Computational analysis of transcriptome and metabolic pathways of malaria parasite *Plasmodium falciparum*

A Thesis

Submitted in partial fulfilment of the requirement

for the award of degree of

DOCTOR OF PHILOSOPHY
IN
BIOTECHNOLOGY

Submitted By

SANJAY KUMAR SINGH
(Reg. No. 901300016)

Department of Biotechnology
Thapar Institute of Engineering and Technology
Patiala - 147004, Punjab, India

(August, 2021)

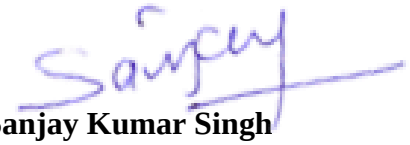
CERTIFICATE

Certified that the thesis "**Computational analysis of transcriptome and metabolic pathways of malaria parasite *Plasmodium falciparum***", which is submitted by **Mr. Sanjay Kumar Singh,** in fulfillment of the requirement for the award of the degree of **Doctor of Philosophy** in the Department of Biotechnology, Thapar Institute of Engineering & Technology, Patiala, is a record of the candidate's own independent and original research work carried out by him under my supervision and guidance. The matter embodied in this thesis has not been submitted in part or full to any other University or Institute for the award of any degree.

(Dr. M. Sudhakara Reddy)
Supervisor
Professor and Head
Department of Biotechnology
Thapar Institute of Engineering & Technology
Patiala, Punjab, India

DECLARATION

I hereby declare that the work which is being presented in this thesis titled "**Computational analysis of transcriptome and metabolic pathways of malaria parasite *Plasmodium falciparum***" submitted by me for the award of the degree of Doctor of Philosophy in the Department of Biotechnology, Thapar Institute of Engineering and Technology, Patiala, is true and original record of my own independent and original research work carried out under Dr. M. Sudhakara Reddy, Professor, Department of Biotechnology, Thapar Institute of Engineering and Technology, Patiala, Punjab, India. The matter embodied in thesis has not been submitted in part or full to any other university or institute for the award of any degree in India or abroad.



Sanjay Kumar Singh

Department of Biotechnology
Thapar Institute of Engineering & Technology
Patiala, Punjab, India

ACKNOWLEDGEMENT

"The best view comes after the hardest climb"

In pursuit of this academic endeavor, I feel that I have been exceptionally fortunate because inspiration, guidance, direction, cooperation, love and care - all came my way in abundance. I feel nostalgic when I look back into my journey and find it difficult to put into words the never-ending support and encouragement of everyone throughout these years.

First and foremost, I acknowledge the grace of ***Almighty GOD*** *who paved my way through all the problems and provided me with the strength required to make this possible.*

With a profound sense of gratitude, I acknowledge my PhD Supervisor, ***Prof. M. Sudhakara Reddy****, Department of Biotechnology, TIET, Patiala for giving me this wonderful opportunity of being a part of his vibrant research team. His pertinent guidance, patronage, prudent advises motivation, enthusiasm, immense knowledge and dedication for research rendered to me during my work, without which the present endeavour would not have achieved the same status. I thank his progressive, energizing and inspirational guidance, unconditional support, admirable dedication throughout my Ph.D. programme.*

I take this opportunity to thank ***Prof. Prakash Gopalan****, Director, TIET, Patiala, for providing all the amenities and guidance for the completion of my research work. I extend my heartfelt word of thanks to* ***Prof. Rafat Siddique****, Dean (Research and Sponsored Projects), TIET, Patiala, for providing motivating research environment during my stay in the Institute.*

I am deeply grateful to Dr. Manoj Baranwal, Dr. Vikas Handa Department of Biotechnology, TIET and Dr. Maninder Singh, Department of Computer Science and Engineering, TIET members of the Doctoral Committee, to monitor my research work from time to time and to make valuable suggestions.

I am very Thankful to Dr. N Das, Dr. Dinesh Goyal, Dr. Moushumi Ghosh and for providing suggestions, for thoughts and taking interest in the progress for this work since inception. I am profoundly thankful to ***Dr. M Vasundhara****, Department of Biotechnology for her valuable suggestions for the progress of my work and* ***Prof. Anil Kumar****, Coordinator, TIFAC-CORE in Agro and Industrial Biotechnology, TIET, for providing facilities, guidance and providing a brilliant environment which is very conducive for research work. I would take this opportunity to acknowledge to all the faculty members of DBT, TIET, Patiala for their constant encouragement for my research work. I am thankful to the office and laboratory staff of Department of Biotechnology for their cooperation and help.*

I express my deep regards to my colleagues ***Mrs. Subhalaxmi Nayak, Mr. Arkadeep Mukherjee, Mrs. Bharti Thakur****,* ***Mr. Sumit Joshi****,* ***Mrs. Shikha Khullar*** *and* ***Mrs. Tanveer Kaur*** *for their great support and valuable suggestions during need of hour. Most of the work would have been incomplete without the sincere support of Lab assistants. So, I owe a word of thanks to* ***Mr. Soni Singh****,* ***Mr. Ram Newal (Babbanji)****,* ***Lallanji*** *and* ***Surinder****.*

I find it difficult to express in words, my deepest sense of indebtedness towards my parents and all other family members. My family withstood all the pressure of this longish period of work with great determination and concern by my side.

Furthermore, I would like to express my sincere thanks to Dr. Vivekanand Jha, Professor, Department of Nephrology, PGIMER, Chandigarh, for providing invaluable assistance,

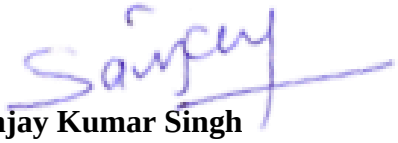
dedicated time, critical feedback and discussions, as well as encouragement on my thesis work. Many thanks!

I sincerely thank the ICMR Biomedical Informatics Centre, PGIMER Chandigarh on this occasion for providing necessary resources to pursue this study.

Special thank goes to my parents, my wife (Mrs. Jaya Singh), my kids (Anshika and Kritansh), friends and colleagues who have provided inspiration, support and encouragement to continue with my work.

Finally, I would like to thank my parents for their understanding, endless patience and constant support.

Last but not the least, I wish to acknowledge all those, whose names have not figured here, but who helped me in any form during the period of my research work.



Sanjay Kumar Singh

Department of Biotechnology
Thapar Institute of Engineering & Technology
Patiala, Punjab, India

List of Publications

Publications in Peer Reviewed Journals

1. Sanjay Kumar Singh, M. Sudhakara Reddy (2019) Investigation of hub genes and their nonsynonymous single nucleotide polymorphism analysis in *Plasmodium falciparum* for designing therapeutic methodologies using next-generation sequencing approach. Indian J Pharmacol 51: 389-99. (**IF: 1.04**)

2. Sanjay Kumar Singh, M. Sudhakara Reddy (2020) Homology modeling, docking, ADMET studies and prediction of deleterious non-synonymous single nucleotide polymorphisms (nsSNPs) of thiamin phosphate synthase: a potential drug target in *Plasmodium falciparum*. Indian J Pharm Sci 82:665-76 (**IF: 0.72**)

3. Sanjay Kumar Singh, M. Sudhakara Reddy (2021) Computational prediction of the effects of non-synonymous single nucleotide polymorphisms in the GPI-anchor transamidase of *Plasmodium falciparum*. Comput Biol Chem 92:107461 (**IF: 1.85**)

4. Sanjay Kumar Singh, M. Sudhakara Reddy. PfIDmap: a database tool for mapping of different identifiers of *Plasmodium falciparum*. J Vector Borne Dis (*Under review*) (**IF: 1.127**)

Papers presented in the conference:

1. NextGen Genomics, Biology, Bioinformatics and Technologies (NGBT) Conference held from October 2- 4, 2017 in Bhubaneswar, Odisha, India organized by SciGenom Research Foundation (SGRF). *Poster presentation*: Functional annotation and pathway analysis of genes differentially expressed in different stages of *Plasmodium falciparum* using RNA-Seq Data

2. Inbix17: Indian conference on Bioinformatics 2017 held from November 7 – 9, 2017 in Jaipur, India organized by Birla Institute of Scientific Research in association with Bioclues. *Poster presentation*: Functional annotation and protein interaction network analyses of RBPs differentially expressed in different stages of Plasmodium falciparum using RNA-Seq Data

Abstract

The drug resistance in malaria parasites is increasingly emerging, so it is important to discover and develop alternative anti-malarial agents against both new and established drug targets. In this study, a comprehensive computational approach was used to derive potential therapeutic targets for *P. falciparum* using RNA-seq data set. The differential expression of genes, functional and pathway enrichment analyses of *P. falciparum* has been apprized in detail. The present study results suggested that PF3D7_0705600, PF3D7_1207100, PF3D7_0508100, PF3D7_1126700, and PF3D7_1234300 hub genes might serve as putative targets for drug designing. These hub genes are showing less mutation and no similarity with human proteins. In addition, the gene finding strategies of this study resulted into a database of different identifiers. This developed database tool (www.cdkd.org/pfidmap/) provides easy mapping of different identifiers related to *P. falciparum*. Functional analysis of the nsSNPs of identified hub genes was undertaken to predict deleterious mutations using various computational approaches and a database has been developed to demonstrate the analysis done by PROVEAN, SIFT, PredictSNP and NetSurfP software, which is available online at www.cdkd.org/pfsnp/. Moreover, the effect of deleterious mutations in glycosylphosphatidylinositol transamidase (GPI-T) subunit GPI8p has been investigated, which could be considered as a potential drug target primarily because of its crucial role in the GPI anchor biosynthesis pathway for the development as well as survival of the parasite. Thiamine phosphate synthase (PfThiE), an essential metabolic gene in the thiamine biosynthesis pathway is also studied and potential inhibitors were identified through docking-based virtual screening along with drug-likeness and ADMET analysis to derive therapeutic targets for *P. falciparum*. Therefore, hub genes identified in this research, GPI8p and PfThiE may be considered as potential drug targets for *P. falciparum*.

Table of Contents

List of figures

List of tables

Abbreviations

- ACT Artemisinin-based combination therapy
- ASA Accessible surface area
- BLAST Basic Local Alignment Search Tool
- BRENDA Braunschweig Enzyme Database
- DAVID Database for Annotation, Visualization and Integrated Discovery
- DDG Free Energy change value
- DEGs Differentially expressed genes
- ET Early trophozoite
- FAC Functional Annotation Clustering
- FPKM Fragments per kilobase per million
- GII Gametocyte stage II
- GO Gene ontology
- GV Gametocyte stage V
- HTML Hypertext Markup Language
- HTTP Hypertext Transfer Protocol
- IDC Intraerythrocytic developmental cycle
- KAAS KEGG Automatic Annotation Server
- KEGG Kyoto Encyclopaedia of Genes and Genomics
- KO KEGG orthology
- LT Late trophozoite
- MCODE molecular complex detection
- MPMP Malaria Parasite Metabolic Pathways
- ncRNA Non-coding RNA
- NDRs Nucleosome depleted regions
- NGS Next Generation Sequence
- nsSNPs Non-synonymous single nucleotide polymorphisms
- NVBDCP National Vector Borne Disease Control Program

- Oo Ookinete
- PASs Polyadenylation sites
- PDB Protein Data Bank
- PfIDmap Identifiers mapping tool for *Plasmodium falciparum*
- PfThiE Thiamine phosphate synthase
- PHP Hypertext Preprocessor
- PredictSNP Consensus classifiers for prediction of disease-related mutations
- PROVEAN Protein Variation Effect Analyzer
- PV Parasitophorous vacuole
- PVM Parasitophorous vacuole membrane
- QC Quality control
- R Ring stage
- RBCs Red blood cells
- RCSB Research Collaboratory for Structural Bioinformatics
- RI Reliability Index
- RNA-Seq RNA sequencing
- SBH Single-directional best hit
- Sc Schizont
- SIFT Sorting Intolerant From Tolerant
- SNAP2 Screening for Non-acceptable Polymorphisms
- SNPs Single nucleotide polymorphisms
- SQL Structured Query Language
- STRING Search Tool for the Retrieval of Interacting Genes/Proteins
- TNF Tumor Necrosis Factor
- WHO World Health Organization

INTRODUCTION

Chapter 1

Introduction

1.1 General Introduction

Malaria, a deadly infectious disease, is caused by intracellular single-celled parasites belongs to the genus *Plasmodium* and is transmitted by infected female Anopheles mosquitoes through the bites. Malaria remains one of the world's deadliest infectious diseases with up to one million deaths each year and has been recognized as one of the most powerful evolutionary selection in the human genome. There are five separate *Plasmodium* species that are capable of infecting humans; *P. falciparum, P. vivax, P. malariae, P. ovale,* and more recently *P. knowlesi*. The most severe virulent malaria leading to death is caused by *P. falciparum,* especially in children under the age of five (Le Roch et al. 2012).

After seven years of international effort first draft of the *P. falciparum* genome was published. Using the Sanger method and the chromosome shotgun technique, the genome was sequenced (Gardner et al. 2002). It was originally estimated that the size of the genome was 22.8 Mb, divided into 14 chromosomes. In addition to its nuclear genome, the parasite contains 6 kb and 35 kb circular DNA found in its mitochondria and plant related apicoplast respectively. The *P. falciparum* genome remains to be the most AT-rich genome. The overall (A+T)-composition is 80.6%, and can rise to 95% in introns and intergenic regions. It contains 6372 genes and 5524 protein-coding genes (genome version: 06-01-2010, http://plasmodb.org/plasmo/).

Despite efforts to develop vaccines and drugs to combat malaria, vaccine escape and drug resistance continues to be a problem (Hay et al. 2010; Imwong et al. 2010; Cohen et al. 2010). The ability to compare whole genomes could assist in these efforts, as genetic variation and recombination have been shown to facilitate antigen diversity, immune escape and evolution

of anti-malarial drug resistance (Imwong et al. 2010; Frank et al. 2008; Mackinnon and Marsh 2010; Ferreira et al. 2004; Good and Doolan 2010).

Traditional first-line therapies like chloroquine and pyrimethamine/sulphadoxine have lost their efficacy in most countries, which resulted in the development of new and more effective anti-malarial medicines, like artemisinin-based combination therapy (ACT) (Goswami et al. 2013). In recent years, parasite resistance to artemisinins has been detected. Drug resistance is increasingly emerging in malaria parasites, so it is important to identify and develop alternative anti-malarial agents against both new and existing drug targets.

High-throughput RNA sequencing (RNA-seq) approaches are used to perform transcriptome assays. Next Generation Sequence (NGS) technologies have a variety of advantages compared to microarrays, including single base pair resolution, low background signal, a wide dynamic range of expression levels across which transcripts can be detected, lower sample requirements for starting RNA, and no restriction on detecting transcripts that do not conform to a genome previously sequenced (Wang et al. 2009). The study of RNA dynamics within a cell has been revolutionised by RNA-Seq, which applies high-throughput sequencing technology to the transcriptome of an organism (Wang et al. 2009). The existence and quantity of the transcripts can be measured through millions of short read sequences. It has been shown that RNA-Seq has a greater dynamic range than microarrays for gene expression levels (Wilhelm et al. 2008) and allows scientists to display the transcriptome at the resolution of single nucleotides. RNA-seq technology allows the precise identification of isoforms of genes, events of translocation, changes of nucleotides and post-transcriptional base modifications.

Several pathway databases are currently documenting metabolic pathways and gene signalling networks, such as KEGG (Kyoto Encyclopedia of Genes and Genomes) (Ogata et al. 1999), BioCyc (Caspi et al. 2016), BioCarta (http://www.biocarta.com), and Reactome (Joshi-Tope et al. 2005), providing the potential for more nuanced and useful research.

The present study aimed to identify putative drug targets in *P. falciparum*. A comprehensive approach has been used to derive potential therapeutic targets for *P. falciparum* using RNA-seq data set. A web tool PfIDmap was created to retrieve different identifiers of from various databases related to *P. falciparum*. The differential expression of genes, functional and pathway enrichment analysis of *P. falciparum* were apprised in detail. Several bioinformatics approaches were used to characterize the hub genes. A computational approach was undertaken to systematically analyse the functional consequences of deleterious nsSNPs in identified hub genes. A metabolic enzyme, thiamine phosphate synthase (PfThiE) was studied to identify SNP, functional analysis, stability analysis, conservation analysis, and 3D structure modelling. In addition, the 3D structure of PfThiE was used for the screening of compounds, ADMET analysis and molecular docking studies. The ultimate purpose of this study is to identify the putative drug targets in *P. falciparum*.

1.2 Gap in study

Malaria is a caused by a mosquito-borne eukaryotic protozoan parasite of the genus *Plasmodium*. In most countries, conventional first-line therapies such as chloroquine and pyrimethamine / sulphadoxine have lost their effectiveness, resulting in the production of new and more powerful anti-malarial medications, such as artemisinin-based combination therapy (ACT) (Goswami et al. 2013). In recent years, parasite resistance to artemisinins has been detected. Though efforts have been made to develop vaccines and medicines to tackle malaria, the issue of vaccine escape and drug resistance still persists (Hay et al. 2010). Drug resistance is increasingly emerging in malaria parasites, so it is important to identify and develop alternative anti-malarial agents against both new and existing drug targets. A systematic approach to derive potential therapeutic targets for *P. falciparum* using RNA-seq data set was used in this analysis. The differential expression, functional and pathway enrichment analysis

of *P. falciparum* genes were evaluated in detail. In addition, a computational approach was undertaken to systematically analyse the nsSNPs to predict deleterious mutations.

1.3 Specific objectives of present investigation

1. To identify novel gene/isoform/exon prediction from *P. falciparum* genome
2. Differential expression analysis of *P. falciparum*
3. Identification of SNPs from various strains of *P. falciparum*
4. Computational analysis of metabolic pathways of *P. falciparum*

REVIEW OF LITERATURE

Chapter 2

Review of Literature

2.1 Malaria

Malaria, a vector-borne disease is caused by intracellular single-celled parasites belongs to the genus *Plasmodium*. Female *Anopheles* mosquitoes transmit malaria from person to person. Malaria antigen has recently been found in remains of Egyptian origin in 3200 and 1304 BC (Miller et al. 1994). Indian Vedic-era writings (1500 to 800 BC) named malaria the "king of diseases" (Gelband et al. 2004). The original Italian word 'mala aria' means 'bad air' and was used in the 1700s to describe the symptoms of the disease and related circumstances. In 1880 Alphonse Laveran (1845-1922) observed the first gametocyte of malaria in the blood of a French soldier in Algeria, a finding that earned the 1907 Nobel Prize for medicine. Sir Ronald Ross (1857-1932) of Britain, an army surgeon employed in Secunderabad India, proved in 1897 that mosquitoes spread malaria. This achievement earned him Knighthood and the 1902 Nobel Prize. The World Mosquito Day (August 20) commemorates the discovery of the connection between mosquitoes and malaria transmission in 1897 by Sir Ronald Ross (Gelband et al. 2004). Shortt and Garnham solved the third piece of the human malaria puzzle in 1948, where mosquito-inoculated sporozoites underwent early production in the human host.

Like other protozoa, in their two-host life cycle, plasmodia go through a variety of stages. *P. falciparum, P. vivax, P. knowlesi, P. malariae,* and *P. ovale* are different species that cause disease in humans. The most dangerous type of malaria causing agent in humans is *P. falciparum*. The reason for the virulence of *P. falciparum* lies in the ability to invade red blood cells (RBCs) of all ages that cause very high parasitaemia, attain high multiplication rates (up to 24 merozoites vs 8-10 *P. vivax* merozoites) and enhanced growth, and the ability to adhere to vascular endothelium through the sequestration process (Rowe et al. 2009). *Plasmodium*

falciparum expresses variant surface antigens on the membrane of the erythrocyte, allowing the parasite to bind to specific receptors on endothelial host cells. This so-called sequestration helps to avoid clearing out the spleen of the infected red blood cells (Miller et al. 1994; Kyes et al. 2001; Ashley et al. 2018). The erythrocyte membrane protein 1 (PfEMP1) of *P. falciparum* mediates cytoadherence, a series of proteins exported to the contaminated erythrocyte surface through clonal variants and encoded by the family of var genes (Ashley et al. 2018).

The most serious malignant malaria is *P. falciparum*, especially in children under the age of five (Le Roch et al. 2012). Nearly 95% of the population in India lives in endemic malaria areas, and 80% of the identified malaria in India is restricted to 20% of the population areas in rural, hilly, remote or hard-to-reach areas (Sharma 2012). Conventional first-line therapies such as chloroquine and pyrimethamine/sulphadoxine have lost their efficacy in most nations, resulting in the development of new and more potent anti-malarial medicines, such as artemisinin-based combination therapy (ACT) (Goswami et al. 2013). Though efforts are being made to develop vaccines, medicines to eradicate malaria, the issue of vaccine escape, and drug resistance is persist (Hay et al. 2007).

2.1.1 Malaria burden

According to World Malaria Report 2019 by World Health Organization (WHO), about 228 million malaria cases and 4,05,000 deaths were reported in 87 nations in 2018. In 2018, *P. falciparum* reported for 99.7% of confirmed cases of malaria in the African region of WHO, comprising 50% of cases in the South East Asia region of WHO, 71% in the Eastern Mediterranean and 65% in the Western Pacific region. In 2018, 6 countries accounted for more than half of all malaria cases worldwide: Nigeria (25%), the Democratic Republic of Congo (12%), Uganda (5%), and Côte d'Ivoire, Mozambique, and Niger (4% each). India carries the

global burden of malaria by 4% and contributing 87% of South East Asia 's overall malaria cases (Ghosh and Rahi 2019). As the seventh largest country in the world by geographic area, India is the second most populated country with a population of 1,339 billion. In India, an estimated 95% of the population lives in areas where transmission of malaria has been recorded or where climate conditions favour transmission. Eighty percent of India's malaria is limited to 20% of the population, mostly indigenous people live in hilly, difficult and inaccessible terrain (Sharma 2012). According to UN figures, in India there is a chance of malaria in around 1,251 billion (93.4%) population (Kumar 2019). The geographical distribution by country of malaria incidence is shown in the Fig. 2.1.

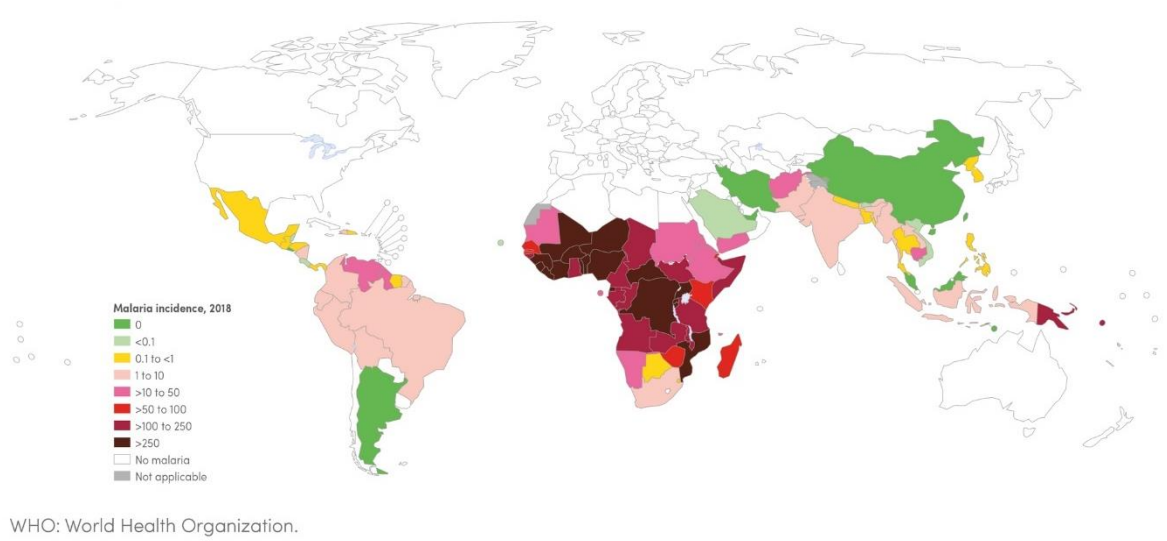


Fig. 2.1 Map of malaria case incidence rate (cases per 1000 population at risk) by country, 2018 (WHO 2019)

It is responsible for up to US$ 200 billion of annual economic toll and leads to the global challenge of antimicrobial resistance (Ghosh and Rahi 2019). The National Vector Borne Disease Control Program (NVBDCP) reported a total of 338494 confirmed malaria cases which include 156940 cases due to *P. falciparum* malaria, and 77 associated deaths in 2019 (NVBDCP 2020).

2.1.2 The agent of malaria

Malaria occurs when a mosquito infected with the parasite *Plasmodium* bites humans. *Plasmodium falciparum*, *P. vivax*, *P. knowlesi*, *P. malariae*, and *P. ovale* are five species of malarial parasite that cause disease in humans. These species are closely related to each other, indicating that adaptation to humans has taken place separately many times during the genus history. Nevertheless, it is still unknown when these associations began, and where they came from (Prugnolle et al. 2011). Especially the origin of *P. falciparum* remains a highly debated subject. Recently it has been shown that *P. falciparum* is belongs to the *Laverania* subgenus. Otto et al. (2018) showed that the characteristics of *P. falciparum* which make it the only Laverania member capable of infecting and spreading in humans. Liu et al. (2010) indicated that *P. falciparum* is of gorilla rather than chimpanzee, bonobo or ancient human descent. Apicomplexa, a large group of unicellular protozoans, are distinguished by a dense electron structure at the merozoite's apical pole, the invasive form of the parasite, which enables the parasite to invade and establish itself in the host cells. In addition, most Apicomplexa parasites have an apicoplast, a secondary endosymbiotic vestigial plasmid that harbours essential biochemical pathways and is indispensable for the growth of parasites (Lim and McFadden 2010).

Both *P. vivax* and *P. ovale* that develop hypnozoites, dormant liver stages that can cause chronic disease to rebound long after the initial infection occurs. In comparison, *P. malariae* and *P. falciparum* do not form hypnozoites, but *P. malariae* has recorded recurrences of permanently disposable blood stages (Dembélé et al. 2014). Parasite sporozoite inoculation occurs through bite of infected blood-feeding genus *Anopheles* female mosquitoes. There are more than 30 Anopheles subspecies capable of transmitting *Plasmodium* parasites but *A. gambiae* and *A. funestus* are the most important sub species that account for the highest transmission rate in Africa (Tuteja 2007). *A. stephensi* is a major vector of malaria in India's

urban areas, and is primarily spread in other countries, including Thailand, Myanmar, Pakistan, China, Afghanistan, Iran and Iraq (Tyagi et al. 2017).

The most dangerous form of malaria causative agent for humans is *P. falciparum*. *P. falciparum* sporozoites migrate to the liver where they grow exponentially in hepatocytes to form schizonts. These schizonts rupture after an incubation period to release merozoites that in turn invade the erythrocytes. Mosquitoes receive *Plasmodium* gametocytes with a blood meal, and they undergo another reproductive process within the mosquito before being moved to another human host. The clinical symptoms of malaria are primarily attributed to the replication of asexual stages in human blood, but only through the development of sexual stages, called gametocytes, is the transmission to mosquitoes achieved (Meibalan and Marti 2017).

2.1.3 Pathogenesis and pathophysiology of malaria

Pathogenesis, the way a disease evolves, for a clinical illness, involving human malaria is a complicated story with several participants, environments and possible outcomes (Milner 2018). Genetic traits, which protect people from the disease, can be found in the human genome, as Haldane first observed in the case of sickle cell trait (Kwiatkowski 2005). The process of the parasites occurs in both humans (asexual stages) and mosquitoes (sexual stages). Sporozoites, the malaria parasite's infectious form, are poured into a human host by the saliva of an *Anopheles* mosquito. Within minutes, these sporozoites enter the hepatocytes, take on a new form and multiply. The blood-stage parasites recognized as merozoites are released when the hepatocytes rupture. Released merozoites invade new red blood cells, where the parasites undergo repeated rounds of growth, replication, egress, and invasion. *Plasmodium* uses haemoglobin, which is an essential source of nutrition and energy, to expand and replicate within human red blood cells. Nevertheless, this process produces toxic heme, which is aggregated by the parasite into an insoluble biocrystal called hemozoin. This molecule sequestration in various organs (liver, spleen, and brain), possibly leading to the development

of immunopathogenesis of malaria (Moxon et al. 2020). Individuals diagnosed with *P. falciparum* can have moderate (e.g., fever, chills, sweating, vomiting, nausea, malaise) or extreme clinical symptoms (e.g., respiratory distress, pulmonary oedema, severe anaemia, acute renal failure, cerebral malaria) (Degarege et al. 2019).

Cyclic fever has also been known as a common symptom of malaria, even before *Plasmodium* parasites were recognized as the disease's etiological agents. The duration from initial malaria infection to the emergence of symptoms (incubation period), usually varies. A brood of schizonts matures 48 hours in *P. vivax* and *P. ovale* malaria, and the periodicity of fever is tertian ("tertian malaria"), while in *P. malariae,* fever occurs every 72 hours ("quartier malaria"). Fever may occur every 48 hours in *falciparum* malaria, but it is typically sporadic, showing no distinct periodicity (Crutcher and Hoffman 1996).

2.1.4 The lifecycle of *P. falciparum*

The life cycle of *P. falciparum* parasites can be divided into two stages, alternating between mosquitoes and humans (Fig. 2.2). The cycle can be divided into three sequential multiplication stages: two schizogony stages (asexual multiplication) in the vertebrate host, first in hepatocytes, then in RBCs, and one sporogony phase (sexual multiplication) in the mosquito.

In order to start the asexual cycle in humans, an infected female Anopheles mosquito injects sporozoites during a blood meal in the new human host. Sporozoites, the parasite's motile and infective form, travel through the host’s bloodstream from the mosquito’s salivary glands to the liver, where they divide and differentiate in hepatocytes to form hepatic schizonts.

The pre-erythrocytic stages are called the sporozoites and the hepatic stages. Such schizonts rupture after an incubation time of about 10 days to release up to 10,000 merozoites that in turn invade the erythrocytes. When merozoites are in the bloodstream, they invade erythrocytes and

may either develop via an asexual period of 48 hours, or develop sexually. These merozoites start the erythrocytic stage by invading and replicating red blood cells (RBCs). In erythrocytes,

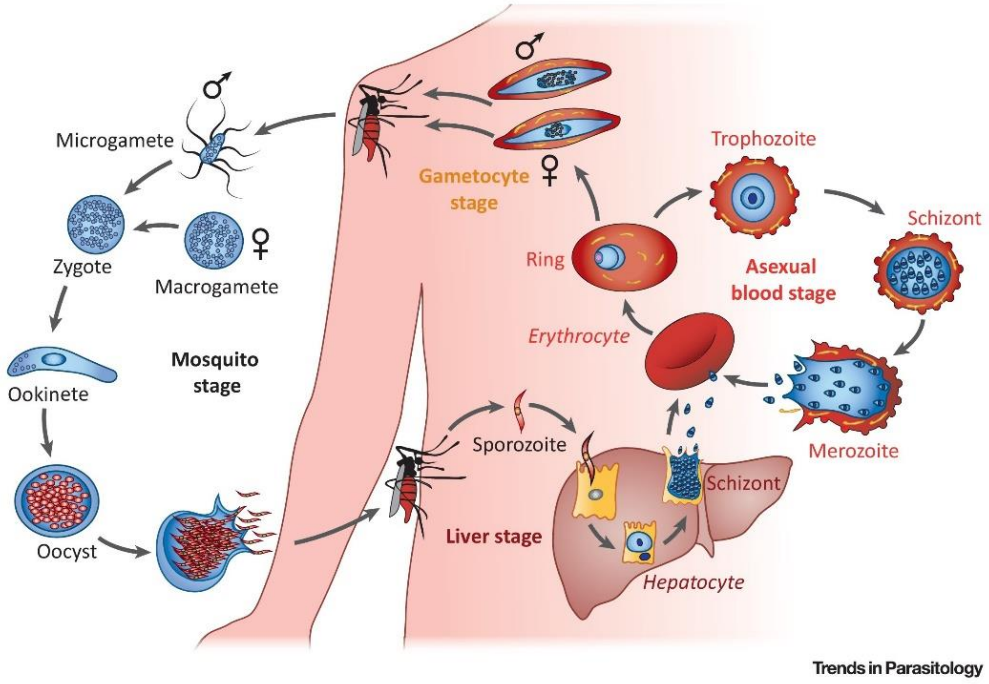


Fig. 2.2 Schematic depiction of the life cycle of *P. falciparum*. Sporozoites are inserted into the dermis from the bite of an infected Anopheles mosquito. Sporozoites migrate to the liver through the blood stream and invade to hepatocytes. They grow into thousands of merozoites and released to infect erythrocytes into the bloodstream. Asexual reproduction is undergone by the parasite. A small proportion grows into gametocytes, which is the sexual form of the parasite. An Anopheles mosquito may take up gametocytes and undergo sexual reproduction with a second blood meal (Reprinted from Maier et al. 2019 with permission from Elsevier).

the 48 h asexual development is complex, with three successive morphological stages (Ring, trophozoite and schizont stages) (Bannister et al. 2000). These parasites move from ring to trophozoite stages followed by a multinucleate schizont stage, which finally generates up to 32 merozoites that after rupture of the host cell re-infect new erythrocytes. Some of these merozoites grow into sexual stage gametocytes that are taken up during their blood meal by the female Anopheles mosquito. They re-emerge as sporozoites in their salivary glands after 10-12 days of development in the mosquito midgut and re-enter human circulation after fresh infection after mosquito bite (Dantzler et al. 2015; Josling and Llinas 2015).

2.1.4.1 The asexual blood stage

During 48-hour intraerythrocytic developmental cycle (IDC), merozoite, ring, trophozoite, and schizont stage parasites can be differentiated microscopically (Fig. 2.3).

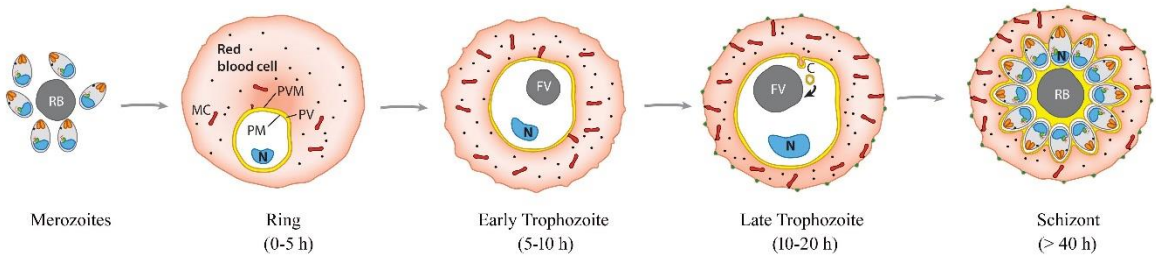


Fig. 2.3 The blood stage cycle of *P. falciparum*: Graphically the four major blood stages are represented by merozoite, ring, trophozoite (early and late), and schizont. The parasite encapsulates itself in a parasitophorous vacuole membrane during invasion. The ring stage delivers proteins to the parasite and produces Maurer's clefts. [Abbreviations: **h**: hours post invasion; **N**: nucleus; **PV**: parasitophorous vacuole; **PVM**: parasitophorous vacuolar membrane; **MC**: Maurer's cleft; **PM**: parasite membrane; **FV**: food vacuole; **C**: cytostome; **RB**: residual body (Boddey and Cowman 2013)

The non-motile merozoites via proteins located on the merozoite surface invades RBCs through a first contact and the resulting use of the machinery based on actin-myosin (Baum et al. 2008). The RBC invasion occurs in minutes in order to minimize interaction with the host immune system. Invaginating the RBC membrane and forming a parasitophorous vacuole membrane (PVM) creating the parasitophorous vacuole (PV) create the framework for the further growth of the parasite during invasion. Following invasion, the parasite becomes the ring stage and induces the first steps of modification of the host cell. The parasite therefore expresses a variety of proteins, which induce structures such as Maurer's clefts or Tubovesicular Network (TVN) (Atkinson and Aikawa 1990). The parasite proliferates from the ring to the trophozoite stage within the PV and increases its metabolism to establish a suitable place for intraerythrocytic survival.

To acquire space for growth and amino acid supply, the parasite uses proteolytically degraded haemoglobin, and the toxic haematin by-product is processed into a crystalline form known as

haemozoin and deposited in the food vacuole (Goldberg 2013). In addition, during the trophozoite period, which lasts around 22 to 36 hours after the invasion, the parasite initiates DNA replication, the number of ribosomes increases, and the endoplasmatic reticulum (ER) extends. The parasite turns into a schizont stage parasite after 36 hours of post-invasion and occupies most of the host cell. Finally, during this stage, it generates up to 32 daughter merozoites. The host cell ruptures and released merozoites invade new RBCs after 48 hours of post invasion. While the asexual number of parasites during the blood stage increases tremendously, a small proportion of merozoites exit the asexual replication cycle and evolves into sexual forms of the parasite.

2.1.4.2 Gametocytogenesis (Development of sexual stages)

Gametocytes are specific sexual precursor cells that mediate malaria parasite transmission from a mammalian host to the Anopheles mosquito. The formation of gametocytes in *P. falciparum* takes about 10 to 12 days, with all the merozoites of one schizont dedicated to either grow into gametes or begin their asexual cycle (Guttery et al. 2015; Amoah at al. 2020). Unlike the asexual blood-stages that are responsible for malaria's clinical outcome, gametocytes do not cause clinical manifestations. The process of changing from the asexual stage of blood to the gametocyte is called gametocyte commitment.

It is assumed that development of gametocytes takes place sometime before schizogony, in which each schizont produces a progeny of merozoites that mature in the blood stage into either sexual form (male or female gametocytes) or asexual parasites. The predominant female-biased sex ratio reported in malaria parasites is due to the development of a higher percentage of dedicated female schizonts than their male counterpart (Smith et al. 2000; Silvestrini et al. 2000). Experiments on *P. falciparum* and *P. berghei* have showed that gametocytogenesis commitment is triggered by activating the AP2-G transcription factor (Kafsack et al. 2014; Sinha et al. 2014). Gametocytes are the only malaria parasite type that can be transmitted to

the mosquito vector (Ngotho et al. 2019). The *P. falciparum* gametocytes subsequently undergo five developmental stages (I–V) (Fig. 2.4) over the course of 9-12 days that are morphologically discernible (Ngotho et al. 2019).

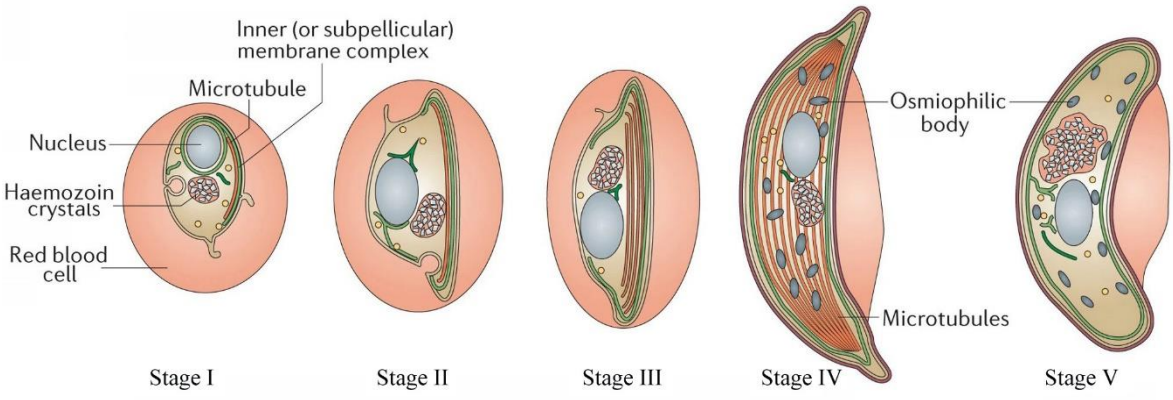


Fig. 2.4 Schematics picture of different stages of gametocytes of *P. falciparum*: Production of a mature gametocyte *P. falciparum* takes place over 10–12 days and is typically divided into five morphologically distinct stages (Reprinted from Josling and Llinás 2015 with permission from Springer Nature)

Various studies revealed that only mature gametocyte phases are found in the blood circulation in some species, whereas immature gametocyte sequesters in host tissues, especially bone marrow and spleen (Joice et al. 2014; De Niz et al. 2018), while in the peripheral blood vessels, stage V circulates (Tibúrcio et al. 2013; Joice et al. 2014). Stage I gametocytes look very close to asexual trophozoites and cannot be morphologically distinguished; nevertheless, genetic reporters can identify them as their transcriptome starts to vary from asexual parasites. Slight changes in appearance become evident in stage II, when the parasite assumes the shape of a lemon or oat grain with one pointing end while the parasite shape resembles the letter 'D' in stage III. During stage IV, gametocytes can be differentiated between male and female, both sexes possessing an elongated thin structure with pointy ends. The pigment tends to be scattered in male gametocytes, while it is denser in females. Because of the crescent or falciform shape from which the name '*falciparum*' (from the Latin '*falx*' (sickle) and '*parere*' (to give birth) derives, of all stages, stage V gametocytes are the most distinguishable. Stage V gametocytes

are located in the peripheral blood (Smalley et al. 1981), where they take another 3 days to become infectious to mosquitoes (Kuehn and Pradel 2010; Henry et al. 2019) and can be taken up by a blood-sucking female *Anopheles* mosquito.

2.1.4.3 The sexual phase of life cycle

When gametocytogenesis is complete, a female anopheles mosquito takes on mature male and female stage -V gametocytes while feeding the blood. Mature gametocytes can circulate several days in the human blood, maximizing their chances of transmission to mosquitoes. If ingested, gametocytes rapidly convert to male (microgamete) and female (macrogamete) gametes in response to environmental signals such as pH increase, temperature decrease, and xanthurenic acid exposure (Billker et al. 1997).

Proteases are used by male and female gametocytes to exit the RBCs a few minutes after reaching the midgut of the mosquito and distinguish between 8 microgametes and one macrogamete, respectively (Sologub et al. 2011; Venugopal et al. 2020). The fertile gametes have evolved within around 20 min. After fertilization, the zygote transforms within 1 day into an infective motile ookinete that leaves the lumen of the midgut and thus marks the end of the sexual phase of malaria. The ookinete crosses midgut wall's epithelial layer to form an oocyste. Parasites in the oocyst undergo third asexual replication process in order to create thousands of sporozoites released into the haemolymph. Sporozoites which enter the mosquito's salivary glands bind and invade the gland, remain until transmitted via a mosquito bite to a new vertebrate host to restart the process (Venugopal et al. 2020).

2.1.4 Diagnosis of malaria infection

Rapid and precise diagnosis is key to successful malaria treatment. Diagnosis of malaria involves the detection of parasites or antigens/products in the blood of patients. Quick diagnosis not only alleviates suffering, but also declines community transmission. In many countries, diagnosis and treatment delays are the leading causes of death (CDC 2008;

Tangpukdee 2009). Health doctors usually have a clinical diagnosis of malaria. This approach is less expensive and more commonly used. Clinical diagnosis is depending on the signs and symptoms of the patients, as well as on physical observations during examination. The earliest malaria signs include fever, headache, weakness, chills, dizziness, vomiting, nausea, abdominal pain, diarrhea, muscle pain and fatigue (Landier et al. 2016). Malaria is diagnosed in the laboratory using various techniques, such as traditional microscopic diagnosis by staining thin and thick peripheral blood smear, quantitative buffy coat (QBC) method, rapid diagnostic tests and molecular diagnostic methods such as polymerase chain reaction (PCR).

More than a century later, microscopy remains the gold standard for laboratory diagnosis in the detection of malaria parasites and the identification of *Plasmodium* species in Giemsa-stained thick blood films (for malaria parasite detection) and thin blood films (for species identification) (Bharti et al. 2007). Molecular diagnostic methods such as PCR have an advantage over manual microscopy and RDT serodiagnosis. This method is reliable and can be used to diagnose malaria.

2.1.5 Prevention and treatment

Uncomplicated *P. falciparum* malaria, especially in nonimmune individuals, can progress rapidly to serious illness and death. In treating uncomplicated malaria caused by *P. falciparum*, the WHO recommends artemisinin-based combination therapies (ACTs). The key benefit of the combination treatment is that artemisinin eliminates the majority of malaria parasites rapidly and dramatically, and the companion medication removes the remaining small number of parasites (WHO 2015, 2018b; Naing et al. 2019). It is therefore necessary for a malaria diagnosis to be made soon after the onset of symptoms of malaria and for antimalarial treatment to be started without delay. While insecticides, vector control, bed nets, and antimalarial medicines are otehr most effective tools for disease prevention.

