFERENCES

Abrahamsen, M.S., Templeton, T.J., Enomoto, S., Abrahante, J.E., Zhu, G., Lancto, C.A., Deng, M., Liu, C., Widmer, G., Tzipori, S., Buck, G.A., Xu, P., Bankier, A.T., Dear, P.H., Konfortov, B.A., Spriggs, H.F., Iyer, L., Anantharaman, V., Aravind, L., and Kapur, V. (2004). Complete genome sequence of the apicomplexan, Cryptosporidium parvum. Science *304*, 441-445.

Allali-Hassani, A., Pan, P.W., Dombrovski, L., Najmanovich, R., Tempel, W., Dong, A., Loppnau, P., Martin, F., Thornton, J., Edwards, A.M., Bochkarev, A., Plotnikov, A.N., Vedadi, M., and Arrowsmith, C.H. (2007). Structural and chemical profiling of the human cytosolic sulfotransferases. PLoS biology *5*, e97.

Alonso, P.L., Brown, G., Arevalo-Herrera, M., Binka, F., Chitnis, C., Collins, F., Doumbo, O.K., Greenwood, B., Hall, B.F., Levine, M.M., Mendis, K., Newman, R.D., Plowe, C.V., Rodriguez, M.H., Sinden, R., Slutsker, L., and Tanner, M. (2011). A research agenda to underpin malaria eradication. PLoS medicine *8*, e1000406.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166-169.

Andersen, J.S., and Mann, M. (2006). Organellar proteomics: turning inventories into insights. EMBO reports *7*, 874-879.

Andrews, K.T., Fisher, G., and Skinner-Adams, T.S. (2014). Drug repurposing and human parasitic protozoan diseases. International journal for parasitology Drugs and drug resistance *4*, 95-111.

Arhel, N., and Kirchhoff, F. (2010). Host proteins involved in HIV infection: new therapeutic targets. Biochimica et biophysica acta *1802*, 313-321.

Ariey, F., Witkowski, B., Amaratunga, C., Beghain, J., Langlois, A.C., Khim, N., Kim, S., Duru, V., Bouchier, C., Ma, L., Lim, P., Leang, R., Duong, S., Sreng, S., Suon, S., Chuor, C.M., Bout, D.M., Menard, S., Rogers, W.O., Genton, B., Fandeur, T., Miotto, O., Ringwald, P., Le Bras, J., Berry, A., Barale, J.C., Fairhurst, R.M., Benoit-Vical, F., Mercereau-Puijalon, O., and Menard, D. (2014). A molecular marker of artemisinin-resistant Plasmodium falciparum malaria. Nature *505*, 50-55.

Aslett, M., Aurrecoechea, C., Berriman, M., Brestelli, J., Brunk, B.P., Carrington, M., Depledge, D.P., Fischer, S., Gajria, B., Gao, X., Gardner, M.J., Gingle, A., Grant, G., Harb, O.S., Heiges, M., Hertz-Fowler, C., Houston, R., Innamorato, F., Iodice, J., Kissinger, J.C., Kraemer, E., Li, W., Logan, F.J.,

Miller, J.A., Mitra, S., Myler, P.J., Nayak, V., Pennington, C., Phan, I., Pinney, D.F., Ramasamy, G., Rogers, M.B., Roos, D.S., Ross, C., Sivam, D., Smith, D.F., Srinivasamoorthy, G., Stoeckert, C.J., Jr., Subramanian, S., Thibodeau, R., Tivey, A., Treatman, C., Velarde, G., and Wang, H. (2010). TriTrypDB: a functional genomic resource for the Trypanosomatidae. Nucleic acids research *38*, D457-462.

Aurrecoechea, C., Barreto, A., Brestelli, J., Brunk, B.P., Caler, E.V., Fischer, S., Gajria, B., Gao, X., Gingle, A., Grant, G., Harb, O.S., Heiges, M., Iodice, J., Kissinger, J.C., Kraemer, E.T., Li, W., Nayak, V., Pennington, C., Pinney, D.F., Pitts, B., Roos, D.S., Srinivasamoorthy, G., Stoeckert, C.J., Jr., Treatman, C., and Wang, H. (2011). AmoebaDB and MicrosporidiaDB: functional genomic resources for Amoebozoa and Microsporidia species. Nucleic acids research *39*, D612-619.

Aurrecoechea, C., Brestelli, J., Brunk, B.P., Carlton, J.M., Dommer, J., Fischer, S., Gajria, B., Gao, X., Gingle, A., Grant, G., Harb, O.S., Heiges, M., Innamorato, F., Iodice, J., Kissinger, J.C., Kraemer, E., Li, W., Miller, J.A., Morrison, H.G., Nayak, V., Pennington, C., Pinney, D.F., Roos, D.S., Ross, C., Stoeckert, C.J., Jr., Sullivan, S., Treatman, C., and Wang, H. (2009a). GiardiaDB and TrichDB: integrated genomic resources for the eukaryotic protist pathogens Giardia lamblia and Trichomonas vaginalis. Nucleic acids research *37*, D526-530.

Aurrecoechea, C., Brestelli, J., Brunk, B.P., Dommer, J., Fischer, S., Gajria, B., Gao, X., Gingle, A., Grant, G., Harb, O.S., Heiges, M., Innamorato, F., Iodice, J., Kissinger, J.C., Kraemer, E., Li, W., Miller, J.A., Nayak, V., Pennington, C., Pinney, D.F., Roos, D.S., Ross, C., Stoeckert, C.J., Jr., Treatman, C., and Wang, H. (2009b). PlasmoDB: a functional genomic database for malaria parasites. Nucleic acids research *37*, D539-543.

Aurrecoechea, C., Brestelli, J., Brunk, B.P., Fischer, S., Gajria, B., Gao, X., Gingle, A., Grant, G., Harb, O.S., Heiges, M., Innamorato, F., Iodice, J., Kissinger, J.C., Kraemer, E.T., Li, W., Miller, J.A., Nayak, V., Pennington, C., Pinney, D.F., Roos, D.S., Ross, C., Srinivasamoorthy, G., Stoeckert, C.J., Jr., Thibodeau, R., Treatman, C., and Wang, H. (2010). EuPathDB: a portal to eukaryotic pathogen databases. Nucleic acids research *38*, D415-419.

Baker, D.A. (2010a). Malaria gametocytogenesis. Molecular and biochemical parasitology *172*, 57-65.

Baker, D.A. (2010b). Malaria gametocytogenesis. Molecular and biochemical parasitology *172*, 57-65.

Baldauf, H.M., Pan, X., Erikson, E., Schmidt, S., Daddacha, W., Burggraf, M., Schenkova, K., Ambiel, I., Wabnitz, G., Gramberg, T., Panitz, S., Flory, E.,

Landau, N.R., Sertel, S., Rutsch, F., Lasitschka, F., Kim, B., Konig, R., Fackler, O.T., and Keppler, O.T. (2012). SAMHD1 restricts HIV-1 infection in resting CD4(+) T cells. Nature medicine *18*, 1682-1687.

Baragana, B., Hallyburton, I., Lee, M.C., Norcross, N.R., Grimaldi, R., Otto, T.D., Proto, W.R., Blagborough, A.M., Meister, S., Wirjanata, G., Ruecker, A., Upton, L.M., Abraham, T.S., Almeida, M.J., Pradhan, A., Porzelle, A., Martinez, M.S., Bolscher, J.M., Woodland, A., Norval, S., Zuccotto, F., Thomas, J., Simeons, F., Stojanovski, L., Osuna-Cabello, M., Brock, P.M., Churcher, T.S., Sala, K.A., Zakutansky, S.E., Jimenez-Diaz, M.B., Sanz, L.M., Riley, J., Basak, R., Campbell, M., Avery, V.M., Sauerwein, R.W., Dechering, K.J., Noviyanti, R., Campo, B., Frearson, J.A., Angulo-Barturen, I., Ferrer-Bazaga, S., Gamo, F.J., Wyatt, P.G., Leroy, D., Siegl, P., Delves, M.J., Kyle, D.E., Wittlin, S., Marfurt, J., Price, R.N., Sinden, R.E., Winzeler, E.A., Charman, S.A., Bebrevska, L., Gray, D.W., Campbell, S., Fairlamb, A.H., Willis, P.A., Rayner, J.C., Fidock, D.A., Read, K.D., and Gilbert, I.H. (2015). A novel multiple-stage antimalarial agent that inhibits protein synthesis. Nature *522*, 315-320.

Beiting, D.P., and Roos, D.S. (2011). A systems biological view of intracellular pathogens. Immunological reviews *240*, 117-128.

Bissantz, C., Folkers, G., and Rognan, D. (2000). Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. J Med Chem *43*, 4759-4767.

Bjorndal, A., Deng, H., Jansson, M., Fiore, J.R., Colognesi, C., Karlsson, A., Albert, J., Scarlatti, G., Littman, D.R., and Fenyo, E.M. (1997). Coreceptor usage of primary human immunodeficiency virus type 1 isolates varies according to biological phenotype. Journal of virology *71*, 7478-7487.

Blader, I.J., Manger, I.D., and Boothroyd, J.C. (2001). Microarray analysis reveals previously unknown changes in Toxoplasma gondii-infected human cells. The Journal of biological chemistry *276*, 24223-24231.

Bollard, M.E., Stanley, E.G., Lindon, J.C., Nicholson, J.K., and Holmes, E. (2005). NMR-based metabonomic approaches for evaluating physiological influences on biofluid composition. NMR in biomedicine *18*, 143-162.

Bordbar, A., Lewis, N.E., Schellenberger, J., Palsson, B.O., and Jamshidi, N. (2010). Insight into human alveolar macrophage and M. tuberculosis interactions via metabolic reconstructions. Molecular systems biology *6*, 422.

Bourzac, K. (2014). Infectious disease: Beating the big three. Nature *507*, S4-7.

Bozdech, Z., Llinas, M., Pulliam, B.L., Wong, E.D., Zhu, J., and DeRisi, J.L. (2003). The transcriptome of the intraerythrocytic developmental cycle of Plasmodium falciparum. PLoS biology *1*, E5.

Brass, A.L., Dykxhoorn, D.M., Benita, Y., Yan, N., Engelman, A., Xavier, R.J., Lieberman, J., and Elledge, S.J. (2008). Identification of host proteins required for HIV infection through a functional genomic screen. Science *319*, 921-926.

Bringaud, F., Biran, M., Millerioux, Y., Wargnies, M., Allmann, S., and Mazet, M. (2015). Combining reverse genetics and NMR-based metabolomics unravels trypanosome-specific metabolic pathways. Molecular microbiology.

Bruce, J.W., Ahlquist, P., and Young, J.A. (2008). The host cell sulfonation pathway contributes to retroviral infection at a step coincident with provirus establishment. PLoS pathogens *4*, e1000207.

Bunnik, E.M., Chung, D.W., Hamilton, M., Ponts, N., Saraf, A., Prudhomme, J., Florens, L., and Le Roch, K.G. (2013). Polysome profiling reveals translational control of gene expression in the human malaria parasite Plasmodium falciparum. Genome biology *14*, R128.

