COMPUTATIONAL INVESTIGATION OF THE MOLECULAR BASIS OF SUSCEPTIBILITY AND RESILIENCE IN DIFFERENT MACAQUE SPECIES INFECTED WITH MALARIA CAUSING *PLASMODIUM* PATHOGENS

A Dissertation
Presented to
The Academic Faculty

by

Anuj Gupta

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Biological Sciences

Georgia Institute of Technology
December 2021

COMPUTATIONAL INVESTIGATION OF THE MOLECULAR BASIS OF SUSCEPTIBILITY AND RESILIENCE IN DIFFERENT MACAQUE SPECIES INFECTED WITH MALARIA CAUSING *PLASMODIUM* PATHOGENS

Approved by:

Dr. Eberhard O. Voit, Advisor
Department of Biomedical Engineering
Georgia Institute of Technology

Dr. Peng Qiu
Department of Biomedical Engineering
Georgia Institute of Technology

Dr. I. King Jordan
School of Biological Sciences
Georgia Institute of Technology

Dr. Mary R. Galinski
School of Medicine
Emory University

Dr. Mark P. Styczynski
School of Chemical and Biomolecular Engineering
Georgia Institute of Technology

Date Approved: November 30, 2021

This work is dedicated to the loving memory of Baba.

“गुरु गोविन्द दोऊ खड़े, काके लागू पाय|

बलिहारी गुरु आपने, गोविन्द दियो बताय||”

ACKNOWLEDGEMENTS

I would begin by expressing my deepest gratitude to my advisor, Dr. Eberhard O. Voit, for the invaluable advice and continuous support during my Ph.D., for his patience, motivation, and immense knowledge. Dr. Voit has been an amazing mentor and has been central to not only my research work but also to my life in the last 5 years. Thank you for your guidance, for giving me the opportunity to study and learn a myriad of topics, and for providing the most nurturing environment a student could ask for. I would like to also thank Dr. King Jordan and Dr. Mary Galinski for their guidance. Dr. Jordan, thank you for giving me the platform to dive into a career in scientific research. Dr. Galinski, thank you for your leadership during MaHPIC and HAMMER studies that supported the collaborations and scientific discussions that made these projects so successful. Thank you, Mary, for teaching me about malaria and guiding me along my Ph.D. journey. Additionally, I am grateful to Dr. Mark Styczynski and Dr. Luis Fonseca for their valuable input whenever I needed guidance or ran into trouble. I also want to thank the numerous members of the MaHPIC consortium who put in their valuable time and effort in building one of the most remarkable malaria projects. I would also like to thank my fellow researchers and friends at Georgia Tech who helped me along the way. Special thanks to Ankit Srivastava, Ashis Pati, Daniel Olivença, Han-Ting Chou and Mohit Agarwal for being there every step of the way. Finally, I would like to thank my family. My mom, dad, brother, sister-in-law and my grand-mom have been a constant source of energy, motivation and support.

Thank you everyone.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

CLP	common lymphoid progenitor
DAMP	damage associated molecular pattern
DC	dendritic cells
DEG	differentially expressed genes
dpi	days post inoculation
DRG	differentially responding genes
FDR	false discovery rate
GAS	gamma activated sequence
GSEA	gene set enrichment analysis
HAMMER	host acute models of malaria to study experimental resilience
HSPC	hematopoietic stem and progenitor cell
ISRE	interferon-stimulated response element
MaHPIC	malaria host pathogen interaction center
Mf	*Macaca fascicularis*
Mm	*Macaca mulatta*
NES	normalized enrichment score
NHP	non-human primate
NK	natural killer cell
PAMP	pathogen associated molecular pattern
PCA	principal component analysis
Pk	*Plasmodium knowlesi*
PRR	pattern recognition receptor
RP	ribosomal protein
TF	transcription factor
TP	time point

SUMMARY

Malaria has a complex pathology with varying manifestations and symptoms, effects on host tissues, and different degrees of severity and ultimate outcome, depending on the causative *Plasmodium* pathogen species. The studies in this dissertation analyze consequences of transcriptomic changes in the blood of two closely related macaque species (*Macaca mulatta* and *Macaca fascicularis*) in response to acute primary infection by *Plasmodium knowlesi*. *P. knowlesi* is an emanant zoonotic pathogen that causes acute severe infection in humans. Although the two macaque species are very closely related to each other and to humans, the infection in *M. mulatta* is fatal, unless aggressively treated (infections in humans are treatable if detected early), whereas *M. fascicularis* develops a chronic, but tolerable infection.

The comparative analysis described here suggests that a reason for this stark difference in outcome is that the two hosts differ in immune cell programs and the expression of important genes. Specifically, the analyses establish a delayed pathogen detection in *M. mulatta* followed by extended inflammation that overwhelms this monkey's immune response. By contrast, *M. fascicularis* was found to detect the pathogen earlier and to control the inflammation. Additionally, *M. fascicularis* limits cell proliferation pathways until peak infection, presumably in an attempt to reinforce recovery through the adaptive immune system. To compliment this transcriptomics analysis, a gene expression-based metabolic modeling approach was developed that combines multi-omics knowledge to offer a molecular interpretation to biological systems. This approach helped with the interpretation of changes in a well-known inflammation biomarker, Kyn/Trp ratio, and

relate it to differences in immune response and cell proliferation. In-depth analysis of observed differences reveals that pattern-recognition receptor (PRR) signaling pathways are crucial for detection of pathogen and transcriptomic differences in early liver phase of infection revealed an early detection in *M. fascicularis*. Correlation analysis between host and pathogen transcripts reveals a pathogenic surface antigen, *SICAvar* Type 1, as an important regulator throughout the infection. The log phase of infection in hosts is similar, with macrophages and monocytes responsible for innate immune responses. During this phase, *M. mulatta* shows higher inflammation signals with upregulated inflammasome IL6-JAK-STAT3 signaling and IL10 expression, which continues to peak-infection phase. In contrast, *M. fascicularis* controls inflammation, presumably by means of the p53 pathway, which is distinctly downregulated near the peak of infection, thereby enabling adaptive immunity with various cell proliferation pathways that aid CD4+ T-cells and memory B-cells. Integrative metabolic modeling shows the potential role of tryptophan metabolism in regulating inflammation and stress response.

A complete understanding of the exact dynamics of the immune response is difficult to reach. Nonetheless, the results in this dissertation provide clear indication toward processes that underlie an effective immune response. Thus, this study may pave the way for future immune strategies toward treating malaria and identifies multiple points of intervention that are apparently responsible for a balanced and effective immune response.

CHAPTER 1. INTRODUCTION

1.1 The Scourge of Malaria

Malaria has plagued humanity since the dawn of civilization and is one of the world's deadliest infectious diseases, with an estimated 229 million cases and 409,000 deaths reported in 2019[1,2]. Studies on its origins in humans suggest that the disease was acquired from great apes over 10000 years ago[3]. Malaria has been studied scientifically since the late 1800's, when the disease was blamed on "bad air" (Italian: *mal' aria*), and great progress has been made toward understanding and treating the disease since the identification of the infecting agent as mosquito-borne *Plasmodium* parasites[4-6]. Yet, malaria and zoonotic pathogens still persist throughout the world, and resistance to anti-malarial drugs is increasing[7,8] Targeted studies of the fundamental molecular and physiological mechanisms of *Plasmodium* infections in the liver, and subsequently in the blood, and their effects on the host continue to be direly needed.

1.2 *Plasmodium* – a twist around every turn

Malaria is caused by protozoan parasites of the genus *Plasmodium*, within the phylum *Apicomplexa*. More than 200 species of *Plasmodium* have been identified to infect various reptiles, birds, and mammals. Out of these, only six species of *Plasmodium* adapted to humans and are recognized as causing malaria: *P. cynomolgi*, *P. falciparum*, *P. knowlesi*, *P. malariae, P. ovale* and *P. vivax* [9].

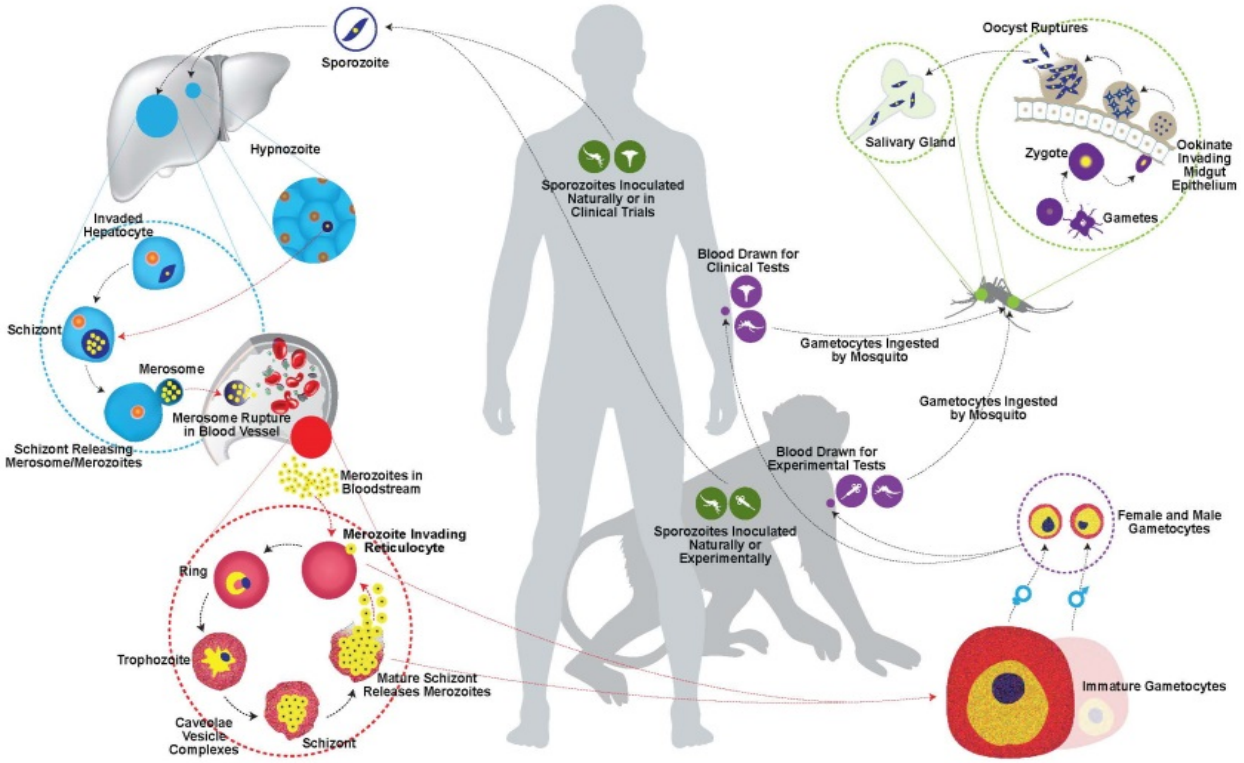


Figure 1.1- Multi-stage life cycle of *Plasmodium* pathogen in comparable human and macaque hosts. Reprinted from *Advances in Parasitology*, Volume 81, Chapter 1 *Plasmodium vivax*: Modern Strategies to Study a Persistent Parasite's Life Cycle, by M. R. Galinski, E. V. S. Meyer, J. W. Barnwell, Pages 1-26, 2013, with permission from Elsevier[10].

The *Plasmodium* parasite has a complex, multistage life cycle which spreads across the mosquito vectors and the vertebrate hosts (Figure 1.1). A specialized set of over 5000 genes and their proteins help *Plasmodium* invade and grow in multiple cell types while often evading host immune systems[11,12]. Their life cycle consists of an exogenous sexual phase (sporogony), which takes place in *Anopheles* mosquitoes, and an endogenous asexual phase (schizogony). The second phase begins with pre-erythrocytic schizogony when sporozoites are injected by mosquitoes into the blood and quickly travel to host liver cells where they undergo development and multiplication, thereby forming a tissue schizont that contains thousands of merozoites. This exoerythrocytic schizogony lasts about 6 days (in *P. falciparum* and *P. knowlesi*) to 16 days (in *P. malariae*). This

liver phase is followed by a blood phase, which is characterized by clinical manifestations and brought about by bursting of mature schizonts, releasing merozoites. The blood cycle begins when merozoites invade erythrocytes and intracellularly convert to trophozoites, thus ending the prepatent phase. The trophozoites mature within the erythrocytes into merozoite containing blood schizonts, a process that takes between 24 and 72 hours, depending on pathogen species. The infected erythrocytes then rupture and release a new wave of merozoites, which continue to infect other erythrocytes. Additionally, during the process of schizogony, some of the merozoites convert to sexual gametocytes which complete the *Plasmodium* lifecycle once they are ingested by a female anopheline mosquito.

1.3 *Plasmodium knowlesi* – challenges and opportunities

The prevalence of *P. falciparum* and *P. vivax* in humans has driven the acquisition of enormous knowledge regarding these species. By contrast, the number of *P. knowlesi* cases has increased in the past few decades, but many details of the zoonotic potential of this parasite are yet to be discovered.

P. knowlesi has long been known to infect long-tailed and pig-tailed macaques but has been identified as a pathogen for malaria in humans only a few decades ago [13]. In fact, naturally acquired human infections were extremely rare and often misinterpreted as morphologically similar to those caused by *P. malariae*[13-15]. The zoonotic nature of *P. knowlesi* not only makes it a growing public health concern, but also provides a unique opportunity to perform comparative analyses leading to insights into the immune response of macaque and human hosts. *P. knowlesi* infections range from mild to potentially fatal, if not readily treated.

Even though *P. knowlesi* was identified and isolated before 1930, its zoonotic potential was initially underestimated[13]. The situation changed in 1999 when a large group of human *P. knowlesi* infections was discovered. Subsequent studies confirmed the majority of several previously identified *P. malariae* cases as *P. knowlesi* infections[16]. Comparative genomics of *P. knowlesi* from macaque and human hosts ascertains that the most recent common ancestor was active between 100,000 and 500,000 years ago, which makes *P. knowlesi* as old or older than both *P. falciparum* and *P. vivax*[17]. Multiple other studies have suggested *P. knowlesi* to be zoonotic since humans settled in Southeast Asia 70,000 years ago[13].

P. knowlesi has a life cycle and molecular features that mostly resemble other *Plasmodium* parasites. However, notable differences include the absence of dormant hypnozoites, which are responsible for latency and relapse in human malarial infections caused by closely related species *P. vivax, P. ovale* and *P. cynomolgi*. Another major difference is a 24-h erythrocytic cycle with quotidian fever patterns, which makes it uniquely rapid among primate-infecting *Plasmodium* species[18]. *P. knowlesi* can cause both uncomplicated and severe malaria in humans with about 10% severe and 2% fatal infections[19].

In recent times, *P. knowlesi* has been extensively used for malarial research of host-parasite interactions and vaccine and drug development[20]. Its ability to infect rhesus macaques, the most used model primates in research, is probably the reason for this prevalence. The pathogen can also be used for *in vitro* research and can be maintained in cultured red blood cells of both humans and macaques[20]. Moreover, recent advances in molecular genetics permit transfection experiments that can be used to genetically modify *P. knowlesi*, therewith increasing its utility as a model organism[21].

1.4 Animal models for malaria

Numerous cross-sectional blood sample examinations from individuals with varying levels of malarial disease have been carried out around the world for decades, both for diagnosis and to conduct research[22]. However, the investigative repertoire for studying the temporal changes associated with disease progression in humans is relatively scarce[23-25]. The main reason is of an ethical nature, as treatment of patients is typically mandated as soon as they are diagnosed. Consequently, the main source of human blood samples for research on natural infections has been from active case detection and symptomatic individuals seeking treatment. It is evident that deeper knowledge of malaria and its progression concurrent with the launching of an immune response by the host is a prerequisite for rationally developing new measures for preventing the disease, treating patients, and improving patient outcomes. While immunity and vaccine research is advancing in the context of controlled human malaria infection (CHMI) studies initiated with sporozoites[26-34], performing longitudinal molecular studies with human volunteers has numerous logistical and ethical challenges[35].

Rodent malaria models have been widely used to expand our understanding of *Plasmodium* infections[36-38]. While they offer a spectrum of advantages, the differences in human and mouse or rat physiology present some drawbacks. Nonhuman primate (NHP) macaque models are much closer to humans, and the clinical presentation of malaria and subsequent immune responses are more similar between humans and macaques[20,39-42]. As a consequence, macaques have become important alternatives to rodent models for explaining different host-pathogen interactions, not only for malaria, but various other diseases, including those caused by HIV/SIV and SARS-CoV-2[43,44]. Indeed, infections of macaques are widely accepted as robust *in vivo* models for human malaria, with comparable liver- and blood-stage cycles[45,46].

1.5 So close (Evolutionary), yet so far: Varying degree of susceptibility and resilience

Macaques belong to the genus of Old World monkeys and consist of 22 diverse species. Of these, *Macaca mulatta* (Mm; rhesus macaques) and *Macaca fascicularis* (Mf; long-tailed, cynomolgus, or kra macaques) are the most widely used NHP model species[41]. Long before *P. knowlesi* became a zoonotic concern, Knowles *et. al.*[47] identified *Macaca fascicularis* (Mf) as a natural host for *P. knowlesi* infection. They also reported that it causes a mild infection in Mf, but a fatal infection in Mm.

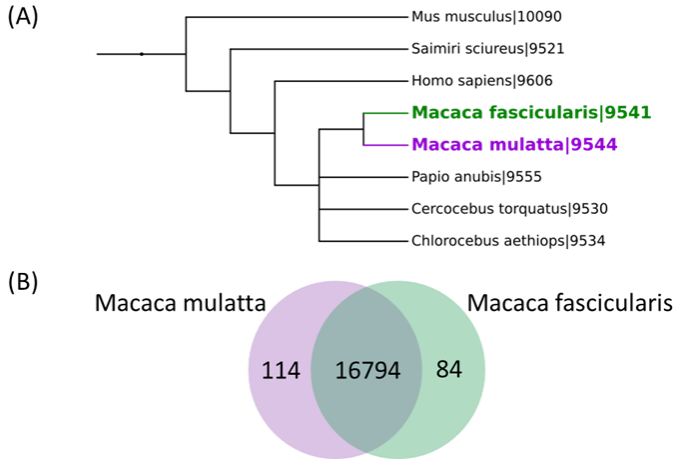


Figure 1.2- Overall genomic similarity and evolutionary separation of *M. mulatta*, *M. fascicularis*, and related species. (A) Pertinent part of the phylogenetic tree, assembled using the phylot web tool[48,49]. (B) Venn diagram of homologous genes between Mm and Mf.

While Mm and Mf are closely related (Figure 1.2) NHP species with reported interbreeding and shared geographical locations[50], an important difference must be noted. Namely, Mf co-evolved with *P. knowlesi* within a large geographical area of Southeast Asia, whereas the distribution of Mm overlaps with *P. knowlesi* only slightly (Figure 1.3). Arguably as a consequence, Mf shows signs of the disease but survives with a low-level infection that can become chronic, whereas Mm becomes severely ill with escalating life-threatening parasitemia

and succumbs unless treated[45,46,51,52]. The molecular and physiological basis of this stark difference is unknown; it is the overall topic of this dissertation. Studies analyzing these differences in other disease models have begun to show that Mf in many cases launches an arguably more effective immune response[53-56]. Therefore, the comparison of Mf and Mm during malarial infection, which is addressed in this work, will not only help explain details of the *P. knowlesi* infection but also provide a better understanding of the control of the biological programs that differentiate the immune responses in these two species. The results will offer insights into the details of these responses and may also point to molecular targets that might lead to improved treatments for malaria and possibly other infections.

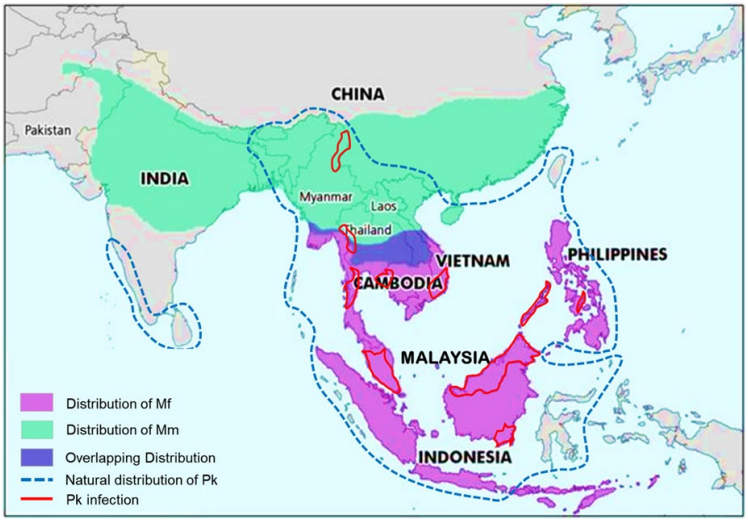


Figure 1.3- Distribution of *M. mulatta* and *M. fascicularis* macaques and the *Plasmodium knowlesi* pathogen. Regions of geographical distribution of Mm and Mf, overlaid with areas where *P. knowlesi* (Pk) is present and able to infect the macaques (dashed blue lines). Areas with observed *P. knowlesi* infections are marked with red lines[57-59].

1.6 MAHPIC and HAMMER: understanding resilience against malaria

Between 2012 and 2019, under the auspices of research contracts from the U.S. National Institute of Allergy and Infectious Diseases (NIAID) and the U.S. Defense Advanced Research

Projects Agency (DARPA), the Malaria Host-Pathogen Interaction Center[60] (MaHPIC) and Host Acute Models of Malaria to study Experimental Resilience (HAMMER) projects collected and deposited in NIAID-supported Bioinformatics Resource Centers (BRCs) large clinical, parasitological, immunological and multi-omic datasets from longitudinal infections of macaque species infected with *P. coatneyi*, *P. cynomolgi* or *P. knowlesi*[60,61]. These investigations yielded a large collection of datasets from over 30 macaque infections and controls, including frequent samplings of blood and bone marrow[60,62-66]. The first two simian malaria parasite species are excellent models of corresponding human infections by the parasites *P. falciparum* and *P. vivax*, respectively[39,67], while *P. knowlesi* reflects aspects of both, depending on the question[20,68]. *P. cynomolgi* and *P. knowlesi* infect both humans and NHPs and constitute zoonotic public health concerns in Southeast Asia, where humans and macaque species coexist[8,69-72]. Comparative longitudinal macaque infection studies of the type performed by these consortia can help focus analyses on significant molecular features that play crucial roles in determining the course of the disease. Furthermore, because host-pathogen interactions in macaques closely mimic those in humans, findings from macaques have great potential for establishing a rational basis for new therapeutic targets and interventions, including host-directed therapies.

1.7 Moving forward

The rich data resource assembled by MaHPIC and HAMMER allows us to pursue several avenues of research, some of which are as follows:

1.7.1 Genetic predisposition

Natural selection has been identified as an important basis for adaptation[73,74]. In 1949 Haldane *et al.* hypothesized that red blood cell (RBC) disorders like sickle-cell anemia and various

thalassemias were more prevalent in malaria endemic tropical regions due to natural selection of these traits[75]. Follow-up studies confirmed specific genetic mutations correlated with these geographical locations and demonstrated that they provided resistance against malaria[76]. Since then, multiple other examples have shown selective pressure to shape genomes, including the human genome. Malaria is arguably one of the oldest known disease and greatest cause for morbidity and mortality that resulted in the evolution of adaptive traits that helped humans survive over the past millennia[77].

The two macaque hosts Mm and Mf are most widely used in NHP studies. They have often been used to study common infections and differences in their responses[78-81]. One of the starkest difference occurs with *P. knowlesi* malarial infections: The natural host, Mf, survives with relatively mild parasitemia while Mm succumbs to severe infection. With such closely related species one would expect a similar immune response, suggesting the hypothesis that the selective pressure on the natural host over multiple millennia altered the host genotype of Mf to boost its resistance.

Addressing this hypothesis, I analyzed some of the MaHPIC and HAMMER data, beginning with a comparative transcriptomics study of the two hosts, which was specifically designed to shed light on fundamental differences leading to pathogen tolerance in Mf. Blood transcriptomics is a very rich resource and can directly be used to infer details about the host immune response. It can help us understand key aspects of infection, like modes of detecting invading parasites, differences in the innate and adaptive immune response, metabolic differences, and the role of inflammation in controlling it.

The insights gained from these analyses could have multiple impact. First, the zoonosis of *P. knowlesi* and the similarities in the immune responses of several primates could direct the search for drug targets and interventions boosting tolerance or resistance in humans. Second, strong results could support researchers in their choice of best-suited model organisms for specific research purposes. The latter aspect is important because the known genetic differences and similarities are presently considered in biomedical experiments only marginally, even though their influence could be considerable[82].

1.7.2 Variations in immune response

A priori, one might expect the immune responses in comparable species to be similar for common infections. But the two hosts analyzed here demonstrate drastically different outcomes[53,83]. The challenge to understanding these differences is the complexity of malarial disease. Nonetheless, I will demonstrate in this dissertation that the dramatic difference in overall outcome is the result of several small divergences between the two host species responses throughout the infection process, which are a function of both time and host tissue.

As discussed before, the malarial infection presumably acts as selective pressure in shaping host genetics. Additionally, the host-pathogen interactions increase the genetic diversity in both organisms over time. Multiple other infection studies have suggested that proteins are the most likely targets at the host-pathogen interface[84]. They primarily include a wide range of *Plasmodium* antigens, as well as components of the innate immune system responsible for pathogen detection. Differences in these detection mechanisms, in turn, trigger slightly different immune responses. Each immune response is a complex system of mechanisms spreading across multiple organs, tissues and cell types. As a consequence, the relatively subtle response differences between the

hosts are at risk of being lost in the myriad of processes that the immune system simultaneously activates. Within the blood tissue, these include functionalities of individual cell populations that are involved in the continuous detection of parasitemia and in transmitting corresponding signals. The first line of defense is the innate immune response, which is later followed by an adaptive immune response involving T cells, B cells, cytokines and antibodies. Specifically, foreign molecules associated with pathogens are first detected by macrophages which signal other cells through cytokines. Along with neutrophils, they phagocytize the pathogens. The cytokines activate various inflammatory pathways while antigen presenting cells (dendritic cells, B cells and macrophages) activate T cells and B cells. These cells finally attempt to disarm the pathogens either directly or through antibodies. These immune processes are accompanied by metabolic processes, which are responsible for maintaining essential functions, such as the supply of energy.

It seems reasonable to expect that a deeper knowledge of these processes, their functions and interdependences can greatly benefit our understanding of the immune responses and of triggers that cause these subtle but crucial differences.

1.7.3 Customizing methodologies

Comparative transcriptomics analysis between species often requires the development of new statistical strategies, which are frequently inspired by existing tools, and the modification of existing strategies to enable answers to key questions regarding the involved molecules. Because Mm and Mf have a similar genetic makeup, the comparison of their transcriptome has a chance to reveal crucial differences in responses, but characterizing subtle differences is often obstructed by limitations of the existing methodologies. For example, raw sequences are typically mapped to respective reference genomes. In our case, both genomes have been annotated, but Mf has only

recently been annotated by using the Mm reference genome. The quality of genome annotation improves with iterated analyses and one must be cautious when making direct comparisons for homologous genes. Another challenge is the small sample size for these NHP experiments. All statistical analyses are affected by bias, which is stronger for small samples, and care is needed when claims are made regarding the significance of alleged differences. Finally, NHPs are fairly well-studied research models, but any comparative experiments might suffer from the fact that most of the background datasets were created using human data. An example is the establishment of most response gene sets.

Supporting transcriptomics analyses, the integration of multi-omics data can help rationalize and explain biological processes with greater clarity and reliability. Diseases are a prime example. They always represent specific cases of malfunctioning within a complex system. Whereas it is often feasible to observe and possibly treat the symptoms of a disease, it is much more challenging to identify and characterize its molecular root causes. In particular, it is imperative to understand how small changes in transcriptomic regulations can bring about changes in functional protein products. Several broad types of methodologies have been developed to link transcriptomics data not only to protein profiles but also to metabolic events. Most of these methods rely on associations within datasets and in the recent past, several machine learning and dynamic modelling methods have been proposed for this purpose[85]. The sheer sizes of many –omics datasets pose challenges as well. Modern-day computers can process huge amounts of data, but it is no longer easy to discern valid information or true signals in the data from uncertainties, variability, and noise. This aspect is critical, because every individual is different and many differences in gene expression within a diseased patient cohort may simply be manifestations of their genetic make-up and health histories.

The malaria model of *M. mulatta*, *M. fascicularis* and *P. knowlesi* presents a unique opportunity for the combined analysis of such large datasets. In this work, I perform such an analysis by developing, adapting and finetuning methodologies that systematically interpret multi-omics data and by focusing on molecular functions that contrast relatively small differences within the complex host-pathogen system of malaria.

1.8 Summary

The combination of NHP models of *P. knowlesi* malaria has immense potential toward understanding of host immune response, pathogenesis and host pathogen interactions that govern the disease progression and ultimate outcome. The MaHPIC and HAMMER consortium have gathered valuable multi-omics data by several longitudinal malarial studies. The studies described in this dissertation address aspects of malarial infections that govern the molecular interactions between hosts and parasites, including resilience, tolerance, and pathogenicity, as well as specific metabolic and immune responses of the hosts. Accordingly, the research work is structured into three specific projects that are described in following chapters.

Chapter 2: Transcriptomics analysis identifies key differences during Plasmodium knowlesi *infection.*

In Chapter 2, we compare the blood transcriptomic profiles of two closely related macaque species (*M. mulatta*; abbreviated as Mm, and *M. fascicularis*; Mf) in response to infection with a common parasite species (*P. knowlesi*; Pk) throughout parallel infection trajectories. Using standardized transcriptomics techniques, I quantify and differentiate various phases of the *Plasmodium* infection and explore differential responses in the two host species. Special emphasis is placed on immune mechanisms and inflammation control. This study identifies key signaling

pathways that control various phases of infection, including the early sensing of pathogenic material and the transition from an immediate immune response to control of inflammation. The study sheds light on species-specific responses to *P. knowlesi* infection and identifies molecular mechanisms that help Mf survive the infection while its close relative Mm succumbs to the disease if not treated.

Chapter 3: Integration of metabolic modelling and transcriptomics to explain metabolomics changes during infection.

Chapter 3 introduces methods of computational systems biology for interpreting transcriptomic changes during malarial infection in terms of metabolic changes, by mapping gene expression metrics onto alterations in enzyme activities and investigating, with dynamic models, the consequent changes in important pathway systems. The specific focus is on purine and tryptophan metabolism because of their importance for various aspects of *Plasmodium* infections. This modeling strategy enables us to characterize metabolic changes in comparison with observed transcriptomics changes. The chosen approach has the potential of elucidating the mechanistic functioning of various biological pathways by bridging the gap between transcriptomics and metabolomics data.

Chapter 4: Molecular mechanisms that explain key differences in progression of Plasmodium knowlesi *infection.*

Using transcriptomics and metabolomics data from multiple MaHPIC and THoR experiments, I investigate in Chapter 4 the key findings from Chapter 2 in order to interpret and explain the molecular mechanisms responsible for pathogen detection, immune response and inflammation. I identify pathogen antigens that are key to host pathogen interactions and

differentiate the two hosts based on pathogen detection and protein kinase signaling while highlighting the importance of Ca^{2+} homeostasis in immune response. Finally, using methodology developed in Chapter 3, I describe the molecular mechanism that regulates inflammatory biomarker, Kyn/Trp ratio, thereby establishing a clearer role of tryptophan metabolism for the balance of inflammation and cell proliferation.

Taken together, these studies not only give us mechanistic insights into the progression of *Plasmodium* infections but also help us identify candidate targets and timing regimens, which might be of importance for the development of future clinical or pharmaceutical interventions. They give us crucial insights into molecular mechanisms behind susceptibility and resilience in host immune response.

CHAPTER 2. TRANSCRIPTOMICS ANALYSIS IDENTIFIES KEY DIFFERENCES DURING *PLASMODIUM KNOWLESI* INFECTION[1]

2.1 Summary

Plasmodium knowlesi, a model malaria parasite, is responsible for a significant portion of zoonotic malaria cases in Southeast Asia and must be controlled to avoid disease severity and fatalities. However, little is known about the host-parasite interactions and molecular mechanisms in play during the course of *P. knowlesi* malaria infections, which also may be relevant across *Plasmodium* species. Here we contrast *P. knowlesi* sporozoite-initiated infections in *Macaca mulatta* and *Macaca fascicularis* using whole blood RNA-sequencing and transcriptomic analysis. These macaque hosts are evolutionarily close, yet malaria-naïve *M. mulatta* will succumb to blood-stage infection without treatment, whereas malaria-naïve *M. fascicularis* controls parasitemia without treatment. This comparative analysis reveals transcriptomic differences as early as the liver phase of infection, in the form of signaling pathways that are activated in *M. fascicularis*, but not *M. mulatta*. Additionally, while most immune responses are initially similar during the acute stage of the blood infection, significant differences arise subsequently. The observed differences point to prolonged inflammation and anti-inflammatory effects of IL10 in *M. mulatta*, while *M. fascicularis* undergoes a transcriptional makeover towards cell proliferation, consistent with its recovery. Together, these findings suggest that timely detection of *P. knowlesi* in *M. fascicularis*, coupled with control of inflammation while initiating the replenishment of key cell populations, helps contain the infection. Overall, this study described in this chapter points to specific genes

[1] Gupta, A., Styczynski, M.P., Galinski, M.R. *et al.* Dramatic transcriptomic differences in *Macaca mulatta* and *Macaca fascicularis* with *Plasmodium knowlesi* infections. Sci Rep (2021)

and pathways that could be investigated as a basis for new drug targets that support recovery from acute malaria.

2.2 Introduction

Malaria presents a constant burden to public health of over 100 countries worldwide[1]. Of the five *Plasmodium* species known to cause malarial infection in humans, *Plasmodium knowlesi* is unique and interesting in its own way[86]. *P. knowlesi* infections in humans are almost exclusively reported in Southeast Asian countries and includes travelers from these countries as well[13,87]. The major reason for these infections to be confined within Southeast Asia is the geographical distribution of its natural hosts and vectors[13,19]. *P. knowlesi* malaria ranges from mild, uncomplicated to severe and is accompanied by fevers and chills in humans[13].

Natural hosts of the species *Macaca fascicularis* (Mf) do not exhibit severe disease while the rarer hosts *Macaca mulatta* (Mm), which are typically used for experimentation, experience severe infections with high parasitemia levels and death if not readily treated. What makes this stark difference particularly interesting is the fact that Mm and Mf are closely related with reported successful interbreeding in geographical overlaps and in captivity[88]. Malarial infections have been the most common and well-studied examples of selective pressure in humans. Due to their long history with Mf, it is not surprising to hypothesize evolutionary changes in Mf that make them resilient to *P. knowlesi* infection. This makes comparative longitudinal studies of the two closely related hosts, Mm and Mf, with *P. knowlesi* infection an intriguing avenue to understand host immune responses.

Transcriptomics studies provide the most detailed understanding of molecular mechanisms at play in various biological processes. With the current repertoire of analyses, transcriptomics

data may be used not only to make inferences regarding the expression of genes, but also for enrichments of biological processes, studies of the involvement of key transcription factors and many other insightful analyses. For *P. knowlesi* infections, it can help be useful to establish and confirm various phases of infection, characterize the immune response at different time points and contrast the two hosts on subtle differences.

Even though the phases of malarial infection remain roughly similar between Mm and Mf, several characteristic differences have been noted in various studies[51,83]. For instance, Mf experienced lower and more controlled parasitemia compared to Mm[83,89]. In addition, transcriptomics analysis can help distinguish finer differences between the immune response of the two hosts such as earlier inflammatory response in Mf[89]. Such analyses could help with the identification of key transcription factors and leukocyte cell types at play during the infection and of the key determinants in terms of immune responses and inflammation that boost resistance in Mf.

For the analyses described in this chapter, we used peripheral blood transcriptomics data from cohorts of *Macaca mulatta* (Mm) and *Macaca fascicularis* (Mf), which were infected with sporozoites of the Malayan strain of *P. knowlesi*[90]. We studied the host responses prior to and from the time of parasite inoculation through the development of liver-stage schizonts containing invasive merozoites followed by the release of these merozoites from infected hepatocytes into the bloodstream. Cyclical merozoite invasion and multiplication within red blood cells (RBCs) caused the progression of malarial illness and disease manifestations, presented elsewhere[83].

The biggest challenge that this transcriptomics analysis faces is the small sample sizes of the macaque experiments. These cannot be completely avoided but stricter significance, false

discovery rates (FDR) and fold change thresholds can greatly reduce their impact and help make better conclusions. Another challenge is with respect to the innate variability among the host species and the availability of genetic and transcriptomics resources for these analyses. These can be addressed by analyzing the genetic differences and using human data resources that are still much more reliable than those for non-human primates (NHPs).

2.3 Methods

2.3.1 Experimental Setup and Data Collection

For this analysis, four male Mm and seven male Mf were infected with sporozoites of the Malayan strain of *P. knowlesi*[90]. The animals were observed daily during baseline periods, when inoculated with sporozoites, and throughout the infection. Female monkeys were excluded to avoid confounding effects of menstruation. Blood samples were collected at pre-defined TPs between 1 pm – 3 pm for both hosts, when the *P. knowlesi* cycle presents predominantly—if not exclusively—ring forms in circulation (Figure 2.1). The Mf experiment underwent an initial unsuccessful sporozoite inoculation, and a new pre-infection baseline (TP2B) was established. The sporozoite re-inoculation was conducted approximately 80 days after the failed inoculation, and as shown in Appendix Figure A.1.11, the failed inoculation did not have any apparent effect on the subjects and the observed transcriptomes. All pre-infection samples were used to establish baseline expression.

The present study describes a secondary data analysis, while experimental details are described in Peterson *et al.*[83]. All experiments involving NHPs were performed at the Yerkes National Primate Research Center (YNPRC), an AAALAC International-accredited facility. All procedures followed ARRIVE guidelines and were approved by Emory's IACUC and the Animal

Care and Use Review Office (ACURO) of the US Department of Defense and followed accordingly. The Emory's IACUC approval number was PROTO201700484 - YER-2003344-ENTRPR-A.

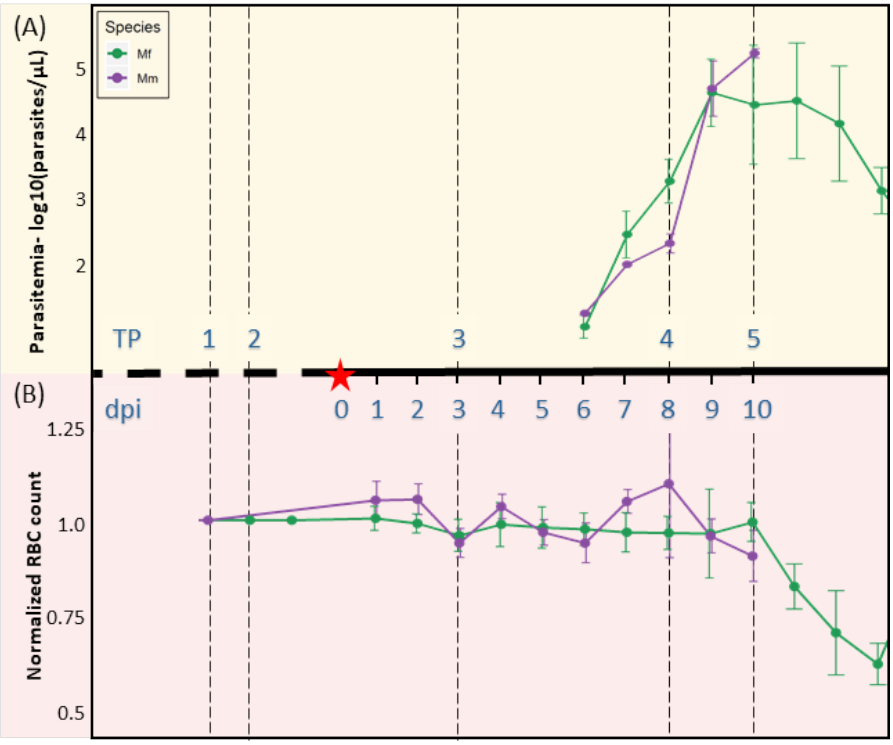


Figure 2.1- Timeline of Mm and Mf infection with *P. knowlesi*, along with parasitemia levels and RBC counts. The x-axis depicts the time points (TPs) and days post inoculation (dpi). The red star represents the day of sporozoite inoculation (*i.e.*, dpi 0). Data at TP1 were collected more than a month before TP2. (A) Parasitemia: The y-axis shows average parasitemia levels throughout the *P. knowlesi* infection on a log10 scale. Parasitemia levels were measured as parasites/µL. (B) Normalized RBC counts: The y-axis shows the ratio of mean RBC counts with respect to pre-infection levels.

After infection with sporozoites at Day 0, parasitemia in both species became patent 6 days post inoculation (dpi) (Figure 2.1). Mf self-controlled the infection, while parasites in Mm kept rising, as expected, with no evidence of the animals controlling the infection[83]. In the Mm species, untreated parasitemia can escalate rapidly such that the majority of all RBCs become infected, which could result in certain death of the animal. In our study, the monkeys were monitored

carefully with blood smear readings taken twice daily during the acute stage of the infection and finally at 10 or 11 dpi when the animals had approximately 1% parasitemia and were euthanized for pathological measurements[83]. Red blood cell numbers and parasitemia levels are presented in Figure 2.1 with details in Peterson *et al.*[83]. Other details of the experiments have been reported in publicly available databases: Mm (referenced as Experiment 06) at https://plasmodb.org/plasmo/app/static-content/PlasmoDB/mahpic.html and https://www.ncbi.nlm.nih.gov/bioproject/524357, and Mf (referenced as Experiment 07) at https://plasmodb.org/plasmo/app/static-content/PlasmoDB/mahpic.html and https://www.ncbi.nlm.nih.gov/bioproject/526495. Additionally, a clinical and histopathological analysis of these cohorts can be found elsewhere[83]. The transcriptomics data for both hosts can be found in the Gene Expression Omnibus (GEO accession numbers: GSE127079, GSE128115). PlasmoDB and NCBI-Bioproject are public databases that are freely accessible to anyone without specific permission.

2.3.2 Orthology Analysis and Gene Similarity Scores

The reference Mm genome, corresponding transcripts and annotations were obtained from Zimin *et al*.[91]. The data can be downloaded from the reference site[92]. The corresponding reference genome files for Mf were obtained from NCBI annotation release 101[93] and can be downloaded from the NCBI ftp server[94].

Nucleotide sequences from the corresponding transcripts (fasta) file were used to detect reciprocal best hits. This correspondence was achieved with a reciprocal BLAST protocol described in reference[95] and used to identify orthologous transcripts. Unfortunately, the reciprocal-best-hits method does not guarantee orthology and is prone to shortcomings like its handling of

gene duplications. Thus, to estimate the similarity between two orthologous sequences, transcripts differing in length by more than 50 bp were removed to avoid manual curation. Additionally, transcripts with less than 85 percent identity were removed. Finally, to calculate the evolutionary distance between homologous genes, a robust and widely accepted metric of sequence similarity was used[96]. The similarity score for the transcripts was calculated using BLAST alignment scores. These scores are calculated by assigning a value to each pair of nucleotides and then summing these values[97]. These scores were then normalized by the lengths of the transcripts to obtain the similarity scores for each pair of homologous genes.

2.3.3 Read Mapping and Gene Expression Analysis

Samples were sequenced using Illumina Hi-seq 3000. For each host, the reads were mapped using STAR (version 2.5.2b)[98] against corresponding references (*cf.* sources in Orthology analysis above). For each species, a composite reference genome was assembled using STAR index, and further raw RNA-seq reads were mapped to the combined reference using STAR.

Raw reads were normalized for library size, sequencing depth and composition using the DESeq2[99] standard library size normalization method (estimateSizeFactors function). Custom R scripts were used to implement the DESeq2 normalizations and create PCA plots using variance stabilized transformation (vst function) of the normalized data. The major genes contributing to each PC were extracted using the largest absolute values of components for the eigenvectors for each PC. The enrichment *p*-value for each gene set was calculated using the top 200 genes in a hypergeometric test.

2.3.4 Differential Expression and Differential Response

Differential expression of genes was calculated using DESeq2[99]. First, we filtered low-abundance genes by removing genes which had low read counts in more than 20% of the samples, as is standard procedure. For differential expression analysis, the samples were modeled using species as the major factor and infection-TPs as a secondary factor for a subset of samples for each host. For differential response a species:infection-TPs interaction term was added to all the samples.

Since DESeq2 models the expression data as a negative binomial distribution, dispersion was estimated using the 'estimateDispersions' function, and differential expression was calculated using Wald's test (nbinomWaldTest). DESeq2 functions adjust the *p*-value using the Benjamini-Hochberg method. For differential expression analysis, data were contrasted on the TP infection state with respect to baseline. For differential responses, the samples were contrasted on the interaction term. Representative examples of differentially expressed and differentially responding genes are highlighted in Appendix Figure A.1.3. The DESeq2 package in R was used[100].

2.3.5 Gene Set Enrichment Analysis

Differentially expressed genes were analyzed for the enrichment of gene sets using the GSEA toolkit (version 3.0)[101] by the Broad Institute. The gene sets used for the analysis were Hallmark[102] and Gene Ontology (GO)[103,104]. The pre-ranked GSEA module of the toolkit was used for the analyses, where genes were ranked on an adjusted *p*-value and the sign of the fold-change.

We performed the ranked analysis in two ways. First, to retain only robust signals, we selected genes that were significant (adjusted $p < 0.01$ and log2(fold-change) > 1) and used their

ranked list. Second, for comparisons between infection TPs between species, we selected all genes and used their ranked list. This step reduced biases in enrichment of gene sets as the same genes would be present in each set and the enrichment score (ES) would be calculated by their ranks. The toolkit calculates an ES for each of the gene sets that demonstrates overrepresentation of the gene set at the top/bottom of the ranked gene list. GSEA uses a weighted standard Kolmogorov–Smirnov statistic to calculate the ES. To account for different sizes of gene sets and correlations between gene sets and expression data, a normalized ES was calculated by considering 1,000 permutations of ES, calculated by randomly assigning phenotypes to samples. Finally, false positives were restricted by applying a false discovery rate (FDR) correction[105] and using the threshold FDR < 0.25.

To elucidate the transition between TP4 and TP5 further, ranked gene set analysis was performed (Figure 2.4). All gene sets identified as Hallmark and GO Biological Processes were ranked based on their normalized enrichment scores (NES) for each species at both TP4 and TP5. For the Hallmark sets (Figures 2.4C, 4D), a rank flow plot was created to visualize changes in ranks between TP4 to TP5 for each species. For the GO sets (Figure 2.4E), we refined the approach to identify which among the important gene sets undergo a transition between TP4 and TP5 for Mf but remain essentially unchanged in Mm. We achieved this by first calculating the rank difference for each gene set from TP4 to TP5 as shown by a violin plot. Each data point in the distribution represents a GO gene set and the quantitative value (y-axis) is the rank change (Rank @TP4 – Rank @ TP5). We narrowed our analysis to the most significant gene sets (rank < 200 for at least one species/TP). We then analyzed the gene sets which remain almost unchanged in Mm (absolute rank difference < 100) but are significantly changed in Mf (absolute rank difference > 1000). This filtering resulted in lists of gene sets that were then summarized by removing

redundant GO terms using REVIGO[106]. The results are summarized in blue and red boxes in Figure 2.4E.

GO-Net[107] and REVIGO[106] were also used in Figure 2.4E, S6, S8 and S9 to summarize GO results and use their hierarchical structure for inference. Custom R scripts were created to plot heatmaps and bar plots for GSEA results; the scripts are available in github package binf.gsea.visualizations at (https://github.com/LBSA-VoitLab/packages). To analyze the hierarchical structure of GO annotations, we used the Cytoscape (v 3.4) plugin Bingo (v 3.0.3)[108] and the treemaps were formed using REVIGO[106].

2.3.6 Upstream Targets and Motifs

Transcription factors and upstream regulators were analyzed with iRegulon[109] and TRRUST db[110]. While iRegulon predicts TF-targets using ChIP-seq data, the TRRUST database is built on highly curated TF-target associations acquired from the literature. Cumulative results of both applications were used. In some cases, the results were quite diverse, thereby leading to a larger list of TFs (*e.g.*, for Mf at TP3).

The iRegulon (v1.3) plugin for Cytoscape (v3.4) was used to predict TFs and gene set motifs. iRegulon implements a genome-wide ranking and recovery approach to detect enriched TFs and motifs. It looks for cis-regulatory sequences among co-expressed genes. For this particular analysis, we used the 10K PWMs (position weight matrices) database with NES > 3 and FDR < 0.001 on motif similarity.

TRRUST uses a sentence-based text mining approach and is very well curated. Customized R scripts and packages were created to analyze and merge the results and then create a regulatory

network; the scripts are available in binf.trrust github package at https://github.com/LBSA-VoitLab/packages.

2.3.7 Modular Transcriptome Repertoire

The third generation of the modular transcriptional framework was used as described in Altman *et al.*[111]. Module-level details such as transcripts and annotations were obtained from the corresponding supplementary material of Altman *et al.*[111]. We identified 382 uniquely annotated modules, but in some cases more than one module had a similar functional annotation. A ranked list using differential expression for each host species followed by enrichment analysis of each of these modules was established using the method explained above for GSEA[105]. The modules that were considered for this analysis had a $p < 0.01$ (adjusted) and FDR < 0.25. The scripts are available in binf.modular github package at https://github.com/LBSA-VoitLab/packages.

2.3.8 Cell Population Markers

To gauge the changes of various cell populations, we performed enrichment analysis of these populations and subpopulations. The Marker Gene database from the Atlas of Human Blood Cells[112] was used to compile cell markers for various cell populations. This database consists of 43 transcriptional cell clusters that are profiled from single-cell deep sequencing. Ranked lists of DEGs were used for enrichment, which was calculated using the method explained above for GSEA. The database was downloaded from the source and custom scripts were created to analyze and obtain enrichment scores, as detailed in Subramanian *et. al.*[101]. A cutoff adjusted *p*-value of 0.05 was chosen to select enriched cell populations.

2.4 Results

The overall goal of this analysis is to compare and characterize the temporal whole blood transcriptional programs launched by the two macaque species in response to the infection.

2.4.1 Plasmodium knowlesi *infection causes different gene expression patterns in* Macaca mulatta *(Mm) and* Macaca fascicularis *(Mf)*

As expected, gene expression repertoires and levels change in both monkey species during the course of experimentally introduced *P. knowlesi* sporozoite infections, both as the *P. knowlesi* parasites multiply within hepatocytes and following their cyclical replication in host RBCs and parasitemia patency (Figure 2.1A, 6 dpi (days post inoculation)). To characterize these changes, we compared RNA-seq data generated from samples acquired from both macaque species at baseline and at specific infection time points (TPs). Principal component analysis (PCA) of the whole-blood gene expression patterns (Figure 2.2A) shows a clear separation of the two species, as well as between samples taken before infection (TP1 and TP2) and during the fast rise and approach of peaking parasitemia (TP4 and TP5, respectively 8 and 10 dpi). Samples taken shortly after the inoculation of sporozoites, during the pre-patent period (TP3, 3 dpi), cluster mostly with pre-infection samples. Most of the variance (PC1 = 51.9%) shown is due to the host species, while the second major axis of variance (PC2 = 14.11%) separates pre-infection (TP1, TP2) from acute-parasitemia samples (TP4, TP5). Analysis of pre-infection TP1 and TP2 samples with pre-patent TP3 samples (Appendix Figure A.1.1) demonstrates dominance of inter-individual variability over short-term transcriptomic changes during the initial phase of the infection.

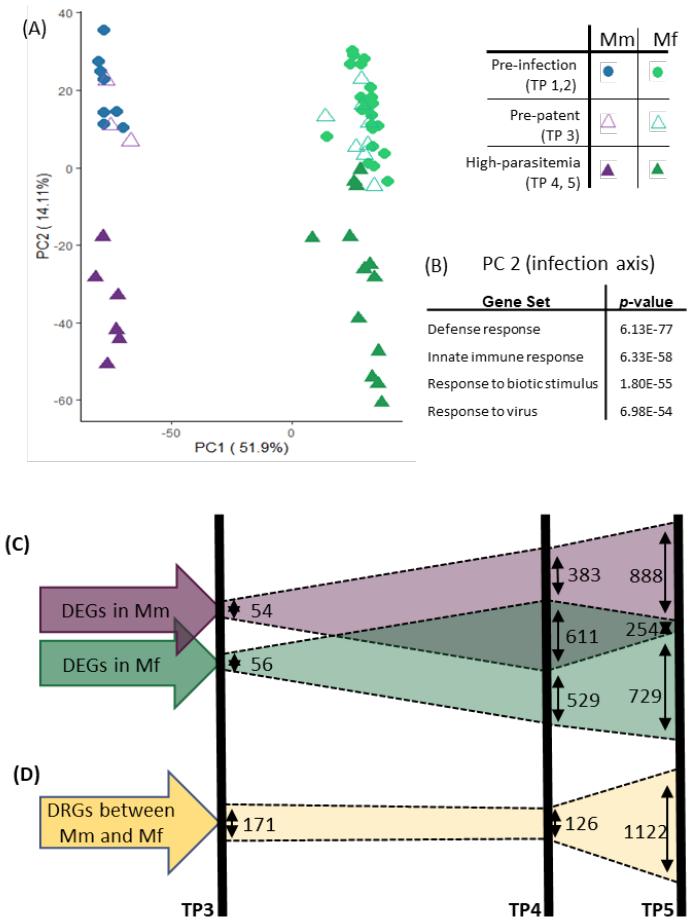


Figure 2.2- Transcriptomics patterns of whole-blood gene expression of all samples at time points TP1 – TP5 from Mm and Mf. (A) Principal component analysis (PCA) of all samples from Mm and Mf. PC1 captures inter-species variance in expression profiles, while PC2 captures temporal variance in expression profiles. Pre-infection samples TP1 and TP2 (before infection) form species-specific clusters that are separate from the samples at TP4 and TP5, reflecting peaking parasitemia. TP3 (3 days after inoculation of sporozoites, and prior to blood-stage parasitemia, i.e., the pre-patent period) samples cluster with the pre-infection samples TP1 and TP2. (B) Top GO gene sets over-represented in PC2. (C) Differentially expressed genes (DEGs) and (D) Differentially responding genes (DRGs) in TP3-TP5. The numbers of DEGs change significantly between TP3 (~50) and TP4 (~1000-1140) for both Mm (purple) and Mf (green), and then remain similar between TP4 and TP5 (~1000-1100) for each species. However, the number of common DEGs between Mm and Mf decreases substantially from TP4 (611) to TP5 (254). Concomitantly, the DRGs between TP4 and TP5 (gold) increase substantially.

It is worth noting that the same PC2 axis crisply separates pre-infection (TP1, TP2) and acute infection (TP4, TP5) samples for both host species even though the disease progression is

different, suggesting fundamentally similar transcriptional responses. To gain deeper insights into the details of PC2, we identified enriched Gene Ontology (GO) gene sets along this axis (Figure 2.2B). The most enriched gene sets are associated with defense and innate immune responses ($p \approx 10^{-70}$) and with a response to cytokine and biotic stimuli (Appendix Figure A.1.2).

To characterize key features in response to the *P. knowlesi* infection in each host species, we identified significant intra-species and inter-species transcriptional changes. For intra-species analyses, we identified differentially expressed genes (DEGs) for each infection time point (TP3, TP4 and TP5) compared to the pre-infection baseline samples (TP1 and TP2). For inter-species analyses, we identified genes that responded differentially to infection, as explained further in our supplementary data (Appendix Figure A.1.3). These differentially responding genes (DRGs) provided contrasting differences in infection responses between the species at each time point.

As early as TP3, when the infection is still confined to the liver, some changes in blood transcriptome were identified in each host species. These are statistically significant, although they are quantitatively much smaller than the changes that occur when parasitemia is rising and peaking (TP4 and TP5, respectively). While many DEGs are shared between the two species at TP4, there is greater divergence in the transcriptional profiles at TP5, as indicated by fewer shared DEGs (Figure 2.2C). This divergence is supported by the results of our inter-species analyses, where we observe only a few DRGs up to TP4, but a substantial increase in DRGs at TP5 (Figure 2.2D).

2.4.2 Evolutionary distance of homologous genes does not account for the differential responses observed in the Macaca mulatta (Mm) and Macaca fascicularis (Mf) host species

We hypothesized that there might be a relationship between the evolutionary divergence of homologous Mm and Mf genes and their expression profiles in the two hosts, since such

divergence could be explained by evolutionary pressure and possibly underpin the differences in the control of parasitemia and in the different outcomes observed between the species. To test this hypothesis, we estimated the evolutionary distance between homologous genes in the two species (Appendix Table A.2.7) and compared the similarity scores of DRGs at each time point.

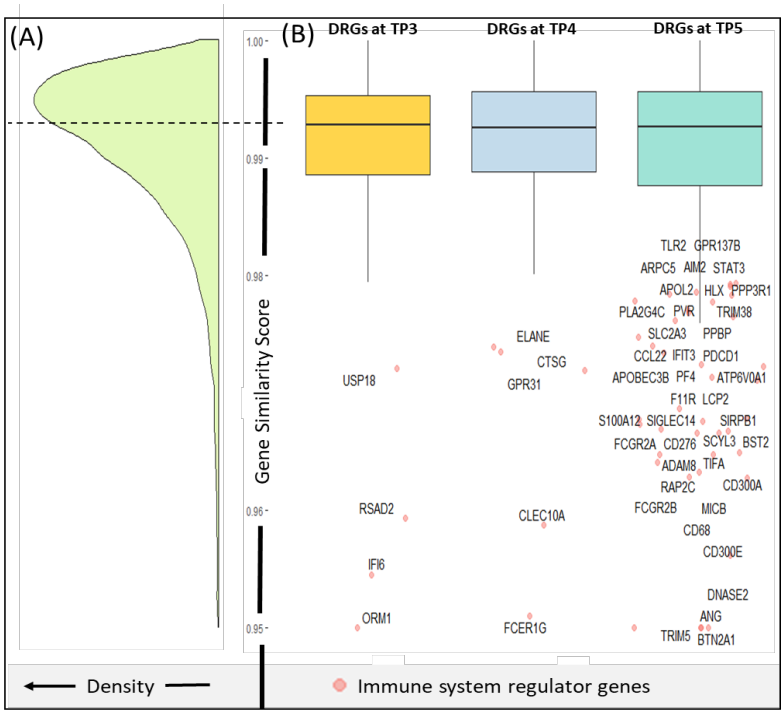


Figure 2.3- Evolutionary distance of homologous genes. (A) Density plot showing the distribution of genes across similarity score of ~15000 homologous genes. (B) Box plot comparing the gene similarity score of DRGs at TP3, TP4 and TP5. Although the mean similarity score across DRGs at TP4 and TP5 is not different as compared to all genes, GO annotation of the outlier genes (Red points) suggests numerous genes being involved with immune regulation.