Established anti-malarial therapies include a variety of medications that administered alone or in combination and can be grouped based on their drug class (Table 2.10). The first antimalarial medicine, extracted from the bark of the Cinchona tree in 1820, was Quinine. It has been found highly effective against late blood stages (Achan et al. 2011). In 2009, 31 African countries recommended quinine as a second-line treatment for non-malaria, 38 for the first-line treatment of extreme malaria, and 32 for the first-trimester treatment of malaria (WHO 2009).

In circumstances where artemisinin is not available, it is still on the WHO Model List of Essential Medicines (MLEM) for treating severe malaria. For any parasite sensitive to the medication, chloroquine is the preferred treatment. But in many regions of the world, malaria-causing parasites are chloroquine-resistant, and treatment is no longer an effective remedy. ACTs are the most effective antimalarial drugs available today, by combining 2 active ingredients with specific mechanisms of action. Five ACTs recommended by WHO include: artemether-lumefantrine (AL), artesunate- amodiaquine (ASAQ), artesunate- mefloquine (ASMQ), artesunate plus sulfadoxine-pyrimethamine (ASSP) and dihydroartemisinin-piperaquine (DHP) for the treatment of uncomplicated *P. falciparum* malaria (WHO 2018b; Naing et al. 2019). RTS,S / AS01 is the world's first malaria vaccine to be shown to have partial immunity against malaria in young children (Laurens 2020). A drug safety review of the antimalarial drug hydroxychloroquine was performed for SARS COV 2 epidemic prophylaxis (Singh et al. 2020).

The drug class Cinchona alkaloids (Quinine and Quinidine) accumulate and form toxic haem complexes in food vacuoles, while Primaquine was thought to inhibit the parasite's oxidative metabolism. Phenanthrenes and derivatives (Halofantrine) causes parasite membrane damage by forming cytotoxic complexes whereas benzene and substituted derivatives (Sulfadoxine, Sulfamethoxypyridazine and Proguanil) and Diazines (Pyrimethamine) can inhibit synthesis of

Table 2.1 Anti-malarial FDA approved drugs (Ramakrishnan et al. 2017)

Drug name	Drug class	Anti-malarial activity	Side effects
Quinine Quinidine	Cinchona alkaloids	Accumulates in food vacuoles and forms toxic haem complexes	Side effects include hearing impairment, rashes, vertigo, vomiting and in some cases neurotoxicity
Mefloquine	Quinolines and derivatives		Nausea, dizziness, diarrhoea, bradycardia and neurotoxicity
Chloroquine			May cause psoriasis
Amodiaquine			Vomiting, dizziness and in some cases hepatic disorders
Primaquine		Believed to block oxidative metabolism in the parasite	Anorexia, vomiting, cramps and anaemia
Halofantrine	Phenanthrenes and derivatives	Causes parasite membrane damage by forming cytotoxic complexes	Nausea, diarrhoea, itching and high cardiotoxicity
Sulfadoxine Sulfamethoxypyridazine	Benzene and substituted derivatives	Inhibit synthesis of folates	Skin reactions (rare)
Proguanil			Very few: hair loss and mouth ulcers
Pyrimethamine	Diazines		Occasional rashes
Tetracycline	Tetracyclines	Inhibits translation	–
Doxycycline			Depression of bone growth and gastrointestinal disturbances
Clindamycin	Carboxylic acids and derivatives	Inhibits protein synthesis	Nausea, vomiting and cramps
Azithromycin	Macrolides and analogues		May cause angioedema and jaundice
Artemisinin	Lipids and lipid-like molecules	Believed to affect mitochondrial electron transport chain or disrupt cellular redox cycling or inhibition of haem metabolism	Nausea, anorexia, dizziness and neurotoxicity
Atovaquone	Naphthalenes	Affects mitochondrial electron transport chain	May cause rashes, diarrhoea and headache

folates. Tetracyclines (Tetracycline and Doxycycline) inhibits translation while carboxylic acids and derivatives (Clindamycin) and Macrolides and analogues (Azithromycin) inhibits protein synthesis. Artemisinin and its derivatives believed to affect electron transport chain of mitochondria or disrupt cellular redox cycling or inhibition of haem metabolism (Balint 2001). Atovaquone is used in the treatment of children and adults with uncomplicated malaria as a fixed-dose combination with proguanil (Malarone) or as chemoprophylaxis to avoid malaria in travellers by influencing the electron transport chain of mitochondria (Nixon et al. 2013).

2.2 *Plasmodium falciparum* genome

An international initiative to sequence the *P. falciparum* genome was initiated in 1996 with the hope that the sequence of the genome would open new avenues for research (Hoffman et al. 1997). After seven years, the first draft of the *P. falciparum* genome was published in 2002 by sequencing through the Sanger technique and the chromosome shotgun technique (Gardner et al. 2002). The genome of *P. falciparum* is approximately 130 times smaller than the genome of the human (Sibley 2019). It was originally estimated that the size of the genome was 22.8 Mb, divided into 14 chromosomes. In addition to its nuclear genome, the parasite comprises 6 kb and 35 kb of circular DNA in its mitochondria and plant related apicoplast, respectively. The *P. falciparum* genome is AT-rich genome today and the overall (A+T)-composition is 80.6% and can increase in introns and intergenic regions to 95%. It consists of 6372 genes and 5524 protein-coding genes (genome version: 06-01-2010, http://plasmodb.org/plasmo/).

In the first draft, approximately 60 per cent of genes did not have adequate homology to allocate predicted functional annotation to characterised genes from model organisms and were instead annotated as “hypothetical”. Most of this terminology has been discarded, and "unknown" function and "putative" function (where function is predicted, but confirmation of experiments is required) genes now have different annotations. Genome annotation has revealed major insights into the biology of the parasites.

For example, in contrast to other unicellular, non-parasitic eukaryotes, the *P. falciparum* genome has fewer genes that encode transporters and metabolic enzymes; it lacks genes encoding enzymes for de novo amino acid and purine synthesis; and 1.3% of its genes are associated with immune evasion and cell adhesion (Gardner et al. 2002). During the erythrocytic lifecycle, the parasite expresses much of its genome cyclically and during this period, 5% of the annotated *P. falciparum* genes have clonally variant expression (Duffy et al. 2013).

2.3 Transcriptome (RNA-Seq) data analysis

Sanger sequencing has been used to sequence a significant fraction of the genomes commonly used in modern databases, along with the human genome. A variety of sequencing technologies are established in the past few years and able to produce more sequence production compared to traditional Sanger sequencing. These approaches are collectively referred as next-generation sequencing (NGS). These technologies, such as the Solexa Genome Analyzer (Illumina, San Diego), SOLiD platform (Applied Biosystems; Foster City; CA, USA), 454 Genome Sequencers (Roche Applied Science, Basel) and HeliScope Single Molecule Sequencer (Helicos, Cambridge, MA, USA) have been released as commercial products. These technologies generate reads of 25 to 250 bps in length and up to 40 million reads per run (Barba et al. 2014). The sequencing with NGS approaches of expressed RNAs is abbreviated as RNA-Seq, a number of studies appeared that applied RNA-Seq to transcriptomes of various species (Wang et al. 2009; Mortazavi et al. 2008; Cloonan et al. 2008). These papers and reviews (Wang et al. 2009; Costa et al. 2011; Grabherr et al. 2011) have established RNA-Seq experiments that provide in-depth transcriptional landscape knowledge with unprecedented sensitivity, throughput, and outperforming previous expressed sequence tags (EST) sequencing and microarray splicing techniques.

RNA sequencing (RNA-seq) has become most widely used technology in the last decade because of declining costs and popularization of shared-resource sequencing core in many research institutions. Since the discovery of the function of RNA as the primary intermediate between genome and proteome, distinct core activities in molecular biology have been transcript identification and quantification of gene expression. The protocol of RNA-seq starts with the conversion of RNA population into cDNA fragments to be sequenced. Using a high-throughput platform, each cDNA fragment is subsequently sequenced or "read" after fragmentation, adapter ligation, and index ligation. Illustration of typical RNA-seq experiment provided in Fig. 2.5.

RNA-Seq is a recently established transcriptome profiling method that uses technologies for deep sequencing. Our outlook on the degree and complexity of transcriptome has already been altered by studies using this method. RNA-Seq also gives a much more reliable measurement of transcript levels and their isoforms than other techniques. A population of RNA (total or fractionated, such as poly$(A)^+$) is typically converted to a cDNA fragment library with adaptors attached to one or both ends. To obtain short sequences from one end (single-end sequencing) or both ends (pair-end sequencing), each molecule, with or without amplification, is then sequenced in a high-throughput manner. The evaluation of host-parasite interactions in malaria through transcriptome analysis offers a model for understanding the role of systemic host-pathogen interactions in general (Lee et al. 2018). Technical and bioinformatic advances have enabled the association of RNA expression with fundamental biology, immunity, pathogenesis, diagnosis and prognosis to be analyzed increasingly in depth (Lee et al. 2018). Whole blood (leukocytes and RBCs) may be used as a source for both host and parasite cells, and transcriptomic analyses can be performed in the same sample to test one cell type or both cell type (Daily et al. 2004).

High-throughput RNA sequencing (RNA-seq) methods perform transcriptome assays. Next Generation Sequence (NGS) technologies have a range of benefits when compared with microarrays. Single base pair resolution, low background signal, a wide dynamic range of expression levels over which transcripts can be detected, smaller sample criteria for starting

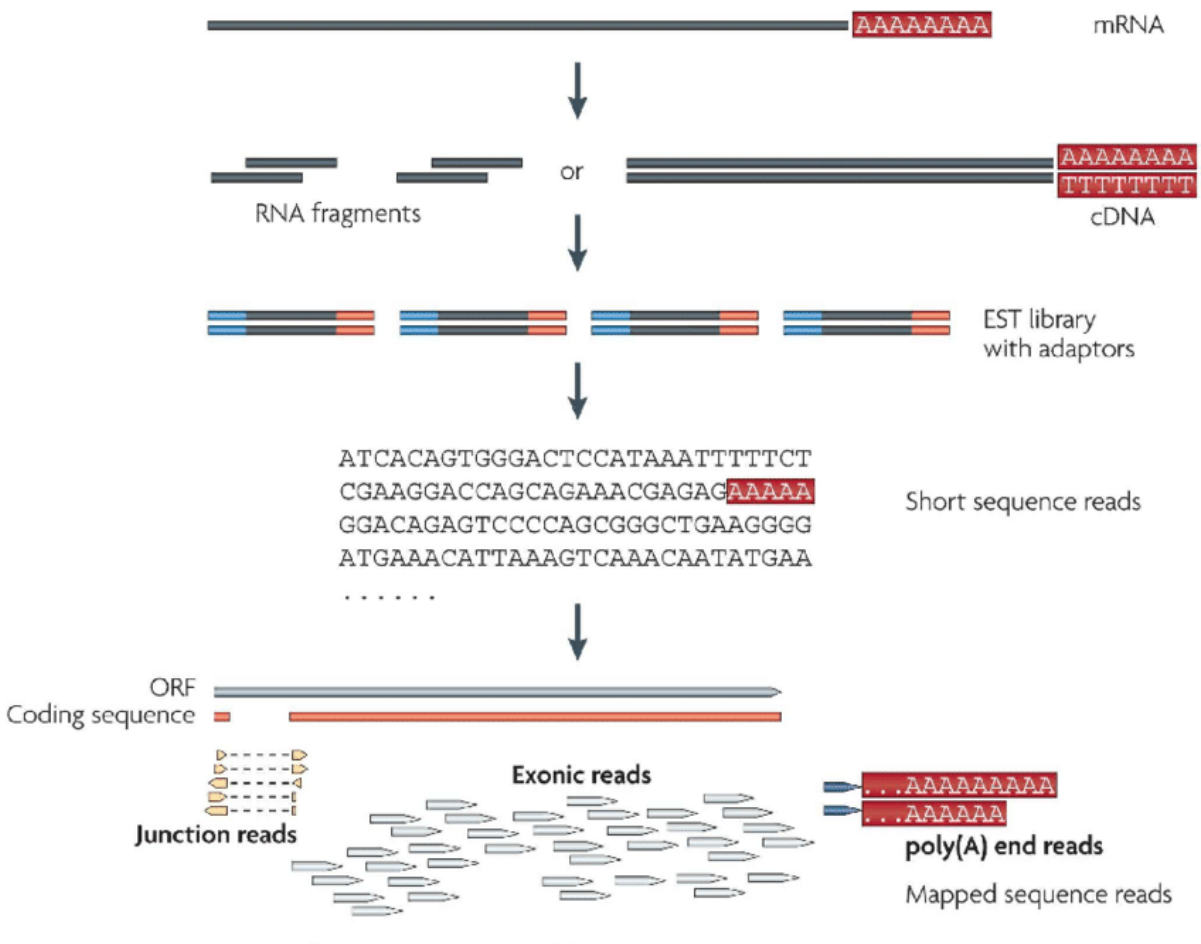


Fig. 2.5 A typical RNA-seq experiment (Wang Z et al. 2009)

RNA, and no restriction on detecting transcripts that do not conform to the genome previously sequenced are included in NGS technologies (Wang et al. 2009).

The study of RNA dynamics within a cell has been revolutionised by RNA-Seq, which applies high-throughput sequencing technology to the transcriptome of an organism (Wang et al. 2009). The existence and quantity of transcripts can be measured by millions of short read

sequences. RNA-Seq has been shown to have a greater dynamic range than microarrays for gene expression levels (Wilhelm et al. 2008) and it helps researchers at single nucleotide resolution to view the transcriptome. RNA-seq technology allows gene isoforms, translocation events, nucleotide variants and post-transcriptional base modifications to be described in detail.

RNA-Seq has clear advantages over the current methods on NGS platforms. First, RNA-Seq, unlike hybridization-based technologies, is not limited to the detection of existing transcripts, thus enabling new splice isoforms to be detected, characterised and quantified. It helps researchers to determine the proper annotation of genes, gene transcriptional boundaries, and expressed Single Nucleotide Polymorphisms (SNPs). The low 'background signal', the lack of an upper limit for quantification and, therefore, the wider dynamic range of expression levels across, which transcripts can be detected are other advantages of RNA-Seq compared to microarrays. High reproducibility levels for both technical and biological replicates are also present in RNA-Seq data.

Available transcriptomic (RNA-seq) data sets for *P. falciparum*

Le Roch data set

Le Roch et al (2003) proposed a strategy for gene function annotation with the help of expression profiling. Gene expression profiling during seven time points in *P. falciparum* of the erythrocytic asexual stage (early and late ring stage, early and late trophozoite, early and late schizont, and merozoite), the sexual stage (gametocytes present in human blood) and the mosquito stage (salivary gland sporozoite) were performed. In a one-color approach, gene transcript levels with custom built high-density oligonucleotide arrays covering predicted coding and non-coding sequences were studied. The study showed that genes with similar roles exhibit similar profiles of expression and therefore be used to further annotate functionally uncharacterized genes through microarray experiments.

Bártfai data set

Bártfai et al (2010) used deep-sequencing technology to investigate *P. falciparum* epigenome (ChIP-seq) and transcriptome (RNA-seq) at multiple stages of intraerythrocytic development (iRBC cycle). They performed RNA-seq on 8 time-points and detailed genome-wide localization maps and a systematic analysis of three epigenetic characteristics followed by in-depth transcriptome sequencing during the intraerythrocytic growth of the *P. falciparum*. Their ChIP-seq study revealed that a particular nucleosome subtype containing the histone variant H2A.Z is found in euchromatic intergenic regions of the *P. falciparum* genome.

Sorber data set

Sorber et al. (2011) provided a transcriptome-wide characterization of intraerythrocytic splicing events captured on a single highly synchronous culture by RNA-Seq data from four timepoints. Their study reported 977 new 5' GU-AG 3' and 5 new 5' GC-AG 3' junctions and over 30 antisense and alternative splicing events including the complex transcriptional arrangements in the parasite.

Otto data set

Otto et al. (2010) applied RNA-Seq to seven time points from the asexual intraerythrocytic developmental cycle (IDC) of *P. falciparum* to capture characteristics associated with all expressed RNA transcripts and to quantify splicing dynamics that occur during parasite development. Their RNA-Seq findings corroborated well with the parallel DNA microarray analysis performed on the same samples, and provided better resolution over the process based microarray. These data show new features of the transcriptional landscape of *P. falciparum* and greatly enhanced the understanding of the red blood cell stage transcriptome of the parasite. They also identified different alternate splicing events and further on 5 'and 3' UTRs, which provided additional insight into human pathogen gene regulation.

Bunnik data set

Bunnik et al. (2013) compared next-generation sequencing data from steady-state mRNA and polysome-associated mRNA to gain a better understanding of mechanisms that control gene expression at the translational level during the asexual cell cycle of *P. falciparum*. They have harvested parasites immediately after the Red Blood Cell invasion at early ring stage (0 h), as well as in the trophozoite (18 h) and schizontal (36 h) stages. After the comparison of steady-state mRNA and polysomal mRNA expression clusters, they found substantial number of genes transcribed in the trophozoite and/or schizont stages of the cell cycle. They reported that more than 50% of genes expressed during the malaria parasite asexual cycle showed some form of translational regulation, ranging from a partial change in translation levels to a pause in translation of 18 hours or more compared to transcriptional activity.

López-Barragán data set

López-Barragán et al (2011) have constructed seven Illumina RNA-seq single-end libraries such as two gametocyte stages (GII and GV), ookinete (Oo), and four erythrocyte stage time points representing ring (R), early trophozoite (ET), late trophozoite (LT), and schizont (Sc). To systematically examine antisense transcripts, they also sequenced strand-specific cDNA libraries from ET, Sc, GII and GV. In gene models, they observed more than one thousand additional errors, alternatively spliced events including phase-specific alternatively spliced genes, and G and Oo phase antisense transcripts. Their data showed that in the sexual stages, antisense RNA plays a role in controlling gene expression.

Siegel data set

Siegel et al (2014) conducted a strand-specific transcriptomic study to discover widespread transcription of natural antisense transcripts (NATs) in *P. falciparum*. For parasites across the intraerythrocytic developmental cycle (IDC) as well as for separately purified nuclear and

cytosolic RNA fractions, they produced a total of 11 strand-specific transcriptome profiles. Their study revealed transcription from almost 80% of the *Plasmodium* genome and transcription of antisense for 24% of genes, many of which are controlled by development. They discovered that antisense transcripts are not evenly distributed in CDSs, but are heavily enriched at the 3' ends. They suggested that in *Plasmodium*, antisense levels do not correlate with sense transcript levels for the majority of genes due to RNA copying mechanism or nucleosome depleted regions (NDRs) function. Their findings indicated that non-coding RNA (ncRNA) in *P. falciparum* is biologically significant rather than just a result of noisy transcriptional control.

2.4 Functional consequences of non-synonymous SNPs

A single nucleotide polymorphism, or SNP, is a variation between individuals at a single location in a DNA sequence that is not only linked to genes, but may also occur in non-coding DNA regions. While a single SNP does not cause a disorder but certain SNPs are associated with certain diseases. Current epidemiology, medicine, and pharmaceutical genomics are making a great deal of effort, concentrating on the discovery of genetic variations, especially SNPs that are involved in common and complex diseases (Lee et al. 2008). Prioritization of SNPs based on their potential deleterious functional effects is important due to the large number of SNPs present in the genomes (Bhatti et al. 2006). Due to lack of experimental evidence in most SNPs, it becomes difficult to authenticate their deleterious effects. Many bioinformatics tools that predict the presumed deleterious effects of these SNPs have been developed and widely used (Rebbeck et al. 2004). Typically, these computational methods scrutinise if the SNP inhabits functional genomic regions such as exons, splice sites, or regulatory sites for transcription. Accordingly, the potential associated functional effects can be predicted by the SNP using a variety of methods of machine learning (Lee et al. 2008). However, to thoroughly analyse the functional implications of SNPs, it takes a considerable

amount of time and effort to independently apply various methods and carefully interpret the effects (Lee et al. 2008).

A single nucleotide polymorphism (SNP) in a genome is a significant source of variation. SNPs may result in affecting protein function by decreasing protein solubility or by destabilizing structure of a protein (Kucukkal et al. 2015). nsSNPs are single amino acid replacements in a protein sequence that can lead to a missense mutation or nonsense mutation. The change in protein sequence can subsequently affect the structure of protein and its interactions with the protein, and may have functional effects (Yates and Sternberg 2013).

SNPs have recently received considerable attention due to their potential as markers for genetic mapping and studying molecular evolution and dynamics of populations (Sachidanandam et al. 2001; Reich et al. 2001). A dense SNP map for chromosome 3 has been developed in *P. falciparum* with hundreds of SNPs (Mu et al. 2002), and multiple polymorphic sites on chromosome 2 were also identified through microarray hybridisation (Volkman et al. 2002). In *P. falciparum*, the genetic diversity exists in the form of single nucleotide polymorphism (SNPs), microsatellite repeats, insertions, deletions, and a variety of gene duplication. Several studies have been reported to investigate the SNPs of *P. falciparum* (Subudhi et al. 2015; Breglio et al. 2018). For designing strategies to combat the disease, the study of genetic variation in malaria parasites has functional significance.

2.5 Metabolic pathways analysis

Analysis of the metabolic pathways is the discovery and study of functional routes within metabolic networks (Schilling et al. 1999; Schuster et al. 2000). There is no doubt that it is important to understand the core metabolic pathways, that developments in molecular biology have greatly increased our understanding in the last few decades and revealed a more complex image of metabolism. The reactants, products and intermediates of an enzymatic reaction are

classified as metabolites, formed by a series of enzyme-catalyzed chemical reactions. In the development of new effective drugs and vaccines, understanding the cellular processes and interactions between cellular components is instrumental.

Intuitively, a pathway should be a series of linked reactions; providing an exact concept of 'meaningful' is much more difficult, and would include both physiological and biotechnological aspects (Klamt and Stelling 2003). Metabolic pathway data can be considered to consist of three levels: metabolites form the lowest tier; reactions are based on metabolites, and pathways are based on reactions. Numerous pathway databases such as KEGG (Ogata et al. 1999), BioCyc (Caspi et al. 2016), BioCarta (http://www.biocarta.com), and Reactome (Joshi-Tope et al. 2005) currently describe metabolic pathway and gene signaling networks providing the potential for a more nuanced and useful study. A unique identifier is given for each of the reactions and compounds found in KEGG, either starting with ‘R’ for reaction or ‘C’ for compound, followed by five digits.

PlasmoDB (http:/www.plasmodb.org) is the official database of malaria parasite genome project and includes the completed *P. falciparum* strain 3D7 genome and its official annotation as given by the consortium’s members for genome sequencing. In recent years, a variety of databases has been developed to store and organise the ever-increasing volume of metabolism data. The aim of metabolic path finding techniques is to use data on enzymatic reactions and chemical compounds input and output to identify appropriate pathways. There are two main kinds of online databases containing this kind of information. The first is database of metabolic pathways, which organise the reactions into pathways and concentrate on the relationship between these pathways and the other is a database of enzymes based on individual enzymes and their properties.

In order to find antibacterial drug targets, a chokepoint study and the essentiality of a reaction have been combined (Kim et al. 2010). Metabolic potential is a crucial determinant for the

development and growth, infectivity and maintenance of a pathogen (Taylor et al. 2013; Umeda et al. 2011). The enzymes forming a pathogen's metabolic network are therefore possible targets for drug development. Certain enzymes form nodes in the metabolic network that are particularly vulnerable and are thus attractive drug-targeting candidates. These enzymes are important for the pathogen's survival as these proteins are an integral part of the reaction that creates or consumes specific substrates, which are specific to the pathogen and are involved in various pathways. In metabolic pathways, targeting chokepoint enzymes has led to new treatments for autoimmune disorders, cancers and infectious diseases (Tyagi et al. 2019).

Studying metabolic pathways, especially chokepoint reactions within specific pathways, is a systematic way of identifying new targets. Specifically, the parasite is supposed to be harmed by targeting enzymes that either uniquely generate or consume a substrate called 'chokepoint enzymes'. If carefully chosen, using insight into both the host and the pathogen's biology and metabolic requirements, such targets have the potential to selectively damage the parasites without inappropriate host side effects. It is difficult to manually monitor all cellular processes because of the number of interactions between these biological entities in an organism. The representation helps us to determine the protein encoded by a gene, modified types of specific proteins, and how subunits assemble to form complexes of proteins. Using the BioCyc webserver to classify essential protein, metabolic chokepoint analysis was performed using the criteria: exclude reactions found in humans, exclude reactions catalysed by more than one enzyme, and limit the reaction found in multiple pathways (Yeh et al. 2004; Caspi et al. 2016). Therefore, after metabolic chokepoint analysis, the list of targets obtained were further compared with the non-homologous protein sets, and only those that are present in both sets are retained. For each of these organisms, analysing the result obtained from BioCyc gains insight into a list of unique proteins located in multiple pathways, so targeting some of these proteins will lead to the non-functioning of multiple pathways.

In many pathogenic species, chokepoint studies have been used for drug target identification (Taylor 2013). The chokepoint analysis was performed in two separate experiments to establish novel drug targets for two parasites: the mitochondrial protist, *Entamoeba histolytica* (Singh et al. 2007), and the malaria-causing protozoan parasite, *P. falciparum* (Yeh et al. 2004). In order to find specific drug targets for *Pseudomonas aeruginosa* (Perumal et al. 2009) (a common bacterium that causes infections) and *Bacillus anthracis* (Rahman et al. 2006) (the bacterium that causes anthrax), two further studies have applied chokepoint analysis. Another research examining the drug targets of *P. falciparum* assessed the essentiality of a reaction in a pathway by removing a silico reaction and evaluating whether the metabolic network to achieve the same endpoint (Fatumo et al. 2009) could identify an alternative pathway.

Functional annotations of gene products make it possible to assemble metabolic pathways that demonstrate how proteins function in conjunction to generate cellular compounds or transmit information. The software environment for Pathway Tools has been used to create pathway/genome databases for various prokaryotic and eukaryotic organisms (http://biocyc.org; Karp et al. 2002). There are several databases of pathways that describe the metabolite and enzyme interconnection, such as KEGG, WIT, and MetaCyc, within an organism. In diagrammatic form, the Malaria Metabolic Pathways (http:/sites.huji.ac.il/malaria) shows existing knowledge of malaria metabolism. In a frame-based representation, the underlying formal ontology describes an array of various concepts such as genes, proteins, compounds, reactions and pathways (Karp 2000).

Yeh et al. (2004) defined 216 enzymatic activities as catalysing chokepoint reactions with a chokepoint analysis, assuming that each enzyme has only one active site, unless annotated as multifunctional. If an enzyme catalysed at least one chokepoint reaction, it would be suggested as a promising drug target. For example, d-aminolevulinate dehydratase (ALAD) has been recognized as a chokepoint in *P. falciparum* (Yeh et al., 2004). Indeed, in heme biosynthesis,

d-aminolevulinate dehydratase is involved and can act as a valid antimalarial target (Bonday et al., 2000). Inactivating chokepoints may be shown to contribute to the destruction of an organism. If it is possible to inhibit the enzyme that catalyses a reaction which creates or consumes a specific compound, the entire pathway would be blocked, leading to the accumulation of specific substrate or the organism starving for a specific product (Yeh et al. 2004). Palumbo et al. (2007) showed that lethality corresponds to a lack of alternate pathways in a network that has been interrupted by a blocked enzyme, further support the concept of chokepoints and essentiality. Therefore, in order to identify and develop alternative anti-malarial drug targets, the study of differential gene expression, functional and pathway enrichment analyses using RNA-seq data is important. In addition, there is practical importance in the study of genetic variation in malaria parasites. There is no doubt that understanding the core metabolic pathways is important, that advances in molecular biology have increased our understanding significantly.

MATERIALS AND METHODS

Chapter 3

Materials and Methods

This section covers datasets, databases, tools for computational analysis, a detailed overview of the approaches and analytical techniques used in the current thesis. Fig. 3.1 illustrates the overall approach followed in this study.

3.1 Datasets

Many studies were carried out using transcriptomic data of *P. falciparum*. Present study concentrated primarily on data sets from López-Barragán (Accession number SRP009370) (López-Barragán 2011), as it comprises all the seven stages (Ring, Early trophozoite, Late trophozoite, Schizont, Gametocyte stage II, Gametocyte stage V and Ookinete) of the *P. falciparum* 3D7 parasite. The RNA-seq data set of all stages was obtained from NCBI database (https://www.ncbi.nlm.nih.gov/sra/) for analysis. SRA Toolkit modules vdb-validate and fastq-dump were used respectively to validate the integrity of downloaded SRA data and to convert SRA data into fastq format.

vdb-validate

vdb-validate SRR364849

It validates the integrity of downloaded SRA and if reported as as "ok" and "consistent", then the data were correct. If validation fails, then the data should be re-acquired from the SRA.

fastq-dump (https://ncbi.github.io/sra-tools/fastq-dump.html)

It converts SRA data into fastq format. There are many data formatting and filtering options available in it. When converted into fastq format, it provides information about spots.

fastq-dump ERR006186

Read 6755509 spots for ERR006186.sra

Written 6755509 spots for ERR006186.sra

Seven bidirectional reads from the 3D7 parasite have been taken for further investigation. The nsSNPs information of the *P. falciparum* genes were collected from PlasmoDB database (https://plasmodb.org/plasmo/). The protein sequences were obtained from the UniProtKB database (https://www.uniprot.org/) in the FASTA format.

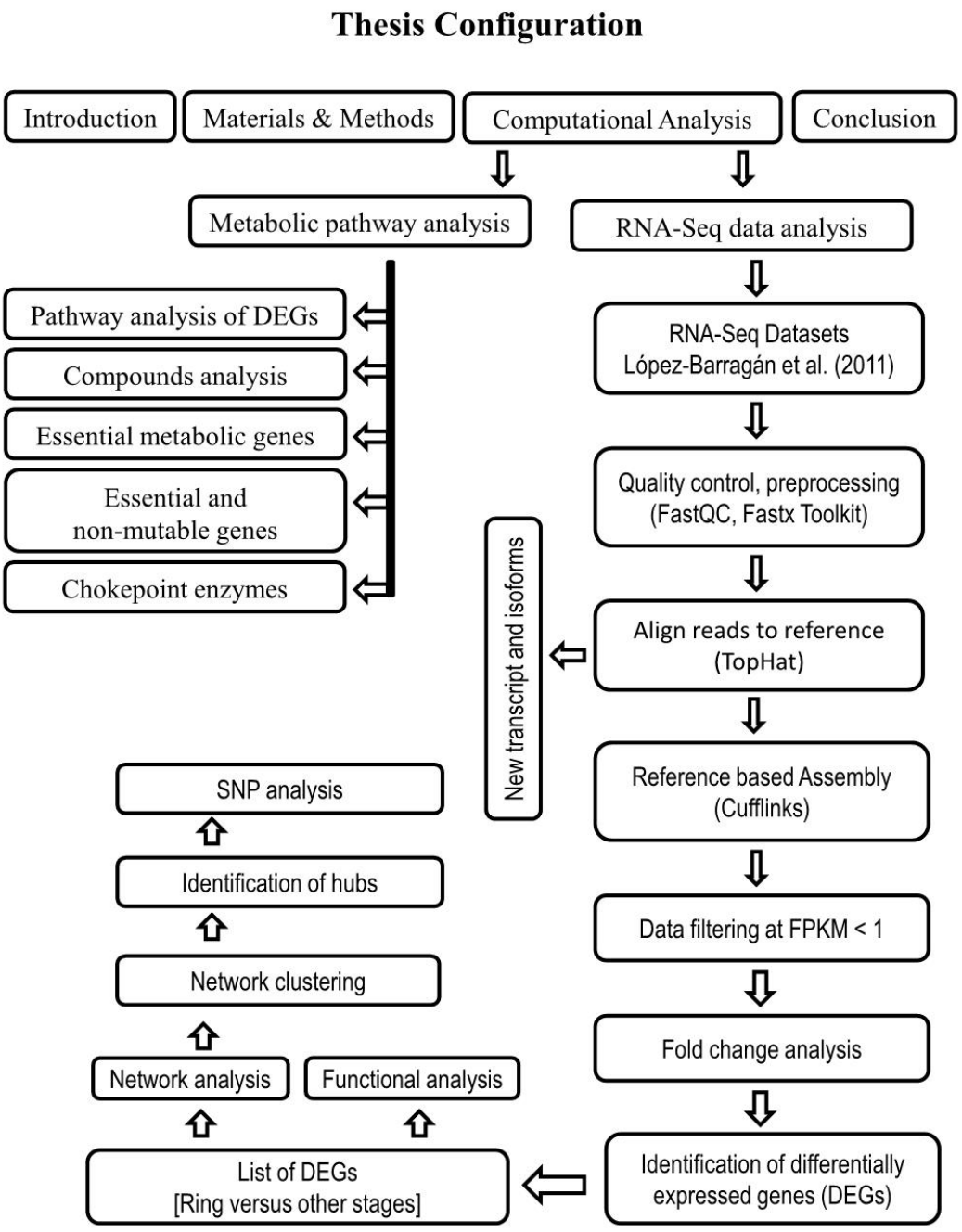


Fig. 3.1 The schematic representation of the overall workflow of the study

3.2 Quality control and pre-processing

If the sequencing data have a lot of noise, it needs to be pre-processed before the further downstream analysis. The low quality reads and adapter used for sequencing have been removed. The quality of sequence data obtained from high throughput sequencing pipelines were tested by using FastQC tool (http://www.bioinformatics.babraham.ac.uk/projects/ fastqc/). FastQC was implemented to determine the quality of the raw data, enabling evaluation of overall and per-base output for each sample read. The FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) was followed to clean the reads with low base-call quality using a quality filter tool.

As the initial stage of the routine RNA-seq workflow, quality control (QC) of raw data is important, the FastQC (Andrews 2010) tool was used to assess the quality of raw data, allowing the overall and per-base quality to be measured for each read in each sample. Generally, a QC value of 30 is considered safe for the downstream analysis. Care has been taken that at least 75% of the reads must have the desired QC values. The final call is based on the QC equal to or better than Q30C100. Any type of read trimming may be advisable prior to aligning the RNA-seq data, depending on the RNA-seq library construction strategy. Adapter trimming and quality trimming provide two popular trimming techniques. Trimming of adapter and low quality reads are followed by the QC check once again to have the desired QC values. Adapter trimming involves eliminating the sequence of the adapter by masking unique sequences used during the creation of libraries. Generally, quality trimming eliminates the ends of reads where base quality scores have decreased to a degree such that sequence errors prevent reads from aligning and the subsequent mismatches. As most recent sequencers provide raw data where the adapters are already trimmed, the adapter trimming step is usually not required. In contrast, depending on the research technique used, quality trimming may be a necessary step. The FASTX-Toolkit (Hannon 2010) was used for this purpose. After quality control, these reads

were mapped to the *P. falciparum 3D7* (PlasmoDB version 7.1) genome sequence. Mapping of reads to the reference was done by TopHat tool. The default values were used for mapping.

Following commands were used to clean reads with low base-call quality using a quality filter tool from the FASTX Toolkit with:

These commands are optimized based on the QC values of the raw reads.

fastq_quality_filter -q 20 -p 95 -Q 33 -i /home/sanjay/SRP009370/BD/SRR364849.fastq -o /home/sanjay/SRP009370/BD/SRR364849.trimmed.fastq

fastq_quality_filter -q 20 -p 95 -Q 33 -i /home/sanjay/SRP009370/BD/SRR364848.fastq -o /home/sanjay/SRP009370/BD/SRR364848.trimmed.fastq

fastq_quality_filter -q 20 -p 95 -Q 33 -i /home/sanjay/SRP009370/BD/SRR364847.fastq -o /home/sanjay/SRP009370/BD/SRR364847.trimmed.fastq

fastq_quality_filter -q 20 -p 95 -Q 33 -i /home/sanjay/SRP009370/BD/SRR364843.fastq -o /home/sanjay/SRP009370/BD/SRR364843.trimmed.fastq

fastq_quality_filter -q 20 -p 90 -Q 33 -i /home/sanjay/SRP009370/BD/SRR364840.fastq -o /home/sanjay/SRP009370/BD/SRR364840.trimmed.fastq

fastq_quality_filter -q 20 -p 75 -Q 33 -i /home/sanjay/SRP009370/BD/SRR364838.fastq -o /home/sanjay/SRP009370/BD/SRR364838.trimmed.fastq

Where,

[-q N] = Minimum quality score to keep.

[-p N] = Minimum percent of bases that must have [-q] quality.

[-i INFILE] = FASTA/Q input file. default is STDIN.

[-o OUTFILE] = FASTA/Q output file. default is STDOUT.

3.3 Sequence reads alignment

The reference genome of *P. falciparum* 3D7 (PlasmoDB version 7.1), which was available in the database, was used to map the sequence reads of all the seven stages such as Ring, Early

trophozoite, Late trophozoite, Schizont, Gametocyte stage II, Gametocyte stage V and Ookinete with the Bowtie tool. Following steps were followed:

The Bowtie index and bowtie2-build: It builds a Bowtie index from a set of DNA sequences

It outputs a set of 6 files with suffixes .1.bt2, .2.bt2, .3.bt2, .4.bt2, .rev.1.bt2, and .rev.2.bt2

Plasmodium falciparum Chromosomes: NC_000521.fna, NC_000910.fna, NC_004314.fna, NC_004315.fna, NC_004316.fna, NC_004317.fna, NC_004318.fna, NC_004325.fna, NC_004326.fna, NC_004327.fna, NC_004328.fna, NC_004329.fna, NC_004330.fna, NC_004331.fna

TopHat tool version: 2.0.14 (https://ccb.jhu.edu/software/tophat/) (Trapnell 2012) was used for further analysis. The minimum anchor length was seven base pairs for reads present at each side and a maximum size of intron 800 bp. Other mapping parameters were kept as default. The output of TopHat was filtered to keep only reads mapped from ring to gametocyte stages with 0 mismatches and up to 1 mismatch in Ookinete stage to maximize the accuracy. The resulted BAM file from TopHat, containing the mapped locations of the input reads were used for transcript assembly and quantification.

3.4 Transcript assembly

The read alignments were processed by Cufflinks v2.0.0 to identify novel gene, isoform and exon (Trapnell et al. 2010). The expression profiles of the transcripts were recorded in terms of FPKM (Fragments Per Kilobase Million) values. The command “cufflinks [options] <aligned_reads.(sam/bam)>” is used for the quantification of the transcripts. The SAM file format is a standard short read alignment format. The SAM format is helpful in the linking of custom tags to individual alignments.

The Cuffcompare is used to compare the Cufflinks assemblies to reference annotation files and help to sort out desired genes from known ones. In order to classify new transcripts, Cufflinks

also uses the reference annotation-based transcript (RABT) assembly method (Roberts et al. 2011) to assemble against an established reference annotation. Cufflinks GTF from the sample is matched against the reference GTF, and sample isoforms were tagged as overlapping, matching, or novel where appropriate.

3.5 Transcript quantification

The Cufflinks (http://cole-trapnell-lab.github.io/cufflinks/) tools were used to further evaluate the matched reads using a multifasta files (*Plasmodium_falciparum*.fa), which improves the reliability of abundance of transcript by detecting bias and an algorithm for correction (Trapnell 2012). The relative affluence of transcripts has been depicted as fragments/kilobase of exon/ million fragments mapped (FPKM), which is analogous to single-read Reads Per Kilobase of transcript, per Million mapped reads (RPKM) (Trapnell et al., 2010).

FPKM = fragments per kilobase per million

= [# of fragments]/[length of transcript in kilo base]/[million mapped reads]

= [# of fragments]/([length of transcript]/1000)/([total reads]/10^6)

3.6. Fold change analysis

Fold change is regularly used in the analysis of gene expression data from RNA-seq experiments to measure changes in the level of expression of a gene and to compare gene expression between two sets of arrays, e.g. case and control sets.

Fold change = value 2 FPKM / value 1 FPKM

Log_2 (Fold change) = LOG (Fold change, 2)

To compare the gene expression between Ring and other stages, fold change analysis of genes commonly expressed between two different samples was performed. All the genes were manually verified in PlasmoDB database on the basis of gene locus. Fold change was calculated as the ratio of ring (R) vs early trophozoite (ET), R vs late trophozoite (LT), R vs schizont (Sc), R vs gametocyte stage II (GII), R vs gametocyte stage V (GV) and R vs ookinete (Oo) groups.

Poorly expressed genes were removed from the dataset by eliminating genes with FPKM value < 2 in all the stages.

3.7 Annotation of differentially expressed genes

After calculating the fold change ratio of each gene, $\log_2$ (Fold change) was calculated for every gene by using below statistical formula:

Log_2(Fold change) = LOG(Fold change, 2)

In this study, the total fold change of ≥ 2 was considered to classify the differentially expressed genes. Based on this definition, up-regulated and down-regulated genes were identified in each group. Poorly expressed genes were removed from the dataset by eliminating genes with fold change value < 2 in all the stages. After that a list of non-redundant genes, which differentially expressed was created by combining all identified genes of ring and other stages and the duplicate genes were removed. Further, genes which are expressed differentially were manually checked in the PlasmoDB (Aurrecoechea et al. 2009) and UniProt database (UniProt Consortium 2018). Annotation of the differentially expressed genes were performed by several methods such as homology based, where we associated the function as of the known homologous sequences already annotated.

3.8 Functional analysis

DAVID web server was used to identify and select significantly enriched gene ontology terms and pathways (Huang 2009). The functional annotation was determined with DAVID program (http://david.abcc.ncifcrf.gov/home.jsp). It mainly provides p-value, Fold Enrichment, Bonferroni, Benjamini and FDR. Those terms with count number of ≥ 5 genes and a p-value of ≤ 0.05 were chosen for analysis. In DAVID, the terms Gene Ontology (GO) cellular component (CC), biological process (BP) and molecular function (MF) were used to classify improved biological topics in lists of genes which differentially expressed.

3.9 Pathway mapping

The KEGG Automatic Annotation Server (KAAS) has been used to map the pathway (Moriya 2007). Using the single-directional best hit (SBH) method for orthology assignment, the amino acid sequences of genes, which were up regulated and down regulated were submitted as input to the KAAS server. KAAS offers functional gene annotation in the database KEGG GENES through a similarity search tool of BLAST for a manually curated orthology group sets. For data sets mapped to one of reference pathways of KEGG, KAAS assigned a KEGG Orthology (KO) number to the genes. The mapping criteria used were default as of BLAST. E-value, query coverage and identify percentage values were used for filtering of the results.

3.10 Investigation of protein-protein interactions

The Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) program was used to study the functional relationships between DEGs (Szklarczyk 2017). The Gene Ontology terms, molecular function, cellular component, biological process and pathways of KEGG were used to create the interaction network using reference organism *P. falciparum*. Prior to analysis, each gene was mapped to the relevant identifier in STRING so that all genes in the list could be validated. The list of multiple protein names was uploaded with advance setting parameters like full string network type, 0.4 median confidence score and 5% FDR stringency. The physical and functional interactions, co-expression, co-location, pathways, predicted interactions and protein domain similarity were included in the network relationship between genes. The network has been filtered by eliminating all interactions with weights below 0.1.

3.11 Identification of hub genes

Cytoscape 3.4.0, a data integration and network visualization bioinformatics package was used to identify hub genes by measuring distribution of node degrees via Network Analyzer plugin (Smoot 2011). The input data was manually checked before running Cytoscape clustering tool. Clusters were created using Cytoscape molecular complex detection (MCODE) plug-in. It

detects regions, which are highly interconnected in the network (Bader 2003). The statistical criteria for MCODE are as follows: K score =2, Cutoff degree =2, Cutoff node score = 0.2 as the default. In the current network, top five genes with the largest distribution were considered as hubs.

3.12 Identification of non-synonymous SNPs (nsSNPs)

The polymorphism information for hub genes identified for *P. falciparum* in this study (PF3D7_0324900, PF3D7_1306000, PF3D7_1439500, PF3D7_0705600, PF3D7_1207100, PF3D7_0508100, PF3D7_1126700 and PF3D7_1234300) was retrieved from PlasmoDB database (https://plasmodb.org/plasmo/) by using default parameters. The protein sequences were retrieved from the NCBI GenBank and UniProtKB databases. Specific keyword searches were used in the mining of these sequences at the e-value more than zero. Only non-synonymous single nucleotide polymorphisms (nsSNPs) were used for further analysis. Mining of the non-synonymous SNPs were performed from the public databases.

3.13 Prediction of deleterious non-synonymous SNPs of hub genes

Hub genes are the genes with maximum connectivity in the network, the threshold value for the hub genes are 10% connectivity.

Following web based tools were used to predict the whether a substitution in amino acid affects the biological functions of a protein.