Bushman, F.D., Malani, N., Fernandes, J., D'Orso, I., Cagney, G., Diamond, T.L., Zhou, H., Hazuda, D.J., Espeseth, A.S., Konig, R., Bandyopadhyay, S., Ideker, T., Goff, S.P., Krogan, N.J., Frankel, A.D., Young, J.A., and Chanda, S.K. (2009). Host cell factors in HIV replication: meta-analysis of genome-wide studies. PLoS pathogens *5*, e1000437.

Butler, C.L., Lucas, O., Wuchty, S., Xue, B., Uversky, V.N., and White, M. (2014). Identifying novel cell cycle proteins in Apicomplexa parasites through co-expression decision analysis. PloS one *9*, e97625.

Capewell, P., Monk, S., Ivens, A., MacGregor, P., Fenn, K., Walrad, P., Bringaud, F., Smith, T.K., and Matthews, K.R. (2013). Regulation of Trypanosoma brucei Total and Polysomal mRNA during Development within Its Mammalian Host. PloS one *8*.

Capper, M.J., O'Neill, P.M., Fisher, N., Strange, R.W., Moss, D., Ward, S.A., Berry, N.G., Lawrenson, A.S., Hasnain, S.S., Biagini, G.A., and Antonyuk, S.V. (2015). Antimalarial 4(1H)-pyridones bind to the Qi site of cytochrome bc1. Proceedings of the National Academy of Sciences *112*, 755-760.

Caro, F., Ahyong, V., Betegon, M., and DeRisi, J.L. (2014). Genome-wide regulatory dynamics of translation in the asexual blood stages. eLife *3*.

Carter, R., and Mendis, K.N. (2002). Evolutionary and historical aspects of the burden of malaria. Clin Microbiol Rev *15*, 564-594.

Chang, M.W., Ayeni, C., Breuer, S., and Torbett, B.E. (2010). Virtual screening for HIV protease inhibitors: a comparison of AutoDock 4 and Vina. PloS one *5*, e11955.

Chapman, E., Best, M.D., Hanson, S.R., and Wong, C.H. (2004). Sulfotransferases: structure, mechanism, biological activity, inhibition, and synthetic utility. Angewandte Chemie *43*, 3526-3548.

Chavali, A.K., Blazier, A.S., Tlaxca, J.L., Jensen, P.A., Pearson, R.D., and Papin, J.A. (2012). Metabolic network analysis predicts efficacy of FDA-approved drugs targeting the causative agent of a neglected tropical disease. BMC systems biology *6*, 27.

Chavali, A.K., Whittemore, J.D., Eddy, J.A., Williams, K.T., and Papin, J.A. (2008). Systems analysis of metabolism in the pathogenic trypanosomatid Leishmania major. Molecular systems biology *4*, 177.

Cho, Y., Ioerger, T.R., and Sacchettini, J.C. (2008). Discovery of novel nitrobenzothiazole inhibitors for Mycobacterium tuberculosis ATP phosphoribosyl transferase (HisG) through virtual screening. J Med Chem *51*, 5984-5992.

Choe, H., Li, W., Wright, P.L., Vasilieva, N., Venturi, M., Huang, C.C., Grundner, C., Dorfman, T., Zwick, M.B., Wang, L., Rosenberg, E.S., Kwong, P.D., Burton, D.R., Robinson, J.E., Sodroski, J.G., and Farzan, M. (2003). Tyrosine sulfation of human antibodies contributes to recognition of the CCR5 binding region of HIV-1 gp120. Cell *114*, 161-170.

Clark, I.A., al Yaman, F.M., and Jacobson, L.S. (1997a). The biological basis of malarial disease. International journal for parasitology *27*, 1237-1249.

Clark, I.A., AlYaman, F.M., and Jacobson, L.S. (1997b). The biological basis of malarial disease. International journal for parasitology *27*, 1237-1249.

Cleary, M.D., Singh, U., Blader, I.J., Brewer, J.L., and Boothroyd, J.C. (2002). Toxoplasma gondii asexual development: identification of developmentally regulated genes and distinct patterns of gene expression. Eukaryotic cell *1*, 329-340.

Connor, R.I., Chen, B.K., Choe, S., and Landau, N.R. (1995). Vpr is required for efficient replication of human immunodeficiency virus type-1 in mononuclear phagocytes. Virology *206*, 935-944.

Cooper, M.A., and Shlaes, D. (2011). Fix the antibiotics pipeline. Nature *472*, 32-32.

Corish, P., and Tyler-Smith, C. (1999). Attenuation of green fluorescent protein half-life in mammalian cells. Protein engineering *12*, 1035-1040.

Cox, J., and Mann, M. (2011). Quantitative, high-resolution proteomics for data-driven systems biology. Annual review of biochemistry *80*, 273-299.

Cui, L., Mharakurwa, S., Ndiaye, D., Rathod, P.K., and Rosenthal, P.J. (2015). Antimalarial Drug Resistance: Literature Review and Activities and Findings of the ICEMR Network. The American journal of tropical medicine and hygiene *93*, 57-68.

da Cruz, F.P., Martin, C., Buchholz, K., Lafuente-Monasterio, M.J., Rodrigues, T., Sonnichsen, B., Moreira, R., Gamo, F.J., Marti, M., Mota, M.M., Hannus, M., and Prudencio, M. (2012). Drug screen targeted at Plasmodium liver stages identifies a potent multistage antimalarial drug. The Journal of infectious diseases *205*, 1278-1286.

Dandapani, S., and Marcaurelle, L.A. (2010). Grand challenge commentary: Accessing new chemical space for 'undruggable' targets. Nature chemical biology *6*, 861-863.

Date, S.V., and Stoeckert, C.J., Jr. (2006). Computational modeling of the Plasmodium falciparum interactome reveals protein function on a genome-wide scale. Genome research *16*, 542-549.

Daubener, W., Spors, B., Hucke, C., Adam, R., Stins, M., Kim, K.S., and Schroten, H. (2001). Restriction of Toxoplasma gondii growth in human brain microvascular endothelial cells by activation of indoleamine 2,3-dioxygenase. Infection and immunity *69*, 6527-6531.

Delves, M., Plouffe, D., Scheurer, C., Meister, S., Wittlin, S., Winzeler, E.A., Sinden, R.E., and Leroy, D. (2012). The activities of current antimalarial drugs on the life cycle stages of Plasmodium: a comparative study with human and rodent parasites. PLoS medicine *9*, e1001169.

Derbyshire, E.R., Mota, M.M., and Clardy, J. (2011). The next opportunity in anti-malaria drug discovery: the liver stage. PLoS pathogens *7*, e1002178.

Derbyshire, E.R., Prudencio, M., Mota, M.M., and Clardy, J. (2012). Liver-stage malaria parasites vulnerable to diverse chemical scaffolds. Proceedings of the National Academy of Sciences of the United States of America *109*, 8511-8516.

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., and Winzeler, E.A. (2010).

Whole-genome sequencing and microarray analysis of ex vivo Plasmodium vivax reveal selective pressure on putative drug resistance genes. Proceedings of the National Academy of Sciences of the United States of America *107*, 20045-20050.

Dong, C.K., Urgaonkar, S., Cortese, J.F., Gamo, F.J., Garcia-Bustos, J.F., Lafuente, M.J., Patel, V., Ross, L., Coleman, B.I., Derbyshire, E.R., Clish, C.B., Serrano, A.E., Cromwell, M., Barker, R.H., Jr., Dvorin, J.D., Duraisingh, M.T., Wirth, D.F., Clardy, J., and Mazitschek, R. (2011). Identification and validation of tetracyclic benzothiazepines as Plasmodium falciparum cytochrome bc1 inhibitors. Chemistry & biology *18*, 1602-1610.

Duncan, R. (2004). DNA microarray analysis of protozoan parasite gene expression: outcomes correlate with mechanisms of regulation. Trends in parasitology *20*, 211-215.

Durmus, S., Cakir, T., Ozgur, A., and Guthke, R. (2015). A review on computational systems biology of pathogen-host interactions. Frontiers in microbiology *6*, 235.

Ehrenkaufer, G.M., Weedall, G.D., Williams, D., Lorenzi, H.A., Caler, E., Hall, N., and Singh, U. (2013). The genome and transcriptome of the enteric parasite Entamoeba invadens, a model for encystation. Genome biology *14*, R77.

Falany, C.N. (1997). Enzymology of human cytosolic sulfotransferases. FASEB journal : official publication of the Federation of American Societies for Experimental Biology *11*, 206-216.

Farrell, A., Coleman, B.I., Benenati, B., Brown, K.M., Blader, I.J., Marth, G.T., and Gubbels, M.J. (2014). Whole genome profiling of spontaneous and chemically induced mutations in Toxoplasma gondii. BMC genomics *15*, 354.

Farzan, M., Mirzabekov, T., Kolchinsky, P., Wyatt, R., Cayabyab, M., Gerard, N.P., Gerard, C., Sodroski, J., and Choe, H. (1999). Tyrosine sulfation of the amino terminus of CCR5 facilitates HIV-1 entry. Cell *96*, 667-676.

Fatkenheuer, G., Pozniak, A.L., Johnson, M.A., Plettenberg, A., Staszewski, S., Hoepelman, A.I., Saag, M.S., Goebel, F.D., Rockstroh, J.K., Dezube, B.J., Jenkins, T.M., Medhurst, C., Sullivan, J.F., Ridgway, C., Abel, S., James, I.T., Youle, M., and van der Ryst, E. (2005). Efficacy of short-term monotherapy with maraviroc, a new CCR5 antagonist, in patients infected with HIV-1. Nature medicine *11*, 1170-1172.

Feist, A.M., Herrgard, M.J., Thiele, I., Reed, J.L., and Palsson, B.O. (2009). Reconstruction of biochemical networks in microorganisms. Nat Rev Microbiol *7*, 129-143.

Flannery, E.L., Chatterjee, A.K., and Winzeler, E.A. (2013). Antimalarial drug discovery - approaches and progress towards new medicines. Nature reviews Microbiology *11*, 849-862.

Fonager, J., Franke-Fayard, B.M., Adams, J.H., Ramesar, J., Klop, O., Khan, S.M., Janse, C.J., and Waters, A.P. (2011). Development of the piggyBac transposable system for Plasmodium berghei and its application for random mutagenesis in malaria parasites. BMC genomics *12*, 155.

Foth, B.J., Zhang, N., Chaal, B.K., Sze, S.K., Preiser, P.R., and Bozdech, Z. (2011). Quantitative time-course profiling of parasite and host cell proteins in the human malaria parasite Plasmodium falciparum. Molecular & cellular proteomics : MCP *10*, M110 006411.

Franke-Fayard, B., Trueman, H., Ramesar, J., Mendoza, J., van der Keur, M., van der Linden, R., Sinden, R.E., Waters, A.P., and Janse, C.J. (2004). A Plasmodium berghei reference line that constitutively expresses GFP at a high level throughout the complete life cycle. Mol Biochem Parasitol *137*, 23-33.