The evolutionary hypothesis was ultimately rejected given a gene similarity score of all genes that was not significantly different (Kolmogorov-Smirnov p-value > 0.1) from that of the three sets of DRGs (Figure 2.3). However, it is interesting to note that genes involved in regulation of immune system processes were overrepresented in the outliers at each time point, with a hypergeometric p-value of 3.9×10^{-18} and 3.29-fold over-enrichment in the outliers compared to

expectation in all immune system genes. Of particular interest are immune system related genes (FCER1G and ELANE), cytokines and growth factors (CCL22, CTSG, PF4 and PPBP), transcription factors (STAT3, TRIM38 and HLX) and various cell differentiation markers (BST2, CD276, CD300A, CD68, CLEC10A, F11R, FCGR2A, FCGR2B, PDCD1, PVR, SIRPB1 and TLR2).

2.4.3 Gene set enrichment analysis indicates distinctive gene expression profiles between the species by TP5

To identify well-defined, significantly enriched gene sets with respect to *P. knowlesi* infection, we performed gene set enrichment analysis (GSEA) using significant DEGs (adjusted $p < 0.01$, log2 (absolute fold change) > 1) at each time point. We used Hallmark gene sets from the Molecular Signature database[102] for an overview analysis (Appendix Table A.2.1) followed by a detailed analysis using GO annotated biological pathways. Intriguingly, by TP3 a few DEGs were identified that only showed significant enrichment in the Mf. Although the number of DEGs is similar in both species, Mf has several Hallmark gene sets significantly enriched (Appendix Table A.2.1). Importantly, these results suggest that by 3 dpi, when parasites solely reside in the liver, Mf is already mounting interferon (IFN)-mediated immune responses (IFNα and IFNγ) against the parasite. Relevant genes involved in the response include IRF7, CCL22, CXCL12 and PML. In-depth GO analysis (Appendix Figure A.1.4) indicates that the IFN responses are characterized by pathways known to regulate viral genome replication and the cytoplasmic pattern recognition receptor (PRR) signaling pathway, known to indicate the presence of foreign genetic material. Differences between the two species are less evident by TP4, when parasitemia is rising. Subsequently, both species show pronounced enrichment for IFNα and IFNγ immune responses

and several signaling pathways, including NFκB and IL6-JAK-STAT3 (Appendix Figure A.1.5, Appendix Table A.2.1).

In-depth GO analysis shows that while many aspects of the response at TP4 are similar for the two species, including the typical response to viruses, Type-1 IFN production, and urea catabolism, some notable differences exist (Appendix Figure A.1.6). For instance, Mm exhibits prominent regulation of calcium ion transport along with changes in certain metabolic pathways including amino acid metabolism, protein catabolism and cytokine metabolism. In Mf, by contrast, cellular amide metabolic processes (urea catabolism) and additional immune response pathways triggered by foreign organisms dominate the response. Interestingly, Mf also shows enrichment of an adaptive immune pathway.

At TP5, which is only 2 days later than TP4 (Figure 2.1A), the differences between the species in their most significantly enriched pathways become much more profound (Figure 2.4, Appendix Figures A.1.5, A.1.7; Appendix Table A.2.1). Mm continues to express mainly immune response genes related to cytokine secretion, leukocyte activation and responses to the presence of foreign organisms, while the Mf gene expression profile shifts dramatically.

Interestingly, some of the adaptive immune pathways previously only identified in Mf at TP4 (Appendix Figure A.1.6) are evident in Mm at TP5 (Appendix Figure A.1.7). Beyond these differences, the two hosts have some pathways in common at TP5, mostly associated with immune responses to inflammation and metabolic pathways regulating cell cycle and protein modification.

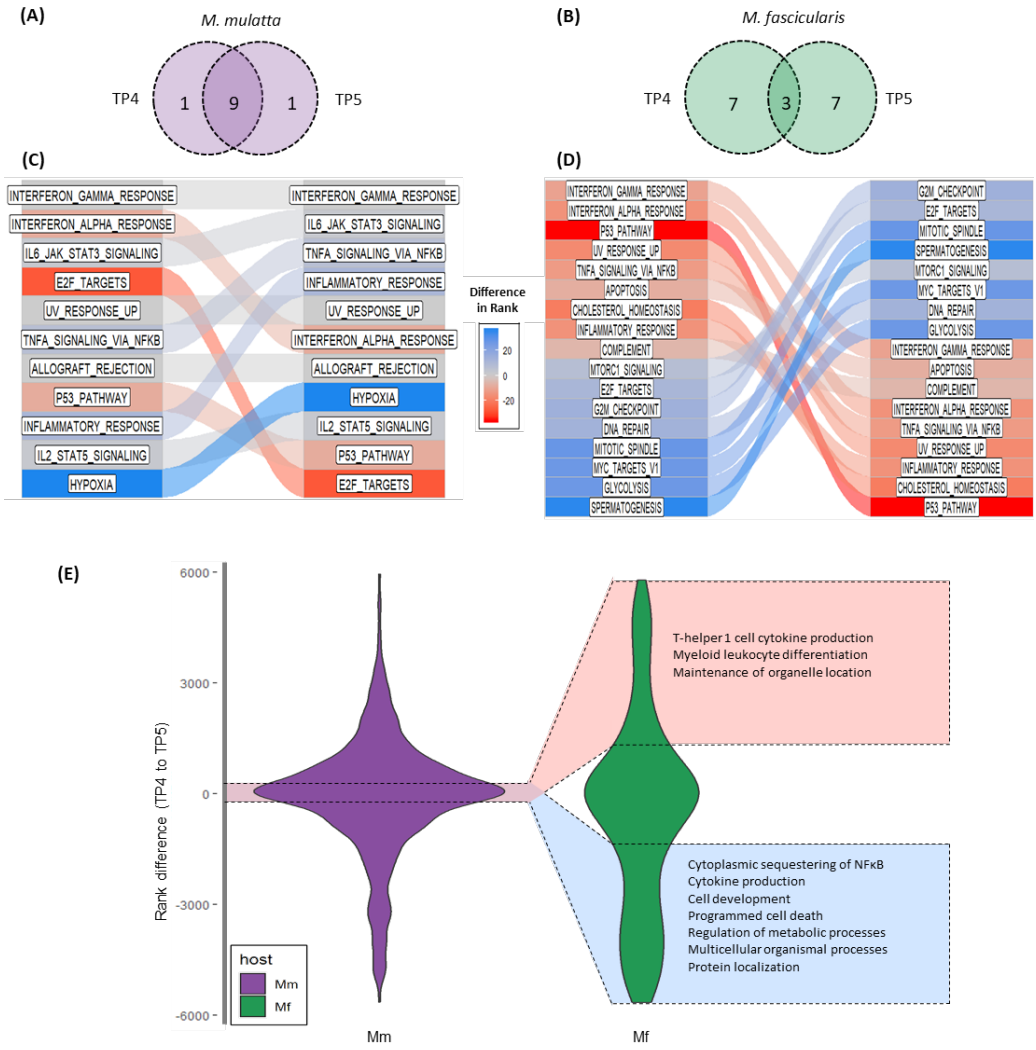


Figure 2.4- Changes in enriched pathways from TP4 to TP5 in Mm and Mf. (A, B): Venn diagrams of top 10 enriched pathways at TP 4 and TP5 for Mm and Mf, respectively. In Mm, 9 out of the top 10 are common between the two TPs. In Mf, only 3 out of the top 10 are common. (C, D): Ranking, from top to bottom, of most enriched pathways in Mm and Mf, respectively, in TP4 and TP5 (left to right). For each host, the changes in ranks of the top pathways between TP4 to TP5 are shown. Each list contains the union of the most significantly changed gene sets, ordered by their enrichment score. The top ranked gene set is the most enriched. Connectors show changes in rank for each gene set between the two TPs. Red shading indicates increases in rank and hence decreased enrichment from left (TP4) to right (TP5). Blue shading indicates decreases in rank and thus increased enrichment from TP4 to TP5. Color shades are proportional to relative changes in rank considering all gene sets (not just the ones shown). For instance, the darker shade of red for p53 pathway in panel D represents a sharp decrease in rank. (E): Violin plots of rank differences, representing the importance of gene sets, between TP4 and TP5 in the two host species. The vertical axis represents the distribution of rank differences between TP4 and TP5 for GO

gene sets in Mm and Mf. Each data point in the two distributions represents one GO gene set (see Methods for details). The transition from TP4 to TP5 in Mm is characterized by a distribution (shown in purple) with 0 mean and a broad, rather than narrow distribution which corresponds to relatively small changes in the importance of most gene sets. By contrast, the distribution of ranks in Mf (shown in green) has a much narrower distribution at the mean, with heavy tails, indicating many more changes in rank, overall. Both red and blue domains represent pathways that are important in Mm during both TP4 and TP5. In Mf, the red domain represents pathways that were not important at TP4 (higher ranked / lower enrichment) but become more important at TP5 (lower ranked / higher enrichment), while the blue domain represents pathways that were important at TP4 but become less important at TP5. The striking difference between these distributions demonstrates that Mf alters and refocuses its gene expression profile between TP4 and TP5 towards cell proliferation, etc. In contrast, Mm's gene expression remains almost unchanged, still emphasizing the immune response.

Upon further analysis of the differences between TP4 and TP5 (Figure 2.4), a dramatic contrast emerges: the response of Mm at TP4 and TP5 is very similar, whereas Mf shifts into a different phase of its response after TP4, characterized by what seems like an effort towards rehabilitation and recovery. The immune responses in Mf are still evident at TP5, but gene expression in this species notably shifts toward cell proliferation and cell division functions, highlighted by DNA replication, chromosome segregation, organelle fission and localization pathways. One of the noticeable enrichment changes in Mf is in the p53 pathway (Figure 2.4D), with the timely enrichment of this stress response pathway by TP4, and its ceased enrichment by TP5. This observation can be regarded as a precursor for DNA replication, cell division and cell proliferation pathways[113]. Moreover, we were able to extract more information about previously identified cell cycle related pathways (Figure 2.4E). Myeloid leukocyte differentiation hints at upregulation of specific cell populations in Mf. It is worth noticing that Mf downregulates cytokine production while upregulating Th1 cell cytokine production, which is a part of the adaptive immune response. Enrichment of the JAK-STAT pathway (Appendix Figure A.1.5) in Mm along with its cross-regulation with both IL10 and IL6 suggests duplicated efforts[114]. Differential expression of the SOCS3 gene in both Mm and Mf suggests upregulation of the JAK-STAT

pathway but the two-fold regulation could mean impaired functionality in Mm. Overall, as discussed further below, the GSEA reveals a prominent difference in molecular responses that might be responsible for the diverse outcomes of the two macaque hosts.

In sum, Mf detects the pathogen sooner than Mm and is able to balance its immune response and inflammation in the face of higher parasitemia levels. This earlier response is consistent with the conclusions of Peterson *et al.*[83] from their analysis of clinical, parasitological and immune response data for these infected animals. Interestingly, in Mm, a strong enrichment of Ca^{2+} ion transport might be playing an important role in pathogen survival as it maintains Ca^{2+} homeostasis and aids the parasite's Ca^{2+}-based signaling, which is critical for parasite growth and differentiation within infected RBCs, and their egress and invasion of new host RBCs during the blood stage of the disease[115]. The transition from TP4 to TP5 highlights key differences in immune responses between the two species. Gene expression at TP4 has significant similarities between the two species, although there are some key differences. Then the species' responses diverge. Figures 2.4, Appendix Figures A.1.5 and A.1.8 demonstrate that while Mm does not seem to change its response much, Mf mounts a radical makeover in expression profiles in important pathways between TP4 and TP5, and these changes appear to be among the most crucial differences observed between the two species. In contrast to Mf's response, Mm displays fairly small changes among the most enriched gene sets during the transition from TP4 to TP5, exceptions being an increased focus on hypoxia-related genes and decreased importance of E2F target genes at TP5.

2.4.4 Analysis of transcription factors reveals prominent regulators that distinguish the immune response of the two hosts

To characterize the differences in gene expression further, we identified the transcription factors (TFs) and other regulatory proteins that might orchestrate changes in these coordinated gene programs. These TFs and the gene networks they control can be found by searching for upstream regulators of DEGs in the databases iRegulon[109] and TRRUST DB[110] (Sec. 2.3.6). The most significantly enriched TFs, according to a combination of the two databases, are shown in Figure 2.5A; an exhaustive list of iRegulon TFs identified is presented in Appendix Table A.2.3.

Consistent with the results presented above, the transcriptional data from Mf indicate significant enrichment for TF activity as early as TP3, and this response is absent in Mm (Figures 2.5A, 2.6). In particular, IRF7 and IRF4 are enriched in Mf at TP3 along with TRIM25, STAT2 and STAT1. In contrast to Mf, the corresponding Mm data only indicate slight enrichment of NFIL3 activity, which might suggest the emergence of precursors of common helper innate lymphoid cells; however, the signal is too weak to warrant a definitive claim.

Several other pertinent TFs are enriched at TP4, especially for Mm, where the parasitemia continues to rise unabated. In Mm, across TP4 and TP5, STAT3 is persistently activated, which is brought about by inflammatory cytokines. At least in humans, and presumably also in NHPs, enrichment of NFκB signaling related TFs – NFκB1, REL and RELA points toward the canonical NFκB signaling pathway[116]. The persistent activation of the anti-apoptotic pro-inflammatory NFκB pathway along with the opposing p53 pathway suggests unduly extended inflammation might contribute to Mm's severe, and indeed life-threatening systemic illness in response to the acute *P. knowlesi* blood-stage infection. Mm continues to show these immune responses at TP5

with additional activation of STAT4, a sign of IFN production by dendritic cells (DCs)[117]. In contrast, and paralleling the observations in gene sets, the immune response of Mf at TP5 is showing signs of recovery, with enrichment of E2F4, TFDP1, FOXM1, E2F1 and TP53, along with other TFs that are involved in balancing quiescence and cell cycle activation. The Mf furthermore enhances cell cycle related pathways by promoting both early stage (E2F1 and TFDP1) and late stage (FOXM1) cell cycle processes.

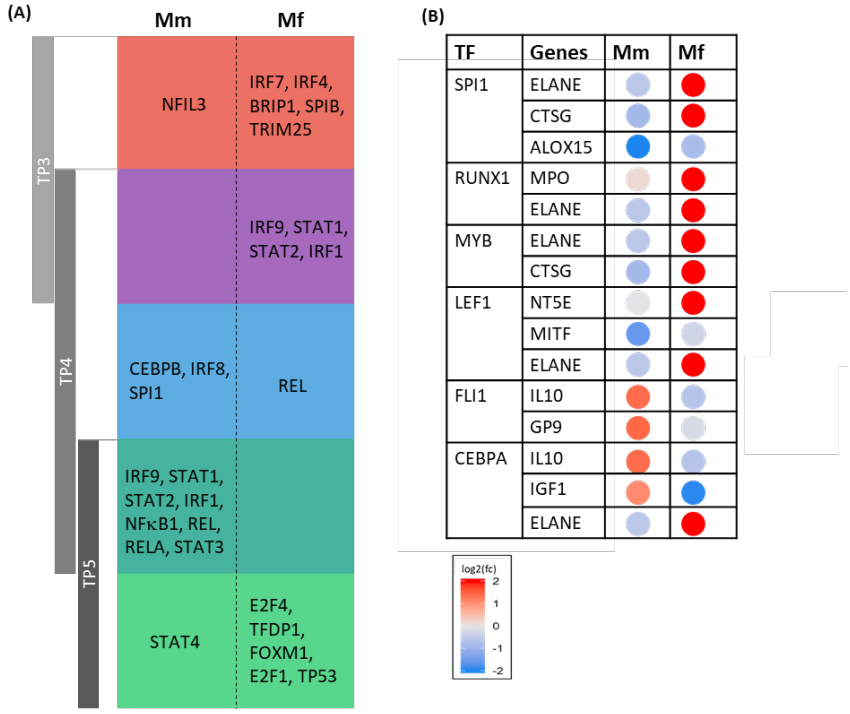


Figure 2.5- Enriched TFs. (A) Most significantly enriched TFs at each TP for the two host species. Mm has a disjointed set of TFs between TP3 and TP4 along with substantial overlap between TP4 and TP5. Mf, on the other hand, has substantial overlap between TP3 and TP4, which is then disjointed with TP5. (B) TFs enriched by DRGs at TP4 along with the corresponding DRGs. The heatmap shows differential expression of these genes at TP4 with respect to baseline.

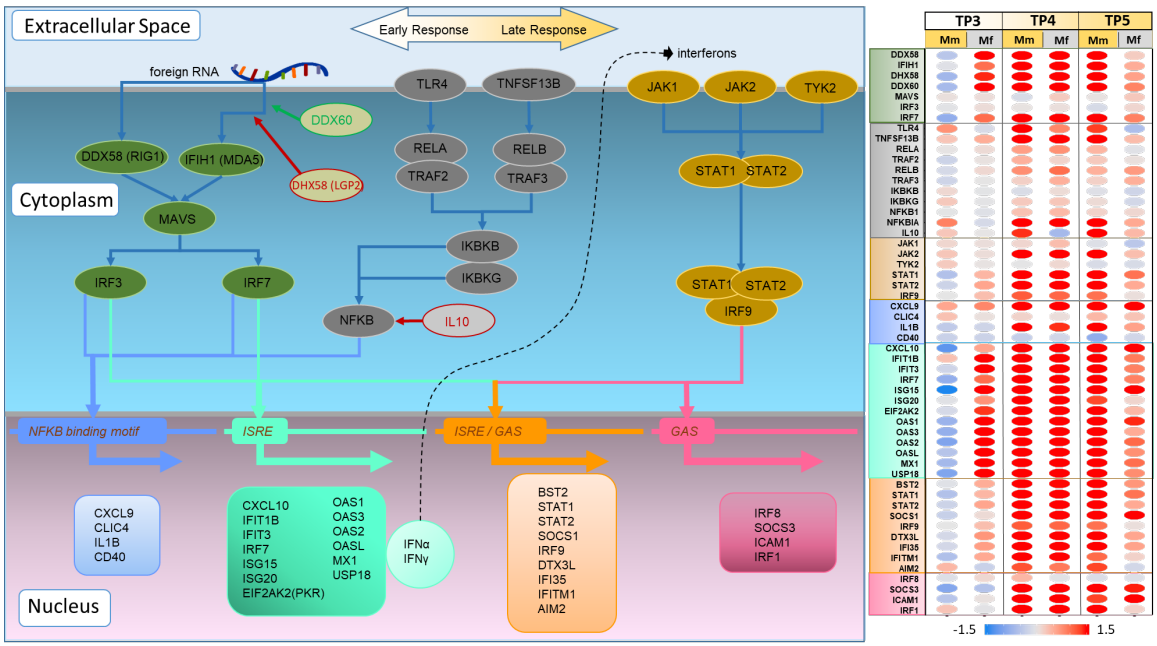


Figure 2.6- Transcription factors and associated genes of signaling pathways during the immune response. The TFs and genes shown are associated with differences in DEGs between early and late responses to infection. DEGs and their magnitudes are shown on the right for TP3, TP4, and TP5, with columns for each of Mm and Mf. Red dots signify up-regulation and blue dots down-regulation. In addition to DEGs at TP4 and TP5, the RIG1/MDA5-mediated PRR signaling pathway is included, as it is significantly different at TP3. IRF7-regulated genes expressed by ISRE also show significant changes. Finally, NFκB signaling, mediated by REL and RELA, plays a crucial role in controlling inflammation, and the ongoing strongly differential expression of the corresponding genes (ISRE- and GAS-regulated) at TP5 in Mm suggests a prolonged inflammatory response.

The intersection of TF sets at TP4 with those at TP3 and TP5 in both species (Figure 2.5A) distinguishes phases of infection in the two hosts. In Mf, TP4 shares many active TFs with TP3, but not TP5. This dramatic shift in TF profiles by TP5 in Mf could predictably allow the host to counteract the expansion phase of *P. knowlesi* by slowing down the inflammatory response and initiating recovery pathways. Mm, by contrast, essentially lacks an immune response at TP3. In Mm, TP3 also does not share any enriched TFs with TP4, whereas the sets of active TFs overlap substantially between TP4 and TP5, suggesting that the delayed TF program is sustained until TP5, resulting in continued inflammation and/or the lack of an appropriate immune response.

To highlight the main driving factors that differentiate the immune response in the two hosts at TP4, we identified the TFs enriched by the DRGs (Figure 2.5B). These include TFs like RUNX1, SPI1, LEF1, FLI1 and CEBPA, which play important roles in hematopoiesis and lymphocyte differentiation. Interestingly, the affected genes suggest involvement of RUNX1, SPI1 and LEF1 in Mf while FLI1 and CEBBPA play a crucial role in Mm, including upregulation of IL10.

2.4.5 Modular transcriptional repertoire analysis provides further insights on gene expression differences between the Macaca mulatta (Mm) and Macaca fascicularis (Mf) host species

The analysis of transcriptional modules yielded general agreement with our prior results, along with some surprising new insights (Figures 2.7, Appendix Figures A.1.9, A.1.10; Appendix Tables A.2.4, A.2.5, A.2.6). TP3 shows significant enrichment of IFN modules in Mf (M10.1, M15.127 and M8.3; Figure 2.7). Type-1 IFN module M8.3 suggests induction of antiviral effector genes like MX1 in addition to IFN modules M15.127 and M10.1, which may indicate immunopathology with host-pathogen interactions. M15.127 and M10.1 also point to a pathogen associated molecular pattern (PAMP), specifically, double-stranded RNA. Mm meanwhile shows a relatively modest enrichment of module M15.113, which is related to IL-1-mediated inflammation (Appendix Figure A.1.9).

TP4 shows much similarity between the two hosts with respect to inflammation, IFN and cytokine-related modules (M13.16, M13.1, M10.1, M13.17, M15.127, M15.64, M15.86 and M8.3). However, some modules show distinguishing behavior. Down-regulation of the B-cell module M13.27 suggests possible inhibition of T cells with downregulation of CD96 and LY9. Among the erythrocyte modules, the difference in M13.30 suggests a relative difference in

hematopoiesis by megakaryocyte erythroid progenitor markers like BLVRB, SLC25A39, HBM and HBQ1. The prostanoids module M8.2 suggests differences in platelet activation through enrichment of genes like PPBP, GP9, and others. Certain cell cycle related modules also suggest differentiating biological behaviors between the two hosts. Mm shows downregulation of a mitosis-related module (M13.32), while its upregulation in Mf suggests cell division.

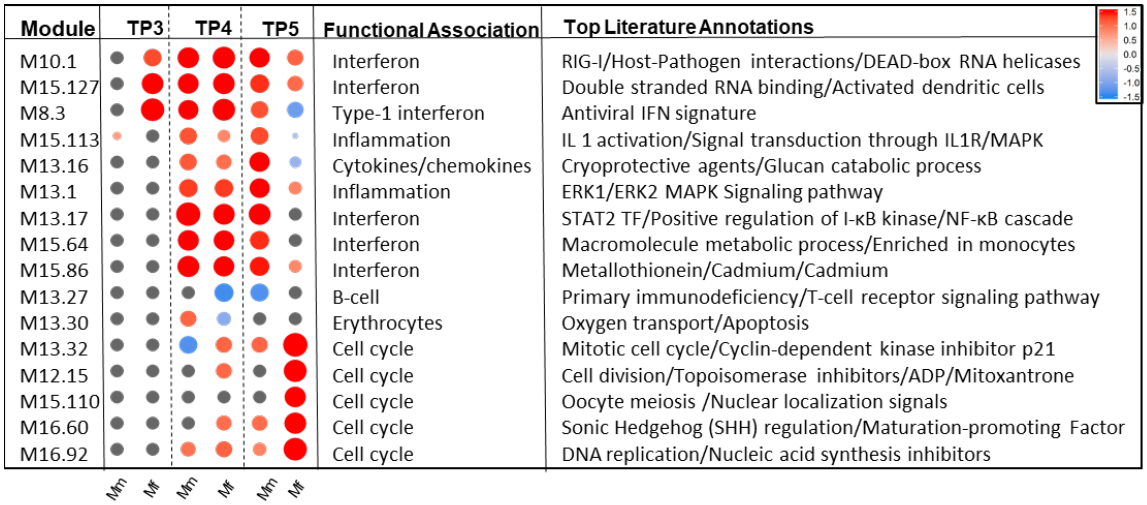


Figure 2.7- Heat map of most enriched transcription modules at each TP for both hosts. The heat map represents the enrichment score (ES) for each of the modules. Relevant information related to each module, for example pertaining to functional associations and literature annotations, is presented on the right; further details and the entire list of modules with enrichment is presented in the Appendix A, including Appendix Table A.2.4.

Mf exhibits several enriched cell cycle-related modules that become even more significant at TP5 (M12.15, M13.32, M15.110, M16.60 and M16.92). At TP5, Mm still expresses enriched IFN modules that are now downregulated in Mf as compared to TP4. These enrichments align well with previous results showing that the inflammatory immune response in Mf subsides by TP5.

2.4.6 Cell population markers

To complement these results, we analyzed single-cell reference gene markers to explore hematopoiesis and identify enriched cell types in blood samples (Table 1). These gene marker data indicate that non-classical monocytes are enriched in both hosts at TP4. Anti-inflammatory effects of non-classical monocytes include vascular homeostasis and are the first line of defense in terms of pathogen detection and clearance[118]. Mm shows enrichment of progenitor cells for both NK and B cells. By TP5, Mm exhibits enrichment of erythrocytes and neutrophils as well. Mm may amplify cytokine signaling using NK cells. Activation of intermediate monocytes suggests their involvement in dealing with oxidative stress created by the infection. Interestingly, enrichment of hematopoietic stem and progenitor cells (HSPCs) in Mf is concentrated to the G2-M cell cycle phases suggesting proliferation of immature myeloid progenitors.

Table 1- Gene markers indicate enriched cell populations (and sub-populations) in the two species at different TPs during the infection. HSPC: hematopoietic stem and progenitor cell; NK cells: natural killer cells; CLP: common lymphoid progenitor

	Mm	Mf
TP4	- HSPC - Pre-B/NK - Monocyte - Non-Classical - B cell - Pro-B - NK cell - NKP	- Monocyte - Non-Classical
TP5	- Erythrocyte - ERY1, ERY/GRA2 - NK cells - Cytokine NK - Monocyte - Intermediate - Neutrophil - Meta-Myelocyte/ Mature Neutrophil, Myelocyte	- Monocyte - Pre-Monocyte - B cell - Cycling Pre-B - NK cell - CLP - HSPC - G2M

2.5 Discussion

The comparison described in this chapter between a disease-resilient host (Mf) and a highly vulnerable host (Mm), both infected with the same parasite species (*P. knowlesi*), was designed to yield insights into the host transcriptional programs associated with biological pathways that play significant roles in such infections. The identification and characterization of critical molecular and cellular differences between resilient and non-resilient hosts is an important step toward understanding the mechanisms of host-parasite responses.

The Mm and Mf species are separated evolutionarily by fewer than three million years and their geographical distribution areas overlap slightly (Figures 1.1, 1.2). We speculated that querying the evolutionary distance of homologous genes could potentially offer insights into the resilience of Mf. However, no correlation emerged in the two species between the sequence-level similarity of homologous genes and their differential response as parasitemia was rising. Interestingly, nevertheless, was the finding that the immune system-related DRGs are overrepresented in the outliers. While we performed a detailed analysis and interpretation of DRGs, a comprehensive evolutionary analysis might help identify key immune regulation checkpoints that should be queried in future work as they point to fundamental evolutionary differences in the immune responses of the two hosts.

Our analyses revealed that the immune system of Mf (but not Mm) senses the presence of foreign organisms as early as 3 dpi. It is not surprising that Mf shares many common TFs with Mm at TP4, but it is interesting to note that prominent TFs generally associated with immune responses are solely activated in Mf as early as TP3 (Figure 2.5). The first line of response in Mf is detection of infected hepatocytes via cytoplasmic PRR (cytokine) signaling via MDA5 and

RIG1, which activates the TFs IRF3 and IRF7[119,120]. These signaling pathways activate the innate immune response, led by IFNα, which starts almost simultaneously with the pro-inflammatory response that is led by IFNγ[121,122]. In the context of responses to viruses, this type of detection is known to cause IFNα-mediated downregulation of viral genome replication[123]. Taken together, this early cytokine signaling along with IFN responses appears to be a crucial response signature in Mf that is missing or delayed in Mm, potentially rendering Mm more vulnerable to the infection (Figure 2.6). This chain of events is revisited with further consequences in Chapter 4.

Upregulation of cytoplasmic PRRs and the MDA5 signaling pathway in Mf at TP3 also marks the onset of a pro-inflammatory innate immune response led by IFNγ signaling. This response, along with regulation by certain transcription factors, including IRF7, STAT1, STAT2, and IRF4, leads to elevated cytokine production. IRF7 efficiently activates both IFN-α and IFN-β genes[124,125] via the Interferon Sensitive Response Element (ISRE), which is clearly shown by corresponding gene expression that is activated in Mf at TP3 (Figure 2.6). Interestingly, IRF7 has a short half-life (~0.5–1 hr)[126] due to its susceptibility to ubiquitin-dependent degradation[127]. The labile nature of IRF7 may represent a mechanism critical to rendering the entire IFN gene-induction process transient, preventing overexpression of IFNs and harm to the host. Activation of the lymphoid-specific enhancer Spi-B transcription factor (SPIB) in the blood could either indicate the production of Type-1 IFN by plasmacytoid dendritic cells (pDCs) or IgM by mature B cells[128], both of which have significant roles in the immune response.

The *P. knowlesi* infection transitions from hepatocytes to RBCs after about 5 days of parasite multiplication in the liver[45,46]. The following period of parasite multiplication in RBCs and the rising anemia and parasitemia are characterized by the most significant responses in terms

of the number of DEGs in both species compared to baseline expression. All transcriptomic immune response signatures are consistent with the presence of pathogens in the blood, and judging by the small number of DRGs at TP4, both species launch a similar response. Against this background of similarity in transcriptomic responses, subtle differences are apparently very important.

In both species, the rise in parasitemia at TP4 is marked by an elevated immune response via IFNα and IFNγ. The IFNγ-inducible genes are carefully orchestrated by the transcription factors STAT1, STAT2, and IRF9 to respond appropriately to the specific needs of the cell. This immune response is complemented by the IL-6 regulated JAK-STAT3 signaling pathway, which controls cytokines like erythropoietin, thrombopoietin and G-CSF and thereby may be involved in dealing with pathological conditions like anemia, thrombocytopenia, and neutropenia or, alternatively, the generation of antibodies through B-cell and plasma cell differentiation.

The innate immune system in the host spearheads the immune response not only by developing protective immunity but also by aiding the host in dealing with pathogenesis[129], and the initial pro-inflammatory response attempts to clear the infection. However, when elevated and prolonged, inflammation leads to physiological deterioration and increasing severity of various pathological conditions. The fact that pro-inflammatory genes of the IFNγ pathway are more strongly elevated at TP4 in Mm than Mf suggests prolonged inflammation only in Mm, despite similar levels of parasitemia in both species at that time point. IL-10 is a chief anti-inflammatory receptor and affects various pathways to inhibit inflammatory cytokines like IL-6, IL-1 and TNFα. Substantial upregulation of IL-10 combined with enriched B-cell subpopulations (Pro B and Pre B) at this phase of infection is again suggestive of Mm's attempt to stem the inflammation and

fight the infection. In contrast, the presence of significantly enriched non-classical monocytes along with differentially expressed TNFα, IL1β and CCL3L3 is indicative of a direct response to the pathogen via the MyD88-MEK pathway[130]. Non-classical monocytes are known to produce more TNFα and less IL10, and timing of their enrichment might be crucial for the outcome.

Time point 5, representing the final days of high parasitemia, reveals the most dramatic differences between the species. Mm continues with its prolonged immune response with proinflammatory signatures, while Mf initiates a program of cell proliferation with the transcription of multiple genes involved in DNA replication and repair, mitosis, and cell cycle progression; all these suggest the onset of a recovery phase in the Mf. Enrichment of HSPCs in their G2M phase could support replenishment of lost cells and reinforcement of the immune response. Notably in Mf, the p53 pathway is significantly downregulated at TP5 in comparison to TP4, consistent with the goal of preventing p53 targets from hindering these cell proliferation pathways[131]. The gene TP53 is known to be activated in response to DNA damage and oxidative stress. It assists in apoptosis of damaged RBCs and maintains adult stem cell niches. In patients with malaria, TP53 has been shown to modulate inflammatory responses to infection[132]. Hypoxia activates this p53 signaling, and, indeed, hypoxia levels are higher in Mm than Mf, suggesting that higher hypoxia levels, sustained for a prolonged time, may contribute to the extended stress response of the p53 pathway. If so, the p53 pathway might be an important yardstick for inflammation. Sustained inflammation and a delay in upregulation of cell development pathways in Mm is in fact revealing with regard to the mechanistic explanation for Mm's deterioration, compared to Mf's resilience and recovery, with the added major concern that infected Mm are rapidly running out of healthy, uninfected RBCs.

Unlike GSEA, which uses curated gene sets to define a function or process, the 'modular transcriptional repertoire' is derived from multiple (large-scale data) samples that display perturbed responses caused by various diseases or pathogens and pertain to a particular tissue (in our case white blood cells (WBCs))[133]. Analysis of whole modules rather than individual genes can increase our ability to detect interesting changes by decreasing the impact of multiple hypothesis testing and considering the inter-dependence of different transcript profiles. Interestingly, pre-patent phase modules in Mf are known to be associated with the detection of viruses (Appendix Table A.2.4), especially influenza, which has been frequently confused with malarial infection due to early infection symptoms[134]. In line with duplicated efforts of IL10 and IL6 with respect to the JAK-STAT signaling pathway, differential responses of inflammatory modules related to the ERK1/2 MAPK signaling pathway suggest a possible role of SOCS3 or STAT3 in the extension of inflammation[135,136].

The transcription differences between two closely related species, infected with the same parasite, offer hints to why one species when infected with *P. knowlesi* faces severe life-threatening disease and requires aggressive treatment, whereas the other becomes sick but rebounds without the provision of antimalarial drugs, as further detailed in Peterson *et al.*[83]. A noteworthy component of the difference appears to be the delayed detection of the parasites in Mm and the consequently delayed initial immune response in Mm, which comes too late for this species to recover.

Our analysis has identified interesting changes in molecular profiles between Mm and Mf. Not surprisingly, it has shortcomings, which are by and large due to the infrequency of sampling in our longitudinal study, which in turn was dictated by regulatory blood draw limitations. Since most of the differentiating factors between the two macaque species point to the critical timing of

pathogen detection and the timely switch of Mf's transcriptomics program toward recovery, an iterative longitudinal study with denser sampling around these times would most likely allow more refined and definitive claims. Such experiments are warranted, particularly since our study results highlight promising prospects that one might use for the development of future anti-malarial treatments and vaccines. Many potential adjuvants for anti-malarial vaccines are under investigation, which mostly seek to target PRR signaling via TLR agonists[137]. Our analysis provides further mechanistic support for the application of such vaccines. During the later phases of the infection, IL10- and p53-related pathways could provide interesting drug targets. While it might be challenging to control IL10 due to its numerous and diverse roles, targeting p53 has already been demonstrated to attenuate malarial inflammation and protect from fever[132].

We analyze the differences in PRR signaling, inflammation and p53 mechanism in more details in Chapter 4 to explain their molecular mechanisms better. Metabolomics data created by MaHPIC experiments could also be integrated with transcriptomics analysis to provide additional insights into these mechanisms. As an example, in Chapter 3 we develop a transcriptomics-based metabolic modeling method that combines two rich data resources. We develop the same methodology, using a different experimental setup for malarial infection in macaques, and apply the procedure in Chapter 4 to compare, contrast and extend the findings described there.

CHAPTER 3. INTEGRATION OF METABOLIC MODELING AND TRANSCRIPTOMICS TO EXPLAIN METABOLOMICS CHANGES DURING MALARIAL INFECTION[2]

3.1 Summary

Disease represents a specific case of malfunctioning within a complex system. Whereas it is often feasible to observe and possibly treat the symptoms of a disease, it is much more challenging to identify and characterize its molecular root causes. Even in infectious diseases that are caused by a known parasite, it is often impossible to pinpoint exactly which molecular profiles of components or processes are directly or indirectly altered. However, a deep understanding of such profiles is a prerequisite for rational, efficacious treatments. Modern -omics methodologies are permitting large-scale scans of some molecular profiles, but these scans often yield results that are not intuitive and difficult to interpret. For instance, the comparison of healthy and diseased transcriptome profiles may point to certain sets of involved genes, but a host of post-transcriptional processes and regulatory mechanisms renders predictions regarding metabolic or physiological consequences of the observed changes in gene expression unreliable. Here we present proof of concept that dynamic models of metabolic pathway systems may offer a tool for interpreting transcriptomic profiles measured during disease. We illustrate this strategy with the interpretation of expression data of genes coding for enzymes associated with purine metabolism. These data were obtained during infections of rhesus macaques (*Macaca mulatta*) with the malaria parasite *Plasmodium cynomolgi* or *P. coatneyi*. The model-based interpretation reveals clear patterns of

[2] Tang, Y., Gupta, A., *et al.* Metabolic modeling helps interpret transcriptomic changes during malaria. Molecular Basis of Disease (2018)

flux redistribution within the purine pathway that are consistent between the two malaria pathogens and are even reflected in data from humans infected with *P. falciparum*.

3.2 Introduction

In contrast to many other complex diseases, such as cancer, Crohn's or metabolic syndrome, infectious diseases have the distinction of a clear root cause: a pathogen has invaded the body and was not stopped by the host's natural immune defenses. If the pathogen can be eliminated with medical or pharmaceutical means, the disease has a straightforward cure. However, in many cases this is not directly possible, or it requires a relatively long period of time, during which the patient is at risk of deteriorating, with possibly lethal consequence. In these cases, the root cause becomes almost immaterial, and it is the complex system of interactions between the pathogen and the host that needs to move to the center of attention [138]. The intriguing aspect of these interactions is that we often have no real understanding of which specific subsystems in the host or the pathogen are turned on or off, so that any molecular characterization of the disease, or any attempt of a targeted intervention, becomes an enormous challenge. As a case in point, while malaria is initially a disease of the blood, it quickly affects other tissues and organ functions and triggers uncounted responses of the host's defense systems. To measure the complete molecular state of a person is therefore an unsurmountable problem. In order to better understand the molecular functions and changes during infection, this chapter discusses a methodology that benefits from reliable omics data and metabolic modeling to provide a tangible understanding of biological systems.

The presence of a pathogen definitively triggers numerous cellular and humoral immune responses, but the chains of events leading to these responses have remained obscure. The reasons for these gaps in our understanding are manifold. First and foremost, the immune system is

exceedingly complicated. It contains uncounted components in the form of different immune cells and specific proteins, such as immunoglobulins and cytokines, whose roles are not always exactly known. Even a coarse overview of some key aspects of malaria render it evident that the disease is systemic and that large numbers of physiological subsystems interact in a life-or-death effort to control the disease.

The complexity of the host responses poses a grand challenge. Pathological measurements of pathogens and RBCs help to monitor the infection and characterize the degree of malarial anemia. Such measurements can then be used to develop models of specific aspects of the disease[139-141]. Physiological markers like body temperature and blood cell counts reflect the severity of the disease at a high level.

However, if the goal of an analysis is to identify specific drug targets, a much deeper understanding of the molecular events during the infection is required. This necessity of a better characterization of processes is problematic, because it is generally much more difficult to measure events than states. As a consequence, malaria research, as well as other disease investigations, usually must resort to measuring molecular profiles. The good news is that the –omics revolution has rendered it possible to assess molecular profiles incomparably more comprehensively than just a couple of decades ago. For instance, we can relatively easily and reliably measure the expression of most genes, and in the process distinguish between host and pathogen genes. Modern mass spectrometry has rendered it possible to establish profiles of many thousands of native and foreign metabolites and their break-down products, even though it is not always entirely clear how such high-throughput results are to be interpreted. While not yet as definitive as genome analyses, proteomics, metabolomics, and lipidomics offer a glimpse into the abundances of subclasses of

metabolites, proteins, and lipids. Taken together, modern biology allows us to convert small volumes of biological samples into enormous datasets.

The sheer sizes of -omics datasets pose challenges that are new to the field of biology[142]. Namely, it is no longer easy to discern valid information or true signals in the data from uncertainties, variability, and noise. In our case study of malaria, every monkey is a unique individual, and many differences in gene expression within a cohort may simply be manifestations of their genetic make-up and health histories. As a pertinent example, which we will discuss later in this chapter, seemingly similar macaques responded rather differently to infections, as can humans, with some individuals suffering relatively lightly, some very severely, and some not even surviving despite anti-malarial drug interventions[66]. Clearly, humans and monkeys are dynamical entities whose features change over time and reflect prior exposures and experiences. Also, of course, all experiments are burdened with certain inaccuracies, which may not be fully characterizable. As a consequence of these and other complications, the expression of a given gene or protein at a given time point may be suggestive of a biomarker of disease, but it may also be a spurious event.

In this chapter, we describe, as a proof of principle, a computational strategy for approaching the complex questions raised in the previous paragraphs. We will use malaria as an example and discuss how different types of experiments and computational analyses have shed light on unforeseen aspects of the disease. We use several well-studied datasets created by the MaHPIC consortium[60] to develop and test the methodology. Additionally, in Chapter 4 we use this methodology in comparative *P. knowlesi* infections to understand specific details of tryptophan metabolism. It is obvious that we cannot obtain complete or definitive answers to the questions we

had asked at the beginning of our analysis. However, the strategies developed here have practical applications, as discussed here and in next chapter.

3.3 Methods

3.3.1 Generic Data-Based Characterization of a Complex Disease

Extracting information from large -omics datasets has been compared to "drinking from a firehose[143]." Yet, even comprehensive attempts to measure pertinent data are not always sufficient. Within the context of our case study, our Malaria Host Pathogen Interaction Center (MaHPIC; www.systemsbiology.emory.edu/) has been collecting -omics datasets to characterize the processes accompanying malarial infections in non-human primates (NHPs). Although well-equipped and well-funded for a 5-year period, this effort has been encountering complicated obstacles that are typical for investigations of complex diseases.

Even specifically with respect to the –omics of blood, which is much easier to obtain than measurements from other tissue samples, the following limitations arise. First, issues of ethics and animal welfare restrict blood draws from macaques, for example, to 10 milliliters / kilogram / month, or 6 milliliters / kilogram / month if the animal is anemic. This regulation results in a spacing of measurement time points that obviously precludes the assessment of immediate metabolic host responses to the emergence of pathogens in the bloodstream. In fact, one is led to assume that metabolism, measured in this manner, is always in a steady state. It is permissible to obtain blood from the monkeys daily through standard procedures involving ear pricks, where no anesthesia is required, but it surprisingly turned out that blood from this source is metabolically quite different from venous blood[144].

Second, multi-omics approaches are often envisioned to include genomics, proteomics, metabolomics, and maybe other measurements from the same source at the same time. In our case, it is of course possible to subject blood samples to these different -omics measurements. However, these measurements shed light on different blood components. RBCs, which are affected most directly, have no nuclei or mitochondria and therefore no DNA. Thus, "blood genomics" automatically and necessarily excludes about 99% of all blood cells, as it is restricted to WBCs and the parasites that infect and grow in the RBCs. By contrast, plasma proteomics is dominated by typical proteins like albumins, while membrane-proteomic measurements from infected RBCs are often outnumbered by the host RBC membrane proteins. Metabolomics suffers from the fast time scale of metabolic profiles that change very quickly, even with quenching. Thus, the "different -omics from the same source" turn out to yield a heterogeneous gemisch of information. Notwithstanding, this information is very valuable, but it is by far not as straightforward, comprehensive and indicative of the same molecular events as one might naïvely assume.

In spite of this unavoidable, natural heterogeneity, the different -omics measurements can be used to identify correlations and associations with methods of statistical machine learning. For instance, MaHPIC scientists analyzed the transcriptional response in the bone marrow (BM) of macaques of the species *Macaca mulatta* to infection with the malaria parasite *Plasmodium cynomolgi* to test the overarching hypothesis that infected animals during relapses exhibit substantial molecular changes similar to those observed during the primary infection [145]. Contrary to this hypothesis, very little impact was found regarding the transcriptional profiles in the BM. As described elsewhere[146], the researchers then correlated gene- and pathway-level changes in the transcriptional profiles during peak parasitemia with immunological and chemical profiles and found that differentially expressed genes in the BM were primarily related to ongoing

inflammatory responses that were measurable in the periphery and dominated by interferon signatures. The analysis ultimately suggested that malarial anemia is driven by monocyte-associated suppression of the transcription factors GATA1/2 in erythroid progenitors, resulting in disrupted erythropoiesis and insufficient erythropoietic output. Further study led to the creation of an atlas of tissue-specific Multi -omics Relatedness Networks (MORNs) of malaria, which exhibit differential dynamics of the host-immune response to *P. cynomolgi* infection in the peripheral blood and BM[146]. When the MORNs were integrated with immunological profiles, including cytokine profiles, and clinical traits to bridge molecular mechanisms with disease outcomes, a positive correlation of PD-1 and mTOR signaling emerged with PD-1+ central memory CD8+ T cells and PD-1+ B cells, which suggested and involvement of the combination of these signals in immune memory.

3.3.2 Direct Data Generation and Analysis

3.3.2.1 Details of Induced Malarial Infections

Details of data acquisition and raw data analysis for infections of rhesus macaques (*M. mulatta*) with *P. cynomolgi* were recently described [145]. Here, we use the same names for the macaques as in [145], namely, RFa14, RFv13, RIc14, RMe14, and RSb14, for easy comparisons. Corresponding details for infections with *P. coatneyi* will be published elsewhere. The names of the macaques in this study were RCs13, RTi13, RUn13, RWr13 and RZe13. For the following, it is sufficient to present the most pertinent highlights.

a. Parasites. Infections were introduced using *P. cynomolgi B/M strain or P. coatneyi* sporozoites and carried out for approximately 100 days. Baseline control measurements were taken before infections were introduced. Each infection was started with an inoculum of about 2,000 *P.*

cynomolgi or 100 *P. coatneyi* sporozoites that were freshly dissected from the salivary glands of *Anopheles* mosquitos bred at a laboratory of the Centers of Disease Control and Prevention (CDC). Some auxiliary data analyzed here pertained to human infections with *P. falciparum*, which shares morphological and biological features most closely with the macaque parasite *P. coatneyi*.

b. Monkeys. The rhesus macaques (*M. mulatta*) used for the infections under investigation were malaria-naïve males, which had been born and raised at the Yerkes National Primate Research Center at Emory University. The limitation to one gender eliminated potential blood loss issues associated with the female menstrual cycle, which could have confounded efforts to characterize malarial anemia. All experimental methods followed standard protocols that were approved by Emory's Institutional Animal Care and Use Committee (IACUC). The monkeys were followed throughout the infections and through relapses in the case of *P. cynomolgi*, if they occurred within the 100-day experimental period. Details for *P. cynomolgi* were described in [145], and the corresponding details for *P. coatneyi* are published elsewhere[63,147].

c. Time Point Sampling. In each experiment, the first sample was taken at time point 1 (TP1), which served as a baseline control. TP2 was chosen at peak parasitemia and TP3 immediately after this peak. TP4 – TP7 were spread out over the remaining days during the 100-day experiments; some were coarsely aligned with recurring parasitemias, others with phases in between.

3.3.2.2 Transcriptome Analysis

a. Library preparation for RNA-Seq. RNAs were extracted from whole blood using Tempus-Spin RNA isolation kits and from BM mononuclear cells using Qiagen RNEasy Mini-Plus kits, according to the manufacturer's instructions. Sufficient quality of RNA samples was confirmed using an Agilent Bioanalyzer. Approximately 1 μg of total RNA per sample was reverse-

transcribed into double-stranded cDNA. Strand-specific libraries were generated using Illumina TruSeq Stranded mRNA Sample Prep kits. For quality control, spike-in RNAs with known GC proportion and concentration (ERCC Spike-In Control, Life Technologies) [148] were added to each library to constitute 1% of total RNA.

b. Quantification of gene expression. Libraries were sequenced on the Illumina HiSeq 2000 at the Yerkes Genome Core. RNA-Seq reads were aligned to a reference genome (assembly of *M. mulatta* version 4.0, GenBank accession number PRJNA214746 ID: 214746) using Tophat2 with default parameters [149,150]. Reads mapping to multiple genomic locations were excluded from the analysis to ensure high-confidence mapping. Transcript abundance was inferred at the level of annotated genes using HTSeq v0.5.4 [151]. Data reliability was assessed by quality control: linear correlation of spike-in control abundance with known concentration; confirmation of strand-specificity of controls as 99.9%; and confirmation of the absence of 3' bias in the controls with the RSeqC software [152]. Gene expression was normalized to library size with the R package DESeq (version1.10.1; [153]), using default parameters.

c. RNA-Seq Data Processing. To minimize noise, features below a minimum FPKM (Fragments Per Kilobase of transcript per Million mapped reads) cutoff value of 32 were excluded from further analysis. Gene expression data were log2-transformed. A large variance in gene expression was observed among animals. Supervised Normalization of Microarrays (SNM) [154] was used to remove the variance associated with the animal effect. The SNM model was used with "longitudinal time points" as the biological variable, representing the course of pre-infection, acute primary infection, post-peak, inter-relapse and relapses (in the case of *P. cynomolgi*) [145], and "animal" as the adjustable variable.

d. Differential Expression Analysis (DEA). DEA analysis was performed using analysis of variance in the JMP Genomics software (SAS Institute Inc., Cary, NC). "Individual animal" was set as random effect and the "longitudinal time point" as a fixed effect. Differences in gene expression across all genes among the fixed effects were tested. Benjamini-Hochberg false discovery rate (FDR; [155]) corrections were used to adjust for multiple hypothesis testing; FDR ≤ 0.05 was used as the significance threshold.

e. Human transcriptome analysis. In order to assess whether the NHP results on changes in purine metabolism might be relevant for human malaria, we took advantage of a human malaria study [156] and downloaded the human transcriptome data of this study from the Gene Expression Omnibus (GEO accession number GSE67184). Briefly, 12 individuals in this study volunteered in a challenge trail. Blood samples were taken at two time points, namely prior to a *P. falciparum* challenge (baseline) and at the day of diagnosis (infection). Data processing and normalization were described in the original paper. Normalized expression values for selected genes were used for modeling in the same way as described for the NHP models.

3.3.2.3 Gene Set Enrichment Analysis

To identify pathway-level changes in gene expression, we performed Gene Set Enrichment Analysis (GSEA; [101,157]), which determines the statistical significance of the frequency of changes in expression for specific sets of genes within a larger set of genes. GSEA leads to a ranked list of pathways that are likely to be most affected, as judged by differentially expressed genes.

3.3.2.4 Dynamic modeling

To interpret the implications of transcriptional changes during the infection, we employed a kinetic model of purine metabolism that the Voit Lab had developed, diagnosed, and validated a number of years ago [158-161]. This model was formulated within the modeling framework of Biochemical Systems Theory[142,162-167] and has the format of a Generalized Mass Action system, where all processes are represented as products of power-law functions. It contains 16 metabolites and 37 fluxes, as well as a large number of regulatory signals. A diagram of the model structure is shown in Figure 3.1 and the equations are presented in the *Appendix B.3*. It is evident that the pathway system is tightly regulated through numerous inhibiting and activating signals. The mathematical formulation of the model was established for purine metabolism in humans, and we assume here that this implementation is applicable to NHPs as well. We therefore use the model without changes.

To achieve a coarse interpretation of the transcriptomics data in this study, we supposed that fold-changes in the expression of pertinent coding genes correspond to the same fold-changes in enzyme amounts. Though this assumption ignores issues of post-transcriptional regulation, it is expected that it may be qualitatively, and possibly quantitatively, appropriate, at least approximately. Thus, the differential expression of each gene was incorporated as a corresponding change in enzyme amount. For missing values, where a transcript had not been measured, we retained the original parameters. In cases of isozymes, the gene expression values were averaged.

In the original model formulation, the enzyme activities were lumped into apparent rate constants. Therefore, the differential expression of each gene was modeled as a corresponding change in its respective reaction rate constant parameter. All other parameters were retained at the

same values as at the original steady state. The model equations were then integrated to obtain a new steady state, and the variable concentrations and fluxes of the system were studied. This analysis was performed for every monkey and every time point during its infection.

As an example, consider the dynamics of xanthosine monophosphate (XMP; X_7), an intermediate between inosine monophosphate (IMP; X_2) and the pool of guanosine phosphates (GTP, GDP, GMP; "GXP"; X_8). The production of XMP depends on the substrate, IMP, and is inhibited by both XMP and GXP. It is catalyzed by the enzyme IMP dehydrogenase (IMPD). The dynamics of XMP is formulated in the Generalized Mass Action (GMA) model as

$$\dot{X}_7 = 1.2823\, X_2^{0.15} X_7^{-0.09} X_8^{-0.03} - 0.3738\, X_4^{0.12} X_7^{0.16} \qquad (1)$$

(see Appendix B.3 for other equations). In this type of model, all factors contributing to a process are modeled as power-law functions, in which the exponent, called the *kinetic order*, quantifies the strength of the effect of the variable on the production term. The kinetic order associated with the substrate, IMP, in the term for the production of XMP is positive, whereas the two inhibitory signals, from XMP and GXP, are negative. The product of these power-law functions is multiplied by a rate constant. The amount or activity of the catalyzing enzyme, IMPD, is not explicit in the synthesis term, but subsumed in this rate constant. The degradation term of X_7 is constructed similarly.

To account for a change in the expression of the gene coding for IMPD, we replace the value 1.2823 of the rate constant with 1.2823 × (fold-change in IMPDH1) (see Figure 3.2). This procedure was repeated for all significant changes in transcriptomics at a given time point for a given animal.

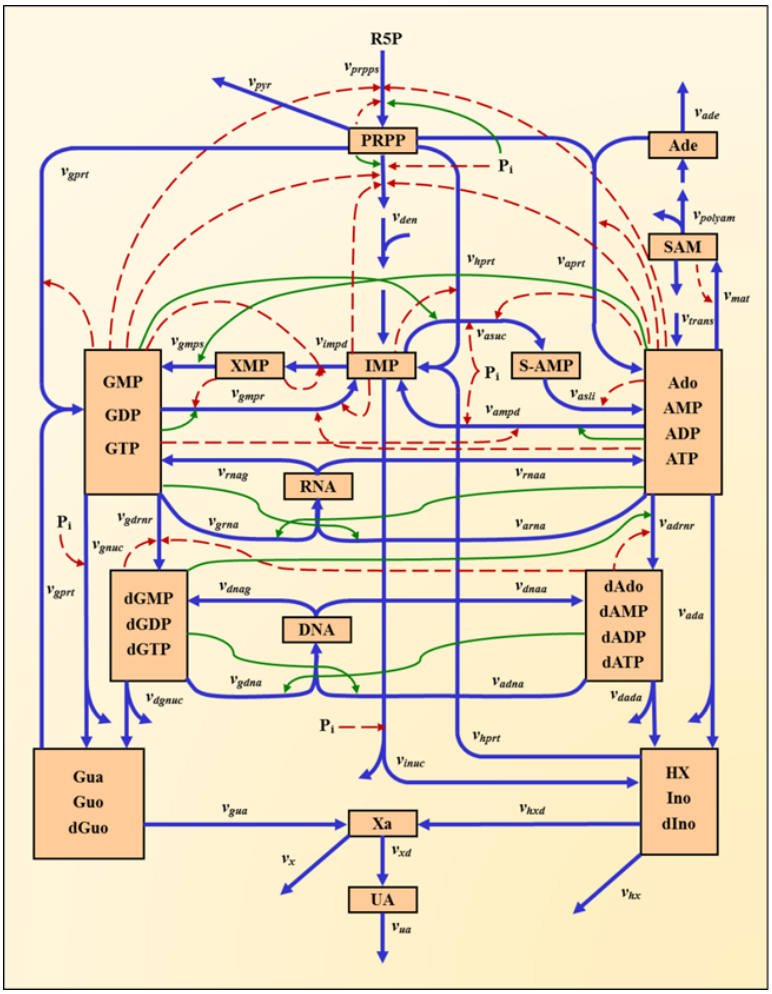


Figure 3.1- Model diagram of purine metabolism (adapted from [160]). Peach-colored boxes contain metabolites or pools of metabolites. Blue arrows indicate enzyme-catalyzed reactions or transport steps. Subscripted quantities *v* denote enzymes primarily responsible for associated steps. Solid green and dashed red arrows represent activating or inhibiting signals, respectively. Abbreviations of metabolite pools: PRPP, phosphoribosylpyrophosphate; IMP, inosine monophosphate; S-AMP, adenylosuccinate; Ado, adenosine; AMP, adenosine monophosphate; ADP, adenosine diphosphate; ATP, adenosine triphosphate; SAM, S-adenosyl-L-methionine; Ade, adenine; XMP, xanthosine monophosphate; GMP, guanosine monophosphate; GDP, guanosine diphosphate; GTP, guanosine triphosphate; dAdo, deoxyadenosine; dAMP, deoxyadenosine monophosphate; dADP, deoxyadenosine diphosphate; dATP, deoxyadenosine triphosphate; dGMP, deoxyguanosine monophosphate; dGDP, deoxyguanosine diphosphate; dGTP, deoxyguanosine triphosphate; RNA, ribonucleic acid; DNA, deoxyribonucleic acid; HX, hypoxanthine; Ino, inosine; dIno, deoxyinosine; Xa, xanthine; Gua, guanine; Guo, guanosine; dGuo, deoxyguanosine; UA, uric acid; R5P, ribose-5-phosphate. The Supplements contain a list of the reactions names and their abbreviations.