- **SIFT** (Sorting Intolerant from Tolerant) (http://sift.bii.a-star.edu.sg/www/SIFT_seq_submit2.html) (Kumar 2009)
- **PROVEAN** (Protein Variation Effect Analyzer) (http://provean.jcvi.org/) (Choi 2015)
- **PredictSNP** (Consensus classifiers for prediction of disease-related mutations) (https://loschmidt.chemi.muni.cz/predictsnp/) (Bendl 2014)
- **SNAP2** (Screening for Non-acceptable Polymorphisms) (https://rostlab.org/services/snap2web/) (Hecht et al. 2015)

SIFT and PROVEAN are tools to determine how amino acid substitution affects protein function in case the score is lower than threshold value. The PROVEAN tool was used to perform BLAST hits clustering. For each supporting sequence, a delta alignment score was calculated. In order to calculate the final PROVEAN score, average scores were generated across clusters (Choi 2015). Deleterious is considered a score of -2.5 or higher, when anything below this cut-off rating, which has an effect neutral. SIFT tool based on the physical properties of amino acids and homology of sequences determines whether substitution of an amino acid can influence the function of protein (Kumar 2009). The sequence tool SIFT makes SIFT predictions for a particular sequence of proteins in FASTA format. The sequence of protein queries and interest substitutions of nsSNPs and hub genes with default parameters have been submitted to http://sift.bii.a-star.edu.sg/www/SIFT_seq_submit2.html. SIFT program predicts substitutions with values < 0.05 to be detrimental. PredictSNP has been explicitly designed to combine the projected outcomes of several tools to form a prediction of consensus (Bendl 2014). For the predictions, prediction tools use a list of variants in the protein sequence as input. By using various biophysical features, evolutionary knowledge and several features of structure, SNAP2 determines whether or not an SNP is likely to alter the function of proteins. It provides prediction results in the form of effect or neutral and a score ranging from -100 to 100.

3.14 Prediction of mutation impacts on protein stability

I-Mutant 2.0 (http://folding.biofold.org/i-mutant/i-mutant2.0.html) was used for the analysis of protein stability and alterations by taking into account the SNPs (Capriotti et al. 2005). Protein sequence, temperature (25^{o}C), pH (7.0), and detailed SNP data are the input parameters for this tool. It provides prediction in the form of either Reliability Index (RI) or Free Energy change value (DDG).

3.15 Prediction of conservation of amino acids

ConSurf (consurf.tau.ac.il/), a Bayesian algorithm was used to evaluate the evolutionary stability of amino acid positions in the protein (Ashkenazy et al. 2016). Amino acid sequence of protein was submitted to the server because no known structure was available. The 3D structure of the protein was predicted by using MODELLER and multiple sequence alignment (MSA) was generated by ConSurf. Conserved regions were predicted using conservation scores and a colour scheme and further divided into different nine-degree scales. The score of conservation is 1-4 for were considered as variable, 5-6 as intermediate and 7-9 as conserved.

3.16 Prediction of solvent accessibility and secondary structure

NetSurfP program (http://www.cbs.dtu.dk/services/NetSurfP) was used to predict protein accessible surface area (ASA), surface accessibility and secondary structure. The NetSurfP simultaneously predicts accuracy for each prediction by calculating the Z-score. This approach contains two types of neural networks; the first type networks are based on secondary structure predictions and sequence profiles, with two outputs with respect to buried or exposed and in combination with sequence profiles. The other networks use these outputs as inputs and are trained to assess the relative surface exposure of each amino acid residues (Petersen et al. 2009). The normal and predicted SNP sequences in fasta format were submitted for prediction to NetSurfP. For prediction of secondary structure and accessible surface area, protein encoding gene in normal and its predicted SNP substitutions have been uploaded individually to NetSurfP web server. Microsoft Excel® 2016 has converted the most predicted secondary structure probabilities from NetSurfP to a single letter code representing helical (H), β-strand (E) and coil (C).

3.17 Prediction and validation of 3D structure of thiamine phosphate synthase

The Protein Data Bank (https://www.rcsb.org/) do not have the 3D structure of thiamine phosphate synthase (PfThiE). Therefore, SWISS-MODEL (https://swissmodel.expasy.org/)

was used to build 3D structure of protein by submitting FASTA protein sequence (Waterhouse et al. 2018). SWISS-MODEL is a fully automated server that uses the crystal structure of similar protein as a template to predict 3D protein structures. Depending on global model quality estimation (GMQE) and qualitative model energy analysis (QMEAN) values, the most reliable 3D structure was selected. Further, Verify3D, ERRAT and PROCHECK tools available on SAVES v5.0 (https://servicesn.mbi.ucla.edu/SAVES/) were used to validate the predicted 3D model of protein. According to Verify3D criteria, at least 80% of the amino acids should have scored >= 0.2 in the 3D/1D profile. ERRAT generally produce values around 95% or higher for good high-resolution structures and around 91% for lower resolutions (2.5 to 3A). In PROCHECK validation, a good quality model would be expected to have over 90% in the most favoured regions.

3.18 Alignment of model and the template structure

The Dali (http://ekhidna2.biocenter.helsinki.fi/dali/) program was used for comparing 3D structure of proteins (Holm and Rosenström 2010). This provides four options for structure comparisons as PDB search, PDB25, Pairwise and All against all. Pairwise structure comparison was used to compare template structure and modelled structure. It also provides secondary amino acid structure of a protein by means of DSSP.

3.19 Screening of compounds

Several thiamine phosphate synthase (EC 2.5.1.3) inhibitors and their analogues were taken from BRENDA (BRaunschweig ENzyme DAtabase), PubChem (https://pubchem.ncbi.nlm.nih.gov/), ZINC (Sterling and Irwin 2015) (https://zinc.docking.org/) and DrugBank (https://www.drugbank.ca/) compound databases. The 3D structure of protein modelled by SWISS-MODEL was uploaded to MTiOpenScreen (Labbé et al. 2015) (http://bioserv.rpbs.univ-paris-diderot.fr/services/MTiOpenScreen/) for screening of drug-like

compounds. MTiOpenScreen conducts automatic virtual ligand screening, based on AutoDock Vina docking. This enables a curated library of small compounds to be screened in order to identify compounds that are likely to bind to a given protein receptor. This comprises five in-house prepared libraries, containing drug-like molecules. There are many compounds library filters available for screening customization.

3.20 Molecular Docking

AutoDock Vina (Trott and Olson 2010) in PyRx 0.8 (Dallakyan and Olson 2010) was used for molecular docking. ZINC database was used for retrieval of compounds in Structure Data File (SDF) format. Open Babel (http://openbabel.org) tool was used to convert various file formats. PyRx program was used initially to minimize compounds energy and convert all molecules to AutoDock Ligand (PDBQT) format. The compounds without any predefined binding sites were docked against the entire surface of protein. The outcomes of docking results were reported in the form of binding energy. LigPlot^{+} program was used for the analysis of post-docking results (Laskowski and Swindells 2011). Using PyMOL, the docked complexes, which showed lowest binding affinity values were further analyzed hydrogen and hydrophobic bond interaction analysis.

3.21 Molecular features analyses

The ADME (Absorption, Distribution, Metabolism and Excretion) and drug-likeness predictions of compounds were carried out using SwissADME (Daina et al. 2017) (http://www.swissadme.ch/). The SMILES of compounds were used in SwissADME web tool as input. Further, ADMET and the pharmacokinetic properties were evaluated using admetSAR (http://lmmd.ecust.edu.cn/admetsar2) (Yang et al. 2019) web server to ensure the druggability potential of compounds.

3.22 Computational analysis of metabolic pathways of *P. falciparum*

3.22.1 KEGG (Kyoto Encyclopedia of Genes and Genomes) database

KEGG, is a knowledge base that was developed and maintained by the Kanehisa lab (https://www.genome.jp/kegg/) in Kyoto (Kanehisa and Goto, 2000; Kanehisa et al. 2017). It was originally created in 1995 for biological interpretation by KEGG pathway mapping of fully sequenced genomes. KEGG is an integrated database that consists of 15 manually curated and a computationally created database in four categories (Systems, Genomic, Chemical and Health Information). PATHWAY, BRITE, and MODULE are the databases in the Systems Information category. The databases in the genomic information category are KEGG ORTHOLOGY (KO), GENOME, GENES and SSDB. COMPOUND, GLYCAN, REACTION, RCLASS and ENZYME, which are collectively called KEGG LIGAND are the databases in the category for chemical information. The health information category consists of DISEASE, DRUG, DGROUP and ENVIRON databases.

3.22.2 Malaria Parasite Metabolic Pathways (MPMP)

Malaria Parasite Metabolic Pathways (MPMP) (http://mpmp.huji.ac.il/) is a database for the functional genomics of intraerythrocytic *P. falciparum*. The MPMP database consists of maps and tables, which organize the parasite genome's annotated genes in a functional sense. The necessary information was collected from KEGG database while constructing MPMP. This includes maps in various categories, including 171 for Biochemistry, 52 for Cell-Cell Interactions, 65 for Drug, 184 for Genetic Information Analysis, 109 for Morphology and Pathology, and 162 for Physiology. It also includes different search parameters including the search for map analysis.

3.22.3 BioCyc database

The BioCyc database contains organism specific 16,822 Pathway/Genome Databases (PGDBs) and various software tools for exploring them. The BioCyc database family is managed by SRI International, based in Menlo Park, California. For thousands of species they provide reference to genome and metabolic path information (Caspi et al. 2016).

Chokepoint reaction analysis was performed for *P. falciparum* using BioCyc webserver Chokepoint reaction finder using exclude reactions found in human, exclude reaction catalyzed by more than one enzyme and limit to the reaction found in multiple pathways criteria. A chokepoint reaction is a reaction that is either the unique consumer of a given metabolite, or producer of a metabolite. Theoretically, chokepoint reactions make good drug targets because disabling the reaction disrupts the metabolite's only pathway.

3.23 Database and Tools development

3.23.1 Database of nsSNPs for hub genes identified from *P. falciparum* RNA-seq data

Database of nsSNPs for hub genes identified from *P. falciparum* RNA-seq data was developed to show the analysis done by PROVEAN, SIFT, PredictSNP and NetSurfP software. The SNP data predicted to be damaging or deleterious by using any two out of three tools (PROVEAN, SIFT and PredictSNP). This can in turn affect function of protein and prediction of secondary structures and solvent accessibility of the hub genes using NetSurfP. Database was developed using MySQL Server as back-end and the front-end is built using PHP, HTML, JavaScript with user-friendly search environment using a range of options, such as simple searches and advanced searches. Since they are open-source tools and platform independent, Apache, MySQL and PHP technologies were chosen. Besides these benefits, multithreading and multi-user environments are provided by MySQL.

3.23.2 PfIDmap: a database tool for mapping of different identifiers of *P. falciparum*

Identifier Mapping is a growing challenge in the bioinformatics workflows. It requires integration of experimental data from different sources. Different identifiers related to *P. falciparum* were retrieved from different databases. PfIDmap has been designed as a client - server architecture running on an Apache server with PHP, HTML and JavaScript as the front end and MySQL as the back end using the relational data model. Description of PfIDmap architecture is provided in Fig. 3.2.

These technologies were preferred as they are open-source software and platform independent. This database contains direct links to many protein databases like PlasmoDB, UniProtKB, NCBI Entrez, NCBI GeneBank and RefSeq Protein. Protein sequences can also be retrieved in FASTA format. This identifier mapping tool can encourage the mapping of different identifiers from different databases relevant to all *Plasmodium* species in the future.

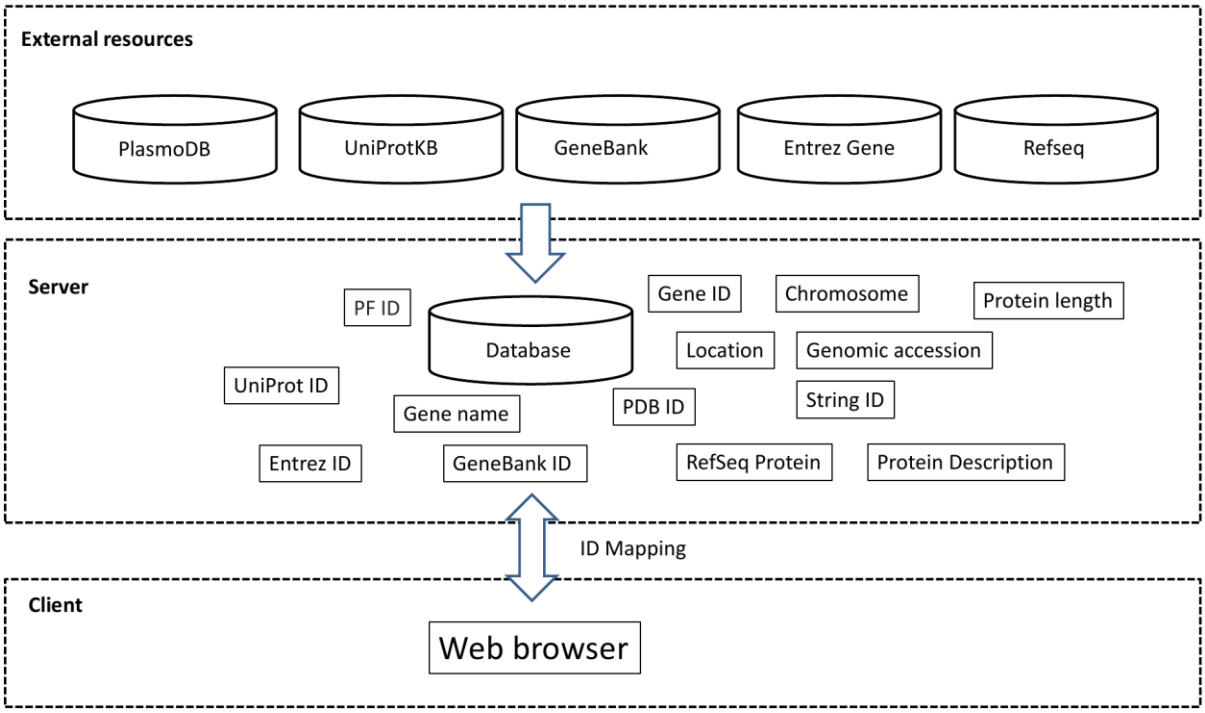


Fig. 3.2 Architecture of PfIDmap. Different identifiers like PlasmoDB Gene ID, Entrez ID, UniProt ID, PDB ID, GeneBank, RefSeq Protein and String ID can be mapped using ID Mapping from web browser

RESULTS AND DISCUSSION

Chapter 4

Results and Discussion

In this study, a comprehensive computational approach was used to derive potential therapeutic targets for *P. falciparum* using RNA-seq data set. The differential expression of genes, functional and pathway enrichment analyses of *P. falciparum* has been apprized in detail. The hub genes identified in the present study might serve as putative targets for drug designing. Functional analysis of the nsSNPs of identified hub genes was undertaken to predict deleterious mutations using various computational approaches. Moreover, the effect of deleterious mutations in glycosylphosphatidylinositol transamidase (GPI-T) subunit GPI8p has been investigated. Thiamine phosphate synthase (PfThiE), an essential metabolic gene in the thiamine biosynthesis pathway is also studied and potential inhibitors were identified through docking-based virtual screening along with drug-likeness and ADMET analysis to derive therapeutic targets for *P. falciparum.*

A. Gene expression and SNPs analysis to predict drug targets

Transcriptome assembly from mRNA high-throughput sequencing (RNA-seq) is an effective method for detecting differences in gene expression and sequences for both model and non-model organisms (Wang 2009; Ozsolak 2011). Many studies were carried out using transcriptomic data of *P. falciparum*. Some of the transcriptomic studies are listed in table 4.1. In the present study, we mainly concentrated on data sets of López-Barragán (Accession number SRP009370), as it comprises all seven stages of *P. falciparum* 3D7 parasite.

The RNA-Seq data set for all the stages of *P. falciparum* such as Ring SRR364849, Early trophozoite SRR364848, Late trophozoite SRR364847, Schizont SRR364843, Gametocyte stage II SRR364840, Gametocyte stage V SRR364838 and Ookinete SRR364834 were downloaded from NCBI SRA for analysis (Table 4.2).

Table 4.1 Available transcriptomic (RNA-seq) data sets of *P. falciparum*

Title	Study	Runs	Bytes	Bases
GSE52867: *Plasmodium falciparum* NF54 Transcriptome	SRP033414	06	14.61 Gb	21.66 G
Transcription profile of the *Plasmodium falciparum* intraerythrocytic cycle	SRP003507	08	5.68 Gb	8.37 G
Plasmodium falciparum RNA-Seq in Blood Stage	ERP000069	17	8.99 Gb	8.86 G
Plasmodium falciparum Transcriptome and Translatome	SRP021890	07	10.23 Gb	16.64 G
Plasmodium falciparum strain: 3D7 Transcriptome or Gene expression	SRP017623	08	5.79 Gb	9.14 G
A global strand-specific RNAseq analysis reveals high levels of natural antisense transcripts	ERP001849	12	38.08 Gb	52.70 G
***Plasmodium falciparum* Transcriptome or Gene expression**	**SRP009370**	**11**	**3.65 Gb**	**4.96 G**
RNA-Seq Analysis of Splicing in *Plasmodium falciparum* Uncovers New Splice Junctions, Alternative Splicing, and Splicing of Antisense Transcripts	SRP003615	05	11.85 Gb	9.99 G

SRA Toolkit modules vdb-validate and fastq-dump were used respectively to validate the integrity of downloaded SRA data and to convert SRA data into fastq format. All the seven SRA files downloaded from NCBI SRA database were firstly validated by vdb-validate module and then converted to fastq format.

Ring, early trophozoite and late trophozoite stages have read length of 35 bp while Schizont, GII and GV have 36 bp length and ookinete have reads with 51 bp.

4.1 Quality control and pre-processing

The initial stage of the standard RNA-seq workflow should be the QC of raw data. FastQC was implemented to determine the quality of the raw data, enabling evaluation of the overall and

Table 4.2 *Plasmodium falciparum* Transcriptome or Gene expression of seven stages

Stage name	Run	No. of read	Length (bp)
Ring	SRR364849	13456136	35
Early trophozoite	SRR364848	14505855	35
Late trophozoite	SRR364847	14599502	35
Schizont	SRR364843	9861818	36
GII	SRR364840	9294474	36
GV	SRR364838	10641921	36
Ookinete	SRR364834	55140294	51

per-base output for each sample read. Quality trimming typically eliminates the ends of reads where base quality scores have decreased to a point that prevents sequence errors and the subsequent mismatches from aligning reads. FASTX Toolkit, a quality filter tool was used to clean the reads with low base-call quality.

FastQC was used to check quality of downloaded RNA-seq data set for all the stages and accordingly, reads with low base-call quality were cleaned using a quality filter tool from the FASTX Toolkit. The final result of quality control is depicted in table 4.3 and these trimmed FASTAQ files were used for further studies.

4.2 Sequence reads alignment

Read mapping or alignment is the next step after obtaining a high-quality data from quality control and pre-processing. The reference genome of *P. falciparum* 3D7 (PlasmoDB version 7.1) was used to map the sequence reads with TopHat (Trapnell 2009). Without a reference annotation, TopHat can identify splice junctions. Potential exons are identified by mapping RNA-seq reads to the reference genome and then map the reads against these junctions to confirm them, using mapping information.

Table 4.3 FastQC (Quality control and preprocessing) result of seven stages

	FASTQ				TRIMMED FASTQ			
File name	**Total sequence**	**Total flagged as poor quality**	**Sequence length**	**%GC**	**Total sequence**	**Total flagged as poor quality**	**Sequence length**	**%GC**
SRR364849	13456136	0	35	46	9065759	0	35	44
SRR364848	14505855	0	35	42	10296994	0	35	41
SRR364847	14599502	0	35	39	10329150	0	35	38
SRR364843	9861818	0	36	50	8150329	0	36	50
SRR364840	9294474	0	36	36	3580874	0	36	35
SRR364838	10641921	0	36	24	4779419	0	36	23
SRR364834	714768	0	50-51	29	714768	0	50-51	29

Bowtie index was built from a set of DNA sequences of *P. falciparum* chromosomes. RNA-seq reads for every stage were mapped for *P. falciparum* genome by using TopHat program (table 4.4). The output of TopHat was filtered to keep only reads, which were mapped with 0 mismatches from ring to gametocyte stages, and up to 1 mismatch in Ookinete stage to maximize the accuracy.

In the directory where it was invoked, the TopHat script creates a number of files.

- **accepted_hits.bam** (A SAM format list of read alignments. SAM is a compact format for short read alignment that is increasingly being adopted)
- **junctions.bed** (A UCSC BED track reported by TopHat for junctions. Each junction consists of two BED blocks connected, where each block is as long as the maximum overhang of the junction spanning every reading. The number of alignments covering the junction is the ranking)

- **insertions.bed** and **deletions.bed** (Insertions and deletions for UCSC BED tracks reported by TopHat)

Table 4.4 Sequence reads alignment to reference using TopHat tool

Sample name	Run	Total Read	Mapped Read	%
Ring	SRR364849	13456136	5775553	42.92
Early trophozoite	SRR364848	14505855	7696256	53.05
Late trophozoite	SRR364847	14599502	7581273	51.92
Schizont	SRR364843	9861818	2073101	21.02
GII	SRR364840	9294474	3015618	32.44
GV	SRR364838	10641921	4090958	38.44
Ookinete	SRR364834	714768	711080	1.29

Ring, Early trophozoite, Late trophozoite, Schizont, GII, GV and Ookinete stage reads were mapped to reference genome with 42.92%, 53.05%, 51.92%, 21.02%, 32.44%, 38.44% and 1.29%, respectively.

4.3 Novel gene identification and differential expression analysis

Transcriptome analysis from RNA-seq data includes two sub-problems, i.e., transcript identification and the quantification of gene expression. If no annotation is available for a species of interest, or novel transcript discovery is required, isoform structures must first be created from RNA-seq data (Bernard et al. 2015). Single strain sequencing, either DNA or RNA, can help to identify genes and transcripts unique to the population. It is only possible to identify a gene as new on the basis of comparison with other organisms.

After the discovery of a gene, by looking for potential homologues, one needs to figure out whether this gene is new. Current biological information is defined by sequence databases and used to identify a gene as a novel gene (Klasberg et al. 2016). The overall methodology adopted in novel gene identification and differential expression analysis is depicted in Fig. 4.1.

Fig. 4.1 Workflow diagram for novel gene prediction and differential expression analysis

4.3.1 Transcript assembly

The Cufflinks suite assembles these reads into transcripts as well as quantifies these transcripts in single or multiple experiments. For Cufflinks, the BAM file from TopHat which contains the mapped locations of the input reads is taken as input. These are located in the file accepted_hits.bam in the TopHat output folder for the sample.

Cufflinks produces three output files:

1. transcripts.gtf: Transcriptome assembly file

2. isoforms.fpkm_tracking : Expression values for the transcripts expressed
3. genes.fpkm_tracking: Expression values for the genes expressed

Cuffcompare identified several transcripts with novel splicing and categorized all identified transcripts as complete match, potentially novel isoform, or unknown.

These transcripts fetched by their class codes e.g.

awk '$22 ~ /j/ { print }' cuffcmp.combined.gtf > Alternatively_spliced.gtf

and need to do some filtering e.g. length of transcripts more than 200

awk '{ if ($5-$4>200) print $0 }' Alternatively_spliced.gtf > Alternatively_spliced_200.gtf

The summary of Cuffcompare results are presented in table 4.5. Cuffcompare detected a total of 6931 transcripts for complete match of intro chain that code as equal sign (=) in output file while 4326 transcripts were potentially novel isoforms coded as ‘j’ and intergenic transcript as ‘u’ have 1839 transcripts. We have used all the detected 4326 potentially novel isoforms for further studies. Out of 4326 potentially novel isoforms, 1104 unique gene list was retrieved. These 1104 genes were searched in PlasmoDB database and these genes were available in this database.

In the present study Cuffcompare found some transcripts with novel splicing among all identified transcripts listed as intergenic, complete match, novel or unknown. A total of 4326 novel splice isoforms were detected by Cuffcompare. We employed rigorous bioinformatics research to minimize these novel splicing events to a more accurate conclusion and to minimize the number of false positives. To check novel isoforms, we examined all potentially novel isoforms using PlasmoDB and also carried out blastp of translated protein sequences of transcripts of novel splices to analyze changes in amino acids in the protein sequence (Sheynkman et al. 2013). In summary, predicting novel genes by computational methods is based on sequence properties, identification of known elements, and comparison with sister

Table 4.5 Summary of results of Cuffcompare obtained in this study

Cuffcompare class	Code*	Number transcripts	Percentage
Complete match of intro chain	=	6931	39.42
Multiple classifications	.	622	3.54
Contained in the reference	c	0	0
Possible pre-mRNA fragment	e	875	4.98
Transcript falling within a reference intron	i	67	0.38
Potentially novel isoforms	j	4326	24.6
Generic overlap with a reference transcript	o	64	0.36
Possible polymerase run-on fragment	p	2742	15.59
Intergenic transcript	u	1839	10.46
Exonic overlap on opposite strand	x	118	0.67
Repeat	r	0	0
Overlapping intron transfrag in the other strand	s	0	0

*Code related to Cuffcompare class

species. Since the discovery of the function of RNA as the primary intermediate between genome and proteome, transcript recognition and quantification of gene expression have been distinct core activities in molecular biology. New computational methods for predicting accurate gene structures with the peculiar properties of novel genes are on the verge of progress and will open up new perspectives (Klasberg et al. 2016).

4.3.2 Transcript quantification

Transcripts were assembled and analyzed in FPKM for their relative expression levels by Cufflinks tool after sequencing reads mapped to the TopHat reference genome. Between Ring and other stages, fold change analysis was done to compare the gene expression (Table 4.6). The data with FPKM values equivalent to zero were removed and remaining values were

subjected to further analysis. As a result, 7517 genes from ring (R), 6799 genes from early trophozoite (ET), 7482 genes from late trophozoite (LT), 5102 genes from schizont (Sc), 8731 genes from gametocyte stages (GII), 8831 genes from gametocyte stages (GV), and 5155 genes from ookinete (Oo) stages were identified. The non-expressed genes in stages R, ET, LT, Sc, GII, GV, Oo contains 209, 332, 291, 851, 381, 397, 758, respectively. Most highly expressed genes were found in GV stage while less genes were expressed in Sc stage.

Table 4.6 RNA-Seq gene expression distribution for the seven time points studied

	R	**ET**	**LT**	**Sc**	**GII**	**GV**	**Oo**
Highly expressed genes (≥ 500 FPKM)	441	392	463	273	593	619	479
Medium expressed genes (≥ 10 FPKM to 500 FPKM)	4693	3964	5783	2924	5511	6873	3935
Lowly expressed genes (< 10 FPKM)	2383	2443	1236	1905	2627	1339	741
Total expressed genes	7517	6799	7482	5102	8731	8831	5155
Non expressed genes	209	332	291	851	381	397	758

The obtained expression values of all the stages were merged by using Cuffmerge program. After merging the expression values of all samples, the Cuffdiff program was used to detect differentially expressed or regulated genes between samples. Only genes with differential expression were evaluated for further investigation. The statistical analysis of differentially expressed genes between Ring and other stages are presented in Appendix I.

All the expressed genes were chosen for further analysis. Then six groups RvET, RvLT, RvSc, RvGII, RvGV and RvOo were created and the common genes between RvET - 4402; RvLT - 4354; RvSc - 4022; RvGII - 3917; RvGV - 2988 and RvOo – 4009 have been identified (Fig. 4.2 and Table 4.7).

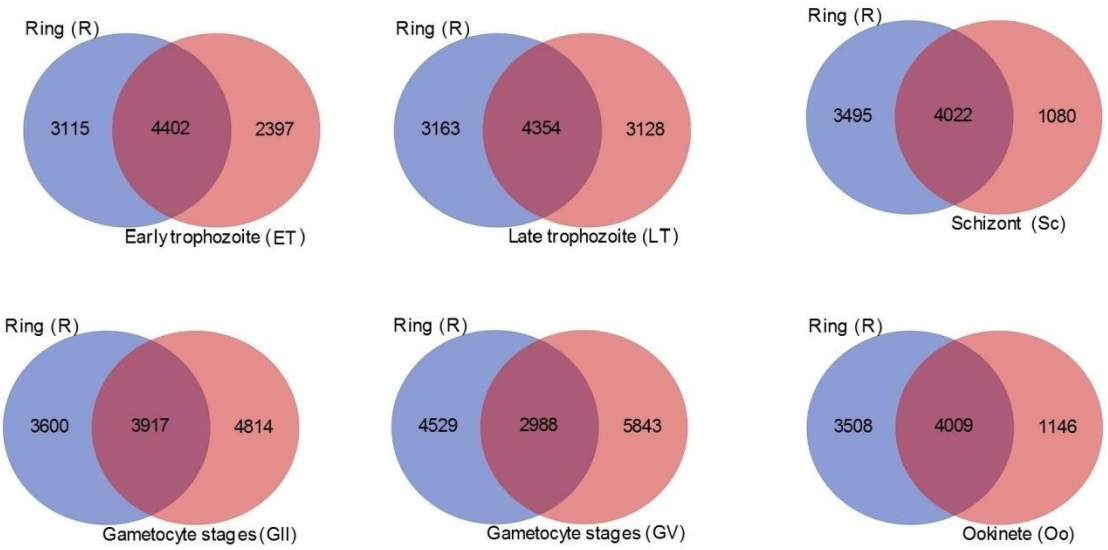


Fig. 4.2 Common genes between Ring (R) and other stages. Blue color showing genes only in the Ring (R) stage while pink color showing unique genes in other stages (ET, LT, Sc, GII, GV and Oo). Dark pink in the center showing common genes between RvET, RvLT, RvSc, RvGII, RvGV and RvOo stages.

Table 4.7 Genes which differentially expressed in different stages of *P. falciparum*

Sample	**Ring (R) and other stages**		
	Common genes	**Up-regulated (≥ 2 fold)**	**Down-regulated (≥ 2 fold)**
Early trophozoite (ET)	4402	731	1711
Late trophozoite (LT)	4354	1895	901
Schizont (Sc)	4022	1249	1686
Gametocyte stages (GII)	3917	897	1910
Gametocyte stages (GV)	2988	1116	1064
Ookinete (Oo)	4009	2070	825

A total of 4402 common genes were expressed between the R and ET stages, and 3115 and 2397 genes were different in R and ET stages, respectively. While, a total of 4354 common genes were expressed between R and LT stages, and 3163 and 3128 genes were expressed

specifically in R and LT stages, respectively. Between R and Sc stages, a total of 4022 common genes were expressed and 3495 and 1080 genes were expressed specifically in the R and Sc stages. Between R and GII stages, a total of 3917 common genes were expressed and 3600 and 4814 genes were expressed specifically within the R and GII stages, respectively. A total of 2988 common genes were expressed between the R and GV stages, and 4529 and 5843 genes were expressed specifically in the R and GV stages, respectively. Between the R and Oo stages, a total of 4009 common genes were expressed, and 3508 and 1146 genes were unique in the R and Oo stages, respectively.

4.3.3 Fold change analysis

Analysis of expression of one gene relative to other gene in two group was determined. The fold change in each group was estimated by using these common genes, which was defined as the FPKM value ratio of RvET, RvLT, RvSc, RvGII, RvGV and RvOo groups. For the fold change analysis, 4402 genes between RvET, 4354 between RvLT, 4022 between RvSc, 3917 between RvGII, 2988 between RvGV, and 4009 between RvOo groups were sorted. All these common genes between Ring (R) and other stages were used for differential expression analysis.

4.3.4 Annotation of differentially expressed genes (DEGs)

After calculating the fold change ratio of each gene, $\log_2$(Fold change) was calculated for every gene. In this study, the total fold change of ≥ 2 was considered to classify the differentially expressed genes. Poorly expressed genes were removed from the dataset by eliminating genes with fold change value < 2 in all the stages. Based on this, up-regulated and down-regulated genes were identified in each group. On the basis of above criteria, there are 2442 DEGs between RvET, 2796 between RvLT, 2935 between RvSc, 2807 between RvGII, 2180 between RvGV, and 2895 between RvOo groups were sorted out for the analysis (Table 4.7).

In order to see the up and down regulated genes between Ring and other stages, these DEGs were further analyzed and found that 731 and 1711 were up-regulated and down-regulated, respectively between the Ring and the early trophozoite group. Between the Ring and the late trophozoite group, 1895 and 901 were up-regulated and down-regulated, respectively while 1249 and 1686 were up-regulated and down-regulated, respectively between the Ring and the Schizont group. Between the Ring and the GII group, respectively, 897 and 1910 were up-regulated and down-regulated and 1116 and 1064 were up-regulated and down-regulated between the Ring and the GV group. Between the Ring and the Ookinete group, 2070 and 825 were up-regulated and down-regulated, respectively.

4.3.5 Functional analysis

A list of non-redundant genes which differentially expressed was created from the RNA-seq data by combining all identified genes of ring and other stages and the duplicate gene were removed. The functional annotation cluster analysis was conducted by DAVID tool on the differentially expressed genes (DEGs). The GO terms biological process, molecular functions and cellular components were used for interpretation and only those terms with count number of ≥ 5 genes and p-value of ≤ 0.05 were selected.

Five top GO terms with significant *p*-values for each group from functional analysis are presented in table 4.8**.** Only those gene IDs which were mapped via DAVID tool are used for further study. From this data, it is clear that the GO terms are enriched in RvET2209, RvLT2546 and RvGV1990 genes represent functions necessary for host cell plasma membrane, antigenic variation and pathogenesis (Table 4.8A, 4.8B & 4.8E). The enhanced GO terms in RvSc2735 including functions related to pathogenesis, single organismal cell-cell adhesion, receptor activity and cell adhesion molecule binding (Table 4.8C) and RvGII2594 include functions related to single organismal cell-cell adhesion, pathogenesis, receptor activity

Table 4.8 Five top enriched gene ontology (GO) terms detected by DAVID in genes expressed differentially between A) RvET2209, B) RvLT2546, C) RvSc2735, D) RvGII2594, E) RvGV1990 and F) RvOo2726

Gene Ontology (GO) term	Total genes	P-value	Fold Enrichment	Bonferroni	Benjamini	FDR
A) RvET2209						
GO:0009405 pathogenesis	60	1.50E-11	1.976	8.95E-09	4.48E-09	2.22E-08
GO:0004872 receptor activity	55	2.49E-10	1.955	1.45E-07	7.24E-08	3.67E-07
GO:0020033 antigenic variation	112	2.61E-10	1.578	1.55E-07	3.88E-08	3.85E-07
GO:0020002 host cell plasma membrane	122	7.27E-09	1.479	2.09E-06	6.95E-07	9.65E-06
GO:0020013 modulation by symbiont of host erythrocyte aggregation	55	1.09E-07	1.766	6.53E-05	1.31E-05	1.62E-04
B) RvLT2546						
GO:0016337 single organismal cell-cell adhesion	55	1.25E-08	1.724	8.38E-06	8.38E-06	1.87E-05
GO:0009405 pathogenesis	57	8.41E-06	1.531	0.005634146	0.00188159	0.012609
GO:0004872 receptor activity	53	1.69E-05	1.532	0.011208563	0.00562007	0.025317
GO:0020033 antigenic variation	111	2.22E-04	1.275	0.138760499	0.03665691	0.332848
GO:0020002 host cell plasma membrane	117	0.0027931	1.207	0.599326439	0.14138505	3.715881
C) RvSc2735						
GO:0016337 single organismal cell-cell adhesion	51	1.41E-04	1.441	0.100416159	0.10041616	0.21427
GO:0020035 cytoadherence to microvasculature, mediated by symbiont protein	44	4.92E-04	1.439	0.309017199	0.16874625	0.74645
GO:0050839 cell adhesion molecule binding	46	0.0018620	1.355	0.732751468	0.73275147	2.775761
GO:0009405 pathogenesis	53	0.00804	1.283	0.997682364	0.86766287	11.57191
GO:0004872 receptor activity	49	0.05020	1.199	1	0.99989018	54.0647
D) RvGII2594						
GO:0016337 single organismal cell-cell adhesion	50	7.85E-05	1.484	0.055250518	0.05525052	0.118849

GO:0020035 cytoadherence to microvasculature, mediated by symbiont protein	44	1.16E-04	1.512	0.080864209	0.0412843	0.176275
GO:0050839 cell adhesion molecule binding	45	6.05E-04	1.426	0.343766442	0.34376644	0.907874
GO:0004872 receptor activity	49	0.00977134	1.290	0.998923777	0.96719416	13.75421
GO:0009405 pathogenesis	50	0.015754517	1.272	0.99998984	0.94354259	21.38123
E) RvGV1990						
GO:0009405 pathogenesis	56	3.05E-09	1.880	1.87E-06	6.24E-07	4.52E-06
GO:0004872 receptor activity	53	3.57E-09	1.904	1.92E-06	1.92E-06	5.19E-06
GO:0020002 host cell plasma membrane	114	3.83E-09	1.540	1.19E-06	5.93E-07	5.14E-06
GO:0020033 antigenic variation	106	2.00E-08	1.522	1.23E-05	3.07E-06	2.97E-05
GO:0020013 modulation by symbiont of host erythrocyte aggregation	53	5.99E-07	1.734	3.68E-04	7.36E-05	8.88E-04
F) RvOo2726						
GO:0016337 single organismal cell-cell adhesion	47	9.00E-04	1.416	0.496203095	0.4962031	1.364404
GO:0020035 cytoadherence to microvasculature, mediated by symbiont protein	39	0.008244114	1.361	0.998163281	0.87753437	11.85945
GO:0020030 infected host cell surface knob	38	0.008860222	1.363	0.953182542	0.78362658	11.43261
GO:0050839 cell adhesion molecule binding	40	0.04538152	1.243	1	0.99999995	50.49853
GO:0009405 pathogenesis	47	0.056291162	1.214	1	0.99935639	58.66678

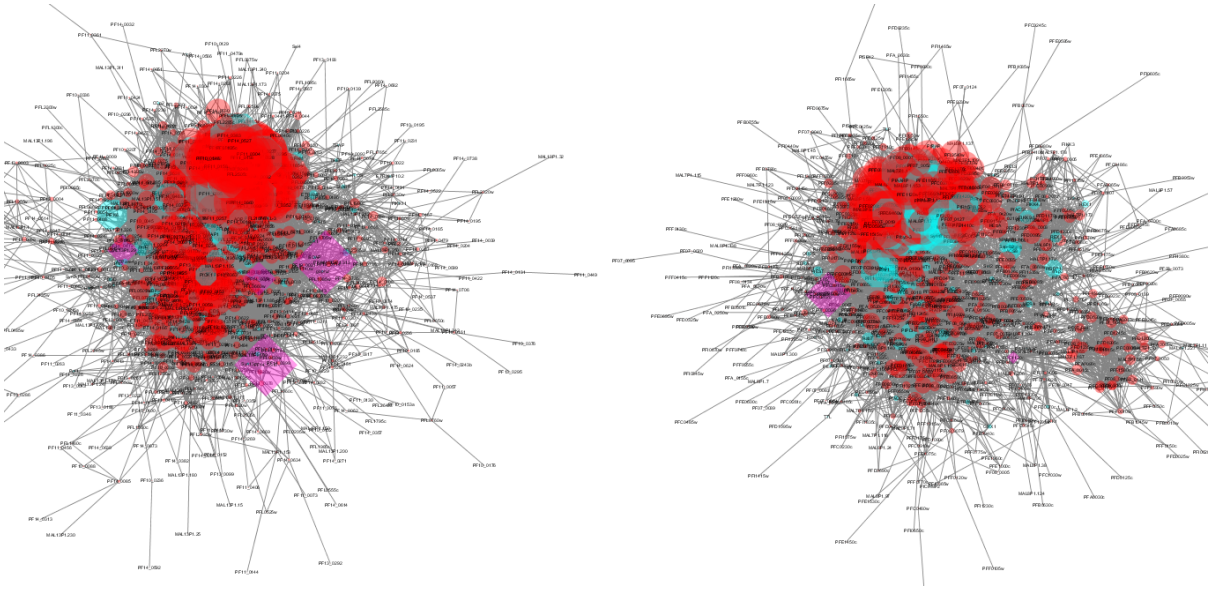
and cell adhesion molecule binding (Table 4.8D). Similarly, functions related single organismal cell-cell adhesion, pathogenesis and cell adhesion molecule binding are enriched in RvOo2726 samples (Table 4.8F).

4.3.6 Investigation of protein-protein interactions

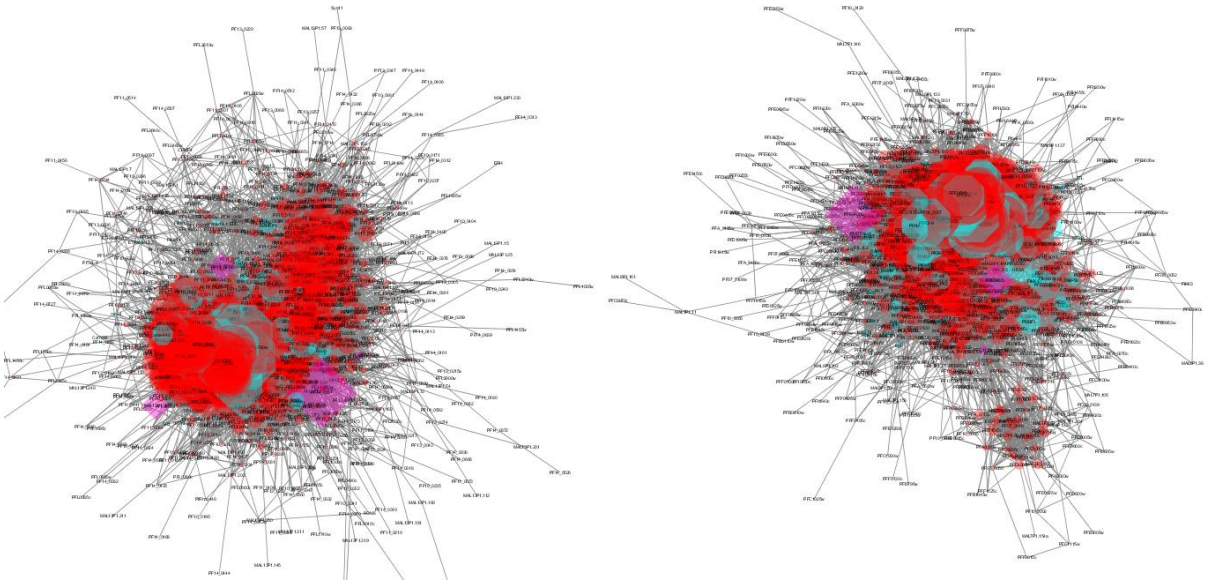
STRING program was used to investigate DEGs interactions. Only those genes showing significant interactions with weights higher than 0.4 were selected for network analysis. A network between DEGs was constructed for all the six groups viz., RvET2209, RvLT2546, RvSc2735, RvGII2594, RvGV1990 and RvOo2726. The six interaction networks resulted from STRING were then subjected to Cytoscape. The network consists of 1647 nodes and 23970 edges for RvET2209, 1969 nodes and 36152 edges for RvLT2209, 2236 nodes and 48514 edges for RvSc2735, 2010 nodes and 30113 edges for RvGII2594, 1473 nodes and 17986 edges for RvGV1990 and 2095 nodes and 37531 edges for RvOo2726. All genes in the network are represented in circles and their interactions represent edges. The interaction networks of all six groups are shown in Fig. 4.3.