Franklin, B.S., Parroche, P., Ataide, M.A., Lauw, F., Ropert, C., de Oliveira, R.B., Pereira, D., Tada, M.S., Nogueira, P., da Silva, L.H., Bjorkbacka, H., Golenbock, D.T., and Gazzinelli, R.T. (2009). Malaria primes the innate immune response due to interferon-gamma induced enhancement of toll-like receptor expression and function. Proceedings of the National Academy of Sciences of the United States of America *106*, 5789-5794.

Friedrich, B.M., Dziuba, N., Li, G., Endsley, M.A., Murray, J.L., and Ferguson, M.R. (2011). Host factors mediating HIV-1 replication. Virus research *161*, 101-114.

Fruh, K., Finlay, B., and McFadden, G. (2010). On the road to systems biology of host-pathogen interactions. Future microbiology *5*, 131-133.

Gajria, B., Bahl, A., Brestelli, J., Dommer, J., Fischer, S., Gao, X., Heiges, M., Iodice, J., Kissinger, J.C., Mackey, A.J., Pinney, D.F., Roos, D.S., Stoeckert, C.J., Jr., Wang, H., and Brunk, B.P. (2008). ToxoDB: an integrated Toxoplasma gondii database resource. Nucleic acids research *36*, D553-556.

Gamage, N., Barnett, A., Hempel, N., Duggleby, R.G., Windmill, K.F., Martin, J.L., and McManus, M.E. (2006). Human sulfotransferases and their role in chemical metabolism. Toxicological sciences : an official journal of the Society of Toxicology *90*, 5-22.

Ganesan, S., Chaurasiya, N.D., Sahu, R., Walker, L.A., and Tekwani, B.L. (2012). Understanding the mechanisms for metabolism-linked hemolytic toxicity of primaquine against glucose 6-phosphate dehydrogenase deficient human erythrocytes: evaluation of eryptotic pathway. Toxicology *294*, 54-60.

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., and Barrell, B. (2002). Genome sequence of the human malaria parasite Plasmodium falciparum. Nature *419*, 498-511.

Gething, P.W., Elyazar, I.R.F., 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.L., Price, R.N., Mueller, I., Baird, K., and Hay, S.I. (2012). A Long Neglected World Malaria Map: Plasmodium vivax Endemicity in 2010. PLoS neglected tropical diseases *6*.

Ghorbal, M., Gorman, M., Macpherson, C.R., Martins, R.M., Scherf, A., and Lopez-Rubio, J.J. (2014). Genome editing in the human malaria parasite Plasmodium falciparum using the CRISPR-Cas9 system. Nature biotechnology *32*, 819-821.

Glatt, H. (2000). Sulfotransferases in the bioactivation of xenobiotics. Chemico-biological interactions *129*, 141-170.

Gobert, G.N., Moertel, L.P., and McManus, D.P. (2005). Microarrays: new tools to unravel parasite transcriptomes. Parasitology *131*, 439-448.

Goff, S.P. (2007). Host factors exploited by retroviruses. Nature reviews Microbiology *5*, 253-263.

Goldfless, S.J., Wagner, J.C., and Niles, J.C. (2014). Versatile control of Plasmodium falciparum gene expression with an inducible protein-RNA interaction. Nature communications *5*, 5329.

Gomes, A.R., Bushell, E., Schwach, F., Girling, G., Anar, B., Quail, M.A., Herd, C., Pfander, C., Modrzynska, K., Rayner, J.C., and Billker, O. (2015). A genome-scale vector resource enables high-throughput reverse genetic screening in a malaria parasite. Cell host & microbe *17*, 404-413.

Gonzalez-Aseguinolaza, G. (2009). Malaria vaccine: the latest news from RTS,S/AS01E vaccine. Expert Rev Vaccines *8*, 285-288.

Gu, W.G., and Chen, X.Q. (2014). Targeting CCR5 for anti-HIV research. Eur J Clin Microbiol *33*, 1881-1887.

Gulick, R.M., Lalezari, J., Goodrich, J., Clumeck, N., DeJesus, E., Horban, A., Nadler, J., Clotet, B., Karlsson, A., Wohlfeiler, M., Montana, J.B., McHale, M., Sullivan, J., Ridgway, C., Felstead, S., Dunne, M.W., van der Ryst, E., Mayer, H., and Teams, M.S. (2008). Maraviroc for previously treated patients with R5 HIV-1 infection. The New England journal of medicine *359*, 1429-1441.

Gulick, R.M., Su, Z.H., Flexner, C., Hughes, M.D., Skolnik, P.R., Wilkin, T.J., Gross, R., Krambrink, A., Coakley, E., Greaves, W.L., Zolopa, A., Reichman, R., Godfrey, C., Hirsch, M., and Kuritzkes, D.R. (2007). Phase 2 study of the safety and efficacy of vicriviroc, a CCR5 inhibitor, in HIV-1-infected, treatment experienced patients: AIDS clinical trials group 5211. Journal of Infectious Diseases *196*, 304-312.

Gunasekera, K., Wuthrich, D., Braga-Lagache, S., Heller, M., and Ochsenreiter, T. (2012). Proteome remodelling during development from blood to insect-form Trypanosoma brucei quantified by SILAC and mass spectrometry. BMC genomics *13*, 556.

Hall, N., Karras, M., Raine, J.D., Carlton, J.M., Kooij, T.W., Berriman, M., Florens, L., Janssen, C.S., Pain, A., Christophides, G.K., James, K., Rutherford, K., Harris, B., Harris, D., Churcher, C., Quail, M.A., Ormond, D., Doggett, J., Trueman, H.E., Mendoza, J., Bidwell, S.L., Rajandream, M.A., Carucci, D.J., Yates, J.R., 3rd, Kafatos, F.C., Janse, C.J., Barrell, B., Turner, C.M., Waters, A.P., and Sinden, R.E. (2005). A comprehensive survey of the Plasmodium life cycle by genomic, transcriptomic, and proteomic analyses. Science *307*, 82-86.

Hao, L.H., He, Q.L., Wang, Z.S., Craven, M., Newton, M.A., and Ahlquist, P. (2013). Limited Agreement of Independent RNAi Screens for Virus-Required Host Genes Owes More to False-Negative than False-Positive Factors. PLoS computational biology *9*.

Haug, K., Salek, R.M., Conesa, P., Hastings, J., de Matos, P., Rijnbeek, M., Mahendraker, T., Williams, M., Neumann, S., Rocca-Serra, P., Maguire, E., Gonzalez-Beltran, A., Sansone, S.A., Griffin, J.L., and Steinbeck, C. (2013). MetaboLights--an open-access general-purpose repository for metabolomics studies and associated meta-data. Nucleic acids research *41*, D781-786.

Hawn, T.R., Shah, J.A., and Kalman, D. (2015). New tricks for old dogs: countering antibiotic resistance in tuberculosis with host-directed therapeutics. Immunological reviews *264*, 344-362.

Hebbring, S.J., Adjei, A.A., Baer, J.L., Jenkins, G.D., Zhang, J., Cunningham, J.M., Schaid, D.J., Weinshilboum, R.M., and Thibodeau, S.N. (2007). Human SULT1A1 gene: copy number differences and functional implications. Human molecular genetics *16*, 463-470.

Hemmerich, S., and Rosen, S.D. (2000). Carbohydrate sulfotransferases in lymphocyte homing. Glycobiology *10*, 849-856.

Hildebrandt, M.A., Carrington, D.P., Thomae, B.A., Eckloff, B.W., Schaid, D.J., Yee, V.C., Weinshilboum, R.M., and Wieben, E.D. (2007). Genetic diversity and function in the human cytosolic sulfotransferases. The pharmacogenomics journal *7*, 133-143.

Hopkins, A.L. (2008). Network pharmacology: the next paradigm in drug discovery. Nature chemical biology *4*, 682-690.

Hortin, G.L., Schilling, M., and Graham, J.P. (1988). Inhibitors of the sulfation of proteins, glycoproteins, and proteoglycans. Biochemical and biophysical research communications *150*, 342-348.

Hrecka, K., Hao, C., Gierszewska, M., Swanson, S.K., Kesik-Brodacka, M., Srivastava, S., Florens, L., Washburn, M.P., and Skowronski, J. (2011). Vpx relieves inhibition of HIV-1 infection of macrophages mediated by the SAMHD1 protein. Nature *474*, 658-661.

Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and Applications of CRISPR-Cas9 for Genome Engineering. Cell *157*, 1262-1278.

Huh, A.J., and Kwon, Y.J. (2011). "Nanoantibiotics": a new paradigm for treating infectious diseases using nanomaterials in the antibiotics resistant era. Journal of controlled release : official journal of the Controlled Release Society *156*, 128-145.

Hunter, C.A., and Sibley, L.D. (2012). Modulation of innate immunity by Toxoplasma gondii virulence effectors. Nature Reviews Microbiology *10*, 766-778.

Hyde, J.E. (2002). Mechanisms of resistance of Plasmodium falciparum to antimalarial drugs. Microbes and infection / Institut Pasteur *4*, 165-174.

Ignowski, J.M., and Schaffer, D.V. (2004). Kinetic analysis and modeling of firefly luciferase as a quantitative reporter gene in live mammalian cells. Biotechnology and bioengineering *86*, 827-834.

Ikadai, H., Shaw Saliba, K., Kanzok, S.M., McLean, K.J., Tanaka, T.Q., Cao, J., Williamson, K.C., and Jacobs-Lorena, M. (2013). Transposon mutagenesis identifies genes essential for Plasmodium falciparum gametocytogenesis. Proceedings of the National Academy of Sciences of the United States of America *110*, E1676-1684.

Imielinski, M., Baldassano, R.N., Griffiths, A., Russell, R.K., Annese, V., Dubinsky, M., Kugathasan, S., Bradfield, J.P., Walters, T.D., Sleiman, P., Kim, C.E., Muise, A., Wang, K., Glessner, J.T., Saeed, S., Zhang, H., Frackelton, E.C., Hou, C., Flory, J.H., Otieno, G., Chiavacci, R.M., Grundmeier, R., Castro, M., Latiano, A., Dallapiccola, B., Stempak, J., Abrams, D.J., Taylor, K., McGovern, D., Western Regional Alliance for Pediatric, I.B.D., Silber, G., Wrobel, I., Quiros, A., International, I.B.D.G.C., Barrett, J.C., Hansoul, S., Nicolae, D.L., Cho, J.H., Duerr, R.H., Rioux, J.D., Brant, S.R., Silverberg, M.S., Taylor, K.D., Barmuda, M.M., Bitton, A., Dassopoulos, T., Datta, L.W., Green, T., Griffiths, A.M., Kistner, E.O., Murtha, M.T., Regueiro, M.D., Rotter, J.I., Schumm, L.P., Steinhart, A.H., Targan, S.R., Xavier, R.J., Consortium, N.I.G., Libioulle, C., Sandor, C., Lathrop, M., Belaiche, J., Dewit, O., Gut, I., Heath, S., Laukens, D., Mni, M., Rutgeerts, P., Van Gossum, A., Zelenika, D., Franchimont, D., Hugot, J.P., de Vos, M., Vermeire, S., Louis, E., Belgian-French, I.B.D.C., Wellcome Trust Case Control, C., Cardon, L.R., Anderson, C.A., Drummond, H., Nimmo, E., Ahmad, T., Prescott, N.J., Onnie, C.M., Fisher, S.A., Marchini, J., Ghori, J., Bumpstead, S., Gwillam, R., Tremelling, M., Delukas, P., Mansfield, J., Jewell, D., Satsangi, J., Mathew, C.G., Parkes, M., Georges, M., Daly, M.J., Heyman, M.B., Ferry, G.D., Kirschner, B., Lee, J., Essers, J., Grand, R., Stephens, M., Levine, A., Piccoli, D., Van Limbergen, J., Cucchiara, S., Monos, D.S., Guthery, S.L., Denson, L., Wilson, D.C., Grant, S.F., Daly, M., Silverberg, M.S., Satsangi, J., and Hakonarson, H. (2009). Common variants at five new loci associated with early-onset inflammatory bowel disease. Nature genetics *41*, 1335-1340.