Figure 3.2- Heat map of changes in gene expression at different time points (TP2 – TP7) during *P. cynomolgi* infection. Shown here are transcriptomic changes in the bone marrow of monkey RSb14, relative to gene expression at TP1. TP3, which immediately follows the peak of infection (TP2), exhibits the strongest changes. Other expression patterns were similar for different monkeys, infected with the same parasite (*P. cynomolgi*; see Appendix Figure B.1.1, B.1.2, B.1.3) or even a different parasite (*P. coatneyi*). Results are shown on a log-10 scale.

3.4 Results

3.4.1 Interpretation of Blood Transcriptomics

Transcriptomics is generally viewed to be the most reliable source of -omics information. However, its output is not always easy to interpret in terms of phenotypical outcomes, as gene expression is two or three steps removed from metabolic, immunological, or physiological manifestations. Faced with this challenge, we employed computational modeling to interpret transcriptomic information.

Gene Set Enrichment Analysis of transcriptional data during malarial infection showed significant changes in the expression of genes associated with the purine pathway, leading us to choose this pathway as an example for interpreting transcriptomic changes with metabolic models. This pathway is furthermore of interest, as it has been implicated in a variety of inflammation studies (see Discussion).

Specifically, we considered gene expression at seven time points (TPs). Measurements at TP1, which preceded the infections, were used as "baseline control" values to which other TP measurements were normalized. TP2 coincided with the peak of infection. Interestingly, the most significant changes in purine metabolism transcriptional profiles occurred at TP3, a few days after peak parasitemia (Figure 3.2). The expression patterns more or less returned to the baseline at later time points.

Quite similar transcriptomic patterns were obtained for NHPs infected with different *Plasmodium* species (*P. cynomolgi* and *P. coatneyi*), and for data from a human cohort (see later). In the cases of rhesus macaques, the strength of the pattern correlated with the severity of

parasitemia [145]: animals with higher parasitemia at TP3 showed more pronounced patterns than the animals with lower parasitemia.

The altered expression of numerous genes coding for enzymes within purine metabolism is interesting, but difficult to interpret, because altered gene expression levels do not reveal much regarding changes in the metabolic concentration or flux profile, which in turn could lead to new functional insights or hypotheses. For example, simultaneous up-regulation of separate genes coding for the production and for the degradation of a metabolite could result in no net changes in the level of that metabolite.

Thus, our first goal was to assess to what degree transcriptomic changes alone are predictive of changes in metabolic fluxes. Such predictions are complicated due to the fact that changes in enzyme activities lead to changes in fluxes, which cause alterations in the concentrations of metabolites, which in turn, as substrates or regulators, affect the magnitudes of fluxes. Because purine metabolism constitutes a complex, highly regulated pathway system, intuition regarding the ultimate consequences of quantitative alterations in transcripts is quickly overwhelmed.

We therefore employed the kinetic model of purine metabolism described in the Section Materials and Methods and implemented significant fold-changes in the expression of pertinent coding genes as the same fold-changes in enzyme amounts (Figure 3.2). A priori, one might expect that a plot of the magnitudes of changed fluxes at the steady state would exhibit a clear positive trend. However, this trend, while it does exist, is quite weak (Figure 3.3), with a correlation coefficient of 0.1228; the difference between this value and 0 itself is significant with a correlation

coefficient of 0.00069. These findings suggest that transcriptomic data alone are insufficient to predict flux distributions in this system (see Figure 3.4).

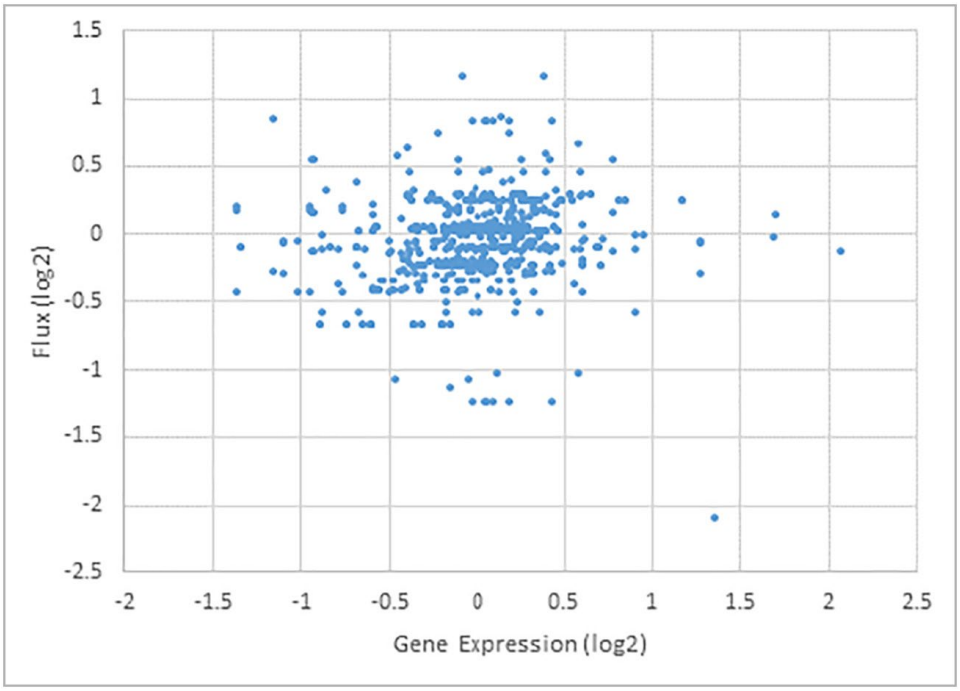


Figure 3.3- Limited predictability of fluxes from transcriptomics data alone. Fold-changes in steady-state fluxes, computed with the dynamic model, are plotted against fold-changes in gene expression in monkey RIc14 for time points TP2, …, TP7 during infection with P. cynomolgi. Both axes are presented on a log2 scale, so that 1 represents a two-fold change. One notes horizontal lines, which correspond to several genes associated with the same pathway.

3.4.2 Changes in Purine Metabolism Following Peak Infection

Faced with the observation of unreliable predictability of metabolic fluxes from transcriptional changes, we explored what the kinetic model, implemented with altered transcriptional profiles, might suggest about metabolic responses during malarial infection. While we analyzed changes in purine metabolism for all measured time points, the most substantial changes were noticed immediately following the peak of infection (TP3).

Although quite coarse, the model-based interpretation of the transcriptomics data revealed striking results. In the experiments with *P. cynomolgi* [145], two particular animals (RFa14 and RMe14) displayed severe symptoms and needed and received sub-curative treatment during the acute primary infection in order to aid in the control of the parasites and disease manifestations; one monkey (RFv13) actually had to be euthanized ultimately due to kidney failure[65].

By contrast, two other animals of the same cohort, RSb14 and RIc14, recovered without treatment, while experiencing much higher parasitemia. It is to be expected that these two untreated animals probably suffered from more severe inflammation at TP3. Particularly strong changes were observed in the fluxes from PRPP to IMP, from GMP/GDP/GTP to IMP, from IMP to HX/Ino/dIno, from HX/Ino/dIno to Xa, and the excretion of HX and Xa, which were elevated more than two-fold at TP3 compared to baseline. The concentration of the HX/Ino/dIno pool also increased more than two-fold at TP3 compared to baseline. By contrast, the two animals with lower parasitemia showed no changes in these fluxes or metabolite concentrations (data not shown).

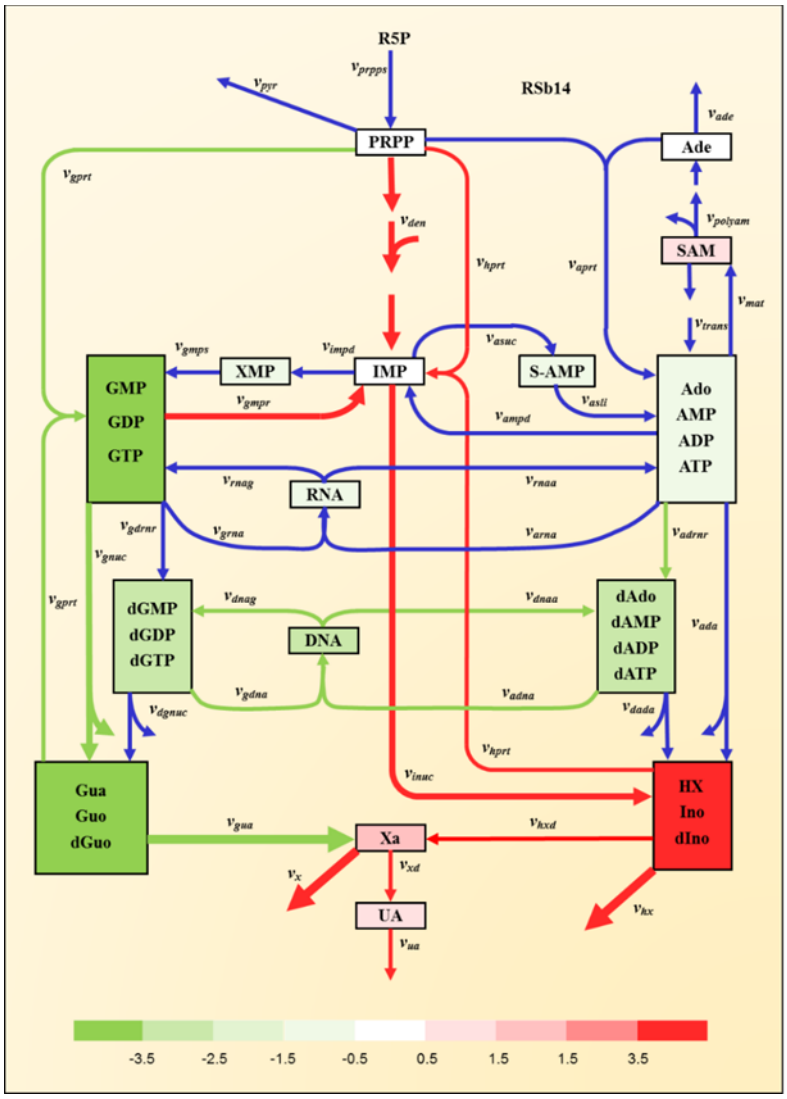


Figure 3.4- Identification of changes in fluxes and concentrations within purine metabolism immediately after peak infection (TP3) in monkey RSb14, which was infected with *P. cynomolgi*. All metabolites or metabolite pools are color-coded to show fold-increases or decreases in concentrations between post-peak and baseline: red colors represent increased concentrations at post-peak, while green colors represent decreased concentrations at post-peak, according to the log-2 color bar. White boxes indicate no significant changes. Red and green colors of fluxes represent up- or down-regulation at post-peak compared to baseline. Only fluxes with fold-changes greater than 2 are colored; others are shown in blue. The degree of change is indicated by the line thickness. For simplicity, regulatory signals are not shown (cf. Figure 3.1).

In a similar experiment with *P. coatneyi* infections[63,147], the macaques RWr13 and RTi13 exhibited higher parasitemia at TP3 and also showed much more pronounced changes in fluxes in

the purine model than two other animals with lower parasitemias (Appendix Figures B.1.4, B.1.5, B.1.6). Intriguingly, the same purine fluxes were increased as in the *P. cynomolgi* infection: from PRPP to IMP, from GMP/GDP/GTP to IMP, from IMP to HX/Ino/dIno, from HX/Ino/dIno to Xa, and the excretion of Xa and HX. These fluxes at TP3 were elevated more than two-fold compared to the baseline value before the infection (TP1). Also as in the *P. cynomolgi* infection, the concentration of HX/Ino/dIno increased more than two-fold at TP3 compared to baseline.

Taken together, the profile analysis of purine fluxes and metabolites computed with the model, based on transcriptomics data, reveals a distinct signature in the affected purine pathway. In the *P. coatneyi* infection experiment at TP3, clustering of the profiles of the 37 simulated fluxes in the model grouped the animals with high and low parasitemias separately (Figure 3.4A). Similar results were seen in clustering profiles of simulated concentrations of the 16 metabolites (Figure 3.4B). The *P. cynomolgi* infection experiments yielded the same results: both TP3 profiles of model fluxes (Figure 3.4C) and metabolite concentrations (Figure 3.4D) clustered the high-parasitemia and low-parasitemia animals separately.

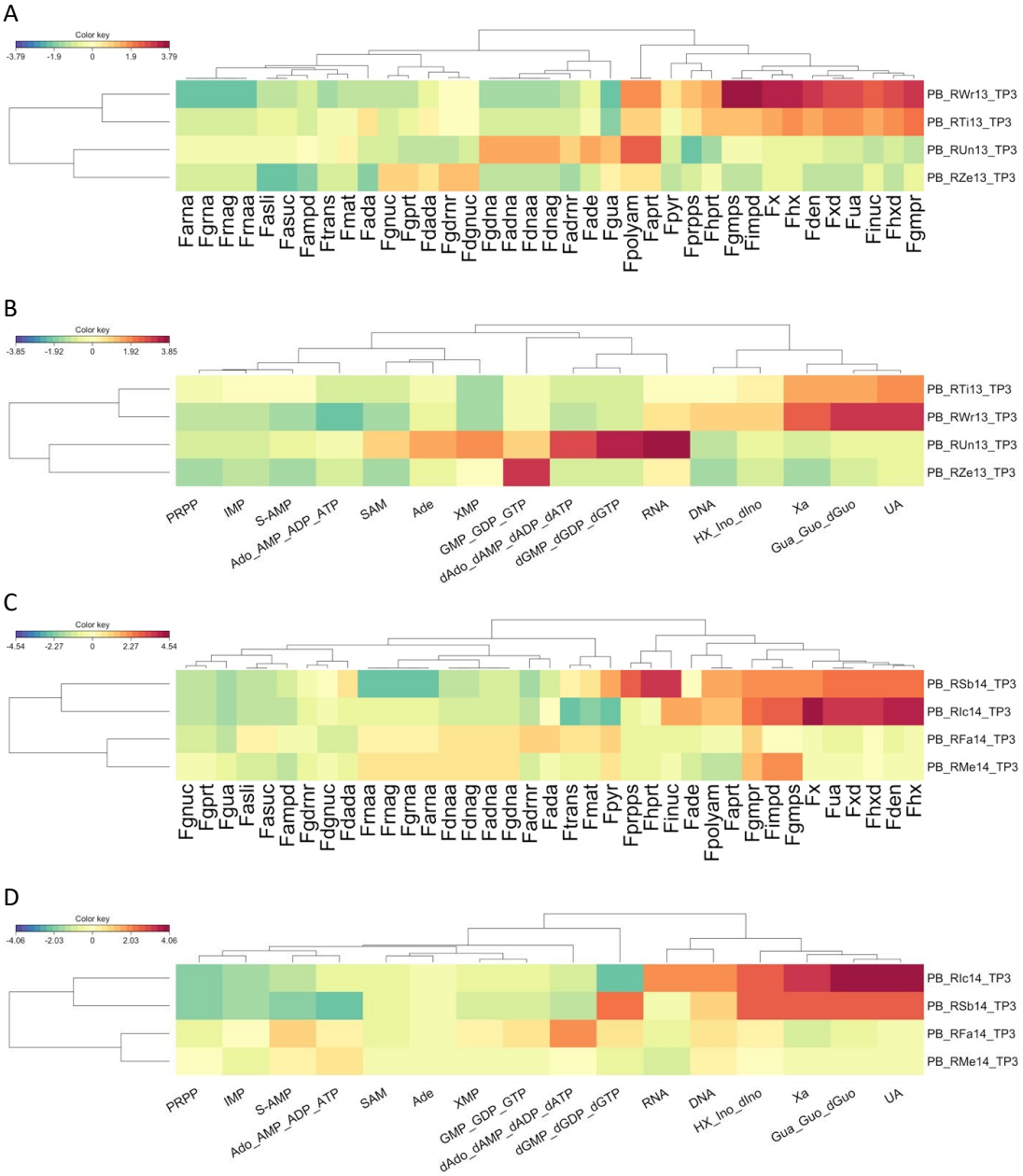


Figure 3.5- Heat maps of simulated fluxes and metabolites immediately following the peak of infection. The results show that animals with higher parasitemia are clustered together, separately from two animals with lower parasitemia. This finding is true in both *P. coatneyi* and *P. cynomolgi* infections. (A) Heat map of model-inferred fluxes in *P. coatneyi* infections. (B) Heat map of model-inferred metabolites in *P. coatneyi* infections. (C) Heat map of model-inferred fluxes in *P. cynomolgi* infections. (D) Heat map of model-inferred metabolites in *P. cynomolgi* infections.

3.4.3 Changes in Human Purine Metabolism during Malaria

To assess whether these results on macaques had any relevance for human malaria, we analyzed transcriptomics data from a human study [156], in which volunteers were enrolled for a sporozoite challenge trial with *P. vivax*, which is closely related to the macaque parasite *P. cynomolgi*. RNASeq analysis was performed on 12 individuals. Blood samples were taken before the challenge (baseline) and at the day of diagnosis, which here is considered as a time point during infection. Normalized expression values of pertinent genes were used for simulation. Quite strikingly, we observed similar pattern of purine metabolism in some of individuals as in the severe animals (Figure 3.5). In particular, the same fluxes we observed in macaques exhibited significant alterations in humans, namely, the fluxes from GMP/GDP/GTP to IMP, from IMP to HX/Ino/dIno, from HX/Ino/dIno to Xa, as well as the excretion of HX and Xa. These results suggest that molecular mechanisms perturbing purine metabolism during malaria are conserved from NHPs to humans.

3.4.4 Changes in Purine Metabolism during Chronic Infection

As can also be the case with *P. cynomolgi*, *P. coatneyi* infections can become chronic, thereby causing persistent inflammation. Model simulations across all time points during primary and chronic infection with *P. coatneyi* showed purine-related responses in monkeys with higher parasitemia (RTi13 and RWr13) that were different from those with low parasitemia. Specifically, we found prolonged increases in fluxes from IMP to HX/Ino/dIno and the excretion of HX/Ino/dIno and Xa during the entire time from acute primary infection to recrudescence patency. These two animals also showed prolonged increases in the concentration of HX/Ino/dIno across

all seven time points (Appendix Figures B.1.7, B.1.8). These changes were not observed for acute infections with *P. cynomolgi*. These findings are interesting, because HX and Ino have been shown to be associated with inflammation [168,169].

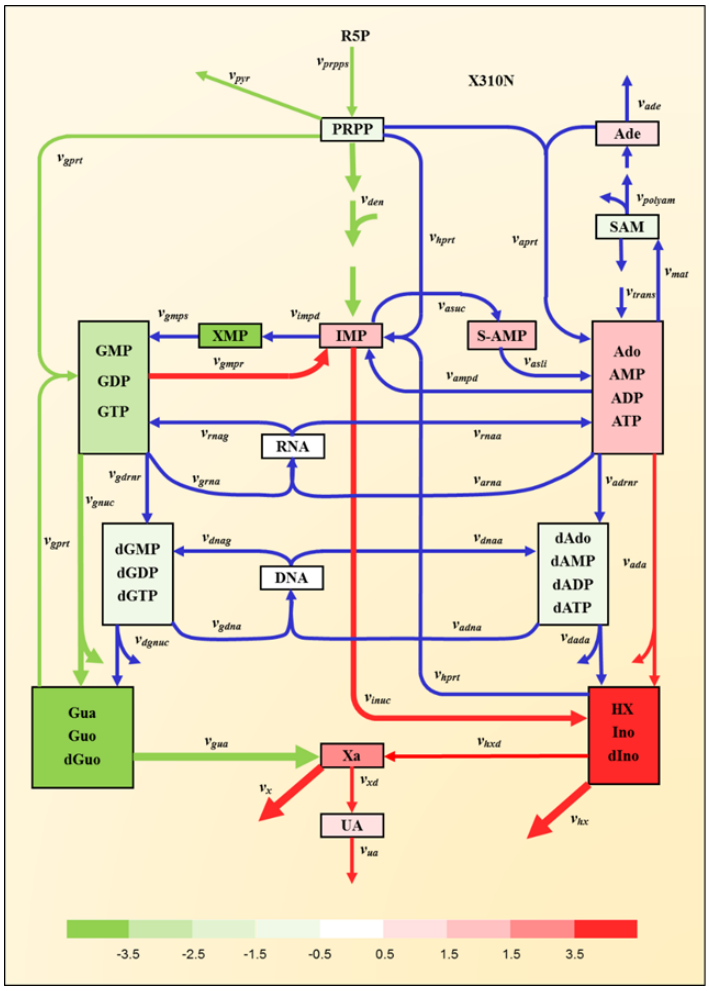


Figure 3.6- Changes in fluxes and concentrations within purine metabolism before and during infection in a human volunteer challenge trial with *P. vivax*. All metabolites or metabolite pools[156] are color-coded to show fold-increases or decreases in concentrations between infection and baseline: red colors represent increased concentrations at post-peak, while green colors represent decreased concentrations at post-peak, according to the log-2 color bar. White boxes indicate no significant changes. Red and green colors of fluxes represent up- or down-regulation at post-peak compared to baseline. Only fluxes with fold-changes greater than 2 are colored; others are shown in blue. The degree of change is indicated by the line thickness. For simplicity, regulatory signals are not shown (Figure 3.1).

3.5 Discussion

Disease reflects the malfunctioning of a very complex system. To combat disease effectively, we need to develop a deep understanding of the molecular signatures and events that the malfunctioning system exhibits. Especially within the setting of personalized disease and medicine, these signatures and their trends must be characterized in individualized detail, so that population-averaged information may be substituted with patient-specific parameters [170,171]. In the past, experimental and clinical limitations allowed the quantification of only very restricted subsets of relevant biomarker signatures. The -omics revolution has fundamentally changed this situation, because single experiments can easily yield thousands of data points, each of which could potentially be a biomarker. The new challenge accompanying these experiments is that it is often difficult to interpret the phenotypical ramifications of changes in patterns of molecular biomarkers. Here, we take as a pertinent example high-throughput transcriptional data, widely viewed to be comprehensive and useful. However, clinical manifestations in terms of metabolic or physiological aberrations, which may have their root causes in transcriptional changes, are several steps removed from the expression of genes so that the specific consequences of altered gene expression are difficult to intuit. Here, we address this issue by proposing an interpretation of genomics information through a dynamic model that allows us to translate changes at the transcript level into alterations in metabolic signatures.

We used as an example for our demonstration the responses of hosts to infections with malaria parasites. More precisely, we focused on purine metabolism, which was identified by gene set enrichment analysis as significantly affected during infection. While the expression of many genes associated with this pathway system was indeed changed to some degree, the functional impacts of this expression profile could not be interpreted easily. By contrast, using the

transcriptome changes as corresponding changes in enzyme amounts in a dynamic model of purine metabolism and integrating this model to the steady state revealed a pattern of flux rearrangement within the pathway system. This pattern was consistent among different monkeys and even among infections with different *Plasmodium* parasites and in part correlated with parasitemia levels. The main results of this flux rearrangement were identified as an increased production and excretion of inosine (IN), hypoxanthine (HX) and xanthine (XA), as well as less dramatic changes in a few other compounds.

The fact that the same pattern was seen in many of the infected monkeys, and was somewhat corroborated by human malaria data, directly leads to new hypotheses regarding the synthesis and utilization of specific purine compounds. This type of hypothesis generation is valuable, but it is clear that the mathematical model alone is not in a position to provide a possible rationale for such changes. Nonetheless, the literature documents a number of observations regarding the roles of purines in malaria, and these may or may not turn out be relevant and/or explanatory. Two classes of observations with particular pertinence are the following.

3.5.1 Purine compounds are often associated with inflammation

It has been known for a while that the end product of the pathway, uric acid, is known to stimulate immune activation, even in the absence of bacteria or other stimuli [168,169]. At the same time, IN inhibits inflammatory cytokine production [172] and, upon metabolic stress, cells release IN into the extracellular space, where elevated inosine levels are present in various inflammatory states. Along the same lines, HX and XA are increased in synovial fluid during inflammatory arthritis [173]. It has also been shown that adenosine and IN have anti-inflammatory effects [174-176].

More generically, a recent review described the role of purinergic signaling in the immune system[177].

Specifically with respect to malaria, hypoxanthine accumulates in infected RBCs [178]. When these release merozoite progeny, the parasites precipitate uric acid, which is released into the blood stream [179]. HX is also released into circulation, where it may be converted into uric acid. Uric acid precipitates are considered key inflammation signals in malaria [178]. They are highly inflammatory molecules and serve as danger signals for the innate immune system [168]. In particular, uric acid precipitates move into micro-vessels, where they stimulate immune cells to produce IL-6, IL-8, IL-10, TNFα, sTNFRII, MCP-1, and IP-10 [180]. These inflammatory cytokines are considered important components of the host's inflammatory reaction to *Plasmodium* infection and a major cause of malaria pathogenesis.

3.5.2 The Plasmodium parasite needs large amounts of purines for proliferation

Plasmodium does not have the metabolic machinery to synthesize purines *de novo*[181]. However, it propagates very quickly within RBCs and thus requires rapid large-scale synthesis of DNA, RNA and of ubiquitous factors like ATP and GTP. In particular, HX is essential for *Plasmodium* growth and commonly used as a required reagent in parasite cultures; its concentration in infected RBCs is much higher than in uninfected ones [178,179]. An article by Downie *et al.* [181] summarized alternative mechanisms in *Plasmodium* for providing purines. Most importantly, the parasite relies on salvaging purine compounds, which are mainly funneled through HX to IMP. HX is transported from plasma into the cytoplasm and can also be produced from adenosine or IN. To augment the typical salvage observed in mammals, which is facilitated by the enzymes HGXPRT and APRT, *P. falciparum* has an additional enzyme, PfHGXPRT, that utilizes

XA instead of HX or adenine, respectively. As an aside, enzymes involved in purine metabolism have been proposed as potential drug targets [182].

Our results and these documented observations are intriguing and puzzling at the same time. The consistency among species and the correlation with the degree of parasitemia suggest that the changes are not merely spurious occurrences. Also, at first glance, the results seem to suggest that more hypoxanthine and inosine are produced because the parasites require them. However, such a casual interpretation of the results requires caution and careful consideration. After all, the parasites invade RBCs, which do not have nuclei and therefore cannot respond with transcriptional changes as they are observed in the WBCs. This difference in involvement in the host response raises a complicated question regarding causality of events: Do the parasites deplete purine compounds like ATP in RBCs, which then signal the need for purines, to which WBCs respond with transcriptomic and metabolic alterations? Of course, the model cannot help with a biological interpretation of such speculations, the computational inferences, or the data themselves. In Chapter 4 we use this methodology to understand and interpret differences in tryptophan metabolism. The results thus obtained were corroborated by the metabolomics measurements. It is remarkable that a metabolic model analysis of the type shown here can lead to new hypotheses regarding the targeted rerouting of fluxes, and regarding a rationale for changes in transcriptomics. It is now up to wet lab experimentalists to test these hypotheses.

CHAPTER 4. MOLECULAR MECHANISMS THAT EXPLAIN KEY DIFFERENCES IN PROGRESSION OF *PLASMODIUM KNOWLESI* INFECTION[3]

4.1 Summary

Malaria has a complex pathology with varying manifestations and symptoms, effects on host tissues, and different degrees of severity and ultimate outcome, depending on the causative *Plasmodium* pathogen and host species. In Chapter 2, we compared the peripheral blood transcriptomes of two macaque species (*M. mulatta*; abbreviated as Mm, and *M. fascicularis*; Mf) in response to acute primary infection by *Plasmodium knowlesi*. Although these two species are very closely related, the infection in *M. mulatta* is fatal, unless aggressively treated, whereas *M. fascicularis* develops a chronic, but tolerable infection in the blood. As a reason for this stark difference, our analysis in Chapter 2 suggests delayed pathogen detection in *M. mulatta* followed by extended inflammation that eventually overwhelms this monkey's immune response. By contrast, *M. fascicularis* detects the pathogen earlier and controls the inflammation. Additionally, *M. fascicularis* limits cell proliferation pathways until the peak of an infection, presumably in an attempt to reinforce recovery.

In this chapter, we focus on molecular mechanisms underlying the key differences in the host and parasite responses and their coordination. We begin with a correlation analysis between host and pathogen transcripts that reveals a similar pattern-recognition receptor (PRR) signaling

[3] Gupta, A. Galinski, MR. Voit, EO. Dynamic control balancing cell proliferation and inflammation is crucial for an effective immune response to malaria. Frontiers in Molecular Biosciences (submitted)

mechanism in the two hosts, which detects pathogenic antigens, including among others the Schizont-Infected Cell Agglutination (SICA) variant proteins that are expressed from the *SICAvar* gene family. In both hosts, PRR signaling initiates a cytokine response, but there are important differences. The first is a limited immune response in *M. mulatta*, due to interferon beta (IFNβ)-mediated inflammation, which is induced by expression of IL10 and IL6-STAT3. Second, transcriptomic evidence suggests B cell activation and reactive oxygen species (ROS) as initiators of this inflammation. Differently in *M. fascicularis*, the key feature seems to be a humoral immune response that is activated by T helper cells.

During the initial parasitemic log phase of the infection, both hosts experience elevated inflammation which, however, is reduced in *M. fascicularis* as it nears the peak of infection. One biomarker for inflammation is the upregulation of the enzyme indoleamine 2,3-dioxygenase (IDO), which causes the redistribution of fluxes within tryptophan (Trp) metabolism. IDO furthermore plays a central role in regulating inflammation, immune tolerance and cell proliferation. A second upregulated enzyme of the Trp pathway, kynurenine 3-monooxygenase (KMO), emerged as a central regulator for balancing NAD-mediated cell proliferation and immune activity via aryl hydrocarbon receptor (AhR) signaling. The difference in KMO activity in the two hosts is clearly associated with higher levels of NADPH and cell proliferation in *M. mulatta*.

Downstream analysis of AhR signaling adds a further layer of regulation, due to its competition with hypoxia inducible factor 1-alpha (HIF1α). Owing to their similar molecular structures, both AhR and HIF1α form complexes with the aryl hydrocarbon receptor nuclear translocator (ARNT) protein, and both these complexes regulate multiple common genes. These

downstream genes include immune related genes like IL6 and IFNγ, as well as FOS genes that are involved in cell proliferation.

In addition to controlling cell proliferation through Trp metabolism, *M. fascicularis* regulates cell proliferation through various mechanisms up to the time of peak parasitemia. For instance, it downregulates ribosomal biosynthesis, which in turn enhances the proliferation suppressor p53 pathway. This mechanism is clearly reflected in the downregulation of all ribosomal proteins in *M. fascicularis* during the log phase of parasite growth and expansion. Later, near peak infection, this regulation is diminished, which presumably aids the adaptive immune response.

A complete understanding of the exact dynamics of the immune response is difficult to reach. Nonetheless, this comparative analysis provides clear suggestions of processes that underlie an effective immune response. Thus, this study identifies multiple points of intervention that are apparently responsible for a balanced and effective immune response and thereby paves the way toward future immune strategies for treating malaria.

4.2 Introduction

In Chapter 2 we analyzed the gene programs with which Mm and Mf respond to a *P. knowlesi* infection, initiated with infectious sporozoites. This comparative analysis revealed numerous transcriptomic similarities, but also notable differences (Figure 4.1). In particular, Mf, but not Mm, apparently detects this pathogen as early as the liver phase of the infection, prior to the parasite infecting the blood, and this correspondingly activates beneficial signaling pathways early on. Later in the infection, significant differences arise in each monkey's immune responses, which in Mm lead to extended inflammatory activities and prolonged inflammation. By contrast, Mf

contains the infection and controls inflammation by undergoing a transcriptional makeover toward cell proliferation that accompanies its recovery.

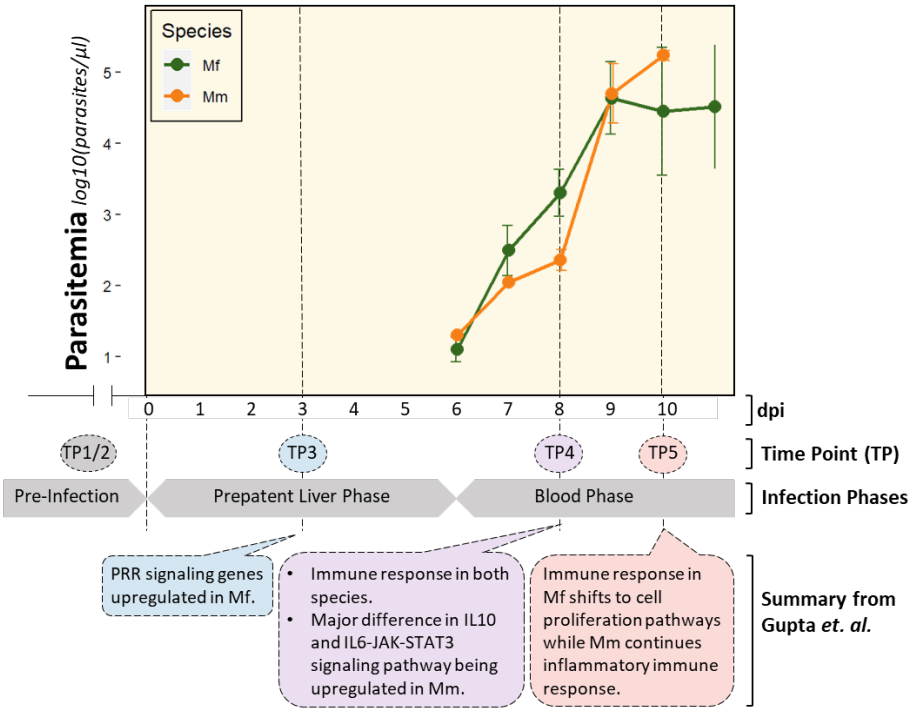


Figure 4.1- Timeline and progression of *P. knowlesi* infection in Mm and Mf. The figure shows the levels of parasitemia and corresponding time-points and days post infection on x-axis. The x-axis also contains various infection phases and key observations from Chapter 2.

In this chapter, we shed light on some of the molecular mechanisms governing the different gene programs and thus the ultimate fates of the two macaque species. In particular, the study identifies and quantifies: differences in the detection of the pathogen, associated differences in the immune response, differences in cell proliferation that directly affect the immune response and indirectly inflammation and, finally, differences in pathways that regulate inflammation.

The detection of malarial parasites by the host immune system is driven by parasite-encoded surface proteins, including among others the Schizont-Infected Cell Agglutination (SICA) variant proteins[183] that are expressed from the *SICAvar* gene family[184-187]. The antigenicity and variability

provided by these various proteins stimulates the production of antibody repertoires and immunogenicity that have been widely studied in the context of vaccine development[188-191]. The *Plasmodium* pathogen multiplies within infected red blood cells (iRBCs) and once matured these cells burst releasing new merozoite progeny that infect other RBCs. This process generates pathogen- and damage-associated molecular patterns (PAMPs and DAMPs), which in turn stimulate various pattern recognition receptor (PRR) signaling pathways in macrophages, monocytes, neutrophils and dendritic cells, and execute various immune mechanisms via protein kinase cascades[192,193]. Co-expression analysis has been shown to be instrumental in determining these host-pathogen interactions[194]. The neutrophils and macrophages not only target foreign content for phagocytosis but also trigger the inflammatory and adaptive immune response. Our previous analysis showed much stronger inflammation in Mm compared to Mf, which launches extensive measures to control cell proliferation. The balance between these pro- and anti-inflammatory mechanisms appears to be the key to resilience, and a deeper understanding of the underlying mechanisms is therefore of utmost importance[195]. The energy-intensive nature of these processes makes metabolic processes like glycolysis and Trp metabolism close accomplices in regulating the overall physiological dynamics. We apply the methodology developed in Chapter 3 to better understand the variations in Trp metabolism. Furthermore, the feedback loop of Trp metabolism and AhR signaling in controlling inflammatory cytokines is essential for this balance.

The complete dynamics of the entire immune response is obviously difficult to comprehend in full detail, as this response is systemic and involves uncounted facets, some evident but others subtle. Thus, while our comparative analysis clearly cannot convey a complete picture of all chains of causes and effects governing the responses by the two macaque species, it offers a first glimpse into some of the same and some of the differentiating processes evoked by the two monkey species.

The study thereby opens a new avenue toward potential future strategies of immune-based malaria treatments and provides multiple promising candidates for interventions targeting a balanced and effective immune response.

4.3 Methods

4.3.1 Experimental setup and data pre-processing (RNA seq / LC-MS)

The analysis described here expands on Chapter 2 and previously published studies[53] with details about individual processes. In a nutshell, four male Mm and seven male Mf were infected with *P. knowlesi* sporozoites. Several samples were extracted before (baseline) and after inoculation with sporozoites (pre-patent – TP3, log-phase – TP4 and peak-phase – TP5). These blood samples were used for transcriptomics and metabolomics analysis (Figure 4.1).

To assess the transcriptome, samples were sequenced using Illumina Hi-seq 3000, mapped using STAR and normalized using DESeq2. Details of the process are described in Chapter 2.

For metabolomics analyses, the plasma samples were quantified using the AbsoluteIDQ p180 kit (Biocrates Life Sciences AG). Specifically, the metabolites were quantified using SCIEX Exion LC and a QTRAP 5500 mass spectrometer in only positive ionization mode with each sample injected using a separation column. Specific details of the process can be found with the corresponding submissions of MaHPIC data to PlasmoDB (https://plasmodb.org/plasmo/app/static-content/PlasmoDB/mahpic.html) with the Mm dataset available at MTBLS824 and the Mf dataset at MTBLS822 from the MetaboLights repository.

4.3.2 Enrichment Analyses

Differential expression (DE) of genes was calculated using DESeq2. Genes with low read counts were removed from analysis. The genes were modeled using the design – *Species* + *TimePoint* + *Species:TimePoint* and DE was calculated using Wald's test.

Gene set enrichment analysis was performed using the GSEA toolkit (version 4.0) of the Broad Institute. The gene sets used for the analysis were Hallmark[196], Reactome[197], ImmuneSigDB[198] and Gene Ontology[199,200]. The pre-ranked GSEA module of the toolkit[101] was used for the analysis, and all genes were ranked based on inverse of adjusted *p*-values and the sign of fold changes. Files of custom gene sets (gmt files) were created using R to contrast enrichment scores between comparable data sets. To compare gene sets across the two species and account for representation bias in individual gene sets, rank scores for all genes were used to calculate enrichment scores (ES), which were adjusted by normalization of gene set sizes. Gene sets with small (<15) and large (>500) overlaps were filtered out. This normalized enrichment score (NES) was used to contrast various gene sets.

Enrichment analysis for targets of AhR, AHRR and HIF1A was performed similarly to the method described above. The gene sets for target genes for each were created using ChIP-Atlas[201] with ±5 Kb overlap with the transcription start site. NES values for each subset described in the *Results* were calculated with the method described above.

4.3.3 Weighted gene co-expression network analysis (WGCNA)

Weighted gene co-expression network analysis (WGCNA) was performed using the WGCNA package (version 1.70-3) [202,203] in R to describe correlation patterns among genes. The

analysis was performed in multiple ways to serve different purposes. The differences arose in the subsets of samples in datasets used for each analysis. First, for co-expression networks with both host and pathogen genes, only infection TPs (TP4 and TP5) were used for both hosts, as there are no pathogen transcripts at baseline and TP3. Next, to differentiate host-specific differences, subsets of each host for different infection TPs were used. Finally, all TPs for both hosts were used with host-only genes to form co-expression networks among host genes.

WGCNA analysis begins with creation of a Pearson correlation matrix of the expression of all gene pairs. These were used to filter highest correlated pairs where required. This step was followed by the creation of an approximately scale-free adjacency matrix, using a power function. The soft threshold parameter (B) for the power function in each case was determined based on the criterion of approximate scale-free topology, as described in the software manual[203]. The topological overlap matrix (TOM) was calculated to quantify the degree of overlap in shared neighbors. Finally, modules were created using a dynamic tree cut algorithm in WGCNA. To characterize each of the modules, module eigengenes and GO annotations were calculated. To calculate the similarities between various modules, Pearson correlation between eigengene vectors was used.

4.3.4 Deconvolution of cell populations

Cibersortx[204] was used to analyze gene expression data to obtain an estimation of abundances of individual cell types from mixed cell populations in the various blood samples. The LM22 signature matrix[205] was used as a cell type reference profile. Previously DESeq2 normalized expression data for all samples were used to estimate the abundances of the 22 cell types from whole blood.

To contrast various groups, the lmFit function (limma package) in R was used to model the cell populations as *Species* + *TimePoint* + *Species:TimePoint* and the eBayes function was used to compute log fold changes, *t* statistics, *p*-values and adjusted *p*-values, using the Benjamini-Hochberg method.

4.3.5 Dynamic modeling of tryptophan metabolism

To understand the implications of transcriptomic changes during *P. knowlesi* infection, we used a well-established tryptophan metabolic model[206] and adjusted its parameters to represent changes in enzymatic activities in accordance to changes in the expression of corresponding genes. This methodology is described in detail in Chapter 3 and Tang *et. al.*[207].

The model was originally developed for liver tissue and had to be adapted for blood. Due to the lack of tissue specific enzyme concentration data, we used gene expression data for individual tissues (in this case blood *vs.* liver data from the GTEx project[208]) to form a crude estimate of enzymatic concentration. Each reaction rate v in the model is described with the Michaelis-Menten rate function

$$v = \frac{V_{\max} . S}{K_m + S}, \qquad \mathbf{1}$$

where $V_{\max}$ is the maximum reaction rate, S is the substrate concentration and K_m is the Michaelis constant. According to our assumption of proportionality between gene expression and enzyme activity[209], $V_{\max}$ is a function of enzyme concentration and enzymatic turnover K_{cat}. Since enzyme concentration is difficult to calculate, mRNA levels were used as approximate quantities:

$$V_{\max} = F.mRNA.K_{cat}. \quad \textbf{2}$$

Here, F is a factor that converts expression values into enzyme concentrations and $mRNA$ is the measured expression.

Once the parameters were updated, the model was simulated to a steady state to obtain baseline metabolite concentrations and fluxes for the blood model.

Next, the kinetic parameters were updated by a factor corresponding to the fold change in gene expression in order to obtain the appropriate enzymatic activity, similar to Eq. 1 and Eq. 2. For each case, the model was simulated to the steady state of all metabolite concentrations and fluxes were used for comparison of different scenarios.

For flux control analysis[210] (Eq. 3), the control coefficients were calculated as

$$C_{v_i}^{S} = \frac{d \ln J}{d \ln v_i}, \quad \textbf{3}$$

where $C_{v_i}^{S}$ is the flux control coefficient for the pathway flux J with small changes in enzyme activity v_i of step i.

4.4 Results

The analysis described in this chapter directly follows from results obtained in Chapter 2 (Figure 4.1) with details published in Peterson *et. al.*[211] . In this longitudinal study of *P. knowlesi* infections in Mm and Mf, peripheral blood and bone marrow samples were collected at various time points (TPs), including baseline (before infection), pre-patent (TP3 or three days post inoculation; dpi), log-phase (TP4 or 8 dpi) and peak-phase (TP5 or 10 dpi). The first signs of

parasitemia were observed 6 dpi, and the infection increased exponentially thereafter. The Mm subjects were euthanized by 10 dpi, at the time parasitemias were escalating to dangerous levels, to carry out necropsies and characterize the infected tissues. We previously observed that Mf shows very early signs of parasite detection by 3 dpi[53]. Even though the immune response was found to be similar between the hosts during the log-phase of the blood infection, Mf was found to switch its response by the peak phase towards cell proliferation, which we concluded is a sign of recovery. In this chapter we address these and other findings to shed additional light on the molecular mechanisms governing these processes.

4.4.1 Correlated NHP host and P. knowlesi transcripts suggest common signaling mechanisms and the expression of key pathogenic proteins, including SICA antigens.

It is to be expected that a mammalian host senses the presence of a parasite based on the detection of pathogenic macromolecules or signals from infected erythrocytes, which trigger signaling pathways in the host that in turn control the gene programs governing a systemic immune response (Figure 4.2). In this section, we analyze the sensing-signaling process by means of co-expression networks, functional annotation, and logistic regression analysis.

4.4.1.1 Co-expression networks of host and parasite genes

Genes with similar functionality often have correlated expression profiles, which may be identified using co-expression network analysis[212]. We adapted this approach by combining both host and pathogen transcripts in a weighted correlation network analysis (WGCNA)[203] in order to identify modules of host and pathogen genes that act in synchrony. We refer to these modules based on their "hub genes." Specifically, the analysis resulted in three types of modules: (A) Host modules consisting exclusively of host genes; (B) Host majority modules with both host and

parasite genes, but with a majority of host genes; and (C) *P. knowlesi* majority modules with both host and parasite genes, but with a majority of *P. knowlesi* genes (Appendix Figure C.1.1).

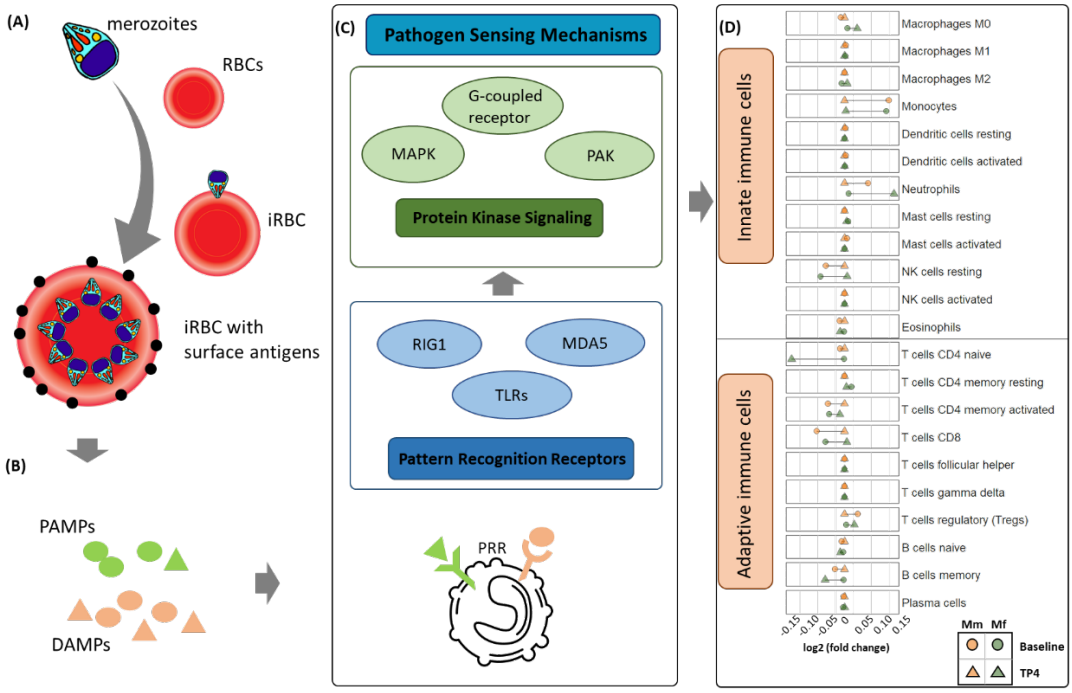


Figure 4.2- Chain of events during the blood phase of *P. knowlesi* infection. (A) Pathogen and RBCs: Once released from the liver into the blood stream, merozoites invade uninfected RBCs leading to Infected RBCs (iRBC) with exposed surface antigens (schizont-infected cell agglutination antigens – SICA; black). The iRBCs are partially eliminated by macrophages, which leads to the production of (B) pathogen/danger-associated molecular patterns (PAMPs/DAMPs). These PAMPs/DAMPs are sensed by other immune cells through (C) Pathogen Sensing Mechanisms (PRR signaling), which activate various protein kinase signaling pathways. These signaling pathways are responsible for (D) immune responses that are mediated through various leukocytes.

Genes involved with essential functions form well-defined modules, namely SNRPD2 for ribosomal assembly, TOP1 for the immune response, and RACGAP1 for mitotic cell cycle functions (Appendix Table C.2.1). It is not surprising that most of the differentially expressed genes (DEGs) during the parasitemic log phase belong to the TOP1 module, followed by C1D and ATAD3A modules (with insignificant functional annotation) for both hosts. It is worth noting that the NF2 (tRNA metabolic process), SNRPD2 and RACGAP1 modules are most highly correlated

with the TOP1 module (Pearson correlation between eigengenes with $p < 0.01$, corrected for false discovery rate (FDR)), suggesting close orchestration between these essential functions. The interactions of host and pathogen genes are most evident through interactions between Type C modules and Type A or Type B modules. Significantly high correlations between host and *P. knowlesi* genes (Appendix Table C.2.2) are found in modules ATAD3A (Type B) and PKNOH_S08507800 (Type C). Interestingly, 23 out of the 26 highly correlated *P. knowlesi* transcripts belong to Schizont-Infected Cell Agglutination variant antigen (*SICAvar*) Type 1 genes[185-187] (Appendix Table C.2.3). The corresponding SICA variant antigens, which are expressed on the surface of infected erythrocytes and associated with virulence[213,214], show high correlations with several important host genes, including IL10, ELK4 and HSPA6. This suggests that *SICAvar* Type 1 transcripts play a role in regulating inflammation in the host, for example, directly through IL10 expression and indirectly by regulating stress signals through HSPA6 expression.

4.4.1.2 Parasite gene expression affecting host co-expression modules

In order to create well defined modules for the hosts that capture correlational trends across all TPs, we used WGCNA with all host samples (including Baseline and TP3), while excluding *P. knowlesi* genes (Appendix Table C.2.4). Logistic regression followed by functional enrichment identified key modules changing during infection in both host species. The defense response module FBXO6, and modules GFRA2 and RASGEF1A (with insignificant functional annotation), were the most different during the log phase. Modules that were different included RPS19 (SRP-dependent co-translational protein targeting to membrane) and NR1H3 (cell activation involved in immune response). Integrating module membership data with host-pathogen transcripts correlation data (Sec 4.4.1.1) highlighted *Plasmodium* proteins that affected each module. Most noticeable are

SICAvar Type 1[186,215], Trp-rich antigen[216] and KIR-like proteins[186] affecting modules RPS19 and NR1H3, which differentiated the two hosts. Additionally, high correlation of hemoglobin complex module EPB42 with several pathogenic ribosomal proteins (Appendix Table C.2.4) suggests a possible mechanism for digesting hemoglobin as an essential nutrition source for the pathogen[217].

4.4.1.3 Host-specific integration with parasite gene expression

Constructing co-expression networks from expression profiles of both hosts maximizes the power to detect co-regulated host and pathogen genes that are consistent with both hosts, but the approach is not likely to identify genes that differentiate the hosts. To identify *P. knowlesi* genes that might be important for individual hosts, co-expression networks were formed using individual host data (Appendix Tables C.2.5-C.2.10).

In Mm, it is not surprising that immune response and energy production related modules (RAB11A, CCL1 and COBRA1; Appendix Table C.2.5) account for the most DEGs during the log phase. These modules are highly correlated with pathogen proteins that are crucial for *Plasmodium* development, such as the AP2 family[218] and Kinesin-5[219], along with protein trafficking proteins, such as v-SNARE Vti1p and vacuolar protein sorting-associated protein 18. The most significantly correlated pairs of host and pathogen genes (Appendix Table C.2.6, C.2.7) include host genes IFNG, CXCL1 and CXCL6 and pathogen genes of *SICAvar* Type 1, as well as genes coding for calcium or potassium channel proteins.

In Mf, by contrast, the formed modules did not have significant functional annotation (Appendix Table C.2.8). Most of the DEGs during the log phase were part of modules TMEM164 and GIMAP7. Both these modules consist of genes for multiple important functions, including immune response and metabolism, suggesting a close orchestration of these modules. *Plasmodium*

proteins that are correlated include *SICAvar* Type 1 and FAD synthetase. Interestingly, the host genes most correlated with *P. knowlesi* genes (Appendix Table C.2.9, C.2.10) are responsible for wound healing and coagulation. These platelet-related host proteins (PF4, GP1BB, GP5 and GP9) are highly correlated with proteins containing a *P. knowlesi* heme/steroid binding domain. Platelet-related wound healing proteins and their correlated *P. knowlesi* transcripts might be associated with increased resilience as well.

4.4.1.4 PRR signaling

The co-expression network analysis is able to detect important host-pathogen relationships that appear to be crucial for the two hosts. In order to focus on detection of the pathogen by a host, we modified the analysis to create a customized module of PRR signaling related genes, which allowed us to identify *Plasmodium* proteins that interact with their products. The most positively correlated pathogen proteins include KIR-like protein[186] and thioredoxin-like protein[220,221]. Among the genes that negatively co-express with the PRR module are *SICAvar* Type II, AP-1 complex subunit sigma and histones H2A/H2B. The host genes most highly correlated with *P. knowlesi* genes include pathogen detection genes like IFIT3, PLA2G4C, MX1, OASL, DDX60, OAS2, RSAD2, MX2, DHX58, IFIH1, STAT1, FAS and TLR4 (for functional annotations[222], see Appendix Table C.2.11). As reported in Chapter 2, it is worth noting that most of these signaling genes are upregulated in Mf at TP3, which further supports the hypothesis of earlier pathogen detection in Mf than in Mm.

For many mammalian hosts, the PRR signaling pathway has been credited for detecting PAMPs or DAMPs, DNAs, and other large molecules[223-225]. In our case, the activity of this pathway is consistent with expression of PRR-related genes in both hosts during the appropriate

infection time points (TPs; Appendix Table C.2.12). Here we concentrate on the log phase of infection because this phase is associated with the most similar features between the two hosts, and any observed differences might highlight critical processes. It is interesting to note that many of the PRR signaling genes are in fact different between Mm and Mf. This difference implies that even though the more general PRR signaling pathway is activated in response to the detected pathogen, the specifics of the pathway operation are apparently different, which could be due either to the detected pathogenic content or the interpretation of the signaling event by the host's immune responses. The major genes differentiating the specifics of PRR signaling include TLR5, NLRP6, TNIP3, SLC15A4, SLC15A3, CD36 and CD300A. We had reported enriched pathways before, but the specific differences are not as easy to deduce (Appendix Table C.2.13). Such differences are evident in subsets of the TLR signaling cascade, especially in TICAM1/RIP1 mediated IKK complex recruitment. Higher expression of corresponding ubiquitination genes (UBE2D1, UBA52, RPS27A and BIRC2) in Mm suggests activation of NFκB[226], which might be responsible for stronger inflammation.

These differences between hosts are carried forward toward responses by networks of protein kinases. We examined enrichment of several protein kinase cascades including MAPK, G protein coupled (GPC) receptor systems and p21 activated cascades. Both host species exhibit higher activation of atypical cytokine activated MAPK4/6 signaling involving PAK (p21 activated kinases)[227,228] in comparison to the typical stress activated p38/MAPK signaling pathway (Appendix Figure C.1.2A). Major differences in protein kinase activities are associated with higher inhibition activity in Mm, which is probably due to peptidyl tyrosine dephosphorylation. Although not fully understood, this pathway has been implicated in both pro- and anti- cell proliferation roles[229] and might be responsible for downstream differences in p53 and HSP27 related cell cycle

activity. Differential regulation of protein kinase C activity (Appendix Figure C.1.2B) might explain these differences[230]. GPC receptor signaling is highlighted in both similarities and differences in the two hosts, with higher positive enrichment of purinergic nucleotide signaling (P2Y) and a much negative enrichment of glutamate receptor signaling. GPC purinergic nucleotide receptor activity not only explains the upregulation of purine metabolism in malaria, but also points to a potential role of macrophages[231]. Macrophage production is significantly upregulated in both hosts during log phase (Appendix Table C.2.14A,B) with same direction of fold change as P2Y signaling. Finally, Ca^{2+} sensing GPC glutamate receptor shows differences between the two hosts in both binding and signaling pathways suggesting the importance of calcium homeostasis in both hosts.

4.4.2 Ribosomal proteins control p53 pathway

In Chapter 2 we had suggested that control over the p53 pathway during the log phase of the infection as a crucial difference between the immune responses of the two host species. Binding of p53 to its target response element leads to the expression of a multitude of genes with a spectrum of functions, including cell cycle growth arrest, DNA repair, cellular senescence and apoptosis[232] (Figure 4.3). An active p53 pathway also protects cells against reactive oxygen species (ROS) through antioxidant genes like TP53INP1[233] (Appendix Figure C.1.3). Indeed, the early response of this pathway in Mf (at TP4) might be crucial in saving cells from apoptosis via PIG3[234] (TP53I3 gene). Co-expression analysis in Sec 4.4.1.2 reveals several modules controlling these functions, including RPS19, CEP55, COX6C, AHCY, UBA2 and ZNF395 (Appendix Table C.2.4).