4.3.7 Identification of hub genes

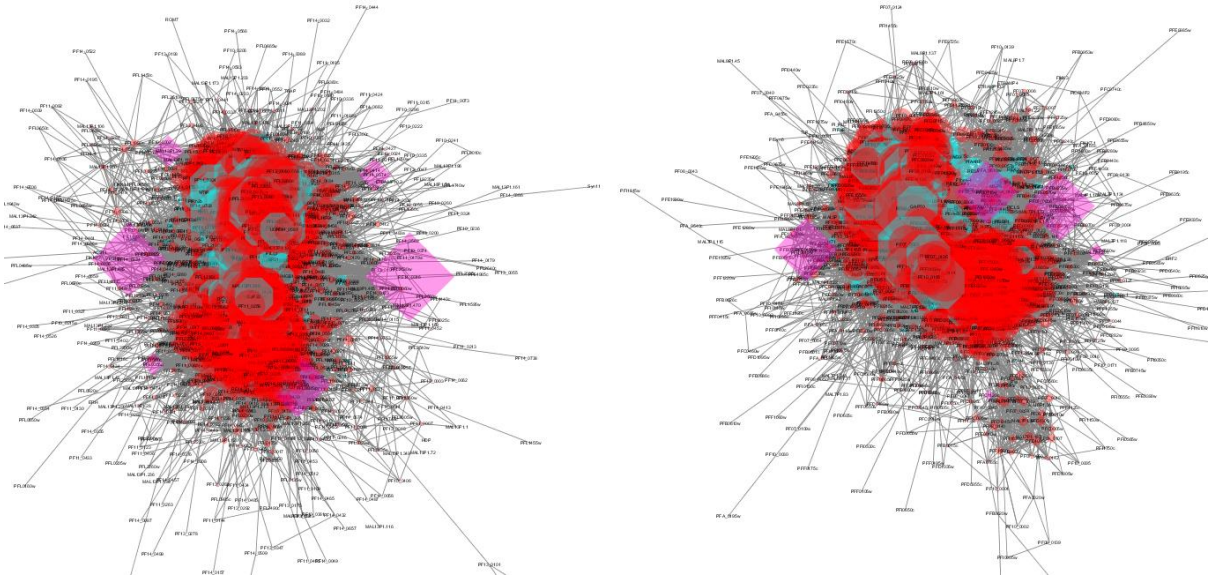
Hub genes in candidate modules are classified as genes with high correlation. Highly connected hub genes play important roles in the biological processes in a module. Proteins that are hubs in networks of interaction appear to be essential. Further, all six networks have been analysed by using Network Analyzer and MCODE modules available in Cytoscape. Network Analyzer calculate the node degrees in the network, whereas the MCODE module creates the clusters in the network. The higher node degrees were regarded to be more significant genes and were referred as hub genes. PF3D7_0324900, PF3D7_1306000, PF3D7_1439500, PF3D7_0705600, PF3D7_1207100, PF3D7_0508100, PF3D7_1126700 and PF3D7_1234300 genes were considered as hubs genes in the network constructed from 4196 *P. falciparum* genes by Network Analyzer and MCODE plugins of Cytoscape. The hub genes in the PPI network



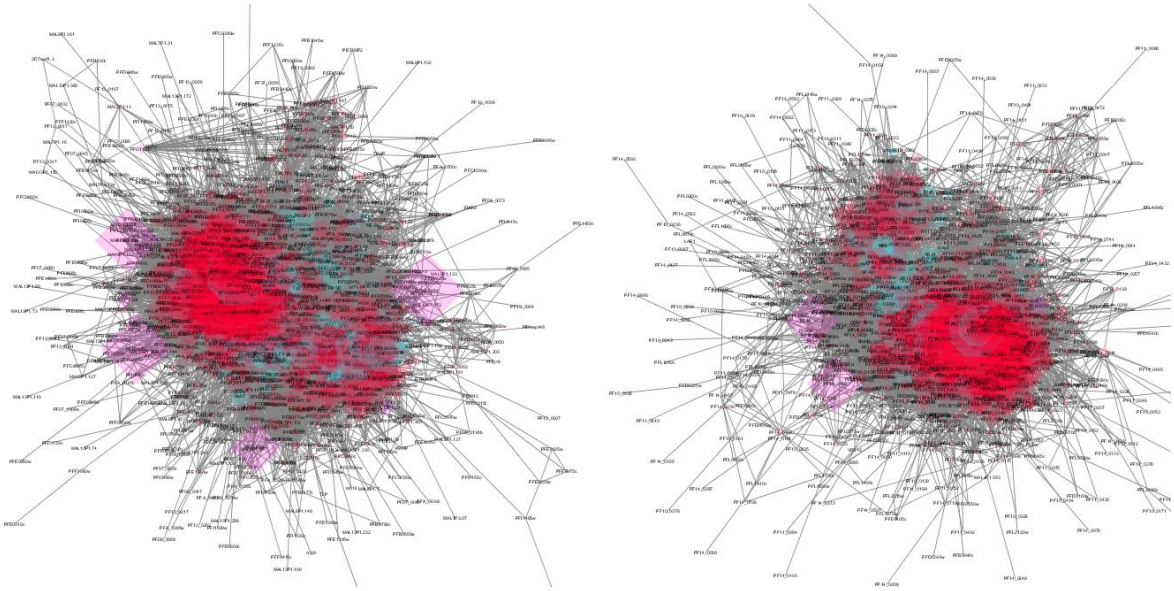
(a)



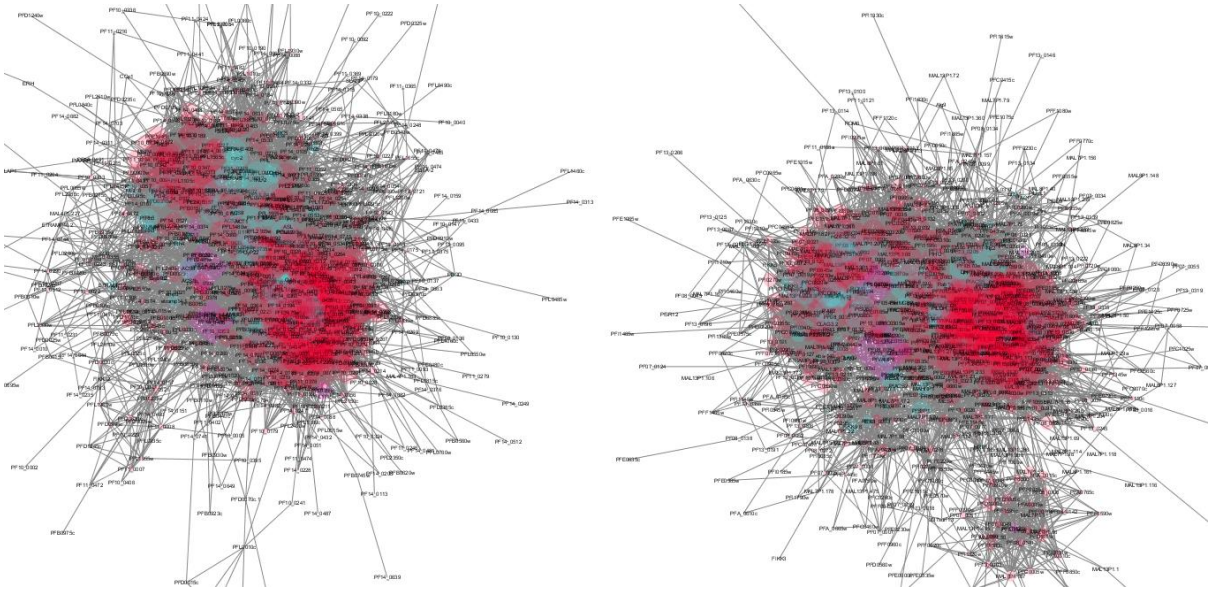
(b)



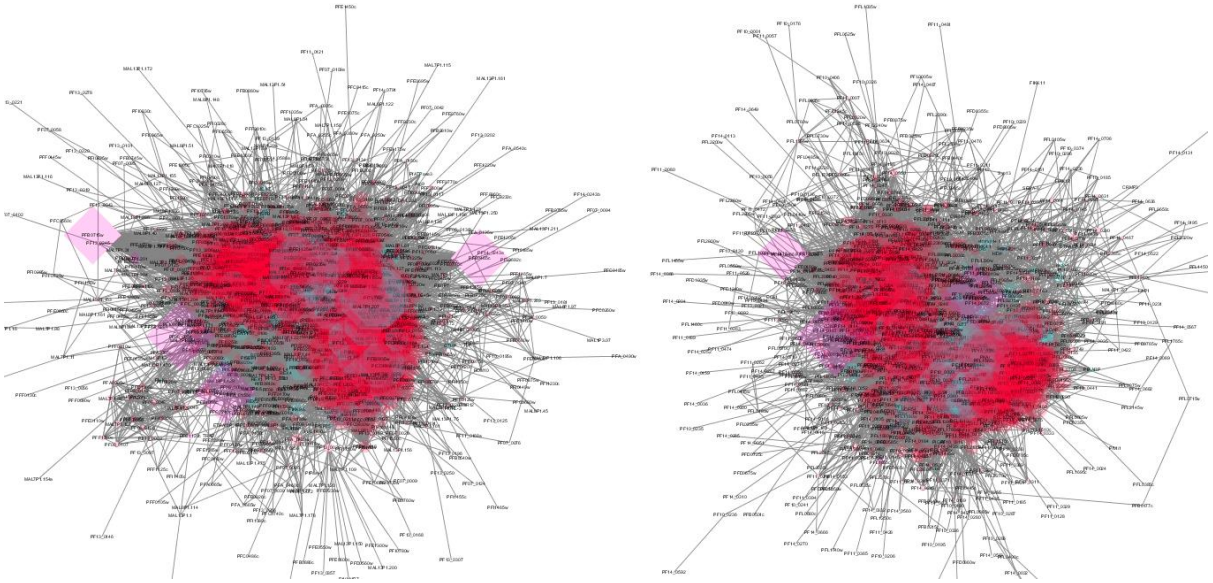
(c)



(d)



(e)



(f)

Fig. 4.3 Network of all six groups. (a) Interaction networks between R-ET DEGs and additional related genes; (b) Interaction network between R-LT DEGs and additional related genes; (c) Interaction network between R-Sc DEGs and additional related genes; (d) Interaction network between R-GII DEGs and additional related genes; (e) Interaction network between R-GV DEGs and additional related genes; (f) Interaction network between R-Oo DEGs and additional related genes. A red node indicates query genes and the genes predicted by STRING are shown in cyan. The top hubs are shown as purple diamonds.

has greater connectivity. Hub genes are considered functionally important because these are highly interconnected with nodes in a system. Therefore, these can act as putative targets for drug designing. The interaction between these hubs and their first neighbours were presented as Fig. 4.4.

PF3D7_0324900 (Erythrocyte membrane protein 1, PfEMP1)

The *P. falciparum* erythrocyte membrane protein 1, PfEMP1 (PF3D7_0324900) is one of the potential hubs found in this study, mediates the adhesion of infected erythrocytes (IE) to various host cells in the vascular lining during the blood stage of the infection with malaria. STRING results predicted the functional association partner of PFC1120c protein with PFD0005w, PFD0020c, PFD0055w, PFD0630c, PFD0635c, PFD0995c, PFD1000c, PFD1005c, PFD1015c, PFD1235w, PFD1245c, PFF1580c, PFF1590w, PFF1595c, PFI1830c, PFL0020w, PFL1955w and PFL1960w. Out of these proteins, PFD0055w and PFF1590w are rifin and rest of the proteins are erythrocyte membrane protein 1, PfEMP1.

Knob assembly may result in new ways of inhibiting PfEMP1 presentation on infected RBCs (Looker et al. 2019). Strategies for overcoming PfEMP1 antigenic diversity would provide an exciting new opportunity for the development of malaria vaccines (Chan et al. 2014).

PF3D7_0508100 (SET domain protein)

SET domain protein, (PF3D7_0508100) enables putative histone-lysine N-methyltransferase activity. Two types of Protein methyltransferase enzymes (PMTs) are present in eukaryotic cells; lysine specific and arginine specific. They were both associated with a variety of diseases including cancer, neurodegenerative and inflammatory diseases, PMT enzymes have emerged as a target class against human disease for drug discovery (Copeland et al. 2009). STRING results predicted the functional association of PFE0400w protein with ASP, Pf38, Pf41, PFD0190w, PFD0385w, PFD0390c, PFD0505c, PFD0530c, PFD0535w, PFD0665c,

PFD0745c, PFD0790c, PFD0800c, PFD0940w, PFD1050w, PFD1100c, PFE0085c, PFE0175c, PFE0265c, PFE0305w, PFE0355c, PfSUB1, RAP2, RAP3 and ROM4.

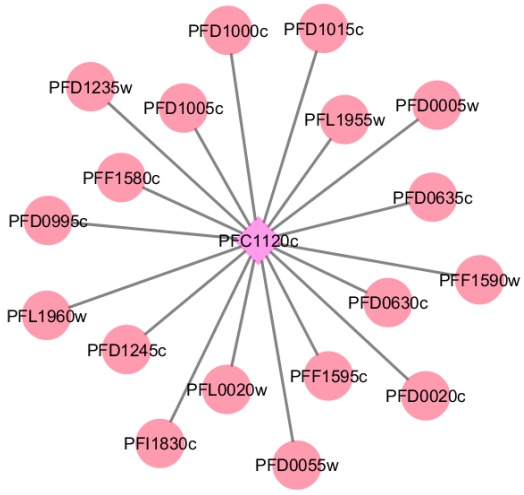


Fig. 4.4 (a) Interaction between PF3D7_0324900 and their first neighbours. A red node indicates query genes and the genes predicted by Search Tool for the Retrieval of Interacting Genes/Proteins are shown in cyan. The hub is shown as purple diamond.

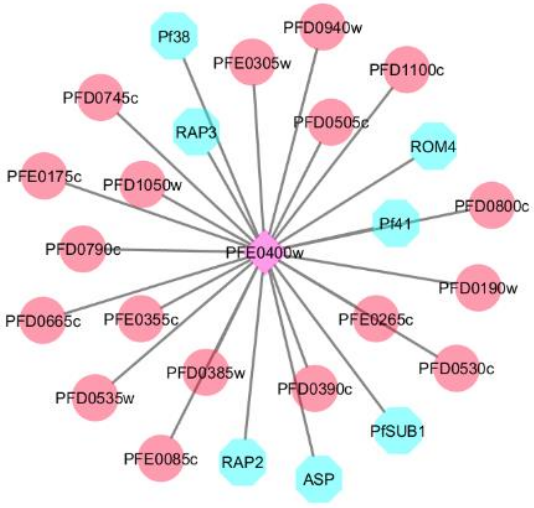


Fig. 4.4 (b) Interaction between PF3D7_0508100 and their first neighbours. A red node displays query genes and cyan is shown in the genes predicted by the Search Tool for the Retrieval of Interacting Genes / Proteins. The hub is seen as a diamond in purple.

PF3D7_0705600 (RNA helicase)

Hub gene RNA helicase, putative (PF3D7_0705600) play a variety of essential roles including cell development and cell growth. Helicases are significant unwinding enzymes that are needed in the malaria parasite for nearly all the nucleic acid metabolism. RNA helicases could be used

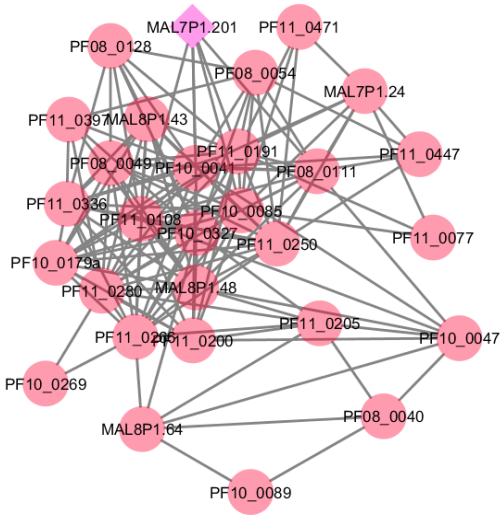


Fig. 4.4 (c) Interaction between PF3D7_0705600 and their first neighbours. In the genes predicted by the Search Tool for the Retrieval of Interacting Genes / Proteins, a red node shows query genes and cyan is shown. In purple, the hub is shown as a diamond.

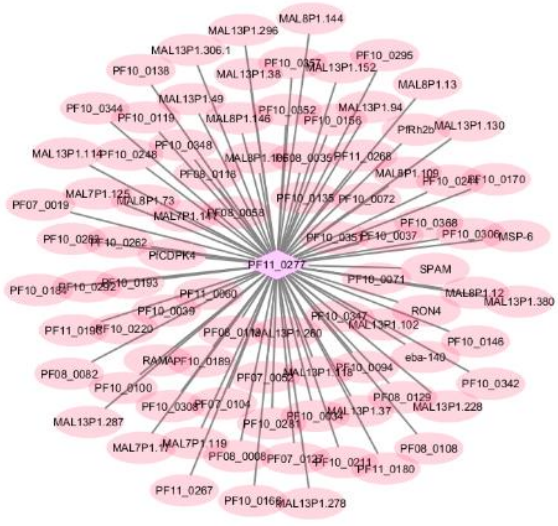


Fig. 4.4 (d) Interaction between PF3D7_1126700 and their first neighbours. A red node shows query genes and the genes predicted by Search Tool for the Retrieval of Interacting Genes/Proteins are shown in cyan. The hub is shown as purple diamond in cluster

as rational biochemical targets to develop new anti-parasite therapies and solve the drug resistance problem (Marchat et al. 2015).

PF3D7_1126700 (Autophagy-related protein 23)

The gene PF3D7_1126700 that was also detected as a hub codes for autophagy-related protein 23, putative. The ATG18 autophagy-related protein controls the biogenesis of apicoplast in apicomplex parasites and depletion of ATG18 in *P. falciparum* resulted in delayed death (Bansal et al. 2017).

PF3D7_1207100 (Small subunit rRNA processing factor)

Small subunit rRNA processing factor, putative (PF3D7_1207100) involved in the maturation of SSU-rRNA from tricistronic rRNA transcript was also found as a hub. It is a proteases group of enzymes that play key roles in the development and invasion of parasites. The ability to design particular protease inhibitors makes them promising objectives for drugs (Lilburn et al. 2011).

DNA polymerase epsilon subunit B, putative (PF3D7_1234300) is involved in the DNA-dependent DNA replication and enables DNA-directed DNA polymerase activity. It has been noted that Pol epsilon's function during replication is to extend the leading strand (Pursell et al. 2007).

PF3D7_1306000 (Conserved *Plasmodium* protein)

PF3D7_1306000 is a conserved *Plasmodium* protein with unknown function was found as a hub. This gene was found as essential along with druggability index 0.5 in TDR Targets database. STRING results predicted the functional association of MAL13P1.29 protein with m26-32-10, MAL13P1.118, MAL13P1.114, MAL13P1.126, MAL13P1.18, MAL13P1.152, MAL13P1.215, MAL13P1.278, MAL13P1.130, MAL13P1.202 and MAL13P1.103.

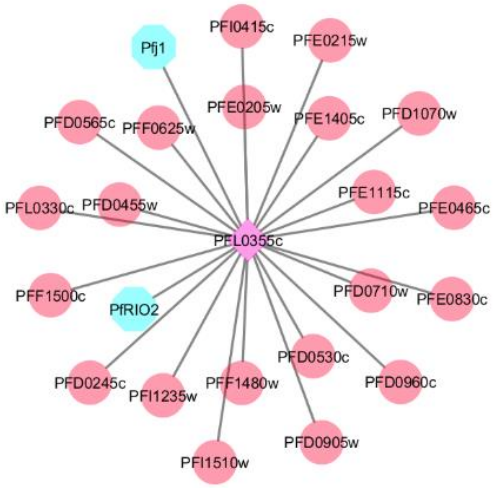


Fig. 4.4 (e) Interaction between PF3D7_1207100 and their first neighbours. The genes predicted by Search Tool for the Retrieval of Interacting Genes/Proteins are shown in cyan and a red node specifies query genes. The hub is shown as purple diamond.

PF3D7_1234300 (DNA polymerase epsilon subunit B)

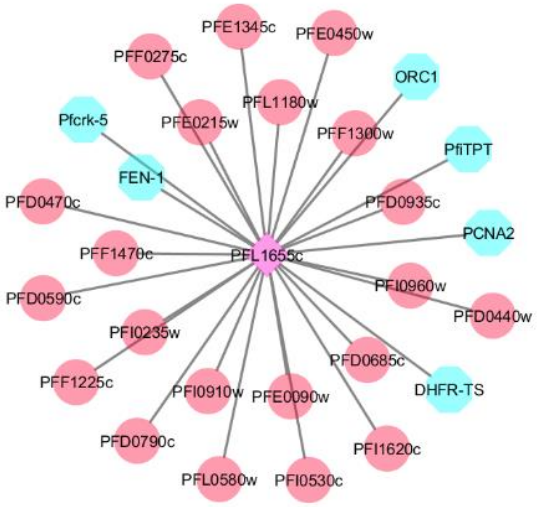


Fig. 4.4 (f) Interaction between PF3D7_1234300 and their first neighbours. A red node indicates query genes and the genes predicted by Search Tool for the Retrieval of Interacting Genes/Proteins are shown in cyan. The hub is shown as purple diamond.

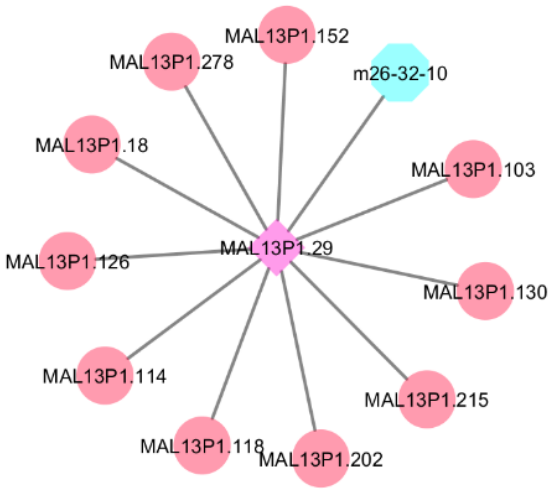


Fig. 4.4 (g) Interaction between PF3D7_1306000 and their first neighbours. A red node specifies query genes and the genes projected by Search Tool for the Retrieval of Interacting Genes/Proteins are shown in cyan. The hub is shown as purple diamond in each cluster

PF3D7_1439500 (oocyst rupture protein 2)

Another hub CCAAT-binding transcription factor or oocyst rupture protein 2 (ORP2) (PF3D7_1439500) was involved in sporozoite egress. Sporozoite egress from the oocyst can be blocked by deleting the N-terminal histone fold domain of ORP2 (Siden-Kiamos et al. 2018).

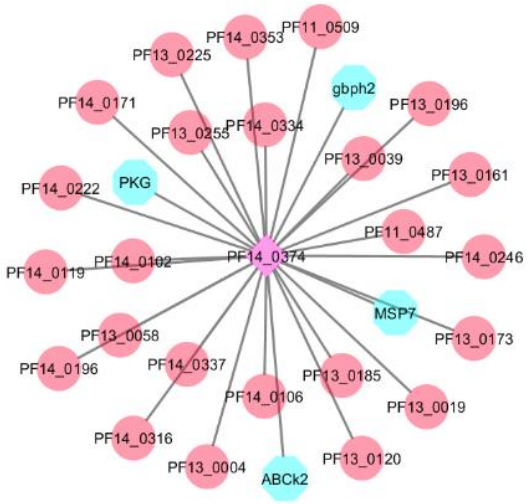


Fig. 4.4 (h) Interaction between PF3D7_1439500 and their first neighbours. A red node indicates query genes and the genes predicted by Search Tool for the Retrieval of Interacting Genes/Proteins are shown in cyan. The hub is shown as purple diamond in each cluster.

The differential expression of genes, functional and pathway enrichment analysis of *P. falciparum* were apprized in detail. A non-redundant list of 4196 genes was used for the functional annotation cluster analysis by the DAVID tool. The most important enriched GO terms identified in these genes by DAVID functional cluster analysis consist of functions required for the host cell plasma membrane, antigenic variation and pathogenesis. The node degree was calculated for each gene in the network to explore the functional roles of genes involved in different processes and interaction networks were created using Network Analyzer and MCODE plugins of Cytoscape.

PF3D7_0324900, PF3D7_1306000, PF3D7_1439500, PF3D7_0705600, PF3D7_1207100, PF3D7_0508100, PF3D7_1126700 and PF3D7_1234300 genes were considered as hubs genes in the network constructed from 4196 *P. falciparum* genes by Network Analyzer and MCODE plugins of Cytoscape. Hub genes are considered functionally important because these are highly interconnected with nodes in a system. Therefore, these can act as putative targets for drug designing.

4.4 Identification of deleterious SNPs

Functional analysis of the nsSNPs of identified hub genes was undertaken to predict deleterious mutations using various computational approaches. A single nucleotide polymorphism (SNP) is a significant source of variance in a genome. SNPs may result in affecting protein function by decreasing protein solubility or by destabilizing the structure of a protein (Kucukkal et al., 2015). The overall methodology adopted in the SNP analysis is depicted in Fig. 4.5.

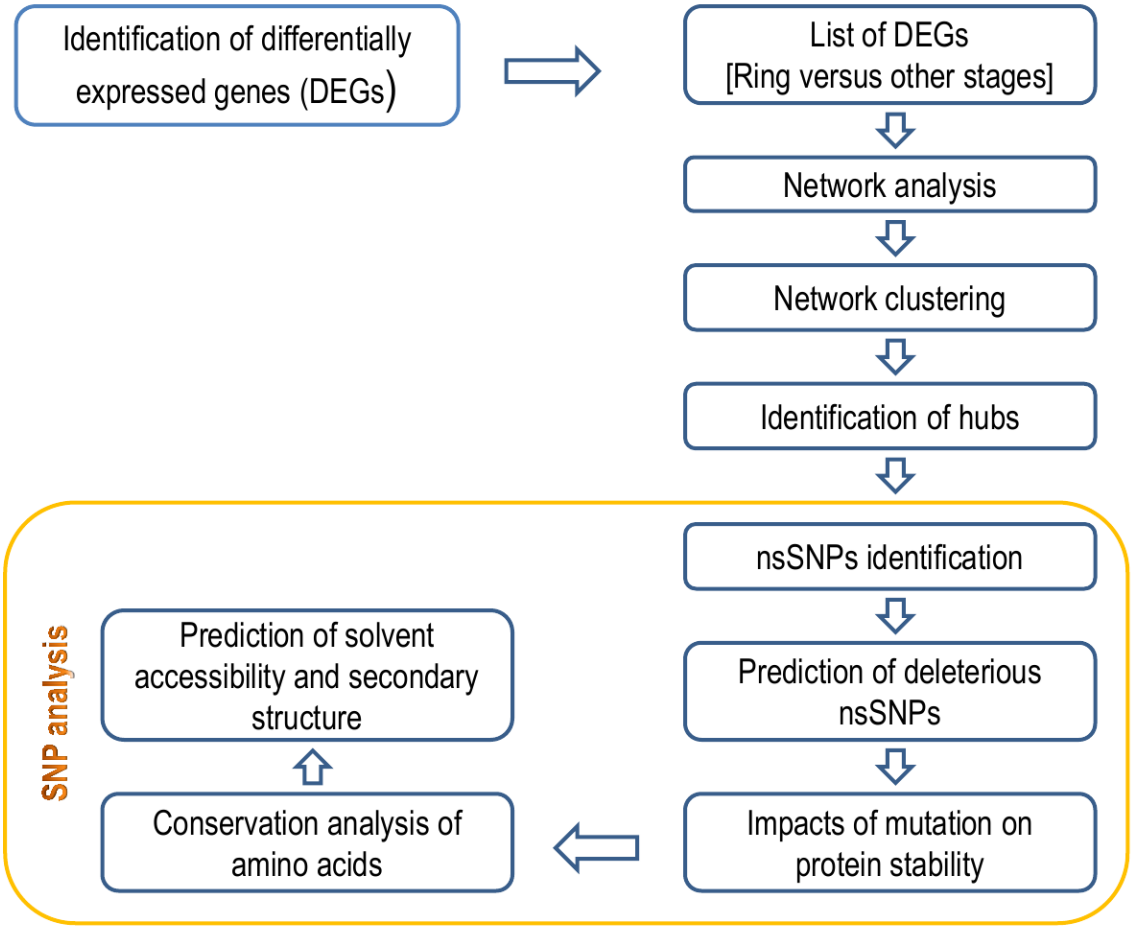


Fig. 4.5 Workflow diagram for SNP analysis

4.4.1 Identification of non-synonymous (nsSNPs)

The PlasmoDB database was used to retrieve the nsSNPs and SNPs for the identified hub genes. The PF3D7_0324900 have highest SNPs and nsSNPs information i.e., 2416 SNPs and 1606 nsSNPs from the database amongst all hub genes; whereas PF3D7_1439500, PF3D7_0508100, PF3D7_1306000, PF3D7_0705600, PF3D7_1207100, PF3D7_1126700 and PF3D7_1234300 have 171, 128, 101, 62, 38, 28 and 11 nsSNPs respectively. The total number of nsSNPs and SNPs recognized for the hub genes are illustrated in Fig. 4.6.

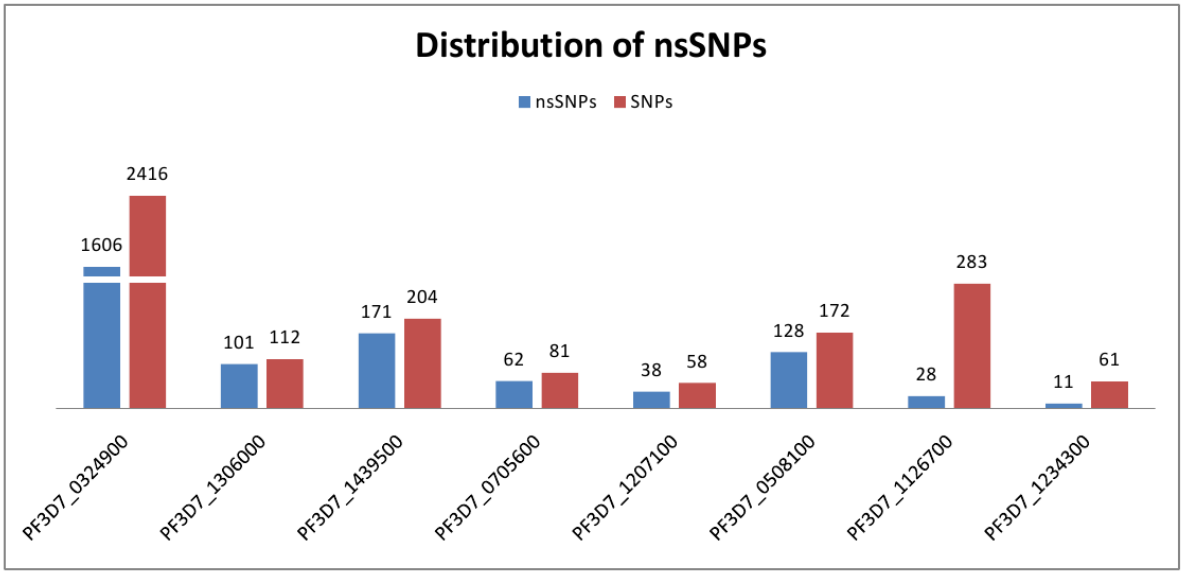


Fig. 4.6 Distribution of total number of single nucleotide polymorphisms and nonsynonymous single nucleotide polymorphisms identified for *P. falciparum* hub genes

4.4.2 Analysis of snSNPs for protein function

Deleterious coding nsSNPs were predicted by using SIFT, PROVEAN and PredictSNP tools. SIFT is a sequence homology-based tool used to classify substitutions for amino acids. The SIFT tool predicts whether substitution of an amino acid can affect protein function for a given FASTA protein sequence or not. The SIFT predicts substitutions with values < 0.05 to be deleterious. The SIFT sequence tool predicted 60, 52, 38, 33, 28, 19, 7 and 2 positions, which affected protein function for PF3D7_1439500, PF3D7_1306000, PF3D7_0508100, PF3D7_0705600, PF3D7_0324900, PF3D7_1207100, PF3D7_1126700 and PF3D7_1234300, respectively. To predict the final PROVEAN score, a delta alignment score is calculated for every supporting sequence and mean value across the clusters. By default, a score of -2.5 or above of it was taken as deleterious, while anything short of this cut-off rating has considered as neutral. In the hub genes, the PROVEAN protein tool predicted 40, 4 and 2 positions to be deleterious for PF3D7_0324900, PF3D7_1306000, PF3D7_1207100 and 1 position for PF3D7_0705600 and PF3D7_1234300, respectively. However, PROVEAN tool did not find any deleterious mutation in PF3D7_1439500, PF3D7_0508100 and PF3D7_112670.

PredictSNP unambiguously designed to combine outcomes of several methods, mostly to annotate disease-variant relationships. According to PredictSNP, 18, 16, 11, 9, 6, 4, 2 and 1 mutations to be deleterious for PF3D7_1306000, PF3D7_0324900, PF3D7_1439500, PF3D7_0705600, PF3D7_1207100, PF3D7_0508100, PF3D7_1126700 and PF3D7_1234300. It was confirmed and verified at least by two tools used in the study that 25 positions for PF3D7_0324900, 20 for PF3D7_1306000, 11 for PF3D7_1439500, 10 for PF3D7_0705600, 6 for PF3D7_1207100, 4 for PF3D7_0508100, 2 for PF3D7_1126700 and 1 for PF3D7_1234300 were identified to influence the function of protein (see Appendix II).

4.4.3 Solvent accessibility and secondary structure prediction

Further, the secondary structures and solvent accessibility of the hub genes was investigated using NetSurfP-1.1. The SNPs, which were predicted by at least two tools to have a negative effect on the function of protein, were used for the analysis of secondary structure and solvent accessibility. Moreover, the data were screened by selecting the residues, which showed change in ASA greater than or equal to 10 $Å^2$ from buried state to exposed state and also exposed state to buried state and its secondary structure change. The information related to ASA and secondary structure are shown individually in Appendix III.

In PF3D7_0324900, N615K mutation showed an ASA change to exposed state from buried but T1745P, S1747R, S2024R, P2026S and E2065K mutations showed the opposite change to buried state from exposed state. There was a change in most of the confirmations from Coil (C) to Alpha-Helix (H), due to these mutations (Table 4.9).

Due to mutations Y779D and Y862N, an ASA change to exposed state from buried was shown in PF3D7_1306000 whereas D1113Y showed an inverse change. Conformations in Y779D and D1113Y were changed from Alpha-Helix to Coil, while in Y862N, Beta-strand (E) was

Table 4.9 NetSurfP results of predicted hub genes: Residues that showed ASA change of ≥ 10 $Å^2$ with change in ASA from buried to exposed state and vice versa and also showed change in their secondary structure

	Mutation	**Class change**		**Conformation Change**					
		B - E	E - B	C - E	C - H	E - C	E - H	H - C	H - E
PF3D7_0324900	N615K	5	3	1	-	-	-	-	-
	T1745P	4	4	-	1	-	-	-	-
	S1747R	1	2	-	1	-	-	-	-
	S2024R	1	2	-	1	-	-	-	-
	P2026S	-	3	-	1	-	-	-	-
	E2065K	1	2	1	-	1	-	-	-
PF3D7_1306000	Y779D	1	-	-	-	-	-	1	-
	Y862N	1	-	-	-	1	-	-	-
	D1113Y	2	1	-	-	-	-	1	-
PF3D7_1439500	N186H	83	25	5	8	2	-	2	-
	S312N	61	54	4	4	3	-	2	-
	S313N	64	69	7	6	2	-	1	1
	N334S	85	25	2	6	2	-	3	-
	Y338N	69	54	7	1	2	-	2	-
	S348C	78	33	4	7	1	-	2	-
	Y403S	70	68	2	12	2	-	2	-
	T411I	59	66	7	7	3	-	2	-
	S445L	69	59	4	9	5	-	2	-
	R570G	67	58	5	8	5	-	2	-
	N746K	69	73	9	8	2	-	3	-
PF3D7_0705600	N124Y	3	2	1	1	-	-	-	-
	E623V	24	24	2	2	1	-	8	4
	A626D	20	26	2	2	1	-	7	4
	D645Y	22	27	2	5	1	-	6	5
	T688R	20	22	2	6	1	-	8	5
	T852S	14	13	1	1	2	-	1	1
PF3D7_1207100	N235K	6	15	1	3	-	-	1	-
	R277W	3	9	-	2	-	-	-	-
	D377H	16	8	-	-	-	1	2	-
	P528S	4	3	-	2	-	-	-	-
	N661Y	5	8	-	1	-	-	1	-
PF3D7_0508100	I800T	30	45	-	4	-	-	7	2
	I1103K	43	51	3	5	2	1	10	2
	S1411P	47	58	2	8	-	2	15	3
	Y1474H	55	64	2	10	-	-	13	3
PF3D7_1234300	S434Y	3	4	-	1	-	-	-	-

changed to Coil (C). Mutations in the residue position N186H, S312N, N334S, Y338N, S348C, Y403S, S445L and R570G in PF3D7_1439500 showed a change to exposed state from buried and mutations S313N, T411I and N746 K showed a change from buried in the majority of the

cases. In most cases, it showed changes in secondary structure from coil to helix and coil to beta-strand. In PF3D7_0705600, mutations showed almost the same change to exposed state from buried state and vice versa. It shows changes in secondary structure from helix to coil and helix to beta-strand in most cases and also from coil to helix in some cases. Due to mutation D377H, an ASA change primarily from buried state to exposed state was shown in PF3D7_1207100 whereas N235K, R277W, P528S and N661Y show an inverse change. In R277W mutation, both F274 and E275 showed change in C to H conformation. D377H showed change in conformation mainly from H to C. In P528S mutation, both L337 and Y340 showed change from C to H conformation. In N661Y mutation, L337 showed change from C to H and I410 showed an opposite change in conformation. Mutations in the residues I800T, I1103K, S1411P and Y1474H, changes in PF3D7_0508100 showed mostly from exposed to buried state. In most of the cases, it shows changes in secondary structure from helix to coil & vice versa. Some also change from helix to beta-strand conformations. In PF3D7_1234300, S434Y mutation, I137 showed change in conformation from coil to helix. In PF3D7_1126700, no significant change was detected (Table 4.9).

The number of mutations observed for each gene are:

- 6 (N615K, T1745P, S1747R, S2024R, P2026S and E2065K) for PF3D7_0324900
- 3 (Y779D, Y862N and D1113Y) for PF3D7_1306000
- 11 (N186H, S312N, S313N, N334S, Y338N, S348C, Y403S, T411I, S445L, R570G and N746K) for PF3D7_1439500
- 6 (N124Y, E623V, A626D, D645Y, T688R and T852S) for PF3D7_0705600, 5 (N235K, R277W, D377H, P528S and N661Y) for PF3D7_1207100
- 4 (I800T, I1103K, S1411P and Y1474H) for PF3D7_0508100
- Only one for PF3D7_1234300

These mutations showed a change from buried to exposed state and vice versa with difference in ASA change of ≥ 10 Å2 and show change in their secondary structure in these proteins by NetSurfP tool.

4.4.4 Database of nsSNPs for identified hub genes

Additionally, the SNP data that are predicted to be deleterious or damaging by using any two tools out of three (PROVEAN, SIFT and PredictSNP) and which may in turn affect protein function for hub genes is available online in the form of a simple database available at www.cdkd.org/pfsnp/.

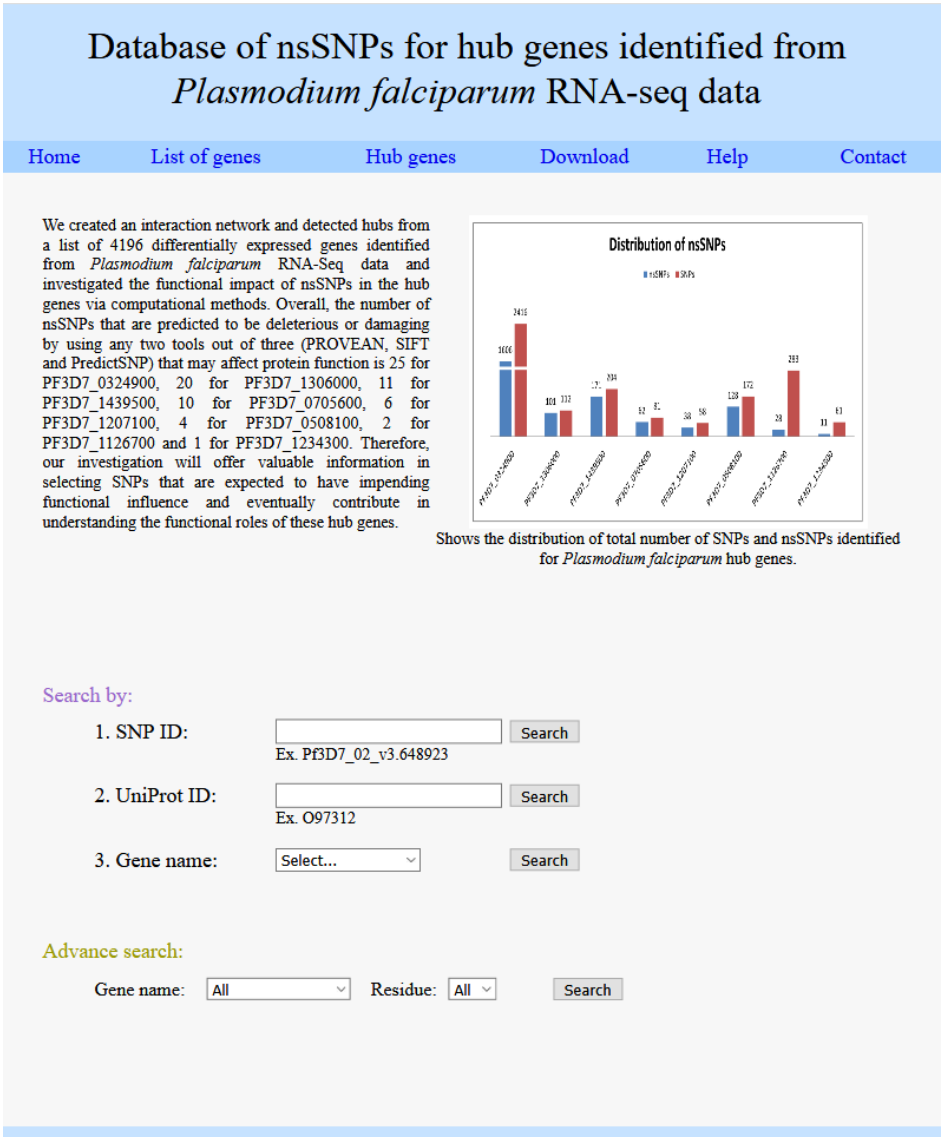
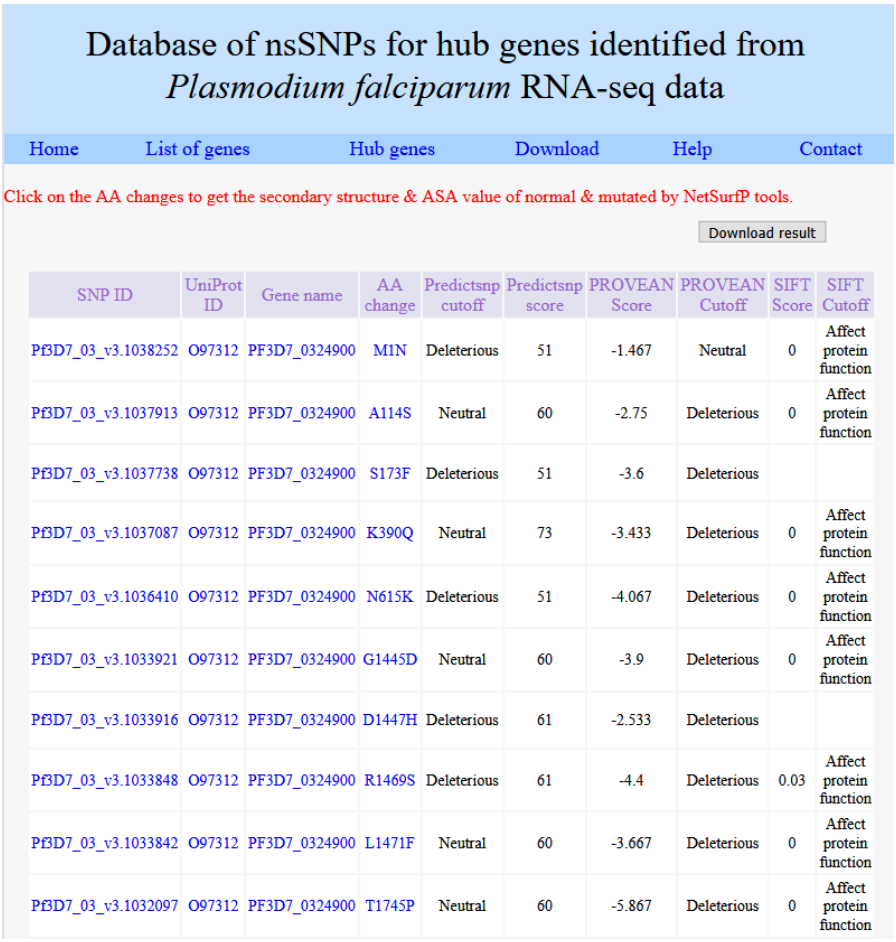


Fig. 4.7 Home page of database of nsSNPs for identified hub genes

An interaction network was created and detected hubs from a list of 4196 differentially expressed genes identified from *P. falciparum* RNA-seq data and analysed the functional impact of nsSNPs in the hub genes using computational methods. Overall, nsSNPs that were predicted to be deleterious or damaging by using any two tools out of three (PROVEAN, SIFT and PredictSNP) can influence protein function is 25 for PF3D7_0324900, 20 for PF3D7_1306000, 11 for PF3D7_1439500, 10 for PF3D7_0705600, 6 for PF3D7_1207100, 4 for PF3D7_0508100, 2 for PF3D7_1126700 and 1 for PF3D7_1234300. Home page of this database tool is depicted in Fig. 4.7.