Inacio, P., Zuzarte-Luis, V., Ruivo, M.T.G., Falkard, B., Nagaraj, N., Rooijers, K., Mann, M., Mair, G., Fidock, D.A., and Mota, M.M. (2015). Parasite-induced ER stress response in hepatocytes facilitates Plasmodium liver stage infection. EMBO reports *16*, 955-964.

Ingmundson, A., Alano, P., Matuschewski, K., and Silvestrini, F. (2014). Feeling at home from arrival to departure: protein export and host cell remodelling during Plasmodium liver stage and gametocyte maturation. Cell Microbiol *16*, 324-333.

Itoe, M.A., Sampaio, J.L., Cabal, G.G., Real, E., Zuzarte-Luis, V., March, S., Bhatia, S.N., Frischknecht, F., Thiele, C., Shevchenko, A., and Mota, M.M. (2014). Host cell phosphatidylcholine is a key mediator of malaria parasite survival during liver stage infection. Cell host & microbe *16*, 778-786.

Jamshidi, N., and Palsson, B.O. (2007). Investigating the metabolic capabilities of Mycobacterium tuberculosis H37Rv using the in silico strain iNJ661 and proposing alternative drug targets. BMC systems biology *1*, 26.

Janse, C.J., Franke-Fayard, B., Mair, G.R., Ramesar, J., Thiel, C., Engelmann, S., Matuschewski, K., van Gemert, G.J., Sauerwein, R.W., and Waters, A.P. (2006). High efficiency transfection of Plasmodium berghei facilitates novel selection procedures. Molecular and biochemical parasitology *145*, 60-70.

Jaramillo, M., Gomez, M.A., Larsson, O., Shio, M.T., Topisirovic, I., Contreras, I., Luxenburg, R., Rosenfeld, A., Colina, R., McMaster, R.W., Olivier, M., Costa-Mattioli, M., and Sonenberg, N. (2011). Leishmania repression of host translation through mTOR cleavage is required for parasite survival and infection. Cell host & microbe *9*, 331-341.

Jayaswal, S., Kamal, M.A., Dua, R., Gupta, S., Majumdar, T., Das, G., Kumar, D., and Rao, K.V.S. (2010). Identification of Host-Dependent Survival Factors for Intracellular Mycobacterium tuberculosis through an siRNA Screen. PLoS pathogens *6*.

Jeelani, G., Sato, D., Husain, A., Escueta-de Cadiz, A., Sugimoto, M., Soga, T., Suematsu, M., and Nozaki, T. (2012). Metabolic profiling of the protozoan parasite Entamoeba invadens revealed activation of unpredicted pathway during encystation. PloS one *7*, e37740.

Jensen, B.C., Sivam, D., Kifer, C.T., Myler, P.J., and Parsons, M. (2009). Widespread variation in transcript abundance within and across developmental stages of Trypanosoma brucei. BMC genomics *10*, 482.

Joice, R., Narasimhan, V., Montgomery, J., Sidhu, A.B., Oh, K., Meyer, E., Pierre-Louis, W., Seydel, K., Milner, D., Williamson, K., Wiegand, R., Ndiaye, D., Daily, J., Wirth, D., Taylor, T., Huttenhower, C., and Marti, M. (2013). Inferring developmental stage composition from gene expression in human malaria. PLoS computational biology *9*, e1003392.

Jones, N.G., Thomas, E.B., Brown, E., Dickens, N.J., Hammarton, T.C., and Mottram, J.C. (2014). Regulators of Trypanosoma brucei cell cycle progression and differentiation identified using a kinome-wide RNAi screen. PLoS pathogens *10*, e1003886.

Joyce, A.R., and Palsson, B.O. (2006). The model organism as a system: integrating 'omics' data sets. Nature reviews Molecular cell biology *7*, 198-210.

Kafsack, B.F., and Llinas, M. (2010). Eating at the table of another: metabolomics of host-parasite interactions. Cell host & microbe *7*, 90-99.

Karanis, P., and Aldeyarbi, H.M. (2011). Evolution of Cryptosporidium in vitro culture. International journal for parasitology *41*, 1231-1242.

Kaushansky, A., Ye, A.S., Austin, L.S., Mikolajczak, S.A., Vaughan, A.M., Camargo, N., Metzger, P.G., Douglass, A.N., MacBeath, G., and Kappe, S.H.I. (2013). Suppression of Host p53 Is Critical for Plasmodium Liver-Stage Infection. Cell reports *3*, 630-637.

Kawashima, H., Petryniak, B., Hiraoka, N., Mitoma, J., Huckaby, V., Nakayama, J., Uchimura, K., Kadomatsu, K., Muramatsu, T., Lowe, J.B., and Fukuda, M. (2005). N-acetylglucosamine-6-O-sulfotransferases 1 and 2 cooperatively control lymphocyte homing through L-selectin ligand biosynthesis in high endothelial venules. Nature immunology *6*, 1096-1104.

Ke, H., Lewis, I.A., Morrisey, J.M., McLean, K.J., Ganesan, S.M., Painter, H.J., Mather, M.W., Jacobs-Lorena, M., Llinas, M., and Vaidya, A.B. (2015). Genetic investigation of tricarboxylic acid metabolism during the Plasmodium falciparum life cycle. Cell reports *11*, 164-174.

Kelley, B.P., Yuan, B., Lewitter, F., Sharan, R., Stockwell, B.R., and Ideker, T. (2004). PathBLAST: a tool for alignment of protein interaction networks. Nucleic acids research *32*, W83-88.

Keshavjee, S., and Farmer, P.E. (2012). Tuberculosis, drug resistance, and the history of modern medicine. The New England journal of medicine *367*, 931-936.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome biology *14*, R36.

Kim, H.U., Kim, S.Y., Jeong, H., Kim, T.Y., Kim, J.J., Choy, H.E., Yi, K.Y., Rhee, J.H., and Lee, S.Y. (2011). Integrative genome-scale metabolic analysis of Vibrio vulnificus for drug targeting and discovery. Molecular systems biology *7*, 460.

King, A.V., Welter, B.H., Koushik, A.B., Gordon, L.N., and Temesvari, L.A. (2012). A genome-wide over-expression screen identifies genes involved in phagocytosis in the human protozoan parasite, Entamoeba histolytica. PloS one *7*, e43025.

Klaassen, C.D., and Boles, J.W. (1997). Sulfation and sulfotransferases 5: the importance of 3'-phosphoadenosine 5'-phosphosulfate (PAPS) in the regulation of sulfation. FASEB journal : official publication of the Federation of American Societies for Experimental Biology *11*, 404-418.

Kolev, N.G., Tschudi, C., and Ullu, E. (2011). RNA interference in protozoan parasites: achievements and challenges. Eukaryotic cell *10*, 1156-1163.

Konig, R., Chiang, C.Y., Tu, B.P., Yan, S.F., DeJesus, P.D., Romero, A., Bergauer, T., Orth, A., Krueger, U., Zhou, Y., and Chanda, S.K. (2007). A probability-based approach for the analysis of large-scale RNAi screens. Nature methods *4*, 847-849.

Konig, R., Zhou, Y., Elleder, D., Diamond, T.L., Bonamy, G.M., Irelan, J.T., Chiang, C.Y., Tu, B.P., De Jesus, P.D., Lilley, C.E., Seidel, S., Opaluch, A.M., Caldwell, J.S., Weitzman, M.D., Kuhen, K.L., Bandyopadhyay, S., Ideker, T., Orth, A.P., Miraglia, L.J., Bushman, F.D., Young, J.A., and Chanda, S.K. (2008). Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication. Cell *135*, 49-60.

Koushik, A.B., Welter, B.H., Rock, M.L., and Temesvari, L.A. (2014). A genomewide overexpression screen identifies genes involved in the phosphatidylinositol 3-kinase pathway in the human protozoan parasite Entamoeba histolytica. Eukaryotic cell *13*, 401-411.

Kramer, K.L., and Yost, H.J. (2003). Heparan sulfate core proteins in cell-cell signaling. Annual review of genetics *37*, 461-484.

Kronstad, J.W. (2006). Serial analysis of gene expression in eukaryotic pathogens. Infectious disorders drug targets *6*, 281-297.

Kuhen, K.L., Chatterjee, A.K., Rottmann, M., Gagaring, K., Borboa, R., Buenviaje, J., Chen, Z., Francek, C., Wu, T., Nagle, A., Barnes, S.W., Plouffe, D., Lee, M.C., Fidock, D.A., Graumans, W., van de Vegte-Bolmer, M., van Gemert, G.J., Wirjanata, G., Sebayang, B., Marfurt, J., Russell, B., Suwanarusk, R., Price, R.N., Nosten, F., Tungtaeng, A., Gettayacamin, M., Sattabongkot, J., Taylor, J., Walker, J.R., Tully, D., Patra, K.P., Flannery, E.L., Vinetz, J.M., Renia, L., Sauerwein, R.W., Winzeler, E.A., Glynne, R.J., and Diagana, T.T. (2014). KAF156 is an antimalarial clinical candidate with potential for use in prophylaxis, treatment, and prevention of disease transmission. Antimicrobial agents and chemotherapy *58*, 5060-5067.

Laguette, N., Sobhian, B., Casartelli, N., Ringeard, M., Chable-Bessia, C., Segeral, E., Yatim, A., Emiliani, S., Schwartz, O., and Benkirane, M. (2011). SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx. Nature *474*, 654-657.

Lahav, T., Sivam, D., Volpin, H., Ronen, M., Tsigankov, P., Green, A., Holland, N., Kuzyk, M., Borchers, C., Zilberstein, D., and Myler, P.J. (2011). Multiple levels of gene regulation mediate differentiation of the intracellular pathogen Leishmania. FASEB journal : official publication of the Federation of American Societies for Experimental Biology *25*, 515-525.

Lam, S.S., Martell, J.D., Kamer, K.J., Deerinck, T.J., Ellisman, M.H., Mootha, V.K., and Ting, A.Y. (2015). Directed evolution of APEX2 for electron microscopy and proximity labeling. Nature methods *12*, 51-54.

Law, G.L., Korth, M.J., Benecke, A.G., and Katze, M.G. (2013). Systems virology: host-directed approaches to viral pathogenesis and drug targeting. Nature reviews Microbiology *11*, 455-466.