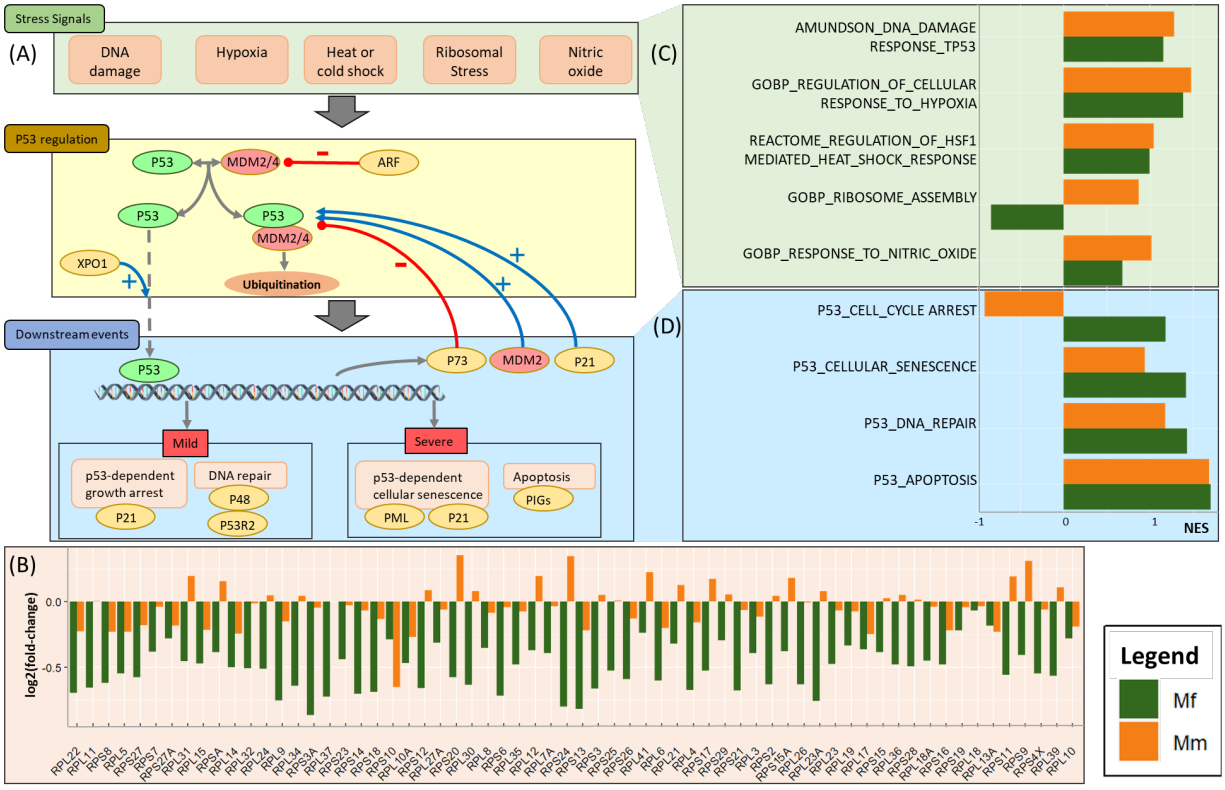


Figure 4.3- Involvement of the p53 pathway. (A) Schematic of cause, regulation, and effect of p53 pathway activation during the log phase of a malarial infection. This includes the key stress signals that engage the p53 regulation pathway and results in regulation of downstream events. (B) Barplot for log2 fold-change of expression of ribosomal proteins at TP4 from baseline comparing both hosts. (C) Barplot for normalized enrichment scores (NES) of stress signals involved in p53 activation. (D) Barplot for normalized enrichment scores of p53 related downstream events.

As observed in the previous section, certain MAPK signaling mechanisms potentially regulate the p53 pathway. Generically, the p53 pathway is operational in the presence of molecular stresses and depends on their severity as well as other factors. Cellular stress signals that activate a p53 response include hypoxia, DNA damage, ribosomal and oxidative stresses, among others[235]. These stresses are, of course, not independent of each other and manifest in an interrelated manner. In the case of a *P. knowlesi* infection, this interdependence can be seen in the enrichment of associated genes. In particular, the enrichment analysis demonstrates that ribosomal stress is a differentiating factor between the two host species, with stress caused by substantial

downregulation of the ribosomal assembly complex in Mf (Figure 4.3C). This downregulation is achieved through the activity of various RNA polymerases (Appendix Figure C.1.4), and the significant downregulation of PLOR1C, POLR2E, POLR2A, POLR2J and POLR2L in Mf at TP4 suggests that these genes might be crucial for the control of p53.

The downregulation of associated ribosomal proteins (RPs) in Mf at TP4 (Figure 4.3B, Appendix Figure C.1.5) is indicative of alterations in ribosomal biosynthesis that results in unassembled RPs and 5S rRNA, which binds to the p53 inhibitors MDM2 and MDM4 and thereby prevents p53 degradation[236,237]. As a consequence, p53 facilitates translation from its mRNA internal ribosome entry site (IRES). Indeed, the co-expression network analysis reveals high correlation of MDM4 with *SICAvar* Type 1 transcripts, which suggests direct control that might be crucial in this regulation.

As a consequence of RP downregulation, the p53 pathway in Mf is upregulated, which is reflected in higher levels of enrichment. However, the less pronounced changes in cell cycle arrest and DNA repair appear to be the strongest differentiating factors between the two species. The important genes involved in these processes include CDKN1A (p21), E2F7, PML and MDM2 (upregulated) and TP73 (downregulated). Other genes that are downregulated in Mf near the peak of infection include TP73, BTG2, PCBP4, RRM2B and XPC.

In addition to the direct action of p53, cell cycle arrest is regulated by p21 and mediated by AP-1. Specifically, AP-1 forms a complex with FOS, a process governed by activation of JNK signaling, which leads to the expression of JUN and formation of the FOS-AP-1 complex. This complex not only aids in various cellular events but also blocks p53-mediated cell cycle arrest due to p21 activation[238]. An observed significant downregulation of genes associated with FOS and

JUN in Mf is probably due to downregulation of the activity of the signaling protein MAPK8IP1 during the log phase, which prevents AP-1 from blocking p21 activity.

Another facet of p53 pathway activation and control is provided by the transcription factor HSF1 (heat shock factor 1). Dysregulation of ribosomal biosynthesis processes leads to proteotoxic stress and a balance between these processes must be maintained[239]. In comparison to Mf, Mm has higher ribosomal biosynthesis and senses higher proteotoxic stress during the log phase of infection, and these processes are further increased near the peak (Appendix Figure C.1.6). Of note in this context is the differential expression of chaperone-mediated protein folding genes – HSPA1A, HSPA8, DNAJB1 and FKBP4. This expression results in upregulation of HSF1 target genes in Mm. p53 has been shown to form multi-chaperone complexes with HSPA1, DNAJB1 or HSPA8, while FKBP4 is essential for its transport to the nucleus[240]. Among the apoptotic targets, ATF3 enrichment at both TP4 and TP5 in Mm highlights differences with Mf. Other notable apoptotic targets include BAG3, MAP4K1, CFLAR and CASP10, along with FOS and JUN.

Finally, the control over cell proliferation in Mf is eased near the peak, which leads to upregulation of adaptive immune cells and, in particular CD4 memory activated and follicular helper cells (Appendix Table C.2.14F), both of which enhance the adaptive immune response by supporting B cells and CD8 T cells[241,242].

4.4.3 Effects of inflammation on immune response and cell proliferation

Control of cell proliferation in Mf during the log phase of infection constitutes a stark contrast to the elevated inflammatory response in Mm. This difference can be seen clearly in the enrichment of several inflammatory pathways, elevated expression of inflammatory genes, inflammasomes and inflammatory biomarkers like the kynurenine (Kyn)/Trp ratio (see Appendix

Figure C.1.7, C.1.8 and next section). Co-expression analysis revealed that most of the inflammatory genes are part of the innate immune module (FBX06 module). Not surprisingly, this module is most significantly changed (logistic regression) in both hosts during log phase. Although the fold-change for this module is similar between the hosts, the lower adjusted p value (q) and higher log-odds (B) suggest a stronger role of this module in the innate immune response in Mf (q < 3e-17 / B > 32) in comparison to Mm (q < 1e-10 / B > 17). This module further highlights the differences between the hosts near peak infection as Mm (q < 7e-7 / B > 8) maintains its immune response while Mf (q < 4e-5 / B > 1) does not. The module that is most highly correlated to FBX06 is FYB, whose functional annotation points to neutrophil activation and intracellular vesicle transport. Also, worth noting is that both these modules are negatively correlated to ribosomal biosynthesis and localization modules RPS19, ZNF395, CHD6 and RASGEF1A. This implies an important, sustained balance between immune related inflammation and control over the cell cycle.

Several similarities in inflammation gene sets are found between the hosts, especially with respect to an LPS-like inflammatory response, which probably is a symptom of the NLRP3 inflammasome (Appendix Figure C.1.7A). This phenomenon might be attributed to significant upregulation of monocytes and monocyte-derived pro-inflammatory M1 macrophages (Appendix Table C.2.14A,B, Figure 4.2E), which are the first-line cells expressing inflammasome genes[243] (Appendix Figure C.1.8). Important differences are detected in the inflammatory response cytokine production and an antigenic stimulus (Appendix Figure C.1.7B). Even though the positive regulation of these functions is similarly enriched in the two hosts, as seen in the important genes NOD2, GPX1 and IL12B, the negative regulation shows a distinct and opposing enrichment. The main distinguishing genes include IL10, NLRP6 and ABCD1.

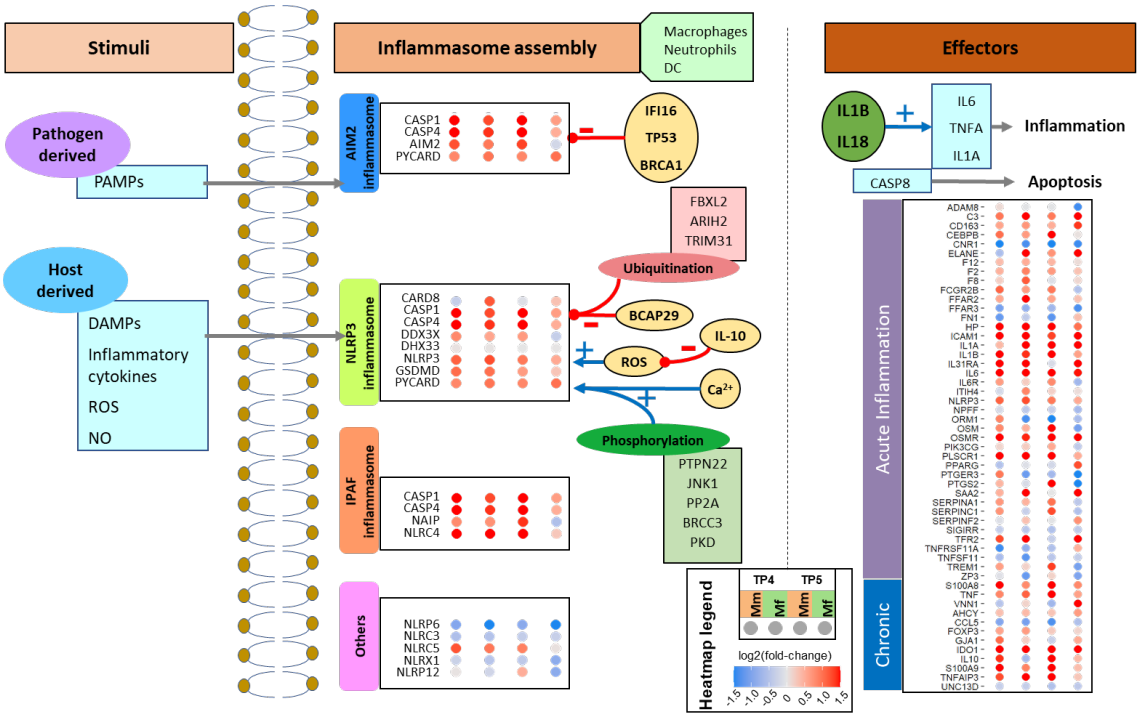


Figure 4.4- Schematic of inflammasome assembly process. Various host and pathogen derived stimuli are responsible for initiation of the inflammasome assembly process. Based on the stimuli, multiple inflammasome assembly processes may get triggered, namely AIM2, NLRP3 and IPAF inflammasome assembly in macrophages, neutrophils and dendritic cells. The NLRP3 assembly process is crucial in both acute and chronic inflammation. This process is very closely regulated by various signals and processes including ROS and Ca^{2+} homeostasis. This control results in several effectors executing important processes like inflammation and apoptosis. Different effectors are activated during acute and chronic phases. Some of the key effector genes include IL6, IL1A, ILIB, S100A8, IL10 and IDO1.

The two hosts show similar enrichment of the chronic inflammatory response; however, Mm has a higher acute inflammatory response (Appendix Figure C.1.7B), reaffirming the stronger inflammation in Mm during log phase. The chronic inflammation changes near the peak and is mostly driven by crucial genes like IL10, IDO1, TNF, TNFAIP3 and CXCL13 (Figure 4.4).

Exploration of innate immune components of inflammation reveals a crucial difference in S100 proteins. Ca^{2+} sensing S100 proteins have a wide range of functionality that includes cell apoptosis, proliferation and inflammation[244]. Differential upregulation of S100A8, S100A9,

S100A16 and S100P in Mm suggests a potential role of Ca^{2+} in inflammation[245], while upregulation of S100A4, S100A2 and S100A3 in Mf suggests possible regulation of p53[246,247]. Since neutrophils release S100A8/A9 during inflammation, their differential expression together with that of genes in the FYB module mediates Ca^{2+} signaling, which positively regulates NLRP3 inflammasome assembly and therefore pro-inflammatory activity of NFκB and AP-1[248,249]. This inflammatory activity is further exacerbated by master regulator DDX3X[250]. Further enrichment of processes specifically associated with innate and adaptive immune processes reveals an interesting pattern that succinctly differentiates the responses of the two hosts. Namely, TLR4 signaling is stimulated by Ca^{2+} via S100 proteins (S100A8 and S100A9), which enhances the inflammatory activity of NLRP3 inflammasome. These inflammatory pathways are responsible for IFNβ regulation and IL6 production (Figure 4.5).

Further exploration of differences in immunological signatures reveals several important similarities and differences between the two hosts (Appendix Figure C.1.9). Both hosts show significant enrichment towards FOXP3+ CD4+ naïve T-reg cells (GSE37533[251], GSE42021[252]), with gene sets pointing to the strongest enrichment of a thymic T-reg subset of intermediate maturation, $CD24^{int}$. A related important difference appears between the two hosts: Mf has a higher mature ($CD24^{low}$) subset while Mm has a higher immature ($CD24^{hi}$) subset. Both hosts have enriched naïve B cells (GSE42724[253]), even though this change could not be confirmed by deconvolution analysis. Differences can be seen in gene sets derived from IL6 and IL10 stimulation as well. Taken together, these signatures suggest downregulation of key genes in Mf which is not observed in Mm. Important genes that seem to be regulating this process in both Mm and Mf include IL6, IL6R, TGFB3, IL23A, IL10 and SOCS3.

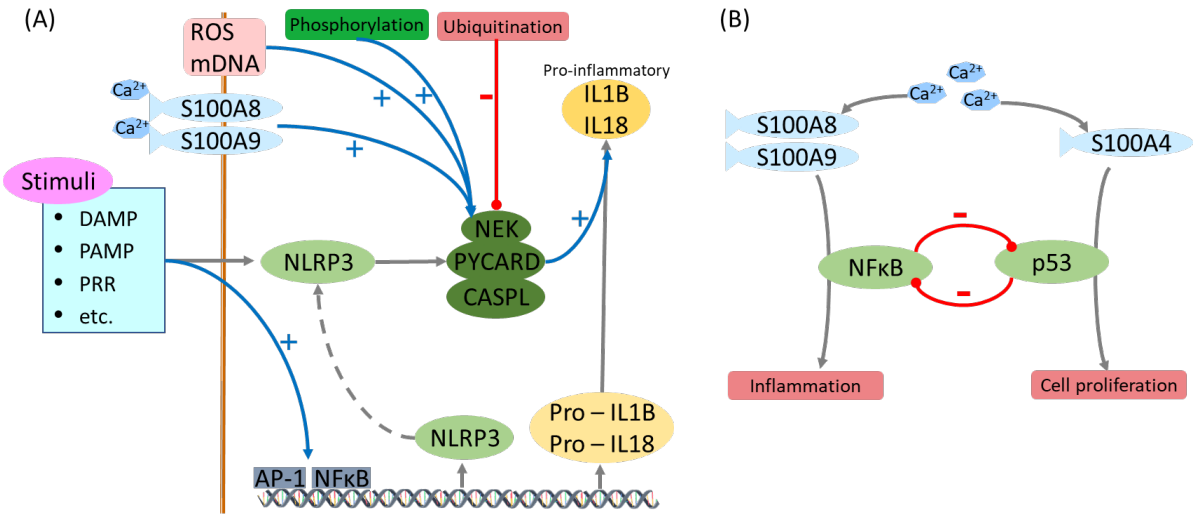


Figure 4.5- Schematic showing regulation of NLRP3 inflammasome assembly and its relationship with Ca^{2+}. (A) Detailed schematic of NLRP3 inflammasome signaling including initiation and regulation of the assembly process followed by pro-inflammatory effectors. Ca^{2+} transporters S100A8 and S100A9 are upregulated in Mm, which suggests a corresponding proinflammatory response. (B) Balance and cross-regulation of p53 and NFKB showing importance of Ca^{2+} homeostasis. While S100A8 and S100A9 are upregulated in Mm, S100A4 is upregulated in Mf. Based on the enrichment of respective inflammatory and cell proliferation pathways in Mm and Mf, it is clear that Ca^{2+} homeostasis plays a central role it its balance.

Although not conclusive, pre-infection state differences in cell populations point to eventual differences in the immune response. For instance, at baseline, Mm has significantly more naïve CD4+ T cells while Mf has higher levels of neutrophils[254] (Figure 4.2E, Appendix Table C.2.14C). Although there is no significant relative difference in cell populations during the log phase of infection between the hosts (Appendix Table C.2.14E), these initial pre-infection differences persist (Appendix Table C.2.14D) and may be a key differentiating factor in the immune response.

4.4.4 Changes in tryptophan metabolism suggests higher NAD metabolism

Metabolomic and transcriptomic analysis of the Mf and Mm hosts revealed prominent differences in the expression of genes associated with Trp metabolism at TP4 and TP5 (Figure 4.6C). Trp metabolism can coarsely be divided into pathways responsible for serotonin and melatonin, NAD^+ and Kyn synthesis. Serotonin and related compounds are not of interest in the present context, and their concentrations in blood are very low. NAD^+ metabolism plays a crucial role in cellular energy regulation as well as the handling of ROS. The Kyn pathway is responsible for the biosynthesis of several metabolites that play key roles in immunomodulation and inflammation.

Both hosts had lower Trp and higher Kyn levels in the peripheral blood during the log and peak phases of infection in comparison with baseline levels (TP1 and TP2), as observed previously with *P. coatneyi* infection of rhesus macaques[255]. Of special note here, however, is that the Kyn/Trp ratio, a known inflammatory biomarker, is reduced in Mf near peak infection, whereas it remains at the same level as during the log phase in Mm (Figure 4.7).

IFNγ signaling is responsible for upregulating the expression of IDO[256-258], which converts Trp into Kyn (Figure 4.6A). Even though IFNγ signaling is upregulated in both hosts, a higher degree of signaling in Mf near the log and peak phases results in higher IDO expression, which thereby leads to higher conversion of Trp to Kyn in Mf than in Mm (Figure 4.6C). One might expect that this increased activity should lead to a higher level of Kyn. Yet, we observe lower levels of Kyn and a lower Kyn/Trp ratio in Mf (Figure 4.7), which however is easily explained by the increased activity of the subsequent enzymes KYNU and KMO in the Kyn utilization pathway, which ultimately lead to higher NAD^+ biosynthesis and immunomodulatory activities. In addition,

earlier downregulation of AhR and AADAT in Mf suggests potential differences in AhR signaling (see next section).

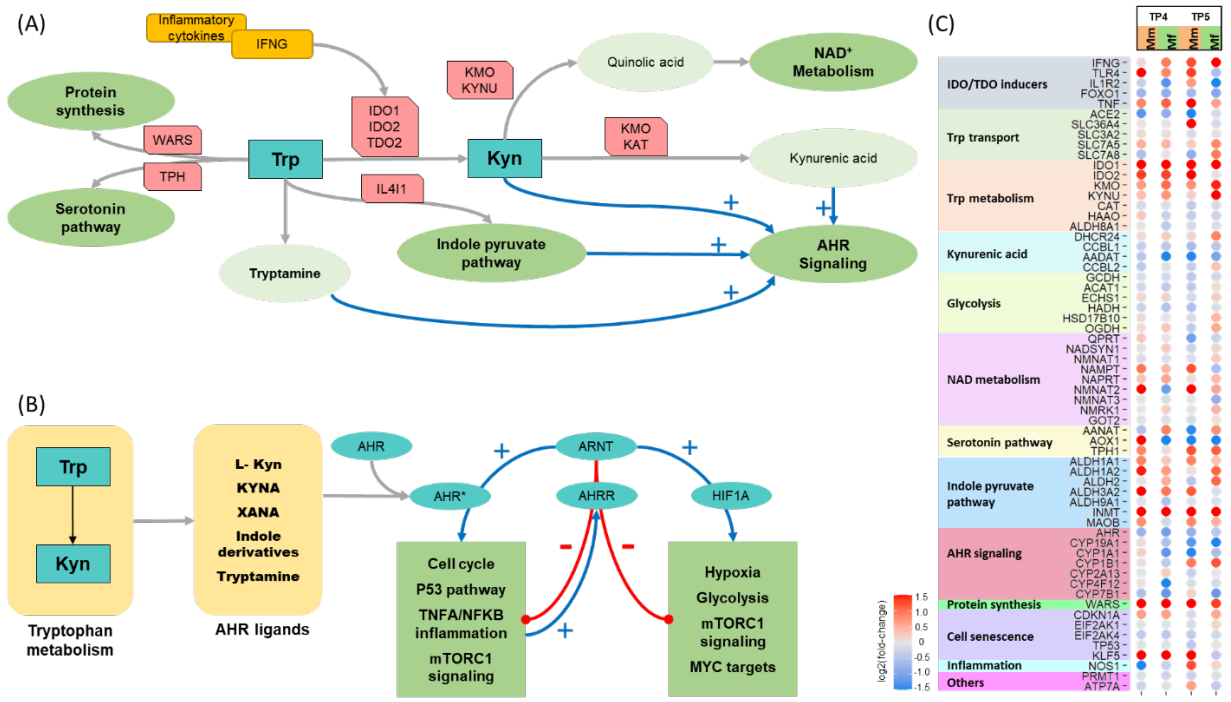


Figure 4.6- Schematic of tryptophan metabolism, its regulators and downstream effects. (A) Schematic showing key features of tryptophan metabolism. Inflammatory cytokines like IFNγ control the rate limiting step (IDO1). Products of this pathway are involved in important processes like AHR signaling and NAD metabolism. (B) Schematic showing AHR signaling. The key regulators are kynurenine derived ligands of AHR and competition with AHRR and HIF1A for ARNT. These, in combination, control several important processes like the cell cycle, inflammation, hypoxia, and others. (C) Heatmap of differential expression of significant genes involved in tryptophan metabolism comparing the two hosts across TP4 and TP5.

In order to elucidate the role of Trp and Kyn levels in the blood and then understand changes in Trp-Kyn metabolism in white blood cells (WBCs) during the infection, we adopted existing metabolic models of Trp metabolism in brain and liver[259] and adapted them to reflect Trp metabolism in WBCs using methodology described in Chapter 3. With this adapted model we can clearly differentiate Trp-Kyn metabolism in blood from brain and liver (Appendix Figure C.1.10).

The model confirms that over 90% of Trp in whole blood is channeled toward Kyn through the activity of IDO, compared to tryptophan-2,3-Dioxygenase (TDO) in the liver model. Kyn itself is the substrate for different reactions, and flux control analysis[260,261] reveals that KMO is the most important control point for Kyn utilization. Extension of the blood model during the log phase of the infection shows for both host species that the transport of both Trp and Kyn through the cell membrane is lowered in comparison to the baseline (Figure 4.7B). This finding is interesting as it potentially leads to serum levels of Trp and Kyn in both hosts that change significantly during this phase and lead to a higher Kyn/Trp ratio (Figure 4.7A). At the same time, these increases are accompanied by a major reshuffling of fluxes, which affects the metabolite concentrations inside the cells.

Specifically, Trp can be metabolized through six reactions, among which the pathway toward Kyn is most important, based on relative fluxes (Figure 4.6A). Indeed, if the Trp concentration inside the cells is decreased, the effluxes out of the Trp pool are also decreased, with the notable exception of the Kyn pathway, which receives essentially a normal influx from Trp. This flux is important, because the pathway later leads to the formation of quinolinic acid, which is a precursor of NAD^+ and thus affects energy metabolism and redox handling. At this juncture, the differential expression of KMO becomes even more important as a control point for Trp metabolism: here, it causes a higher flux in Mf toward NAD^+ synthesis (Figure 4.7C). This enhanced flux from Trp to Kyn is compensated in both species by decreased fluxes from Trp toward protein synthesis and serotonin production (Appendix Figure C.1.11).

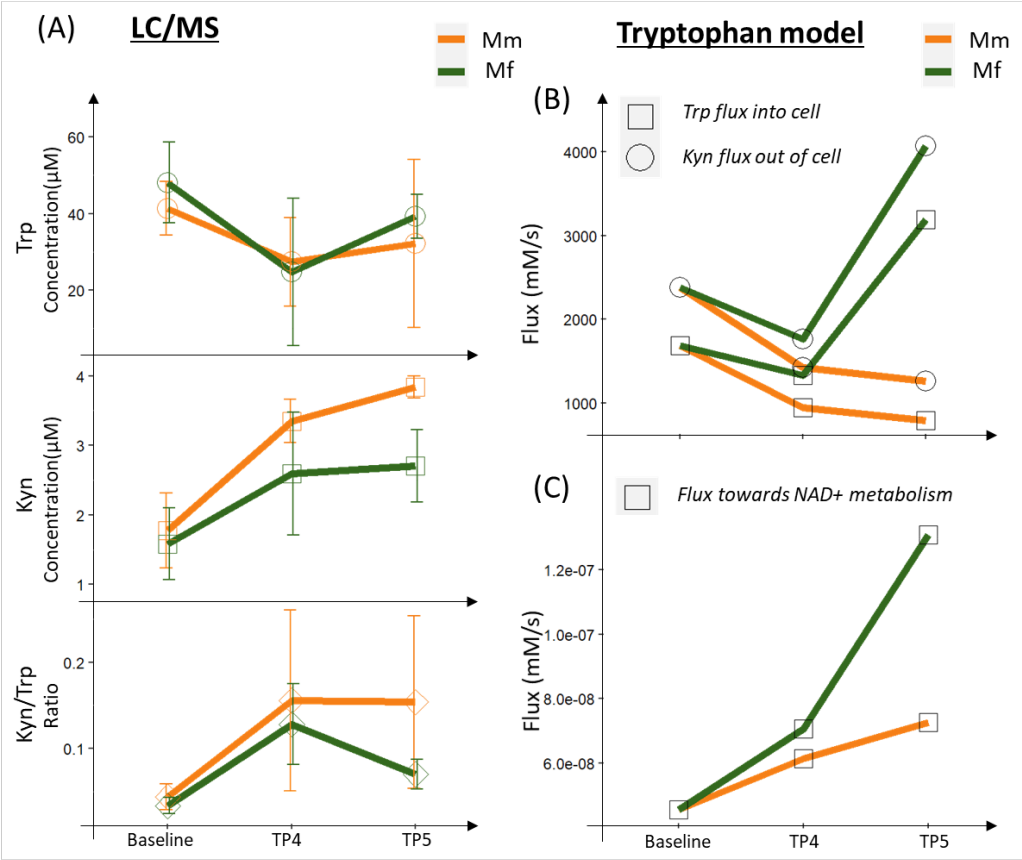


Figure 4.7- Metabolomics (LC-MS measurements) and model predictions for Tryptophan metabolism. (A) LC/MS measurements for Trp, Kyn and Kyn/Trp ratio across infection timepoints comparing Mm and Mf. (B) Fluxes predicted by Trp model for Trp consumed and Kyn produced by blood cells. (C) Trp model predicted flux towards NAD metabolism.

Near the peak of infection, the differences in the two hosts are particularly pronounced, with Trp and Kyn transported through the cell membrane at higher rates in Mf than in Mm. As a result, the fluxes through the Kyn pathway are higher in Mf (Appendix Figure C.1.12). At the same time, the Kyn/Trp ratio is lower in Mf during this phase of the infection, presumably due to the enhanced activity of KMO (Figure 4.7A). It is also worth noting that the higher flux toward NAD^+ metabolism persists in Mf (Figure 4.7C). Furthermore, the concentrations of other Kyn compounds remain high, and these are potential ligands of the AhR (see next section), which ultimately serves as a transcription factor for numerous genes. (Figure 4.6B).

Among the other effluxes out of Trp, indole-pyruvate and tryptamine pathways are also responsible for AhR activation[262,263] (next section). Trp is incorporated into proteins via tryptophanyl-tRNA synthestases (WARS proteins), a process that directly links Trp sensing to p53 activation[264]. Changes in these fluxes during infection further show the central role of Trp metabolism (Appendix Figure C.1.14A,B).

4.4.5 AhR signaling and the role of AhRR in controlling AhR and HIF1A signaling

AhR belongs to the basic helix–loop–helix-PER-ARNT-SIM (bHLH-PAS) superfamily of transcription factors where multiple other members interact with each other and therefore affect each other's functionality. Prominent members include AhRR, ARNT and HIF1α (Appendix Figure C.1.13).

Most of the biologically active intermediates of the Kyn pathway, as well as several other compounds, can act as ligands for AhR[265,266] (Figure 4.6B, Appendix Table C.2.15), which makes this receptor a central control point for multiple physiological changes, *e.g.*, in heme degradation, hypoxia and Trp metabolism. Once a ligand binds, AhR can form a complex with the nuclear transporter ARNT, which is translocated to the nucleus. Once in the nucleus, the AhR-ARNT complex binds to the ARE promoter region of numerous genes.

Different Kyn derivatives may act as ligands for AhR, and their dynamics differentiates the Mf and Mm hosts during infection. Additionally, IL4I1 activity leading to indole-pyruvate derivatives from Trp synthesis also activates AhR[267]. Similarly, multiple other ligands have been associated with AhR activity, and some of these may constitute further differences between the two host responses (Appendix Figure C.1.14C-F). For instance, *Plasmodium*'s consumption of hemoglobin releases heme, which is metabolized (Appendix Figure C.1.15). Certain AhR ligands

that are derived from heme metabolism[268,269] may point to additional differences between the two hosts.

Another level of control of AhR signaling occurs through the competition between AhR, AhRR and HIF1α for ARNT and thus for transport into the nucleus and binding to their corresponding response elements. Given the molecular similarity of the competitors, it is not surprising that most of the downstream genes are simultaneous targets of both the AhR-ARNT and the HIF1α-ARNT complexes (Appendix Table C.2.16A).

Exploring these targets through an enrichment analysis shows that both complexes act quite similarly during log phase (Appendix Figure C.1.16B, TP4). Yet, several differences emerge near the peak of infection (Appendix Figure C.1.16B, TP5). The most pronounced differences emerge with respect to higher upregulation of AhR targets and HIF1α targets in Mf, while targets of the AhR repressor AhRR are downregulated in Mf. It appears that the relative hypoxia stress is quite different between the two hosts (Appendix Figure C.1.17), but it is unclear how the balance is achieved between these complexes and their corresponding genes.

To shed light on the interference among these complexes, many of which share numerous common targets, we calculated enrichment of each subset of these targets (Appendix Figure C.1.16). Specifically, we divided the targets into three major groups (AhR targets, HIF1α targets, and AhR and HIF1α targets) and compared them with and without the AhRR binding site to account for repressor activity (Appendix Figure C.1.16A).

The effect of AhRR on AhR targets is quite clear in Mf, with lower enrichment of targets at both TP4 and TP5, as opposed to almost no effect in Mm. Corresponding effects of AhRR on

HIF1α targets are not easily identified. At TP4, HIF1α targets with the AhRR binding site are more enriched than without AhRR. At TP5, AhRR containing HIF1α targets are enriched more in Mm and less in Mf.

As there are multiple levels of regulation, it is difficult to predict the activity of these targets without further experimentation. However, one may try to elucidate the specific functionality of these targets by identifying the key genes along with their functional annotation. The transcription factor complexes in question are associated with a wide range of genes with diverse functionality (Figure 4.6). Functional annotation of AhR and HIF1α targets shows their involvement in key process like the p53 pathway, heme metabolism, cell cycle related pathways, and immune related IFNγ and NFκB pathways (Appendix Table C.2.16B). The complex nature of this response makes it difficult to elucidate the specifics and differences during a *P. knowlesi* infection, but the activity of individual genes suggests potential outcomes. Their roles in immune and inflammatory processes are evident in the activity of genes like OASL, STAT3, IRF5, IL6, DDIT4, NRF2, REL, and LAG3. These IFNγ signaling genes create a positive feedback loop, because IFNγ directly regulates IDO expression, which leads to enhanced levels of the AhR ligand Kyn. The control over cell proliferation is evident in the operation of p53 and other cell cycle related genes like MXD1, FOS, BCL6, GADD45A, and CREBRF. Another possible contributor with respect to malarial infection is heme metabolism with target genes including CCND3, BLVRB, and KLF1.

4.5 Discussion

Malaria has haunted mankind throughout its history. Even after several decades of active research, malaria continues to be a severe global health concern with over 400K fatalities and about 3.2 billion people at risk annually[1]. Among the six species of *Plasmodium* known to cause

malaria in humans, *P. knowlesi* has become recognized as a major zoonosis in Southeast Asia[270-273]. A *P. knowlesi* infection in humans may range from mild to severe, with 6-10% of the cases considered severe[16,274]. A deeper knowledge of the details of *P. knowlesi* infections could provide a crucial basis for understanding the immune responses in general and for comparing resilient and severe malarial responses in particular. As a zoonotic species, *P. knowlesi* has the advantage that it can be studied in different NHP species[20,211]. Among these NHP models, *Macaca mulatta* (Mm) and *M. fascicularis* (Mf) provide unique advantages specifically for comparing *P. knowlesi* infections with different disease progression. Namely, even though Mm and Mf are evolutionarily very close, Mm, once infected, suffers from increasing parasitemia, which is in almost all cases fatal if not treated, whereas Mf controls parasitemia and escapes death without treatment[211,275,276]. These dramatic differences provide unparalleled opportunities to study the details of host physiology and immune responses in the context of host-parasite interactions and explore mechanisms of resilience in human malaria, and to potentially relate the findings to other diseases that may also show drastically different possible outcomes.

In Chapter 2, we had established crucial differences in the transcriptomics of the two hosts that ultimately determined the outcome in terms of susceptibility and resilience. As also noted in the clinical assessment by Peterson *et al.*[211], transcriptomics analysis showed that Mf detects the pathogen earlier than Mm, and even though both host species mount a similar immune response, Mf starts controlling inflammation as early as the log phase of infection[53]. Subsequently, Mf switches the immune response towards cell proliferation pathways, which presumably aids recovery[53]. The current analysis explores the key findings further and explains the molecular functions that determine the mild or fateful outcome. Interestingly as well, early detection of the

parasites by the Mf animals is also consistent with a rise in temperature in this species immediately upon patency, by 7 dpi[211].

The results discussed in this chapter show consequential differences in signaling mechanisms beginning with the early detection of the presence of *P. knowlesi* pathogens by Mf. Once the merozoites invade the RBCs, they transform the iRBC and express different antigenic forms of surface molecules in an attempt to escape the immune response[183,277,278]. Specifically, antigenic variation of *P. knowlesi* SICA proteins is a main factor responsible for chronicity in Mm[278] (and reviewed in Galinski *et al.*[213]). Moreover, expression of *SICAvar* genes in *P. coatneyi* have been shown to change as chronic rhesus monkey infections are established, also suggesting a role for metabolites in regulating these changes[255]. Our correlation analysis of host and pathogen transcripts sheds light on possibly involved *SICAvar* Type 1 genes[185-187] along with correlated host genes. The specific correlations of individual transcripts from this large pathogen gene family – with 136 *SICAvar* members[187] – could shed light on its transcripts and their variable gene expression, which may trigger different antibody responses. Additionally, correlations with host genes, especially the differentially responding IL10 and HSPA6 genes, can help associate parasite markers with the host immune response.

On the host side, differences in the mechanisms for pathogen detection and PRR signaling pathways are surprisingly subtle. However, these differences are magnified downstream with MAPK signaling. There is a close relation of these signaling cascades, especially the GPC activity with the p53 pathway and cell cycle[279-281]. Ca^{2+} drives intracellular communication and interacts with GPC receptors to regulate various aspects of the cell cycle, and by extension, regulates inflammation and apoptosis during infection. This regulation is even further augmented by inflammasome activity (Figure 4.4). Specifically, some of the Ca^{2+} binding S100 proteins

(S100A8, A9 and A4) might be differentiating factors between the two hosts. While S100A8 and S100A9 aid the inflammasome assembly, S100A4 assists with the regulation of the p53 pathway. Additionally, the inflammasome assembly process is regulated by multiple other factors including ROS, IL10 and transcription factor AP-1. These factors do not only relay the stress response but also seem to be important in regulating the p53 pathway.

The most strongly differentiating factor between the two species appears to be the control of cell proliferation by Mf during log phase via the p53 pathway, along with subsequent inhibition that leads to recovery. Similar stresses can trigger both inflammation and cell proliferation, but it appears that it is the stress related to fundamental ribosomal assembly that causes the inhibition of cell proliferation in Mf through the p53 pathway. Several upstream kinases have been shown to cause this stress. Since ribosomal assembly is one of the most energy intensive functions, inhibition of this fundamental function to conserve energy seems likely[239]. Of course, that is not the sole purpose. In particular, we observe that ribosomal assembly leads to differences in p53 based cell cycle arrest and DNA repair in Mf. The interrelatedness of this pathway with p21, AP-1 and HSF1 activity provides additional regulators that might be responsible for balancing cell proliferation with inflammation.

Another known inflammation biomarker, the Kyn/Trp ratio, shows surprisingly deep integration with these processes. Even though the induction of IDO in malarial infection is quite often discussed[282-285], its biological significance for the immune response is in general poorly understood. Nonetheless, application of methodology discussed in Chapter 3 helped us to develop a mathematical model to show that the direct upregulation of IDO through IFNγ signaling quite clearly changes the Kyn/Trp ratio during the infection (Figure 4.6). This metabolic model is able to shed light on several important, although indirect implications, such as the importance of KMO

and KYNU in regulating fluxes, redirection of fluxes towards NAD^+ metabolism, and metabolite pools of kynurenine compounds as ligands for AhR. In summary, Trp metabolism diverts the fluxes towards the essential functions, and especially NAD^+ metabolism and protein synthesis. The higher activity in Mf also indicates that this host maintains essential functions in spite of the inflammation. Further analysis into the kynurenines shows an impactful control of AhRR in regulating both AhR and HIF1α related signaling. This process includes a competitive effect of multiple stresses, hypoxia and infection induced damage and cytokine response in determining the overall outcome.

Although this analysis dives deep into multiple molecular mechanisms that play crucial roles in permitting resilience of the host, it only paints a crude image of the immune response over time. For example, a more detailed longitudinal and immunologically based analysis of *SICAvar* gene expression and switching of SICA proteins in each host (and with different parasite species[255]) is likely to advance our understanding of the different antibody responses and immune evasion mechanisms (reviewed in Galinski *et al.*[213]). The combined analysis of immune response, inflammation and cell proliferation also seems to reveal Ca^{2+} as a crucial factor, which is known to play a role in iRBC egress[286]. If this general finding can be validated and cross referenced with other bacterial and viral infections[287-289], improved understanding of Ca^{2+} homeostasis might lead to novel targets that could naturally aid the immune response against *Plasmodium* infection. Similarly, the metabolic model we employed, adjusted for transcriptional changes during the infection, provides a deeper appreciation of the mechanisms of Trp metabolism and could possibly be extended to identify targets that could predictably adjust metabolism to aid in resilience.

Overall, this work interprets transcriptional data and integrates them in a manner that provides deeper understanding of *Plasmodium* infections. It is hoped to suggest new avenues of studying malaria and identifying valid candidates for future drug development.

CHAPTER 5. CONCLUSIONS AND FUTURE STUDIES

Malaria is one of the world's deadliest infectious diseases, with an impact spreading across half of the total human population and reaching in over 100 countries[1]. Its causative parasite *Plasmodium spec.* has been actively researched for over a century, but an effective vaccine remains elusive. The main reason for this dichotomy is the highly complex lifecycle of *Plasmodium* and its rich antigenic variations.

The studies described in this dissertation aimed at understanding host-parasite interactions between various *Plasmodium* species and two non-human primate (NHP) species – *Macaca mulatta* (Mm) and *Macaca fascicularis* (Mf). Of particular interest was the stark difference between these two monkey species when infected with the same zoonotic – *Plasmodium knowlesi*. The most striking difference between the infections in these two hosts is that Mf survives the infection with mild parasitemia while Mm succumbs to the ensuing severe infection. Mm and Mf are evolutionary very close with reported cases of interbreeding in shared geographical areas[290]. However, it is known that Mf is a natural host of *P. knowlesi* in the wild, while Mm is not[13]. Considering how closely related the two macaque species in our analysis are, one might have envisioned some simple switch in the expression of a few genes that would result in the two disparate fates following *P. knowlesi* infection. Instead, our comparative analysis demonstrates that the differences in responses are rather subtle but widely distributed within the two macaque species. In fact, the main and most dramatic host-species-specific difference we detected lies in the timing of responses, with Mf able to identify the infection faster and to trigger early immune responses that can suppress the infection and support the macaque's recovery.

We performed a comparative longitudinal study between the two species after they had been infected with *P. knowlesi* pathogens. The transcriptomics comparison established distinct crucial differences between the two species. Importantly, Mf showed signs of early detection of the pathogen. A detailed follow-up analysis identified pathogenic surface antigens as responsible for differences in PRR signaling pathways and associated protein kinase signaling. Although these differences are small, high correlations of pathogenic *SICAvar* Type 1 surface antigens with important immune related genes like IL10 and ELK4 suggest a deeper impact.

As the infection progressed in the two species, we observed comparable parasitemia and similar immune responses. For instance, during the initial log phase of infection, the immune response of both species is dominated by genes related to IFNα and IFNγ. This response is carefully orchestrated by the transcription factors STAT1, STAT2, and IRF9. A closer look revealed that during this phase of infection the differences are exclusively related to inflammation. Mm shows increased inflammation with higher expression of IL10, enrichment of IL6-JAK-STAT3 signaling and an altered Kyn/Trp ratio, which is a known biomarker. By contrast, Mf exhibits clear signs of controlling inflammation and cell proliferation by enrichment of the crucial p53 pathway. We explored several probable causes for these key differences in inflammation. Differences in glutamate GPCR activity and NLRP3 inflammasome assembly showed that Ca^{2+} homeostasis plays an important role in maintaining balance between inflammation and cell proliferation. Changes in Ca^{2+} signaling are detected via glutamate GPCRs. Ca^{2+} directly affects the inflammasome assembly process, which regulates inflammation through the IL6 signaling pathway. Metabolic modeling of the tryptophan pathway not only explained the higher Kyn/Trp ratio in Mm but also emphasized possible roles in NAD metabolism and in AhR signaling, which is associated with the regulation of inflammation.

As the infection neared its peak around day 10 of infection, the immune responses between the two hosts diverged noticeably. While Mm continued with an inflammation dominant immune response, downregulation of the p53 pathway in Mf permitted a burst of cell proliferation pathways, which we surmise to be a step toward recovery. This step assists the adaptive immune response with upregulation of memory activated CD4+ T cells and follicular helper T cells.

The studies in this dissertation explore a multitude of avenues in a complicated multiphase immune response. As one might expect, the results are limited by a number of drawbacks, which should be acknowledged. Computational investigations of the type presented here involve numerous statistical analyses, whose impact is compromised by the small sample sizes for experiments with both hosts. Clearly, this issue reduces the statistical significance of some of the results and makes interpretations prone to errors. There is not all that much that can be done to ameliorate the situation. NHP experiments are inherently expensive. Furthermore, due to laws regarding animal welfare, only relatively small numbers of samples, even of blood, may be taken from NHPs within a certain time period. This restriction prevents a more comprehensive observation of important aspects of a *Plasmodium* infection in the same experiment, as some of these changes appear to happen within hours. Furthermore, the Mm monkeys have to be euthanized once their condition worsens if treatment options by the attending vets are not effective and the well-being of the animal with continued infection cannot be assured. Unless the experimental design includes necropsies for access to tissues, as it was the case in our MaHPIC studies, all reasonable treatment options are provided to aid the recovery on an animal. These limitations lead to samples that are scarce, isolated snapshots rather than comprehensive time series. In order to maximize the information within the limited numbers of time points, the chosen time points belonged to key phases of the infection. However, the sparsity of data is an ongoing problem that

can only be addressed with different, complementary experiments and analyses. For example, to answer specific biological questions, given what we know now from the previous studies, the timing of sample acquisition, *e.g.*, for targeted gene expression or metabolomics studies, could be optimized toward the most informative datasets.

Another major challenge pertains to reference genome annotations and gene homology. Even though the use of sequence-based methods is quite efficient and reliable, this method is not indisputably conclusive for determining homology. Only experiments defining functions of individual genes would be able to accomplish a sufficient degree of certainty. At the same time, the proposed sequence-based method to determine the similarity between two homologs has its drawbacks as well. As a consequence, the presented results should not necessarily be considered definite but rather as a means for developing novel hypotheses regarding different functionalities of homologous genes, which then would require specifically designed experiments for verification.

Establishing the true homology of genes for such comparative analyses as ours is instrumental for understanding the true effects of selective pressure. For example, the two hosts discussed in this study are evolutionarily close, but one of the natural hosts, *M. fascicularis*, has evolved with *P. knowlesi* malaria for a long time. It is therefore to be expected that selective pressure has had an effect on its genome. Since sequence similarity does not confirm homology *per se*, specifically designed homology studies are required to validate the preliminary results and advance the field.

Issues of curation are further exaggerated by the lack of macaque-specific databases. Macaques are much less studied than humans and rodents. As a consequence, many of the datasets used for our analyses, including GSEA, cell type markers, or modular repertoire data, come from

human-derived datasets. The evolutionary closeness of the NHP species with humans is supportive of the methods used, but a critical cross-evaluation of the results is required for verification. Another analysis facing this shortcoming is the cell type deconvolution analysis. This analysis relies entirely on cell type marker genes, which were identified in humans. It is reasonable to expect the marker genes to be similar for NHP hosts, because these genes are generally conserved; nonetheless, there is no certainty until the close similarity is independently verified.

The chosen dynamic models also have shortcomings. First, like all models, the models we used are obviously simplifications and abstractions of the complex systems found in nature. Second, the models were designed for humans rather than macaques, and it is a matter of trust to assume that the models are applicable nevertheless. This issue is arguably not as severe though, as the overall structure of such fundamental metabolic systems as purine and tryptophan metabolism is well preserved, and tangential biochemical details in some sense "cancel out" in our comparative analyses of "before-after" differences in responses.

Even though the proposed transcriptomics-aided metabolic modeling provides a robust methodology for the prediction of changes during infection, several improvements and validations would further its utility. As an extension to adapting metabolic models for various tissues, the models could be updated for specific species and even individual subjects, by replacing average parameter values with species-specific or personalized values. This step would be directly in line with concepts of personalized medicine and drug target discovery[170,291]. Before this step is taken, however, dedicated studies to validate its utility are warranted. These could be performed first in controlled *in vitro* and then in *ex vivo* and *in situ* environments. Targeted metabolomics could be used to validate some of the outputs.

Another drawback of this methodology is the direct use of changes in gene expression for adjusting enzymatic activity. The correlation between the expression of a specific gene and the corresponding translated and post-translationally modified protein and its activity is still a matter of discussion[292]. However, it is not to be expected that *increased* gene expression will lead to *reduced* protein availability, or *vice versa*, and as long as there is a moderate positive correlation, the results should hold at least qualitatively; however, there is no guarantee, and only experimental validation can confirm the findings derived from these models. Thus, although there are clear drawbacks and limitations on the computational front, most of these could likely be systematically removed or at least ameliorated with specific improvements to the underlying data.

The results obtained from this study reveal several crucial differences in the two host responses that should be explored further. First, early detection of the pathogen and the role of the *SICAvar* family of surface antigens needs further clarification. The diverse 136-member *SICAvar* gene family has been shown to be responsible for the chronicity of infection in Mm, and the correlations of this family with immune related genes in both hosts should be studied further for a better understanding of host-pathogen interactions.

Second, on the host side, further details of PRR signaling could help us understand specific molecular patterns that are detected by the two hosts. These patterns are followed by differences in protein kinases and subsequently result in differences in inflammation; a step-wise comparative analysis could pinpoint specific triggers and regulations.

The most important step to address initial pathogen detection would be a denser longitudinal dataset especially focusing on the liver phase of infection. The regulatory challenge regarding limitations to the amount of blood collected from NHPs could be partially circumvented by using

a larger pool of monkeys, possibly sampled at staggered time points. Without such denser samples, one must rely on transcriptomics extrapolation and modeling techniques which, if reliable, would be beneficial in bridging current gaps in the data.

A major finding of this work is the importance of the balance between inflammation and cell proliferation. Ca^{2+} homeostasis has long been studied in regulating inflammation. However, a detailed comparative analysis would be needed to explain the small but crucial differences in the molecular mechanisms governing the diverging host responses to malaria. This study provides two key avenues toward testing the involvement of Ca^{2+} ions in inflammation. Activation of glutamate GPCR signaling pathways could be queried by quantification of amino acid residues that are products of phosphorylation. Additionally, discovery of novel downstream targets would not only ascertain the importance of glutamate GPCR signaling but also shed light on its specific effects on inflammation. Similarly, controlled experiments elucidating the inflammasome assembly process and the role of Ca^{2+} ions could help with the detection of key regulators that balance inflammation and cell proliferation.

Along similar lines, the p53 pathway is widely studied for cancer research and considered a major regulator for tumor growth. Detailed comparative analyses in similar species could highlight switches that control inflammation during the log phase of infection and promote cell proliferation thereafter. Indeed, the importance of the p53 pathway in several other diseases warrants follow-up studies dedicated to understanding the molecular mechanisms of this pathway associated specifically with malaria. Detailed transcriptomics and metabolomics studies could help answer some of these questions, and additional computational analyses and -omics data could furthermore help establish exact molecular functioning of these pathways. For example, a systemic modeling approach could be able to identify and characterize identify feedback loops and molecular or

functional switches that regulate Ca^{2+} concentrations and therefore the signaling role of Ca^{2+}. Single-cell analysis could help confirm involvement of specific cell populations in these responses.

Perhaps the most fundamental findings of this study are the innate differences in the two hosts that could have direct implications for their immune responses. Chief among them are differences in neutrophils and CD4 T cells. Some of these observations have been documented generically in the literature[254], but follow-up studies in the context of malaria, showing their implications during the immune response via comparative analyses, could be very informative. As a case in point, the exact role of neutrophils in the immune response is still under debate.

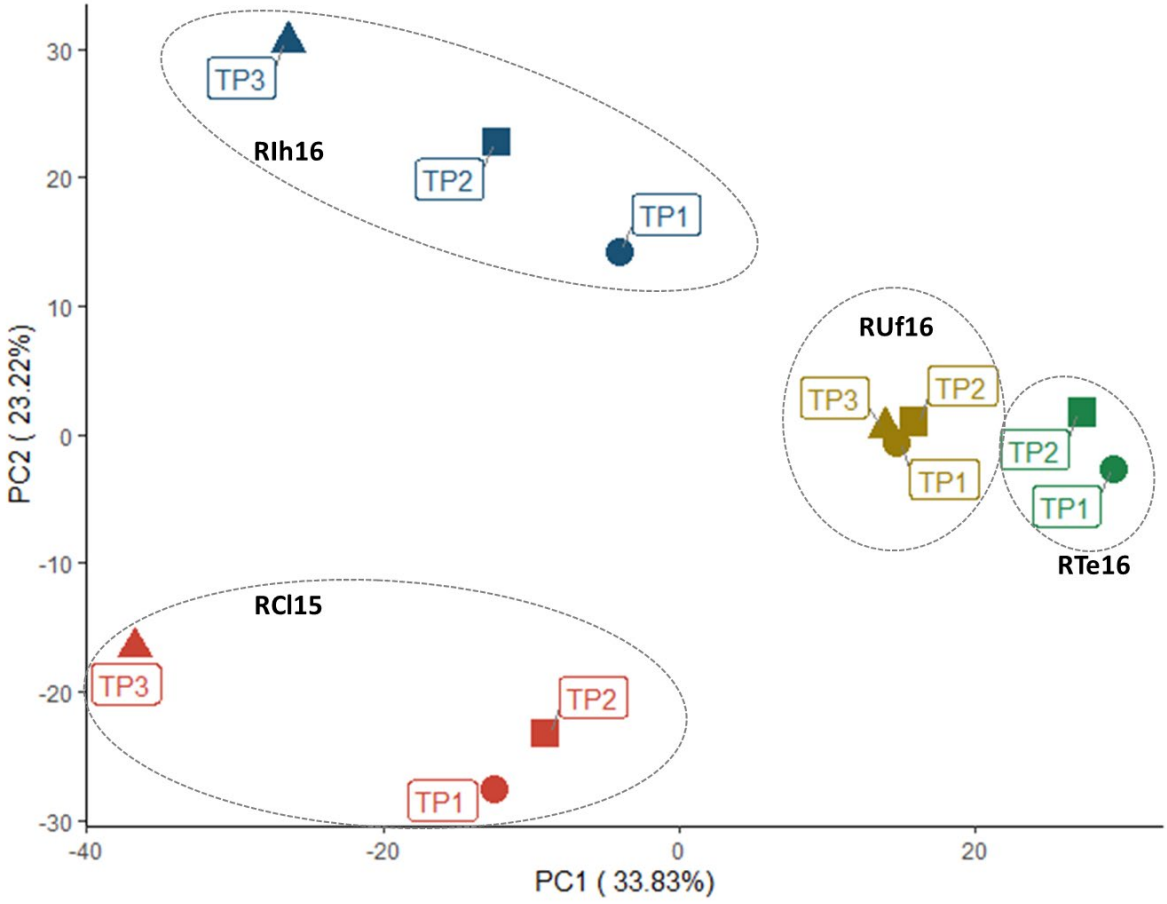
This uncertainty highlights an additional, more general aspect to be considered for designing NHP studies for understanding human infections. NHP models are widely used for human diseases and they have been quite successful. However, there is no general agreement of their specific value, and it is being argued that studies should rather been done directly on humans. Countering this argument, it is not ethically feasible to execute many studies that are allowable with NHPs (in the extreme case, necropsies), and the translation of key findings of experimental and computational analyses on NHPs to human hosts therefore has an unmatched potential for the identification of novel drug targets and the improvement of medical treatment. *Macaca mulatta* and *Macaca fascicularis* are the most widely used NHPs. Still, comparative studies between these monkeys and with humans are still sparse. Dedicated efforts towards understanding the subtle differences between these model organisms have the potential of significant improvements of our understanding of malaria.

In conclusion, the studies presented in this dissertation have enhanced our current insights into the immune response of NHPs to malaria. One of the studies focused on changes in purine

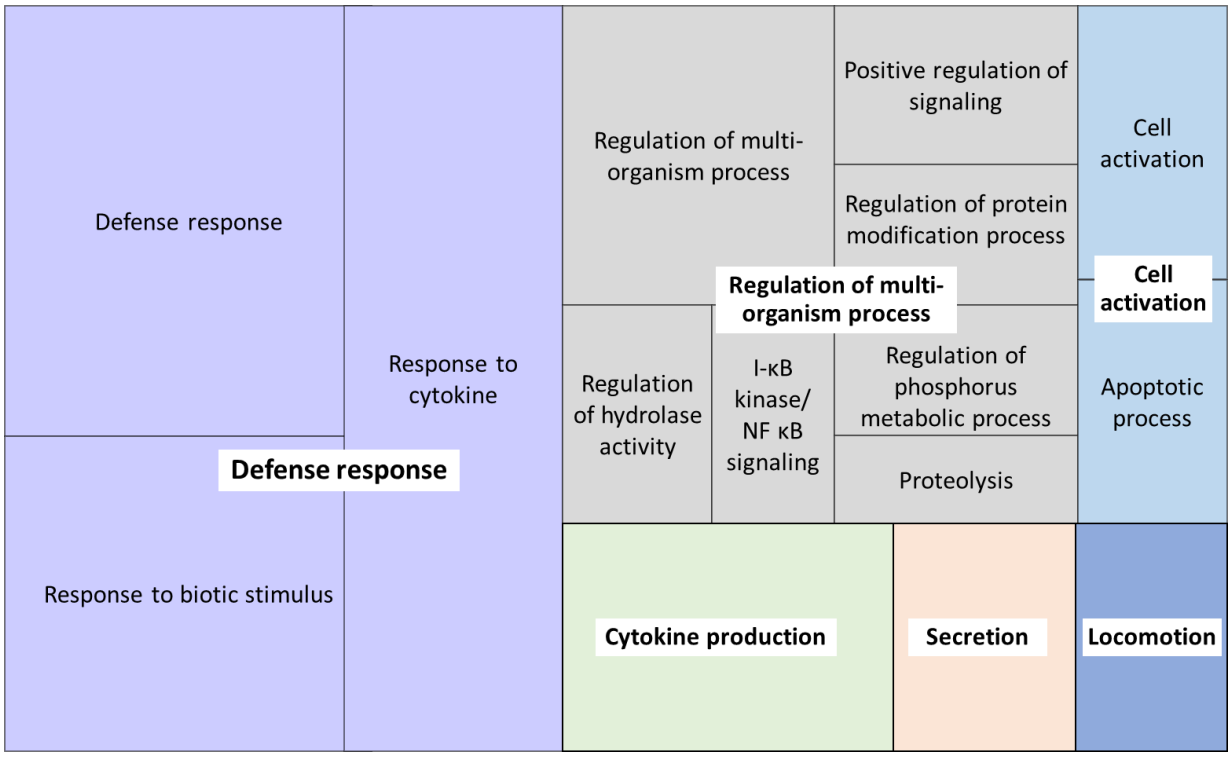
metabolism, while the others addressed a comparison of two host species exhibiting specific differences in molecular responses that result in resilience for *M. fascicularis* but severe infection and death in *M. mulatta*, if not treated. The wealth of data analyzed here is hoped to guide the development of hypothesis driven studies in continuation of our current findings. Furthermore, this work lays the foundation for future comparative analyses that could help answer outstanding questions related to pathogen detection, inflammation, innate and adaptive immune responses and the roles of individual regulators and cell types. This foundation is hoped to aid the furtherance of *P. knowlesi* epidemiology, vaccine development, drug target discovery, diagnosis, and control strategies that might ultimately control and eliminate the threat of malaria.