Database was developed using PHP and JavaScript with user-friendly search environment using a range of options, such as simple searches and advanced searches. Users can search easily with any of the three fields like gene name, UniProt ID or SNP ID. Gene name with corresponding residues is available in advance search. Users can search for particular residues in any gene or genes available in this database (currently the hub genes). For displaying the SNP data, two-step approach is exploited by these two search methods. The first step is to display Gene name, UniProt ID, SNP ID, AA change, PredictSNP cuttoff, Confidence, PROVEAN Score, SIFT Score and cutoff (Fig. 4.8).

The second step will reveal the details of the secondary structure, SNP ID, Residue N, Location, ASA N, Class N, SS N, Residue SNP, SNP, ASA SNP, Class SNP and SS SNP (Fig. 4.9). Users can also customize their search results by choosing any field like Class change, SS change, ASA change or Residue N or any combination for a particular gene. The relevant information is displayed dynamically. In the help section, all headers are defined and hyperlinked to this web portal from the search pages. In addition, every SNP ID is directly linked to its relevant entries in the PlasmoDB database.



Database of nsSNPs for hub genes identified from *Plasmodium falciparum* RNA-seq data

Home | List of genes | Hub genes | Download | Help | Contact

Click on the AA changes to get the secondary structure & ASA value of normal & mutated by NetSurfP tools.

Download result

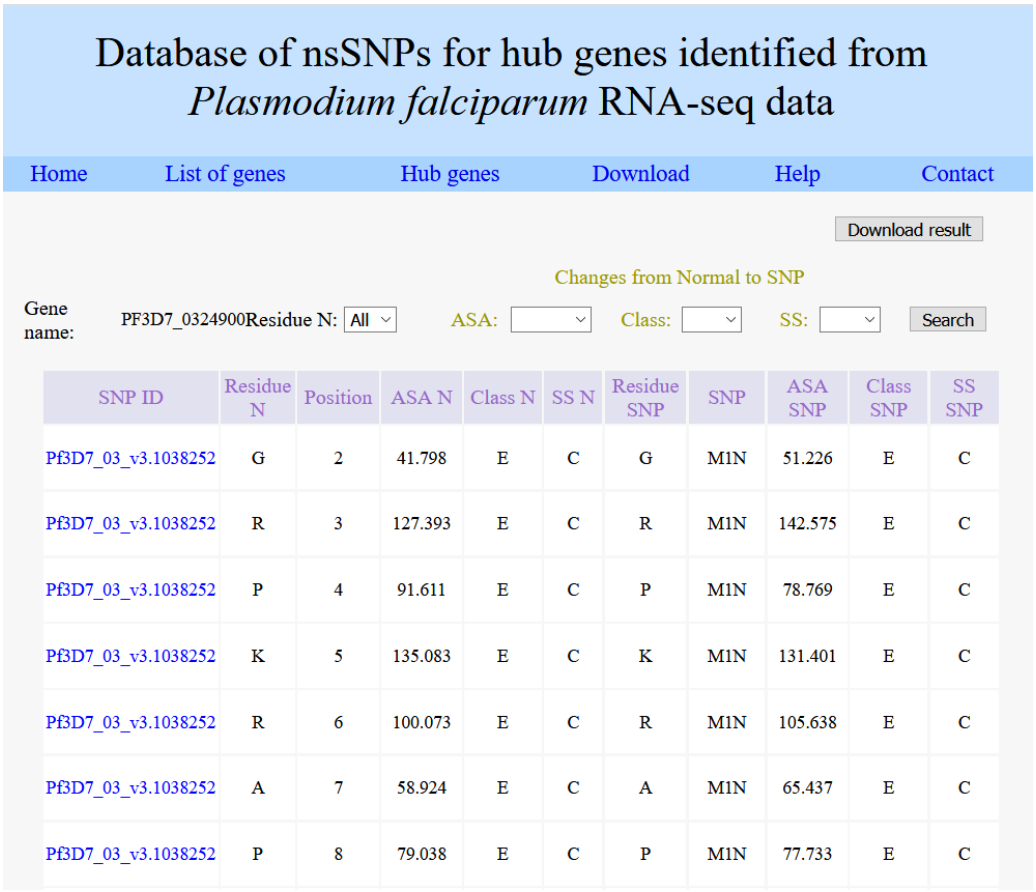
SNP ID	UniProt ID	Gene name	AA change	Predictsnp cutoff	Predictsnp score	PROVEAN Score	PROVEAN Cutoff	SIFT Score	SIFT Cutoff
Pf3D7_03_v3.1038252	O97312	PF3D7_0324900	M1N	Deleterious	51	-1.467	Neutral	0	Affect protein function
Pf3D7_03_v3.1037913	O97312	PF3D7_0324900	A114S	Neutral	60	-2.75	Deleterious	0	Affect protein function
Pf3D7_03_v3.1037738	O97312	PF3D7_0324900	S173F	Deleterious	51	-3.6	Deleterious		
Pf3D7_03_v3.1037087	O97312	PF3D7_0324900	K390Q	Neutral	73	-3.433	Deleterious	0	Affect protein function
Pf3D7_03_v3.1036410	O97312	PF3D7_0324900	N615K	Deleterious	51	-4.067	Deleterious	0	Affect protein function
Pf3D7_03_v3.1033921	O97312	PF3D7_0324900	G1445D	Neutral	60	-3.9	Deleterious	0	Affect protein function
Pf3D7_03_v3.1033916	O97312	PF3D7_0324900	D1447H	Deleterious	61	-2.533	Deleterious		
Pf3D7_03_v3.1033848	O97312	PF3D7_0324900	R1469S	Deleterious	61	-4.4	Deleterious	0.03	Affect protein function
Pf3D7_03_v3.1033842	O97312	PF3D7_0324900	L1471F	Neutral	60	-3.667	Deleterious	0	Affect protein function
Pf3D7_03_v3.1032097	O97312	PF3D7_0324900	T1745P	Neutral	60	-5.867	Deleterious	0	Affect protein function

Fig. 4.8 Prediction of deleterious or damaging from PROVEAN, SIFT and PredictSNP tools

To control the disease in parasites of malaria, the study of genetic variation is of practical importance. In this study, a number of nsSNPs have been identified for nearly all hubs, but few have had an impact on protein functions. In this study, a number of nsSNPs have been identified for nearly all hubs, but few have had an impact on protein functions. PF3D7_0324900, PF3D7_1306000, PF3D7_1439500, PF3D7_0705600, PF3D7_1207100, PF3D7_0508100, PF3D7_1126700 and PF3D7_1234300 genes were annotated by at least two tools to affect protein function. The NetSurfP web server was used to predict the protein secondary structure and surface accessibility in normal for each gene and its predicted SNP substitutions.

Therefore, in selecting SNPs that are supposed to have imminent functional impact, the results of present study gave useful information and ultimately contributed towards understanding the

functional functions of these hub genes. To demonstrate the analysis done by PROVEAN, SIFT, PredictSNP and NetSurfP software, a database has been created, which is available online at URL www.cdkd.org/pfsnp/.



Database of nsSNPs for hub genes identified from *Plasmodium falciparum* RNA-seq data

Home | List of genes | Hub genes | Download | Help | Contact

Download result

Changes from Normal to SNP

Gene name: PF3D7_0324900 Residue N: All ASA: Class: SS: Search

SNP ID	Residue N	Position	ASA N	Class N	SS N	Residue SNP	SNP	ASA SNP	Class SNP	SS SNP
Pf3D7_03_v3.1038252	G	2	41.798	E	C	G	M1N	51.226	E	C
Pf3D7_03_v3.1038252	R	3	127.393	E	C	R	M1N	142.575	E	C
Pf3D7_03_v3.1038252	P	4	91.611	E	C	P	M1N	78.769	E	C
Pf3D7_03_v3.1038252	K	5	135.083	E	C	K	M1N	131.401	E	C
Pf3D7_03_v3.1038252	R	6	100.073	E	C	R	M1N	105.638	E	C
Pf3D7_03_v3.1038252	A	7	58.924	E	C	A	M1N	65.437	E	C
Pf3D7_03_v3.1038252	P	8	79.038	E	C	P	M1N	77.733	E	C

Fig. 4.9 SS and ASA of predicted SNP by NetSurf tool

B. Database tool for identifier mapping

The present study aims to map the identifiers of different databases related to *P. falciparum*. A database tool is designed for easy mapping of different identifiers related to *P. falciparum*. Identifier mapping is a growing challenge in the bioinformatics workflows. It requires integration of experimental data from different sources. A database tool is designed for easy mapping of different identifiers related to *P. falciparum*. PfIDmap database tool can provide mapping of PlasmoDB Gene ID, entrez id, uniprot id, Refseq Protein and string database from either of ids of *P. falciparum*. User can retrieve any identifiers by using above mentioned IDs.

It can also provide information related to chromosome location, gene name, name of protein, genomic accession number. PfIDmap is a step to make as much as possible data easily accessible to scientific community. Each id is directly linked to its relevant entries in the databases.

Data collection and organisation

Different identifiers related to *P. falciparum* were retrieved from different databases. A total of 5687 unique ids were retrieved from the PlasmoDB database, whereas 5670, 5461 and 350 were retrieved from NCBI Entrez, UniProtKB and RCSB PDB databases respectively. 5392 IDs retrieved for both GenBank and RefSeq Proteins. Sequences of *P. falciparum* were collected from PlasmoDB and UniProt databases. Each identifier was mapped to one to another manually. The data in PfIDmap is organized into 13 fields viz. PF ID, Gene ID, Entrez, UniProtKB, PDB, GenBank, RefSeq Protein, Protein length, String ID, chromosome, Location, Gene name, description, genomic accession. Links are provided to access further information, if present in external databases like PlasmoDB, NCBI Entrez, UniProtKB, GenBank, and RefSeq Protein.

Database interface

The PfIDmap interface is designed to allow users to easily navigate the various tools built into the database (Fig. 4.10).

Description

PHP and JavaScript are used for the creation of web interface. User can search easily with any of the three fields (Fig. 4.11), such as Gene ID, Uniprot ID or Entrez ID. User can map their identifiers with the help of eight search options available in ID Mapping (Fig 4.12). Different identifiers like PlasmoDB Gene ID, Entrez ID, UniProt ID, PDB ID, GenBank, RefSeq Protein and String ID can be mapped using ID Mapping (Fig. 4.13). For the functional annotation analysis by DAVID web server, PlasmoDB Gene ID can be mapped to Entrez ID.

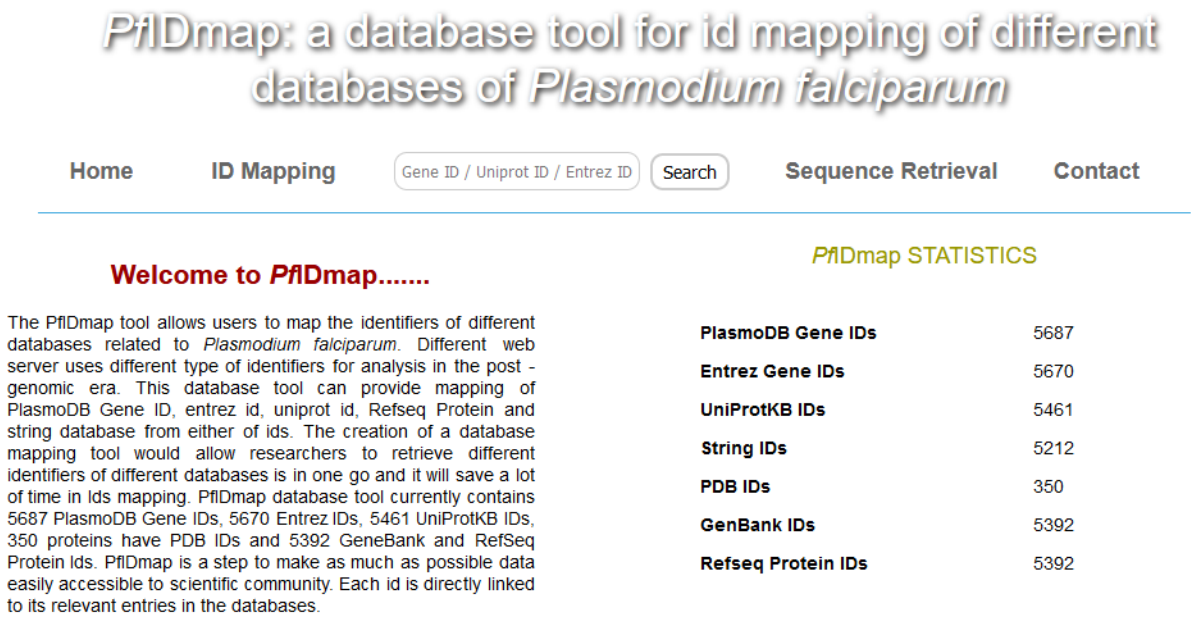


Fig. 4.10 Home page of PfIDmap. Brief description of PfIDmap is available and also displaying statistics of PfIDmap

There is a textbox on header for quick search of Gene ID, Uniprot ID or Entrez ID and an ID mapping link also available on the header. Sequences can be retrieved in FASTA format by using sequence retrieval. Description of database tool available on home page and also contain statistics of different identifiers of different databases used for this database tool.

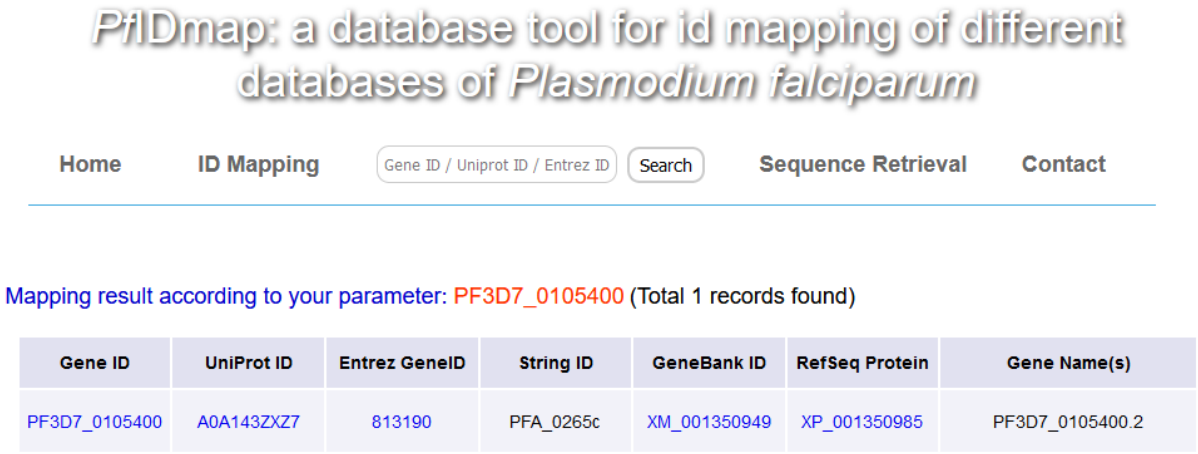


Fig. 4.11 Simple search options in PfIDmap

User can search easily with any of the three fields, such as Gene ID, Uniprot ID or Entrez ID String IDs can be used for the STRING databases for protein-protein interaction (PPI) network analysis and for pathway analysis by KOBAS server (Xie et al. 2011). FASTA format sequences (Fig. 4.14) can be retrieved for pathway analysis by KAAS web server (Moriya et al. 2007).

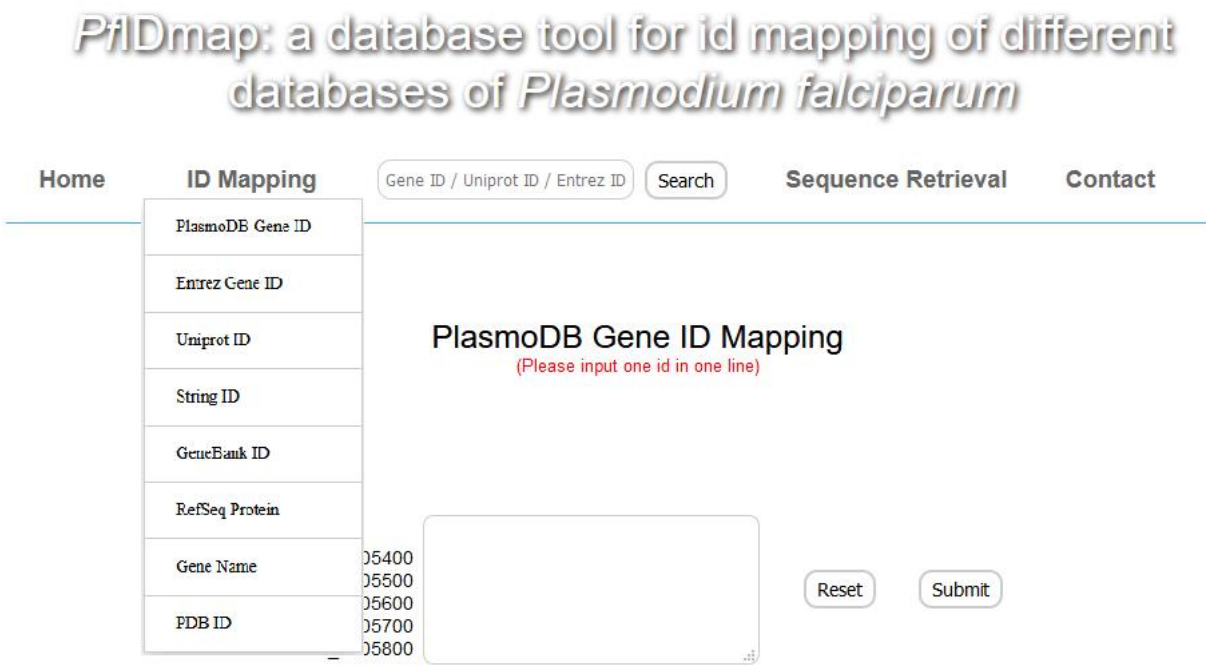
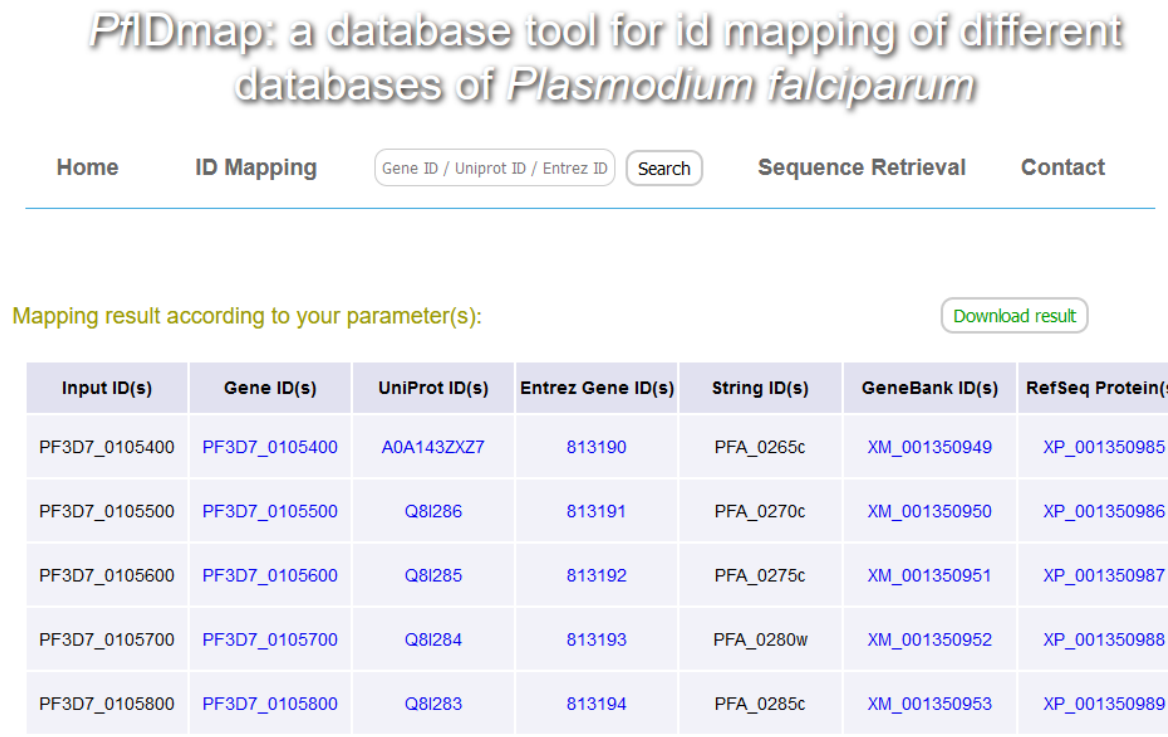


Fig. 4.12 ID mapping options in PfIDmap

User can map their identifiers with the help of eight search options available in ID Mapping

This database tool is designed for easy mapping of different identifiers related to *P. falciparum*. PfIDmap database tool can provide mapping of PlasmoDB Gene ID, entrez id, uniprot id, Refseq Protein and string database from either of ids of *P. falciparum*. PfIDmap database tool currently contains 5687 PlasmoDB Gene IDs, 5670 Entrez IDs, 5461 UniProtKB IDs, 350 proteins have PDB IDs and 5392 GenBank and RefSeq Protein Ids. User can retrieve any identifiers by using above mentioned IDs. It can also provide information related to chromosome location, gene name, name of protein, genomic accession number. PfIDmap is a step to make as much as possible data easily accessible to scientific community. Each id is directly linked to its relevant entries in the databases.

Identifier Mapping is a growing challenge in the bioinformatics workflows. It requires integration of experimental data from different sources. This problem was recognized and a variety of tools were developed to solve it, including BridgeDb framework (van Iersel et al.



Input ID(s)	Gene ID(s)	UniProt ID(s)	Entrez Gene ID(s)	String ID(s)	GeneBank ID(s)	RefSeq Protein(s)
PF3D7_0105400	PF3D7_0105400	A0A143ZXZ7	813190	PFA_0265c	XM_001350949	XP_001350985
PF3D7_0105500	PF3D7_0105500	Q8I286	813191	PFA_0270c	XM_001350950	XP_001350986
PF3D7_0105600	PF3D7_0105600	Q8I285	813192	PFA_0275c	XM_001350951	XP_001350987
PF3D7_0105700	PF3D7_0105700	Q8I284	813193	PFA_0280w	XM_001350952	XP_001350988
PF3D7_0105800	PF3D7_0105800	Q8I283	813194	PFA_0285c	XM_001350953	XP_001350989

Fig. 4.13 Showing result of ID mapping in PfIDmap. Showing mapped result according to search parameters and data can be downloaded in csv format by clicking on download result link.

2010), identifier mapping tools in Cytoscape (Gao et al. 2014; Treister and Pico 2018) DAVID Gene Accession Conversion Tool (Huang et al. 2008), Retrieve/ID mapping tool in UniProt database (Apweiler et al. 2004; UniProt Consortium 2018). PlasmoDB search, DAVID Gene Accession Conversion Tool and UniProt Retrieve/ID mapping tool are currently available as main ID mapping tools for *P. falciparum*.

The mapping of identifiers ignores the problem and significantly limits usability of tool (van Iersel et al. 2010). We assume that integrating identifiers mapping into a tool is better than asking users to manually or with separate software to perform identifiers mapping. That means that we have to deal in the best possible way with the problem of identifier mapping. DAVID web server effectively maps functional annotation analysis using Entrez ID while STRING database tool uses different ids for protein interaction network analysis.

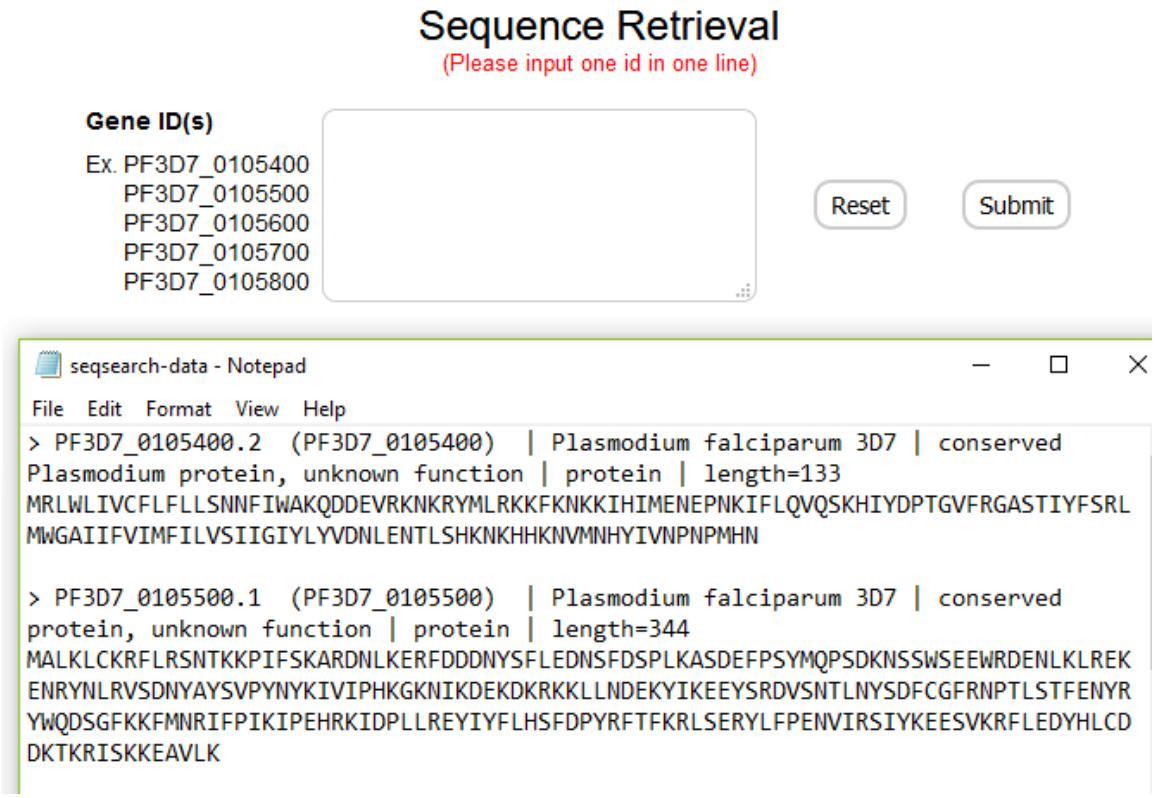


Fig. 4.14 Sequence retrieval result in PfIDmap
Sequences can be downloaded as text file.

The goal of this database tool is to provide easy mapping of different identifiers of different databases related to *P. falciparum*. Researchers can retrieve different identifiers of different databases in one go and it will save a lot of time in ID mapping. PfIDmap is a step to make as much as possible data easily accessible to scientific community. Each id is directly linked to its relevant entries in the databases.

C. Metabolic pathway analysis of *P. falciparum*

The metabolic pathways of malaria parasite were undertaken to find putative drug targets by using different approaches.

4.5 Computational analysis of metabolic pathways

The metabolic pathways of malaria parasite are different from that of a human in a number of ways due to the unique characteristics in the life-cycle of malaria parasite. It is thus very likely

for the malaria parasite to use the specificity of its pathways to devise therapeutic strategies. The metabolic pathway of *P. falciparum* was analyzed by the BIOCYC, MPMP and KEGG databases.

4.5.1 KEGG Automatic Annotation Server (KAAS)

The KEGG Automatic Annotation Server (KAAS) was used to map the pathways of DEGs in all six stages. DEGs amino acid sequences in FASTA format were submitted to the KAAS server. As a result, 277 pathways were predicted for RvET2209, 283 pathways for RvLT2546, 280 pathways for RvSc2735, 285 pathways for RvGII2594, 229 pathways for RvGV1990 and 272 pathways for RvOo2726 were recognised. The top 10 KEGG pathways for each six categories are shown in Table 4.10 and Appendix IV provides a complete list of pathways. It was observed from table 4.10 that most DEGs have been linked with important biological processes, many of which are classified as metabolic pathways, secondary metabolite production pathways, ribosome or being involved in biosynthesis of antibiotics.

4.5.2 Kyoto Encyclopedia of Genes and Genomes (KEGG)

Compounds were retrieved from different pathways of *P. falciparum* and human. All the compounds were compared in both organisms to find compounds only available in *P. falciparum*. List of all available metabolic pathways related to *P. falciparum* was retrieved (Table 4.11).

Compounds were retrieved from all metabolic pathways of human and *P. falciparum* available in KEGG database. A total of 3316 and 1884 compounds were identified in human and *P. falciparum*, respectively. In *P. falciparum*, 166 compounds were uniquely present while comparing above compounds. These 166 compounds have been specifically identified in parasites, and can be targeted.

Table 4.10 Ten top KEGG pathways in each of the six categories

A) RvET		
	Pathway Name	**No. of mapped Genes**
ko01100	Metabolic pathways	102
ko01110	Biosynthesis of secondary metabolites	44
ko03010	Ribosome	43
ko01130	Biosynthesis of antibiotics	32
ko01120	Microbial metabolism in diverse environments	20
ko00230	Purine metabolism	17
ko01200	Carbon metabolism	17
ko00240	Pyrimidine metabolism	14
ko05144	Malaria	13
ko04141	Protein processing in endoplasmic reticulum	12
B) RvLT		
ko01100	Metabolic pathways	146
ko01110	Biosynthesis of secondary metabolites	63
ko01130	Biosynthesis of antibiotics	43
ko00230	Purine metabolism	33
ko01120	Microbial metabolism in diverse environments	27
ko05169	Epstein-Barr virus infection	26
ko00240	Pyrimidine metabolism	26
ko01200	Carbon metabolism	26
ko03030	DNA replication	24
ko00190	Oxidative phosphorylation	23
C) RvSc		
ko01100	Metabolic pathways	129
ko03010	Ribosome	56
ko03040	Spliceosome	55
ko01110	Biosynthesis of secondary metabolites	53
ko01130	Biosynthesis of antibiotics	35
ko00230	Purine metabolism	32
ko00240	Pyrimidine metabolism	30
ko03008	Ribosome biogenesis in eukaryotes	29
ko01120	Microbial metabolism in diverse environments	29
ko03013	RNA transport	27
D) RvGII		
ko01100	Metabolic pathways	137
ko01110	Biosynthesis of secondary metabolites	56
ko03040	Spliceosome	51
ko01130	Biosynthesis of antibiotics	39
ko03010	Ribosome	38
ko01120	Microbial metabolism in diverse environments	29
ko03008	Ribosome biogenesis in eukaryotes	27
ko05169	Epstein-Barr virus infection	27
ko03013	RNA transport	27

ko01200	Carbon metabolism	25
E) RvGV		
ko01100	Metabolic pathways	96
ko01110	Biosynthesis of secondary metabolites	41
ko03010	Ribosome	35
ko03040	Spliceosome	34
ko01130	Biosynthesis of antibiotics	25
ko03008	Ribosome biogenesis in eukaryotes	20
ko00230	Purine metabolism	20
ko01120	Microbial metabolism in diverse environments	19
ko03013	RNA transport	17
ko05016	Huntington's disease	17
F) RvOo		
ko01100	Metabolic pathways	132
ko01110	Biosynthesis of secondary metabolites	60
ko03010	Ribosome	47
ko01130	Biosynthesis of antibiotics	44
ko03040	Spliceosome	36
ko01120	Microbial metabolism in diverse environments	31
ko00230	Purine metabolism	31
ko01200	Carbon metabolism	28
ko03013	RNA transport	27
ko00240	Pyrimidine metabolism	24

Table 4.11 KEGG pathway maps of *P. falciparum* metabolism

Metabolism	Pathway Name	
Global and overview maps	01100	Metabolic pathways
	01110	Biosynthesis of secondary metabolites
	01130	Biosynthesis of antibiotics
	01200	Carbon metabolism
	01210	2-Oxocarboxylic acid metabolism
	01212	Fatty acid metabolism
	01230	Biosynthesis of amino acids
Carbohydrate metabolism	00010	Glycolysis / Gluconeogenesis
	00020	Citrate cycle (TCA cycle)
	00030	Pentose phosphate pathway
	00040	Pentose and glucuronate interconversions
	00051	Fructose and mannose metabolism
	00052	Galactose metabolism
	00500	Starch and sucrose metabolism
	00520	Amino sugar and nucleotide sugar metabolism
	00620	Pyruvate metabolism
	00630	Glyoxylate and dicarboxylate metabolism
	00640	Propanoate metabolism
	00562	Inositol phosphate metabolism
Energy metabolism	00190	Oxidative phosphorylation
	00910	Nitrogen metabolism
Lipid metabolism	00061	Fatty acid biosynthesis
	00062	Fatty acid elongation
	00071	Fatty acid degradation
	00561	Glycerolipid metabolism
	00564	Glycerophospholipid metabolism
	00565	Ether lipid metabolism
	00590	Arachidonic acid metabolism
	01040	Biosynthesis of unsaturated fatty acids
Nucleotide metabolism	00230	Purine metabolism
	00240	Pyrimidine metabolism
Amino acid metabolism	00250	Alanine, aspartate and glutamate metabolism
	00260	Glycine, serine and threonine metabolism
	00270	Cysteine and methionine metabolism
	00280	Valine, leucine and isoleucine degradation
	00310	Lysine degradation
	00220	Arginine biosynthesis
	00330	Arginine and proline metabolism
	00380	Tryptophan metabolism
	00400	Phenylalanine, tyrosine and tryptophan biosynthesis
Metabolism of other amino acids	00440	Phosphonate and phosphinate metabolism
	00450	Selenocompound metabolism
	00480	Glutathione metabolism

Glycan biosynthesis and metabolism	00510	N-Glycan biosynthesis
	00513	Various types of N-glycan biosynthesis
	00514	Other types of O-glycan biosynthesis
	00563	Glycosylphosphatidylinositol (GPI)-anchor biosynthesis
	00604	Glycosphingolipid biosynthesis - ganglio series
Metabolism of cofactors and vitamins	00730	Thiamine metabolism
	00740	Riboflavin metabolism
	00750	Vitamin B6 metabolism
	00760	Nicotinate and nicotinamide metabolism
	00770	Pantothenate and CoA biosynthesis
	00780	Biotin metabolism
	00785	Lipoic acid metabolism
	00790	Folate biosynthesis
	00670	One carbon pool by folate
	00860	Porphyrin and chlorophyll metabolism
	00130	Ubiquinone and other terpenoid-quinone biosynthesis
Metabolism of terpenoids and polyketides	00900	Terpenoid backbone biosynthesis

4.5.3 Malaria Parasite Metabolic Pathways (MPMP)

Malaria Parasite Metabolic Pathways (MPMP) was searched for essential metabolic genes by navigating map analysis menu in search tab and a total of 67 genes were retrieved (Table 4.12). Entrez Gene IDs and PDB ID were retrieved for each gene from respective databases.

Table 4.12 List of essential metabolic genes retrieved from MPMP database

Gene ID	Previously known as	Entrez Gene IDs	PDB ID	Formal Name
PF3D7_0206400	PFB0280w	812659		pentafunctional AROM polypeptide, putative, pseudogene
PF3D7_0206700	PFB0295w	812659		adenylosuccinate lyase
PF3D7_0321200	PFC0935c	814526		UDP-N-acetylglucosamine--dolichyl-phosphate n-acetylglucosaminephosphotransferase, putative
PF3D7_0417200	PFD0830w	9221804		bifunctional dihydrofolate reductase-thymidylate synthase
PF3D7_0513300	PFE0660c	812947	1NW4 1Q1G 1SQ6 2BSX 3ENZ	purine nucleoside phosphorylase

			3FOW 3PHC 6AQS 6AQU	
PF3D7_0603300	PFF0160c	3885966	1TV5 3I65 3I68 3I6R 3O8A 3SFK 4CQ8 4CQ9 4CQA 4ORM	dihydroorotate dehydrogenase
PF3D7_0604700	PFF0230c	3885709		glyoxalase I-like protein gilp
PF3D7_0607300	PFF0360w	3885902		uroporphyrinogen III decarboxylase
PF3D7_0608800	PFF0435w	3885911	3NTJ	ornithine aminotransferase
PF3D7_0615100	PFF0730c	3885811	2FOI 4IGE	enoyl-acyl carrier reductase
PF3D7_0623000	PFF1105c	3885860		chorismate synthase
PF3D7_0626300	PFF1275c	3885996		3-oxoacyl-acyl-carrier protein synthase I/II
PF3D7_0810800	PF08_0095	2655294		hydroxymethyldihydropterin pyrophosphokinase-dihydropteroate synthase
PF3D7_0820700	PF08_0045	2655496		2-oxoglutarate dehydrogenase E1 component
PF3D7_0918900	PFI0925w	813465		gamma-glutamylcysteine synthetase
PF3D7_0922200	PFI1090w	813498		S-adenosylmethionine synthetase
PF3D7_0922400	PFI1100w	813500		para-aminobenzoic acid synthetase
PF3D7_0923800	PFI1170c	813514	4J56 4J57	thioredoxin reductase
PF3D7_1012400	PF10_0121	810279		hypoxanthine-guanine phosphoribosyltransferase
PF3D7_1015800	PF10_0154	810312		ribonucleoside-diphosphate reductase small chain, putative
PF3D7_1022500	PF10_0218	810375		citrate synthase, mitochondrial, putative
PF3D7_1026900	PF10_0409	8445054		biotin-protein ligase 1
PF3D7_1028100	PF10_0275	810432		protoporphyrinogen oxidase
PF3D7_1029600	PF10_0289	810446		adenosine deaminase
PF3D7_1033100	PF10_0322	810479		S-adenosylmethionine decarboxylase/ornithine decarboxylase
PF3D7_1034400	PF10_0334	810491		flavoprotein subunit of succinate dehydrogenase
PF3D7_1108500	PF11_0097	810648		succinyl-CoA synthetase alpha subunit, putative
PF3D7_1113700	PF11_0145	810692		glyoxalase I
PF3D7_1127100	PF11_0282	810829	1VYQ 2Y8C 3T60 3T64 3T6Y 3T70	deoxyuridine 5'-triphosphate nucleotidohydrolase dUTP pyrophosphatase
PF3D7_1128400	PF11_0295	810842		geranylgeranyl pyrophosphate synthase, putative
PF3D7_1129000	PF11_0301	810848	2HTE 2I7C 2PSS 2PT6 2PT9 2PWP	spermidine synthase

			3B7P 3RIE 4BP1 4BP3	
PF3D7_1140000	PF11_0411	810957		carbonic anhydrase
PF3D7_1142400	PF11_0436	810981		coproporphyrinogen-III oxidase HemF
PF3D7_1209600	PFL0480w	811149		porphobilinogen deaminase
PF3D7_1224000	PFL1155w	811283		GTP cyclohydrolase I
PF3D7_1225100	PFL1210w	811294		isoleucine--tRNA ligase, putative
PF3D7_1238600	PFL1870c	811426		sphingomyelin phosphodiesterase
PF3D7_1240000	PFL1940w	811440		3-hydroxyisobutyryl-coenzyme A hydrolase, putative
PF3D7_1246100	PFL2210w	811494		delta-aminolevulinic acid synthetase
PF3D7_1251300	PFL2465c	811545	2WWF 2WWG 2WWH 2WWI 2YOF 2YOG 2YOH	thymidylate kinase
PF3D7_1251700	PFL2485c	811549		tryptophanyl-tRNA synthetase, putative tryptophan--tRNA ligase, putative
PF3D7_1308200	PF13_0044	814023		carbamoyl phosphate synthetase
PF3D7_1324900	PF13_0141	814112	3ZH2	L-lactate dehydrogenase
PF3D7_1325200	PF13_0144	814115		lactate dehydrogenase
PF3D7_1327600	PF13_0159	814129	5LLT 5LM3	nicotinamide/nicotinic acid mononucleotide adenylyltransferase
PF3D7_1332900	PF13_0179	814149		isoleucine--tRNA ligase, putative
PF3D7_1336900	PF13_0205	814174	4J75 4J76 4JFA	tryptophanyl-tRNA synthetase tryptophan--tRNA ligase
PF3D7_1342100	PF13_0229	814196		aconitate hydratase
PF3D7_1343600	MAL13P1.218	813775		UDP-N-acetylglucosamine pyrophosphorylase, putative
PF3D7_1345700	PF13_0242	814208		isocitrate dehydrogenase [NADP], mitochondrial
PF3D7_1351600	PF13_0269	814234	2W40 2W41	glycerol kinase
PF3D7_1354500	PF13_0287	814251		adenylosuccinate synthetase
PF3D7_1364900	MAL13P1.326	813892		ferrochelatase, HemH
PF3D7_1405600	PF14_0053	811635		ribonucleoside-diphosphate reductase small chain, putative
PF3D7_1408000	PF14_0077	811659	5YIA 5YIB 5YIC 5YID 5YIE	plasmepsin II
PF3D7_1416500	PF14_0164	811745		NADP-specific glutamate dehydrogenase
PF3D7_1419300	PF14_0187	811768	2AAW	glutathione S-transferase
PF3D7_1419800	PF14_0192	811773		glutathione reductase
PF3D7_1420600	PF14_0200	811781		pantothenate kinase 1, putative
PF3D7_1437200	PF14_0352	811934		ribonucleoside-diphosphate reductase large subunit, putative
PF3D7_1437400	PF14_0354	811936		pantothenate kinase 2, putative

PF3D7_1440300	PF14_0381	811963		delta-aminolevulinic acid dehydratase porphobilinogen synthase, HemB (4.2.1.24)
PF3D7_1444800	PF14_0425	812007	4TR9	fructose-bisphosphate aldolase
PF3D7_1445100	PF14_0428	812010		histidine--tRNA ligase, putative
PF3D7_1450900	PF14_0484	812066		acetyl-CoA acetyltransferase, putative
PF3D7_1467300	PF14_0641	812223	4Y67 4Y6P 4Y6R 4Y6S 5JAZ 5JBI 5JC1 5JMP 5JMW 5JNL	1-deoxy-D-xylulose 5-phosphate reductoisomerase
PF3D7_1469600	PF14_0664	812246		acetyl-coa carboxylase

Also, a list of 894 essential and non-mutable genes were retrieved by using map analysis menu available on MPMP database server. It eventually disrupts the pathways that are critical for parasite survival by targeting these essential metabolic pathway genes and essential and non-mutable genes.

4.5.4 BioCyc Pathway/Genome Database Collection

BioCyc database contains 208 pathways and 970 enzymatic reactions for *P. falciparum* 3D7 strain. Chokepoint analysis was performed using Chokepoint Reactions menu under Metabolism tab available on BioCyc database.

It is difficult to manually monitor all cellular processes because of the amount of interactions between these biological entities in an organism. Using BioCyc webserver to classify essential proteins, further metabolic chokepoint analysis is performed using the criteria: exclude reactions found in humans, exclude reactions catalysed by more than one enzyme, and limit the reaction found in multiple pathways. Chokepoint reaction finder was used for *P. falciparum* 3D7 by excluding reactions found in human and reactions catalyzed by more than one enzyme. A total of 284 and 290 chokepoints reactions were found respectively on the consuming side and the producing side.

Specifically, the parasite is supposed to be harmed by targeting enzymes that either uniquely generate or consume a substrate called ‘chokepoint enzymes’. If carefully chosen, using insight into both the host and the pathogen’s biology and metabolic requirements, such targets have

the potential to selectively damage the parasites without inappropriate host side effects. The enzymes forming a pathogen's metabolic network are therefore possible targets for drug development.

D. SNPs of the GPI-anchor transamidase

GPI-anchor transamidase (GPI-T) is a potential drug target primarily of its crucial role in the development and survival of the parasite in the GPI anchor biosynthesis pathway. The present investigation was undertaken to explore the plausible effects of nsSNP on the structure and functions of GPI-T subunit GPI8p of *P. falciparum*.

4.6 Prediction of deleterious nsSNPs in the GPI-anchor transamidase

The *P. falciparum* GPI8p (PF3D7_1128700) consist of 1482 bp and 493 amino acids. The GPI8p investigated in this study had a total of 40 SNPs, 34 of which were nsSNPs. Only non-synonymous SNPs have been selected for further analysis as non-synonymous mutations, which may change the protein sequences and ultimately change the structure and function of protein.