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., and Winzeler, E.A. (2003). Discovery of gene function by expression profiling of the malaria parasite life cycle. Science *301*, 1503-1508.

Lemieux, J.E., Gomez-Escobar, N., Feller, A., Carret, C., Amambua-Ngwa, A., Pinches, R., Day, F., Kyes, S.A., Conway, D.J., Holmes, C.C., and Newbold, C.I. (2009). Statistical estimation of cell-cycle progression and lineage commitment in Plasmodium falciparum reveals a homogeneous pattern of transcription in ex vivo culture. Proceedings of the National Academy of Sciences of the United States of America *106*, 7559-7564.

Leroux-Roels, G., Leroux-Roels, I., Clement, F., Ofori-Anyinam, O., Lievens, M., Jongert, E., Moris, P., Ballou, W.R., and Cohen, J. (2014). Evaluation of the immune response to RTS,S/AS01 and RTS,S/AS02 adjuvanted vaccines: randomized, double-blind study in malaria-naive adults. Hum Vaccin Immunother *10*, 2211-2219.

Levin, J.Z., Yassour, M., Adiconis, X., Nusbaum, C., Thompson, D.A., Friedman, N., Gnirke, A., and Regev, A. (2010). Comprehensive comparative analysis of strand-specific RNA sequencing methods. Nature methods *7*, 709-715.

Lewis, N.E., Nagarajan, H., and Palsson, B.O. (2012a). Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. Nature Reviews Microbiology *10*, 291-305.

Lewis, N.E., Nagarajan, H., and Palsson, B.O. (2012b). Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. Nature reviews Microbiology *10*, 291-305.

Li, J.V., Wang, Y., Saric, J., Nicholson, J.K., Dirnhofer, S., Singer, B.H., Tanner, M., Wittlin, S., Holmes, E., and Utzinger, J. (2008). Global metabolic responses of NMRI mice to an experimental Plasmodium berghei infection. Journal of proteome research *7*, 3948-3956.

Li, S., Liberman, L.M., Mukherjee, N., Benfey, P.N., and Ohler, U. (2013). Integrated detection of natural antisense transcripts using strand-specific RNA sequencing data. Genome research *23*, 1730-1739.

Liehl, P., Zuzarte-Luis, V., Chan, J., Zillinger, T., Baptista, F., Carapau, D., Konert, M., Hanson, K.K., Carret, C., Lassnig, C., Muller, M., Kalinke, U., Saeed, M., Chora, A.F., Golenbock, D.T., Strobl, B., Prudencio, M., Coelho, L.P., Kappe, S.H., Superti-Furga, G., Pichlmair, A., Vigario, A.M., Rice, C.M., Fitzgerald, K.A., Barchet, W., and Mota, M.M. (2014). Host-cell sensors for Plasmodium activate innate immunity against liver-stage infection. Nature medicine *20*, 47-53.

Ling, L.L., Schneider, T., Peoples, A.J., Spoering, A.L., Engels, I., Conlon, B.P., Mueller, A., Schaberle, T.F., Hughes, D.E., Epstein, S., Jones, M., Lazarides, L., Steadman, V.A., Cohen, D.R., Felix, C.R., Fetterman, K.A., Millett, W.P., Nitti, A.G., Zullo, A.M., Chen, C., and Lewis, K. (2015). A new antibiotic kills pathogens without detectable resistance (vol 517, pg 455, 2015). Nature *520*.

Logan-Klumpler, F.J., De Silva, N., Boehme, U., Rogers, M.B., Velarde, G., McQuillan, J.A., Carver, T., Aslett, M., Olsen, C., Subramanian, S., Phan, I., Farris, C., Mitra, S., Ramasamy, G., Wang, H., Tivey, A., Jackson, A., Houston, R., Parkhill, J., Holden, M., Harb, O.S., Brunk, B.P., Myler, P.J., Roos, D., Carrington, M., Smith, D.F., Hertz-Fowler, C., and Berriman, M. (2012). GeneDB--an annotation database for pathogens. Nucleic acids research *40*, D98-108.

MacRae, J.I., Sheiner, L., Nahid, A., Tonkin, C., Striepen, B., and McConville, M.J. (2012). Mitochondrial metabolism of glucose and glutamine is required for intracellular growth of Toxoplasma gondii. Cell host & microbe *12*, 682-692.

Mair, G.R., Braks, J.A., Garver, L.S., Wiegant, J.C., Hall, N., Dirks, R.W., Khan, S.M., Dimopoulos, G., Janse, C.J., and Waters, A.P. (2006). Regulation of sexual development of Plasmodium by translational repression. Science *313*, 667-669.

Malmquist, N.A., Moss, T.A., Mecheri, S., Scherf, A., and Fuchter, M.J. (2012). Small-molecule histone methyltransferase inhibitors display rapid antimalarial activity against all blood stage forms in Plasmodium falciparum. Proceedings of the National Academy of Sciences of the United States of America *109*, 16708-16713.

Manfredi, R., Calza, L., Marinacci, G., Cascavilla, A., Colangeli, V., Salvadori, C., Martelli, G., Appolloni, L., Puggioli, C., and Viale, P. (2015). A prospective evaluation of maraviroc administration in patients with advanced HIV disease and multiple comorbidities: focus on efficacy and tolerability issues. Le infezioni in medicina : rivista periodica di eziologia, epidemiologia, diagnostica, clinica e terapia delle patologie infettive *23*, 36-43.

Mariani, R., Chen, D., Schrofelbauer, B., Navarro, F., Konig, R., Bollman, B., Munk, C., Nymark-McMahon, H., and Landau, N.R. (2003). Species-specific exclusion of APOBEC3G from HIV-1 virions by Vif. Cell *114*, 21-31.

Mazumdar, J., and Striepen, B. (2007). Make it or take it: fatty acid metabolism of apicomplexan parasites. Eukaryotic cell *6*, 1727-1735.

McMahon, M.A., Jilek, B.L., Brennan, T.P., Shen, L., Zhou, Y., Wind-Rotolo, M., Xing, S., Bhat, S., Hale, B., Hegarty, R., Chong, C.R., Liu, J.O., Siliciano, R.F., and Thio, C.L. (2007). The HBV drug entecavir - effects on HIV-1 replication and resistance. The New England journal of medicine *356*, 2614-2621.

McNamara, C.W., Lee, M.C., Lim, C.S., Lim, S.H., Roland, J., Nagle, A., Simon, O., Yeung, B.K., Chatterjee, A.K., McCormack, S.L., Manary, M.J., Zeeman, A.M., Dechering, K.J., Kumar, T.R., Henrich, P.P., Gagaring, K., Ibanez, M., Kato, N., Kuhen, K.L., Fischli, C., Rottmann, M., Plouffe, D.M., Bursulaya, B., Meister, S., Rameh, L., Trappe, J., Haasen, D., Timmerman, M., Sauerwein, R.W., Suwanarusk, R., Russell, B., Renia, L., Nosten, F., Tully, D.C., Kocken, C.H., Glynne, R.J., Bodenreider, C., Fidock, D.A., Diagana, T.T., and Winzeler, E.A. (2013). Targeting Plasmodium PI(4)K to eliminate malaria. Nature *504*, 248-253.

Meister, S., Plouffe, D.M., Kuhen, K.L., Bonamy, G.M., Wu, T., Barnes, S.W., Bopp, S.E., Borboa, R., Bright, A.T., Che, J., Cohen, S., Dharia, N.V., Gagaring, K., Gettayacamin, M., Gordon, P., Groessl, T., Kato, N., Lee, M.C., McNamara, C.W., Fidock, D.A., Nagle, A., Nam, T.G., Richmond, W., Roland, J., Rottmann, M., Zhou, B., Froissard, P., Glynne, R.J., Mazier, D., Sattabongkot, J., Schultz, P.G., Tuntland, T., Walker, J.R., Zhou, Y., Chatterjee, A., Diagana, T.T., and Winzeler, E.A. (2011). Imaging of Plasmodium liver stages to drive next-generation antimalarial drug discovery. Science *334*, 1372-1377.

Militello, K.T., Refour, P., Comeaux, C.A., and Duraisingh, M.T. (2008). Antisense RNA and RNAi in protozoan parasites: working hard or hardly working? Molecular and biochemical parasitology *157*, 117-126.

Mishra, B.B., and Tiwari, V.K. (2011). Natural products: an evolving role in future drug discovery. European journal of medicinal chemistry *46*, 4769-4807.

Mohr, I., and Sonenberg, N. (2012). Host translation at the nexus of infection and immunity. Cell host & microbe *12*, 470-483.

Mondor, I., Ugolini, S., and Sattentau, Q.J. (1998). Human immunodeficiency virus type 1 attachment to HeLa CD4 cells is CD4 independent and gp120 dependent and requires cell surface heparans. Journal of virology *72*, 3623-3634.

Monnerat, S., Clucas, C., Brown, E., Mottram, J.C., and Hammarton, T.C. (2009). Searching for novel cell cycle regulators in Trypanosoma brucei with an RNA interference screen. BMC research notes *2*, 46.

Mony, B.M., MacGregor, P., Ivens, A., Rojas, F., Cowton, A., Young, J., Horn, D., and Matthews, K. (2014). Genome-wide dissection of the quorum sensing signalling pathway in Trypanosoma brucei. Nature *505*, 681-685.

Mota, M.M., Pradel, G., Vanderberg, J.P., Hafalla, J.C., Frevert, U., Nussenzweig, R.S., Nussenzweig, V., and Rodriguez, A. (2001). Migration of Plasmodium sporozoites through cells before infection. Science *291*, 141-144.

Mueller, I., Galinski, M.R., Baird, J.K., Carlton, J.M., Kochar, D.K., Alonso, P.L., and del Portillo, H.A. (2009). Key gaps in the knowledge of Plasmodium vivax, a neglected human malaria parasite. The Lancet Infectious diseases *9*, 555-566.

Munch, J., Rajan, D., Schindler, M., Specht, A., Rucker, E., Novembre, F.J., Nerrienet, E., Muller-Trutwin, M.C., Peeters, M., Hahn, B.H., and Kirchhoff, F. (2007). Nef-mediated enhancement of virion infectivity and stimulation of viral replication are fundamental properties of primate lentiviruses. Journal of virology *81*, 13852-13864.

Murry, J.P., Godoy, J., Mukim, A., Swann, J., Bruce, J.W., Ahlquist, P., Bosque, A., Planelles, V., Spina, C.A., and Young, J.A. (2014). Sulfonation pathway inhibitors block reactivation of latent HIV-1. Virology *471-473C*, 1-12.

Nadanaka, S., Ishida, M., Ikegami, M., and Kitagawa, H. (2008). Chondroitin 4-O-sulfotransferase-1 modulates Wnt-3a signaling through control of E disaccharide expression of chondroitin sulfate. The Journal of biological chemistry *283*, 27333-27343.

Nelson, M.M., Jones, A.R., Carmen, J.C., Sinai, A.P., Burchmore, R., and Wastling, J.M. (2008). Modulation of the host cell proteome by the intracellular apicomplexan parasite Toxoplasma gondii. Infection and immunity *76*, 828-844.