APPENDIX A. SUPPLEMENTS FOR CHAPTER 2

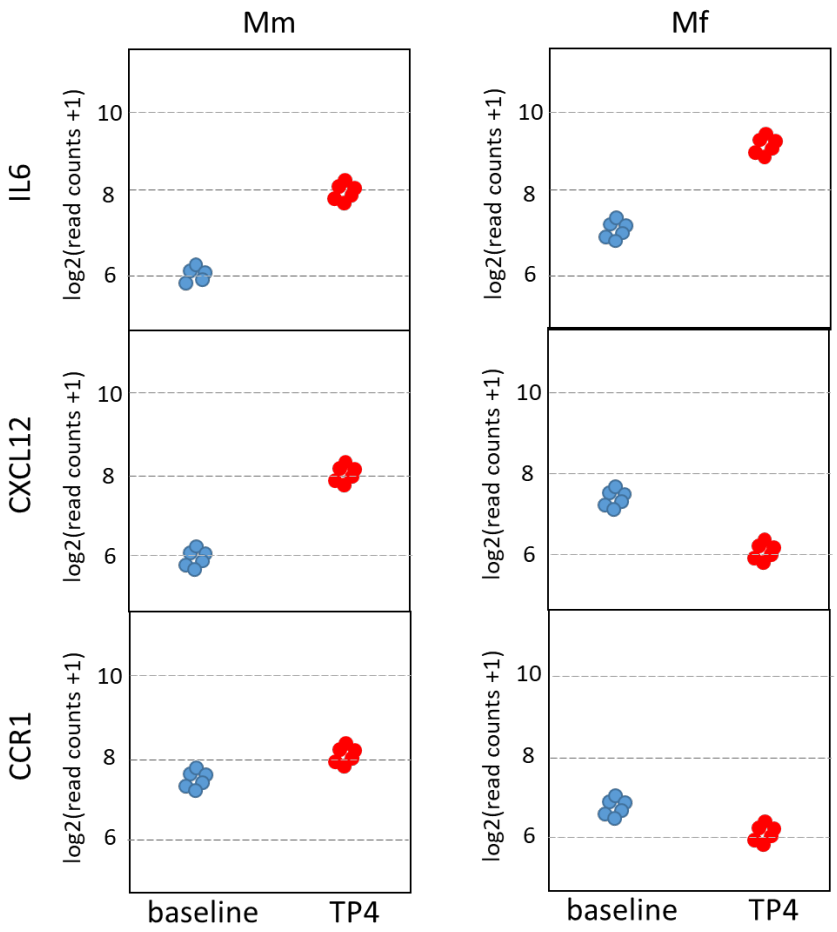
A.1 Supplementary Figures



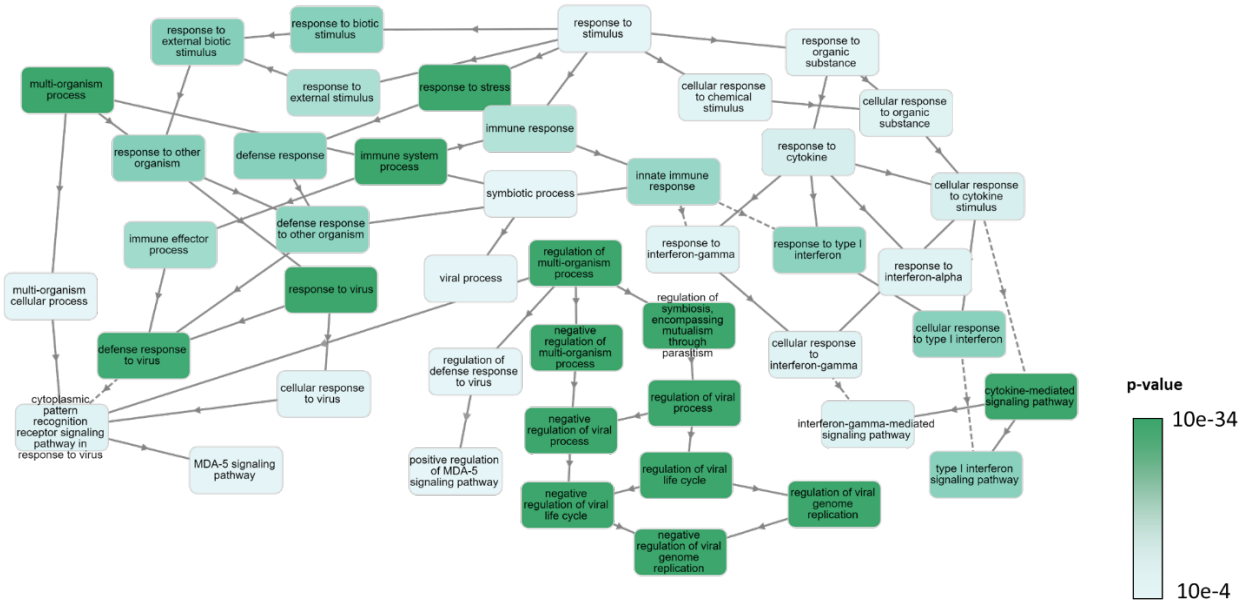
Appendix Figure A.1.1- Principal component analysis (PCA) of *M. mulatta* at three time points. Whole blood samples were taken for transcriptomics analysis of Mm before (TP1 and TP2) and soon after infection (TP3). Any potential differences in TP3 transcriptomics with respect to baseline are dominated by subject-specific variance, which is evident from the clustering of individual subjects (ellipses). RCL15, RTe16, RUf16 and RIh16 represent codes of the macaques from which the blood samples came for this analysis. TP3 data are missing for RTe16.



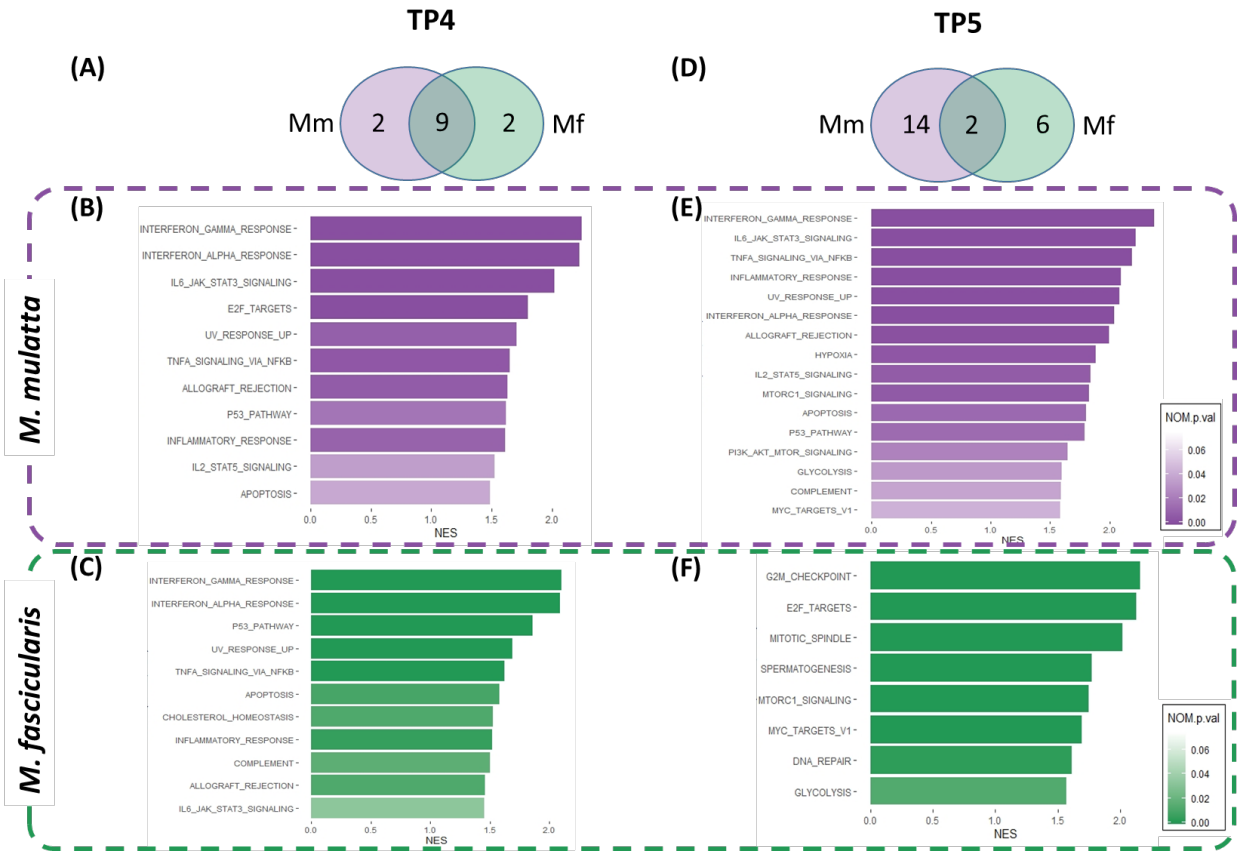
Appendix Figure A.1.2- REVIGO tree map showing top GO annotations from PC2 of the PCA in Figure 2.2. The size of each rectangle is proportional to the –log10(*p*-value) with bigger rectangles representing higher significance. A generic defense response, response to biotic stimuli and response to cytokines are the most significant GO annotations. This visualization summarizes the large set of enriched GO annotations and removes redundant GO annotations, grouping them based on the GO term hierarchy.



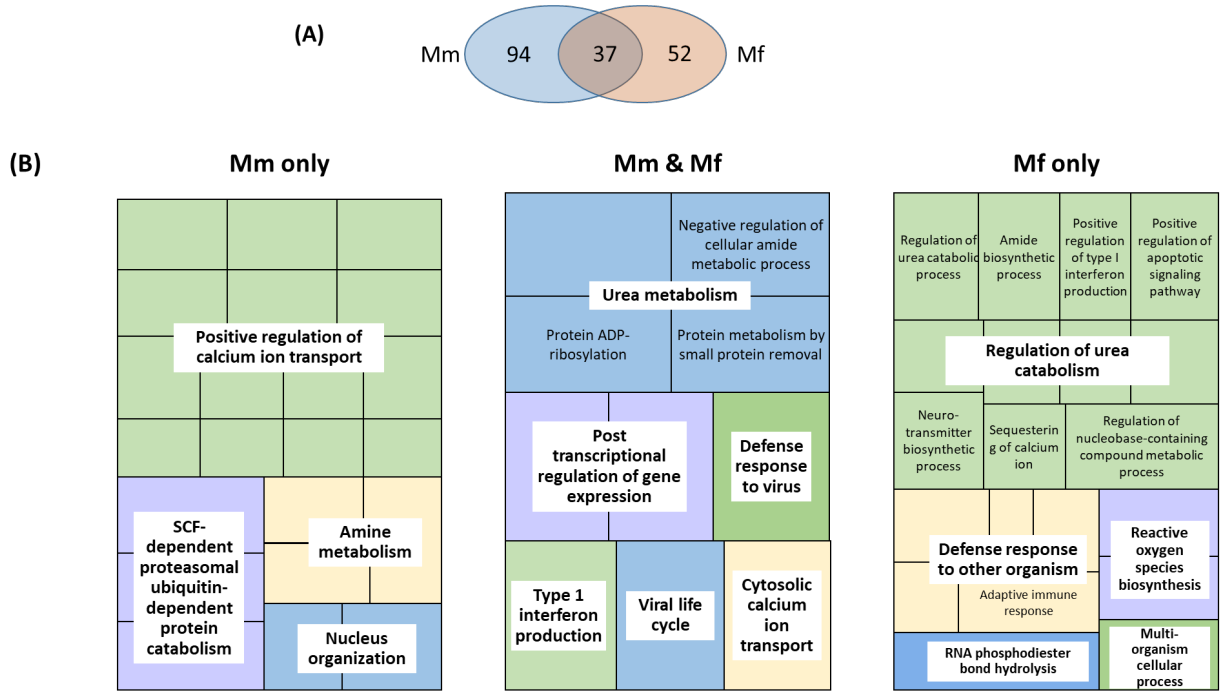
Appendix Figure A.1.3- Example plot explaining DEG *vs.* DRG (Note: These are not actual values; they are presented to highlight the importance of DRGs). IL6 is differentially expressed in Mm and Mf but is not differentially responding (values at TP4 higher than baseline in both species). CXCL12 is differentially expressed in Mm and Mf, and it is also differentially responding. Expression of CCR1 is not significantly different in Mm and Mf but the species are differentially responding, as the (insignificant) changes in each species are in opposite directions.



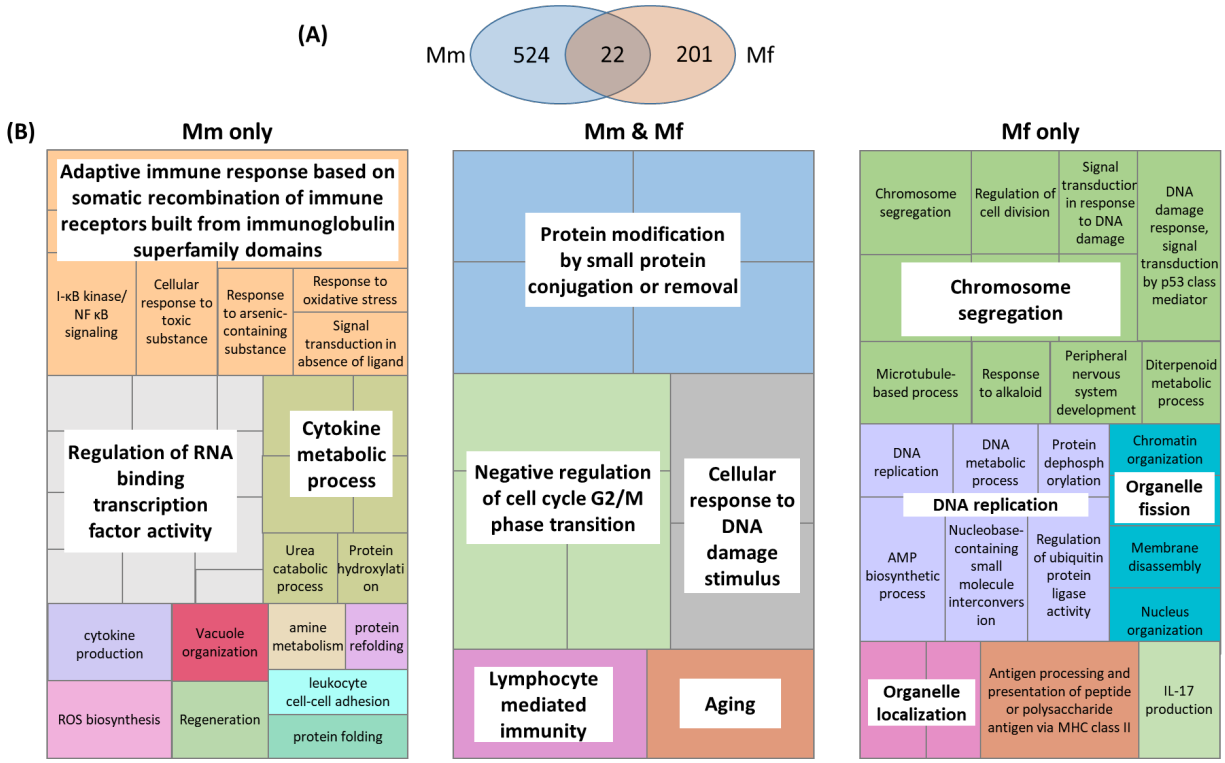
Appendix Figure A.1.4- Hierarchical graph for GO annotations significantly enriched in Mf at TP3 (p <0.001). GO gene-sets offer the advantage of inherent hierarchy, which adds structure and groupings to enriched pathways. The highly enriched gene-sets (p < 0.0001) are colored, with darker colors reflecting higher enrichment. The connecting directional arrows show the hierarchy. At TP3, Mf shows a strong defense response directed against viruses or other invading organisms. The graph was created with the GO-Net web application (https://tools.dice-database.org/GOnet/).



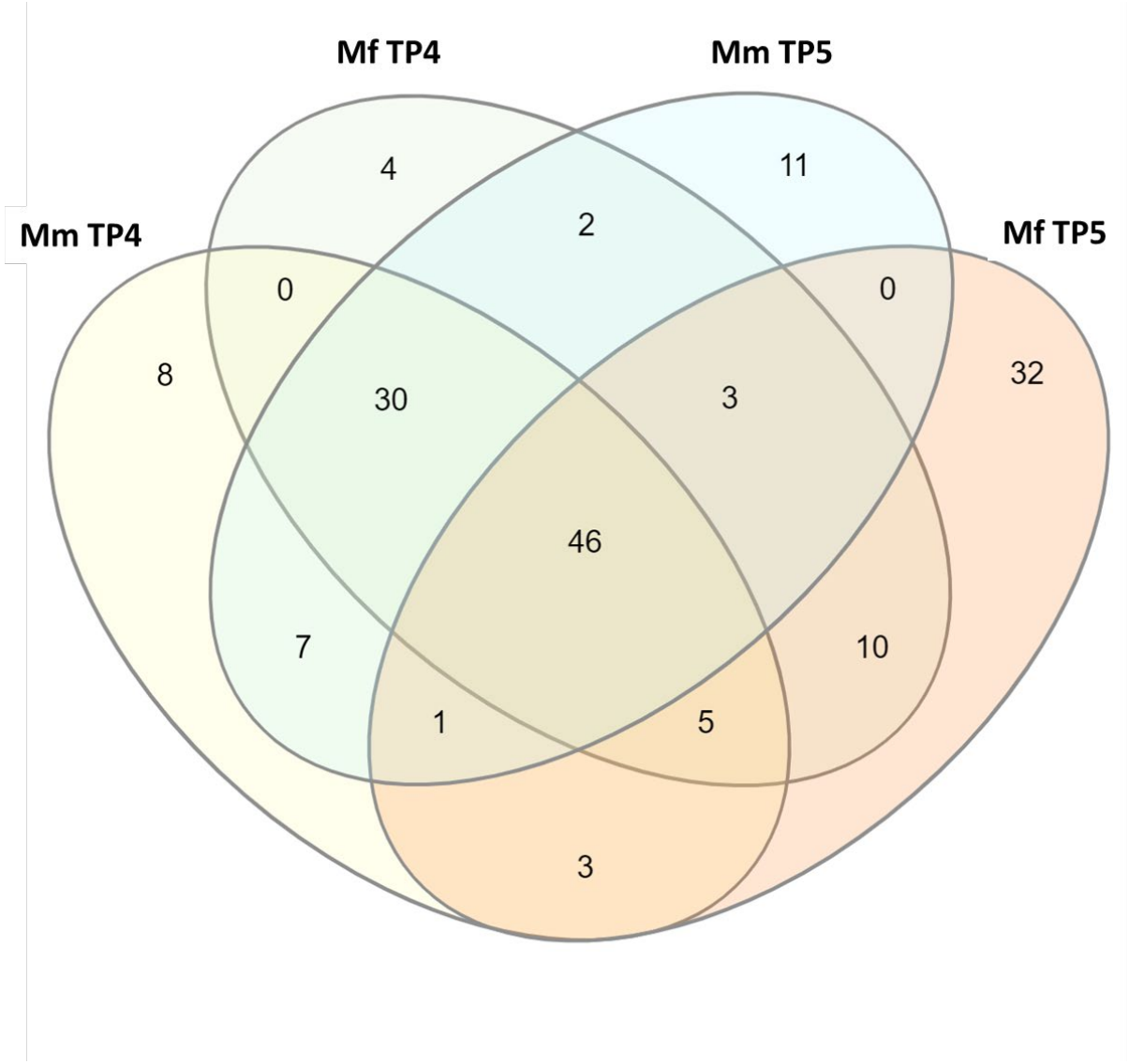
Appendix Figure A.1.5- Results of GSEA for TP4. (A): Venn Diagram showing shared Hallmark pathways between Mm and Mf at TP4 (FDR< 0.25, Appendix Table A.2.1). (B, C): Bar plots showing significantly enriched GO annotations from GSEA of Mm(B) and Mf(C) at TP4. (D): Venn Diagram showing shared Hallmark pathways between Mm and Mf at TP5 (FDR< 0.25). (E, F): Bar plots showing the significantly enriched GO annotations from GSEA of Mm(E) and Mf(F) at TP5. X axis: Normalized Enrichment Scores (NES) colored, colored by significance.



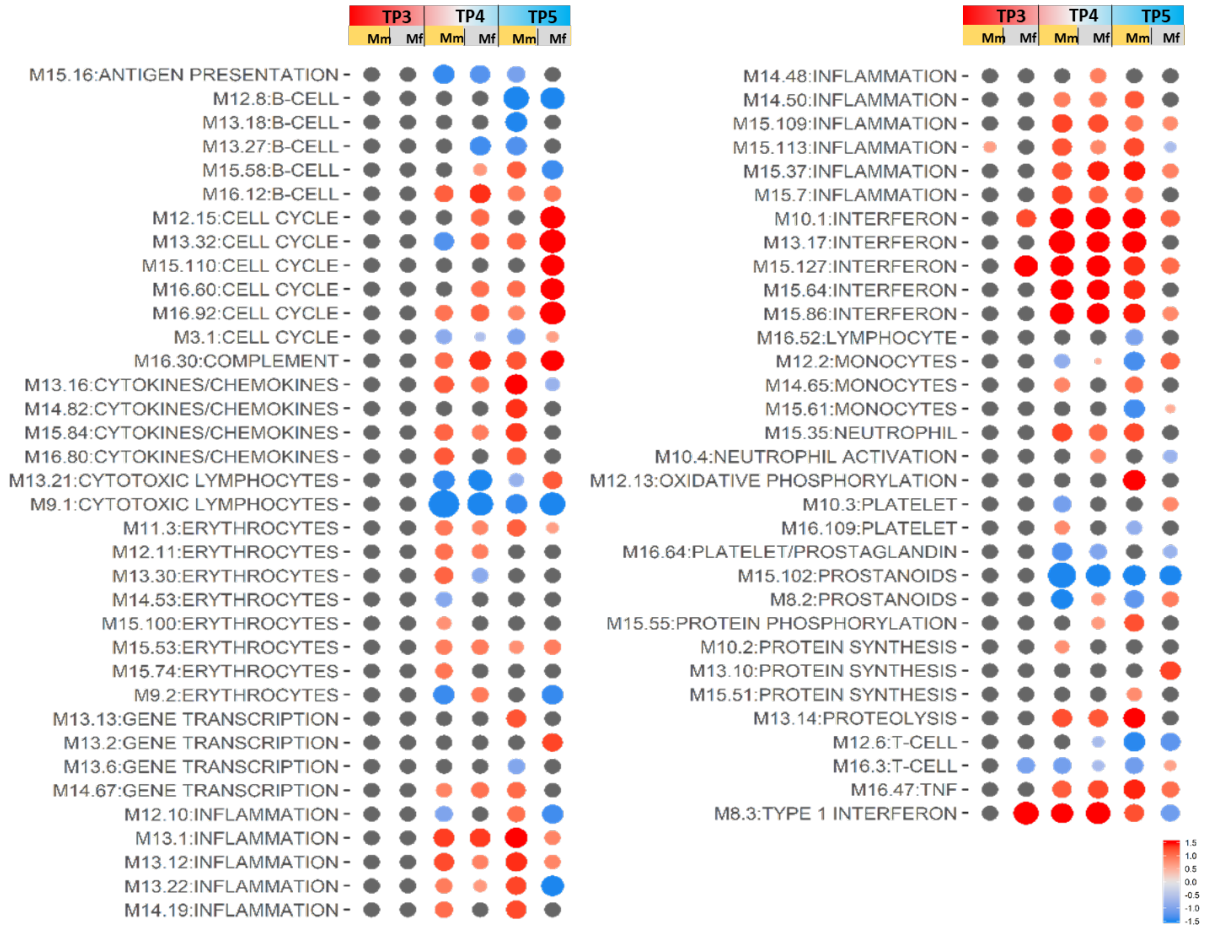
Appendix Figure A.1.6- Most highly enriched pathways at TP4. (A): The Venn diagram shows numbers of common and different most highly enriched GO Biological Processes in Mm and Mf at TP4. (B): Reduced tree maps of enriched GO processes for each group (Mm only, Mm and Mf, Mf only). The GO annotations were summarized using REVIGO to remove redundant annotations for each set.



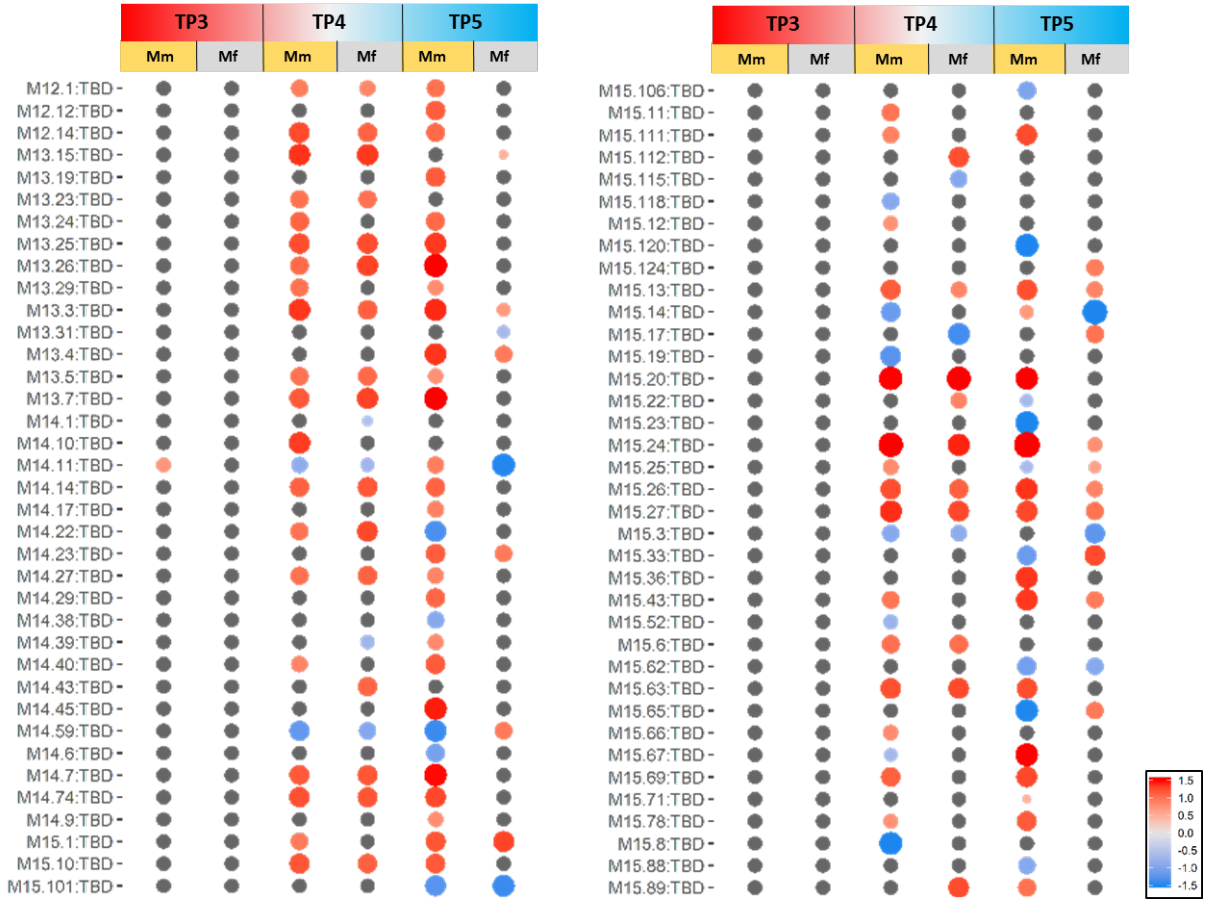
Appendix Figure A.1.7- Most highly enriched pathways at TP5. (A): The Venn diagram shows the numbers of common and different most highly enriched GO Biological Processes of Mm and Mf at TP5. (B): Reduced tree maps of enriched GO processes for each group (Mm only, Mm and Mf, Mf only). The GO annotations were simplified using REVIGO to remove redundant annotations.



Appendix Figure A.1.8- Venn diagram top 100 GO annotations at TP4 and TP5 for Mm and Mf. The intersections of these GO-annotated gene-sets visualize the similarity among the different groups. The majority (76) of pathways are common between Mm at TP4, Mm at TP5 and Mf at TP4. By contrast, Mf at TP5 shows the highest mutually exclusive gene-sets (32), indicating a stark deviation from TP4.

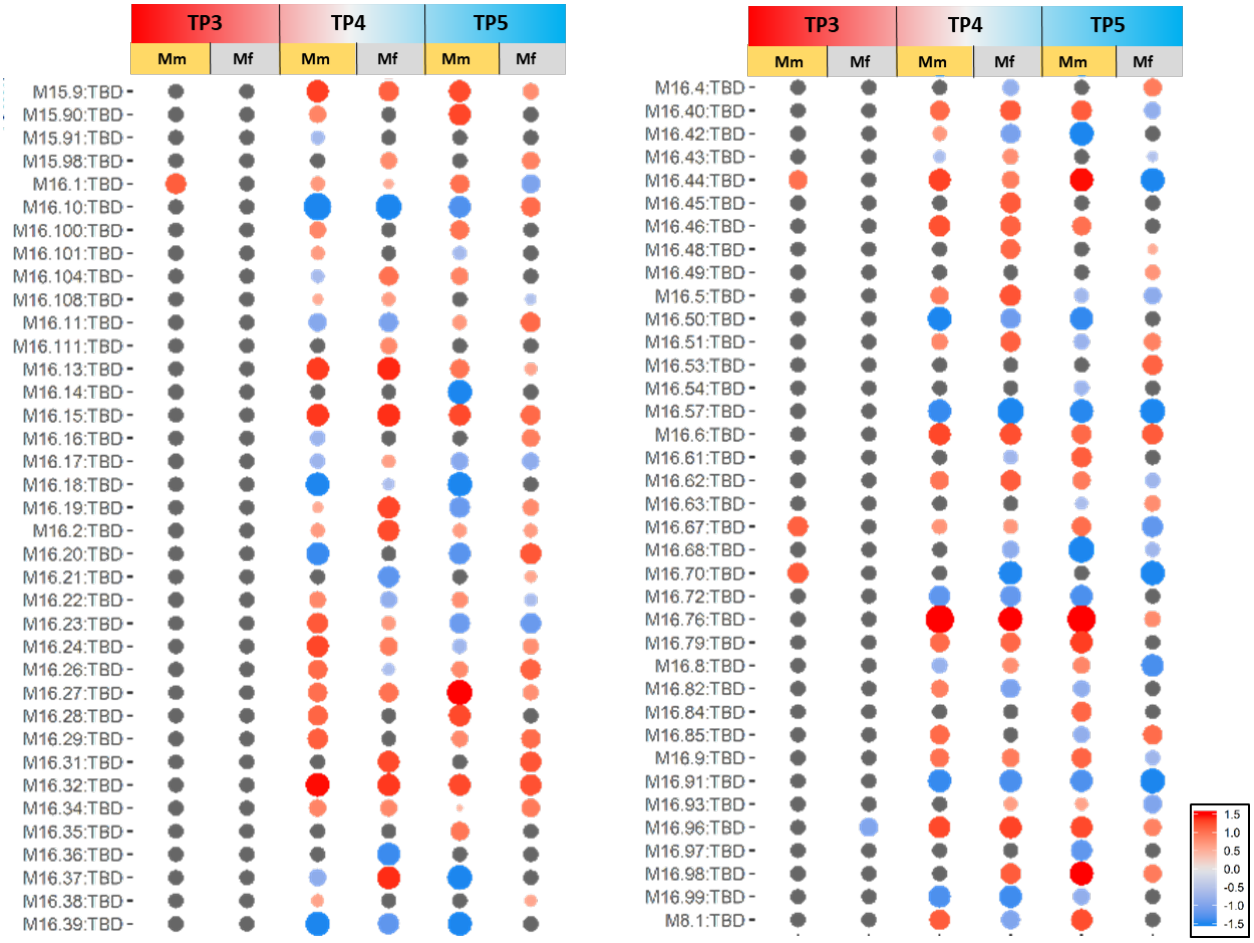


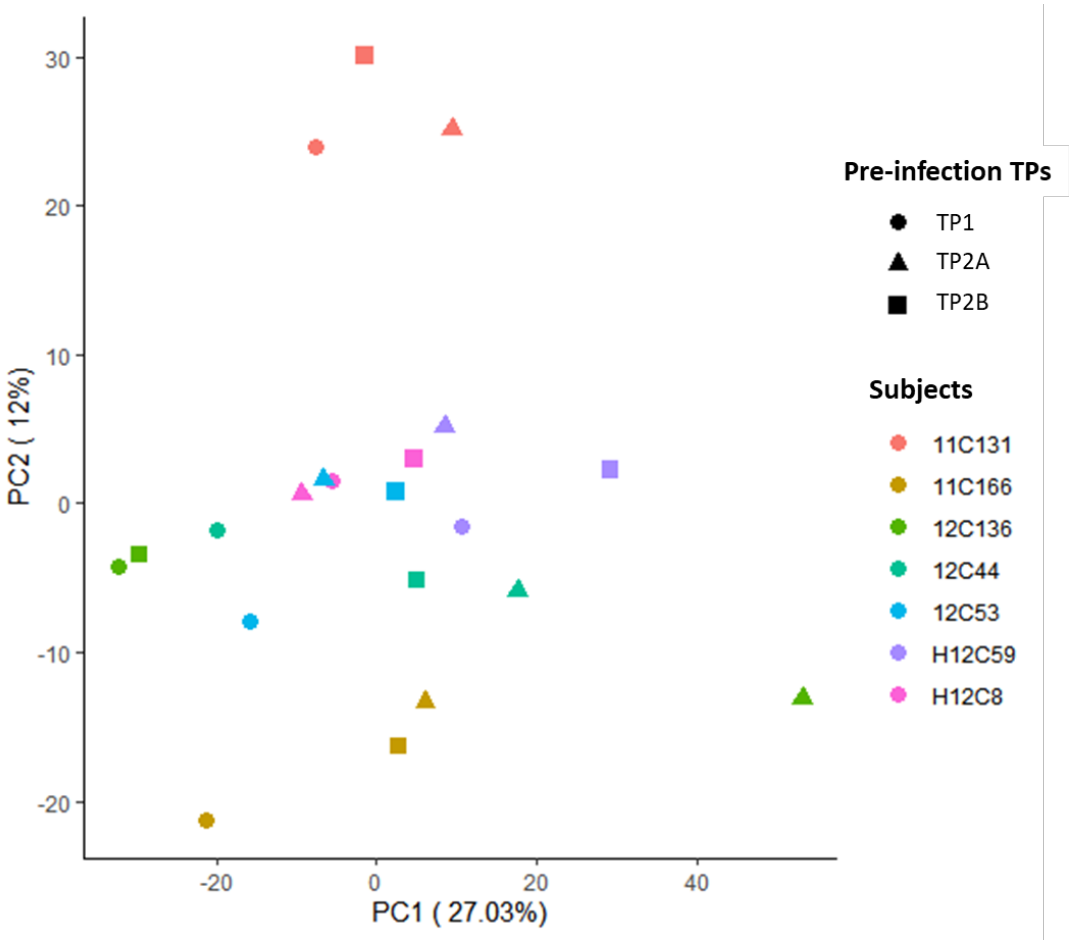
Appendix Figure A.1.9- Comparison of changes in main transcriptome modules of Mm and Mf at TP3, TP4 and TP5. The modular transcriptome analysis presented in the form of a heat map. The modules can be associated with a functional annotation, with details presented in Appendix Table A.2.5. The image consists of all modules that are not as strongly enriched as those in Figure 2.7.



Appendix Figure A.1.10- Heat map of modular transcriptome analysis comparing Mm and Mf at TP3, TP4 and TP5. A functional annotation could not be associated with these modules but additional information regarding association with diseases, as well as enrichment in KEGG and GO can be found in Appendix Table A.2.6.

Appendix Figure A.1.10 (continued)





Appendix Figure A.1.11- PCA of pre-infection samples for Mf. Samples from individual subjects cluster together. TP2A and TP2B do not show any noticeable difference due to failed inoculation.

A.2 Supplementary Tables

Appendix Table A.2.1- GSEA using Hallmark gene sets of Mm and Mf at TP3, TP4 and TP5. Gene sets in orange-shaded boxes have positive enrichment scores; *i.e.*, they were upregulated, while teal-shaded have negative enrichment scores.

M. mulatta		
NAME	**NES**	**FDR q-val**
TP4		
INTERFERON_GAMMA_RESPONSE	2.25	0.00
INTERFERON_ALPHA_RESPONSE	2.22	0.00
IL6_JAK_STAT3_SIGNALING	2.02	0.00
E2F_TARGETS	1.80	0.01
UV_RESPONSE_UP	1.71	0.04
TNFA_SIGNALING_VIA_NFKB	1.65	0.07
ALLOGRAFT_REJECTION	1.63	0.07
P53_PATHWAY	1.62	0.07
INFLAMMATORY_RESPONSE	1.61	0.07
IL2_STAT5_SIGNALING	1.53	0.14
APOPTOSIS	1.49	0.17
TP5		
INTERFERON_GAMMA_RESPONSE	2.37	0.00
IL6_JAK_STAT3_SIGNALING	2.22	0.00
TNFA_SIGNALING_VIA_NFKB	2.19	0.00
INFLAMMATORY_RESPONSE	2.09	0.00
UV_RESPONSE_UP	2.08	0.00
INTERFERON_ALPHA_RESPONSE	2.04	0.00
ALLOGRAFT_REJECTION	2.00	0.01
HYPOXIA	1.88	0.02
IL2_STAT5_SIGNALING	1.84	0.02
MTORC1_SIGNALING	1.83	0.02
APOPTOSIS	1.80	0.03
P53_PATHWAY	1.79	0.03
PI3K_AKT_MTOR_SIGNALING	1.65	0.08
GLYCOLYSIS	1.60	0.10
COMPLEMENT	1.59	0.10
MYC_TARGETS_V1	1.58	0.10

M.fascicularis		
NAME	**NES**	**FDR q-val**
TP3		
INTERFERON_GAMMA_RESPONSE	2.27	0.00
INTERFERON_ALPHA_RESPONSE	1.90	0.00
KRAS_SIGNALING_DN	1.40	0.11
TP4		
INTERFERON_GAMMA_RESPONSE	2.10	0.00
INTERFERON_ALPHA_RESPONSE	2.09	0.00
P53_PATHWAY	1.86	0.00
UV_RESPONSE_UP	1.69	0.01
TNFA_SIGNALING_VIA_NFKB	1.62	0.03
APOPTOSIS	1.58	0.05
CHOLESTEROL_HOMEOSTASIS	1.53	0.08
INFLAMMATORY_RESPONSE	1.52	0.08
COMPLEMENT	1.50	0.09
ALLOGRAFT_REJECTION	1.46	0.13
IL6_JAK_STAT3_SIGNALING	1.46	0.12
MTORC1_SIGNALING	1.41	0.17
ANGIOGENESIS	-1.49	0.12
NOTCH_SIGNALING	-1.54	0.12
UV_RESPONSE_DN	-1.58	0.14
MYOGENESIS	-1.66	0.17
TP5		
G2M_CHECKPOINT	2.16	0.00
E2F_TARGETS	2.13	0.00
MITOTIC_SPINDLE	2.02	0.00
SPERMATOGENESIS	1.77	0.00
MTORC1_SIGNALING	1.74	0.00
MYC_TARGETS_V1	1.69	0.01
DNA_REPAIR	1.61	0.02
GLYCOLYSIS	1.57	0.04

Appendix Table A.2.2- GSEA using hallmark gene sets for Mf *vs.* Mm at TP3, TP4, and TP5.

TP3		
NAME	**NES**	**FDR q-val**
INTERFERON_GAMMA_RESPONSE	3.39	0.00
INTERFERON_ALPHA_RESPONSE	3.37	0.00
KRAS_SIGNALING_DN	1.69	0.06
DNA_REPAIR	1.59	0.09
ESTROGEN_RESPONSE_EARLY	1.23	0.47
G2M_CHECKPOINT	0.98	0.90
ESTROGEN_RESPONSE_LATE	0.82	1.00
INFLAMMATORY_RESPONSE	0.77	1.00
ADIPOGENESIS	0.73	0.99
COMPLEMENT	0.68	0.96
UV_RESPONSE_UP	0.67	0.88
E2F_TARGETS	-0.72	0.83
CHOLESTEROL_HOMEOSTASIS	-0.84	0.72
P53_PATHWAY	-0.90	0.67
UV_RESPONSE_DN	-0.91	0.72
TNFA_SIGNALING_VIA_NFKB	-0.91	0.78
MTORC1_SIGNALING	-0.91	0.85
ALLOGRAFT_REJECTION	-0.94	0.88
XENOBIOTIC_METABOLISM	-1.14	0.54
MITOTIC_SPINDLE	-1.28	0.38
GLYCOLYSIS	-1.30	0.40
KRAS_SIGNALING_UP	-1.33	0.43
EPITHELIAL_MESENCHYMAL_TRANSITION	-1.34	0.51
HYPOXIA	-1.48	0.34
IL2_STAT5_SIGNALING	-1.70	0.10
TGF_BETA_SIGNALING	-1.71	0.20

TP4		
NAME	**NES**	**FDR q-val**
APICAL_JUNCTION	1.28	1.00
COMPLEMENT	1.22	0.86
KRAS_SIGNALING_UP	1.03	1.00
HYPOXIA	1.03	0.87
REACTIVE_OXYGEN_SPECIES_PATHWAY	0.95	0.85
COAGULATION	0.93	0.75
GLYCOLYSIS	0.93	0.64
INFLAMMATORY_RESPONSE	0.81	0.72
XENOBIOTIC_METABOLISM	-0.61	0.93
IL2_STAT5_SIGNALING	-0.69	0.98
EPITHELIAL_MESENCHYMAL_TRANSITION	-0.89	0.79
ADIPOGENESIS	-1.13	0.45
ALLOGRAFT_REJECTION	-1.17	0.48
INTERFERON_ALPHA_RESPONSE	-1.20	0.54
ESTROGEN_RESPONSE_LATE	-1.21	0.69
MYOGENESIS	-1.28	0.78
ESTROGEN_RESPONSE_EARLY	-1.38	1.00

TP5		
NAME	**NES**	**FDR q-val**
E2F_TARGETS	3.89	0.00
G2M_CHECKPOINT	3.56	0.00
MITOTIC_SPINDLE	2.32	0.00
COAGULATION	2.13	0.00
MYC_TARGETS_V2	2.03	0.01
SPERMATOGENESIS	1.98	0.01
MYC_TARGETS_V1	1.69	0.07
KRAS_SIGNALING_DN	1.51	0.17
EPITHELIAL_MESENCHYMAL_TRANSITION	1.51	0.15
ANGIOGENESIS	1.42	0.21
ESTROGEN_RESPONSE_LATE	1.42	0.20
APICAL_JUNCTION	1.36	0.23
OXIDATIVE_PHOSPHORYLATION	1.12	0.51
PROTEIN_SECRETION	1.12	0.49
GLYCOLYSIS	1.06	0.54
BILE_ACID_METABOLISM	1.06	0.51
APICAL_SURFACE	1.04	0.51
ANDROGEN_RESPONSE	0.83	0.84
WNT_BETA_CATENIN_SIGNALING	0.80	0.85
MYOGENESIS	0.79	0.82
COMPLEMENT	-0.96	0.68
MTORC1_SIGNALING	-1.12	0.45
PANCREAS_BETA_CELLS	-1.14	0.45
FATTY_ACID_METABOLISM	-1.22	0.37
PI3K_AKT_MTOR_SIGNALING	-1.27	0.32
P53_PATHWAY	-1.33	0.26
ESTROGEN_RESPONSE_EARLY	-1.36	0.25
APOPTOSIS	-1.37	0.26
UV_RESPONSE_UP	-1.38	0.26
KRAS_SIGNALING_UP	-1.59	0.09
IL6_JAK_STAT3_SIGNALING	-1.68	0.06
INFLAMMATORY_RESPONSE	-1.70	0.06
IL2_STAT5_SIGNALING	-1.71	0.06
HYPOXIA	-1.78	0.04
UNFOLDED_PROTEIN_RESPONSE	-1.83	0.03
DNA_REPAIR	-1.89	0.03
ALLOGRAFT_REJECTION	-1.96	0.02
INTERFERON_ALPHA_RESPONSE	-1.98	0.02
TNFA_SIGNALING_VIA_NFKB	-2.11	0.01
INTERFERON_GAMMA_RESPONSE	-2.32	0.00

Appendix Table A.2.3- Transcription factors for differentially expressed genes, according to iRegulon, with normalized enrichment scores (NES), for the two macaque species at different time points.

TP3

M. mulatta		*M. fascicularis*	
TF	**NES**	**TF**	**NES**
NFIL3	5.859	IRF7	27.983
BNC1	4.549	STAT2	16.764
CDX2	4.478	STAT1	15.805
ZNF274	4.376	IRF4	13.472
AR	4.317	IRF1	5.82
NFIA	4.102	GFI1	4.874

TP4

TF	**NES**	**TF**	**NES**
IRF9	12.404	IRF9	10.142
STAT1	11.231	STAT1	10.021
STAT2	10.671	STAT2	9.849
IRF1	6.6	IRF1	6.111
STAT3	3.949	RF1	3.274
BCL3	3.356	SPIC	3.083
NFATC4	3.193		

TP5

TF	**NES**	**TF**	**NES**
IRF9	11.52	E2F4	8.661
STAT2	9.178	TFDP1	5.329
STAT1	8.76	FOXM1	5.614
IRF1	6.051	SIN3A	4.336
IRF2	4.561	TFDP3	4.236
SPIC	4.119	MYBL2	3.099
BCL3	4.015		
NR2F1	3.615		
GRHL1	3.069		

Appendix Table A.2.4- Modular transcriptome. Highly enriched and functionally annotated modules, complementary to Figure 2.7.

Module ID	Module functional association title	Disease activity up	Disease activity down	Top literature lab annotations	Top matched pathway in KEGG	Top GO term Biological Processes
M10.1	Interferon	HIV, *Burkholderia*, SoJIA, RSV, Influenza, SLE, TB, Transplant		RIG-I/Host-Pathogen interactions/DEAD-box RNA Helicases	RIG-I-like receptor signaling pathway	Response to virus
M15.127	Interferon	HIV, SOJIA, RSV, Influenza, SLE, TB, Transplant		2',5'-oligoadenylate/ Orthomyxoviridae infections/ Double stranded RNA		Immune response
M8.3	Type-1 interferon	HIV, SoJIA, RSV, Influenza, SLE, TB, Transplant		Transcriptome/Transcriptome/Orthomyxoviridae Infections		Immune response
M15.113	Inflammation	Staph, *Burkholderia*, SoJIA, TB		Antirheumatic agents/ IL-1/ IL -1α	MAPK signaling pathway	Intracellular signaling cascade
M13.16	Cytokines/chemokines	Staph, *Burkholderia*, SoJIA, Influenza, TB		Cryoprotective agents/Nicotinic acids/Glycerol	Starch and sucrose metabolism	Glucan catabolic process
M13.1	Inflammation	Staph, *Burkholderia*, SoJIA		Mitogen-activated protein Kinase 1/Mitogen-activated protein kinase 3/ERK1	Acute myeloid leukemia	Regulation of cell proliferation
M13.17	Interferon	HIV, Burkholderia, SOJIA, RSV, Influenza, SLE, TB		STAT2 Transcription Factor/ NLR proteins/ Inflammatory bowel diseases		Positive regulation of I-kB kinase/NF-kB cascade

M15.64	Interferon	HIV, *Burkholderia*, SoJIA, Influenza, SLE, TB		Tripartite motif proteins/ Systemic lupus erythematosus/Thymine		Immune response
M15.86	Interferon	HIV, *Burkholderia*, SoJIA, RSV, Influenza, SLE, TB		Metallothionein/Cadmium/Cadmium		
M13.27	B-cell	MS	Kawasaki, TB, SoJIA, Pregnancy, Staph, *Burkholderia*	Lymphocyte specific protein tyrosine kinase p56(lck)/ZAP-70 protein-tyrosine kinase/LCK	Primary immunodeficiency	Positive regulation of immune system process
M13.30	Erythrocytes	RSV, Transplant, Melanoma	HIV	Biotransformation/Biotransformation/Gases		Oxygen transport
M13.32	Cell cycle	HIV, RSV, SLE, Kawasaki		Aurora kinase/Cyclin A2/Cyclin-dependent kinase inhibitor p21	Cell cycle	Mitotic cell cycle
M12.15	Cell cycle	HIV, SLE, Kawasaki		Topoisomerase inhibitors/ADP/Mitoxantrone		Mitosis
M15.110	Cell cycle	MS, HIV, RSV, SLE		Polo-like kinase/Protein sorting signals/Nuclear localization signals	Oocyte meiosis	Cell cycle
M16.60	Cell cycle	HIV, RSV		Cyclin B1/Cyclin B/Maturation-promoting factor	Oocyte meiosis	Mitosis
M16.92	Cell cycle	HIV, RSV, SLE		Aurora kinases/ Nucleic acid synthesis inhibitors	DNA replication	DNA metabolic process

Appendix Table A.2.5- Modular transcriptome. All functionally annotated modules that are complementary to those in Appendix Figure A.1.9. The complete table can be found at: https://github.com/LBSA-VoitLab/Mm_Mf_analysis/blob/master/Tables/table_S5.txt. A representative sample of the table is as follows:

Module ID	Module functional association title	Disease Activity Up	Disease Activity Down	Top Literature Lab Annotations	Top matched pathway in KEGG	Top GOTERM Biological Processes
M15.16	Antigen presentation		B-Cell deficiency, SoJIA, Staph, *Burkholderia*	Dendritic Cells/Antigen-presenting cells		positive regulation of peptidase activity
M12.8	B-cell		Influenza, HIV, B-Cell Deficiency, TB, SoJIA, Pregnancy, Staph, *Burkholderia*	Antibodies, Monoclonal, Murine-Derived/Antigens, CD20/Antibodies	B cell receptor signaling pathway	B cell activation
M13.18	B-cell	MS	Pregnancy, Staph, *Burkholderia*	Environmental Pollutants/Toxic Actions/Lung Neoplasms	Spliceosome	RNA processing
M13.27	B-cell	MS	Kawasaki, TB, SoJIA, Pregnancy, Staph, *Burkholderia*	Lymphocyte Specific Protein Tyrosine Kinase p56(lck)/ZAP-70 Protein-Tyrosine Kinase/LCK	Primary immunodeficiency	positive regulation of immune system process
M15.58	B-cell			Sulfur compounds/Gene expression regulation, Enzymologic/Biomarkers, Tumor		molting cycle process
M16.12	B-cell			Genome/Mutation/Genome Components		protein localization
M12.15	Cell cycle	HIV, SLE, Kawasaki		Topoisomerase Inhibitors/ADP/Mitoxantrone		mitosis

Appendix Table A.2.6- Modular transcriptome. All modules that do not have a functional association in Appendix Figure A.1.10. The complete table can be found at: https://github.com/LBSA-VoitLab/Mm_Mf_analysis/blob/master/Tables/table_S6.txt. A representative sample of the table is as follows:

Module ID	Module functional association title	Disease Activity Up	Disease Activity Down	Top Literature Lab Annotations	Top matched pathway in KEGG	Top GOTERM Biological Processes
M12.1	TBD	MS	Kawasaki, B-Cell Deficiency, Pregnancy, Staph, *Burkholderia*	Mitochondria Apoptotic/Mitochondria/Cell Respiration		translation
M12.12	TBD	MS, SoJIA, JDM		Peroxides/Oxidative Stress/Superoxide Dismutase	Proteasome	negative regulation of macromolecule metabolic process
M12.14	TBD	SLE	RSV	Intestinal Neoplasms/Colonic Diseases/Colorectal Neoplasms		
M13.15	TBD	Staph		Protein Domains/Protein Structural Elements/Protein Interaction Domains and Motifs	Endocytosis	intracellular signaling cascade
M13.19	TBD	SOJIA	RSV, Transplant	Protein Domains/Protein Structural Elements/Phosphoinositide	Prostate cancer	regulation of transcription

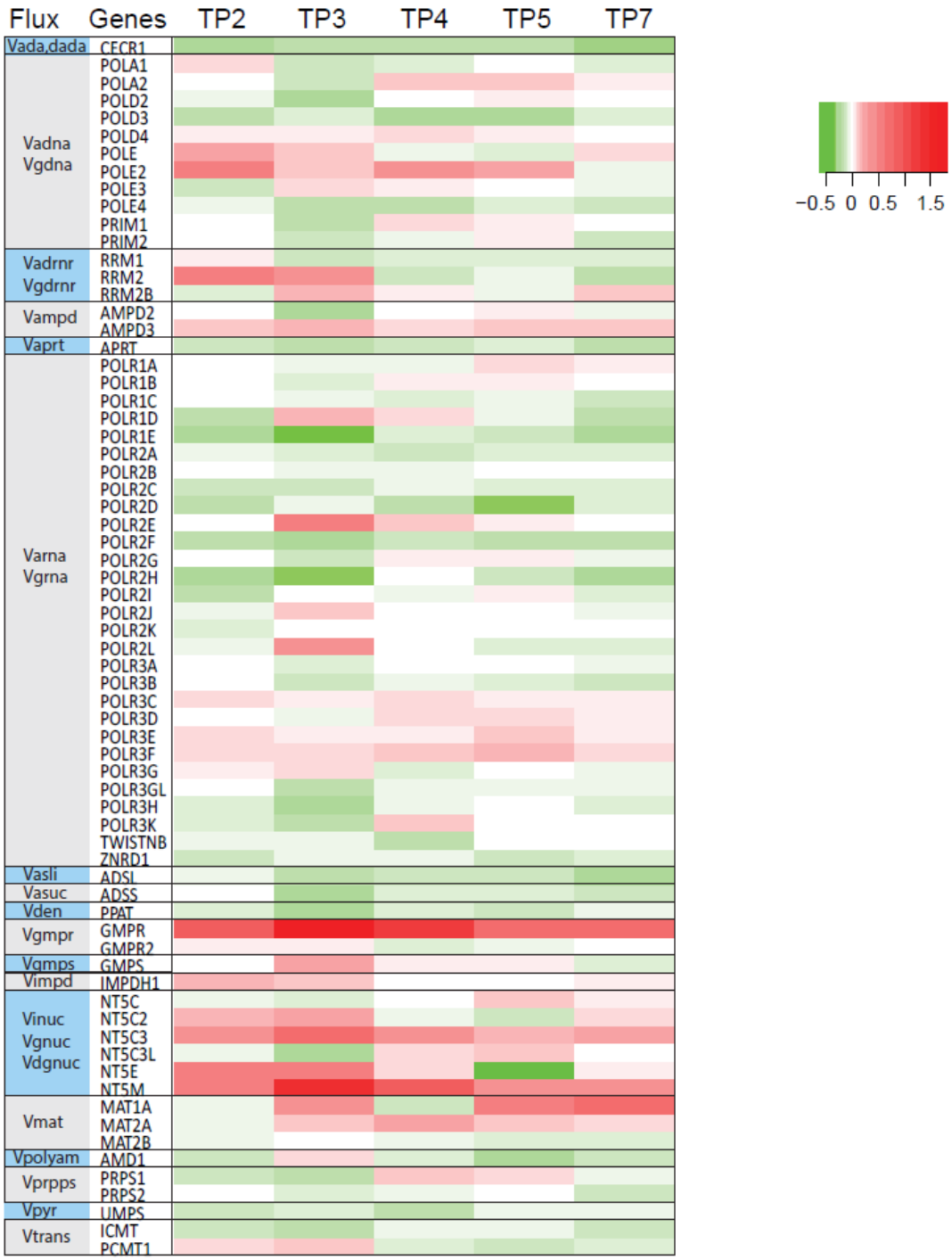
Appendix Table A.2.7- Evolutionary similarity score calculated for homologous genes between Mm and Mf. The complete table can be found at: https://github.com/LBSA-VoitLab/Mm_Mf_analysis/blob/master/Tables/table_S7.txt. A representative sample of gene similarity score for some genes:

Genes	Similarity (max 1)
A1CF	0.998385361
A2M	0.990088106
A2ML1	0.994613313
A4GALT	0.989350913
A4GNT	0.990607389
AAAS	1
AACS	0.968049155
AADAC	0.988264361
AADACL2	0.99402787
AADACL3	0.979161436
AADACL4	0.986658196
AADAT	1
AAED1	0.99751861
AAGAB	0.982069481
AAK1	0.999382462
AAMDC	0.994035785

APPENDIX B. SUPPLEMENTS FOR CHAPTER 3

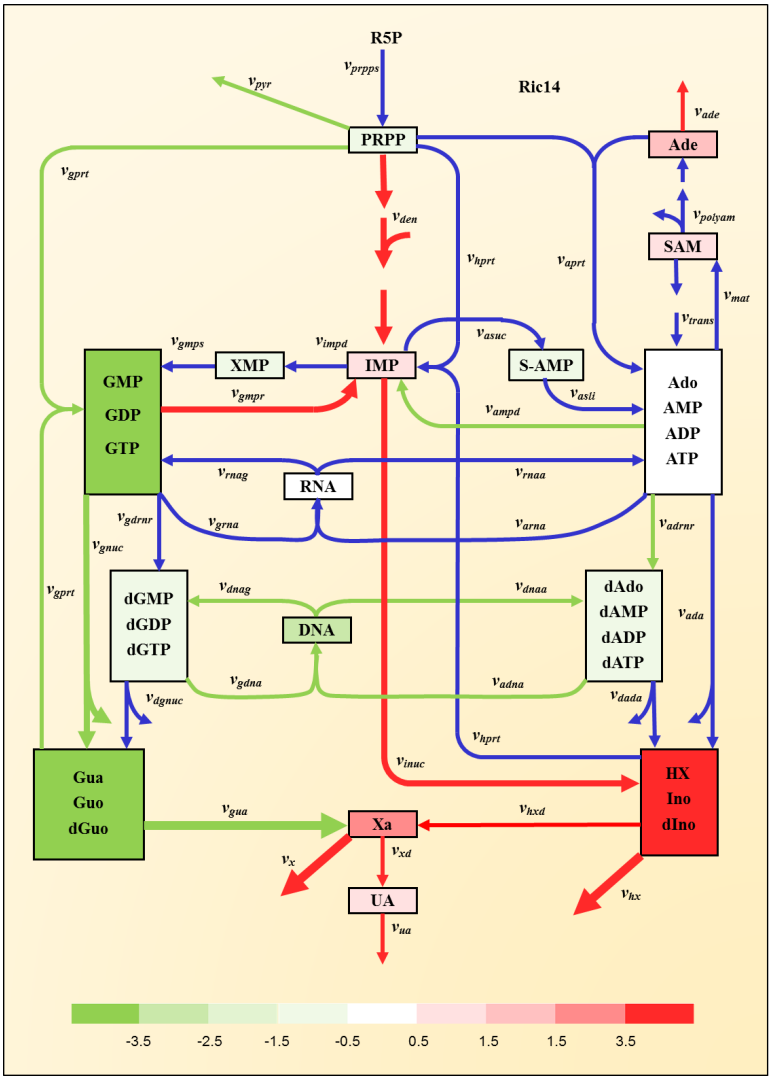
B.1 Supplementary Figures

Appendix Figure B.1.1- Heat map of changes in gene expression at different time points (TP2 – TP7) during *P. cynomolgi* infection. Shown here are transcriptomic changes in the bone marrow of monkey RFa14, relative to gene expression at TP1. TP3, which immediately follows the peak of infection (TP2), exhibits the strongest changes. Results are shown on a log-10 scale.