4.6.1 Prediction of deleterious coding nsSNPs

The SIFT sequence tool predicted a total of 28 variants that had an effect on protein function while 4 variants had no effect on GPI8p. Overall, 4 nsSNPs (T81S, Q121 K, R158 L and T195S) were found to be tolerated with a score of greater than 0.05. Three nsSNPs were recognized as deleterious with a score of 0.01 while the remaining 25 nsSNPs showed a highly deleterious score of 0.00. Two nsSNPs (out of 32) were expected to be deleterious with the PROVEAN tool, with a PROVEAN score below -2.5, and the remaining nsSNPs (30) showed scores above the limit recognizing them as neutral. The PROVEAN tool uses -2.5 for all predictions as a cut-off score. The amino acid sequence of the query protein, mutation positions and desired mutations were submitted using the PredictSNP input page in the FASTA format. According to PredictSNP, 13 mutations were expected to be deleterious while 19 were found

to be in neutral in GPI8p. The findings from the SNAP2 server showed 15 effective variants, while the remaining 17 nsSNPs were neutral. By combining the observations of four prediction tools (SIFT, PROVEAN, PredictSNP and SNAP2), 18 nsSNPs (R124L, N143K, Y145 F, V157I, T195S, K379E, I392K, I437T, Y438H, N439D, Y441H, N442D, N448D, N451D, D457A, D457Y, I458L and N460K) have been shown to influence the protein function by at least two software tools (Table 4.13). These nsSNPs were used for further analyses.

Table 4.13 The nsSNPs that were predicted to affect protein function by at least two programs (SIFT, PROVEAN, PredictSNP and SNAP2) in GPI8p

Amino acid change	SIFT	PROVEAN	PredictSNP	SNAP2
R124L	Affect protein function	Neutral	Neutral	effect
N143K	Affect protein function	Neutral	Deleterious	effect
Y145F	Affect protein function	Deleterious	Deleterious	effect
V157I	Affect protein function	Neutral	Neutral	effect
T195S	Tolerated	Deleterious	Deleterious	neutral
K379E	Affect protein function	Neutral	Neutral	effect
I392K	Affect protein function	Neutral	Neutral	effect
I437T	Affect protein function	Neutral	Deleterious	neutral
Y438H	Affect protein function	Neutral	Deleterious	neutral
N439D	Affect protein function	Neutral	Deleterious	effect
Y441H	Affect protein function	Neutral	Neutral	effect
N442D	Affect protein function	Neutral	Deleterious	effect
N448D	Affect protein function	Neutral	Deleterious	effect
N451D	Affect protein function	Neutral	Deleterious	effect
D457Y	Affect protein function	Neutral	Deleterious	effect
D457A	Affect protein function	Neutral	Deleterious	effect
I458L	Affect protein function	Neutral	Deleterious	neutral
N460K	Affect protein function	Neutral	Deleterious	effect

Mutation T195S was predicted to be tolerated, while it was predicted that the rest of the mutations (R124L, N143K, Y145F, V157I, K379E, I392K, I437T, Y438H, N439D, Y441H, N442D, N448D, N451D, D457Y, D457A, I458L and N460K) would affect protein function

by the SIFT sequence tool.

Only two mutations (Y145F and T195S) were found deleterious by PROVEAN tool while rest of mutations were found to be neutral. The PredictSNP tool found mutations (N143 K, Y145F, T195S, I437 T, Y438H, N442D, N448D, N451D, D457Y, D457A, I458L and N460 K) to be deleterious, while mutations (R124L, V157I, K379E, I392 K and Y441H) were expected to be neutral. Mutations (T195S, I437 T, Y438H and I458L) were found to be neutral by the SNAP2 tool, while mutations (R124L, N143K, Y145F, V157I, K379E, I392K, N439D, Y441H, N442D, N448D, N451D, D457Y, D457A and N460K) were predicted to affect protein function. Only mutation Y145F was predicted to be deleterious by all the four tools.

4.6.2 Prediction of mutation impacts on the stability of proteins

In order to predict the DDG stability and reliability index (RI) upon mutation, the selected variants were subjected to I-Mutant 2.0 web server. According to I-Mutant 2.0, the results on amino acid substitutions predicted either an increase or a decrease in the free energy. All of the modified nsSNPs following mutation resulted decrease in protein stability with a range in the reliability index 1 - 9. Similarly, all the mutations predicted by I-Mutant 2.0 web server to decrease protein stability were also predicted by MuPro server to decrease protein stability. The prediction of changes in stability by I-Mutant 2.0 and MuPro of the 18 selected nsSNPs is provided in Table 4.14.

This finding indicates that the amino acid interactions could be directly or indirectly destabilized by these mutations of GPI8p, leading to functional deviations of the protein.

Table 4.14 I-Mutant and MuPro outcomes for nsSNPs in GPI8p protein

			I-Mutant 2.0			MuPro	
Position	**WT**	**NEW**	**Stability**	**RI**	**DDG**	**Stability**	**DDG**
124	R	L	Decrease	6	-0.25	Decrease	-0.231
143	N	K	Decrease	1	0.36	Decrease	-1.335
145	Y	F	Decrease	2	-0.05	Decrease	-0.569
157	V	I	Decrease	6	-0.53	Decrease	-0.983
195	T	S	Decrease	7	-0.54	Decrease	-1.178
379	K	E	Decrease	2	-0.71	Decrease	-0.602
392	I	K	Decrease	9	-1.19	Decrease	-1.672
437	I	T	Decrease	7	-1.93	Decrease	-2.639
438	Y	H	Decrease	5	-0.39	Decrease	-1.557
439	N	D	Decrease	2	-0.19	Decrease	-0.458
441	Y	H	Decrease	5	-0.09	Decrease	-1.628
442	N	D	Decrease	1	0.15	Decrease	-0.517
448	N	D	Decrease	2	-0.19	Decrease	-0.517
451	N	D	Decrease	2	-0.19	Decrease	-0.734
457	D	Y	Decrease	5	-0.42	Decrease	-0.434
457	D	A	Decrease	8	-0.94	Decrease	-1.187
458	I	L	Decrease	7	-0.75	Decrease	-1.024
460	N	K	Decrease	5	-0.81	Decrease	-0.800

WT: Wild type residue; NEW: Residue after mutation; RI: Reliability index; DDG (Delta-delta-G)

4.6.3 Conservation of amino acids

The results of the ConSurf tool consist of a structural representation of the protein containing the colorimetric conservation score (Fig. 4.15). Out of the 18 most deleterious SNPs, 2 amino acids with conservation score of 9, 1 with conservation score of 8 and 8 with conservation score of 6 were predicted by ConSurf. One amino acid was predicted as average conserved region and 6 residues as variable (Table 4.15).

Table 4.15 Conservation profile of amino acids in GPI8p

Position	Amino Acid	Conservation Score	ConSurf Prediction
124	R124	6	Exposed
143	N143	8	Highly conserved and exposed (f)
145	Y145	9	Highly conserved and buried (s)
157	V157	5	Buried
195	T195	9	Highly conserved and buried (s)
379	K379	1	Exposed
392	I392	4	Buried
437	I437	4	Buried
438	Y438	2	Buried
439	N439	2	Buried
441	Y441	1	Buried
442	N442	6	Buried
448	N448	6	Buried
451	N451	6	Buried
457	D457	6	Buried
457	D457	6	Buried
458	I458	6	Buried
460	N460	6	Buried

Positions N143, Y145 and T195 were predicted in highly conserved regions, hence showing more chances to alter the protein structure. The highly conserved residues are often essential for biological function.

4.6.4 Prediction of solvent accessibility and secondary structure of protein

Secondary structures and solvent accessibility of protein was investigated using NetSurfP-2.0 server. Moreover, the data was filtered by selecting only those residues that showed an ASA change from buried to exposed state and vice versa and change of its secondary structure. Due to mutations N143K, T195S, I392K, I392K, N448D, N451D, D457Y, D457Y, D457Y, D457A, D457A, D457A and N460K residues V366, V366, F345, L352, F345, F345, F345,

N451, Y453, F345, N451, Y453 and F345 respectively showed a change in class from exposed state to buried state and conformation change from coil to helix.

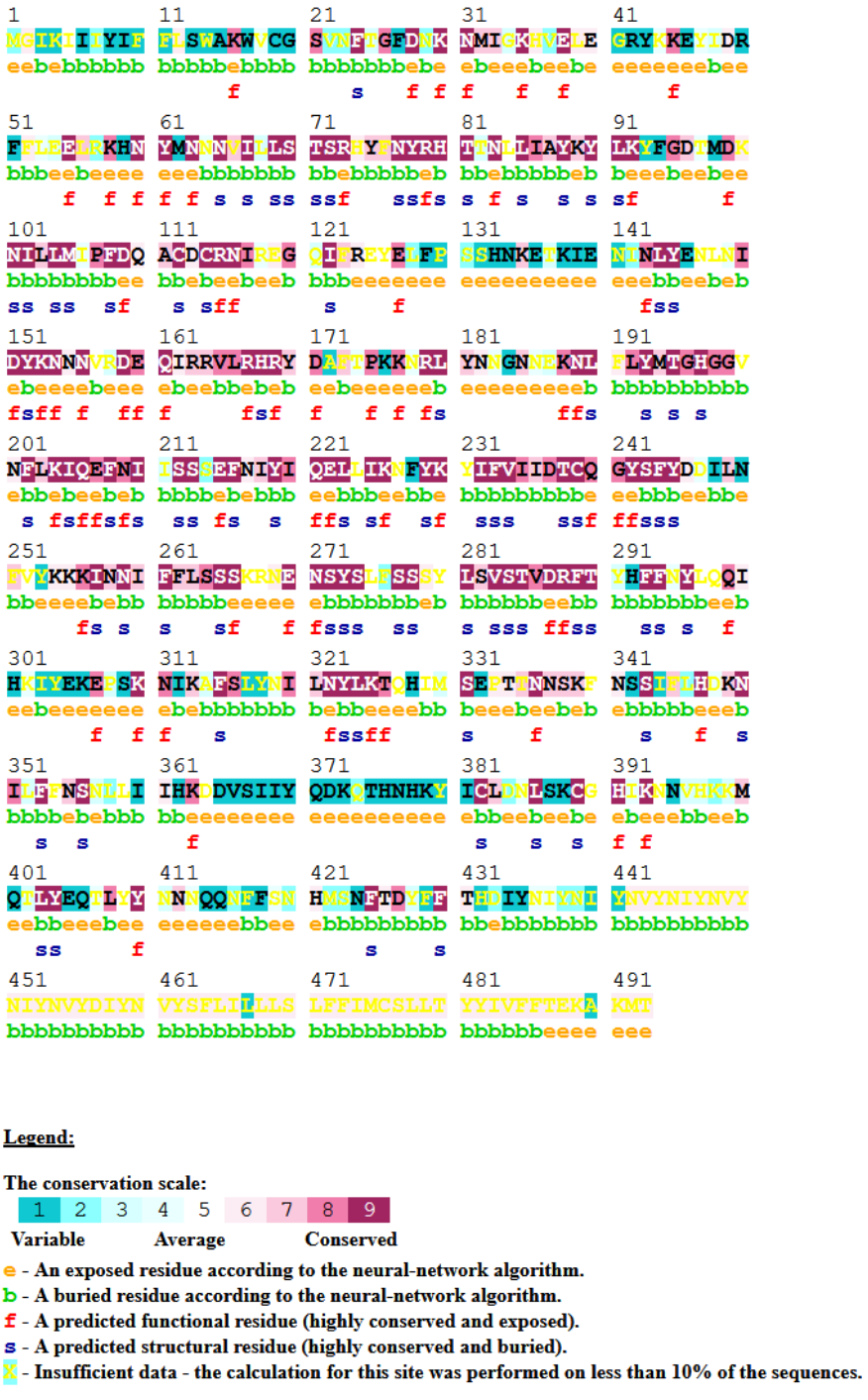


Fig. 4.15 Evolutionary stability of amino acid positions in GPI8p

Unique and conserved regions in the GPI8p protein determined using ConSurf. The color coding bar shows the color scheme representation of the conservation score. Score of conservation is 1–4 for variable, 5–6 for intermediate and 7–9 for conserved.

Table 4.16 Surface accessibility and secondary structure of wild type and mutant variants by NetSurfP

ResidueN	Position	ASA N	Class N	SS N	Residue SNP	SNP	ASA SNP	Class SNP	SS SNP	ASA diff
I	329	45.121	B	H	I	R124L	47.459	E	C	-2.338
V	366	40.7	E	C	V	N143K	35.672	B	H	5.028
V	366	40.7	E	C	V	T195S	36.55	B	H	4.15
F	345	57.367	E	C	F	I392K	38.945	B	H	18.422
L	352	48.69	E	C	L	I392K	44.894	B	H	3.796
F	345	57.367	E	C	F	N448D	41.279	B	H	16.088
F	345	57.367	E	C	F	N451D	41.621	B	H	15.746
F	345	57.367	E	C	F	D457Y	40.489	B	H	16.878
N	451	51.017	E	C	N	D457Y	32.574	B	H	18.443
Y	453	53.696	E	C	Y	D457Y	48.285	B	H	5.411
F	345	57.367	E	C	F	D457A	39.302	B	H	18.065
N	451	51.017	E	C	N	D457A	29.643	B	H	21.374
Y	453	53.696	E	C	Y	D457A	43.361	B	H	10.335
F	345	57.367	E	C	F	N460K	42.659	B	H	14.708

On the other hand, residue I329 due to the mutation R124L showed the opposite change to the exposed state from buried state and also show an opposite change in conformation from helix to coil (Table 4.16).

4.6.5 Protein 3D modeling and structural analysis

The 3D structure of PF3D7_1128700 was created using the available protein sequence with homology-based modelling. If Z-score is higher than 7.5, the respective model will be deemed as good otherwise poor. All the predicted models of PF3D7_1128700 by MUSTER had a Z score above 8.75. Our results stated that all the relevant templates could be regarded as good types (Table 4.17).

Table 4.17 Z score value of different templates analyzed by MUSTER

Rank	Template	Align_length	Coverage	Zscore	Seq_id	Type
1	4fguA	394	0.799	11.297	0.155	Good
2	5h0iA	377	0.764	11.121	0.175	Good
3	5zbiA	390	0.791	11.003	0.138	Good
4	5nijA	396	0.803	10.846	0.167	Good
5	6idvA	392	0.795	10.407	0.156	Good
6	6dhiA	388	0.787	10.265	0.17	Good
7	5nijA1	293	0.594	9.305	0.181	Good
8	5zbiA1	274	0.555	8.85	0.179	Good
9	6dhiA1	282	0.572	8.781	0.213	Good
10	4fguA1	277	0.561	8.753	0.191	Good

Different templates were found based on alignment score

4.6.6 Prediction of protein ligand binding site and protein - protein interactions

The structure predicted by MUSTER with template 4fguA (z-score: 11.29) was taken as input for COACH server. The best ranked active site of the PF3D7_1128700 by COACH with a C-score of 0.14. It was predicted R79, H80, G196, H197, G198, D237, C239, S272, Y273, S274, S284, D287 and R288 as consensus binding residues by using 4aw9A PDB hit.

Additionally, in order to examine the protein-protein interaction of PF11 0298 (Fig. 4.16), we used STRING maps. STRING results predicted the functional association partner of PF11_0298 protein with PF11_0229 (Conserved *Plasmodium* protein), MAL13P1.165 (GPI transamidase subunit PIG-U, putative), MAL13P1.348 (Uncharacterized protein), PFL0685w (Phosphatidylinositol-glycan biosynthesis class O protein, putative), PFL2270w (GPI mannosyltransferase 2), PIG-M (GPI mannosyltransferase I), PFF0915w (N-acetylglucosamine transferase), PF10_0316 (N-acetylglucosaminyl-phosphatidylinositol biosynthetic protein, putative), PF11_0361 (Uncharacterized protein) and Alg9 (Mannosyltransferase-III).

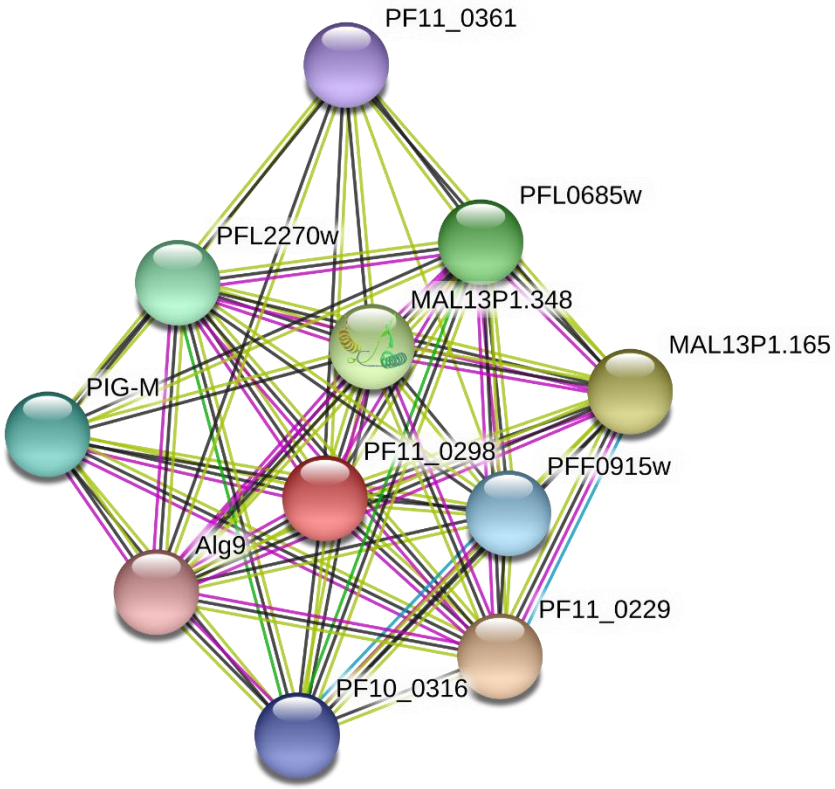


Fig. 4.16 Protein–protein interaction network of PF3D7_1128700 protein shown by STRING. Query proteins and first shell of interactors are shown as colored nodes while second shell of interactors as white nodes. Empty nodes show proteins of unknown 3D structure and filled nodes shows some 3D structure is known or predicted. Different coloured lines represents different interactions, here sky blue and violet lines are used for known interactions and green, red and navy blue for predicted interactions.

Among these proteins, PF11_0229 and MAL13P1.165 are involve in attachment of GPI anchor to protein and PFL0685w, PFL2270w, PIG-M, PFF0915w, PF10_0316 and Alg9 involve in GPI anchor biosynthetic process. These two MAL13P1.348 and PF11_0361 are uncharacterized protein.

Many lower and higher eukaryote glycoproteins are attached to the plasma membrane using a glycosylphosphatidylinositol (GPI) anchor, which is added in the endoplasmic reticulum (ER)

to newly synthesized proteins. GPI transamidase mediates anchoring of GPI in endoplasmic reticulum by replacing the C-terminal GPI attachment signal peptide of a protein with a fully assembled GPI (Liu et al. 2018). In parasitic protozoa, GPI Anchor Proteins, a significant form of membrane proteins, are particularly abundant (Ferguson et al. 1999). After exiting from the salivary glands, sporozoites establish infections are coated by GPI-AP circumsporozoite proteins (CSP) (Wang et al. 2005). Several GPI-APs are also expressed by merozoites that cause significant symptoms such as merozoite surface protein 1 (MSP1) (Das et al. 2015). Merozoites or parasite-infected red blood cells of *P. falciparum* release GPIs that contributes to severe symptoms by causing production of cytokines like TNFα (GPI toxin) (Schofield et al. 2002). GPI-anchor transamidase (GPI-T) is a potential drug target primarily because of its crucial role for the development and survival of the parasite in the GPI anchor biosynthesis pathway. So, the present investigation was undertaken to explore the plausible effects of nsSNP on structure and functions of GPI anchor transamidase of *P. falciparum*.

Various bioinformatics tools were used to investigate the impact of nsSNPs of PF3D7_1128700 gene in this study. Among 34 nsSNPs subjected for functional analysis, 18 nsSNPs (Arg124Leu, Asn143Lys, Tyr145Phe, Val157Ile, Thr195Ser, Lys379Glu, Ile392Lys, Ile437Thr, Tyr438His, Asn439Asp, Tyr441His, Asn442Asp, Asn448Asp, Asn451Asp, Asp457Tyr, Asp457Ala, Ile458Leu, and Asn460Lys) were predicted by at least two or more software tools out of four tools used. Additionally, I-Mutant 2.0 and MuPro both showed a decrease in stability after mutation as a result of these nsSNPs, suggesting to some extent that protein is directly or indirectly destabilized. Only three nsSNPs found in highly conserved regions was predicted by phylogenetic analysis using ConSurf. This finding suggests that most of highly conserved regions are intact. Furthermore, NetSurfP tool revealed that solvent accessibility and secondary structures due to these SNPs were changed mainly from exposed state to buried state and conformation change from coil to helix. The 3D structure of protein

form sequence was generated using MUSTER tool and the best model was taken as input by COACH server to predict protein ligand binding site. It was found that no mutation was present at the predicted ligand binding site. The STRING database results showed that PF3D7_1128700 protein, interact with those proteins which either involve in attachment of GPI anchor to protein or GPI anchor biosynthetic process.

In this study, many predicted deleterious SNPs were identified in GPI8p and evaluated for their possible deleterious effect on the function and stability of the protein. Of the 34 nsSNPs, 18 nsSNPs is predicted to affect protein function through at least two or more software tools out of four tools. Because of these SNPs, the solvent accessibility was primarily altered from exposed to buried state and the conformation of secondary structures from coil to helix. Interestingly, ConSurf prediction suggests that most of region of this protein were highly conserved. Additionally, COACH server found that no mutation was present at predicted ligand binding sites. Inhibition of GPI8p may disrupt the GPI anchor biosynthesis pathways which in turns, prevents GPI anchoring of protein, as anchoring plays a pivotal role in virulence of the parasite. Therefore, the study provides functional and structural impact of nsSNPs and conservation of amino acid positions in the protein, which can be used for further studies to design a therapeutic target that will stabilize the expression of the gene.

E. Thiamine phosphate synthase (PfThiE) as a drug target

Thiamine phosphate synthase (PfThiE) is possible drug targets because of its role and essentialness in the thiamine biosynthesis pathway. The present study aims to model the three-dimensional (3D) structure of thiamine phosphate synthase and to predict the potential inhibitors to derive therapeutic objectives for *P. falciparum*.

4.7 PfThiE: a potential drug target in *P. falciparum*

Drug resistance is increasingly emerging in malaria parasites, so it is important to identify and develop alternative anti-malarial agents against both new and existing drug targets. In

apicomplexan parasites, thiamine biosynthesis offers a potential and exciting chance to achieve such goals, as the pathway is found only in prokaryotes, fungi, and plants, but is not present in mammals (Wrenger et al. 2006, 2008). Thiamine pyrophosphate (Thi-PP) is the active form of vitamin B1, which is a co-factor for various enzymes primarily involved in the metabolism of carbohydrates such as 2-oxoglutarate dehydrogenase, transketolase or pyruvate dehydrogenase. For a few days, the culturing of *P. falciparum* in a thiamine deficient medium showed no adverse effect but a substantial need for 4-amino-5-hydroxymethyl-2-methylpyrimidine (HMP) or thiamine itself for parasite growth was reported after ten days (Wrenger et al. 2006).

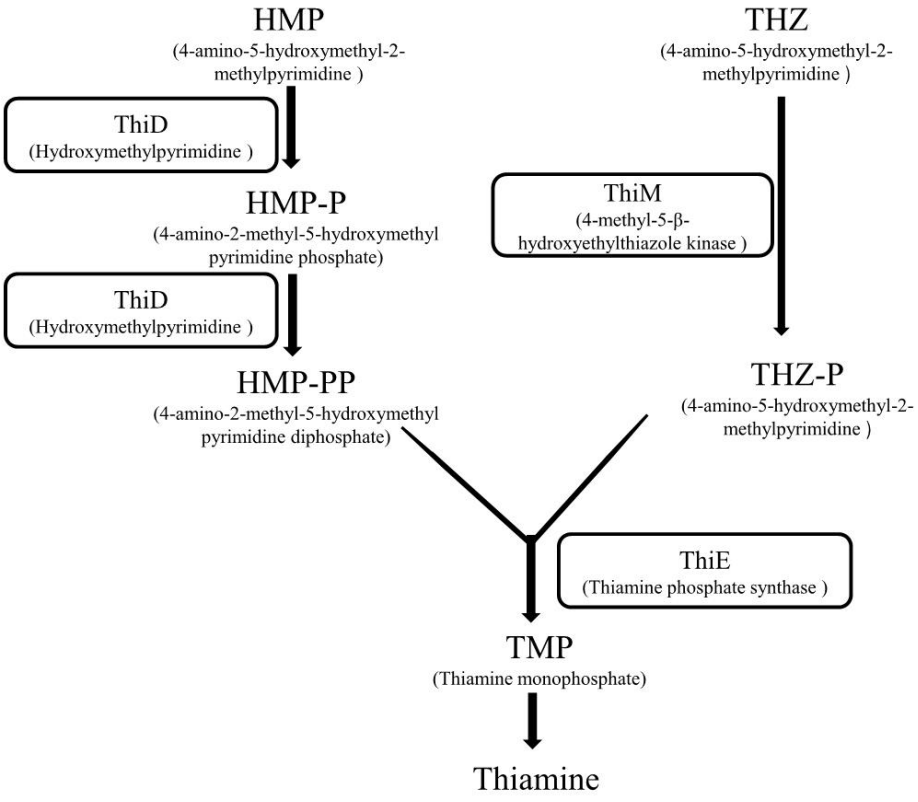


Fig. 4.17 The flow diagram of Thiamine biosynthesis pathway. Thiazole (THZ) and pyrimidine (HMP) moieties are synthesized in separate branches of the pathway. The coupling reaction of THZ-P and HMP-PP using ThiE produces Thiamine monophosphate.

In thiamine biosynthesis pathway (Fig. 4.17), thiazole THZ-P (5-methyl-4-(beta-hydroxyethyl)thiazole phosphate) and pyrimidine HMP-PP (2-methyl-4-amino-5-hydroxymethylpyrimidine pyrophosphate) moieties are combined to yield thiamine phosphate by PfThiE (Wrenger et al. 2008; Zhang et al. 1997). For several enzymes, thiamine is metabolized as an essential cofactor (Chan et al. 2013). So, a novel drug target thiamine phosphate synthase (PfThiE) of *P. falciparum* which is essential enzyme in thiamine biosynthesis was chosen to screen potent anti-malarial drugs. The human host's lack of vitamin biosynthesis signifies that inhibition of the parasite pathways can be a way to particularly interfere with the development of parasites (Müller et al. 2007).

The present study aimed to investigate the possible effects of nsSNPs in PfThiE and their effects on its structure and function, 3D structure formation and prediction of inhibitors for the modelled structure. Till date no reports are available on the effect of deleterious SNPs and docking studies experimentally or computationally on PfThiE of *P. falciparum*. Usually, causative SNPs occur in different forms: those found in the gene coding area and those residing in non-coding areas, such as the regulatory sequences of the gene (Schlauch et al. 2016). Non-synonymous single nucleotide polymorphisms (nsSNPs) lead to variations in the amino acid sequence, as they influence the primary polypeptide directly. These changes are not only associated with their primary sequence modification but can also alter or impair the structure and function of protein in the amino acid sequence. Numerous studies have predicted most deleterious nsSNPs among recorded polymorphisms and understood their effect on protein function, structure, and stability (Desai et al. 2017; Solayman et al. 2017). Subudhi et al. (2015) studied SNPs of *P. falciparum* (Inhibitors have also been studied against Atypical Chemokine Receptor 1 (ACKR1), a receptor that plays a major role in the *P. vivax* and *P. falciparum* host entry mechanisms (Narwani et al. 2018; Horuk 2015). Functional analysis, stability analysis and conservation analysis were performed for PfThiE protein. The 3D structure of PfThiE was

developed and validated. Several potent ligand molecules were identified by virtual screening method and evaluated through ADMET (Absorption, Distribution, Metabolism, Excretion and Toxicity) properties. The interactions between proteins and ligands were studied in this study using molecular docking. In the lack of the molecular structure, the proposed 3D model will be useful in providing a novel target against malaria for structure-based drug design.

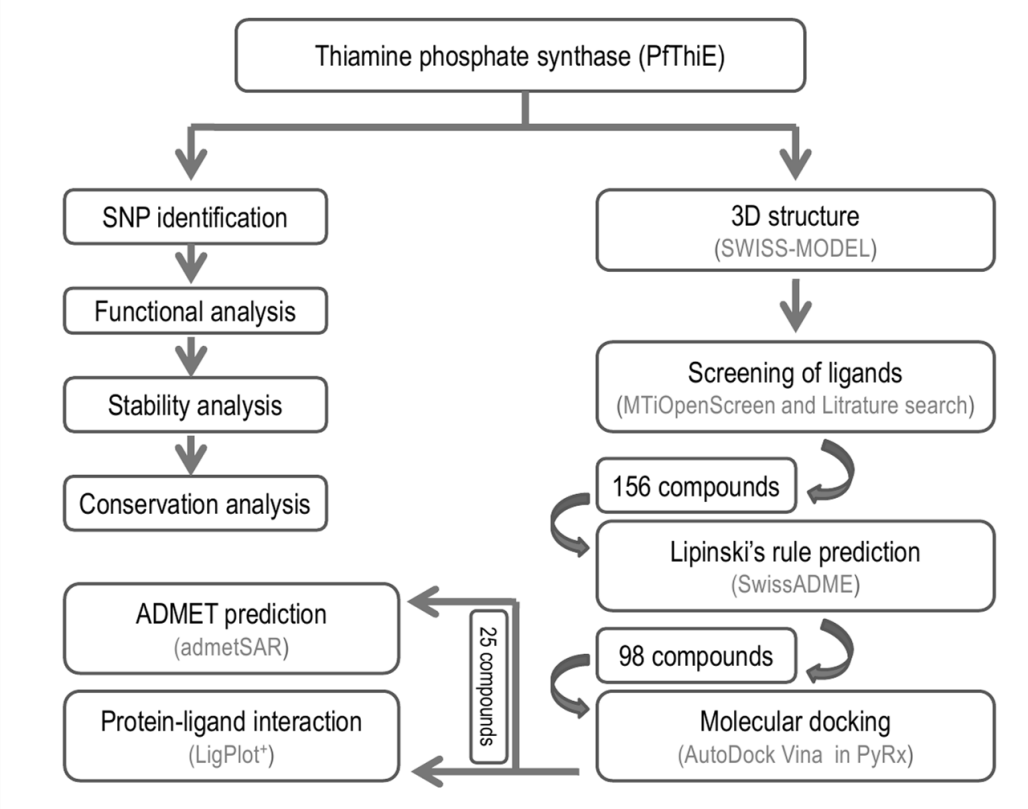


Fig. 4.18 The schematic representation of the work flow for the analysis of PfThiE

There are two key parts of this study: sequence and structure. Sequence part used the protein sequence to identify SNPs, functional analysis, stability analysis, conservation analysis, and 3D structure modelling. The *P. falciparum* thiamine phosphate synthase (PfThiE) consists of total 1617 bp, with a total of 538 amino acids in its protein. The Thiamin phosphate pyrophosphorylase gene studied had a total of 14 nsSNPs in PlasmoDB database. For further study, all these non-synonymous coding SNPs have been chosen. Second part starts with the modelled 3D structure of protein. The 3D structure of protein was used for the screening of

compounds, ADMET analysis and molecular docking studies. The schematic representation of the work flow for the analysis is depicted in Fig. 4.18.

4.7.1 Functional analysis of nsSNPs

To examine whether these SNPs have any impact on protein structure or function of PfThiE, we subjected all nsSNPs to five separate damaging prediction tools. The protein amino acid sequence in FASTA format and list of mutation positions and mutations were submitted to prediction tools for predicting harmful effects.

Table 4.18 The nsSNPs that predicted to affect protein function by SIFT, PROVEAN, PredictSNP and SNAP2 tools in PfThiE

Amino acid change	SIFT	PROVEAN	PredictSNP	SNAP2
A145V	Affect protein function	Deleterious	Neutral	neutral
G165D	Affect protein function	Deleterious	Deleterious	effect
H190P	Affect protein function	Neutral	Neutral	neutral
I220V	Affect protein function	Neutral	Neutral	neutral
N239H	Affect protein function	Neutral	Neutral	neutral
L310V	Affect protein function	Neutral	Neutral	neutral
D311E	Affect protein function	Neutral	Neutral	neutral
S330C	Affect protein function	Neutral	Deleterious	neutral
N355S	Affect protein function	Neutral	Neutral	neutral
G401E	Affect protein function	Neutral	Deleterious	neutral
S427W	Affect protein function	Neutral	Deleterious	effect
I433L	Affect protein function	Neutral	Neutral	neutral
C456F	Affect protein function	Neutral	Deleterious	neutral
D494E	Affect protein function	Neutral	Neutral	neutral

The SIFT sequence tool predicted all of 14 variants that had an effect on protein function in PfThiE. Two nsSNPs (G165D and H190P) with a score of 0.01 and 1.0 nsSNPs (I220V) with a score of 0.02 were deleterious, while the remaining 11 nsSNPs exhibited a deleterious score of 0.00. Two nsSNPs (A145V and G165D) were considered to be deleterious with the

PROVEAN method, with a PROVEAN score below -2.5, and the remaining nsSNPs (12) were recognized as neutral. The PROVEAN method utilizes -2.5 as a cut-off value for all predictions. According to PredictSNP, in PF3D7_0614000, 5 nsSNPs (G165D, S330C, G401E, S427W and C456F) were predicted to be deleterious while 9 were found to be in neutral. The results from the SNAP2 server anticipated that two variants (G165D and S427W) would be effective, while the remaining 12 nsSNPs were intended to be neutral. Only one nsSNP (G165D) was estimated to affect protein function by all prediction tools (SIFT, PROVEAN, PredictSNP and SNAP2) (Table 4.18).

4.7.2 Analysis of mutation effects on the protein stability

I-Mutant further evaluated all 14 SNPs for their effect on the stability of proteins. For each mutation the Reliability Index (RI) was predicted.

Table 4.19 I-Mutant 2.0 outcomes for 14 nsSNPs in the protein PfThiE

			I-Mutant 2.0	
Position	**WT**	**NEW**	**Stability**	**RI**
145	A	V	Decrease	1
165	G	D	Decrease	6
190	H	P	Decrease	5
220	I	V	Decrease	8
239	N	H	Decrease	9
310	L	V	Decrease	9
311	D	E	Decrease	2
330	S	C	Decrease	2
355	N	S	Decrease	7
401	G	E	Decrease	0
427	S	W	Decrease	0
433	I	L	Decrease	2
456	C	F	Decrease	5
494	D	E	Increase	4

WT: Wild type residue; NEW: Residue after mutation; RI: Reliability index

Among the 14 SNPs proposed to predict stability, 13 predicted a decrease in the stability of the protein while one was found to increase the stability. With the exception of D494E, all nsSNPs decreased the stability of proteins with a range in the reliability index (RI) of 0 - 9 after mutation. Analysis of mutation effects on the protein stability of 14 nsSNPs is provided in Table 4.19. This result suggests that these mutations of PfThiE may directly or indirectly destabilize the amino acid interactions, leading to functional protein deviations.

Table 4.20 Conservation profile of amino acids in PfThiE

Position	Amino Acid	Conservation Score	ConSurf Prediction
145	A145	8	Buried
165	G165	6	Exposed
190	H190	1	Exposed
220	I220	1	Exposed
239	N239	4	Exposed
310	L310	6	Buried
311	D311	6	Exposed
330	S330	4	Exposed
355	N355	1	Exposed
401	G401	1	Exposed
427	S427	1	Buried
433	I433	9	Highly conserved and buried (s)
456	C456	7	Buried
494	D494	4	Exposed

4.7.3 Conservation analysis of deleterious nsSNPs

To further explore the possible effects of deleterious nsSNPs, the evolutionary conservation of amino acid residues of PfThiE protein was calculated using ConSurf web server. The ConSurf tool predictions consist of a structural protein representation, which includes the colorimetric conservation score (Fig. 4.19).

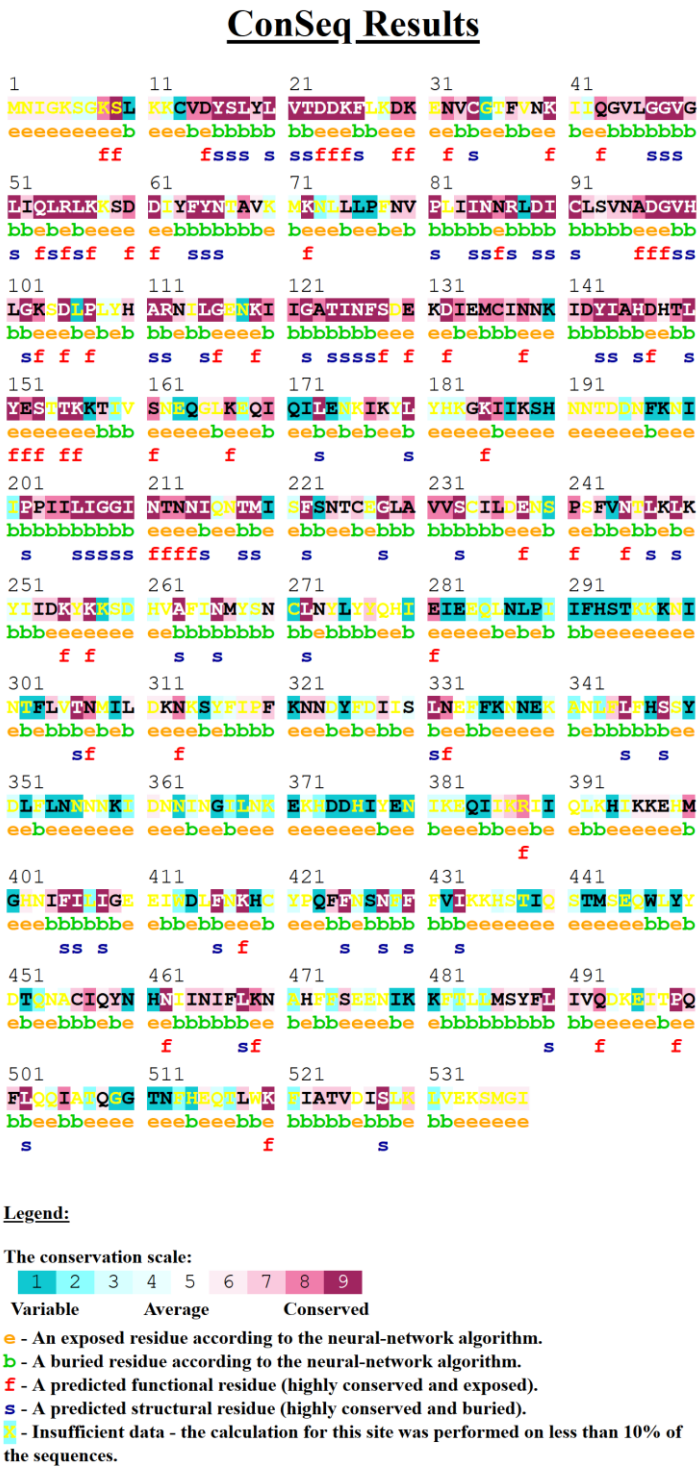


Fig. 4.19 Evolutionary stability of amino acid positions in PfThiE

Unique and conserved regions in the PfThiE protein determined using ConSurf. The color coding bar shows the color scheme representation of the conservation score. Score of conservation is 1–4 for variable, 5–6 for intermediate and 7–9 for conserved.

ConSurf predicted I433, A145 and C456 with conservative score 9, 8 and 7 respectively. Conservation score 6 was projected for G165, L310 and D311 while score 4 was for N239, S330 and D494. However, the remaining 5 amino acids (H190, I220, N355, G401 and S427) with conservative score 1 were predicted in variable region. Positions I433, A145 and C456 were expected in highly conserved regions so mutation in these suggests more possibility of altering the protein structure. The ConSurf findings are presented in Table 4.20.

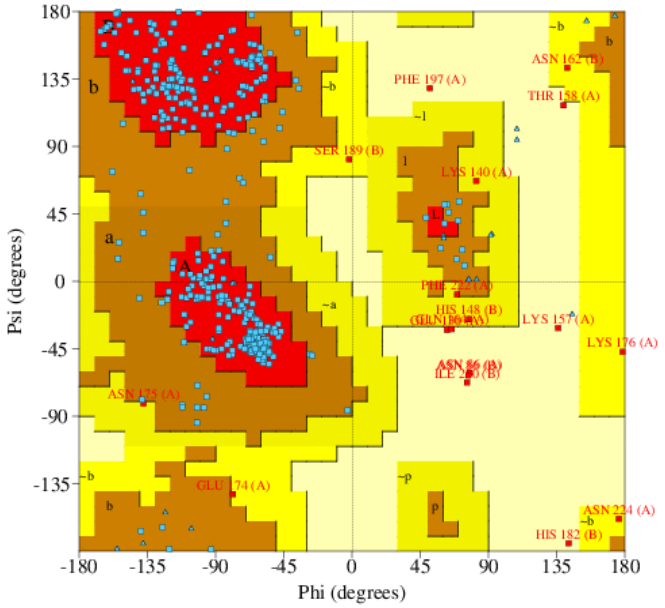
The residues that are highly conserved are sometimes important for biological function. If it is possible to target these conserved residues, the entire pathway would be blocked by disrupting this enzyme.

4.7.4 Protein 3D modeling and structural analysis

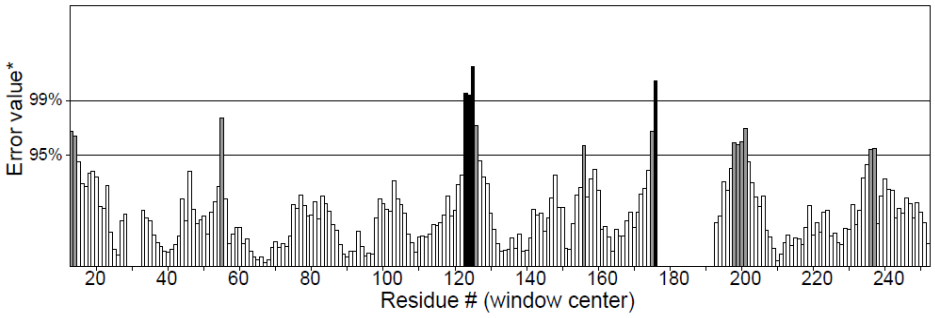
Total three 3D structures of PfThiE were generated by SWISS-MODEL. The best model with GMQE scores 0.22 and QMEAN score −4.71 was generated using thiamine phosphate pyrophosphorylase from *Pyrococcus furiosus* (pdb ID: 1xi3.1.A) as a template for this purpose. The structure for PfThiE was predicted as homo-dimer. Ramachandran plot of the 3D model showed 82.7% of its residues in the core while 13.3% in allowed, 2.0% in generously allowed and 2.0% in disallowed regions (Fig. 4.20A). Overall ERRAT quality score of 80.8 suggested that the structure could be regarded as a good model (Fig. 4.20B). Verify 3D result passed the model with 81.25% of the residues have averaged 3D-1D score > = 0.2 (Fig. 4.20C).

4.7.5 Target and template alignment

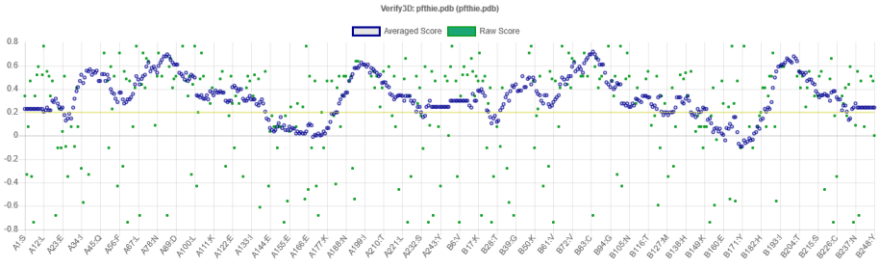
The Dali web server was used for alignment of target and template 3D structure of proteins. Alignment of template and modelled structure and sequence was performed in Dali (Fig. 4.21A & Fig. 4.21B). Alignment score was predicted in the form of Z-score 32.9 and 32.4 for chain A and chain B, respectively. The identical amino acids are labelled with vertical bars. It also provides secondary structure of amino acid of protein by DSSP.