Ng Hublin, J.S., Ryan, U., Trengove, R., and Maker, G. (2013). Metabolomic profiling of faecal extracts from Cryptosporidium parvum infection in experimental mouse models. PloS one *8*, e77803.

Nielsen, T.E., and Schreiber, S.L. (2008). Towards the optimal screening collection: a synthesis strategy. Angewandte Chemie *47*, 48-56.

Niemann, M., Wiese, S., Mani, J., Chanfon, A., Jackson, C., Meisinger, C., Warscheid, B., and Schneider, A. (2013). Mitochondrial outer membrane proteome of Trypanosoma brucei reveals novel factors required to maintain mitochondrial morphology. Molecular & cellular proteomics : MCP *12*, 515-528.

Nilsen, A., LaCrue, A.N., White, K.L., Forquer, I.P., Cross, R.M., Marfurt, J., Mather, M.W., Delves, M.J., Shackleford, D.M., Saenz, F.E., Morrisey, J.M., Steuten, J., Mutka, T., Li, Y., Wirjanata, G., Ryan, E., Duffy, S., Kelly, J.X., Sebayang, B.F., Zeeman, A.M., Noviyanti, R., Sinden, R.E., Kocken, C.H., Price, R.N., Avery, V.M., Angulo-Barturen, I., Jimenez-Diaz, M.B., Ferrer, S., Herreros, E., Sanz, L.M., Gamo, F.J., Bathurst, I., Burrows, J.N., Siegl, P., Guy, R.K., Winter, R.W., Vaidya, A.B., Charman, S.A., Kyle, D.E., Manetsch, R., and Riscoe, M.K. (2013). Quinolone-3-diarylethers: a new class of antimalarial drug. Science translational medicine *5*, 177ra137.

Ning, B., Nowell, S., Sweeney, C., Ambrosone, C.B., Williams, S., Miao, X., Liang, G., Lin, D., Stone, A., Ratnasinghe, D.L., Manjanatha, M., Lang, N.P., and Kadlubar, F.F. (2005). Common genetic polymorphisms in the 5'-flanking region of the SULT1A1 gene: haplotypes and their association with platelet enzymatic activity. Pharmacogenetics and genomics *15*, 465-473.

Nobutaka Kato, E.C., Tomoyo Sakata-Kato, Micah Maetani, Jessica Bastien,, Victoria Corey, D.C., Emily R. Derbyshire, Gillian Dornan, Sandra Duffy, Sean Eckley,, Karin MJ Koolen, T.A.L., Amanda K. Lukens, Emily Lund, Sandra Riera, Bennett C., and Meier, J.M., Branko Mitasev, Eli L. Moss, Morgane Sayes, Yvonne VanGessel, Mathias, J. Wawer, Takashi Yoshinaga, Anne-Marie Zeeman, Vicky M. Avery, Sangeeta N. Bhatia, John, E. Burke, Flaminia Catteruccia, Jon C. Clardy, Paul A. Clemons, Koen J. Dechering10, Jeremy R. Duvall, Michael A. Foley, Fabian Gusovsky, Clemens H. M. Kocken, Marshall L. Morningstar, Benito Munoz, Daniel E. Neafsey, Elizabeth A. Winzeler, Dyann F. Wirth, Christina A. Scherer, Stuart L. Schreiber Diversity synthesis yields multistage antimalarial inhibitors including of a novel target that results in low single-dose cures in mice. (In review).

Nowell, S., Ratnasinghe, D.L., Ambrosone, C.B., Williams, S., Teague-Ross, T., Trimble, L., Runnels, G., Carrol, A., Green, B., Stone, A., Johnson, D., Greene, G., Kadlubar, F.F., and Lang, N.P. (2004). Association of SULT1A1 phenotype and genotype with prostate cancer risk in African-Americans and

Caucasians. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology *13*, 270-276.

O'Halloran, A.M., Patterson, C.C., Horan, P., Maree, A., Curtin, R., Stanton, A., McKeown, P.P., and Shields, D.C. (2009). Genetic polymorphisms in platelet-related proteins and coronary artery disease: investigation of candidate genes, including N-acetylgalactosaminyltransferase 4 (GALNT4) and sulphotransferase 1A1/2 (SULT1A1/2). Journal of thrombosis and thrombolysis *27*, 175-184.

Oehring, S.C., Woodcroft, B.J., Moes, S., Wetzel, J., Dietz, O., Pulfer, A., Dekiwadia, C., Maeser, P., Flueck, C., Witmer, K., Brancucci, N.M., Niederwieser, I., Jenoe, P., Ralph, S.A., and Voss, T.S. (2012). Organellar proteomics reveals hundreds of novel nuclear proteins in the malaria parasite Plasmodium falciparum. Genome biology *13*, R108.

Olszewski, K.L., Morrisey, J.M., Wilinski, D., Burns, J.M., Vaidya, A.B., Rabinowitz, J.D., and Llinas, M. (2009). Host-parasite interactions revealed by Plasmodium falciparum metabolomics. Cell host & microbe *5*, 191-199.

Pache, L., Konig, R., and Chanda, S.K. (2011). Identifying HIV-1 host cell factors by genome-scale RNAi screening. Methods *53*, 3-12.

Palsson, B. (2004). Two-dimensional annotation of genomes. Nature biotechnology *22*, 1218-1219.

Pellegrini, M., Marcotte, E.M., Thompson, M.J., Eisenberg, D., and Yeates, T.O. (1999). Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. Proceedings of the National Academy of Sciences of the United States of America *96*, 4285-4288.

Peterson, D.S., Walliker, D., and Wellems, T.E. (1988). Evidence that a point mutation in dihydrofolate reductase-thymidylate synthase confers resistance to pyrimethamine in falciparum malaria. Proceedings of the National Academy of Sciences of the United States of America *85*, 9114-9118.

Pillai, A.B., Xu, W., Zhang, O., and Zhang, K. (2012). Sphingolipid degradation in Leishmania (Leishmania) amazonensis. PLoS neglected tropical diseases *6*, e1944.

Pino, P., Sebastian, S., Kim, E.A., Bush, E., Brochet, M., Volkmann, K., Kozlowski, E., Llinas, M., Billker, O., and Soldati-Favre, D. (2012). A tetracycline-repressible transactivator system to study essential genes in malaria parasites. Cell host & microbe *12*, 824-834.

Pittman, K.J., Aliota, M.T., and Knoll, L.J. (2014). Dual transcriptional profiling of mice and Toxoplasma gondii during acute and chronic infection. BMC genomics *15*, 806.

Plata, G., Hsiao, T.L., Olszewski, K.L., Llinas, M., and Vitkup, D. (2010). Reconstruction and flux-balance analysis of the Plasmodium falciparum metabolic network. Molecular systems biology *6*, 408.

Plouffe, D., Brinker, A., McNamara, C., Henson, K., Kato, N., Kuhen, K., Nagle, A., Adrian, F., Matzen, J.T., Anderson, P., Nam, T.G., Gray, N.S., Chatterjee, A., Janes, J., Yan, S.F., Trager, R., Caldwell, J.S., Schultz, P.G., Zhou, Y., and Winzeler, E.A. (2008). In silico activity profiling reveals the mechanism of action of antimalarials discovered in a high-throughput screen. Proceedings of the National Academy of Sciences of the United States of America *105*, 9059-9064.

Prommana, P., Uthaipibull, C., Wongsombat, C., Kamchonwongpaisan, S., Yuthavong, Y., Knuepfer, E., Holder, A.A., and Shaw, P.J. (2013). Inducible knockdown of Plasmodium gene expression using the glmS ribozyme. PloS one *8*, e73783.

Prudencio, M., and Mota, M.M. (2013). Targeting host factors to circumvent anti-malarial drug resistance. Current pharmaceutical design *19*, 290-299.

Prudencio, M., Rodrigues, C.D., Ataide, R., and Mota, M.M. (2008a). Dissecting in vitro host cell infection by Plasmodium sporozoites using flow cytometry. Cellular microbiology *10*, 218-224.

Prudencio, M., Rodrigues, C.D., Hannus, M., Martin, C., Real, E., Goncalves, L.A., Carret, C., Dorkin, R., Rohl, I., Jahn-Hoffmann, K., Luty, A.J.F., Sauerwein, R., Echeverri, C.J., and Mota, M.M. (2008b). Kinome-Wide RNAi Screen Implicates at Least 5 Host Hepatocyte Kinases in Plasmodium Sporozoite Infection. PLoS pathogens *4*.

Prudencio, M., Rodriguez, A., and Mota, M.M. (2006a). The silent path to thousands of merozoites: the Plasmodium liver stage. Nature reviews Microbiology *4*, 849-856.

Prudencio, M., Rodriguez, A., and Mota, M.M. (2006b). The silent path to thousands of merozoites: the Plasmodium liver stage. Nature Reviews Microbiology *4*, 849-856.

Raghunathan, A., Shin, S., and Daefler, S. (2010). Systems approach to investigating host-pathogen interactions in infections with the biothreat agent Francisella. Constraints-based model of Francisella tularensis. BMC systems biology *4*, 118.

Ratcliff, A.N., Shi, W., and Arts, E.J. (2013). HIV-1 resistance to maraviroc conferred by a CD4 binding site mutation in the envelope glycoprotein gp120. Journal of virology *87*, 923-934.

Reader, J., Botha, M., Theron, A., Lauterbach, S.B., Rossouw, C., Engelbrecht, D., Wepener, M., Smit, A., Leroy, D., Mancama, D., Coetzer, T.L., and Birkholtz, L.M. (2015). Nowhere to hide: interrogating different metabolic parameters of Plasmodium falciparum gametocytes in a transmission blocking drug discovery pipeline towards malaria elimination. Malaria journal *14*, 213.

Rhee, H.W., Zou, P., Udeshi, N.D., Martell, J.D., Mootha, V.K., Carr, S.A., and Ting, A.Y. (2013). Proteomic Mapping of Mitochondria in Living Cells via Spatially Restricted Enzymatic Tagging. Science *339*, 1328-1331.

Riches, Z., Stanley, E.L., Bloomer, J.C., and Coughtrie, M.W. (2009). Quantitative evaluation of the expression and activity of five major sulfotransferases (SULTs) in human tissues: the SULT "pie". Drug metabolism and disposition: the biological fate of chemicals *37*, 2255-2261.

Riley, E.M., and Stewart, V.A. (2013). Immune mechanisms in malaria: new insights in vaccine development. Nature medicine *19*, 168-178.

Roberts, S.B., Robichaux, J.L., Chavali, A.K., Manque, P.A., Lee, V., Lara, A.M., Papin, J.A., and Buck, G.A. (2009). Proteomic and network analysis characterize stage-specific metabolism in Trypanosoma cruzi. BMC systems biology *3*, 52.

Roderiquez, G., Oravecz, T., Yanagishita, M., Bou-Habib, D.C., Mostowski, H., and Norcross, M.A. (1995). Mediation of human immunodeficiency virus type 1 binding by interaction of cell surface heparan sulfate proteoglycans with the V3 region of envelope gp120-gp41. Journal of virology *69*, 2233-2239.