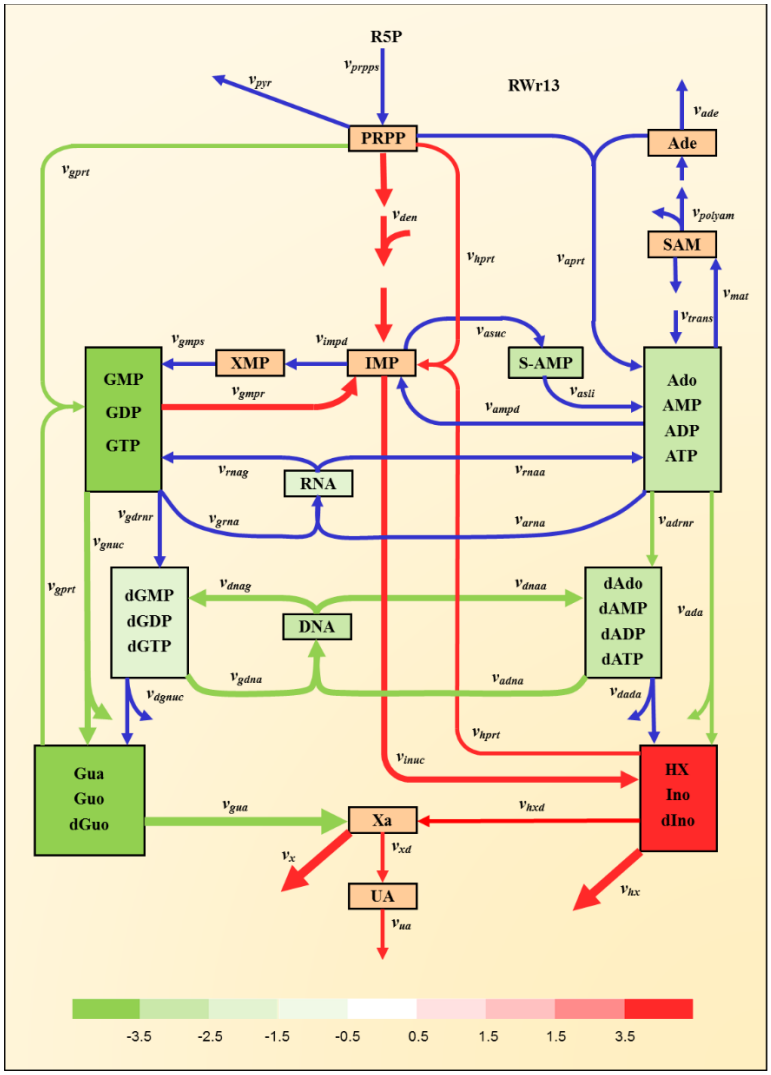


Appendix Figure B.1.2- Heat map of changes in gene expression at different time points (TP2 – TP7) during *P. coatneyi* infection. Shown here are transcriptomic changes in the peripheral blood of monkey RWr13, relative to gene expression at TP1. TP3, which immediately follows the peak of infection (TP2), exhibits the strongest changes. Data at TP6 were not measured. Results are shown on a log-10 scale.

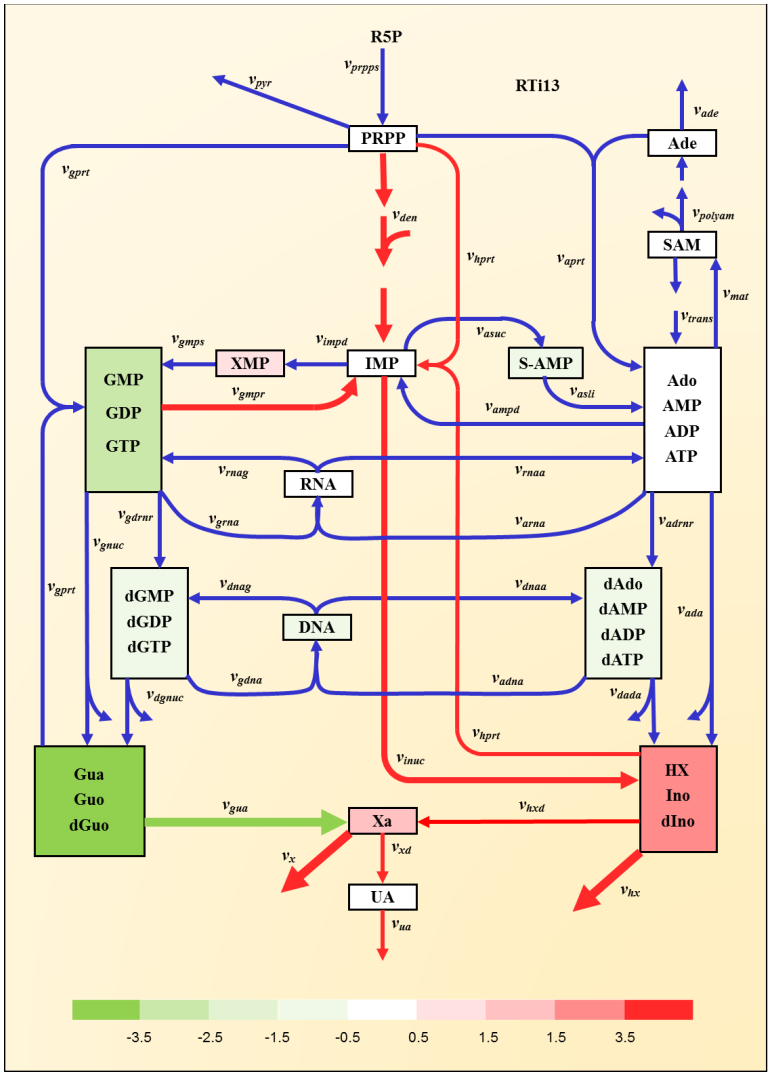
Appendix Figure B.1.3- Heat map of changes in gene expression at different time points (TP2 – TP7) during *P. coatneyi* infection. Shown here are transcriptomic changes in the peripheral blood of monkey RTi13, relative to gene expression at TP1. TP3, which immediately follows the peak of infection (TP2), exhibits the strongest changes. Results are shown on a log-10 scale.



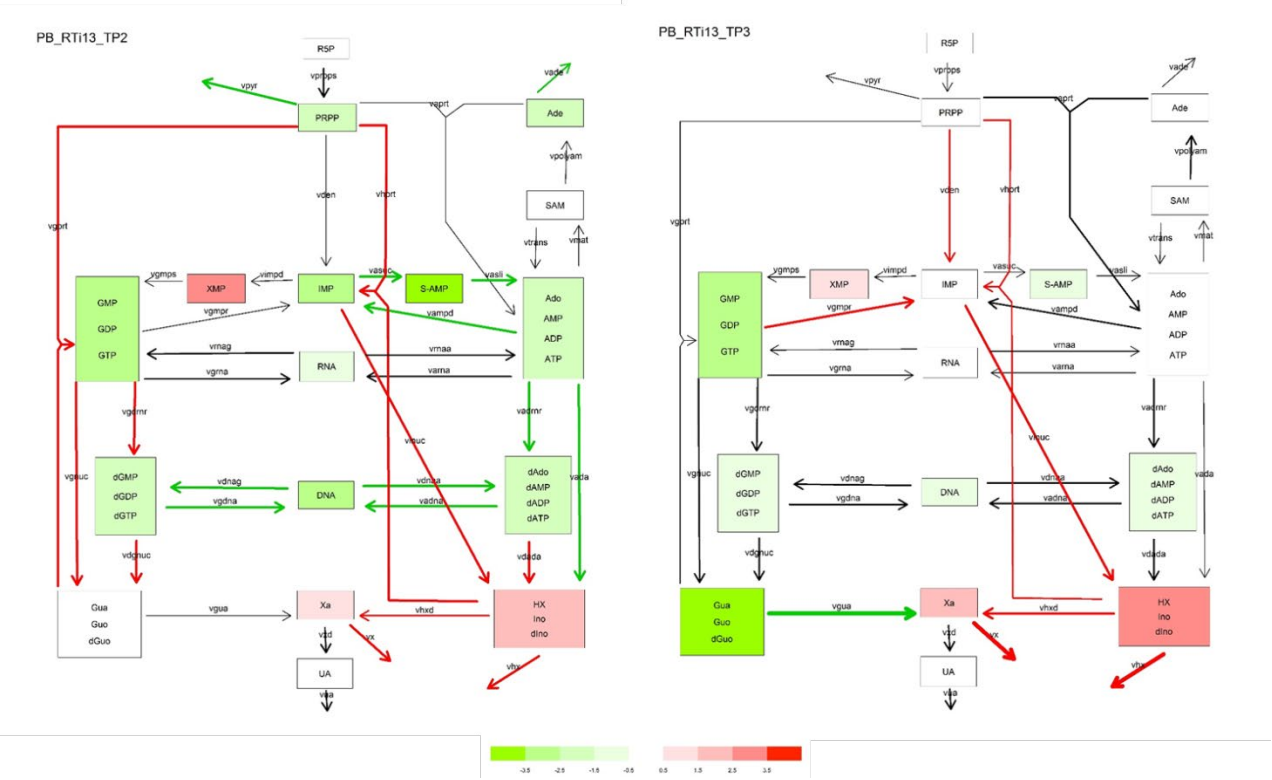
Appendix Figure B.1.4- Identification of key fluxes in purine metabolism. Diagram of purine metabolism in the blood of a *P. cynomolgi* infected macaque immediately after peak infection. This animal, RIc14, controlled the infection without sub-curative treatment.



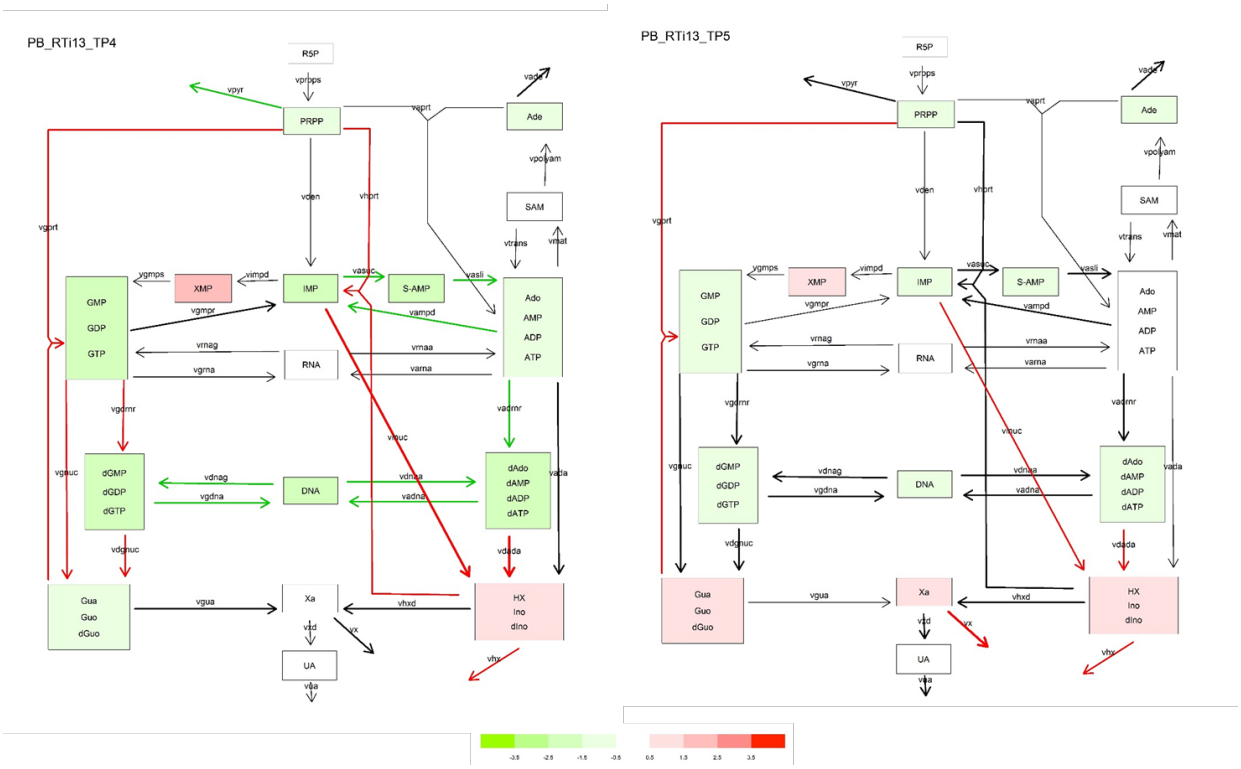
Appendix Figure B.1.5- Identification of key fluxes in purine metabolism. Diagram of purine metabolism in the blood of a *P. coatneyi* infected macaque immediately after peak infection. This animal, RWr13, had a high level of parasitemia.

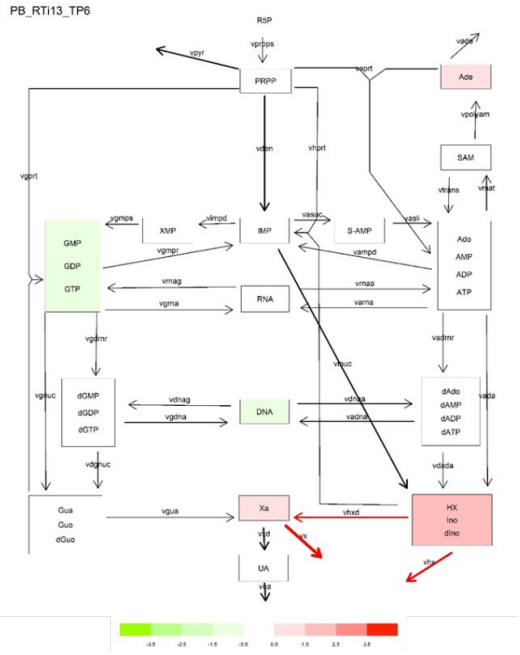


Appendix Figure B.1.6- Identification of key fluxes in purine metabolism. Diagram of purine metabolism in the blood of a *P. coatneyi* infected macaque immediately after peak infection. This animal, RTi13, chad a high level of parasitemia.

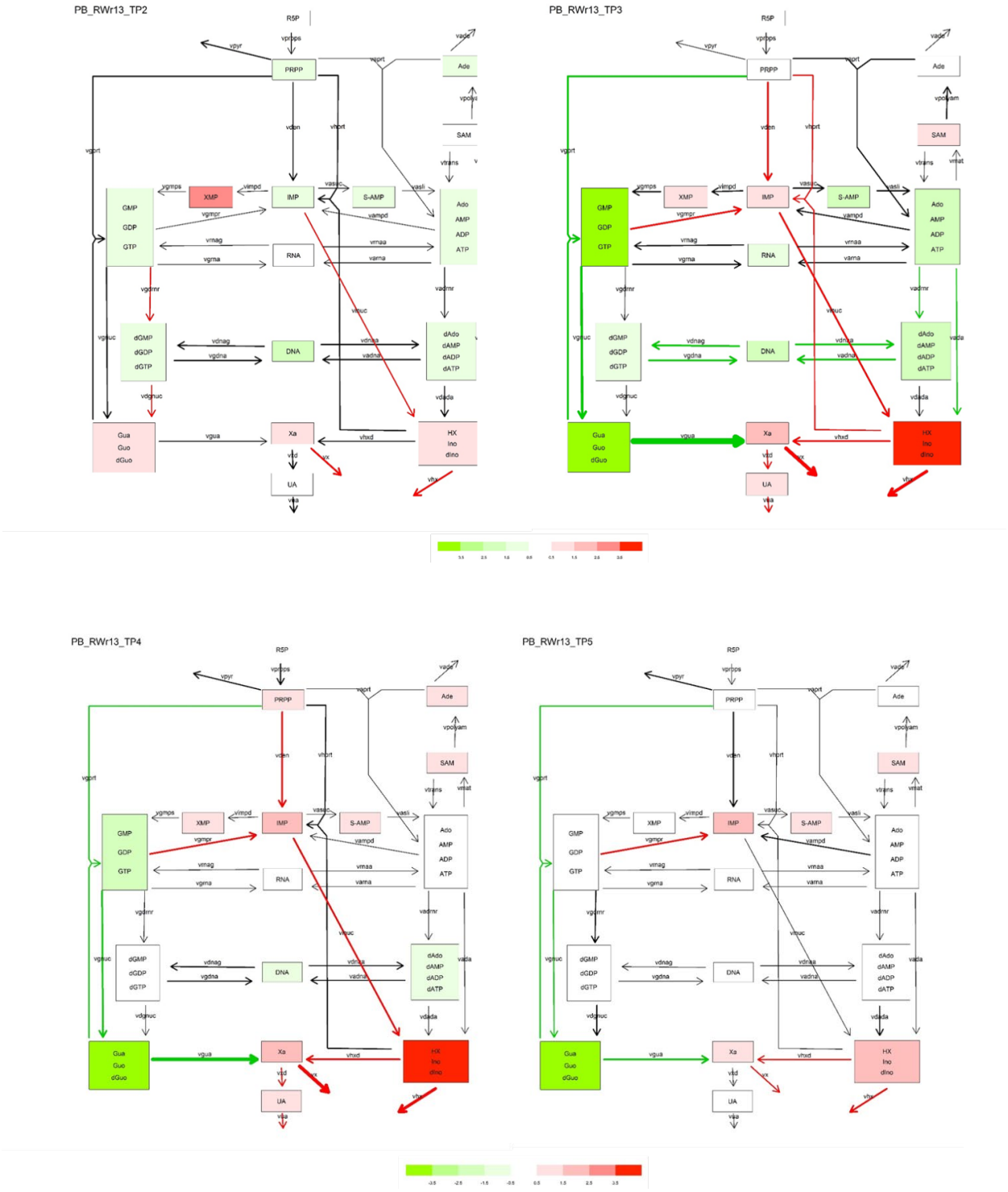


Appendix Figure B.1.7- Changes in key fluxes and metabolites related to chronic infection in malaria. In this *P. coatneyi* infection, animal RTi13 exhibited high parasitemia and showed prolonged increases in flux (v_{inuc}) toward Hypoxanthine (HX), Inosine (Ino) in blood throughout the infection. Shown here: TP2 and TP3.

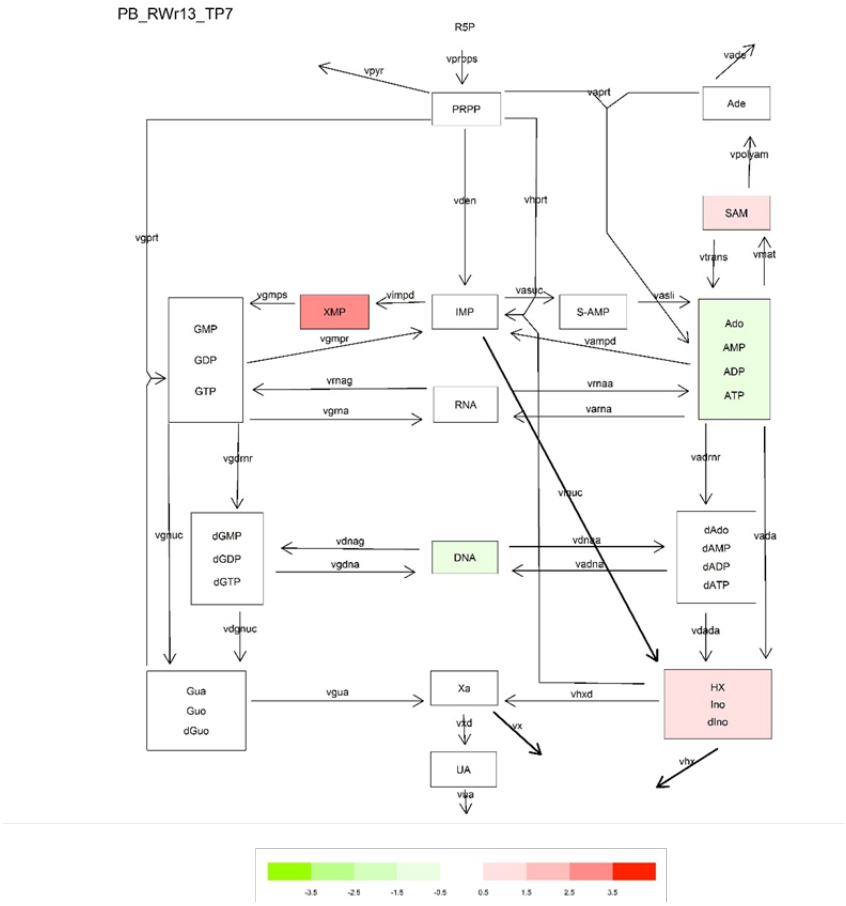




Appendix Figure B.1.8- Changes in key fluxes and metabolites related to chronic infection in malaria. In this *P. coatneyi* infection, animal RTi13 exhibited high parasitemia and showed prolonged increases in flux (v_{inuc}) toward Hypoxanthine (HX), Inosine (Ino) in blood throughout the infection. Shown here: TP4 – TP6; no results were available for TP7.



Appendix Figure B.1.9- Changes in key fluxes and metabolites related to chronic infection in malaria. In this *P. coatneyi* infection, animal RWr13 exhibited high parasitemia and showed prolonged increases in flux (v_{inuc}) toward Hypoxanthine (HX), Inosine (Ino) in blood throughout the infection. Shown here: TP2 – TP5.



Appendix Figure B.1.10- Changes in key fluxes and metabolites related to chronic infection in malaria. In this *P. coatneyi* infection, animal RWr13 exhibited high parasitemia and showed prolonged increases in flux (v_{inuc}) toward Hypoxanthine (HX), Inosine (Ino) in blood throughout the infection. Shown here: TP7; no results were available for TP6.

B.2 Supplementary Tables

Appendix Table B.2.1- Variables and Steady-State Concentrations of the Model of Purine Metabolism.

Abbreviation	Metabolite	Variable	Steady-State Concentration
PRPP	Phosphoribosylpyrophosphate	X_1	5
IMP	Inosine monophosphate	X_2	100
S-AMP	Adenylosuccinate	X_3	0.2
Ado	Adenosine	X_4	2,500
AMP	Adenosine monophosphate		
ADP	Adenosine diphosphate		
ATP	Adenosine triphosphate		
SAM	S-adenosyl-L-methionine	X_5	4
Ade	Adenine	X_6	1
XMP	Xanthosine monophosphate	X_7	25
GMP	Guanosine monophosphate	X_8	400
GDP	Guanosine diphosphate		
GTP	Guanosine triphosphate		
dAdo	Deoxyadenosine	X_9	6
dAMP	Deoxyadenosine monophosphate		
dADP	Deoxyadenosine diphosphate		
dATP	Deoxyadenosine triphosphate		
dGMP	Deoxyguanosine monophosphate	X_{10}	3
dGDP	Deoxyguanosine diphosphate		
dGTP	Deoxyguanosine triphosphate		
RNA	Ribonucleic acid	X_{11}	28,600
DNA	Deoxyribonucleic acid	X_{12}	5,160
HX	Hypoxanthine	X_{13}	10
Ino	Inosine		
dIno	Deoxyinosine		
Xa	Xanthine	X_{14}	5
Gua	Guanine	X_{15}	5
Guo	Guanosine		
dGuo	Deoxyguanosine		
UA	Uric acid	X_{16}	100
R5P	Ribose-5-phosphate	X_{17}	18
P_i	Phosphate	X_{18}	1,400

Appendix Table B.2.2- Enzymes and Steps of the Purine Model, along with Steady-State Flux Rates. Units are µmol x min^{-1} x BW^{-1}.

Flux	Enzyme Name or Step	E.C. Number	Steady-State Flux Rate
v_{prpps}	Phosphoribosylpyrophosphate synthetase (PRPPS)	2.7.6.1.	20.79
v_{gprt}	Hypoxanthine-guanine phosphoribosyltransferase (HGPRT)	2.4.2.8.	3.7
v_{hprt}	Hypoxanthine-guanine phosphoribosyltransferase (HGPRT)	2.4.2.8.	3.7
v_{aprt}	Adenine phosphoribosyltransferase (APRT)	2.4.2.7.	1
v_{den}	"De novo synthesis" (Amidophosphoribosyltransferase; ATASE)	2.4.2.14.	2.39
v_{pyr}	"pyrimidine synthesis"	several enzymes	10
v_{asuc}	Adenylosuccinate synthetase (ASUC)	6.3.4.4.	8
v_{asli}	Adenylosuccinate lyase (ASLI)	4.3.2.2.	8
v_{impd}	IMP dehydrogenase (IMPD)	1.1.1.205.	1.6
v_{gmps}	GMP synthetase (GMPS)	6.3.4.1.	1.6
v_{ampd}	AMP deaminase (AMPD)	3.5.4.6.	5.69
v_{gmpr}	GMP reductase (GMPR)	1.6.6.8.	0.5
v_{trans}	"transmethylation pathway" (Protein O-methyltransferase; MT)	2.1.1.24.	13.99
v_{mat}	Methionine adenosyltransferase (MAT)	2.5.1.6.	15
v_{polyam}	"Polyamine pathway" (S-adenosylmethionine decarboxylase; SAMD)	4.1.1.50.	1.01
v_{ade}	"Adenine oxidation" (xanthine oxidase)	1.2.1.37.	0.01
v_{inuc}	5'-Nucleotidase (5NUC)	3.1.3.5.	2.68
v_{gnuc}	5'-Nucleotidase (5NUC)	3.1.3.5.	4.7
v_{arna}	RNA polymerase (from ATP) (RNAP)	2.7.7.6.	1,980
v_{grna}	RNA polymerase (from GTP) (RNAP)	2.7.7.6.	1,320
v_{rnaa}	RNases (to AMP) (RNAN)	several enzymes	1,980
v_{rnag}	RNases (to GMP) (RNAN)	several enzymes	1,320
v_{dgnuc}	5'(3') Nucleotidase (3NUC)	3.1.3.31.	0.1
v_{ada}	Adenosine deaminase (ADA)	3.5.4.4.	2.1
v_{dada}	Adenosine deaminase (ADA)	3.5.4.4.	0.2
v_{adrnr}	Diribonucleotide reductase (DRNR)	1.17.4.1.	0.2
v_{gdrnr}	Diribonucleotide reductase (DRNR)	1.17.4.1.	0.1
v_{gua}	Guanine hydrolase (GUA)	3.5.4.3.	1.1
v_{adna}	DNA polymerase (from dATP) (DNAP)	2.7.7.7.	10
v_{gdna}	DNA polymerase (from dGTP) (DNAP)	2.7.7.7.	6.8
v_{dnaa}	DNases (to dAMP) (DNAN)	several enzymes	10
v_{dnag}	DNases (to dGMP) (DNAN)	several enzymes	6.8
v_{hx}	"Hypoxanthine excretion"	non-enzymatic	0.05
v_{hxd}	Xanthine oxidase or xanthine dehydrogenase (XD)	1.2.1.37.	1.23
v_{xd}	Xanthine oxidase or xanthine dehydrogenase (XD)	1.2.1.37.	2.3
v_{x}	"Xanthine excretion"	non-enzymatic	0.03
v_{ua}	"Uric acid excretion"	non-enzymatic	2.3

Appendix Table B.2.3- Kinetic Orders of the Model of Purine Metabolism.

f_{ada3} = 0.97	f_{adrnr3} = 0.1	f_{adrnr5} = -0.3	f_{adrnr6}= 0.87
f_{ampd3}=2.7	f_{ampd4}=-0.04	f_{aprt1} = 0.5	f_{aprt3} = -0.8
f_{aprt7} = 0.2	f_{asuc2} = 0.4	f_{asuc3} = -0.21	f_{asuc4} = 0.2
f_{dada5} = 1	f_{den1} = 2	f_{den2} = -0.06	f_{den3} = -0.25
f_{den4} = -0.2	f_{dgnuc6} = 1	f_{gdrnr4} = 0.4	f_{gdrnr5} = -1.2
f_{gdrnr6} = -0.39	f_{gmpr2} = -0.15	f_{gmpr3} = -0.07	f_{gmpr4} = -0.06
f_{gnuc4} = 0.9	f_{gprt1} = 1.2	f_{gprt4} = -1.2	f_{gprt9} = 0.42
f_{gua9} = 0.5	f_{hprt1} = 1.1	f_{hprt2} = -0.89	f_{hprt8} = 0.48
f_{hxd8} = 0.45	f_{hxd10} = -0.22	f_{impd2} = 0.15	f_{impd4} = -0.12
f_{inuc2} = 0.8	$f_{polyam3}$ = 0.9	f_{prpps1} = -0.03	f_{prpps3} = -0.45
f_{prpps4} = -0.04	f_{ua11} = 2.21	f_{xd8} = -0.55	f_{xd10} = 0.78

Appendix Table B.2.4- Rate Constants of the Model of Purine Metabolism.

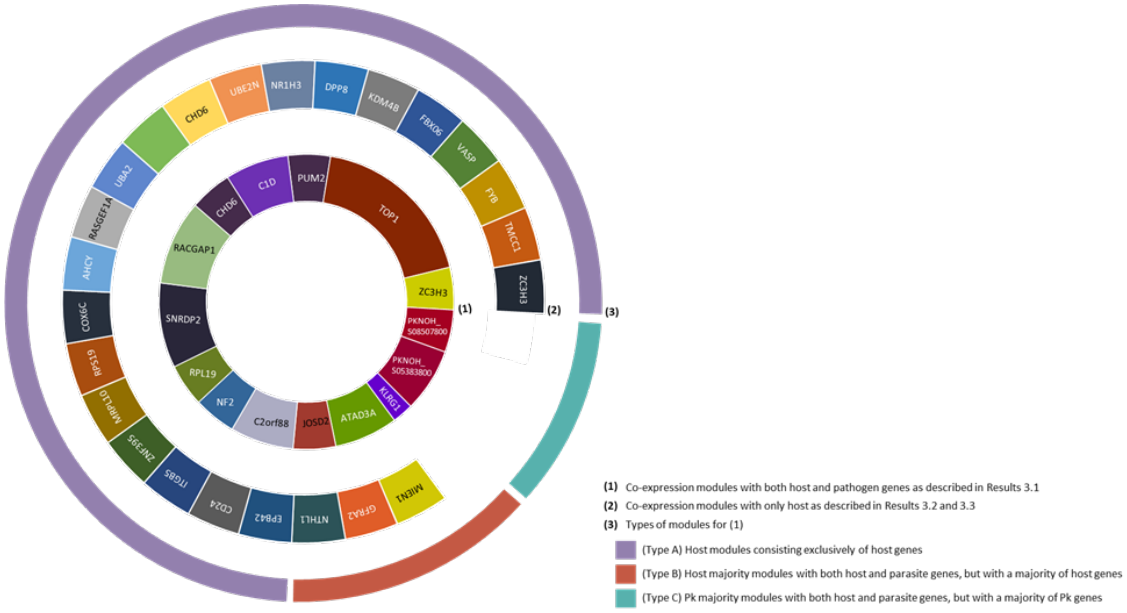
α_{ada} = 0.001062	α_{ade} = 0.01	α_{adna} = 3.2789	α_{adrnr} = 0.0602
α_{ampd}=0.02688	α_{aprt} = 233.8	α_{arna} = 614.5	α_{asli} = 66544.7
α_{asuc} = 3.5932	α_{dada} = 0.03333	α_{den} = 5.2728	α_{dgnuc} = 0.03333
α_{dnaa} = 0.001938	α_{dnag} = 0.001318	α_{gdna} = 2.2296	α_{gdrnr} = 0.1199
α_{gmpr} = 0.3005	α_{gmps} = 0.3738	α_{gnuc} = 0.2511	α_{gprt} = 361.69
α_{grna} = 409.6	α_{gua} = 0.4919	α_{hprt} = 12.569	α_{hx} = 0.003793
α_{hxd} = 0.2754	α_{impd} = 1.2823	α_{inuc} = 0.9135	α_{mat} = 7.2067
α_{polyam} = 0.29	α_{prpps} = 0.898	α_{pyr} = 1.2951	α_{rnaa} = 0.06923
α_{rnag} = 0.04615	α_{trans} = 8.8539	α_{ua} = 0.00008744	α_{x} = 0.0012
α_{xd} = 0.949			

B.3 Model Equations

```
X1'=    0.898 X1^-.03 X4^-.45 X8^-0.04 X17^.65 X18^.7 - 5.2728 X1^2 X2^-0.06 X4^-0.25
X8^-0.2 X18^-0.08 - 233.8 X1^.5 X4^-0.8 X6^.75   -   361.69 X1^1.2 X8^-1.2 X15^.42   -
12.569 X1^1.1 X2^-.89 X13^.48 -1.2951X1^1.27

X2'=   5.2728 X1^2 X2^-0.06 X4^-0.25 X8^-0.2 X18^-0.08 + 12.569 X1^1.1 X2^-.89 X13^.48
+ 0.02688 X4^.8 X8^-0.03 X18^-0.1   +   0.3005 X2^-.15 X4^-.07 X7^-0.76 X8^.7 - 3.5932
X2^.4 X4^-0.24 X8^0.2 X18^-0.05   -   1.2823 X2^.15 X7^-0.09 X8^-0.03 - 0.9135 X2^.8 X18^-
0.36

X3'= 3.5932 X2^.4 X4^-.24 X8^.2 X18^-.05   -   66544.7 X3^.99 X4^-0.95



X4' = 66544.7 X3^.99 X4^-0.95   +   233.8 X1^.5 X4^-0.8 X6^.75   +   0.06923 X11     +
8.8539 X5^.33   -   7.2067 X4^.2 X5^-.6   -   0.02688 X4^.8 X8^-0.03 X18^-0.1   -   614.5
X4^0.05 X8^0.13   -   0.001062 X4^.97   -   0.0602 X4^.1 X9^-.3 X10^0.87

X5' = 7.2067 X4^.2 X5^-0.6   -   8.8539 X5^.33   -   0.29 X5^.9

X6' = 0.29 X5^.9   -   233.8 X1^.5 X4^-0.8 X6^.75   -   0.01 X6^0.55

X7' = 1.2823 X2^.15 X7^-0.09 X8^-0.03   -   0.3738 X4^0.12 X7^0.16

X8' =   0.3738 X4^0.12 X7^0.16   +   361.69 X1^1.2 X8^-1.2 X15^.42   +   0.04615 X11 - 0.3005
X2^-.15 X4^-.07 X7^-0.76 X8^.7   -   409.6 X4^0.05 X8^.13   - 0.2511 X8^0.9 X18^-0.34   -
.1199 X8^.4 X9^-1.2 X10^-.39

X9' = 0.0602 X4^.1 X9^-.3 X10^0.87   +   0.001938 X12   -   3.2789 X9^.42 X10^.33   -   0.03333
X9

X10' = .1199 X8^.4 X9^-1.2 X10^-.39   +   .001318 X12   -   2.2296 X9^.42 X10^.33   -   0.03333
X10

X11' = 614.5 X4^0.05 X8^0.13   +   409.6 X4^0.05 X8^.13   -   0.06923 X11   -   0.04615 X11

X12' = 3.2789 X9^.42 X10^.33   +   2.2296 X9^.42 X10^.33   -   .001318 X12   -   .001938 X12

X13' = 0.001062 X4^.97   +   0.03333 X9   +   0.9135 X2^.8 X18^-0.36   -   12.569 X1^1.1 X2^
-.89 X13^.48   -   0.003793 X13^1.12 -   0.2754 X13^.65

X14' = 0.2754 X13^.65   +   0.4919 X15^.5   -   0.949 X14^.55   -   0.0012 X14^2

X15' =  0.2511  X8^0.9  X18^-0.34    +    0.03333  X10    -    361.69  X1^1.2  X8^-1.2  X15^.42
- 0.4919 X15^.5

X16' = .949 X14^.55   -   .00008744 X16^2.21
```

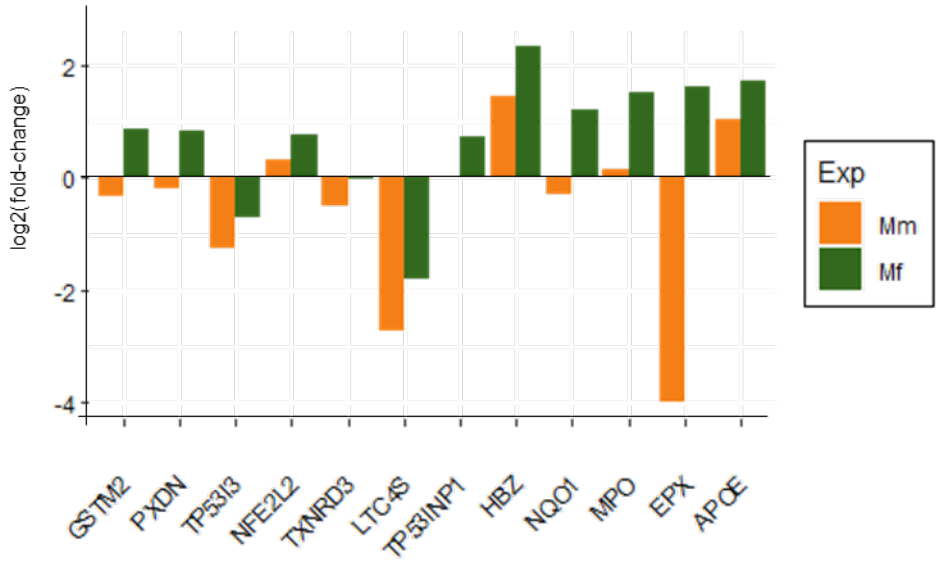
APPENDIX C. SUPPLEMENTS FOR CHAPTER 4

C.1 Supplementary Figures

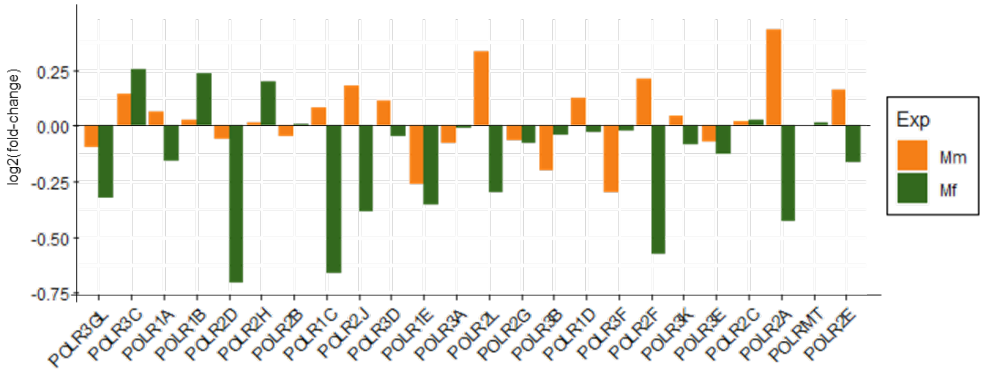


Appendix Figure C.1.1- Illustration for WGCNA showing different analyses and their corresponding modules. Modules in (1) and (2) have been radially adjusted to be closest to modules most similar in the two analyses. (1) is corresponding to analysis for Sec 4.4.1.1. (2) is corresponding to the analysis for Sec 4.4.1.2 and 4.4.1.3.

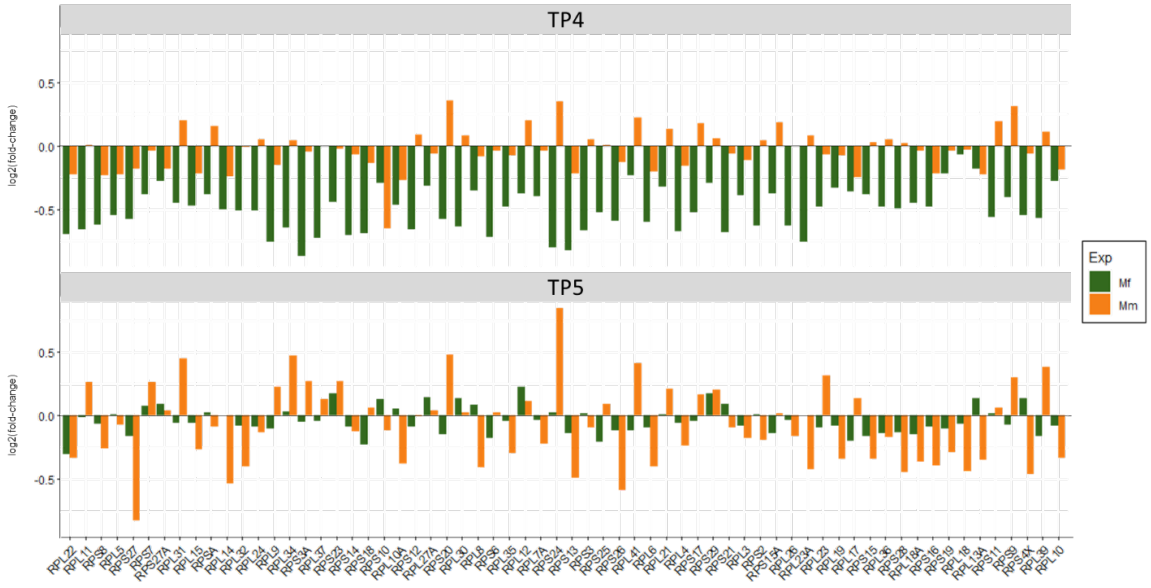
Appendix Figure C.1.2- Barplot for enrichment of genesets related to protein kinase signaling. (A) Genesets that are similarly enriched in both Mm and Mf during log phase (TP4). (B) Genesets that are differently enriched in Mm and Mf during log phase (TP4).



Appendix Figure C.1.3- Barplot of fold changes in the expression (at TP4) of genes involved in protection against ROS through antioxidants. These genes have also been shown to interact with the p53 pathway.



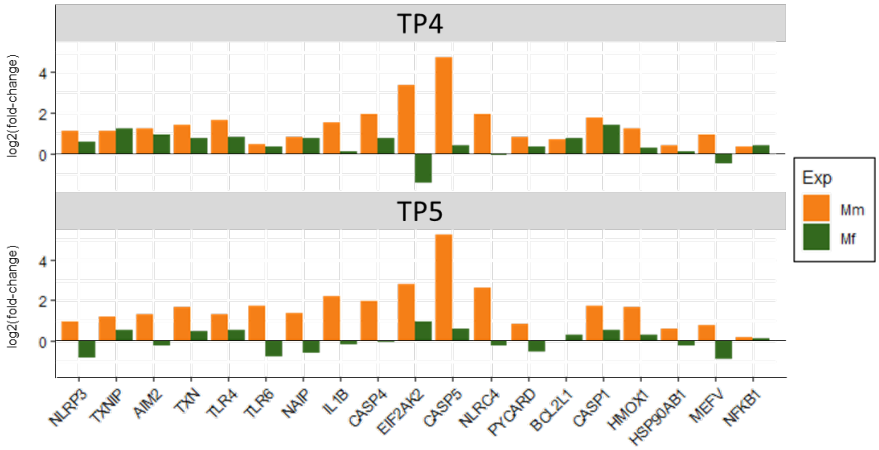
Appendix Figure C.1.4- Barplot of fold changes in the expression of RNA polymerase associated genes (at TP4) that are responsible for ribosomal biosynthesis. Worth noting are PLOR1C, POLR2E, POLR2A, POLR2J and POLR2L, as they correspond to downregulation of ribosomal biosynthesis in Mf at TP4.



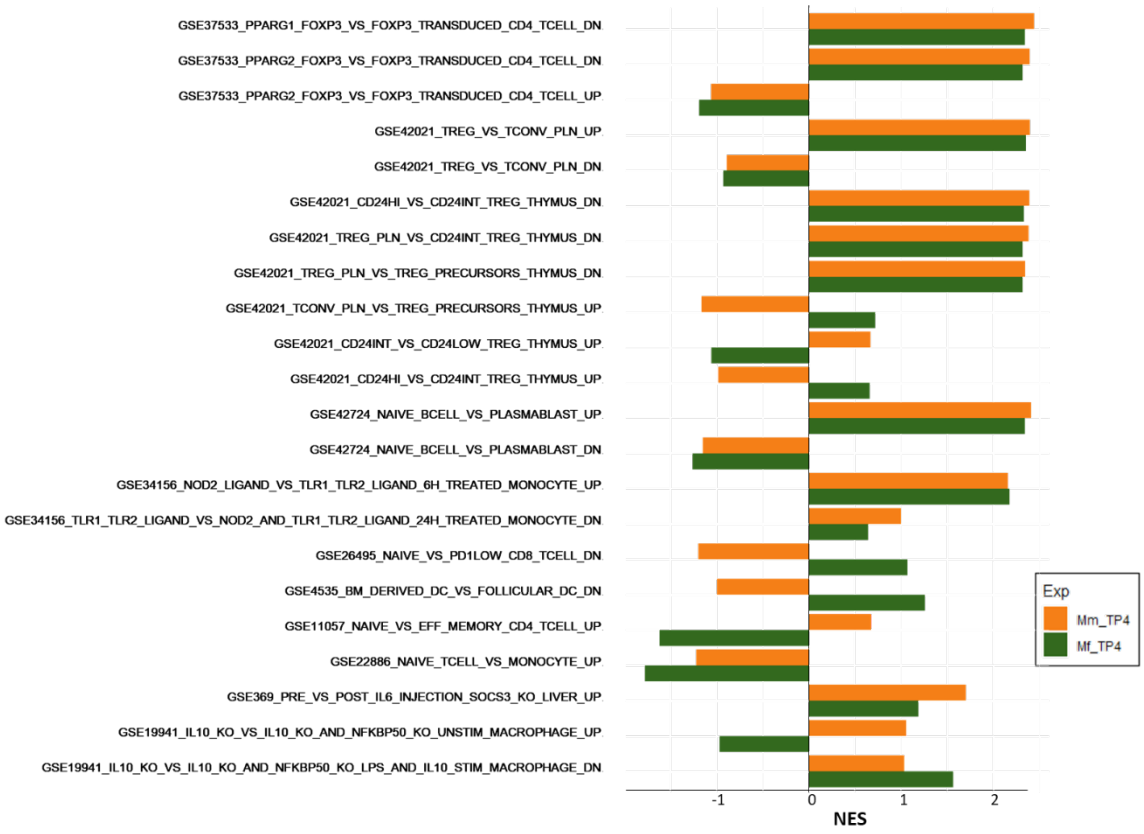
Appendix Figure C.1.5- Barplot of fold changes in gene expression of Ribosomal Proteins (RPs), comparing both hosts across TP4 and TP5. This comparison highlights the differences between TP4 and TP5.

Appendix Figure C.1.6- Barplot of fold changes in the expression of genes coding for heat shock factor (HSF) related proteins (chaperones) and signals for stress, Hsf1 and apoptotic target genes, comparing Mm and Mf across TP4 and TP5.

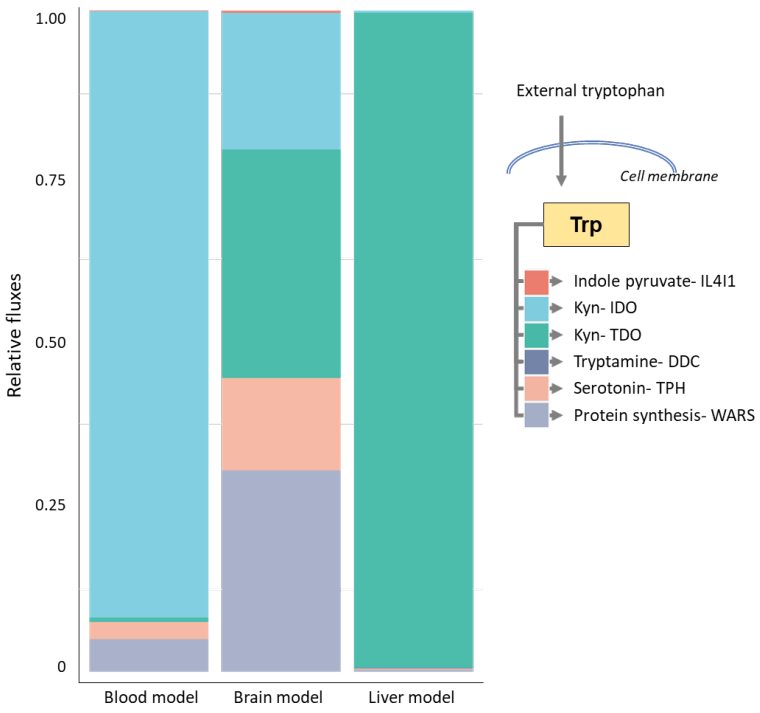
Appendix Figure C.1.7- Barplot of enrichment scores of inflammatory genesets during log phase of infection (TP4). The plot highlights (A) similar (light blue) and (B) distinguishing (light orange) genesets between Mm and Mf.



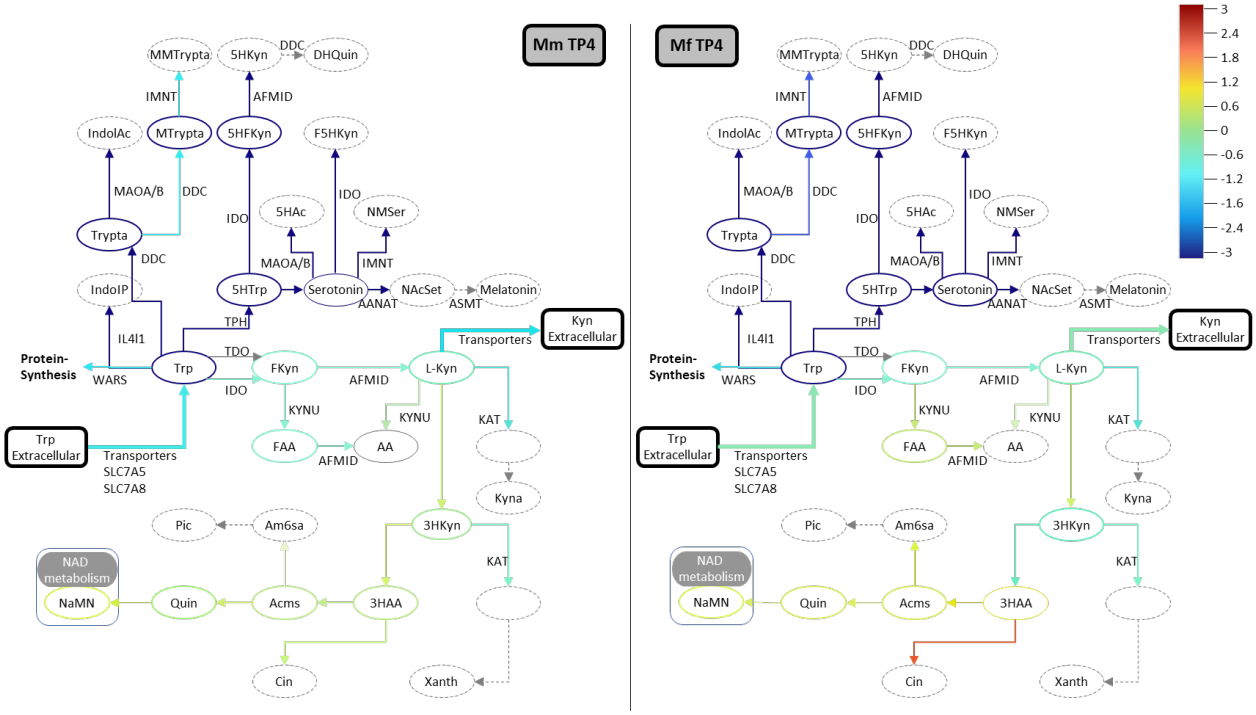
Appendix Figure C.1.8- Barplot of fold changes in the expression of genes involved in the inflammasome complex and its regulation across TP4 and TP5, differentiating Mm from Mf.



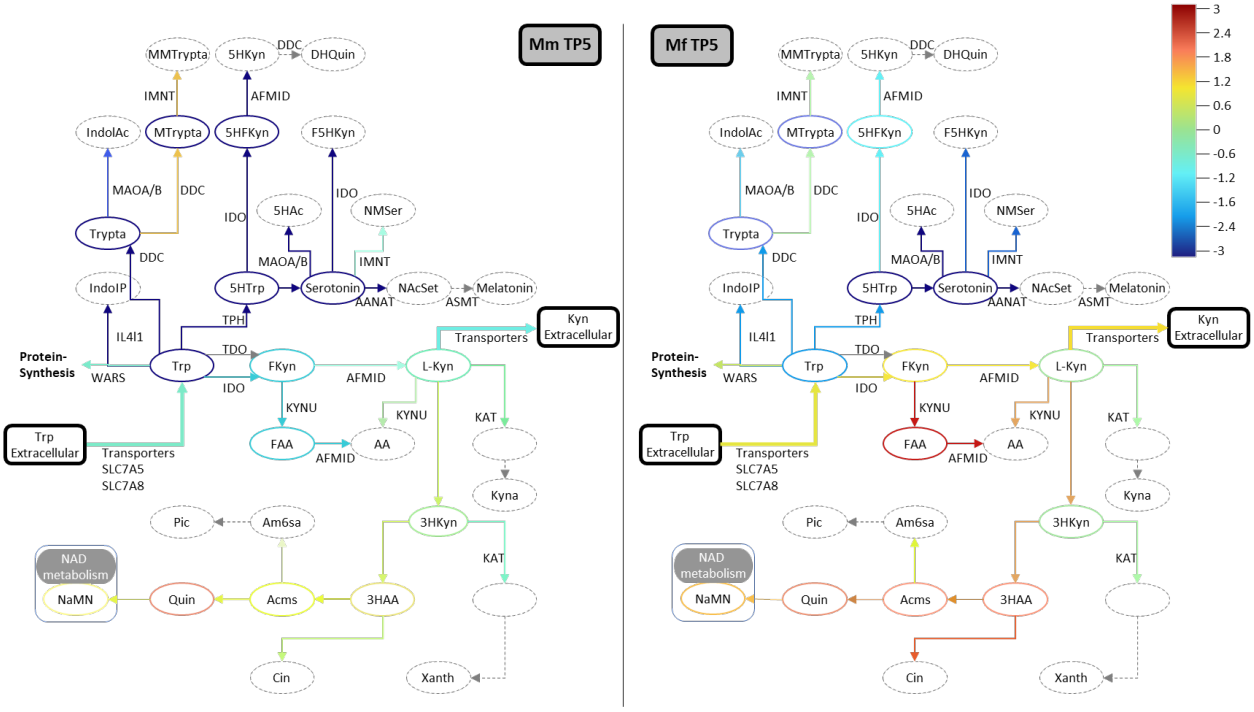
Appendix Figure C.1.9- Barplot of enrichment scores of important immunologic signatures. The plot shows similarities and differences between Mm and Mf during log phase.



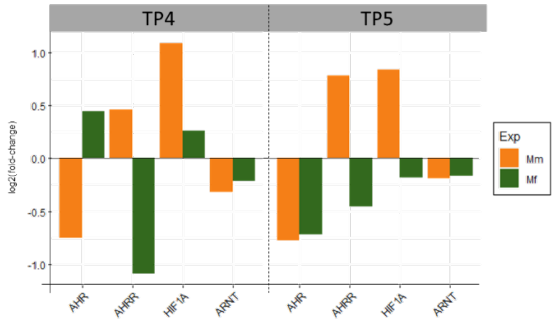
Appendix Figure C.1.10- Comparison of flux distributions through Trp in blood, brain and liver models.



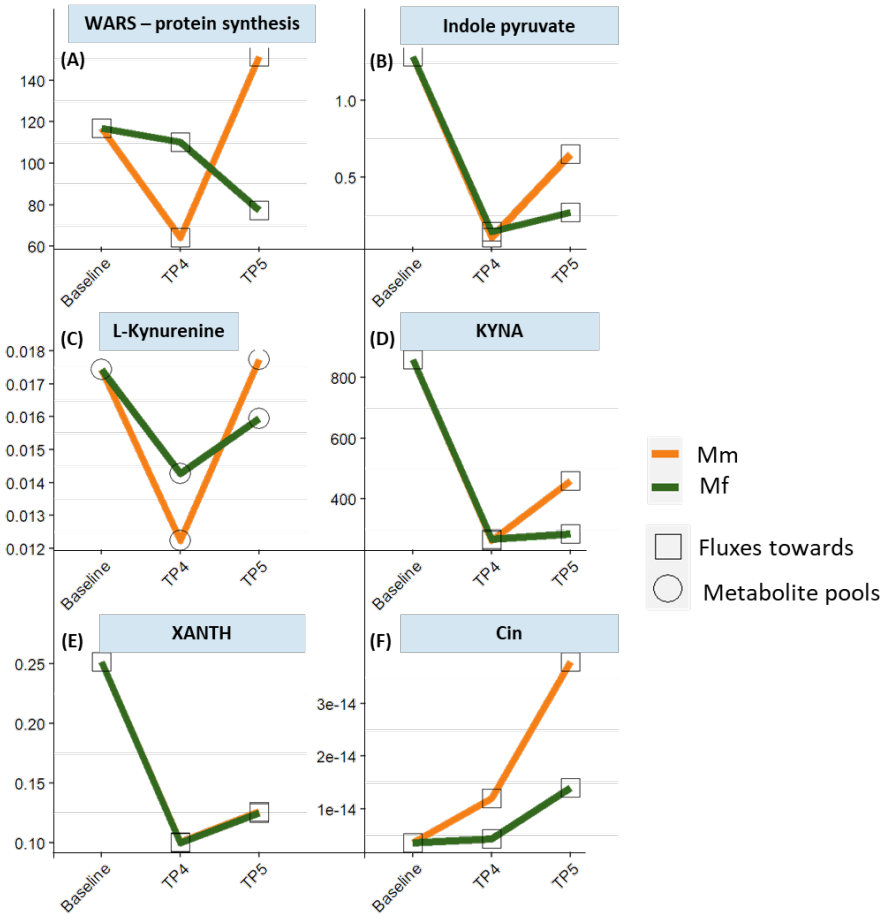
Appendix Figure C.1.11- Tryptophan metabolism model adjusted for enzymatic activity for Mm and Mf at TP4. The colored arrows show corresponding changes in fluxes (log2 fold change) while colored ovals show corresponding changes in metabolite concentrations (log2 fold change) in comparison to the baseline as predicted by the model.