(A)



(B)



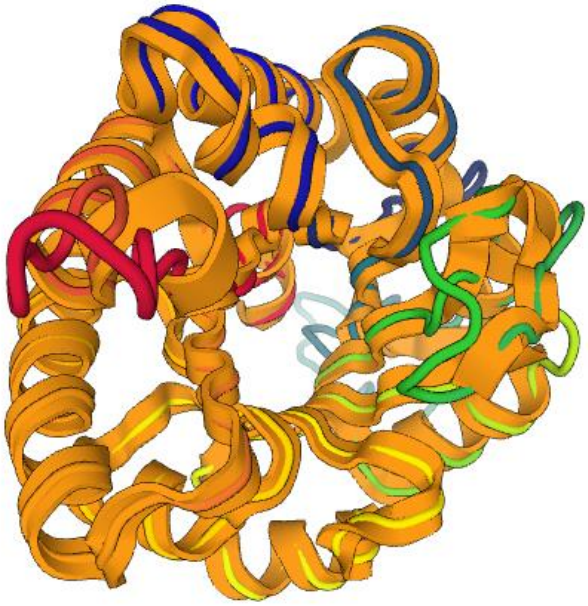
(C)

Fig. 4.20 Structure validation of the PfThiE

(A) Ramachandran plot shows the residue in most favored regions (82.7%), additional allowed regions (13.3%), generously allowed regions (2.0%) and disallowed regions (2.0%). (B) ERRAT plot, the overall quality factor is 92.73% and (C) the QMEAN Z-score is −4.71.

```
DSSP  LLLLllLLLEEEELLLEeehhhleeeELHHHHHHHHHHLLLLEEEELLLLLLHHHHHHHH
Query SLKKcvDYSLYLVTDDKflkdkenvcGTFVNKIIQGVLGGVGLIQLRLKKSDDIYFYNTA   60
ident  |       ||  ||              |        ||   || | |       |
Sbjct NLRN--KLKLYVITDRR--------lKPEVESVREALEGGATAIQMRIKNAPTREMYEIG   50
DSSP  LHHH--HLLEEEELLLL--------lLLHHHHHHHHHHLLLLEEEELLLLLLHHHHHHHH


DSSP  HHHHHHHHHHLLLEEEELLHHHHHHLLLLEEEELLLLLLHHHHHHHHlLLLEEEEEELLl
Query VKMKNLLLPFNVPLIINNRLDICLSVNADGVHLGKSDLPLYHARNILgENKIIGATINFs  120
ident      |            | |   | | |||| ||   | |    |   |    | ||||
Sbjct KTLRQLTREYDALFFVDDRVDVALAVDADGVQLGPEDMPIEVAKEIA-PNLIIGASVYS-  108
DSSP  HHHHHHHHHLLLEEEEELLHHHHHHHLLLEEEELLLLLLHHHHHHHL-LLLEEEEEELL-


DSSP  lLHHHHHHHHLLLLEEEELlLLLLLllllLLLLL-LHHHHHHHLLlllllllllllll
Query dEKDIEMCINNKIDYIAHDhTLYESttkkTIVSNE-QGLKEQIQILenkikylyhkgkii  179
ident              ||                   ||
Sbjct -LEEALEAEKKGADYLGAG-SVFPT----DARVIGlEGLRKIVESV--------------  148
DSSP  -HHHHHHHHHHLLLEEEEE-LLLLL----LLLLLHhHHHHHHHHHL--------------


DSSP  lllllllllllllLLLEEEELLLLLLLHHHHHLLlllLLLLEEELHHHHLLLLLLLlLHH
Query kshnntddnfkniIPPIILIGGINTNNIQNTMISfsnTCEGLAVVSCILDENSPSFvNTL  239
ident               |   |||||  |             | || |           |
Sbjct -------------KIPVVAIGGINKDNAREVLKT---GVDGIAVISAVMGAEDVRK-ATE  191
DSSP  -------------LLLEEEELLLLLLLHHHHHLL---LLLEEEELHHHHLLLLHHH-HHH


DSSP  HHHHHHHHL--
Query KLKYIIDKY--  248
ident  |  |
Sbjct ELRKIVEEVlg  202
DSSP  HHHHHHHHHhl
```

(A)



(B)

Fig. 4.21 Sequence and structure alignment of target and templet

(A) Alignment sequences of template and modelled protein. (B) Alignment structures of template and modelled protein.

4.7.6 Screening of compounds

Several compound databases were searched for thiamine phosphate synthase inhibitors, and 12 from BRENDA and 6 from Drug Bank were taken. Also 39 top Inhibitors against *Mycobacterium tuberculosis* thiamin phosphate synthase from Khare et al. (2011) study was retrieved from PubChem database. MTiOpenScreen screened top 100 drug-like compounds from 10,000 compound libraries that may be inhibitors for thiamine phosphate synthase. 156 compounds were screened from all searches. These 156 drug-like compounds were retrieved in SDF format.

4.7.7 Molecular properties analysis

The ADME and drug-likeness predictions of 156 compounds were carried out using SwissADME. Out of 156 compounds, 98 follows the Lipinski's rule of five is provided as Appendix V.

4.7.8 Molecular Docking and interaction analysis

Molecular Docking was performed by using AutoDock Vina in PyRx 0.8 with 98 compounds which followed the Lipinski's rule. PyRx was initially used to minimize compounds energy and convert all molecules to AutoDock Ligand (PDBQT) format. The value of the grid box was set to center_x = 17.7657, center_y = 21.3541, center_z = 69.9140 while size_x = 65.1899, size_y = 58.2979, and size_z = 83.6737. The default exhaustiveness value was 8. All 98 compounds without any specified binding sites were docked against whole surface of the protein. Table 4.21 lists the outcomes of docking results that were shown as binding affinity less than -8.0 in at least one pose. The graphical representation of the 25 best screened compounds is depicted in the Fig. 4.22. LigPlot^{+} software was used to predict all residues that interact with the cofactors of protein.

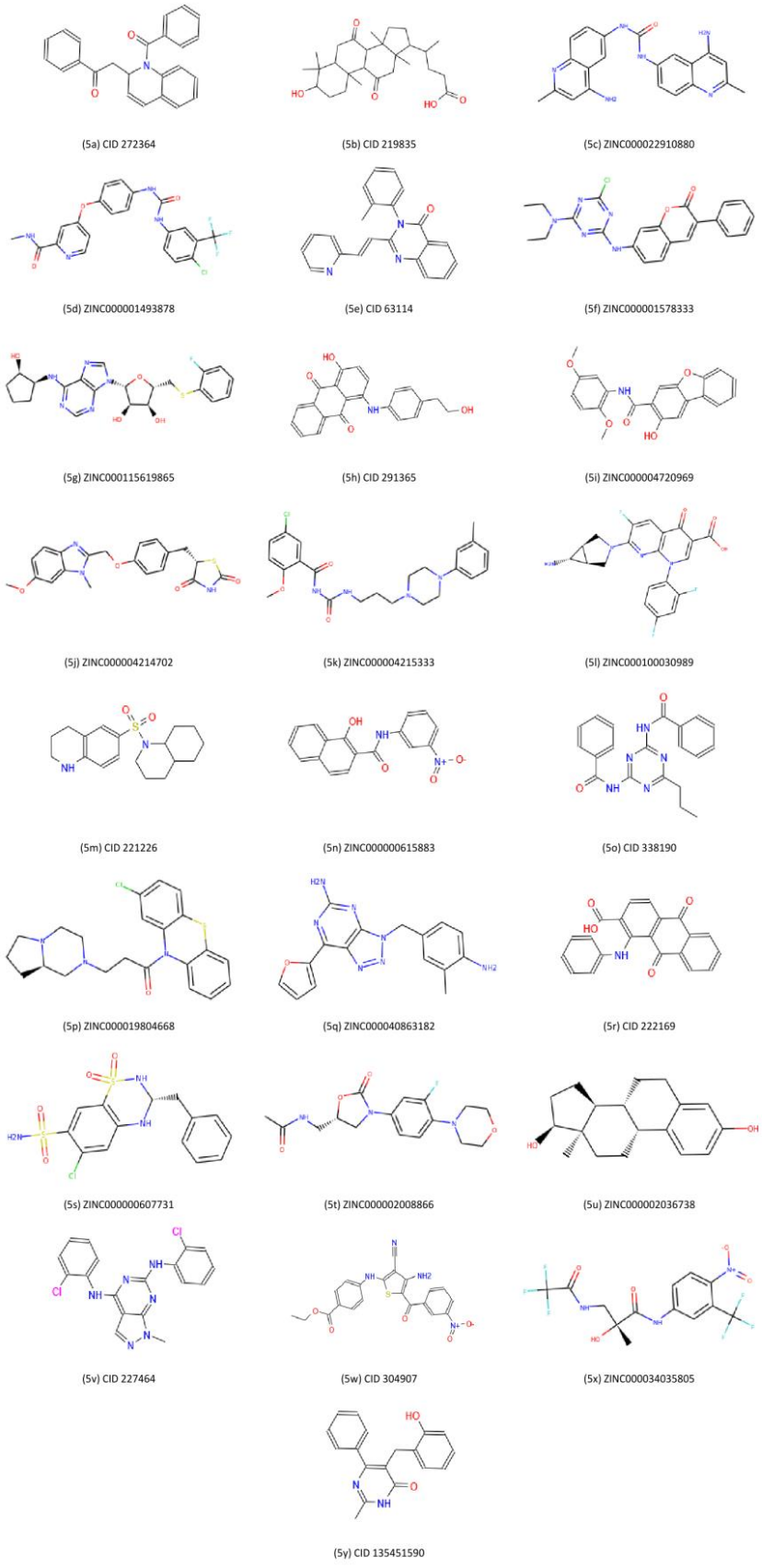
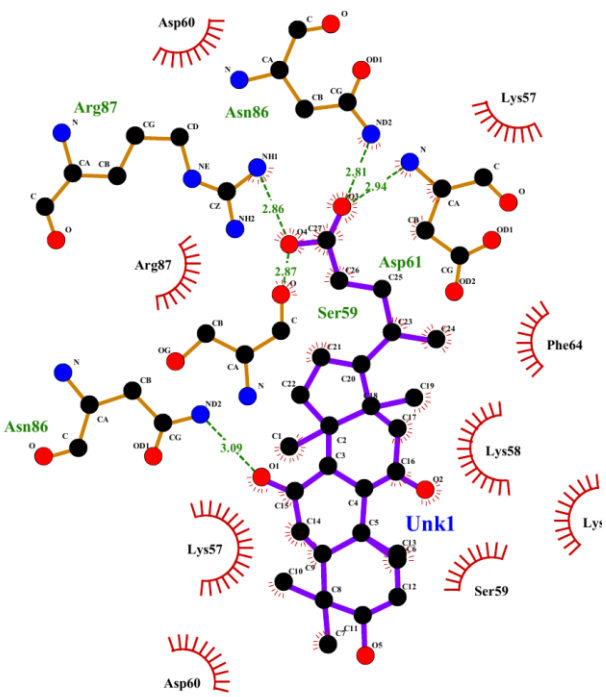


Fig. 4.22 Graphical representations of the best screened inhibitors for PfThiE

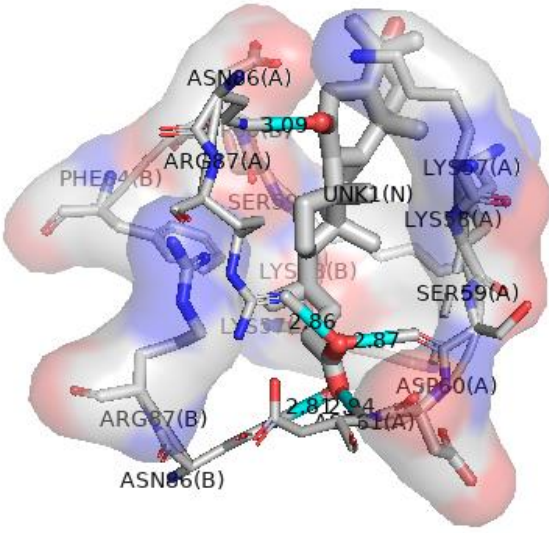
Table 4.21 Docking results of different poses showing binding affinity with less than -8.0 in at least one pose

Compounds	**Binding affinity (kcal/mol)**								
	Pose 1	Pose 2	Pose 3	Pose 4	Pose 5	Pose 6	Pose 7	Pose 8	Pose 9
CID 272364	-11.2	-11.2	-11.1	-10.8	-10.7	-10.7	-10.7	-10.6	-10.4
CID 219835	-9.6	-8.1	-7.9	-7.7	-7.6	-7.6	-7.6	-7.5	-7.3
ZINC000022910880	-9.3	-9.1	-9	-8.9	-8.9	-8.8	-8.8	-8.8	-8.3
ZINC000001493878	-9	-8.7	-8.7	-8.5	-8.4	-8.4	-8.3	-8	-8
CID 63114	-9	-8.6	-8.4	-8.3	-8.3	-7.4	-7.3	-7.3	-7.3
ZINC000001578333	-8.9	-8.3	-8.3	-8.2	-8	-7.9	-7.8	-7.7	-7.7
ZINC000115619865	-8.8	-8.5	-8.3	-8.2	-8	-7.9	-7.8	-7.7	-7.7
CID 291365	-8.8	-8.4	-8.3	-8.2	-7.9	-7.8	-7.8	-7.5	-7.2
ZINC000004720969	-8.7	-8.5	-8.4	-8.1	-8.1	-7.8	-7.6	-7.4	-7.3
ZINC000004214702	-8.6	-8.5	-8.2	-8.2	-8.1	-8.1	-8	-7.8	-7.6
ZINC000004215333	-8.6	-8.5	-8.4	-8.3	-8.2	-7.6	-7.5	-7.4	-7.4
ZINC000100030989	-8.6	-8	-7.9	-7.6	-7.4	-7.3	-7.3	-7.3	-7.1
CID 221226	-8.6	-8.2	-8.1	-8	-7.6	-7.5	-7.4	-6.8	-6.8
ZINC000000615883	-8.6	-8.5	-8.2	-7.9	-7.7	-7.6	-7.6	-7.5	-7.4
CID 338190	-8.5	-7.8	-7.8	-7.8	-7.8	-7.7	-7.3	-7.3	-7.3
ZINC000019804668	-8.4	-7.4	-7.3	-7.3	-7.2	-7.2	-7	-7	-7
ZINC000040863182	-8.4	-8.2	-8.1	-7.8	-7.6	-7.6	-7.6	-7.6	-7.3
CID 222169	-8.4	-8.1	-7.8	-7.5	-7.4	-7.4	-7.2	-7.1	-7.1
ZINC000000607731	-8.3	-7.7	-7.6	-7.5	-7.3	-7	-6.9	-6.9	-6.9
ZINC000002008866	-8.2	-7.9	-7.8	-7.8	-7.6	-7.3	-7.3	-7.1	-7.1
ZINC000002036738	-8.2	-7.7	-7.7	-7.3	-7.2	-7.2	-7	-6.9	-6.7
CID 227464	-8.2	-8.2	-7.7	-7.7	-7.6	-7.5	-7.3	-7.2	-7.2
CID 304907	-8.2	-8.2	-8.1	-7.9	-7.7	-7.7	-7.6	-7.5	-7.5
ZINC000034035805	-8.1	-8	-7.9	-7.7	-7.7	-7.7	-7.6	-7.6	-7.6
CID 135451590	-8.1	-7.3	-7	-6.9	-6.7	-6.6	-6.5	-6.5	-6.5

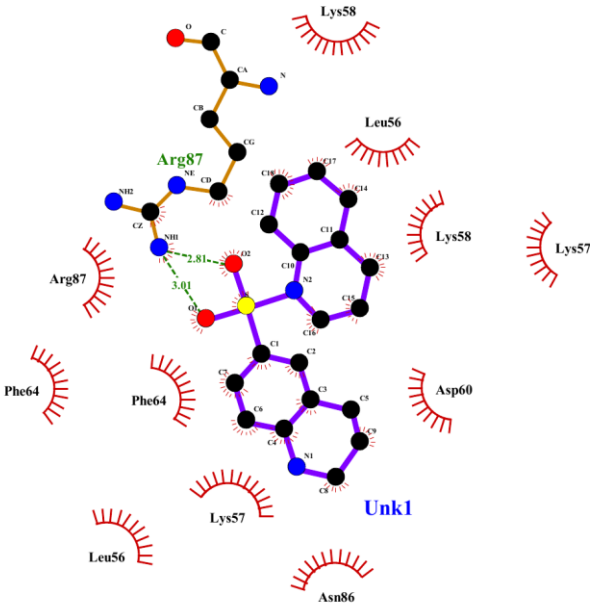
Compound 5b had hydrogen bonds with amino acids SER59, ASP61, ASN86 AND ARG87, whereas hydrophobic interaction with LYS57, LYS58, SER59, ASP60, PHE64 and ARG87 residues (Fig. 4.23A and 4.23B). Compound 5m had only one hydrogen bond with amino acid ARG87, while LEU56, LYS57, LYS58, ASP60, PHE64, ARG86 and ARG87 residues had hydrophobic interaction (Fig. 4.23C and 4.23D). Amino acids LYS57 and ASN86 were connected to compound 5u with hydrogen bonds, while residues LYS58, SER59, PHE64 and ARG87 had hydrophobic interactions (Fig. 4.23E and 4.23F).


Asp60
Arg87
Asn86
Lys57
Arg87
Asp61
Ser59
Phe64
Asn86
Lys58
Unk1
Lys57
Ser59
Asp60

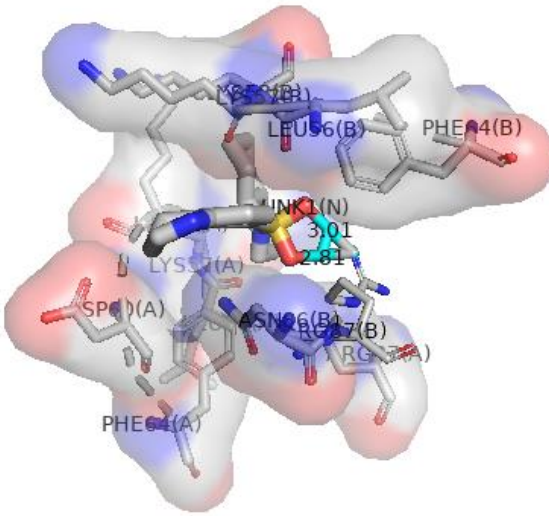
(A)


ASN86(A)
3.09
ARG87(A)
PHE64(B)
UNK1(N)
LYS57(A)
LYS58(A)
SER59(A)
2.86
2.87
ASP60(A)
ARG87(B)
ASN86(B)

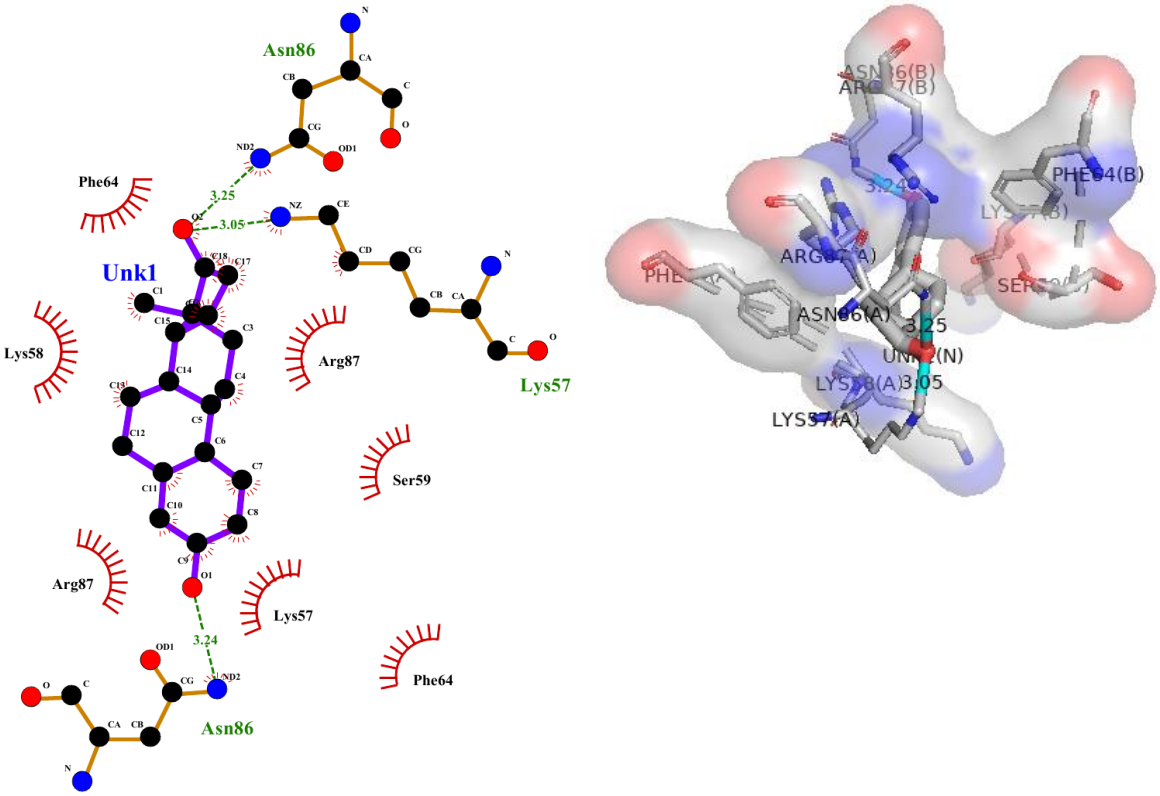
(B)


Lys58
Leu56
Arg87
Lys58
Lys57
Arg87
2.81
3.01
Phe64
Phe64
Asp60
Unk1
Lys57
Leu56
Asn86

(C)


LEU56(B)
PHE64(B)
UNK1(N)
3.01
2.81
PHE64(A)

(D)



(E) **(F)**

Fig. 4.23 Molecular docking interactions between compound and the binding sites of PfThiE (A) 2D model of the interactions between 5b and PfThiE; (B) 3D model of the interactions between 5b and the binding sites of PfThiE; (C) 2D model of the interactions between 5m and PfThiE; (D) 3D model of the interactions between 5m and the binding sites of PfThiE; (E) 2D model of the interactions between 5u and PfThiE; (F) 3D model of the interactions between 5u and the binding sites of PfThiE.

4.7.9 Evaluation of ADMET and pharmacokinetic properties

Twenty-five best docked ligands were further subjected to admetSAR for evaluation of ADMET properties (Table 4.22). Out of 25 compounds, 14 compounds not exhibited toxicity to AMES. Blood Brain Barrier penetration was shown by all except 5b, 5s and 5u. There was no Hepatotoxicity shown in only three compounds. The water solubility of all docked compounds is greater than -3.05. Compounds 5a, 5b, 5m, and 5v demonstrated Caco2 permeability. Intestinal absorption (human) was observed in all compounds. The acute oral toxicity of the maximum compounds was estimated as class III while 5d and 5s were indicated

as II and IV, respectively. Eye corrosion and Eye irritation was not observed in any compounds. Top 25 compounds exhibited dock score values between −8.1 and −11.2 kcal/mol. 14 of the 25 compounds did not exhibit AMES toxicity. Only three compounds (5b, 5 m, and 5u) had all the good ADMET properties. Among the compounds being screened, 5b (ZINC000003953801), 5m (ZINC000001686969), and 5u (ZINC000002036738) had the high binding affinity and good ADMET properties. These compounds may be act as potential inhibitors against thiamine phosphate synthase (PfThiE), an essential enzyme in the thiamine biosynthesis.

A broad understanding of the functional site of a protein is an absolute prerequisite for knowing its mode of action at the molecular level (Innis et al. 2004). The treatment of the parasite could be done effectively if the essential enzymes of this parasite is specifically targeted. In order to fully define a metabolic pathway, it is important to recognise the components of the metabolic pathway (reactants, enzymes, products and reactions) and their relationships. There are several benefits of targeting metabolic pathways on its own.

Similarly, using Docking and *in silico* ADMET analysis, Singh et al. (2013) studied *P. falciparum* SAH hydrolase and identified putative inhibitors against it and Thillainayagam et al. (2018) studied the anti-malarial effects of epoxyazadiradione and its chemical derivatives using a molecular docking strategy against *P. falciparum*. Every step of the pathway is well validated as an important function for the growth of pathogen. We have analysed metabolic pathways by using KAAS, KEGG, MPMP and BioCyc and identifies essential reactions as potential drug targets in the metabolic network of *P. falciparum*.

Table 4.22 ADMET and pharmacokinetic properties of 3 best compounds

Model Name	5b	5m	5u
Ames mutagenesis	-	-	-
Acute Oral Toxicity (c)	III	III	III
Androgen receptor binding	+	+	+
Aromatase binding	+	-	+
Avian toxicity	-	-	-
Blood Brain Barrier	-	+	-
BRCP inhibitior	-	-	+
Biodegradation	-	-	-
BSEP inhibitior	+	-	-
Caco-2	+	+	+
Carcinogenicity (binary)	-	-	-
Carcinogenicity (trinary)	Non-required	Non-required	Danger
crustacea aquatic toxicity	-	-	+
CYP1A2 inhibition	-	-	+
CYP2C19 inhibition	-	+	-
CYP2C9 inhibition	-	-	-
CYP2C9 substrate	-	-	+
CYP2D6 inhibition	-	-	-
CYP2D6 substrate	-	+	+
CYP3A4 inhibition	-	+	-
CYP3A4 substrate	+	+	+
CYP inhibitory promiscuity	-	+	-
Eye corrosion	-	-	-
Eye irritation	-	-	-
Estrogen receptor binding	+	-	+
Fish aquatic toxicity	+	+	+
Glucocorticoid receptor binding	+	-	+
Honey bee toxicity	+	-	+
Hepatotoxicity	-	-	-
Human either-a-go-go inhibition	+	-	+
Human Intestinal Absorption	+	+	+
Human oral bioavailability	-	+	-
MATE1 inhibitior	-	-	-
micronuclear	-	+	-
Acute Oral Toxicity	3.02	3.50	2.26
OATP1B1 inhibitior	+	+	+
OATP1B3 inhibitior	+	+	+
OATP2B1 inhibitior	-	-	-
OCT1 inhibitior	+	-	+
OCT2 inhibitior	-	-	-
P-glycoprotein inhibitior	-	-	-
P-glycoprotein substrate	-	-	-

PPAR gamma	+	-	-
Plasma protein binding	0.84	1.19	0.95
Subcellular localzation	Mitochondria	Lysosomes	Mitochondria
Tetrahymena pyriformis	0.65	2.17	0.73
Thyroid receptor binding	+	+	+
UGT catelyzed	-	-	+
Water solubility	-3.98	-3.70	-4.78

A total of 166 compounds were uniquely found in *P. falciparum* by comparing human and *P. falciparum* compounds using different approaches from KEGG database, while 67 essential metabolic genes were retrieved from MPMP database by navigating map analysis. Chokepoint analysis was performed using BioCyc database for *P. falciparum* 3D7 and a total of 284 and 290 chokepoints reactions were found respectively on the consuming side and the producing side. These enzymes are essential for pathogen survival, since these proteins form an integral part of the reaction that produces or consumes specific substrates that are specific to pathogen and involved in multiple pathways.

The drug resistance in malaria parasites is increasingly emerging, so it is important to discover and develop alternative anti-malarial agents against both new and established drug targets. In this context, the thiamine phosphate synthase (PfThiE) is a possible drug target primarily due to its role and essentialness in the thiamine biosynthesis pathway.

In apicomplexan parasites, thiamine biosynthesis offers a potential and exciting chance to achieve such goals, as the pathway is found only in prokaryotes, fungi, and plants, but is not present in mammals (Wrenger et al. 2006, 2008). Thiamine pyrophosphate (Thi-PP) is the active form of vitamin B1, which is a co-factor for various enzymes primarily involved in the metabolism of carbohydrates such as 2-oxoglutarate dehydrogenase, transketolase or pyruvate dehydrogenase. For a few days, the culturing of *P. falciparum* in a thiamine deficient medium showed no adverse effect but a substantial need for 4-amino-5-hydroxymethyl-2-methylpyrimidine (HMP) or thiamine itself for parasite growth was reported up for more than

ten days (Wrenger et al. 2006). In the thiamine biosynthesis pathway, thiazole THZ-P (5-methyl-4-(beta-hydroxyethyl)thiazole phosphate) and pyrimidine HMP-PP (2-methyl-4-amino-5-hydroxymethylpyrimidine pyrophosphate) moieties are merged to yield thiamine phosphate by PfThiE (Wrenger et al. 2008; Zhang et al. 1997). For several enzymes thiamine is metabolized as an essential cofactor (Chan et al. 2013). So, a novel drug target thiamine phosphate synthase (PfThiE) of *P. falciparum* which is essential enzyme in thiamine biosynthesis was chosen to screen potent anti-malarial drugs.

Thiamine phosphate synthase inhibitors have been studied in many species including *Mycobacterium tuberculosis* (Khare et al. 2011), *Pyrobaculum calidifontis* (Hayashi et al. 2014), *Escherichia coli* (Kawasaki 1979) and *Zea mays* (Rapala-Kozik et al. 2006) but comprehensive research has not been done in *P. falciparum*. PfThiE is an essential enzyme in thiamine biosynthesis (Liu et al. 2018). The thiamine biosynthesis pathway of the parasite has been proposed as a novel and indispensable antimalarial target (Zhang et al. 1997). So, current research attempted to explore the possible effects of nsSNPs in PfThiE and their effects on its structure and function, 3D structure formation and prediction of inhibitors for the modelled structure.

Rapid adaptation to changes in the environment due to the high mutation rate in *P. falciparum* may result in drug resistance to standard medicines (Wrenger et al. 2006). Thus, new drug targets are needed to develop potential inhibitors against the disease. In the present study, the impact of nsSNPs of PfThiE was investigated using various bioinformatics tools. All of the 14 nsSNPs obtained were submitted to functional analysis. SIFT tool predicted that all 14 variants had an effect on protein function. All four tools predicted G165D to affect protein function. 5 nsSNPs (G165D, S330C, G401E, S427W and C456F) by PredictSNP, 2 nsSNPs (A145V and G165D) by PROVEAN and also 2 nsSNPs (G165D and S427W) by SNAP2 considered deleterious. Furthermore, after mutation as a consequence of these nsSNPs, I-Mutant 2.0

displayed a decrease in stability except for D494E, indicating to some degree that the protein is directly or indirectly destabilized. Phylogenetic analysis using ConSurf predicted that only two nsSNPs were found in highly conserved regions. This result shows that the majority of highly conserved residues are stable.

The pharmaceutical industry increasingly utilizes computational techniques to minimize time and financial costs in the drug discovery and development process. In this study, a computational approach was used to systematically evaluate the nsSNPs to predict deleterious mutations and after that 3D model of *P. falciparum* thiamine phosphate synthase was developed using the X-ray crystal structure of *Pyrococcus furiosus* thiamine phosphate pyrophosphorylase as the crystal structure of PfThiE was not available. Various validation methods found the overall structure to be a good model. 156 potential inhibitor compounds were screened by using computational screening techniques. 98 compound followed Lipinski's rule of five and these were chosen for molecular docking studies. Top 25 compounds exhibited dock score values between −8.1 and −11.2 kcal/mol. 14 of the 25 compounds did not exhibit AMES toxicity. Only three compounds (5b, 5 m, and 5u) had all the good ADMET properties. Compounds (5a, 5b, 5e, 5h, 5i, 5 m, 5n, 5o, 5r, 5v, 5w, and 5y) investigated by Khare et al. (2011) as inhibitors against *Mycobacterium tuberculosis* thiamine phosphate synthase had demonstrated strong binding affinity to PfThiE. All the three compounds (5b, 5 m, and 5u) had hydrogen bonds and hydrophobic interaction with amino acids. Further, these compounds have good ADMET properties.

Plasmodium falciparum thiamine phosphate synthase (PfThiE) is an essential enzyme in the thiamine biosynthesis which is not present in humans that can be considered as potential drug target to combat malaria menace. It is one of the few untouched targets for developing anti-malaria drugs. The PfThiE molecular model was developed in the present study by using crystal structure of *Pyrococcus furiosus* thiamine phosphate pyrophosphorylase as a template.

Potential ligands were tried to be identified through docking-based virtual screening with drug-likeness and ADMET analysis. In this analysis, various bioinformatics approaches were also used to examine the effect of non-synonymous SNPs of PfThiE. ConSurf prediction suggests that no mutation was present in the binding site of this protein. Among the compounds being screened, 5b (ZINC000003953801), 5m (ZINC000001686969), and 5u (ZINC000002036738) had the high binding affinity and good ADMET properties. Also, for further confirmation of the protein target and potential ligands, experimental characterization is also required.

In this study, a comprehensive computational approach was used to derive potential therapeutic targets for *P. falciparum* using RNA-seq data set. The differential expression of genes, functional and pathway enrichment analyses of *P. falciparum* has been apprized in detail. The present study results suggested that PF3D7_0705600, PF3D7_1207100, PF3D7_0508100, PF3D7_1126700, and PF3D7_1234300 hub genes might serve as putative targets for drug designing. These hub genes are showing less mutation and no similarity with human proteins. In addition, the gene finding strategies of this study resulted into a database of different identifiers. This developed database tool (www.cdkd.org/pfidmap/) provides easy mapping of different identifiers related to *P. falciparum*. Functional analysis of the nsSNPs of identified hub genes was undertaken to predict deleterious mutations using various computational approaches and a database has been developed to demonstrate the analysis done by PROVEAN, SIFT, PredictSNP and NetSurfP software, which is available online at www.cdkd.org/pfsnp/. Moreover, the effect of deleterious mutations in glycosylphosphatidylinositol transamidase (GPI-T) subunit GPI8p has been investigated, which could be considered as a potential drug target primarily because of its crucial role in the GPI anchor biosynthesis pathway for the development as well as survival of the parasite. Thiamine phosphate synthase (PfThiE), an essential metabolic gene in the thiamine biosynthesis pathway is also studied and potential inhibitors were identified through docking-

based virtual screening along with drug-likeness and ADMET analysis to derive therapeutic targets for *P. falciparum*. Additionally, experimental characterization is also important for further confirmation of the protein target and potential ligands.

SUMMARY

Summary

Malaria is caused by intracellular single-cell parasites belonging to the genus *Plasmodium,* one of the world's most destructive infectious diseases. Malaria appears to be a major concern in developing countries. Notwithstanding attempts to improve malaria-fighting vaccines and medicines, vaccine escape and drug resistance remain a problem. Important advances have been made in the past few years that will greatly lead to malaria prevention, which include the researching of *P. falciparum, A. gambiae* and human genomes, the development of new medicines and candidates for new vaccines, the production and application of combination therapy, periodic preventive care and malaria control at home.

In this study, a comprehensive approach was used to derive potential therapeutic targets for *P. falciparum* using RNA-seq dataset. The differential expression of genes, functional, and pathway enrichment analysis of *P. falciparum* was apprized in detail. It was observed that most DEGs have been linked with important biological processes, many of which are classified as metabolic pathways, secondary metabolite production pathways, ribosome or being involved in biosynthesis of antibiotics. The hub genes viz., PF3D7_0705600, PF3D7_1207100, PF3D7_0508100, PF3D7_1126700, and PF3D7_1234300 identified in this study serves as putative targets for drug designing. These hub genes showed less mutation and no similarity with human proteins and act as lysine methyltransferases, transcription, translation, RNA splicing, and other important cellular pathways in *P. falciparum*. Moreover, this study also provides clusters of hub genes and their network pathways analysis, which can be used further studies to devise a therapeutic target that stabilizes their gene expression. In addition, nsSNPs and their functional impact of these hub genes were also calculated. Hub genes identified in this study may serve as potential targets to develop therapy to suppress the pathogenic action of *P. falciprum* through experimental techniques.

Further, a database has been developed to show the analysis done by PROVEAN, SIFT, PredictSNP and NetSurfP software, which is available online at URL www.cdkd.org/pfsnp/. Database was developed using PHP and JavaScript with user-friendly search environment using a range of options, such as simple searches and advanced searches. This study resulted into a database of different identifiers related to *P. falciparum* from different databases. A database tool is designed for easy mapping of different identifiers related to *P. falciparum*. This database tool can provide mapping of PlasmoDB Gene ID, entrez id, uniprot id, Refseq Protein and string database from either of ids of *P. falciparum*.

GPI-anchor transamidase (GPI-T) is a potential drug target primarily of its crucial role in the development and survival of the parasite in the GPI anchor biosynthesis pathway. The present investigation was undertaken to explore the plausible effects of nsSNP on the structure and functions of GPI-T subunit GPI8p of *P. falciparum*. Of the 34 nsSNPs, 18 nsSNPs were predicted by at least two software as affecting protein function. Hence, solvent accessibility was primarily altered from exposed to buried state and the conformation of secondary structures were changed from coil to helix. Interestingly, ConSurf prediction suggests that most regions of this protein were highly conserved. Additionally, COACH server found no mutation was at predicted ligand binding sites. Inhibition of GPI8p may disrupt the GPI anchor biosynthesis pathways which in turns, prevents GPI anchoring of protein. Therefore, this study provides data on the functional and structural impact of nsSNPs and conservation of amino acid positions in the protein. These findings can be used to develop therapy to suppress the pathogenic action of *P. falciparum*.

Thiamine phosphate synthase (PfThiE) was studied because of its role and essentialness in the thiamine biosynthesis pathway. It is one of the few untouched targets for developing anti-malaria drugs. There are two key parts of this study: sequence and structure. Sequence part used the protein sequence to identify SNP, functional analysis, stability analysis, conservation

analysis, and 3D structure modelling. The PfThiE molecular model was developed in the present study by using crystal structure of *P. furiosus* thiamine phosphate pyrophosphorylase as a template. Potential ligands were tried to be identified through docking-based virtual screening with drug-likeness and ADMET analysis. In this analysis, various bioinformatics approaches were also used to examine the effect of non-synonymous SNPs of PfThiE. ConSurf prediction suggests that no mutation was present in the binding site of this protein. Among the compounds being screened, 5b (ZINC000003953801), 5m (ZINC000001686969), and 5u (ZINC000002036738) had the high binding affinity and good ADMET properties. Also, for further confirmation of the protein target and potential ligands, experimental characterization is also required.

In summary, the research presented in the current study used a comprehensive computational approach to derive potential therapeutic targets for *P. falciparum* using RNA-seq data set. The differential expression of genes, functional and pathway enrichment analyses of *P. falciparum* has been apprized in detail. The present study results suggested that PF3D7_0705600, PF3D7_1207100, PF3D7_0508100, PF3D7_1126700, and PF3D7_1234300 hub genes might serve as putative targets for drug designing. Functional analysis of the nsSNPs of identified hub genes was undertaken to predict deleterious mutations and a database has been developed to demonstrate the analysis done by PROVEAN, SIFT, PredictSNP and NetSurfP software, which is available online at www.cdkd.org/pfsnp/. In addition, the gene finding strategies of this study resulted into a database of different identifiers. This developed database tool (www.cdkd.org/pfidmap/) provides easy mapping of different identifiers related to *P. falciparum*. Moreover, the effect of deleterious mutations in glycosylphosphatidylinositol transamidase (GPI-T) subunit GPI8p has been investigated, which could be considered as a potential drug target primarily because of its crucial role in the GPI anchor biosynthesis pathway for the development as well as survival of the parasite. Thiamine phosphate synthase

(PfThiE), an essential metabolic gene in the thiamine biosynthesis pathway is also studied and potential inhibitors were identified through docking-based virtual screening along with drug-likeness and ADMET analysis to derive therapeutic targets for *P. falciparum*. Therefore, hub genes identified in this research, GPI8p and PfThiE may be considered as potential drug targets for *P. falciparum*.

Future studies

i) The nsSNPs proposed in this study may be further targeted using experimental methods to understand the structural and functional relations of hub mutants.

ii) Developed identifiers mapping tool can encourage the mapping of different identifiers from different databases relevant to all *Plasmodium* species in the future.

iii) To study the genotype investigations, phenotype investigations and pharmacogenetic studies.

iv) Finally, in this study the functional analysis may serve as a worthy model for further exploration of genetically inherited diseases.

REFERENCES

References

Achan J, Talisuna AO, Erhart A, Yeka A, Tibenderana JK, Baliraine FN, Rosenthal PJ, D'Alessandro U (2011) Quinine, an old anti-malarial drug in a modern world: role in the treatment of malaria. Malaria Journal 10: 144

Amoah LE, Acquah FK, Nyarko PB, Cudjoe E, Donu D, Ayanful-Torgby R, Sey F, Williamson KC, Awandare GA (2020) Comparative analysis of asexual and sexual stage *Plasmodium falciparum* development in different red blood cell types. Malaria Journal 19: 200

Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ (2004) UniProt: the universal protein knowledgebase. Nucleic Acids Research 32: D115-9

Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, Ben-Tal N (2016) ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Research 44: W344-W350

Ashley EA, Pyae Phyo A, Woodrow CJ (2018) Malaria. Lancet 391: 1608-1621

Atkinson CT, Aikawa M (1990) Ultrastructure of malaria-infected erythrocytes. Blood Cells 16: 351-368

Aurrecoechea C, Brestelli J, Brunk BP, Dommer J, Fischer S, Gajria B, Gao X, Gingle A, Grant G, Harb OS, Heiges M (2009) PlasmoDB: a functional genomic database for malaria parasites. Nucleic Acids Research 37: D539-43

Balint GA (2001) Artemisinin and its derivatives: an important new class of antimalarial agents. Pharmacology & Therapeutics 90: 261-265

Bannister LH, Hopkins JM, Fowler RE, Krishna S, Mitchell GH (2000) A brief illustrated guide to the ultrastructure of *Plasmodium falciparum* asexual blood stages. Parasitology Today 16: 427-433

Bansal P, Tripathi A, Thakur V, Mohmmed A, Sharma P (2017) Autophagy-related protein ATG18 regulates apicoplast biogenesis in apicomplexan parasites. MBio 8: e01468-17

Barba M, Czosnek H, Hadidi A (2014) Historical perspective, development and applications of next-generation sequencing in plant virology. Viruses 6: 106-136

Bártfai R, Hoeijmakers WA, Salcedo-Amaya AM, Smits AH, Janssen-Megens E, Kaan A, Treeck M, Gilberger TW, Françoijs KJ, Stunnenberg HG (2010) H2A.Z demarcates intergenic regions of the *Plasmodium falciparum* epigenome that are dynamically marked by H3K9ac and H3K4me3. PLOS Pathogens 6: e1001223

Baum J, Gilberger TW, Frischknecht F, Meissner M (2008) Host-cell invasion by malaria parasites: insights from *Plasmodium* and *Toxoplasma*. Trends in Parasitology 24: 557-563

Bendl J, Stourac J, Salanda O, Pavelka A, Wieben ED, Zendulka J, Brezovsky J, Damborsky J (2014) PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. PLOS Computational Biology 10: e1003440

Bernard E, Jacob L, Mairal J, Viara E, Vert JP (2015) A convex formulation for joint RNA isoform detection and quantification from multiple RNA-seq samples. BMC Bioinformatics 16: 1-10

Bharti AR, Patra KP, Chuquiyauri R, Kosek M, Gilman RH, Llanos-Cuentas A, Vinetz JM (2007) Polymerase chain reaction detection of *Plasmodium vivax* and *Plasmodium falciparum* DNA from stored serum samples: implications for retrospective diagnosis of malaria. The American Journal of Tropical Medicine and Hygiene 77: 444-6

Bhatti P, Church DM, Rutter JL, Struewing JP, Sigurdson AJ (2006) Candidate single nucleotide polymorphism selection using publicly available tools: a guide for epidemiologists. American Journal of Epidemiology 164: 794-804

Billker O, Shaw MK, Margos G, Sinden RE (1997) The roles of temperature, pH and mosquito factors as triggers of male and female gametogenesis of *Plasmodium berghei* in vitro. Parasitology 115: 1-7

Breglio KF, Amato R, Eastman R, Lim P, Sa JM, Guha R, Ganesan S, Dorward DW, Klumpp-Thomas C, McKnight C, Fairhurst RM (2018) A single nucleotide polymorphism in the *Plasmodium falciparum* atg18 gene associates with artemisinin resistance and confers enhanced parasite survival under nutrient deprivation. Malaria Journal 17: 391

Boddey JA, Cowman AF (2013) *Plasmodium* nesting: remaking the erythrocyte from the inside out. Annual Review of Microbiology 67: 243-269

Bonday ZQ, Dhanasekaran S, Rangarajan PN, Padmanaban G (2000) Import of host delta-aminolevulinate dehydratase into the malarial parasite: identification of a new drug target. Nature Medicine 6: 898-903

Bunnik EM, Chung DW, Hamilton M, Ponts N, Saraf A, Prudhomme J, Florens L, Le Roch KG (2013) Polysome profiling reveals translational control of gene expression in the human malaria parasite *Plasmodium falciparum*. Genome Biology 14: R128

Capriotti E, Fariselli P, Casadio R (2005) I-Mutant2. 0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Research 33: W306-10

Caspi R, Billington R, Ferrer L, Foerster H, Fulcher CA, Keseler IM, Kothari A, Krummenacker M, Latendresse M, Mueller LA, Ong Q, Paley S, Subhraveti P, Weaver DS, Karp PD (2016) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Research 44: D471-80

CDC (2020) Malaria's Impact Worldwide. CDC website. [Accessed June 19, 2020]. Available at: https://www.cdc.gov/malaria/malaria_worldwide/impact.html.