Ross, L.S., Gamo, F.J., Lafuente-Monasterio, M.J., Singh, O.M., Rowland, P., Wiegand, R.C., and Wirth, D.F. (2014). In vitro resistance selections for Plasmodium falciparum dihydroorotate dehydrogenase inhibitors give mutants with multiple point mutations in the drug-binding site and altered growth. The Journal of biological chemistry *289*, 17980-17995.

Rottmann, M., McNamara, C., Yeung, B.K., Lee, M.C., Zou, B., Russell, B., Seitz, P., Plouffe, D.M., Dharia, N.V., Tan, J., Cohen, S.B., Spencer, K.R., Gonzalez-Paez, G.E., Lakshminarayana, S.B., Goh, A., Suwanarusk, R., Jegla, T., Schmitt, E.K., Beck, H.P., Brun, R., Nosten, F., Renia, L., Dartois, V., Keller, T.H., Fidock, D.A., Winzeler, E.A., and Diagana, T.T. (2010). Spiroindolones, a potent compound class for the treatment of malaria. Science *329*, 1175-1180.

Rout, S., Warhurst, D.C., Suar, M., and Mahapatra, R.K. (2015). In silico comparative genomics analysis of Plasmodium falciparum for the identification of putative essential genes and therapeutic candidates. Journal of microbiological methods *109*, 1-8.

Ruiz, A., and Russell, S.J. (2012). A new paradigm in viral resistance. Cell research *22*, 1515-1517.

Ryoo, J., Choi, J., Oh, C., Kim, S., Seo, M., Kim, S.Y., Seo, D., Kim, J., White, T.E., Brandariz-Nunez, A., Diaz-Griffero, F., Yun, C.H., Hollenbaugh, J.A., Kim, B., Baek, D., and Ahn, K. (2014). The ribonuclease activity of SAMHD1 is required for HIV-1 restriction. Nature medicine *20*, 936-941.

Sachs, J., and Malaney, P. (2002). The economic and social burden of malaria. Nature *415*, 680-685.

Sacks, D., and Sher, A. (2002). Evasion of innate immunity by parasitic protozoa. Nature immunology *3*, 1041-1047.

Saeij, J.P.J., Coller, S., Boyle, J.P., Jerome, M.E., White, M.W., and Boothroyd, J.C. (2007). Toxoplasma co-opts host gene expression by injection of a polymorphic kinase homologue. Nature *445*, 324-327.

Saghatelian, A., and Cravatt, B.F. (2005). Assignment of protein function in the postgenomic era. Nature chemical biology *1*, 130-142.

Sakata, T., and Winzeler, E.A. (2007). Genomics, systems biology and drug development for infectious diseases. Molecular bioSystems *3*, 841-848.

Salman, E.D., Kadlubar, S.A., and Falany, C.N. (2009). Expression and localization of cytosolic sulfotransferase (SULT) 1A1 and SULT1A3 in normal human brain. Drug metabolism and disposition: the biological fate of chemicals *37*, 706-709.

Salzemann, J., Kasam, V., Jacq, N., Maass, A., Schwichtenberg, H., and Breton, V. (2007). Grid enabled high throughput virtual screening against four different targets implicated in malaria. Studies in health technology and informatics *126*, 47-54.

Sams-Dodd, F. (2005). Target-based drug discovery: is something wrong? Drug Discov Today *10*, 139-147.

Savoia, D. (2015). New antimicrobial approaches: reuse of old drugs. Current drug targets.

Schmidt, N.W., Butler, N.S., Badovinac, V.P., and Harty, J.T. (2010). Extreme CD8 T cell requirements for anti-malarial liver-stage immunity following immunization with radiation attenuated sporozoites. PLoS pathogens *6*, e1000998.

Schmittgen, T.D., and Livak, K.J. (2008). Analyzing real-time PCR data by the comparative C(T) method. Nature protocols *3*, 1101-1108.

Schneider, M.V. (2013). Defining systems biology: a brief overview of the term and field. Methods in molecular biology *1021*, 1-11.

Schwegmann, A., and Brombacher, F. (2008). Host-directed drug targeting of factors hijacked by pathogens. Science signaling *1*, re8.

Seibert, C., Cadene, M., Sanfiz, A., Chait, B.T., and Sakmar, T.P. (2002). Tyrosine sulfation of CCR5 N-terminal peptide by tyrosylprotein sulfotransferases 1 and 2 follows a discrete pattern and temporal sequence. Proceedings of the National Academy of Sciences of the United States of America *99*, 11031-11036.

Shah, P., Tiwari, S., and Siddiqi, M.I. (2015). Recent progress in the identification and development of anti-malarial agents using virtual screening based approaches. Combinatorial chemistry & high throughput screening *18*, 257-268.

Shameer, S., Logan-Klumpler, F.J., Vinson, F., Cottret, L., Merlet, B., Achcar, F., Boshart, M., Berriman, M., Breitling, R., Bringaud, F., Butikofer, P., Cattanach, A.M., Bannerman-Chukualim, B., Creek, D.J., Crouch, K., de Koning, H.P., Denise, H., Ebikeme, C., Fairlamb, A.H., Ferguson, M.A., Ginger, M.L., Hertz-Fowler, C., Kerkhoven, E.J., Maser, P., Michels, P.A., Nayak, A., Nes, D.W., Nolan, D.P., Olsen, C., Silva-Franco, F., Smith, T.K., Taylor, M.C., Tielens, A.G., Urbaniak, M.D., van Hellemond, J.J., Vincent, I.M., Wilkinson, S.R., Wyllie, S., Opperdoes, F.R., Barrett, M.P., and Jourdan, F. (2015). TrypanoCyc: a community-led biochemical pathways database for Trypanosoma brucei. Nucleic acids research *43*, D637-644.

Sharma, S., DeOliveira, R.B., Kalantari, P., Parroche, P., Goutagny, N., Jiang, Z., Chan, J., Bartholomeu, D.C., Lauw, F., Hall, J.P., Barber, G.N., Gazzinelli, R.T., Fitzgerald, K.A., and Golenbock, D.T. (2011). Innate immune recognition of an AT-rich stem-loop DNA motif in the Plasmodium falciparum genome. Immunity *35*, 194-207.

Sheehy, A.M., Gaddis, N.C., and Malim, M.H. (2003). The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif. Nature medicine *9*, 1404-1407.

Sidhu, A.B., Verdier-Pinard, D., and Fidock, D.A. (2002). Chloroquine resistance in Plasmodium falciparum malaria parasites conferred by pfcrt mutations. Science *298*, 210-213.

Siegel, T.N., Hon, C.C., Zhang, Q., Lopez-Rubio, J.J., Scheidig-Benatar, C., Martins, R.M., Sismeiro, O., Coppee, J.Y., and Scherf, A. (2014). Strand-specific RNA-Seq reveals widespread and developmentally regulated transcription of natural antisense transcripts in Plasmodium falciparum. BMC genomics *15*, 150.

Silber, A.M., and Pereira, C.A. (2012). Assignment of putative functions to membrane "hypothetical proteins" from the Trypanosoma cruzi genome. The Journal of membrane biology *245*, 125-129.

Silva, A.M., Cordeiro-da-Silva, A., and Coombs, G.H. (2011). Metabolic variation during development in culture of Leishmania donovani promastigotes. PLoS neglected tropical diseases *5*, e1451.

Silvie, O., Greco, C., Franetich, J.F., Dubart-Kupperschmitt, A., Hannoun, L., van Gemert, G.J., Sauerwein, R.W., Levy, S., Boucheix, C., Rubinstein, E., and Mazier, D. (2006). Expression of human CD81 differently affects host cell susceptibility to malaria sporozoites depending on the Plasmodium species. Cellular microbiology *8*, 1134-1146.

Singh, A.P., Buscaglia, C.A., Wang, Q., Levay, A., Nussenzweig, D.R., Walker, J.R., Winzeler, E.A., Fujii, H., Fontoura, B.M.A., and Nussenzweig, V. (2007). Plasmodium circumsporozoite protein promotes the development of the liver stages of the parasite. Cell *131*, 492-504.

Sinka, M.E., Bangs, M.J., Manguin, S., Rubio-Palis, Y., Chareonviriyaphap, T., Coetzee, M., Mbogo, C.M., Hemingway, J., Patil, A.P., Temperley, W.H., Gething, P.W., Kabaria, C.W., Burkot, T.R., Harbach, R.E., and Hay, S.I. (2012). A global map of dominant malaria vectors. Parasit Vectors *5*, 69.

Sinnis, P., De La Vega, P., Coppi, A., Krzych, U., and Mota, M.M. (2013). Quantification of sporozoite invasion, migration, and development by microscopy and flow cytometry. Methods in molecular biology *923*, 385-400.

Snow, R.W., Guerra, C.A., Noor, A.M., Myint, H.Y., and Hay, S.I. (2005). The global distribution of clinical episodes of Plasmodium falciparum malaria. Nature *434*, 214-217.

Spangenberg, T., Burrows, J.N., Kowalczyk, P., McDonald, S., Wells, T.N., and Willis, P. (2013). The open access malaria box: a drug discovery catalyst for neglected diseases. PloS one *8*, e62906.

Srivastava, I.K., Morrisey, J.M., Darrouzet, E., Daldal, F., and Vaidya, A.B. (1999). Resistance mutations reveal the atovaquone-binding domain of cytochrome b in malaria parasites. Molecular microbiology *33*, 704-711.

Suthram, S., Sittler, T., and Ideker, T. (2005). The Plasmodium protein network diverges from those of other eukaryotes. Nature *438*, 108-112.

Swann, J., Jamshidi, N., Lewis, N.E., and Winzeler, E.A. (2015). Systems analysis of host-parasite interactions. Wires Syst Biol Med *7*, 381-400.

Swinney, D.C. (2013). Phenotypic vs. target-based drug discovery for first-in-class medicines. Clinical pharmacology and therapeutics *93*, 299-301.

Tan, D.S. (2005). Diversity-oriented synthesis: exploring the intersections between chemistry and biology. Nature chemical biology *1*, 74-84.

Tan, K.R., Magill, A.J., Parise, M.E., and Arguin, P.M. (2011). Doxycycline for malaria chemoprophylaxis and treatment: report from the CDC expert meeting on malaria chemoprophylaxis. The American journal of tropical medicine and hygiene *84*, 517-531.

Tang, D., Rundle, A., Mooney, L., Cho, S., Schnabel, F., Estabrook, A., Kelly, A., Levine, R., Hibshoosh, H., and Perera, F. (2003). Sulfotransferase 1A1 (SULT1A1) polymorphism, PAH-DNA adduct levels in breast tissue and breast cancer risk in a case-control study. Breast cancer research and treatment *78*, 217-222.

Teubner, W., Meinl, W., Florian, S., Kretzschmar, M., and Glatt, H. (2007). Identification and localization of soluble sulfotransferases in the human gastrointestinal tract. The Biochemical journal *404*, 207-215.

Treeck, M., Sanders, J.L., Elias, J.E., and Boothroyd, J.C. (2011). The phosphoproteomes of Plasmodium falciparum and Toxoplasma gondii reveal unusual adaptations within and beyond the parasites' boundaries. Cell host & microbe *10*, 410-419.