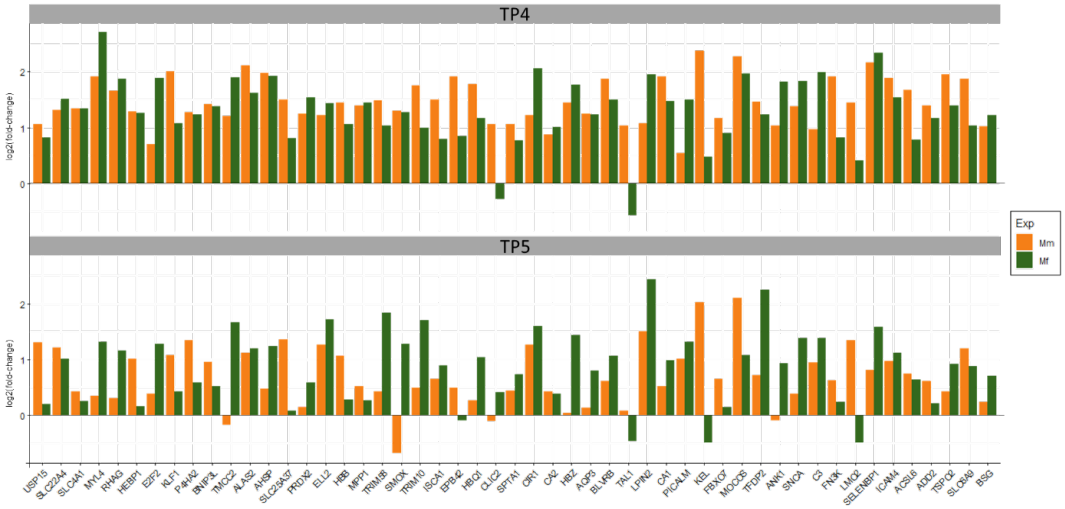
Appendix Figure C.1.12- Tryptophan metabolism model adjusted for enzymatic activity for Mm and Mf at TP5. The colored arrows show corresponding changes in fluxes (log2 fold change) while colored ovals show corresponding changes in metabolite concentrations (log2 fold change) in comparison to the baseline as predicted by the model.



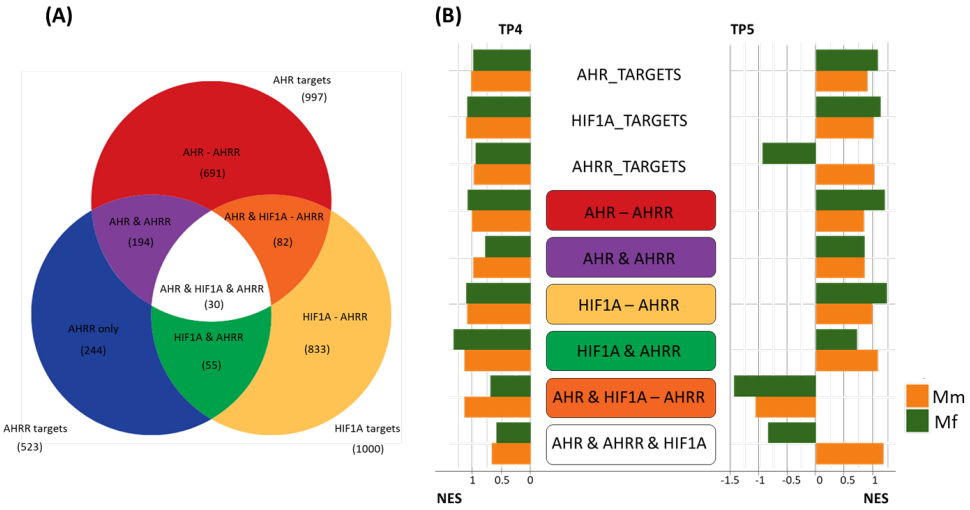
Appendix Figure C.1.13- Changes in bHLH-PAS superfamily genes involved in AhR signaling.



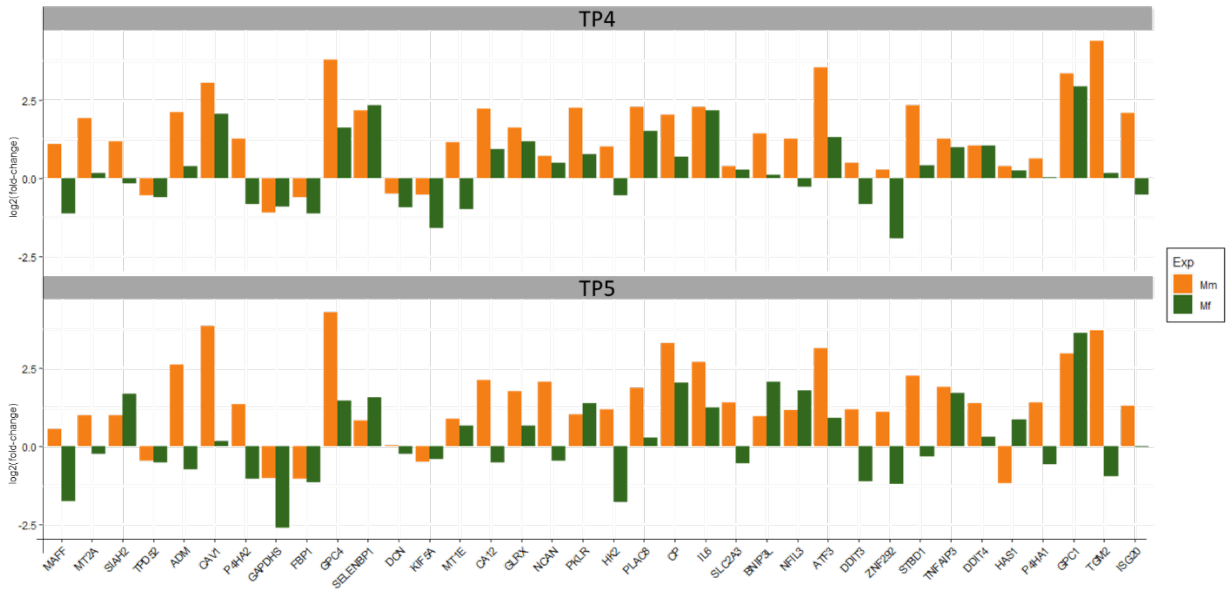
Appendix Figure C.1.14- Model predictions for metabolites and fluxes during infection. (A) WARS driven protein synthesis of tryptophan. (B-F) Tryptophan metabolism derived ligands for AhR measurements (fluxes and metabolite pool) as observed from the model. (B) Tryptophan derived indole pyruvate ligand for AhR. (C-F) Kynurenine derived kynurenines as ligands for AhR.



Appendix Figure C.1.15- Barplots of fold changes in expression of genes related to of heme metabolism comparing the two hosts across TP4 and TP5.



Appendix Figure C.1.16- Results regarding AhR, AhRR, ARNT and HIF1α complexes. (A) Venn diagram for targets for the three complexes involving the aryl hydrocarbon receptor nuclear translocator (ARNT): AhR:ARNT, AhRR:ARNT and HIF1α:ARNT. (B) Enrichment of target genes for members of the bHLH-PAS superfamily of transcription factors during TP4 and TP5. Enrichment of AhR (red) and HIF1α (beige) target genes highlighting the effect of AHRR (TP4). AHR\AHRR: AhR-only targets not affected by AhRR; AHR & AHRR: AHR and AHRR targets; HIF1A\AHRR: HIF1A-only targets not affected by AhRR; HIF1A & AHRR: HIF1A and AHRR targets; HIF1A & AHR\AHRR: HIF1A and AHR targets not affected by AhRR; AHR & AHRR & HIF1A: Targets for HIF1A, AHR and AHRR. Here, "&" represents presences of both components, while "\" represents absence of the second component.



Appendix Figure C.1.17- Barplot of fold changes of hypoxia related genes compared between the two hosts across TP4 and TP5.

C.2 Supplementary Tables

Appendix Table C.2.1- Co-expression modules obtained from WGCNA and their functional annotation (with Pk genes and excluding TP1,2,3) corresponding to Sec 4.4.1.1 (Appendix Figure C.1.1(1)).

Type	Module name/ hub genes	FDR(q-value)	Term Ontology	Term Name
A	SNRPD2	3E-61	MF	structural constituent of ribosome
A	TOP1	7E-45	BP	immune response
A	RACGAP1	2E-31	BP	mitotic cell cycle
A	PUM2	5E-06	BP	vesicle-mediated transport
A	JOSD2	4E-04	BP	heme metabolic process
A	NF2	1E-03	BP	tRNA metabolic process
A	ZC3H3	2E-03	BP	mRNA metabolic process
A	C2orf88	2E-02	CC	integrin complex
C	PKNOH_S08507800	2E-01	BP	pteridine-containing compound metabolic process
B	KLRG1	2E-01	MF	calcium ion binding
C	PKNOH_S05383800	3E-01	BP	regulation of cell activation
A	RPL19	1E+00	MF	UDP-galactosyltransferase activity
A	CHD6	1E+00	BP	positive regulation of bone mineralization
A	C1D	1E+00	MF	ion transmembrane transporter activity
B	ATAD3A	1E+00	MF	snoRNA binding

Appendix Table C.2.2- Host-pathogen transcript pairs with significantly high correlation (r>0.7; FDR p <0.01). Most of the host and pathogen genes belong to modules ATAD3A and PKNOH_S08507800.

Host genes	Pathogen transcripts	Correlation
C8orf59	PKNOH_S060437107	0.71
C8orf59	PKNOH_S09548900	0.75
C8orf59	PKNOH_S100028400	0.75
C8orf59	PKNOH_S140249600	0.72
CDK6	PKNOH_S100028400	0.71
CRIPT	PKNOH_S05369300	0.70
CRIPT	PKNOH_S060437107	0.71
CRIPT	PKNOH_S09548900	0.74
CRIPT	PKNOH_S100028400	0.72
CRIPT	PKNOH_S140249600	0.72
ELK4	PKNOH_S060437107	0.70
ELK4	PKNOH_S09548900	0.71
ELK4	PKNOH_S100028400	0.72
ELK4	PKNOH_S140249600	0.70
FAAH2	PKNOH_S010026907	0.71
FAAH2	PKNOH_S03322200	0.71
FAAH2	PKNOH_S05374800	0.73
FAAH2	PKNOH_S05387900	0.72
FAAH2	PKNOH_S06426900	0.72
FAAH2	PKNOH_S09548900	0.72
FAAH2	PKNOH_S1000070405	0.70
FAAH2	PKNOH_S120125800	0.70
FAAH2	PKNOH_S120148500	0.70
FAAH2	PKNOH_S140249600	0.70
GFPT1	PKNOH_S08492300	0.71
HSPA6	PKNOH_S020314401	0.70
HSPA6	PKNOH_S05369300	0.70
HSPA6	PKNOH_S060437107	0.71
HSPA6	PKNOH_S09548900	0.73
HSPA6	PKNOH_S100028400	0.72
HSPA6	PKNOH_S130186700	0.73
HSPA6	PKNOH_S140249600	0.72

Host genes	Pathogen transcripts	Correlation
IL10	PKNOH_S09548900	0.74
IL10	PKNOH_S100028400	0.72
IL10	PKNOH_S140249600	0.70
KRBOX4	PKNOH_S100028400	0.70
OTUD3	PKNOH_S020314401	0.72
OTUD3	PKNOH_S03322100	0.70
OTUD3	PKNOH_S04352200	0.72
OTUD3	PKNOH_S08471600	0.71
OTUD3	PKNOH_S08492300	0.74
OTUD3	PKNOH_S08508800	0.71
OTUD3	PKNOH_S09551600	0.70
OTUD3	PKNOH_S1000070402	0.72
OTUD3	PKNOH_S100028400	0.72
OTUD3	PKNOH_S140272900	0.70
OTUD3	PKNOH_S140280700	0.70
POLR2K	PKNOH_S09548900	0.72
POLR2K	PKNOH_S100028400	0.72
PSMC6	PKNOH_S08492300	0.72
PSMC6	PKNOH_S09548900	0.75
PSMC6	PKNOH_S100028400	0.73
PSMC6	PKNOH_S140249600	0.74
RGS2	PKNOH_S09548900	0.71
RGS2	PKNOH_S100028400	0.72
SLC16A7	PKNOH_S02306500	0.72
SLC30A6	PKNOH_S09548900	0.73
SLC30A6	PKNOH_S100028400	0.73
SLC30A7	PKNOH_S09548900	0.70
SLC30A7	PKNOH_S100028400	0.73
SUMO2	PKNOH_S08485100	0.72
SUMO2	PKNOH_S09548900	0.71
USP37	PKNOH_S060437105	0.70
USP37	PKNOH_S100028400	0.70
ZRANB2	PKNOH_S100028400	0.71

Appendix Table C.2.3- Blast results showing gene names and accession ids for P. knowlesi transcripts that are highly correlated to host genes. The majority are SICAvar genes.

Pk genes	OTN/GI
Uncharacterized protein PKNOH_S03322100 [Plasmodium knowlesi]	OTN68242.1 GI:1192741552
Uncharacterized protein PKNOH_S140272900 [Plasmodium knowlesi]	OTN64105.1 GI:1192737404
SICAvar type I [Plasmodium knowlesi]	OTN68783.1 GI:1192742095
SICAvar type I [Plasmodium knowlesi]	OTN68579.1 GI:1192741890
SICAvar type I [Plasmodium knowlesi]	OTN68414.1 GI:1192741725
SICAvar type I [Plasmodium knowlesi]	OTN68195.1 GI:1192741505
SICAvar type I [Plasmodium knowlesi]	OTN68043.1 GI:1192741352
SICAvar type I [Plasmodium knowlesi]	OTN67861.1 GI:1192741169
SICAvar type I, partial [Plasmodium knowlesi]	OTN67700.1 GI:1192741008
SICAvar type I [Plasmodium knowlesi]	OTN67596.1 GI:1192740904
SICAvar type I [Plasmodium knowlesi]	OTN67408.1 GI:1192740715
SICAvar type I, partial [Plasmodium knowlesi]	OTN67407.1 GI:1192740714
SICAvar type I [Plasmodium knowlesi]	OTN67394.1 GI:1192740701
SICAvar type I [Plasmodium knowlesi]	OTN66889.1 GI:1192740194
SICAvar type I [Plasmodium knowlesi]	OTN66858.1 GI:1192740163
SICAvar type I [Plasmodium knowlesi]	OTN66731.1 GI:1192740036
SICAvar - type I [Plasmodium knowlesi]	OTN66658.1 GI:1192739963
SICAvar type I, partial [Plasmodium knowlesi]	OTN66461.1 GI:1192739765
SICAvar type I [Plasmodium knowlesi]	OTN66174.1 GI:1192739478
SICAvar type I [Plasmodium knowlesi]	OTN66123.1 GI:1192739426
SICAvar type I [Plasmodium knowlesi]	OTN65758.1 GI:1192739061
SICAvar type I [Plasmodium knowlesi]	OTN65127.1 GI:1192738428
SICAvar type I [Plasmodium knowlesi]	OTN65104.1 GI:1192738405
SICAvar type I [Plasmodium knowlesi]	OTN64516.1 GI:1192737816
putative 40S ribosomal protein S17 [Plasmodium knowlesi]	OTN64312.1 GI:1192737611
SICAvar type I [Plasmodium knowlesi]	OTN63945.1 GI:1192737244

Appendix Table C.2.4- Co-expression modules obtained from WGCNA and their functional annotation (without Pk genes and including TP1,2,3) corresponding to Sec 3.1.2 (Appendix Figure C.1.1(2))

Type	Module name/ hub genes	FDR(q-value)	Ontology Term Name	Pk proteins (+ ively correlated to module)	Pk proteins (- ively correlated to module)	Mm TP4 FDR(q-value)	Mf TP4 FDR(q-value)
A	FBXO6	2.3E-39	defense response	putative Heat shock protein 86 family; putative Pyridoxal 5'-phosphate dependent enzyme class III; putative Mitochondrial import inner membrane translocase subunit TIM17; putative Phosphatidylethanolamine-binding protein; putative FAD synthetase; KIR protein	putative ATPase subunit 9; putative CDGSH iron-sulfur domain-containing protein; putative U6 snRNA-associated Sm-like protein LSm8; putative Syntaxin; Thioredoxin; Proteasome endopeptidase complex; putative AP-1 complex subunit sigma; Peptidyl-prolyl cis-trans isomerase	1E-10	3E-17
A	GFRA2	1.0E+00	cytokine receptor activity			7E-08	5E-09
A	RASGEF1A	1.3E-04	cell periphery	putative Bet3 transport protein; putative U6 snRNA-associated Sm-like protein LSm8; putative 26S proteasome regulatory subunit RPN6; putative Phosphoethanolamine N-methyltransferase ; 6-phosphogluconate dehydrogenase - decarboxylating; Peptidyl-prolyl cis-trans isomerase; putative Ras-related protein Rab-6; putative Cullin-like protein	putative AAA family ATPase; putative Heat shock protein 86 family; putative RNA-binding protein; putative Leucine aminopeptidase; putative Heat shock protein 70	3E-06	2E-09
A	ZNF395	1.0E-03	DNA binding			7E-04	2E-06
A	CHD6	1.0E+00	DNA-binding transcription factor activity			7E-04	4E-07
A	MIEN1	3.3E-04	cellular respiration			1E-03	4E-05
A	EPB42	3.5E-05	hemoglobin complex	putative 60S ribosomal protein L3; putative Ribosomal protein S25; putative Chaperone binding protein; putative 40S ribosomal protein S11; putative 60S Ribosomal protein L44; putative 60S ribosomal protein L38; putative Ribosomal protein L35; putative Trophozoite exported protein 1	putative Regulator of chromosome condensation; putative Lysine--tRNA ligase; putative mitochondrial Acyl carrier protein; putative Sortilin; Casein kinase II subunit beta; ATP synthase subunit gamma	1E-03	2E-04

A	FYB	2.5E-12	cytoplasmic vesicle	KIR-like protein; putative YT521-B-like family protein; putative Pyridoxal 5'-phosphate dependent enzyme class III; putative FAD synthetase; putative ValS protein; putative Ornithine aminotransferase; putative Coatomer alpha subunit	SICAvar type II; putative High molecular weight rhoptry protein 3; Histone H2B; Histone H2A; putative Exosome complex component RRP45; Nucleoside diphosphate kinase	2E-03	5E-06
A	AHCY	2.3E-10	mitochondrial envelope	putative U6 snRNA-associated Sm-like protein LSm8; putative RING zinc finger protein [Plasmodium knowlesi]; Thioredoxin; Proteasome endopeptidase complex; putative AP-1 complex subunit sigma; Peptidyl-prolyl cis-trans isomerase; ATP synthase subunit beta	putative AAA family ATPase; putative Heat shock protein 86 family; putative YT521-B-like family protein; putative Pyridoxal 5'-phosphate dependent enzyme class III; putative FAD synthetase; putative ValS protein; putative Heat shock protein 70; putative Heat shock protein 90;	5E-03	3E-06
A	ITGB5	1.5E-03	blood coagulation			7E-03	3E-03
A	CEP55	7.9E-35	mitotic cell cycle	putative U6 snRNA-associated Sm-like protein LSm8; Serine/threonine-protein phosphatase; putative Phosphoethanolamine N-methyltransferase; Proteasome endopeptidase complex; Proteasome subunit beta type; Peptidyl-prolyl cis-trans isomerase	putative Heat shock protein 86 family; SICAvar type I; KIR protein; SICAvar type II	3E-02	9E-07
A	VASP	1.1E-06	vesicle-mediated transport	putative AAA family ATPase; putative Heat shock protein 86 family; putative YT521-B-like family protein; putative Pyridoxal 5'-phosphate dependent enzyme class III; putative Mitochondrial import inner membrane translocase subunit TIM17; putative FAD synthetase; KIR protein; putative RNA binding protein; putative Heat shock protein 90; putative Calpain; ;	putative ATPase subunit 9; utative U6 snRNA-associated Sm-like protein LSm8; putative Syntaxin; putative Phosphoethanolamine N-methyltransferase; Thioredoxin; putative AP-1 complex subunit sigma; Peptidyl-prolyl cis-trans isomerase	3E-02	3E-04

A	UBA2	2.0E-04	ribosome biogenesis	putative ATPase subunit 9; putative U6 snRNA-associated Sm-like protein LSm8; putative Phosphoethanolamine N-methyltransferase; Proteasome endopeptidase complex; putative AP-1 complex subunit sigma; Peptidyl-prolyl cis-trans isomerase	putative Mannose-6-phosphate isomerase; putative AAA family ATPase; putative Heat shock protein 86 family; 40S ribosomal protein SA; Glycerol-3-phosphate dehydrogenase; putative Leucine aminopeptidase; KIR protein; putative Calpain	1E-01	1E-03
A	DPP8	1.0E+00	negative regulation of vasoconstriction			1E-01	7E-02
A	KDM4B	1.0E-01	sucrose metabolic process			1E-01	2E-02
A	NTHL1	3.4E-02	mitochondrion			1E-01	3E-02
A	MRPL10	9.0E-01	inner membrane			3E-01	3E-02
A	RPS19	5.2E-42	SRP-dependent cotranslational protein targeting to membrane	SICAvar type I; Tryptophan-rich antigen; Nucleoside diphosphate kinase	Thioredoxin-like protein; putative ValS protein; putative Ornithine aminotransferase; putative Coatomer alpha subunit; putative Chaperone; putative ATP-dependent zinc metalloprotease FTSH 1	3E-01	3E-05
A	ZC3H3	1.0E+00	positive regulation of histone deacetylation			4E-01	5E-02
A	CD24	7.4E-03	extracellular matrix			8E-01	3E-02
A	COX6C	2.9E-21	ribonucleoprotein complex	putative 50S ribosomal protein L3 - apicoplast; putative Ubiquitin-conjugating enzyme E2; Ubiquitin carboxyl-terminal hydrolase; Peptidyl-prolyl cis-trans isomerase		8E-01	2E-02
A	NR1H3	2.4E-06	cell activation involved in immune response	putative Acetyltransferase; putative CorA-like Mg2+ transporter protein; putative Ubiquitin conjugating enzyme; putative U6 snRNA-associated Sm-like protein LSm2; putative Mrna cleavage factor-like protein; putative Sortilin; putative Long-chain fatty acid CoA ligase	SICAvar type I; KIR-like protein	8E-01	2E-05
A	UBE2N	4.1E-02	protein-containing complex			8E-01	8E-01
A	TMCC1	3.8E-03	cytoplasmic vesicle			1E+00	1E+00

Appendix Table C.2.5- Co-expression modules obtained from WGCNA with their functional annotation and most correlated P. knowlesi proteins for modules of the host (Mm) (see Section 3.1.3 of the Text)

Type	Module name/ hub genes	P-value (fdr)	Term Ontology	Term Name	Pk proteins (+ ively correlated to module)
B	MPV17L	2E-15	BP	cotranslational protein targeting to membrane	putative Inner membrane complex protein
B	RAB11A	2E-12	BP	innate immune response	putative AP2 family; putative Kinesin-5; putative Vesicle transport v-SNARE protein VTI1; putative Vacuolar protein sorting-associated protein 18
B	COBRA1	2E-10	CC	mitochondrial matrix	putative Zinc finger protein
B	HPD	2E-08	BP	cell division	putative Polyubiquitin binding protein; putative Origin recognition complex subunit 2; Flap endonuclease 1; putative GDP dissociation inhibitor; putative Prohibitin-like protein; Rhomboid-like protein; Proteasome subunit alpha type; putative Leucine carboxyl
A	CCL1	4E-07	BP	immune response	
A	TRIL	7E-03	MF	unfolded protein binding	
B	LRRIQ3	6E-02	MF	immunoglobulin binding	KIR protein
B	AIF1L	6E-02	MF	ubiquitin-protein transferase activity	putative Ham1 family protein; putative Prolyl 4-hydroxylase subunit alpha; Transcription initiation TFIID-like; putative Tryptophan-rich antigen
A	COX5B	1E-01	BP	mRNA processing	
B	MGST1	2E-01	BP	lipid metabolic process	putative Coatomer epsilon subunit; putative Reductase; putative Glutathione synthetase; putative AAA family ATPase; putative Peptide chain release factor; putative AP-3 complex subunit delta
B	BLOC1S2	2E-01	MF	C-C chemokine receptor activity	putative Mitochondrial ribosomal protein S18; putative Serine--tRNA ligase
A	EXOC2	2E-01	BP	transport	
B	WDR47	2E-01	CC	ribonucleoprotein complex	putative Mitotic-spindle organizing protein 1; putative Endopeptidase
B	FASTKD1	3E-01	CC	NatA complex	putative EKC/KEOPS complex subunit BUD32; putative AAA family ATPase ; putative Calcium-dependent protein kinase 3 ; DNA helicase; putative Cysteine repeat modular protein 2; Prefoldin subunit 3
B	BLCAP	4E-01	BP	excretion	
C	PKNOH_S01017300	4E-01	BP	G protein-coupled receptor signaling pathway involved in heart process	
B	PKNOH_S07466700	5E-01	BP	macrophage activation	

B	TTC12	6E-01	BP	gene expression	
A	NUPR1	6E-01	MF	endoribonuclease activity	
B	DLGAP3	7E-01	BP	pentose metabolic process	
B	PKNOH_S09531300	9E-01	BP	response to bacterium	
A	TUB	1E+00	BP	spontaneous synaptic transmission	putative SNARE protein
B	SNIP1	1E+00	BP	positive regulation of cellular senescence	Flap endonuclease 1; putative GDP dissociation inhibitor; putative Cdc2-related protein kinase 1
A	ZNF449	1E+00	BP	spermatogonial cell division	putative DnaJ protein
B	PKNOH_S130192200	1E+00	CC	nuclear pericentric heterochromatin	
B	COPS8	1E+00	BP	response to glycoside	putative RAP protein; Serine/threonine-protein phosphatase; putative Dynein heavy chain
B	ZNF821	1E+00	MF	protein-disulfide reductase activity	putative Liver specific protein 2; putative Inner membrane complex protein
B	ADRBK2	1E+00	CC	motile cilium	putative Dynactin subunit 6
A	GORASP2	1E+00	BP	positive regulation of histone H3-K9 methylation	
A	NKIRAS1	1E+00	CC	male germ cell nucleus	
B	NDFIP2	1E+00	CC	ribosome	putative Bifunctional dihydrofolate reductase-thymidylate synthase; putative ATP-dependent Clp protease subunit; Coatomer subunit beta'
B	NAA30	1E+00	BP	positive regulation of extracellular matrix organization	SICAvar - type I; Cg4-like protein; 1-cys peroxiredoxin ; putative Gas41-like protein
B	FAM150B	1E+00	CC	integral component of membrane	
A	PCYOX1	1E+00	BP	adenohypophysis development	putative Vesicle transport v-SNARE protein VTI1

Appendix Table C.2.6- Blast results showing protein names and accession ids for P. knowlesi transcripts that are highly correlated to the host (Mm).

Pk genes	OTN/GI		Pk genes	OTN/GI
Uncharacterized protein PKNOH_S01023100 [Plasmodium knowlesi]	OTN68692.1 GI:1192742004		putative Ham1 family protein [Plasmodium knowlesi]	OTN68604.1 GI:1192741915
Uncharacterized protein PKNOH_S02306300 [Plasmodium knowlesi]	OTN68482.1 GI:1192741793		putative Acid phosphatase [Plasmodium knowlesi]	OTN67757.1 GI:1192741065
Uncharacterized protein PKNOH_S03332700 [Plasmodium knowlesi]	OTN68261.1 GI:1192741571		putative Asparagine--t RNA ligase [Plasmodium knowlesi]	OTN67225.1 GI:1192740531
Uncharacterized protein PKNOH_S09524600 [Plasmodium knowlesi]	OTN66298.1 GI:1192739602		putative Nucleotide binding protein [Plasmodium knowlesi]	OTN66953.1 GI:1192740259
Uncharacterized protein PKNOH_S09548800 [Plasmodium knowlesi]	OTN66225.1 GI:1192739529		SICAvar type I [Plasmodium knowlesi]	OTN66500.1 GI:1192739804
Uncharacterized protein PKNOH_S110080700 [Plasmodium knowlesi]	OTN65297.1 GI:1192738599		putative GTP-binding protein [Plasmodium knowlesi]	OTN65385.1 GI:1192738687
Uncharacterized protein PKNOH_S120162500 [Plasmodium knowlesi]	OTN65181.1 GI:1192738482		putative Calcium/potassium channel [Plasmodium knowlesi]	OTN65267.1 GI:1192738569
Uncharacterized protein PKNOH_S130207800 [Plasmodium knowlesi]	OTN64462.1 GI:1192737762		putative Acyl-CoA synthetase [Plasmodium knowlesi]	OTN64580.1 GI:1192737880
Uncharacterized protein PKNOH_S140237800 [Plasmodium knowlesi]	OTN64060.1 GI:1192737359		putative DNAJ protein [Plasmodium knowlesi]	OTN63779.1 GI:1192737078

Appendix Table C.2.7- List of key host genes (Mm) which are most highly correlated with pathogen transcripts.

Genes					
NEK6	IL12B	TMEM33	HMMR	SLC7A11	DARS1
PSMA6	AUNIP	HBS1L	ORC5	MGST2	MME
PSMD14	RMI1	PAIP1	CNEP1R1	VEGFA	ITGB1BP1
TGFA	RAD54B	DDX6	CEP97	ATP7A	FBXL13
CENPE	SASS6	CLN5	CENPK	S100A12	FBXO33
BRCA2	NUP43	EIF5A2	CDKN2B	IL15	AMN1
CCNA2	GADD45B	CHAC2	PHLDA1	CXCL6	ZYG11B
SGO2	BIRC3	MRPL42	PBK	NT5E	LYVE1
SPC25	RGS2	MRPL35	SKA1	IER3	ERI1
TACC1	KLHL42	MRPL33	USP3	CD160	TRDMT1
FBXL22	PLD6	SRP9	IFNG	SLC38A2	ACOT13

Appendix Table C.2.8- Co-expression modules obtained from WGCNA with their functional annotation and most correlated P. knowlesi proteins for modules of the host (Mf) (see Section 3.1.3 of the Text)

Type	Module name/ hub genes	P-value (fdr)	Term Ontology	Term Name	Pk proteins (+ ively correlated to module)
C	PKNOH_S120121300	0.06	BP	regulation of respiratory gaseous exchange	
A	PDXDC1	0.14	BP	aspartate metabolic process	
A	STAT4	0.30	BP	musculoskeletal movement	
A	HAUS8	0.30	BP	myoblast differentiation	putative FAD synthetase; putative Anaphase promoting complex subunit 10
A	MGST3	0.33	MF	JAK pathway signal transduction adaptor activity	
A	BLK	0.78	BP	hematopoietic stem cell differentiation	
A	RLTPR	1.00	BP	positive regulation of protein localization to nucleus	
A	GIMAP7	1.00	BP	adrenal gland development	
C	PKNOH_S09550100	1.00	CC	rough endoplasmic reticulum membrane	
A	ADRB1	1.00	BP	mitochondrial electron transport, succinate to ubiquinone	
A	FBN2	1.00	BP	positive regulation of cytokine production involved in inflammatory response	
A	OPRL1	1.00	CC	cell-cell adherens junction	
A	ZNF671	1.00	BP	aggresome assembly	
A	BLOC1S2	1.00	BP	peptide metabolic process	
B	NDUFB7	1.00	MF	hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds, in linear amidines	putative Mitochondrial import inner membrane translocase subunit TIM17; putative RNA binding protein; putative Ornithine aminotransferase

B	GPR126	1.00	BP	primitive streak formation	SICAvar type I; putative Subunit of proteaseome activator complex;
A	DPEP1	1.00	BP	response to stress	
A	TFRC	1.00	MF	stearoyl-CoA 9-desaturase activity	putative Regulator of chromosome condensation; putative Regulator of chromosome condensation
A	DUSP1	1.00	BP	secretory granule organization	putative Acetyltransferase; putative Gamete release protein; putative Cysteine repeat modular protein 2; putative 3'-5' exoribonuclease Csl4-like protein; putative S-adenosyl-L-methionine-dependent methyltransferase
A	PRSS36	1.00	MF	small protein activating enzyme binding	
A	SAMD10	1.00	BP	cellular response to UV	Casein kinase II subunit beta
A	APH1A	1.00	BP	phosphatidylcholine biosynthetic process	
A	ATP6V1G1	1.00	BP	exogenous drug catabolic process	putative Acetyltransferase; Proteasome subunit beta type; putative Regulator of chromosome condensation; putative Apical sushi protein
A	HSD17B6	1.00	CC	ruffle membrane	putative Coatomer epsilon subunit ; putative Regulator of chromosome condensation; Cg8-like protein
A	MBD4	1.00	BP	response to organic cyclic compound	
A	ISCA1	1.00	MF	rRNA methyltransferase activity	KIR protein; KIR-like protein
A	LRP1	1.00	BP	inhibition of cysteine-type endopeptidase activity involved in apoptotic process	putative Regulator of chromosome condensation; putative Importin-7; putative Rhoptry-associated membrane antigen; Cg8-like protein;
A	TMEM164	1.00	BP	intrinsic apoptotic signaling pathway in response to DNA damage	SICAvar type I; putative FAD synthetase
A	MASTL	1.00	MF	steroid dehydrogenase activity	putative Rhoptry-associated membrane antigen; Cg8-like protein; Proteasome subunit beta type

Appendix Table C.2.9- Blast results showing protein names and accession ids for P. knowlesi transcripts that are highly correlated to the host (Mf).

Pk genes	OTN/GI		Pk genes	OTN/GI
Uncharacterized protein PKNOH_S03323600 [Plasmodium knowlesi]	OTN68194.1 GI:1192741504		putative Ras-related protein Rab-6 [Plasmodium knowlesi]	OTN65018.1 GI:1192738319
putative Gluatamate dehydrogenase [Plasmodium knowlesi]	OTN68647.1 GI:1192741959		putative High mobility group protein B4 [Plasmodium knowlesi]	OTN64682.1 GI:1192737982
Heme/steroid binding domain containing protein [Plasmodium knowlesi]	OTN67690.1 GI:1192740998		Superoxide dismutase [Plasmodium knowlesi]	OTN63956.1 GI:1192737255
Lysine--tRNA ligase [Plasmodium knowlesi]	OTN66324.1 GI:1192739628		putative Cop-coated vesicle membrane protein p24 [Plasmodium knowlesi]	OTN63947.1 GI:1192737246

Appendix Table C.2.10- List of key host genes (Mf) that are most highly correlated with pathogen transcripts.

Genes			
TAL1	NRG1	OCIAD2	EFEMP1
GP5	ADRA2A	PF4	GAL
GK5	TIMP3	PF4V1	FOSL1
RARRES1	GAS2L1	GPX3	MS4A8B
SERPINE1	GP1BB	LIPC	FAM20A
TFPI2	GP9	STON2	FAM69C
RHAG	IGFBP2		

Appendix Table C.2.11- Functional Annotation (DAVID[51]) of differentially expressed PRR genes in Mm and Mf showing specific aspects of PRR signaling pathway activated during log phase of infection.

Annotation Cluster 1	Enrichment Score: 9.62				Annotation Cluster 3	Enrichment Score: 2.528		
Category	**Term**	**Fold Enrichment**	**FDR**		**Category**	**Term**	**Fold Enrichment**	**FDR**
UP_KEYWORDS	Antiviral defense	81.3	2E-15		KEGG_PATHWAY	hsa05162:Measles	22.2	1E-04
UP_KEYWORDS	Immunity	22.3	2E-12		UP_SEQ_FEATURE	mutagenesis site	3.4	3E-01
UP_KEYWORDS	Innate immunity	36.1	2E-12					
GOTERM_BP_DIRECT	GO:0051607~defense response to virus	42.4	1E-10		Annotation Cluster 4	Enrichment Score: 2.082600416554552		
GOTERM_BP_DIRECT	GO:0009615~response to virus	57.2	2E-10		**Category**	**Term**	**Fold Enrichment**	**FDR**
GOTERM_BP_DIRECT	GO:0060337~type I interferon signaling pathway	87.5	2E-10		GOTERM_MF_DIRECT	GO:0003725~double-stranded RNA binding	57.7	9E-05
UP_KEYWORDS	Cytoplasm	2.3	4E-02		GOTERM_BP_DIRECT	GO:0045087~innate immune response	11.4	1E-03
UP_KEYWORDS	Nucleotide-binding	3.8	4E-02		GOTERM_MF_DIRECT	GO:0003727~single-stranded RNA binding	49.1	6E-02
INTERPRO	IPR027417:P-loop containing nucleoside triphosphate hydrolase	5.3	5E-02		INTERPRO	IPR027417:P-loop containing nucleoside triphosphate hydrolase	5.3	5E-02
					UP_KEYWORDS	RNA-binding	6.4	7E-02
Annotation Cluster 2	Enrichment Score: 3.00				GOTERM_MF_DIRECT	GO:0004386~helicase activity	24.8	9E-02
Category	**Term**	**Fold Enrichment**	**FDR**		UP_SEQ_FEATURE	domain:Helicase C-terminal	23.2	3E-01
KEGG_PATHWAY	hsa05164:Influenza A	19.8	2E-05		UP_SEQ_FEATURE	domain:Helicase ATP-binding	22.0	3E-01
KEGG_PATHWAY	hsa05162:Measles	22.2	1E-04		INTERPRO	IPR001650:Helicase, C-terminal	21.5	5E-02
KEGG_PATHWAY	hsa05161:Hepatitis B	13.6	5E-02		INTERPRO	IPR014001:Helicase, superfamily 1/2, ATP-binding domain	21.1	5E-02
KEGG_PATHWAY	hsa05168:Herpes simplex infection	10.7	8E-02		UP_KEYWORDS	Hydrolase	3.6	9E-02

Appendix Table C.2.12- Key differentially expressed PRR signaling pathway genes (DEGs) in Mm and Mf and differentially responding genes (DRGs) in Mm vs. Mf during the log phase of infection.

DEGs in Mm and Mf			
CTSK	RIPK2	TLR3	ITGAM
RSAD2	DDX58	TICAM2	NOD2
NMI	TLR4	IRF1	NR1D1
TANK	UBE2D1	HAVCR2	DHX58
IFIH1	PIK3AP1	FLOT1	IFI35
TIFA	IRF7	TNFAIP3	CD300A
ALPK1	TBK1	CAV1	RIOK3
TLR2	UBC	LY96	TICAM1
DDX60	NFKBIA	XIAP	LILRA2

DRGs: Mm vs. Mf		
BIRC2	NLRP6	UBA52
CAV1	NR1H3	UBE2D1
CD300A	PELI1	USP15
CD36	RPS27A	
EPG5	SLC15A3	
FFAR2	SLC15A4	
IRAK2	SLC46A2	
LACC1	TLR5	
LTF	TNIP3	

Appendix Table C.2.13- Reactome[117] pathways enriched by PRR related DEGs in Mm during log phase. (Right): Reactome pathways enriched by PRR related DEGs in Mf during log phase. The contraposition highlights the subtle differences between the PRR pathway of the two hosts along with key responsible genes.

Pathway name	FDR	Submitted entities found
Attenuation phase	3E-07	HSPA1A
HSF1-dependent transactivation	7E-07	HSPA1A
Toll-like Receptor Cascades	9E-07	TICAM2;RIPK2;UBE2D1;LY96;NOD2;TICAM1;TANK;NFKBIA;CTSK;IRF7;TLR5;TLR4;S100A9;TLR3;S100A8;LGMN;TLR2
Diseases of Immune System	3E-06	NFKBIA;LY96;TLR5;TICAM1;S100A9;TLR4;S100A8;TLR3;TLR2
Diseases associated with the TLR signaling cascade	3E-06	NFKBIA;LY96;TLR5;TICAM1;S100A9;TLR4;S100A8;TLR3;TLR2
IRAK4 deficiency (TLR2/4)	2E-05	LY96;S100A9;TLR4;S100A8;TLR2
IKK complex recruitment mediated by RIP1	6E-05	TICAM2;UBE2D1;LY96;TICAM1;TLR4
MyD88 deficiency (TLR2/4)	6E-05	LY96;S100A9;TLR4;S100A8;TLR2
Toll Like Receptor 4 (TLR4) Cascade	1E-04	TICAM2;RIPK2;UBE2D1;LY96;NOD2;TICAM1;TANK;NFKBIA;IRF7;TLR4;S100A9;S100A8;TLR2
Activation of IRF3/IRF7 mediated by TBK1/IKK epsilon	1E-04	TICAM2;IRF7;LY96;TICAM1;TLR4;TANK
Regulation of TLR by endogenous ligand	2E-04	LY96;S100A9;TLR4;S100A8;TLR2
Innate Immune System	5E-04	TNFAIP3;UBE2D1;LY96;NOD2;TANK;IFIH1;CTSK;DHX58;TIFA;TICAM2;DDX58;RIPK2;CD300A;TICAM1;NFKBIA;IRF7;ALPK1
Caspase activation via Death Receptors in the presence of ligand	8E-04	TICAM2;LY96;TICAM1;TLR4
MyD88-independent TLR4 cascade	9E-04	NFKBIA;TICAM2;RIPK2;IRF7;UBE2D1;LY96;NOD2;TICAM1;TLR4;TANK
TRIF(TICAM1)-mediated TLR4 signaling	9E-04	NFKBIA;TICAM2;RIPK2;IRF7;UBE2D1;LY96;NOD2;TICAM1;TLR4;TANK

Pathway name	FDR	Submitted entities found
TICAM1-dependent activation of IRF3/IRF7	6E-06	UBB;IRF7;TICAM1;TANK;TLR3
Toll-like Receptor Cascades	6E-05	FGB;TICAM2;UNC93B1;RIPK2;NOD1;TICAM1;TANK;NFKBIA;UBB;CTSK;IRF7;CD36;TLR3;MAP2K6
Activation of IRF3/IRF7 mediated by TBK1/IKK epsilon	6E-05	TICAM2;UBB;IRF7;TICAM1;TANK
TICAM1, RIP1-mediated IKK complex recruitment	1E-04	UBB;TICAM1;TLR3
IKK complex recruitment mediated by RIP1	2E-04	TICAM2;UBB;TICAM1
Ovarian tumor domain proteases	2E-04	IFIH1;UBB;DDX58;RIPK2;TNIP3;TNFAIP3;NOD1
FLT3 signaling by CBL mutants	2E-04	UBB
activated TAK1 mediates p38 MAPK activation	2E-04	UBB;RIPK2;NOD1;MAP2K6
Toll Like Receptor 3 (TLR3) Cascade	2E-04	NFKBIA;UBB;RIPK2;IRF7;NOD1;TICAM1;TANK;TLR3;MAP2K6
Toll Like Receptor 4 (TLR4) Cascade	3E-04	FGB;NFKBIA;TICAM2;UBB;RIPK2;IRF7;NOD1;CD36;TICAM1;TANK;MAP2K6
MyD88-independent TLR4 cascade	4E-04	NFKBIA;TICAM2;UBB;RIPK2;IRF7;NOD1;TICAM1;TANK;MAP2K6
TRIF(TICAM1)-mediated TLR4 signaling	4E-04	NFKBIA;TICAM2;UBB;RIPK2;IRF7;NOD1;TICAM1;TANK;MAP2K6
Myoclonic epilepsy of Lafora	5E-04	UBB
TICAM1,TRAF6-dependent induction of TAK1 complex	6E-04	UBB;TICAM1;TLR3
Alpha-protein kinase 1 signaling pathway	6E-04	UBB;TIFA;ALPK1

Appendix Table C.2.14- Cibersortx[124] results for deconvolution of various cell populations and their comparison across various cases. For example, panels A and B list changes in cell populations during log phase (TP4) from baseline for Mm and Mf.

(A) Mm TP4 v Baseline		
Cell types	**logFC**	**adj.P.Val**
Dendritic cells activated	0.00	5E-07
Mast cells activated	0.01	1E-06
Macrophages M1	0.00	9E-04
Monocytes	0.12	2E-03
T cells regulatory (Tregs)	0.04	3E-03
T cells CD8	-0.08	3E-03
Eosinophils	-0.01	2E-02
T cells CD4 memory activated	-0.05	3E-02
NK cells resting	-0.05	3E-02
Dendritic cells resting	0.00	8E-02

(B) Mf TP4 v Baseline		
Cell types	**logFC**	**adj.P.Val**
Macrophages M1	0.00	0.00
Monocytes	0.12	0.00
NK cells resting	-0.07	0.00
T cells CD8	-0.05	0.01
T cells CD4 memory activated	-0.04	0.01
T cells CD4 memory resting	0.02	0.01
Dendritic cells activated	0.00	0.01
Mast cells resting	0.01	0.02
Macrophages M2	-0.01	0.04
Plasma cells	0.00	0.17

(C) Baseline Mm vs Baseline Mf		
Cell types	**logFC**	**adj.P.Val**
T cells CD4 naive	-0.15	0.00
Neutrophils	0.14	0.00
Macrophages M0	0.04	0.00
Eosinophils	-0.01	0.00
T cells regulatory (Tregs)	0.03	0.00
B cells memory	-0.05	0.01
Macrophages M2	0.01	0.04
B cells naive	-0.01	0.05
Mast cells resting	0.01	0.21
T cells CD4 memory activated	-0.01	0.63

(D) TP4 Mm vs TP4 Mf		
Cell types	**logFC**	**adj.P.Val**
T cells CD4 naive	-0.14	0.00
Mast cells activated	-0.01	0.00
Dendritic cells activated	0.00	0.00
Macrophages M0	0.05	0.00
Mast cells resting	0.02	0.02
T cells CD4 memory resting	0.03	0.02
Neutrophils	0.08	0.22
T cells CD8	0.03	0.47
Dendritic cells resting	0.00	0.56
B cells memory	-0.03	0.59

(E) DR cell populations at TP4 - Mm vs Mf		
Cell types	**logFC**	**adj.P.Val**
Mast cells activated	-0.01	0.00
Dendritic cells activated	0.00	0.01
T cells regulatory (Tregs)	-0.03	0.10
Eosinophils	0.01	0.16
T cells CD4 memory resting	0.02	0.26
Mast cells resting	0.01	0.48
Macrophages M2	-0.01	0.48
Macrophages M0	0.02	0.62
Dendritic cells resting	0.00	0.68
Neutrophils	-0.05	0.70

(F) Mf TP5 v TP4		
Cell types	**logFC**	**adj.P.Val**
T cells follicular helper	0.00	0.01
T cells CD4 memory activated	0.06	0.01
Macrophages M1	0.00	0.09
NK cells resting	0.04	0.10
Neutrophils	-0.10	0.10
Macrophages M0	-0.02	0.51
Monocytes	0.04	0.51
Mast cells resting	-0.01	0.51
B cells memory	-0.02	0.85
Plasma cells	0.00	0.92

Appendix Table C.2.15- Various ligands for AhR[92,93]

Group	Compound	
Indole Metabolites	Indole	Indole-3-carbinol
	Indolo[3,2-b]carbazole (ICZ)	3,3-diindoylmethane
	2-(Indol-3-ylmethyl)-3,3'-diindolylmethane (Ltr-1)	2-(1'H-indole-3'-carbonyl)-thiazole-4-carboxylic acid methyl ester (ITE)
	3,3'-Diindolylmethane (DIM)	
Tryptophan Metabolites	Kynurenine (Kyn)	5-hydroxy-tryptophan (5HTP)
	Kynyrenic acid (KA)	Tryptamine (TA)
	Zanthurenic acid	Indol-3-acetic Acid (IAA)
	Cinnabarinic acid (CA	3-methylindole (Skatole)
	6-Formylindolo[3,2-b]carbazole (FICZ)	Indole-3-aldehyde (IAld)
	Indoxyl-3-sulfate (I3S)	
Heme-derived	Bilirubin	
	Biliverdin	

Appendix Table C.2.16- (Left) Important target genes for AhR, HIF1α and both AhR &HIF1α reflecting the complexity of the mechanism and its outcome. (Right) List of target genes of AhR and HIF1α that affect important processes like the p53 pathway, heme metabolism, cell cycle and immune system responses, notably IFN-γ and NFκB signaling.

AHR targets	HIF1A targets	AHR and HIF1A targets
TP53	HK2	RUNX1
WRAP53	PKM	PPP1R13L
KAT7	PGAM1	HES1
PARK7	TPI1	BHLHE40
FLCN	HK1	RARA
CDKN1A	ALDOA	ENO1
PIK3R1	GAPDH	UBC
PRKCD	ALDOC	CBX5
JUP	GPI	CGGBP1
FYN	ENO2	DRAP1
TARDBP	PFKP	DEAF1
HYAL2	PFKL	SP2
CDH1	PGM1	CDKN1B
PKIA	PFKFB3	RUVBL2
PKIG	PFKFB4	DOT1L
LIMK2	LDHA	MSH3
CCT5	TPR	EAPP
EIF2AK3	NUP58	MIDEAS
TRIM8	POM121	NMNAT1
CARD10	PDK1	NOCT
ABCA7	PDK3	DDIT4
TSC2	AK4	SLX1A
BMP7	NME1	SLX1B
PIN1	AK2	ZC3H10
RANBP2	NUDT18	MTRNR2L1
LMNA	GUK1	MTRNR2L2
CNEP1R1	RORA	MTRNR2L8
BCL3	RAN	MTRNR2L9
CTNNA1	KDM1A	PSMA1
POLR1A	SHMT2	APOLD1
STAT3	DNAJC30	SOS1
PPP3CA	ATP5PO	SLC6A6
SUN1	PPAT	YWHAG
IPO11	NT5C	INHA
FAM53C	HMGCL	LAG3
TAF8	ADSS1	OBSL1
SIRT7	PAICS	FAXDC2
MCM2	SCD	WDFY2
EME2	GNAI3	NOL3
BRIP1	OXSM	BNIP3L
CDC45	ACOT6	PGK1
FBH1	HAGH	IER3
MCM4	VDAC1	SULT1A4
MCM5	SLC16A3	SULT1A3
NUCKS1	SLC2A1	AP5S1
GINS4	PPP1CB	BOLA2
GINS2	BCKDK	BOLA2B
RECQL	PPP1R3E	IBA57
RMI2	PPP1R3C	CHSY1
PNKP	PPP4R3B	DUX4

P53 pathway	Heme metabolism	Cell cycle	Interferon gamma response	NFKB signaling
TPD52L1	HEBP1	AURKA	TRAFD1	NFE2L2
RXRA	SLC22A4	CDC25C	NCOA3	BCL3
TRAFD1	NFE2L1	CDC14B	CDKN1A	REL
TP63	SLC30A1	TPR	IRF8	DUSP2
RAB40C	DAAM1	XRCC3	STAT1	FOS
FOS	KLF1	FBXO5	OASL	TIPARP
SP1	CDR2	CDK5RAP2	SLC25A28	TGIF1
PPP1R15A	RCL1	LCMT1	STAT3	PPP1R15A
VDR	DCAF11	BUB1B	IRF5	LDLR
H1-2	KHNYN	MAD1L1	BTG1	TRIP10
GADD45A	FOXJ2	KNTC1	RNF213	GADD45A
SLC19A2	PC	FZR1	RAPGEF6	CDKN1A
MAPKAPK3	UROS	CDK2	NAMPT	BCL6
CDKN1A	TNRC6B	CDKN1A	PFKP	ICOSLG
ABCC5	CCND3	TP53	SRI	MXD1
TP53	CTNS	GADD45A	RBCK1	TNFAIP8
BAIAP2	TNS1	SFN	ISG15	PER1
CCNK	BLVRB	CDKN1B	ARL4A	BTG1
CCND3	BNIP3L	E2F4	UPP1	RCAN1
MXD1	NARF	PRMT1		FOSL2
ABAT	TFRC	CNOT1		IER3
BTG1	P4HA2	CNOT8		HES1
COQ8A	SLC2A1	CNOT3		BHLHE40
SFN	CPOX	CNOT10		PNRC1
STOM	MXI1	CDC73		PFKFB3
PMM1	RIOK3	OVOL1		TNIP1
DDIT3	HAGH	WEE1		NAMPT
IER3	ABCB6	PRKDC		KDM6B
DDIT4	RNF19A	FBXO31		SERPINE1
FAM162A	PPP2R5B	TAOK1		NFKB2
POM121	SPTB	TAOK3		KLF10
H2AJ		NOP53		MAFF
HDAC3		RBL1		RELA
NDRG1		KANK2		EFNA1
HMOX1		GJC2		NFIL3
EI24		PSMC3		
UPP1		PSMC4		
		PSMA1		
		RBX1		
		TPRA1		
		TFAP4		
		BABAM1		
		CDC23		
		RAD51B		
		CENPE		
		NEK6		
		HECW2		
		KMT2E		
		PKD1		
		ADAM17		
		CCND3		
		CYP1A1		

REFERENCES

1 World malaria report 2020: 20 years of global progress and challenges. *Geneva: World Health Organization* (2020).

2 Harper, K. & Armelagos, G. The changing disease-scape in the third epidemiological transition. *Int J Environ Res Public Health* **7**, 675-697, doi:10.3390/ijerph7020675 (2010).

3 Sharp, P. M., Plenderleith, L. J. & Hahn, B. H. Ape Origins of Human Malaria. *Annu Rev Microbiol* **74**, 39-63, doi:10.1146/annurev-micro-020518-115628 (2020).

4 Division of Parasitic Diseases and Malaria (CDC). *The History of Malaria, an Ancient Disease*, <https://www.cdc.gov/malaria/about/history/index.html> (2017).

5 Karunamoorthi, K. Malaria vaccine: a future hope to curtail the global malaria burden. *Int J Prev Med* **5**, 529-538 (2014).

6 Kirchner, S., Power, B. J. & Waters, A. P. Recent advances in malaria genomics and epigenomics. *Genome Med* **8**, 92, doi:10.1186/s13073-016-0343-7 (2016).

7 Su, X. Z., Lane, K. D., Xia, L., Sá, J. M. & Wellems, T. E. *Plasmodium* Genomics and Genetics: New Insights into Malaria Pathogenesis, Drug Resistance, Epidemiology, and Evolution. *Clin Microbiol Rev* **32**, doi:10.1128/cmr.00019-19 (2019).

8 Raja, T. N. *et al.* Naturally Acquired Human *Plasmodium cynomolgi* and *P. knowlesi* Infections, Malaysian Borneo. *Emerging Infectious Disease journal* **26**, 1801, doi:10.3201/eid2608.200343 (2020).

9 Sato, S. *Plasmodium*—a brief introduction to the parasites causing human malaria and their basic biology. *Journal of Physiological Anthropology* **40**, 1, doi:10.1186/s40101-020-00251-9 (2021).

10 Galinski, M. R., Meyer, E. V. S. & Barnwell, J. W. in *Advances in Parasitology* Vol. 81 (eds S. I. Hay, Ric Price, & J. Kevin Baird) 1-26 (Academic Press, 2013).

11 Greenwood, B. M. *et al.* Malaria: progress, perils, and prospects for eradication. *J Clin Invest* **118**, 1266-1276, doi:10.1172/JCI33996 (2008).

12 Florens, L. *et al.* A proteomic view of the *Plasmodium falciparum* life cycle. *Nature* **419**, 520-526, doi:10.1038/nature01107 (2002).

13 Singh, B. & Daneshvar, C. Human Infections and Detection of *Plasmodium knowlesi*. *Clin Microbiol Rev* **26**, 165-184, doi:doi:10.1128/CMR.00079-12 (2013).

14 Lee, W. C. *et al.* Hyperparasitaemic human *Plasmodium knowlesi* infection with atypical morphology in peninsular Malaysia. *Malar J* **12**, 88, doi:10.1186/1475-2875-12-88 (2013).

15 Barber, B. E., William, T., Grigg, M. J., Yeo, T. W. & Anstey, N. M. Limitations of microscopy to differentiate *Plasmodium* species in a region co-endemic for *Plasmodium falciparum*, *Plasmodium vivax* and *Plasmodium knowlesi*. *Malar J* **12**, 8, doi:10.1186/1475-2875-12-8 (2013).

16 Singh, B. *et al.* A large focus of naturally acquired *Plasmodium knowlesi* infections in human beings. *Lancet* **363**, 1017-1024, doi:10.1016/s0140-6736(04)15836-4 (2004).

17 Lee, K. S. *et al. Plasmodium knowlesi*: reservoir hosts and tracking the emergence in humans and macaques. *PLoS Pathog* **7**, e1002015, doi:10.1371/journal.ppat.1002015 (2011).

18 Butcher, G. A. & Mitchell, G. H. The role of *Plasmodium knowlesi* in the history of malaria research. *Parasitology* **145**, 6-17, doi:10.1017/s0031182016001888 (2018).

19 Millar, S. B. & Cox-Singh, J. Human infections with *Plasmodium knowlesi*—zoonotic malaria. *Clinical Microbiology and Infection* **21**, 640-648, doi:https://doi.org/10.1016/j.cmi.2015.03.017 (2015).

20 Pasini, E. M., Zeeman, A. M., Voorberg, V. A. N. D. E. R. W. A. & Kocken, C. H. M. *Plasmodium knowlesi*: a relevant, versatile experimental malaria model. *Parasitology* **145**, 56-70, doi:10.1017/s0031182016002286 (2018).

21 de Koning-Ward, T. F., Gilson, P. R. & Crabb, B. S. Advances in molecular genetic systems in malaria. *Nature Reviews Microbiology* **13**, 373-387, doi:10.1038/nrmicro3450 (2015).