Tsigankov, P., Gherardini, P.F., Helmer-Citterich, M., Spath, G.F., and Zilberstein, D. (2013). Phosphoproteomic analysis of differentiating Leishmania parasites reveals a unique stage-specific phosphorylation motif. Journal of proteome research *12*, 3405-3412.

Urbaniak, M.D., Martin, D.M., and Ferguson, M.A. (2013). Global quantitative SILAC phosphoproteomics reveals differential phosphorylation is widespread between the procyclic and bloodstream form lifecycle stages of Trypanosoma brucei. Journal of proteome research *12*, 2233-2244.

van der Schaar, H.M., van der Linden, L., Lanke, K.H.W., Strating, J.R.P.M., Purstinger, G., de Vries, E., de Haan, C.A.M., Neyts, J., and van Kuppeveld, F.J.M. (2012). Coxsackievirus mutants that can bypass host factor PI4KIII beta and the need for high levels of PI4P lipids for replication. Cell research *22*, 1576-1592.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., Zhang, Q., Kodira, C.D., Zheng, X.H., Chen, L., Skupski, M., Subramanian, G., Thomas, P.D., Zhang, J., Gabor Miklos, G.L., Nelson, C., Broder, S., Clark, A.G., Nadeau, J., McKusick, V.A., Zinder, N., Levine, A.J., Roberts, R.J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A.E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T.J., Higgins, M.E., Ji, R.R., Ke, Z., Ketchum, K.A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G.V., Milshina, N., Moore, H.M., Naik, A.K., Narayan, V.A., Neelam, B., Nusskern, D., Rusch, D.B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M.L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y.H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N.N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J.F., Guigo, R., Campbell, M.J., Sjolander, K.V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y.H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M.,

Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., and Zhu, X. (2001). The sequence of the human genome. Science *291*, 1304-1351.

Visvesvara, G.S., and Garcia, L.S. (2002). Culture of protozoan parasites. Clinical microbiology reviews *15*, 327-328.

Wagner, J.C., Platt, R.J., Goldfless, S.J., Zhang, F., and Niles, J.C. (2014). Efficient CRISPR-Cas9-mediated genome editing in Plasmodium falciparum. Nature methods *11*, 915-918.

Waisberg, M., Cerqueira, G.C., Yager, S.B., Francischetti, I.M.B., Lu, J.H., Gera, N., Srinivasan, P., Miura, K., Rada, B., Lukszo, J., Barbian, K.D., Leto, T.L., Porcella, S.F., Narum, D.L., El-Sayed, N., Miller, L.H., and Pierce, S.K. (2012). Plasmodium falciparum merozoite surface protein 1 blocks the proinflammatory protein S100P. Proceedings of the National Academy of Sciences of the United States of America *109*, 5429-5434.

Wang, Y., Spitz, M.R., Tsou, A.M., Zhang, K., Makan, N., and Wu, X. (2002). Sulfotransferase (SULT) 1A1 polymorphism as a predisposition factor for lung cancer: a case-control analysis. Lung cancer *35*, 137-142.

Wang, Y., Utzinger, J., Saric, J., Li, J.V., Burckhardt, J., Dirnhofer, S., Nicholson, J.K., Singer, B.H., Brun, R., and Holmes, E. (2008). Global metabolic responses of mice to Trypanosoma brucei brucei infection. Proceedings of the National Academy of Sciences of the United States of America *105*, 6127-6132.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nature reviews Genetics *10*, 57-63.

Want, E.J., Nordstrom, A., Morita, H., and Siuzdak, G. (2007). From exogenous to endogenous: the inevitable imprint of mass spectrometry in metabolomics. Journal of proteome research *6*, 459-468.

Warren, K., Warrilow, D., Meredith, L., and Harrich, D. (2009). Reverse Transcriptase and Cellular Factors: Regulators of HIV-1 Reverse Transcription. Viruses *1*, 873-894.

Wastling, J.M., Armstrong, S.D., Krishna, R., and Xia, D. (2012). Parasites, proteomes and systems: has Descartes' clock run out of time? Parasitology *139*, 1103-1118.

Wells, T.N., Burrows, J.N., and Baird, J.K. (2010a). Targeting the hypnozoite reservoir of Plasmodium vivax: the hidden obstacle to malaria elimination. Trends in parasitology *26*, 145-151.

Wells, T.N.C., Burrows, J.N., and Baird, J.K. (2010b). Targeting the hypnozoite reservoir of Plasmodium vivax: the hidden obstacle to malaria elimination. Trends in parasitology *26*, 145-151.

Westermann, A.J., Gorski, S.A., and Vogel, J. (2012a). Dual RNA-seq of pathogen and host. Nature reviews Microbiology *10*, 618-630.

Westermann, A.J., Gorski, S.A., and Vogel, J. (2012b). Dual RNA-seq of pathogen and host. Nature Reviews Microbiology *10*, 618-630.

Widmer, G., Corey, E.A., Stein, B., Griffiths, J.K., and Tzipori, S. (2000). Host cell apoptosis impairs Cryptosporidium parvum development in vitro. The Journal of parasitology *86*, 922-928.

Wilby, K.J., Lau, T.T., Gilchrist, S.E., and Ensom, M.H. (2012). Mosquirix (RTS,S): a novel vaccine for the prevention of Plasmodium falciparum malaria. Ann Pharmacother *46*, 384-393.

Winzeler, E.A. (2006). Applied systems biology and malaria. Nature reviews Microbiology *4*, 145-151.

Wittkop, L., Gunthard, H.F., de Wolf, F., Dunn, D., Cozzi-Lepri, A., de Luca, A., Kucherer, C., Obel, N., von Wyl, V., Masquelier, B., Stephan, C., Torti, C., Antinori, A., Garcia, F., Judd, A., Porter, K., Thiebaut, R., Castro, H., van Sighem, A.I., Colin, C., Kjaer, J., Lundgren, J.D., Paredes, R., Pozniak, A., Clotet, B., Phillips, A., Pillay, D., Chene, G., and EuroCoord, C.s.g. (2011). Effect of transmitted drug resistance on virological and immunological response to initial combination antiretroviral therapy for HIV (EuroCoord-CHAIN joint project): a European multicohort study. The Lancet Infectious diseases *11*, 363-371.

Wu, C., Orozco, C., Boyer, J., Leglise, M., Goodale, J., Batalov, S., Hodge, C.L., Haase, J., Janes, J., Huss, J.W., 3rd, and Su, A.I. (2009a). BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. Genome biology *10*, R130.

Wu, Y., Nelson, M.M., Quaile, A., Xia, D., Wastling, J.M., and Craig, A. (2009b). Identification of phosphorylated proteins in erythrocytes infected by the human malaria parasite Plasmodium falciparum. Malaria journal *8*, 105.

Yamagishi, J., Natori, A., Tolba, M.E., Mongan, A.E., Sugimoto, C., Katayama, T., Kawashima, S., Makalowski, W., Maeda, R., Eshita, Y., Tuda, J., and

Suzuki, Y. (2014). Interactive transcriptome analysis of malaria patients and infecting Plasmodium falciparum. Genome research *24*, 1433-1444.

Yates, J.R., Ruse, C.I., and Nakorchevsky, A. (2009). Proteomics by mass spectrometry: approaches, advances, and applications. Annual review of biomedical engineering *11*, 49-79.

Yeung, M.L., Houzet, L., Yedavalli, V.S., and Jeang, K.T. (2009). A genome-wide short hairpin RNA screening of jurkat T-cells for human proteins contributing to productive HIV-1 replication. The Journal of biological chemistry *284*, 19463-19473.

Young, J.A., Fivelman, Q.L., Blair, P.L., de la Vega, P., Le Roch, K.G., Zhou, Y., Carucci, D.J., Baker, D.A., and Winzeler, E.A. (2005). The Plasmodium falciparum sexual development transcriptome: a microarray analysis using ontology-based pattern identification. Molecular and biochemical parasitology *143*, 67-79.

Young, J.A., Johnson, J.R., Benner, C., Yan, S.F., Chen, K., Le Roch, K.G., Zhou, Y., and Winzeler, E.A. (2008). In silico discovery of transcription regulatory elements in Plasmodium falciparum. BMC genomics *9*, 70.

Yu, X., Dhakal, I.B., Beggs, M., Edavana, V.K., Williams, S., Zhang, X., Mercer, K., Ning, B., Lang, N.P., Kadlubar, F.F., and Kadlubar, S. (2010). Functional genetic variants in the 3'-untranslated region of sulfotransferase isoform 1A1 (SULT1A1) and their effect on enzymatic activity. Toxicological sciences : an official journal of the Society of Toxicology *118*, 391-403.

Yu, X., Yu, Y., Liu, B., Luo, K., Kong, W., Mao, P., and Yu, X.F. (2003). Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex. Science *302*, 1056-1060.

Zeeman, A.M., van Amsterdam, S.M., McNamara, C.W., Voorberg-van der Wel, A., Klooster, E.J., van den Berg, A., Remarque, E.J., Plouffe, D.M., van Gemert, G.J., Luty, A., Sauerwein, R., Gagaring, K., Borboa, R., Chen, Z., Kuhen, K., Glynne, R.J., Chatterjee, A.K., Nagle, A., Roland, J., Winzeler, E.A., Leroy, D., Campo, B., Diagana, T.T., Yeung, B.K., Thomas, A.W., and Kocken, C.H. (2014). KAI407, a potent non-8-aminoquinoline compound that kills Plasmodium cynomolgi early dormant liver stage parasites in vitro. Antimicrobial agents and chemotherapy *58*, 1586-1595.

Zhang, C., Xiao, B., Jiang, Y., Zhao, Y., Li, Z., Gao, H., Ling, Y., Wei, J., Li, S., Lu, M., Su, X.Z., Cui, H., and Yuan, J. (2014). Efficient editing of malaria parasite genome using the CRISPR/Cas9 system. mBio *5*, e01414-01414.

Zhang, H., Guo, F., Zhou, H., and Zhu, G. (2012). Transcriptome analysis reveals unique metabolic features in the Cryptosporidium parvum Oocysts associated with environmental survival and stresses. BMC genomics *13*, 647.

Zhang, O., Wilson, M.C., Xu, W., Hsu, F.F., Turk, J., Kuhlmann, F.M., Wang, Y., Soong, L., Key, P., Beverley, S.M., and Zhang, K. (2009). Degradation of host sphingomyelin is essential for Leishmania virulence. PLoS pathogens *5*, e1000692.

Zheng, L., Wang, Y., Schabath, M.B., Grossman, H.B., and Wu, X. (2003). Sulfotransferase 1A1 (SULT1A1) polymorphism and bladder cancer risk: a case-control study. Cancer letters *202*, 61-69.

Zhou, H., Xu, M., Huang, Q., Gates, A.T., Zhang, X.D., Castle, J.C., Stec, E., Ferrer, M., Strulovici, B., Hazuda, D.J., and Espeseth, A.S. (2008). Genome-scale RNAi screen for host factors required for HIV replication. Cell host & microbe *4*, 495-504.

Zumla, A., Nahid, P., and Cole, S.T. (2013). Advances in the development of new tuberculosis drugs and treatment regimens. Nat Rev Drug Discov *12*, 388-404.