22 Long, C. A. & Zavala, F. Immune Responses in Malaria. *Cold Spring Harb Perspect Med* **7**, a025577, doi:10.1101/cshperspect.a025577 (2017).

23 Woodford, J. *et al.* An Experimental Human Blood-Stage Model for Studying *Plasmodium malariae* Infection. *The Journal of Infectious Diseases* **221**, 948-955, doi:10.1093/infdis/jiz102 (2019).

24 McCarthy, J. S. *et al.* Experimentally induced blood-stage *Plasmodium vivax* infection in healthy volunteers. *The Journal of infectious diseases* **208**, 1688-1694, doi:10.1093/infdis/jit394 (2013).

25 McCarthy, J. S. *et al.* A pilot randomised trial of induced blood-stage *Plasmodium falciparum* infections in healthy volunteers for testing efficacy of new antimalarial drugs. *PLoS One* **6**, e21914-e21914, doi:10.1371/journal.pone.0021914 (2011).

26 Tumbo, A.-M. *et al.* Role of human *Pegivirus* infections in whole *Plasmodium falciparum* sporozoite vaccination and controlled human malaria infection in African volunteers. *Virology Journal* **18**, 28, doi:10.1186/s12985-021-01500-8 (2021).

27 Milne, K. *et al.* Mapping immune variation and var gene switching in naive hosts infected with *Plasmodium falciparum*. *Elife* **10**, doi:10.7554/eLife.62800 (2021).

28 Yap, X. Z., McCall, M. B. B. & Sauerwein, R. W. Fast and fierce versus slow and smooth: Heterogeneity in immune responses to *Plasmodium* in the controlled human malaria infection model. *Immunol Rev.* **293**, 253-269, doi:https://doi.org/10.1111/imr.12811 (2020).

29 Hickey, B. *et al.* IMRAS—A clinical trial of mosquito-bite immunization with live, radiation-attenuated *P. falciparum* sporozoites: Impact of immunization parameters on protective efficacy and generation of a repository of immunologic reagents. *PLoS One* **15**, e0233840, doi:10.1371/journal.pone.0233840 (2020).

30 Hoo, R. *et al.* Transcriptome profiling reveals functional variation in *Plasmodium falciparum* parasites from controlled human malaria infection studies. *EBioMedicine* **48**, 442-452, doi:https://doi.org/10.1016/j.ebiom.2019.09.001 (2019).

31 Stanisic, D. I., McCarthy, J. S. & Good, M. F. Controlled Human Malaria Infection: Applications, Advances, and Challenges. *Infection and Immunity* **86**, e00479-00417, doi:10.1128/IAI.00479-17 (2018).

32 van Wolfswinkel, M. E. *et al.* Changes in total and differential leukocyte counts during the clinically silent liver phase in a controlled human malaria infection in malaria-naïve Dutch volunteers. *Malar J* **16**, 457, doi:10.1186/s12936-017-2108-1 (2017).

33 Payne, R. O., Griffin, P. M., McCarthy, J. S. & Draper, S. J. *Plasmodium* vivax Controlled Human Malaria Infection – Progress and Prospects. *Trends in Parasitology* **33**, 141-150, doi:https://doi.org/10.1016/j.pt.2016.11.001 (2017).

34 Sauerwein, R. W., Roestenberg, M. & Moorthy, V. S. Experimental human challenge infections can accelerate clinical malaria vaccine development. *Nature Reviews Immunology* **11**, 57-64, doi:10.1038/nri2902 (2011).

35 Jao, I. *et al.* Deliberately infecting healthy volunteers with malaria parasites: Perceptions and experiences of participants and other stakeholders in a Kenyan-based malaria infection study. *Bioethics* **34**, 819-832, doi:https://doi.org/10.1111/bioe.12781 (2020).

36 Minkah, N. K., Schafer, C. & Kappe, S. H. I. Humanized Mouse Models for the Study of Human Malaria Parasite Biology, Pathogenesis, and Immunity. *Front Immunol.* **9** doi:10.3389/fimmu.2018.00807 (2018).

37 De Niz, M. & Heussler, V. T. Rodent malaria models: insights into human disease and parasite biology. *Current opinion in microbiology* **46**, 93-101, doi:10.1016/j.mib.2018.09.003 (2018).

38 Langhorne, J. *et al.* The relevance of non-human primate and rodent malaria models for humans. *Malar J* **10**, 23, doi:10.1186/1475-2875-10-23 (2011).

39 Joyner, C., Barnwell, J. W. & Galinski, M. R. No more monkeying around: primate malaria model systems are key to understanding *Plasmodium vivax* liver-stage biology, hypnozoites, and relapses. *Front Microbiol* **6** 145, doi:10.3389/fmicb.2015.00145 (2015).

40 Craig, A. G. *et al.* The role of animal models for research on severe malaria. *PLoS Pathog* **8**, e1002401, doi:10.1371/journal.ppat.1002401 (2012).

41 Gardner, M. B. & Luciw, P. A. Macaque Models of Human Infectious Disease. *ILAR Journal* **49**, 220-255, doi:10.1093/ilar.49.2.220 %J ILAR Journal (2008).

42 Aikawa, M. *et al.* A primate model for human cerebral malaria: *Plasmodium coatneyi*-infected rhesus monkeys. *Am J Trop Med Hyg* **46**, 391-397 (1992).

43 Salguero, F. J. *et al.* Comparison of rhesus and cynomolgus macaques as an infection model for COVID-19. *Nature Communications* **12**, 1260, doi:10.1038/s41467-021-21389-9 (2021).

44 Shedlock, D. J., Silvestri, G. & Weiner, D. B. Monkeying around with HIV vaccines: using rhesus macaques to define 'gatekeepers' for clinical trials. *Nat Rev Immunol* **9**, 717-728, doi:10.1038/nri2636 (2009).

45 Coatney, G. R., Collins, W. E., Warren, M. & Contacos, P. G. *The Primate Malarias, e-book [original book published in 1971]* Vol. 381 (Division of Parasitic Diseases, Centers for Disease Control and Protection, 2003).

46 Coatney, G. R. *The Primate Malarias*. (U.S. National Institute of Allergy and Infectious Diseases, 1971).

47 Knowles, R. & Gupta, B. M. D. A Study of Monkey-Malaria, and Its Experimental Transmission to Man. *Ind Med Gaz* **67**, 301-320 (1932).

48 *phyloT : a tree generator*, <https://phylot.biobyte.de/> (2020).

49 *Primates*, <https://ori.hhs.gov/education/products/ncstate/primate.htm> (2021).

50 Kanthaswamy, S. *et al.* Interspecies hybridization and the stratification of nuclear genetic variation of rhesus (*Macaca mulatta*) and long-tailed macaques (*Macaca fascicularis*). *Int J Primatol* **29**, 1295-1311, doi:10.1007/s10764-008-9295-0 (2008).

51 Butcher, G. A., Mitchell, G. H. & Cohen, S. *Plasmodium knowlesi* infections in a small number of non-immune natural hosts (*Macaca fascicularis*) and in rhesus monkeys (*M. mulatta*). *Trans R Soc Trop Med Hyg* **104**, 75-77, doi:10.1016/j.trstmh.2009.05.017 (2010).

52 Butcher, G. A. Models for malaria: Nature knows best. *Parasitology today (Personal ed.)* **12**, 378-382, doi:10.1016/0169-4758(96)10062-4 (1996).

53 Gupta, A., Styczynski, M. P., Galinski, M. R., Voit, E. O. & Fonseca, L. L. Dramatic transcriptomic differences in *Macaca mulatta* and *Macaca fascicularis* with *Plasmodium knowlesi* infections. *Scientific Reports* **11**, 19519, doi:10.1038/s41598-021-98024-6 (2021).

54 Pinski, A. N., Maroney, K. J., Marzi, A. & Messaoudi, I. Distinct transcriptional responses to fatal Ebola virus infection in cynomolgus and rhesus macaques suggest species-specific immune responses. *Emerg Microbes Infect* **10**, 1320-1330, doi:10.1080/22221751.2021.1942229 (2021).

55 El Mubarak, H. S. *et al.* Infection of cynomolgus macaques (*Macaca fascicularis*) and rhesus macaques (Macaca mulatta) with different wild-type measles viruses. *J Gen Virol* **88**, 2028-2034, doi:10.1099/vir.0.82804-0 (2007).

56 Waag, D. M. *et al.* Evaluation of cynomolgus (*Macaca fascicularis*) and rhesus (*Macaca mulatta*) monkeys as experimental models of acute Q fever after aerosol exposure to phase-I Coxiella burnetii. *Lab Anim Sci* **49**, 634-638 (1999).

57 Street, S. L., Kyes, R. C., Grant, R. & Ferguson, B. Single nucleotide polymorphisms (SNPs) are highly conserved in rhesus (Macaca mulatta) and cynomolgus (Macaca fascicularis) macaques. *BMC Genomics* **8**, 480, doi:10.1186/1471-2164-8-480 (2007).

58 Moyes, C. L. *et al.* Defining the Geographical Range of the Plasmodium knowlesi Reservoir. *PLoS Neglected Tropical Diseases* **8**, e2780, doi:10.1371/journal.pntd.0002780 (2014).

59 Singh, B. & Daneshvar, C. Human Infections and Detection of Plasmodium knowlesi. *Clin Microbiol Rev.* **26**, 165-184, doi:doi:10.1128/CMR.00079-12 (2013).

60 *Malaria Host Pathogen Interaction Center (MaHPIC Consortium)*, <http://www.systemsbiology.emory.edu/research/Public%20Data%20Releases/index.html> (2020).

61 DeBarry, J. D. *et al.* Practical Recommendations for Supporting a Systems Biology Cyberinfrastructure. *Data Science Journal* **19**, doi:10.5334/dsj-2020-024 (2020).

62 Joyner, C. J. *et al.* Humoral immunity prevents clinical malaria during *Plasmodium* relapses without eliminating gametocytes. *PLoS Pathog* **15**, e1007974, doi:10.1371/journal.ppat.1007974 (2019).

63 Cordy, R. J. *et al.* Distinct amino acid and lipid perturbations characterize acute versus chronic malaria. *JCI Insight* **4** doi:10.1172/jci.insight.125156 (2019).

64 Tang, Y. *et al.* Integrative analysis associates monocytes with insufficient erythropoiesis during acute *Plasmodium cynomolgi* malaria in rhesus macaques. *Malar J* **16**, 384, doi:10.1186/s12936-017-2029-z (2017).

65 Joyner, C. J. *et al.* Case Report: Severe and Complicated Cynomolgi Malaria in a Rhesus Macaque Resulted in Similar Histopathological Changes as Those Seen in Human Malaria. *Am J Trop Med Hyg* **97**, 548-555, doi:10.4269/ajtmh.16-0742 (2017).

66 Joyner, C. *et al. Plasmodium cynomolgi* infections in rhesus macaques display clinical and parasitological features pertinent to modelling vivax malaria pathology and relapse infections. *Malar J* **15**, 451, doi:10.1186/s12936-016-1480-6 (2016).

67 Lombardini, E. D., Gettayacamin, M., Turner, G. D. & Brown, A. E. A Review of *Plasmodium coatneyi*-Macaque models of severe malaria. *Veterinary pathology* **52**, 998-1011, doi:10.1177/0300985815583098 (2015).

68 Galinski, M. R. *et al. Plasmodium knowlesi*: a superb *in vivo* nonhuman primate model of antigenic variation in malaria. *Parasitology* **145**, 85-100, doi:10.1017/S0031182017001135 (2018).

69 Imwong, M. *et al.* Asymptomatic Natural Human Infections With the Simian Malaria Parasites *Plasmodium cynomolgi* and *Plasmodium knowlesi*. *J Infect Dis* **219**, 695-702, doi:10.1093/infdis/jiy519 (2019).

70 Singh, B. *et al.* Naturally acquired human infections with the simian malaria parasite, *Plasmodium cynomolgi*, in Sarawak, Malaysian Borneo. *International Journal of Infectious Diseases* **73**, 68, doi:10.1016/j.ijid.2018.04.3581 (2018).

71 Anstey, N. M. & Grigg, M. J. Zoonotic Malaria: The Better You Look, the More You Find. *The Journal of Infectious Diseases* **219**, 679-681, doi:10.1093/infdis/jiy520 (2018).

72 Ta, T. H. *et al.* First case of a naturally acquired human infection with *Plasmodium cynomolgi*. *Malar J* **13**, 68, doi:10.1186/1475-2875-13-68 (2014).

73 Karlsson, E. K., Kwiatkowski, D. P. & Sabeti, P. C. Natural selection and infectious disease in human populations. *Nat Rev Genet* **15**, 379-393, doi:10.1038/nrg3734 (2014).

74 Sabeti, P. Natural selection: uncovering mechanisms of evolutionary adaptation to infectious disease. *Nature Education* **1(1):13** (2008).

75 Lederberg, J. J. B. S. Haldane (1949) on Infectious Disease and Evolution. *Genetics* **153**, 1-3 (1999).

76 Allison, A. C. Protection afforded by sickle-cell trait against subtertian malareal infection. *Br Med J* **1**, 290-294, doi:10.1136/bmj.1.4857.290 (1954).

77 Kwiatkowski, D. P. How malaria has affected the human genome and what human genetics can teach us about malaria. *American journal of human genetics* **77**, 171-192, doi:10.1086/432519 (2005).

78 Lu, S. *et al.* Comparison of nonhuman primates identified the suitable model for COVID-19. *Signal Transduction and Targeted Therapy* **5**, 157, doi:10.1038/s41392-020-00269-6 (2020).

79 Garcia, M. A. *et al.* Outbreak of Mycobacterium bovis in a Conditioned Colony of Rhesus (*Macaca mulatta*) and Cynomolgus (*Macaca fascicularis*) Macaques. *Comparative Medicine* **54**, 578-584 (2004).

80 El Mubarak, H. S. *et al.* Infection of cynomolgus macaques (Macaca fascicularis) and rhesus macaques (Macaca mulatta) with different wild-type measles viruses. **88**, 2028-2034, doi:https://doi.org/10.1099/vir.0.82804-0 (2007).

81 Maiello, P. *et al.* Rhesus Macaques Are More Susceptible to Progressive Tuberculosis than Cynomolgus Macaques: a Quantitative Comparison. *Infection and Immunity* **86**, e00505-00517, doi:doi:10.1128/IAI.00505-17 (2018).

82 Roos, C. & Zinner, D. in *The Nonhuman Primate in Nonclinical Drug Development and Safety Assessment* (eds Joerg Bluemel, Sven Korte, Emanuel Schenck, & Gerhard F. Weinbauer) 3-16 (Academic Press, 2015).

83 Peterson, M. S. *et al.* Clinical recovery of *Macaca fascicularis* infected with *Plasmodium knowlesi* (In press). *Malar J*, doi:10.1101/2021.06.28.448877 (2021).

84 Sironi, M., Cagliani, R., Forni, D. & Clerici, M. Evolutionary insights into host–pathogen interactions from mammalian sequence data. *Nature Reviews Genetics* **16**, 224-236, doi:10.1038/nrg3905 (2015).

85 Eicher, T. *et al.* Metabolomics and Multi-Omics Integration: A Survey of Computational Methods and Resources. *Metabolites* **10**, 202, doi:10.3390/metabo10050202 (2020).

86 White, N. J. Plasmodium knowlesi: the fifth human malaria parasite. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **46**, 172-173, doi:10.1086/524889 (2008).

87 Yusof, R. *et al.* High proportion of knowlesi malaria in recent malaria cases in Malaysia. *Malar J* **13**, 168, doi:10.1186/1475-2875-13-168 (2014).

88 Fooden, J. Rhesus and Crab-Eating Macaques: Intergradation in Thailand. *Science* **143**, 363-364, doi:10.1126/science.143.3604.363 (1964).

89 Peterson, M. *Malaria disease severity and resilience: Plasmodium knowlesi infection of Macaca mulatta and Macaca fascicularis*, (2020).

90 Lapp, S. A. *et al.* PacBio assembly of a *Plasmodium knowlesi* genome sequence with Hi-C correction and manual annotation of the SICAvar gene family. *Parasitology* **145**, 71-84, doi:10.1017/S0031182017001329 (2018).

91 Zimin, A. V. *et al.* A new rhesus macaque assembly and annotation for next-generation sequencing analyses. *Biol Direct* **9**, 20, doi:10.1186/1745-6150-9-20 (2014).

92 *University of Nebraska Nonhuman Primate Genome Center: Rhesus macaque (Macaca mulatta)*, <https://www.unmc.edu/rhesusgenechip/index.htm#NewRhesusGenome> (

93 NCBI. *Macaca fascicularis Annotation Release 101*, <https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Macaca_fascicularis/101> (2016, Jan 25).

94 *NCBI Genomes: GCF_000364345.1 Macaca fascicularis 5.0*, <ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/364/345/GCF_000364345.1_Macaca_fascicularis_5.0> (2019, November 01).

95 Schmelling, N. M. *et al.* Minimal tool set for a prokaryotic circadian clock. *BMC Evol Biol* **17**, 169, doi:10.1186/s12862-017-0999-7 (2017).

96 Nei, M. Phylogenetic analysis in molecular evolutionary genetics. *Annu Rev Genet* **30**, 371-403, doi:10.1146/annurev.genet.30.1.371 (1996).

97 Altschul, S. F. & Gish, W. in *Methods in Enzymology* Vol. 266 460-480 (Academic Press, 1996).

98 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).

99 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).

100 Love, M. *DESeq2 v1.12.3*, <https://www.rdocumentation.org/packages/DESeq2/versions/1.12.3> (2019).

101 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).

102 Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417-425, doi:10.1016/j.cels.2015.12.004 (2015).

103 Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29, doi:10.1038/75556 (2000).

104 The Gene Ontology, C. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res* **47**, D330-D338, doi:10.1093/nar/gky1055 (2019).

105 Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 289-300 (1995).

106 Supek, F., Bosnjak, M., Skunca, N. & Smuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6**, e21800, doi:10.1371/journal.pone.0021800 (2011).

107 Pomaznoy, M., Ha, B. & Peters, B. GOnet: a tool for interactive Gene Ontology analysis. *BMC Bioinformatics* **19**, 470, doi:10.1186/s12859-018-2533-3 (2018).

108 Maere, S., Heymans, K. & Kuiper, M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* **21**, 3448-3449, doi:10.1093/bioinformatics/bti551 (2005).

109 Janky, R. *et al.* iRegulon: from a gene list to a gene regulatory network using large motif and track collections. *PLoS Comput Biol* **10**, e1003731, doi:10.1371/journal.pcbi.1003731 (2014).

110 Han, H. *et al.* TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Res* **46**, D380-D386, doi:10.1093/nar/gkx1013 (2018).

111 Altman, M. C. *et al.* A Novel Repertoire of Blood Transcriptome Modules Based on Co-expression Patterns Across Sixteen Disease and Physiological States. *BioRxiv*, 525709, doi:10.1101/525709 (2019).

112 Xie, X. *et al.* Single-cell transcriptomic landscape of human blood cells. *National Science Review*, doi:10.1093/nsr/nwaa180 (2020).

113 Harris, S. L. & Levine, A. J. The p53 pathway: positive and negative feedback loops. *Oncogene* **24**, 2899-2908, doi:10.1038/sj.onc.1208615 (2005).

114 Owen, K. L., Brockwell, N. K. & Parker, B. S. JAK-STAT Signaling: A Double-Edged Sword of Immune Regulation and Cancer Progression. *Cancers (Basel)* **11**, 2002 (2019).

115 Pandey, K. *et al.* Ca(2+) monitoring in *Plasmodium falciparum* using the yellow cameleon-Nano biosensor. *Scientific Reports* **6**, 23454, doi:10.1038/srep23454 (2016).

116 Sun, S.-C. The non-canonical NF-κB pathway in immunity and inflammation. *Nat Rev Immunol* **17**, 545-558, doi:10.1038/nri.2017.52 (2017).

117 Fukao, T. *et al.* Inducible expression of Stat4 in dendritic cells and macrophages and its critical role in innate and adaptive immune responses. *J Immunol* **166**, 4446-4455, doi:10.4049/jimmunol.166.7.4446 (2001).

118 Kapellos, T. S. *et al.* Human Monocyte Subsets and Phenotypes in Major Chronic Inflammatory Diseases. *Frontiers in Immunology* **10** doi:10.3389/fimmu.2019.02035 (2019).

119 Holz, L. E., Fernandez-Ruiz, D. & Heath, W. R. Protective immunity to liver-stage malaria. *Clinical & translational immunology* **5**, e105, doi:10.1038/cti.2016.60 (2016).

120 Mogensen, T. H. Pathogen recognition and inflammatory signaling in innate immune defenses. *Clin Microbiol Rev* **22**, 240-273, Table of Contents, doi:10.1128/CMR.00046-08 (2009).

121 Lee, A. J. & Ashkar, A. A. The Dual Nature of Type I and Type II Interferons. *Front Immunol* **9**, 2061, doi:10.3389/fimmu.2018.02061 (2018).

122 Ockenhouse, C. F. *et al.* Common and Divergent Immune Response Signaling Pathways Discovered in Peripheral Blood Mononuclear Cell Gene Expression Patterns in Presymptomatic and Clinically Apparent Malaria. *Infection and Immunity* **74**, 5561-5573, doi:10.1128/IAI.00408-06 (2006).

123 Samuel, C. E. Antiviral actions of interferons. *Clin Microbiol Rev* **14**, 778-809, table of contents, doi:10.1128/CMR.14.4.778-809.2001 (2001).

124 Marie, I., Durbin, J. E. & Levy, D. E. Differential viral induction of distinct interferon-alpha genes by positive feedback through interferon regulatory factor-7. *EMBO J* **17**, 6660-6669, doi:10.1093/emboj/17.22.6660 (1998).

125 Sato, M. *et al.* Positive feedback regulation of type I IFN genes by the IFN-inducible transcription factor IRF-7. *FEBS Lett* **441**, 106-110, doi:10.1016/s0014-5793(98)01514-2 (1998).

126 Sato, M. *et al.* Distinct and essential roles of transcription factors IRF-3 and IRF-7 in response to viruses for IFN-alpha/beta gene induction. *Immunity* **13**, 539-548, doi:10.1016/s1074-7613(00)00053-4 (2000).

127 Yu, Y., Wang, S. E. & Hayward, G. S. The KSHV immediate-early transcription factor RTA encodes ubiquitin E3 ligase activity that targets IRF7 for proteosome-mediated degradation. *Immunity* **22**, 59-70, doi:10.1016/j.immuni.2004.11.011 (2005).

128 Schotte, R. *et al.* The transcription factor Spi-B is expressed in plasmacytoid DC precursors and inhibits T-, B-, and NK-cell development. *Blood* **101**, 1015-1023, doi:10.1182/blood-2002-02-0438 (2003).

129 Gowda, D. C. & Wu, X. Parasite Recognition and Signaling Mechanisms in Innate Immune Responses to Malaria. *Front Immunol* **9**, 3006, doi:10.3389/fimmu.2018.03006 (2018).

130 Sampath, P., Moideen, K., Ranganathan, U. D. & Bethunaickan, R. Monocyte Subsets: Phenotypes and Function in Tuberculosis Infection. *Frontiers in immunology* **9** doi:10.3389/fimmu.2018.01726 (2018).

131 Gudkov, A. V., Gurova, K. V. & Komarova, E. A. Inflammation and p53: A Tale of Two Stresses. *Genes Cancer* **2**, 503-516, doi:10.1177/1947601911409747 (2011).

132 Tran, T. M. *et al.* A Molecular Signature in Blood Reveals a Role for p53 in Regulating Malaria-Induced Inflammation. *Immunity* **51**, 750-765 e710, doi:10.1016/j.immuni.2019.08.009 (2019).

133 Chaussabel, D. & Baldwin, N. Democratizing systems immunology with modular transcriptional repertoire analyses. *Nat Rev Immunol* **14**, 271-280, doi:10.1038/nri3642 (2014).

134 Lillie, P. J. *et al.* Distinguishing malaria and influenza: early clinical features in controlled human experimental infection studies. *Travel Med Infect Dis* **10**, 192-196, doi:10.1016/j.tmaid.2012.03.008 (2012).

135 Kubo, M. & Motomura, Y. Transcriptional regulation of the anti-inflammatory cytokine IL-10 in acquired immune cells. *Frontiers in Immunology* **3** doi:10.3389/fimmu.2012.00275 (2012).

136 Hovsepian, E., Penas, F., Siffo, S., Mirkin, G. A. & Goren, N. B. IL-10 Inhibits the NF-κB and ERK/MAPK-Mediated Production of Pro-Inflammatory Mediators by Up-Regulation of SOCS-3 in Trypanosoma cruzi-Infected Cardiomyocytes. *PloS One*

8, e79445, doi:10.1371/journal.pone.0079445 (2013).

137 Vasou, A., Sultanoglu, N., Goodbourn, S., Randall, R. E. & Kostrikis, L. G. Targeting Pattern Recognition Receptors (PRR) for Vaccine Adjuvantation: From Synthetic PRR Agonists to the Potential of Defective Interfering Particles of Viruses. *Viruses* **9**, 186, doi:10.3390/v9070186 (2017).

138 Gutierrez, J. B., Galinski, M. R., Cantrell, S. & Voit, E. O. From within host dynamics to the epidemiology of infectious disease: Scientific overview and challenges. *Math Biosci* **270**, 143-155, doi:10.1016/j.mbs.2015.10.002 (2015).

139 Fonseca, L. L. *et al.* Quantifying the removal of red blood cells in *Macaca mulatta* during a *Plasmodium coatneyi* infection. *Malar J* **15**, 410, doi:10.1186/s12936-016-1465-5 (2016).

140 Fonseca, L. L. & Voit, E. O. Comparison of mathematical frameworks for modeling erythropoiesis in the context of malaria infection. *Math. Biosci.* **270**, 224-236 (2015).

141 Fonseca, L. L., Joyner, C., MaHPIC Consortium, Galinski, M. R. & Voit, E. O. *Plasmodium vivax* concealment modeled based on *P. cynomolg*i infections in *Macaca mulatta*. *Malar J*, doi:10.1186/s12936-017-2008-4 (2017).

142 Voit, E. O. Models of data and models of processes in the post-genomic era. *Math Biosci* **180**, 263-274 (2002).

143 Hunter, D. J. & Kraft, P. Drinking from the Fire Hose — Statistical Issues in Genomewide Association Studies. **357**, 436-439, doi:10.1056/NEJMp078120 (2007).

144 Stein, D. F., O'Connor, D., Blohmke, C. J., Sadarangani, M. & Pollard, A. J. Gene expression profiles are different in venous and capillary blood: Implications for vaccine studies. *Vaccine* **34**, 5306-5313, doi:10.1016/j.vaccine.2016.09.007 (2016).

145 Joyner, C. *et al. Plasmodium cynomolgi* infections in rhesus macaques display clinical and parasitological features pertinent to modelling vivax malaria pathology and relapse infections. *Malar J* **15**, 451, doi:10.1186/s12936-016-1480-6 (2016).

146 Tang, Y. *et al.* Multi-omics Integrative Analysis of Acute and Relapsing Malaria in a Non-Human Primate Model of *P. vivax* infection. *bioRxiv*, 564195, doi:10.1101/564195 (2019).

147 Galinski, M. R. Functional genomics of simian malaria parasites and host-parasite interactions. *Brief Funct Genomics* **18**, 270-280, doi:10.1093/bfgp/elz013 (2019).

148 Devonshire, A. S., Elaswarapu, R. & Foy, C. A. Evaluation of external RNA controls for the standardisation of gene expression biomarker measurements. *BMC Genomics* **11**, 662, doi:10.1186/1471-2164-11-662 (2010).

149 Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**, R36, doi:10.1186/gb-2013-14-4-r36 (2013).

150 Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7**, 562-578, doi:10.1038/nprot.2012.016 (2012).

151 Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169, doi:10.1093/bioinformatics/btu638 (2015).

152 Wang, L., Wang, S. & Li, W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **28**, 2184-2185, doi:10.1093/bioinformatics/bts356 (2012).

153 Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol* **11**, R106, doi:10.1186/gb-2010-11-10-r106 (2010).

154 Mecham, B. H., Nelson, P. S. & Storey, J. D. Supervised normalization of microarrays. *Bioinformatics* **26**, 1308-1315, doi:10.1093/bioinformatics/btq118 (2010).

155 Benjamini, Y. & Hochberg, Y. Controlling false discovery rate - a practical and powerful approach to multiple testing. *J. Roy. Soc. Series B-Methodological* **57**, 289-300, doi:10.1007/978-3-319-19992-4_10 (1995).

156 Rojas-Pena, M. L., Vallejo, A., Herrera, S., Gibson, G. & Arevalo-Herrera, M. Transcription Profiling of Malaria-Naive and Semi-immune Colombian Volunteers in a *Plasmodium vivax* Sporozoite Challenge. *PLoS Negl Trop Dis* **9**, e0003978, doi:10.1371/journal.pntd.0003978 (2015).

157 Hung, J. H., Yang, T. H., Hu, Z., Weng, Z. & DeLisi, C. Gene set enrichment analysis: performance evaluation and usage guidelines. *Brief Bioinform* **13**, 281-291, doi:10.1093/bib/bbr049 (2012).

158 Curto, R., Voit, E. O. & Cascante, M. Analysis of abnormalities in purine metabolism leading to gout and to neurological dysfunctions in man. *Biochem J* **329 (Pt 3)**, 477-487 (1998).

159 Curto, R., Voit, E. O., Sorribas, A. & Cascante, M. Validation and steady-state analysis of a power-law model of purine metabolism in man. *Biochem J* **324 (Pt 3)**, 761-775 (1997).

160 Curto, R., Voit, E. O., Sorribas, A. & Cascante, M. Mathematical models of purine metabolism in man. *Math Biosci* **151**, 1-49 (1998).

161 Voit, E. O. *Computational Analysis of Biochemical Systems: A Practical Guide for Biochemists and Molecular Biologists*. (Cambridge University Press, 2000).

162 Savageau, M. A. *Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology*. (Addison-Wesley Pub. Co. Advanced Book Program (reprinted 2009), 1976).

163 Torres, N. V. & Voit, E. O. *Pathway Analysis and Optimization in Metabolic Engineering*. (Cambridge University Press, 2002).

164 Voit, E. O. Biochemical Systems Theory: A review. *Int. Scholarly Res. Network (ISRN – Biomathematics)* **Article 897658**, 1-53 (2013).

165 Knorre W., S. M. Biochemical Systems Analysis. A Study of Function and Design in Molecular Biology. 396 S., 115 Abb., 14 Tab. Reading, Mass. 1976. Addison-Wesley Pbl. Co./Advanced Book Program. £ 26,50. **19**, 149-150, doi:https://doi.org/10.1002/jobm.19790190219 (1979).

166 Voit, E. O. *Canonical Nonlinear Modeling*. (Van Nostrand Reinhold, 1991).

167 Voit, E. O., & Ferreira, A. *Computational analysis of biochemical systems: A practical guide for biochemists and molecular biologists*. (Cambridge University Press, 2000).

168 Ghaemi-Oskouie, F. & Shi, Y. The role of uric acid as an endogenous danger signal in immunity and inflammation. *Curr Rheumatol Rep* **13**, 160-166, doi:10.1007/s11926-011-0162-1 (2011).

169 Leyva, F. *et al.* Uric acid in chronic heart failure: a marker of chronic inflammation. *Eur Heart J* **19**, 1814-1822 (1998).

170 Voit, E. O. & Brigham, K. L. The role of systems biology in predictive health and personalized medicine. *The Open Path. J.* **2**, 68-70 (2008).

171 Voit, E. O. A systems-theoretical framework for health and disease: inflammation and preconditioning from an abstract modeling point of view. *Math Biosci* **217**, 11-18 (2009).

172 Haskó, G. *et al.* Inosine inhibits inflammatory cytokine production by a posttranscriptional mechanism and protects against endotoxin-induced shock. *J Immunol* **164**, 1013-1019 (2000).

173 Gudbjornsson, B., Zak, A., Niklasson, F. & Hallgren, R. Hypoxanthine, xanthine, and urate in synovial fluid from patients with inflammatory arthritides. *Ann Rheum Dis* **50**, 669-672 (1991).

174 da Rocha Lapa, F., da Silva, M. D., de Almeida Cabrini, D. & Santos, A. R. Anti-inflammatory effects of purine nucleosides, adenosine and inosine, in a mouse model of pleurisy: evidence for the role of adenosine A2 receptors. *Purinergic Signal* **8**, 693-704, doi:10.1007/s11302-012-9299-2 (2012).

175 Marton, A. *et al.* Anti-inflammatory effects of inosine in human monocytes, neutrophils and epithelial cells *in vitro*. *Int J Mol Med* **8**, 617-621 (2001).

176 Takahashi, T., Otsuguro, K., Ohta, T. & Ito, S. Adenosine and inosine release during hypoxia in the isolated spinal cord of neonatal rats. *Br J Pharmacol* **161**, 1806-1816, doi:10.1111/j.1476-5381.2010.01002.x (2010).

177 Ferrari, D., McNamee, E. N., Idzko, M., Gambari, R. & Eltzschig, H. K. Purinergic Signaling During Immune Cell Trafficking. *Trends Immunol* **37**, 399-411, doi:10.1016/j.it.2016.04.004 (2016).

178 Gallego-Delgado, J., Ty, M., Orengo, J. M., van de Hoef, D. & Rodriguez, A. A surprising role for uric acid: the inflammatory malaria response. *Curr Rheumatol Rep* **16**, 401, doi:10.1007/s11926-013-0401-8 (2014).

179 Lutgen, P. *Uric acid: emerging as key factor in malaria*, <https://malariaworld.org/blog/uric-acid-emerging-key-factor-malaria> (Accessed 2017).

180 Lopera-Mesa, T. M. *et al.* Plasma uric acid levels correlate with inflammation and disease severity in Malian children with *Plasmodium falciparum* malaria. *PLoS One* **7**, e46424, doi:10.1371/journal.pone.0046424 (2012).

181 Downie, M. J., Kirk, K. & Mamoun, C. B. Purine salvage pathways in the intraerythrocytic malaria parasite *Plasmodium falciparum*. *Eukaryot Cell* **7**, 1231-1237, doi:10.1128/EC.00159-08 (2008).

182 Cassera, M. B., Zhang, Y., Hazleton, K. Z. & Schramm, V. L. Purine and pyrimidine pathways as targets in *Plasmodium falciparum*. *Curr Top Med Chem* **11**, 2103-2115 (2011).

183 Howard, R. J., Barnwell, J. W. & Kao, V. Antigenic variation of *Plasmodium knowlesi* malaria: identification of the variant antigen on infected erythrocytes. *Proceedings of the National Academy of Sciences of the United States of America* **80**, 4129-4133, doi:10.1073/pnas.80.13.4129 (1983).

184 Wahlgren, M., Fernandez, V., Chen, Q., Svard, S. & Hagblom, P. Waves of malarial var-iations. *Cell* **96**, 603-606, doi:10.1016/s0092-8674(00)80569-3 (1999).

185 al-Khedery, B., Barnwell, J. W. & Galinski, M. R. Antigenic variation in malaria: a 3' genomic alteration associated with the expression of a *P. knowlesi* variant antigen. *Mol Cell* **3**, 131-141, doi:10.1016/s1097-2765(00)80304-4 (1999).

186 Pain, A. *et al.* The genome of the simian and human malaria parasite *Plasmodium knowlesi*. *Nature* **455**, 799-803, doi:10.1038/nature07306 (2008).

187 Lapp, S. A. *et al.* PacBio assembly of a *Plasmodium knowlesi* genome sequence with Hi-C correction and manual annotation of the SICAvar gene family. *Parasitology* **145**, 71-84, doi:http://dx.doi.org/10.1017/S0031182017001329 (2018).

188 Ouattara, A. *et al.* Designing malaria vaccines to circumvent antigen variability. *Vaccine* **33**, 7506-7512, doi:10.1016/j.vaccine.2015.09.110 (2015).

189 Ferreira, M. U., da Silva Nunes, M. & Wunderlich, G. Antigenic diversity and immune evasion by malaria parasites. *Clin Diagn Lab Immunol* **11**, 987-995, doi:10.1128/CDLI.11.6.987-995.2004 (2004).

190 Rénia, L. & Goh, Y. S. Malaria Parasites: The Great Escape. *Frontiers in Immunology* **7**, doi:10.3389/fimmu.2016.00463 (2016).

191 França, C. T. *et al.* Identification of highly-protective combinations of *Plasmodium vivax* recombinant proteins for vaccine development. *eLife* **6**, e28673, doi:10.7554/eLife.28673 (2017).

192 Alberts, B. *Molecular biology of the cell*. 4th edn, (Garland Science, 2002).

193 Schroder, K. & Tschopp, J. The inflammasomes. *Cell* **140**, 821-832, doi:10.1016/j.cell.2010.01.040 (2010).

194 Lee, H. J. *et al.* Integrated pathogen load and dual transcriptome analysis of systemic host-pathogen interactions in severe malaria. *Sci Transl Med* **10**, doi:10.1126/scitranslmed.aar3619 (2018).

195 Cicchese, J. M. *et al.* Dynamic balance of pro- and anti-inflammatory signals controls disease and limits pathology. *Immunol Rev* **285**, 147-167, doi:10.1111/imr.12671 (2018).

196 Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417-425, doi:10.1016/j.cels.2015.12.004 (2015).

197 Jassal, B. *et al.* The reactome pathway knowledgebase. *Nucleic Acids Res* **48**, D498-d503, doi:10.1093/nar/gkz1031 (2020).

198 Godec, J. *et al.* Compendium of Immune Signatures Identifies Conserved and Species-Specific Biology in Response to Inflammation. *Immunity* **44**, 194-206, doi:10.1016/j.immuni.2015.12.006 (2016).

199 Gene Ontology, C. The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res* **49**, D325-D334, doi:10.1093/nar/gkaa1113 (2021).

200 Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29, doi:10.1038/75556 (2000).

201 Oki, S. *et al.* ChIP-Atlas: a data-mining suite powered by full integration of public ChIP-seq data. *EMBO Rep* **19**, doi:10.15252/embr.201846255 (2018).

202 Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* **4**, Article17, doi:10.2202/1544-6115.1128 (2005).

203 Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559, doi:10.1186/1471-2105-9-559 (2008).

204 Newman, A. M. *et al.* Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol* **37**, 773-782, doi:10.1038/s41587-019-0114-2 (2019).

205 Newman, A. M. *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* **12**, 453-457, doi:10.1038/nmeth.3337 (2015).

206 Stavrum, A. K., Heiland, I., Schuster, S., Puntervoll, P. & Ziegler, M. Model of tryptophan metabolism, readily scalable using tissue-specific gene expression data. *J Biol Chem* **288**, 34555-34566, doi:10.1074/jbc.M113.474908 (2013).

207 Tang, Y. *et al.* Metabolic modeling helps interpret transcriptomic changes during malaria. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **1864**, 2329-2340, doi:https://doi.org/10.1016/j.bbadis.2017.10.023 (2018).

208 Consortium, G. T. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-585, doi:10.1038/ng.2653 (2013).

209 Tang, Y. *et al.* Metabolic modeling helps interpret transcriptomic changes during malaria. *Biochimica et biophysica acta. Molecular basis of disease* **1864**, 2329-2340, doi:10.1016/j.bbadis.2017.10.023 (2018).

210 Wildermuth, M. C. Metabolic control analysis: biological applications and insights. *Genome Biol* **1**, REVIEWS1031, doi:10.1186/gb-2000-1-6-reviews1031 (2000).

211 Peterson, M. S. *et al.* Clinical recovery of *Macaca fascicularis* infected with *Plasmodium knowlesi*. *Malar J* **(In press)** (2021).

212 Fuller, T. F. *et al.* Weighted gene coexpression network analysis strategies applied to mouse weight. *Mammalian Genome* **18**, 463-472, doi:10.1007/s00335-007-9043-3 (2007).

213 Galinski, M. R. *et al. Plasmodium knowlesi*: a superb in vivo nonhuman primate model of antigenic variation in malaria. *Parasitology* **145**, 85-100, doi:10.1017/S0031182017001135 (2018).

214 Howard, R. J., Barnwell, J. W. & Kao, V. Antigenic variation of *Plasmodium knowlesi* malaria: identification of the variant antigen on infected erythrocytes. *Proc Natl Acad Sci U S A* **80**, 4129-4133, doi:10.1073/pnas.80.13.4129 (1983).

215 Lapp, S. A., Korir, C. C. & Galinski, M. R. Redefining the expressed prototype SICAvar gene involved in *Plasmodium knowlesi* antigenic variation. *Malar J* **8**, 181-181, doi:10.1186/1475-2875-8-181 (2009).

216 Wang, B. *et al.* Immunoprofiling of the tryptophan-rich antigen family in *Plasmodium vivax*. *Infect Immun* **83**, 3083-3095, doi:10.1128/IAI.03067-14 (2015).

217 Counihan, N. A., Modak, J. K. & de Koning-Ward, T. F. How Malaria Parasites Acquire Nutrients From Their Host. *Frontiers in Cell and Developmental Biology* **9**, doi:10.3389/fcell.2021.649184 (2021).

218 Painter, H. J., Campbell, T. L. & Llinás, M. The Apicomplexan AP2 family: integral factors regulating *Plasmodium* development. *Mol Biochem Parasitol* **176**, 1-7, doi:10.1016/j.molbiopara.2010.11.014 (2011).

219 Zeeshan, M. *et al. Plasmodium berghei* Kinesin-5 Associates With the Spindle Apparatus During Cell Division and Is Important for Efficient Production of Infectious Sporozoites. *Frontiers in Cellular and Infection Microbiology* **10**, doi:10.3389/fcimb.2020.583812 (2020).

220 Wang, W. *et al.* A Thioredoxin Homologous Protein of *Plasmodium falciparum* Participates in Erythrocyte Invasion. *Infect Immun* **86**, doi:10.1128/IAI.00289-18 (2018).

221 Yindom, L. M. *et al.* Killer-cell immunoglobulin-like receptors and malaria caused by *Plasmodium falciparum* in The Gambia. *Tissue Antigens* **79**, 104-113, doi:10.1111/j.1399-0039.2011.01818.x (2012).

222 Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* **4**, 44-57, doi:10.1038/nprot.2008.211 (2009).

223 Mogensen, T. H. Pathogen recognition and inflammatory signaling in innate immune defenses. *Clin Microbiol Rev* **22**, 240-273, doi:10.1128/CMR.00046-08 (2009).

224 Roh, J. S. & Sohn, D. H. Damage-Associated Molecular Patterns in Inflammatory Diseases. *Immune Netw* **18**, e27-e27, doi:10.4110/in.2018.18.e27 (2018).

225 Amarante-Mendes, G. P. *et al.* Pattern Recognition Receptors and the Host Cell Death Molecular Machinery. *Frontiers in Immunology* **9**, doi:10.3389/fimmu.2018.02379 (2018).

226 Festjens, N., Vanden Berghe, T., Cornelis, S. & Vandenabeele, P. RIP1, a kinase on the crossroads of a cell's decision to live or die. *Cell Death & Differentiation* **14**, 400-410, doi:10.1038/sj.cdd.4402085 (2007).

227 De la Mota-Peynado, A., Chernoff, J. & Beeser, A. Identification of the atypical MAPK Erk3 as a novel substrate for p21-activated kinase (Pak) activity. *J Biol Chem* **286**, 13603-13611, doi:10.1074/jbc.M110.181743 (2011).

228 Déléris, P. *et al.* Activation loop phosphorylation of ERK3/ERK4 by group I p21-activated kinases (PAKs) defines a novel PAK-ERK3/4-MAPK-activated protein kinase 5 signaling pathway. *J Biol Chem* **286**, 6470-6478, doi:10.1074/jbc.M110.181529 (2011).

229 Kostenko, S., Dumitriu, G. & Moens, U. Tumour promoting and suppressing roles of the atypical MAP kinase signalling pathway ERK3/4-MK5. *Journal of molecular signaling* **7**, 9, doi:10.1186/1750-2187-7-9 (2012).

230 Saha, K., Adhikary, G., Kanade, S. R., Rorke, E. A. & Eckert, R. L. p38δ regulates p53 to control p21Cip1 expression in human epidermal keratinocytes. *J Biol Chem* **289**, 11443-11453, doi:10.1074/jbc.M113.543165 (2014).

231 Barberà-Cremades, M., Baroja-Mazo, A. & Pelegrín, P. Purinergic signaling during macrophage differentiation results in M2 alternative activated macrophages. *Journal of leukocyte biology* **99**, 289-299, doi:10.1189/jlb.1A0514-267RR (2016).

232 Haupt, Y., Robles, A. I., Prives, C. & Rotter, V. Deconstruction of p53 functions and regulation. *Oncogene* **21**, 8223-8231, doi:10.1038/sj.onc.1206137 (2002).

233 Sablina, A. A. *et al.* The antioxidant function of the p53 tumor suppressor. *Nature Medicine* **11**, 1306-1313, doi:10.1038/nm1320 (2005).

234 Lee, J. H. *et al.* The p53-inducible gene 3 (PIG3) contributes to early cellular response to DNA damage. *Oncogene* **29**, 1431-1450, doi:10.1038/onc.2009.438 (2010).

235 Joerger, A. C. & Fersht, A. R. The p53 Pathway: Origins, Inactivation in Cancer, and Emerging Therapeutic Approaches. *Annual Review of Biochemistry* **85**, 375-404, doi:10.1146/annurev-biochem-060815-014710 (2016).

236 Golomb, L., Volarevic, S. & Oren, M. p53 and ribosome biogenesis stress: The essentials. *FEBS Letters* **588**, 2571-2579, doi:https://doi.org/10.1016/j.febslet.2014.04.014 (2014).

237 Haupt, S., Mejía-Hernández, J. O., Vijayakumaran, R., Keam, S. P. & Haupt, Y. The long and the short of it: the MDM4 tail so far. *Journal of Molecular Cell Biology* **11**, 231-244, doi:10.1093/jmcb/mjz007 %J Journal of Molecular Cell Biology (2019).

238 Shaulian, E. & Karin, M. AP-1 in cell proliferation and survival. *Oncogene* **20**, 2390-2400, doi:10.1038/sj.onc.1204383 (2001).

239 Albert, B. *et al.* A ribosome assembly stress response regulates transcription to maintain proteome homeostasis. *eLife* **8**, e45002, doi:10.7554/eLife.45002 (2019).

240 Toma-Jonik, A., Vydra, N., Janus, P. & Widłak, W. Interplay between HSF1 and p53 signaling pathways in cancer initiation and progression: non-oncogene and oncogene addiction. *Cellular Oncology* **42**, 579-589, doi:10.1007/s13402-019-00452-0 (2019).

241 MacLeod, M. K., Clambey, E. T., Kappler, J. W. & Marrack, P. CD4 memory T cells: what are they and what can they do? *Seminars in immunology* **21**, 53-61, doi:10.1016/j.smim.2009.02.006 (2009).

242 Crotty, S. T follicular helper cell differentiation, function, and roles in disease. *Immunity* **41**, 529-542, doi:10.1016/j.immuni.2014.10.004 (2014).

243 Awad, F. *et al.* Impact of human monocyte and macrophage polarization on NLR expression and NLRP3 inflammasome activation. *PLoS One* **12**, e0175336, doi:10.1371/journal.pone.0175336 (2017).

244 Xia, C., Braunstein, Z., Toomey, A. C., Zhong, J. & Rao, X. S100 Proteins As an Important Regulator of Macrophage Inflammation. *Front Immunol* **8**, 1908, doi:10.3389/fimmu.2017.01908 (2017).

245 Wang, S. *et al.* S100A8/A9 in Inflammation. *Front Immunol* **9**, 1298, doi:10.3389/fimmu.2018.01298 (2018).

246 Pan, S. C. *et al.* The p53-S100A2 Positive Feedback Loop Negatively Regulates Epithelialization in Cutaneous Wound Healing. *Scientific Reports* **8**, 5458, doi:10.1038/s41598-018-23697-5 (2018).

247 Boye, K. & Maelandsmo, G. M. S100A4 and metastasis: a small actor playing many roles. *Am J Pathol* **176**, 528-535, doi:10.2353/ajpath.2010.090526 (2010).

248 Wang, S. *et al.* S100A8/A9 in Inflammation. **9**, doi:10.3389/fimmu.2018.01298 (2018).

249 Xia, C., Braunstein, Z., Toomey, A. C., Zhong, J. & Rao, X. S100 Proteins As an Important Regulator of Macrophage Inflammation. *Frontiers in Immunology* **8**, doi:10.3389/fimmu.2017.01908 (2018).

250 Fox, D. & Man, S. M. DDX3X: stressing the NLRP3 inflammasome. *Cell Research* **29**, 969-970, doi:10.1038/s41422-019-0250-8 (2019).

251 Cipolletta, D. *et al.* PPAR-γ is a major driver of the accumulation and phenotype of adipose tissue Treg cells. *Nature* **486**, 549-553, doi:10.1038/nature11132 (2012).

252 Toker, A. *et al.* Active demethylation of the Foxp3 locus leads to the generation of stable regulatory T cells within the thymus. *J Immunol* **190**, 3180-3188, doi:10.4049/jimmunol.1203473 (2013).

253 Covens, K. *et al.* Characterization of proposed human B-1 cells reveals pre-plasmablast phenotype. *Blood* **121**, 5176-5183, doi:10.1182/blood-2012-12-471953 (2013).

254 Koo, B.-S. *et al.* Reference values of hematological and biochemical parameters in young-adult cynomolgus monkey (*Macaca fascicularis*) and rhesus monkey (*Macaca mulatta*) anesthetized with ketamine hydrochloride. *Laboratory Animal Research* **35**, 7, doi:10.1186/s42826-019-0006-0 (2019).

255 Cordy, R. J. *et al.* Distinct amino acid and lipid perturbations characterize acute versus chronic malaria. *JCI Insight* **4**, e125156, doi:10.1172/jci.insight.125156 (2019).

256 Sarkar, S. A. *et al.* Induction of indoleamine 2,3-dioxygenase by interferon-gamma in human islets. *Diabetes* **56**, 72-79, doi:10.2337/db06-0617 (2007).

257 Taylor, M. W. & Feng, G. S. Relationship between interferon-gamma, indoleamine 2,3-dioxygenase, and tryptophan catabolism. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **5**, 2516-2522 (1991).

258 Banzola, I. *et al.* Expression of Indoleamine 2,3-Dioxygenase Induced by IFN-γ and TNF-α as Potential Biomarker of Prostate Cancer Progression. *Frontiers in Immunology* **9**, doi:10.3389/fimmu.2018.01051 (2018).

259 Stavrum, A.-K., Heiland, I., Schuster, S., Puntervoll, P. & Ziegler, M. Model of tryptophan metabolism, readily scalable using tissue-specific gene expression data. *J Biol Chem* **288**, 34555-34566, doi:10.1074/jbc.M113.474908 (2013).

260 Fell, D. *Understanding the control of metabolism*. [Repr.] edn, XII, 301 Seiten Illustrationen, Diagramme (Portland Press, 2007).

261 Fell, D. A. Metabolic control analysis: a survey of its theoretical and experimental development. *The Biochemical journal* **286 (Pt 2)**, 313-330, doi:10.1042/bj2860313 (1992).

262 Hubbard, T. D., Murray, I. A. & Perdew, G. H. Indole and Tryptophan Metabolism: Endogenous and Dietary Routes to Ah Receptor Activation. *Drug Metab Dispos* **43**, 1522-1535, doi:10.1124/dmd.115.064246 (2015).

263 Roager, H. M. & Licht, T. R. Microbial tryptophan catabolites in health and disease. *Nature Communications* **9**, 3294, doi:10.1038/s41467-018-05470-4 (2018).

264 Yu, Y. C., Han, J. M. & Kim, S. Aminoacyl-tRNA synthetases and amino acid signaling. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1868**, 118889, doi:https://doi.org/10.1016/j.bbamcr.2020.118889 (2021).

265 Gutiérrez-Vázquez, C. & Quintana, F. J. Regulation of the Immune Response by the Aryl Hydrocarbon Receptor. *Immunity* **48**, 19-33, doi:10.1016/j.immuni.2017.12.012 (2018).

266 Murray, I. A., Patterson, A. D. & Perdew, G. H. Aryl hydrocarbon receptor ligands in cancer: friend and foe. *Nature Reviews Cancer* **14**, 801-814, doi:10.1038/nrc3846 (2014).

267 Zhang, X. *et al.* Endogenous Indole Pyruvate Pathway for Tryptophan Metabolism Mediated by IL4I1. *Journal of Agricultural and Food Chemistry* **68**, 10678-10684, doi:10.1021/acs.jafc.0c03735 (2020).

268 Phelan, D., Winter, G. M., Rogers, W. J., Lam, J. C. & Denison, M. S. Activation of the Ah receptor signal transduction pathway by bilirubin and biliverdin. *Archives of biochemistry and biophysics* **357**, 155-163, doi:10.1006/abbi.1998.0814 (1998).

269 Kapitulnik, J. & Gonzalez, F. J. Marked endogenous activation of the CYP1A1 and CYP1A2 genes in the congenitally jaundiced Gunn rat. *Mol Pharmacol* **43**, 722-725 (1993).

270 Raja, T. N. *et al.* Naturally Acquired Human *Plasmodium cynomolgi* and *P. knowlesi* Infections, Malaysian Borneo. *Emerg Infect Dis* **26**, 1801-1809, doi:10.3201/eid2608.200343 (2020).

271 Zaw, M. T. & Lin, Z. Human *Plasmodium knowlesi* infections in South-East Asian countries. *Journal of Microbiology, Immunology and Infection* **52**, 679-684, doi:https://doi.org/10.1016/j.jmii.2019.05.012 (2019).

272 Cox-Singh, J. Zoonotic malaria: *Plasmodium knowlesi*: , an emerging pathogen. *Current Opinion in Infectious Diseases* **25**, 530-536, doi:10.1097/QCO.0b013e3283558780 (2012).

273 Barber, B. E., Rajahram, G. S., Grigg, M. J., William, T. & Anstey, N. M. World Malaria Report: time to acknowledge *Plasmodium knowlesi* malaria. *Malar J* **16**, 135, doi:10.1186/s12936-017-1787-y (2017).

274 Daneshvar, C. *et al.* Clinical and Laboratory Features of Human *Plasmodium knowlesi* Infection. *Clinical Infectious Diseases* **49**, 852-860, doi:10.1086/605439 %J Clinical Infectious Diseases (2009).

275 Knowles, R. & Gupta, B. M. D. A Study of Monkey-Malaria, and Its Experimental Transmission to Man. *The Indian medical gazette* **67**, 301-320 (1932).

276 Napier, L. E. & Campbell, H. G. M. Observations on a Plasmodium Infection Which Causes Hæmoglobinuria in Certain Species of Monkey. *The Indian medical gazette* **67**, 246-249 (1932).

277 Biggs, B. A. *et al.* Antigenic variation in *Plasmodium falciparum*. *Proc Natl Acad Sci U S A* **88**, 9171-9174, doi:10.1073/pnas.88.20.9171 (1991).

278 Brown, K. N. & Brown, I. N. Immunity to malaria: antigenic variation in chronic infections of *Plasmodium knowlesi*. *Nature* **208**, 1286-1288, doi:10.1038/2081286a0 (1965).

279 New, D. C. & Wong, Y. H. Molecular mechanisms mediating the G protein-coupled receptor regulation of cell cycle progression. *Journal of molecular signaling* **2**, 2, doi:10.1186/1750-2187-2-2 (2007).

280 Goldsmith, Z. G. & Dhanasekaran, D. N. G protein regulation of MAPK networks. *Oncogene* **26**, 3122-3142, doi:10.1038/sj.onc.1210407 (2007).

281 Zhang, W. & Liu, H. T. MAPK signal pathways in the regulation of cell proliferation in mammalian cells. *Cell Research* **12**, 9-18, doi:10.1038/sj.cr.7290105 (2002).

282 Tetsutani, K., To, H., Torii, M., Hisaeda, H. & Himeno, K. Malaria parasite induces tryptophan-related immune suppression in mice. *Parasitology* **134**, 923-930, doi:10.1017/S0031182007002326 (2007).

283 Hansen, A. M., Driussi, C., Turner, V., Takikawa, O. & Hunt, N. H. Tissue distribution of indoleamine 2,3-dioxygenase in normal and malaria-infected tissue. *Redox Report* **5**, 112-115, doi:10.1179/135100000101535384 (2000).

284 Sanni, L. A. *et al.* Dramatic changes in oxidative tryptophan metabolism along the kynurenine pathway in experimental cerebral and noncerebral malaria. *Am J Pathol* **152**, 611-619 (1998).

285 Dos Santos, R. O. *et al.* A First *Plasmodium vivax* Natural Infection Induces Increased Activity of the Interferon Gamma-Driven Tryptophan Catabolism Pathway. *Front Microbiol* **11**, 400-400, doi:10.3389/fmicb.2020.00400 (2020).

286 Glushakova, S. *et al.* Cytoplasmic free Ca2+ is essential for multiple steps in malaria parasite egress from infected erythrocytes. *Malaria Journal* **12**, 41, doi:10.1186/1475-2875-12-41 (2013).

287 Crespi, B. & Alcock, J. Conflicts over calcium and the treatment of COVID-19. *Evol Med Public Health* **9**, 149-156, doi:10.1093/emph/eoaa046 (2021).

288 Chen, X., Cao, R. & Zhong, W. Host Calcium Channels and Pumps in Viral Infections. *Cells* **9**, doi:10.3390/cells9010094 (2019).

289 Tran Van Nhieu, G., Dupont, G. & Combettes, L. Ca2+ signals triggered by bacterial pathogens and microdomains. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1865**, 1838-1845, doi:https://doi.org/10.1016/j.bbamcr.2018.08.007 (2018).

290 Hamada, Y., San, A. M. & Malaivijitnond, S. Assessment of the hybridization between rhesus (*Macaca mulatta*) and long-tailed macaques (*M. fascicularis*) based on morphological characters. **159**, 189-198, doi:https://doi.org/10.1002/ajpa.22862 (2016).

291 Davis, J. D., Kumbale, C. M., Zhang, Q. & Voit, E. O. Dynamical systems approaches to personalized medicine. *Curr Opin Biotechnol* **58**, 168-174, doi:10.1016/j.copbio.2019.03.005 (2019).

292 Schwanhäusser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337-342, doi:10.1038/nature10098 (2011).