Chan JA, Fowkes FJ, Beeson JG (2014) Surface antigens of *Plasmodium falciparum* infected erythrocytes as immune targets and malaria vaccine candidates. Cellular and Molecular Life Sciences 71: 3633-57

Chan XW, Wrenger C, Stahl K, Bergmann B, Winterberg M, Müller IB, Saliba KJ (2013) Chemical and genetic validation of thiamine utilization as an antimalarial drug target. Nature Communications 4: 2060

Choi Y, Chan AP (2015) PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics 31: 2745-7

Cloonan N, Forrest AR, Kolle G, Gardiner BB, Faulkner GJ, Brown MK, Taylor DF, Steptoe AL, Wani S, Bethel G, Robertson AJ (2008) Stem cell transcriptome profiling via massive-scale mRNA sequencing. Nature Methods 5: 613-619

Copeland RA, Solomon ME, Richon VM (2009) Protein methyltransferases as a target class for drug discovery. Nature Reviews Drug Discovery 8: 724-32

Costa V, Angelini C, D'Apice L, Mutarelli M, Casamassimi A, Sommese L, Gallo MA, Aprile M, Esposito R, Leone L, Donizetti A (2011) Massive-scale RNA-Seq analysis of non ribosomal transcriptome in human trisomy 21. PLoS One 6: e18493

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology 29: 644-652

Crutcher JM, Hoffman SL. Malaria. In: Baron S, editor (1996) Medical Microbiology. 4th edition. Galveston (TX): University of Texas Medical Branch at Galveston. Chapter 83. Available from: https://www.ncbi.nlm.nih.gov/books/NBK8584/

Daily JP, Le Roch KG, Sarr O, Fang X, Zhou Y, Ndir O, Mboup S, Sultan A, Winzeler EA, Wirth DF (2004) In vivo transcriptional profiling of *Plasmodium falciparum*. Malaria Journal. 3:30

Daina A, Michielin O, Zoete V (2017) SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Scientific Reports 7: 42717

Dallakyan S, Olson AJ (2015) Small-molecule library screening by docking with PyRx. Methods in Molecular Biology 1263: 243-250

Dantzler K. W., Ravel D. B., Brancucci N. M., Marti M. (2015). Ensuring transmission through dynamic host environments: host-pathogen interactions in *Plasmodium* sexual development. Current Opinion in Microbiology 26: 17–23

De Niz M, Meibalan E, Mejia P, Ma S, Brancucci NM, Agop-Nersesian C, Mandt R, Ngotho P, Hughes KR, Waters AP, Huttenhower C (2018) *Plasmodium* gametocytes display

homing and vascular transmigration in the host bone marrow. Science Advances 4: eaat3775

Degarege A, Gebrezgi MT, Beck-Sague CM, Wahlgren M, de Mattos LC, Madhivanan P (2019) Effect of ABO blood group on asymptomatic, uncomplicated and placental *Plasmodium falciparum* infection: systematic review and meta-analysis. BMC Infectious Diseases 19: 86

Dembélé L, Franetich JF, Lorthiois A, Gego A, Zeeman AM, Kocken CH, Le Grand R, Dereuddre-Bosquet N, van Gemert GJ, Sauerwein R, Vaillant JC (2014) Persistence and activation of malaria hypnozoites in long-term primary hepatocyte cultures. Nature Medicine 20: 307-312

Desai M, Chauhan JB (2017) Computational analysis for the determination of deleterious nsSNPs in human MTHFD1 gene. Computational Biology and Chemistry 70: 7-14

Duffy MF, Selvarajah SA, Josling GA, Petter M (2014) Epigenetic regulation of the *Plasmodium falciparum* genome. Briefings in Functional Genomics 13: 203-216

Fatumo S, Plaimas K, Mallm JP, Schramm G, Adebiyi E, Oswald M, Eils R, König R (2009) Estimating novel potential drug targets of *Plasmodium falciparum* by analysing the metabolic network of knock-out strains in silico. Infection, Genetics and Evolution 9: 351–358

Gao J, Zhang C, van Iersel M, Zhang L, Xu D, Schultz N, Pico AR (2014) BridgeDb app: unifying identifier mapping services for Cytoscape. F1000Research 3

Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S, Paulsen IT (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*. Nature 419: 498-511

Gelband H, Panosian CB, Arrow KJ, editors (2004) Saving lives, buying time: economics of malaria drugs in an age of resistance. National Academies Press

Ghosh SK, Rahi M (2019) Malaria elimination in India-The way forward. Journal of Vector Borne Diseases 56: 32-40

Goldberg DE (2013) Complex nature of malaria parasite hemoglobin degradation. Proceedings of the National Academy of Sciences. 110: 5283-5284

Goswami D, Baruah I, Dhiman S, Rabha B, Veer V, Singh L, Sharma DK (2013) Chemotherapy and drug resistance status of malaria parasite in northeast India. Asian Pacific Journal of Tropical Medicine 6: 583-8

Guttery DS, Roques M, Holder AA, Tewari R (2015) Commit and Transmit: Molecular Players in *Plasmodium* Sexual Development and Zygote Differentiation. Trends in Parasitology 31: 676-685

Hannon GJ (2010) FASTX-Toolkit. http://hannonlab.cshl.edu/fastx_toolkit

Hay SI, Okiro EA, Gething PW, Patil AP, Tatem AJ, Guerra CA, Snow RW (2010) Estimating the global clinical burden of *Plasmodium falciparum* malaria in 2007. PLOS Medicine 7: e1000290

Hayashi M, Kobayashi K, Esaki H, Konno H, Akaji K, Tazuya K, Yamada K, Nakabayashi T, Nosaka K (2014) Enzymatic and structural characterization of an archaeal thiamin phosphate synthase. Biochimica et Biophysica Acta 1844: 803-9

Hecht M, Bromberg Y, Rost B (2015) Better prediction of functional effects for sequence variants. BMC Genomics 16: S1

Henry NB, Sermé SS, Siciliano G, Sombié S, Diarra A, Sagnon NF, Traoré AS, Sirima SB, Soulama I, Alano P (2019) Biology of *Plasmodium falciparum* gametocyte sex ratio and implications in malaria parasite transmission. Malaria Journal 18: 70

Hoffman SL, Bancroft WH, Michael G, James SL, Burroughs EC, Stephenson JR, Morgan MJ (1997) Funding for malaria genome sequencing. Nature 387: 647

Holm L, Rosenström P (2010) Dli server: conservation mapping in 3D. Nucleic Acids Research 38: W545-549

Horuk R (2015) The Duffy antigen receptor for chemokines DARC/ACKR1. Frontiers in Immunology 6: 279

Huang DW, Sherman BT, Lempicki RA (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Research 37: 1-3

Huang X, Chen XG, Armbruster PA (2016) Comparative performance of transcriptome assembly methods for non-model organisms. BMC Genomics 17: 523

Innis CA, Anand AP, Sowdhamini R (2004) Prediction of functional sites in proteins using conserved functional group analysis. Journal of molecular biology 337: 1053-68

Joice R, Nilsson SK, Montgomery J, Dankwa S, Egan E, Morahan B, Seydel KB, Bertuccini L, Alano P, Williamson KC, Duraisingh MT (2014) *Plasmodium falciparum* transmission stages accumulate in the human bone marrow. Science Translational Medicine 6: 244re5

Josling GA, Llinás M (2015) Sexual development in *Plasmodium* parasites: knowing when it's time to commit. Nature Reviews Microbiology 13: 573-87

Kafsack BF, Rovira-Graells N, Clark TG, Bancells C, Crowley VM, Campino SG, Williams AE, Drought LG, Kwiatkowski DP, Baker DA, Cortés A (2014) A transcriptional switch underlies commitment to sexual development in malaria parasites. Nature 507: 248-252

Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Research 45: D353-61

Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Research 28: 27-30

Karp PD (2000) An ontology for biological function based on molecular interactions. Bioinformatics 16: 269-85

Karp PD, Paley S, Romero P (2002) The pathway tools software. Bioinformatics 18: S225-32

Kawasaki T (1979) Thiamine phosphate pyrophosphorylase. Methods in Enzymology 62: 69-73

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

Kim TY, Kim HU, Lee SY (2010) Metabolite-centric approaches for the discovery of antibacterials using genome-scale metabolic networks. Metabolic Engineering 12: 105-111

Khare G, Kar R, Tyagi AK (2011) Identification of inhibitors against *Mycobacterium tuberculosis* thiamin phosphate synthase, an important target for the development of anti-TB drugs. PLOS One 6: e22441

Klamt S, Stelling J (2003) Two approaches for metabolic pathway analysis?. Trends in Biotechnology 21: 64-69

Klasberg S, Bitard-Feildel T, Mallet L (2016) Computational Identification of Novel Genes: Current and Future Perspectives. Bioinformatics and Biology Insights 10: 121-131

Kucukkal TG, Petukh M, Li L, Alexov E (2015) Structural and physico-chemical effects of disease and non-disease nsSNPs on proteins. Current Opinion in Structural Biology 32: 18-24

Kuehn A, Pradel G (2010) The coming-out of malaria gametocytes. Journal of Biomedicine and Biotechnology 2010: 976827

Kumar A (2019) Some considerable issues concerning malaria elimination in India. Journal of Vector Borne Diseases 56: 25-31

Kumar P, Henikoff S, Ng PC (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nature Protocols 4: 1073-81

Kwiatkowski DP (2005) How malaria has affected the human genome and what human genetics can teach us about malaria. The American Journal of Human Genetics 77: 171-192

Kyes S, Horrocks P, Newbold C (2001) Antigenic variation at the infected red cell surface in malaria. Annual Review of Microbiology 55: 673-707

Landier J, Parker DM, Thu AM, Carrara VI, Lwin KM, Bonnington CA, Pukrittayakamee S, Delmas G, Nosten FH (2016) The role of early detection and treatment in malaria elimination. Malaria Journal 15: 363

Laskowski RA, Swindells MB (2011) LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. Journal of Chemical Information and Modeling 51: 2778-2786

Laurens MB (2020) RTS,S/AS01 vaccine (Mosquirix™): an overview. Human Vaccines & Immunotherapeutics 16: 480-489

Le Roch KG, Zhou Y, Blair PL, Grainger M, Moch JK, Haynes JD, De La Vega P, Holder AA, Batalov S, Carucci DJ, Winzeler EA (2003) Discovery of gene function by expression profiling of the malaria parasite life cycle. Science 301: 1503-8

Le Roch KG, Chung DW, Ponts N (2012) Genomics and integrated systems biology in *Plasmodium falciparum*: a path to malaria control and eradication. Parasite Immunology 34: 50-60

Lee HJ, Georgiadou A, Otto TD, Levin M, Coin LJ, Conway DJ, Cunnington AJ (2018) Transcriptomic Studies of Malaria: a Paradigm for Investigation of Systemic Host-Pathogen Interactions. Microbiology and Molecular Biology Reviews 82: e00071-17

Lee PH, Shatkay H (2008) Ranking single nucleotide polymorphisms by potential deleterious effects. BMC Bioinformatics 9: 1-3

Lilburn TG, Cai H, Zhou Z, Wang Y (2011) Protease-associated cellular networks in malaria parasite *Plasmodium falciparum*. BMC Genomics 12: S9

Lim L, McFadden GI (2010) The evolution, metabolism and functions of the apicoplast. Philosophical Transactions of the Royal Society B: Biological Sciences 365: 749-763

Liu Q, Zhao Y, Zheng L, Zhu X, Cui L, Cao Y (2018) The glycosylphosphatidylinositol transamidase complex subunit PbGPI16 of *Plasmodium berghei* is important for inducing experimental cerebral malaria. Infection and Immunity 86: e00929-17

Liu W, Li Y, Learn GH, Rudicell RS, Robertson JD, Keele BF, Ndjango JB, Sanz CM, Morgan DB, Locatelli S, Gonder MK (2010) Origin of the human malaria parasite *Plasmodium falciparum* in gorillas. Nature 467: 420-425

Liu X, Wang Y, Liang J, Wang L, Qin N, Zhao Y, Zhao G (2018) In-depth comparative analysis of malaria parasite genomes reveals protein-coding genes linked to human disease in *Plasmodium falciparum* genome. BMC Genomics 19: 312

López-Barragán MJ, Lemieux J, Quiñones M, Williamson KC, Molina-Cruz A, Cui K, Barillas-Mury C, Zhao K, Su XZ (2011) Directional gene expression and antisense transcripts in sexual and asexual stages of *Plasmodium falciparum*. BMC genomics 12: 587

Looker O, Blanch AJ, Liu B, Nunez-Iglesias J, McMillan PJ, Tilley L, Dixon MW (2019) The knob protein KAHRP assembles into a ring-shaped structure that underpins virulence complex assembly. PLOS Pathogens 15: e1007761

Maier AG, Matuschewski K, Zhang M, Rug M (2019) *Plasmodium falciparum*. Trends in Parasitology 35: 481-482

Marchat LA, Arzola-Rodríguez SI, Hernandez-de la Cruz O, Lopez-Rosas I, Lopez-Camarillo C (2015) DEAD/DExH-Box RNA Helicases in Selected Human Parasites. Korean Journal of Parasitology 53: 583-95

Meibalan E, Marti M (2017) Biology of Malaria Transmission. Cold Spring Harbor Perspectives in Medicine 7: a025452

Miller LH, Good MF, Milon G (1994) Malaria pathogenesis. Science 264: 1878-1883

Miller RL, Ikram S, Armelagos GJ, Walker R, Harer WB, Shiff CJ, Baggett D, Carrigan M, Maret SM (1994) Diagnosis of *Plasmodium falciparum* infections in mummies using the rapid manual Para 1994Sight™-F test. Transactions of the Royal Society of Tropical Medicine and Hygiene 88: 31-2

Milner DA Jr (2018) Malaria Pathogenesis. Cold Spring Harbor Perspectives in Medicine 8: a025569

Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M (2007) KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Research 35: W182-5

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature Methods 5: 621-628

Moxon CA, Gibbins MP, McGuinness D, Milner DA Jr, Marti M (2020) New Insights into Malaria Pathogenesis. Annual Review of Pathology: Mechanisms of Disease 15: 315-343

Mu J, Duan J, Makova KD, Joy DA, Huynh CQ, Branch OH, Li WH, Su XZ (2002) Chromosome-wide SNPs reveal an ancient origin for *Plasmodium falciparum*. Nature 418: 323-6

Müller S, Kappes B (2007) Vitamin and cofactor biosynthesis pathways in Plasmodium and other apicomplexan parasites. Trends in Parasitology 23: 112-121

Naing C, Whittaker MA, Htet NH, Aye SN, Mak JW (2019) Efficacy of antimalarial drugs for treatment of uncomplicated *falciparum* malaria in Asian region: A network meta-analysis. PLOS One 14: e0225882

Narwani TJ, Pettrucci AK, Abby S, de Brevern AG (2018) A structural affair of atypical Chemokine Receptor 1 and *Plasmodium vivax*. F1000Research 7

Ngotho P, Soares AB, Hentzschel F, Achcar F, Bertuccini L, Marti M (2019) Revisiting gametocyte biology in malaria parasites. FEMS Microbiology Reviews 43: 401-414

Nixon GL, Moss DM, Shone AE, Lalloo DG, Fisher N, O'Neill PM, Ward SA, Biagini GA (2013) Antimalarial pharmacology and therapeutics of atovaquone. Journal of Antimicrobial Chemotherapy 68: 977-985

NVBDCP (2020) Malaria situation in India from 2016. National Vector Borne Disease Control Programme, Available from: https://nvbdcp.gov.in/WriteReadData/l892s/84555671551594036596.pdf (Accessed on August 13, 2020)

Otto TD, Gilabert A, Crellen T, Böhme U, Arnathau C, Sanders M, Oyola SO, Okouga AP, Boundenga L, Willaume E, Ngoubangoye B (2018) Genomes of all known members of a *Plasmodium* subgenus reveal paths to virulent human malaria. Nature Microbiology 3: 687-97

Otto TD, Wilinski D, Assefa S, Keane TM, Sarry LR, Böhme U, Lemieux J, Barrell B, Pain A, Berriman M, Newbold C (2010) New insights into the blood-stage transcriptome of *Plasmodium falciparum* using RNA-Seq. Molecular Microbiology 76: 12-24

Ozsolak F, Milos PM (2011) RNA sequencing: advances, challenges and opportunities. Nature Reviews Genetics 12: 87-98

Palumbo MC, Colosimo A, Giuliani A, Farina L (2007) Essentiality is an emergent property of metabolic network wiring. FEBS Letters 581: 2485-2489

Perumal D, Lim CS, Sakharkar KR, Sakharkar MK (2009) 'Load Points' and 'Choke Points' as Nodes for Prioritizing Drug Targets in *Pseudomonas aeruginosa*. Current Bioinformatics 4: 48–53

Petersen B, Petersen TN, Andersen P, Nielsen M, Lundegaard C (2009) A generic method for assignment of reliability scores applied to solvent accessibility predictions. BMC Structure Biology 9: 51

Pursell ZF, Isoz I, Lundström EB, Johansson E, Kunkel TA (2007) Yeast DNA polymerase epsilon participates in leading-strand DNA replication. Science 317: 127-30

Prugnolle F, Durand P, Ollomo B, Duval L, Ariey F, Arnathau C, Gonzalez JP, Leroy E, Renaud F (2011) A fresh look at the origin of *Plasmodium falciparum*, the most malignant malaria agent. PLOS Pathogens 7: e1001283

Ramakrishnan G, Chandra N, Srinivasan N (2017) Exploring anti-malarial potential of FDA approved drugs: an in silico approach. Malaria Journal 16: 290

Rapala-Kozik M, Olczak M, Ostrowska K, Starosta A, Kozik A (2007) Molecular characterization of the thi3 gene involved in thiamine biosynthesis in Zea mays: cDNA sequence and enzymatic and structural properties of the recombinant bifunctional protein with 4-amino-5-hydroxymethyl-2-methylpyrimidine (phosphate) kinase and thiamine monophosphate synthase activities. Biochemical Journal 408: 149-159

Rebbeck TR, Spitz M, Wu X (2004) Assessing the function of genetic variants in candidate gene association studies. Nature Reviews Genetics 5: 589-97

Rahman SA, Schomburg D (2006) Observing local and global properties of metabolic pathways: 'load points' and 'choke points' in the metabolic networks. Bioinformatics 22: 1767-1774

Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, Lander ES (2001) Linkage disequilibrium in the human genome. Nature 411: 199-204

Rowe JA, Claessens A, Corrigan RA, Arman M (2009) Adhesion of *Plasmodium falciparum*-infected erythrocytes to human cells: molecular mechanisms and therapeutic implications. Expert Reviews in Molecular Medicine 11: e16

Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL, Hunt SE (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409: 928-34

Schilling CH, Schuster S, Palsson BO, Heinrich R (1999) Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. Biotechnology Progress 15: 296-303

Schlauch KA, Khaiboullina SF, De Meirleir KL, Rawat S, Petereit J, Rizvanov AA, Blatt N, Mijatovic T, Kulick D, Palotás A, Lombardi VC (2016) Genome-wide association analysis identifies genetic variations in subjects with myalgic encephalomyelitis/chronic fatigue syndrome. Translational Psychiatry 6: e730

Schuster S, Fell DA, Dandekar T (2000) A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. Nature Biotechnology 18: 326-332

Sharma VP (2012) Continuing challenge of malaria in India. Current Science 102: 678-82

Sheynkman GM, Shortreed MR, Frey BL, Smith LM (2013) Discovery and mass spectrometric analysis of novel splice-junction peptides using RNA-Seq. Molecular & Cellular Proteomics 12: 2341-53

Sibley CH (2019) Genomic analysis informs malaria evolution. Science 365: 752-753

Siden-Kiamos I, Pace T, Klonizakis A, Nardini M, Garcia CRS, Currà C (2018) Identification of *Plasmodium berghei* Oocyst Rupture Protein 2 (ORP2) domains involved in sporozoite egress from the oocyst. International Journal for Parasitology 48: 1127-36

Siegel TN, Hon CC, Zhang Q, Lopez-Rubio JJ, Scheidig-Benatar C, Martins RM, Sismeiro O, Coppée JY, Scherf A (2014) Strand-specific RNA-Seq reveals widespread and developmentally regulated transcription of natural antisense transcripts in *Plasmodium falciparum*. BMC Genomics 15: 150

Silvestrini F, Alano P, Williams JL (2000) Commitment to the production of male and female gametocytes in the human malaria parasite *Plasmodium falciparum*. Parasitology 121: 465-471

Singh A, Kumaravel J, Mahendru D, Yadav M, Kumar H, Prakash A, Medhi B (2020) Hydroxychloroquine drug safety review for the prophylaxis of SARS COV 2 pandemic. International Journal of Pharmaceutical Sciences and Research. 2020.

Singh DB, Gupta MK, Singh DV, Singh SK, Misra K (2013) Docking and in silico ADMET studies of noraristeromycin, curcumin and its derivatives with *Plasmodium falciparum* SAH hydrolase: a molecular drug target against malaria. Interdisciplinary Sciences: Computational Life Sciences 5: 1-2

Singh S, Malik BK, Sharma DK (2007) Choke point analysis of metabolic pathways in *E. histolytica*: a computational approach for drug target identification. Bioinformation 2: 68-72

Sinha A, Hughes KR, Modrzynska KK, Otto TD, Pfander C, Dickens NJ, Religa AA, Bushell E, Graham AL, Cameron R, Kafsack BF (2014) A cascade of DNA-binding proteins for sexual commitment and development in *Plasmodium*. Nature 507: 253-257

Smalley ME, Abdalla S, Brown J (1981) The distribution of *Plasmodium falciparum* in the peripheral blood and bone marrow of Gambian children. Trans R Soc Trop Med Hyg 75: 103-105

Smith TG, Lourenço P, Carter R, Walliker D, Ranford-Cartwright LC (2000) Commitment to sexual differentiation in the human malaria parasite, *Plasmodium falciparum*. Parasitology 121: 127-133

Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T (2011) Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics 27: 431-2

Solayman M, Saleh MA, Paul S, Khalil MI, Gan SH (2017) In silico analysis of nonsynonymous single nucleotide polymorphisms of the human adiponectin receptor 2 (ADIPOR2) gene. Computational Biology and Chemistry 68:175-185

Sologub L, Kuehn A, Kern S, Przyborski J, Schillig R, Pradel G (2011) Malaria proteases mediate inside-out egress of gametocytes from red blood cells following parasite transmission to the mosquito. Cellular Microbiology13: 897-912

Sorber K, Dimon MT, DeRisi JL (2011) RNA-Seq analysis of splicing in *Plasmodium falciparum* uncovers new splice junctions, alternative splicing and splicing of antisense transcripts. Nucleic Acids Research 39: 3820-3835

Sterling T, Irwin JJ (2015) ZINC 15 - Ligand Discovery for Everyone. Journal of Chemical Information and Modeling 55: 2324-2337

Subudhi AK, Boopathi PA, Pandey I, Kaur R, Middha S, Acharya J, Kochar SK, Kochar DK, Das A (2015) Disease specific modules and hub genes for intervention strategies: A co-expression network based approach for *Plasmodium falciparum* clinical isolates. Infection, Genetics and Evolution 35: 96-108

Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, Jensen LJ (2017) The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. Nucleic Acids Research 45: D362-D368

Tangpukdee N, Duangdee C, Wilairatana P, Krudsood S (2009) Malaria diagnosis: a brief review. Korean Journal of Parasitology 47: 93-102

Taylor CM, Wang Q, Rosa BA, Huang SC, Powell K, Schedl T, Pearce EJ, Abubucker S, Mitreva M (2013) Discovery of anthelmintic drug targets and drugs using chokepoints in nematode metabolic pathways. PLOS Pathogens 9: e1003505

Thillainayagam M, Malathi K, Anbarasu A, Singh H, Bahadur R, Ramaiah S (2018) Insights on inhibition of Plasmodium falciparum plasmepsin I by novel epoxyazadiradione derivatives–molecular docking and comparative molecular field analysis. Journal of Biomolecular Structure and Dynamics. doi: 10.1080/07391102.2018.1510342

Tibúrcio M, Silvestrini F, Bertuccini L, Sander AF, Turner L, Lavstsen T, Alano P (2013) Early gametocytes of the malaria parasite *Plasmodium falciparum* specifically remodel the adhesive properties of infected erythrocyte surface. Cellular Microbiology15: 647-659

Trapnell C, Pachter L, Salzberg SL (2009) TopHat: Discovering splice junctions with RNA-Seq. Bioinformatics 25: 1105-1111

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature Protocols 7: 562-78

Treister A, Pico AR (2018) Identifier Mapping in Cytoscape. F1000Research 7

Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. Journal of Computational Chemistry 31: 455-461

Tuteja R (2007) Malaria - an overview. FEBS Journal 274: 4670-4679

Tyagi V, Dhiman S, Sharma AK, Srivastava AR, Rabha B, Sukumaran D, Veer V (2017) Morphometric and morphological appraisal of the eggs of *Anopheles stephensi* (Diptera: Culicidae) from India. Journal of Vector Borne Diseases 54: 151-156

Tyagi R, Elfawal MA, Wildman SA, Helander J, Bulman CA, Sakanari J, Rosa BA, Brindley PJ, Janetka JW, Aroian RV, Mitreva M (2019) Identification of small molecule enzyme inhibitors as broad-spectrum anthelmintics. Scientific Reports 9: 9085

Umeda T, Tanaka N, Kusakabe Y, Nakanishi M, Kitade Y, Nakamura KT (2011) Molecular basis of fosmidomycin's action on the human malaria parasite *Plasmodium falciparum*. Scientific Reports 1:9

UniProt Consortium (2018) UniProt: the universal protein knowledgebase. Nucleic Acids Research 46: 2699

van Iersel MP, Pico AR, Kelder T, Gao J, Ho I, Hanspers K, Conklin BR, Evelo CT (2010) The BridgeDb framework: standardized access to gene, protein and metabolite identifier mapping services. BMC Bioinformatics 11: 1-7

Venugopal K, Hentzschel F, Valkiūnas G, Marti M (2020) *Plasmodium* asexual growth and sexual development in the haematopoietic niche of the host. Nature Reviews Microbiology 18: 177-189

Volkman SK, Hartl DL, Wirth DF, Nielsen KM, Choi M, Batalov S, Zhou Y, Plouffe D, Le Roch KG, Abagyan R, Winzeler EA (2002) Excess polymorphisms in genes for membrane proteins in *Plasmodium falciparum*. Science 298: 216-8

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

Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TA, Rempfer C, Bordoli L, Lepore R (2018) SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Research 46: W296-W303

WHO (2009) Global antimalarial drug policies database - WHO African region. September edn.

WHO (2015) Guidelines For The Treatment of Malaria - 3rd edition. Available at https://apps.who.int/iris/bitstream/handle/10665/162441/9789241549127_eng.pdf?sequence=1

WHO (2018a) Model List of Essential Medicines. Available at https://www.who.int/medicines/publications/essentialmedicines/en/

WHO (2018b) Antimalarial drug efficacy and drug resistance. Available at https://www.who.int/malaria/areas/treatment/drug_efficacy/en/

Wrenger C, Eschbach ML, Müller IB, Laun NP, Begley TP, Walter RD (2006) Vitamin B1 de novo synthesis in the human malaria parasite *Plasmodium falciparum* depends on external provision of 4-amino-5-hydroxymethyl-2-methylpyrimidine. Journal of Biological Chemistry 387: 41-51

Wrenger C, Knöckel J, Walter RD, Müller IB (2008) Vitamin B1 and B6 in the malaria parasite: requisite or dispensable? Brazilian Journal of Medical and Biological Research 41: 82-88

Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, Kong L, Gao G, Li CY, Wei L (2011) KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. Nucleic Acids Research 39: W316-22

Yang H, Lou C, Sun L, Li J, Cai Y, Wang Z, Li W, Liu G, Tang Y (2019) admetSAR 2.0: web-service for prediction and optimization of chemical ADMET properties. Bioinformatics 35: 1067-1069

Yates CM, Sternberg MJ (2013) The effects of non-synonymous single nucleotide polymorphisms (nsSNPs) on protein-protein interactions. Journal of Molecular Biology 425: 3949-3963

Yeh I, Hanekamp T, Tsoka S, Karp PD, Altman RB (2004) Computational analysis of *Plasmodium falciparum* metabolism: organizing genomic information to facilitate drug discovery. Genome Research 14: 917-924

Zhang Y, Taylor SV, Chiu HJ, Begley TP (1997) Characterization of the *Bacillus subtilis* thiC operon involved in thiamine biosynthesis. Journal of Bacteriology 179: 3030-3035

Appendix I

Table A1: Statistical analysis of differentially expressed genes between Ring (R) & other stages

Here R: Ring; ET: Early trophozoite; LT: Late trophozoite; Sc: Schizont; GII: Gametocyte stages II; GV: Gametocyte stages V; Oo: Ookinete; log2(FC): log2 (Fold change); P: p-value (<0.05). Blank space: log2(FC) have no values from fold change.

	R-ET		R-LT		R-Sc		R-GII		R-GV		R-Oo	
locus	**log2(FC)**	**P**	**log2(FC)**	**P**	**log2(FC)**	**P**	**log2(FC)**	**P**	**log2(FC)**	**P**	**log2(FC)**	**P**
1:29732-37349									-6.56716	0.02385		
1:99051-102515	-5.28737	0.01415										
1:178321-182377					7.33368	0.0081						
1:281369-285674							-4.61461	0.03765				
1:290601-292806											3.80787	0.03965
1:363139-363739									6.36462	0.03075		
1:364081-366268			4.1944	0.0377								
1:384133-393243							-17.1431	0.0225	-14.4152	0.01505		
1:465875-473342	3.78924	0.01045										
1:478427-482531											4.25199	0.0489
1:522528-524301							-5.21059	0.04625	-5.02934	0.02725		
1:528828-538073	-3.22579	0.01465	-5.9639	0.0105								
1:119274-121483	3.32774	0.02										
1:131164-134362					4.67426	0.03035						
1:147503-153355	2.89408	0.0486										
1:183056-184457	3.9224	0.0211			4.2581	0.04955						
1:225347-226017					5.75583	0.04965						
1:338254-348100			20.0826	0.02715								
1:372817-377288											4.32515	0.03245
1:489040-491276					-4.50396	0.0253						
1:502107-504017	4.70801	0.00655										
1:506086-508023							-6.8233	0.0365			-6.29688	0.0321
1:609109-616613	-3.42616	0.01945							-6.37728	0.0308		
10:28490-36165									-4.84333	0.0372		
10:63152-64033	3.22231	0.0243										

10:112929-116016	3.82289	0.00625										
10:228179-233269					-3.94527	0.02725						
10:253703-255611											5.36657	0.02675
10:372518-374072									4.30064	0.04875		
10:387026-388441	3.36052	0.04275										
10:510051-515097	4.13508	0.00255										
10:633643-636158			4.80604	0.03115								
10:1030041-1031532											6.1525	0.0025
10:1137999-1158978					-22.4659	0.03185						
10:1181002-1182499			4.10525	0.0476								
10:1217345-1222439					5.59795	0.0162						
10:1257381-1258476					5.14647	0.03275						
10:1335219-1336287	4.88988	0.0295										
10:1339759-1340626									-6.36304	0.0198		
10:1391441-1393127					4.82141	0.0438						
10:1394835-1396593											4.44121	0.0435
10:1399191-1402893					5.42958	0.01945						
10:1407272-1408388					4.41526	0.0463						
10:1413196-1415290					5.4539	0.0232						
10:1420529-1422668	3.39376	0.01465										
10:1423979-1425680					5.49856	0.03445						
10:1427123-1428341			4.07339	0.05005	6.22612	0.02085						
10:1557814-1560695	-2.851	0.04									-4.79302	0.04715
10:1568707-1571814					-4.92986	0.04685	-5.69828	0.03275	-5.87863	0.01495	-5.22758	0.03325
10:1639169-1640517	3.31889	0.04645										
10:51349-52600	4.55935	0.0064										
10:74661-77573											-5.91315	0.0265
10:80758-82065			-5.16613	0.0367								
10:86538-89010			5.66897	0.0126	5.00146	0.0331						
10:96113-97032	-3.37256	0.0423										
10:98432-100393	-3.33004	0.0204										
10:106653-108663					-3.99211	0.0449						
10:149293-151081											-4.00932	0.04485

10:206132-210044	3.21236	0.02865										
10:220018-221354									4.7365	0.03865		
10:256116-257699											4.8338	0.03525
10:261014-267029									6.42319	0.0146		
10:270297-276099									7.56978	0.04285		
10:288334-290826			4.74317	0.0113								
10:301125-305919	4.27829	0.03825										
10:360624-362480			5.60029	0.0271								
10:389502-391164					4.50793	0.0469						
10:405015-410535					-4.53603	0.0352						
10:468022-469535					4.40478	0.03955						
10:482525-484193					-4.49257	0.0489			-6.12257	0.01685	-4.78963	0.03015
10:627042-629039	4.50703	0.002										
10:656707-659026									-5.12017	0.0212	-5.83089	0.03335
10:669200-671852									-4.91781	0.02755		
10:680768-683513			4.92984	0.03495	3.16477	0.04465						
10:688955-690819					5.16083	0.0479						
10:694838-696708											4.15113	0.029
10:808616-808991									5.68171	0.0141		
10:1137999-1158978									-4.24405	0.0233		
10:1176043-1177847			4.41993	0.0361								
10:1211655-1212969					5.80371	0.0337						
10:306382-308657							5.06856	0.04465	5.62204	0.0203		
11:24159-31598	-3.3785	0.0215							-5.58366	0.0346		
11:224307-231015			4.16085	0.0247								
11:247450-248971									-4.35924	0.0439		
11:289666-290286									5.07423	0.0333		
11:356082-360434											4.17909	0.04825
11:527570-528701									-4.85775	0.0311		
11:652577-653975					4.92756	0.04965						
11:744872-745931									4.91415	0.0286		
11:749150-754400							-5.31636	0.0308				
11:938948-939611											5.37155	0.00795

11:1026462-1029719	2.72291	0.0301										
11:1062216-1070996			3.56975	0.03545					4.61719	0.02285		
11:1156434-1158264	2.75151	0.0343										
11:1270321-1272955			5.14183	0.0327								
11:1293853-1295722					4.72983	0.0376						
11:1302170-1307639			-5.18458	0.01665								
11:1367183-1368807											-5.09758	0.03625
11:1441499-1445468					-5.10193	0.02935	-4.57114	0.0381				
11:1446485-1448502											4.80451	0.0396
11:1461283-1461509	3.63659	0.02965			5.82906	0.0193						
11:1540800-1544160							5.49871	0.02225				
11:1548182-1553558					-4.44034	0.0422	-7.58584	0.03565				
11:1608002-1608606					6.64562	0.03305					4.58916	0.0391
11:1774746-1776723											4.41306	0.04505
11:1941344-1942585	-4.52639	0.015							-5.18515	0.04325		
11:1976403-1979801	-4.62848	0.0075					-5.98183	0.0218				
11:2001064-2003310					-6.51968	0.0344	-6.97825	0.02425				
11:2025813-2035883	-4.11683	0.0079							-6.5444	0.04105		
11:32665-42386	-3.09597	0.03045							-5.54907	0.03295		
11:144841-144940	4.80671	0.0092					5.04282	0.0485				
11:182383-187224							-4.11219	0.0459				
11:190800-194359					4.53042	0.0368						
11:328903-335932	2.98965	0.01935										
11:529092-533977			16.6426	0.0142								
11:581248-582697			-4.04534	0.04865								
11:592128-593583									-4.9003	0.029		
11:602981-606917			5.56075	0.02805								
11:738511-741670									-5.28913	0.02615		
11:1008407-1010852					4.62216	0.04285						
11:1040787-1046085					4.72002	0.0491					5.0326	0.02945
11:1195985-1198307	3.31223	0.0245									4.53302	0.041
11:1210683-1211454											6.42949	0.02615
11:1225480-1226350									4.8316	0.048		

11:1227217-1231948	3.74185	0.00965										
11:1359361-1364111					-5.07356	0.0308						
11:1417463-1420068					5.72752	0.0234					5.26736	0.0272
11:1463807-1465055											4.73724	0.0349
11:1499625-1500450							5.03359	0.04965				
11:1600415-1603449	3.20515	0.02745			3.60343	0.03205						
11:1672515-1674554					5.15631	0.0463						
11:1717202-1727319					4.09985	0.02065						
11:1759253-1759919	4.2172	0.01265										
11:1813345-1817807									-25.0548	0.03445		
11:1843641-1853369											4.40974	0.03725
11:1985866-1986636	4.21554	0.00805										
12:16972-24497	-3.24986	0.0241										
12:32702-41940	-3.86005	0.0237										
12:89933-90749	-3.52934	0.01645							-6.12655	0.0295		
12:409847-422459	2.58482	0.0448										
12:520133-525393			5.57563	0.01515								
12:576389-576641	6.85365	0.0257	5.47843	0.03255			6.19599	0.0407				
12:602420-606955					4.72449	0.04115						
12:618490-620235					3.92692	0.04635						
12:621117-621768							6.15184	0.04015				
12:678594-681954					-4.64453	0.03995						
12:745192-756345	2.4121	0.0367										
12:918713-919328			5.13309	0.02085	7.0712	0.0471	4.72576	0.0397				
12:1148101-1151452			4.30729	0.04245	4.78996	0.0363						
12:1199700-1201287			-4.22784	0.0427					-5.13437	0.02545		
12:1267357-1270156					-4.76651	0.0372	-5.34907	0.0278				
12:1458220-1459170			4.57095	0.0315	6.47023	0.01605						
12:1484293-1485899			5.67325	0.01455	6.54386	0.02595						
12:1487071-1488679			4.51553	0.0368	4.83368	0.03745						
12:1589157-1592061							-4.61053	0.0408				
12:1610351-1612730							-4.4869	0.0483				
12:1634462-1643293	2.43326	0.03885	4.12064	0.03395								

12:1655475-1658118	3.92756	0.01875										
12:1694138-1703076									-4.57306	0.04035		
12:1704500-1712479			-5.60795	0.0471					-5.18408	0.0229		
12:1719562-1727445	-3.9245	0.00515	-5.76781	0.0108	-7.92113	0.0496						
12:1734909-1745087			-5.1805	0.00495	-8.21336	0.0088	-8.36835	0.00675	-8.68537	0.00475	-7.73407	0.0041
12:1798630-1799776											5.20919	0.03355
12:1857765-1858074							5.05626	0.0213				
12:1920770-1921901			5.63712	0.02305								
12:1930316-1933509					4.32854	0.02815						
12:1968149-1971061									4.22973	0.0467		
12:2092072-2094217					4.45867	0.0381						
12:2168165-2169744	-5.84292	0.00315									-5.58292	0.02155
12:46787-56805	-3.61071	0.0296	-5.87569	0.03775			-4.82686	0.0353				
12:60874-64888	-4.58445	0.02015	-5.83018	0.00465								
12:76845-79067	4.59734	0.0309							-5.61307	0.0269		
12:81448-84451	4.25115	0.01185										
12:195627-206549	-5.85728	0.0246	-6.11153	0.0172	-7.32416	0.01865						
12:249332-251396	3.36952	0.02										
12:301125-302823							-4.61712	0.04145			-4.15477	0.0494
12:351844-354001									4.55316	0.04725		
12:426386-433030	2.70037	0.0411										
12:590186-592427											-4.45822	0.0379
12:629001-629660	3.2686	0.0228										
12:658805-661695											7.98458	0.0363
12:675448-677569									4.04878	0.02435		
12:690815-699042									-90.7838	0.03995		
12:725833-729830							-4.63003	0.0428			-4.22985	0.04455
12:732166-734952									-4.58788	0.0434		
12:766646-774190									-4.8834	0.0262		
12:871716-876693			-4.51745	0.0198								
12:927816-929983			5.40427	0.04945	7.24082	0.0258						
12:972701-973385	3.40673	0.0434										
12:1134360-1137639									-4.95369	0.02825		

12:1160719-1162951			5.84108	0.02255								
12:1251903-1253362					4.19452	0.0307						
12:1328301-1331362	3.71159	0.02615										
12:1340169-1343303			4.11429	0.04745	6.37391	0.01555						
12:1405169-1405956							4.34955	0.0491				
12:1464205-1467559					5.73852	0.01685						
12:1577147-1585147			-224.403	0.03515			-39.2883	0.0425	-94.0218	0.02505		
12:1981719-1987114							38.4786	0.04805				
12:2095047-2096494							6.87706	0.0497	4.84766	0.03745		
12:2241254-2248946							-4.70499	0.04365	-6.6667	0.02825	-5.84956	0.0382
12:79190-79255	3.52317	0.03925										
12:303391-307170											-5.0109	0.0467
13:21466-28890									-6.09991	0.0433		
13:82239-84779	3.85712	0.0463										
13:98985-101347			-4.32194	0.0465								
13:102944-110979			-4.35434	0.0382								
13:407669-412532							-13.1003	0.0488				
13:449108-449870									6.79669	0.02095		
13:508101-509937											-3.64162	0.0479
13:644710-647258			4.0824	0.04825	5.15015	0.04715						
13:712145-715163			5.1793	0.0485	5.18074	0.02445						
13:720760-735286	4.02281	0.0168										
13:780134-781880											5.2992	0.0265
13:829734-833511							-5.4429	0.0303				
13:916560-918795					4.97876	0.0494						
13:959960-965521									5.05329	0.02955		
13:982811-984002							4.49059	0.0448				
13:1050146-1052401									3.76564	0.0487		
13:1332496-1335790					5.93054	0.0345						
13:1652234-1660202							-13.9704	0.03685				
13:1673729-1676570					5.00748	0.0233						
13:1679723-1681475	2.70819	0.04175										
13:1904460-1920908	-2.8304	0.02585										

13:2055946-2060255											4.60693	0.0373
13:2061669-2063690									-4.38934	0.0236		
13:2097817-2105419											-4.76499	0.0326
13:2126863-2127398			-5.36098	0.03515								
13:2228215-2229581					7.47235	0.0414	5.69867	0.02855				
13:2368214-2374292	4.41332	0.00765										
13:2458401-2460117									6.29226	0.02825		
13:2470204-2477959			5.15063	0.04395								
13:2483086-2483686							5.86474	0.0285				
13:2564890-2567281			4.30433	0.04305								
13:2683709-2688216					-5.37183	0.023						
13:2719023-2722844	2.92409	0.0425										
13:2766249-2769814							-5.23175	0.0387				
13:2784305-2785247			-5.06907	0.0383								
13:2828795-2830236	4.12598	0.0099										
13:2853800-2855166	-3.77288	0.0349										
13:34061-44845	-3.43791	0.01685							-5.76078	0.0369		
13:53400-55555							4.70746	0.0219				
13:112894-113918							-4.94727	0.03555			-4.826	0.0342
13:120517-123451	5.409	0.0133										
13:468433-468865			4.61622	0.0346	5.8948	0.02595	5.52621	0.02305				
13:470796-475204									-4.66991	0.0362		
13:626682-628877	3.54274	0.03985										
13:668561-670947									5.93493	0.01795		
13:770846-773096											4.76275	0.03355
13:890306-894983			4.1312	0.0487	5.09553	0.02965						
13:1092594-1094388									-5.25636	0.0224		
13:1138758-1141762									5.46807	0.04775		
13:1204386-1207401	3.20308	0.0501										
13:1218117-1225736							-4.22361	0.0356				
13:1227572-1234619	3.35415	0.0118										
13:1239023-1242259			4.09293	0.04395								
13:1401094-1402474	4.35146	0.01105										

13:1405928-1408127	3.19605	0.02815										
13:1409246-1410143					4.71455	0.0434						
13:1411297-1411906	-3.73777	0.0265					-4.78826	0.04265	-7.26454	0.03405		
13:1413299-1414145					5.64746	0.03495						
13:1419283-1420339			4.96804	0.02295	6.28619	0.0125						
13:1506976-1521323									243.665	0.04095		
13:1547205-1553502			23.432	0.03925								
13:1561449-1563318											5.1438	0.02885
13:1586000-1586594							4.71525	0.04475				
13:1596935-1598332									6.44923	0.0275		
13:1632632-1632740	6.36715	0.0078	6.06757	0.02405								
13:1646360-1647749											5.54123	0.02935
13:1773468-1778894					4.50234	0.0137						
13:1784157-1785207									5.04747	0.0289		
13:1968220-1970812							-4.62591	0.0401				
13:1998759-2001537											4.8436	0.0159
13:2109927-2113138	-6.34314	0.0059	-6.60911	0.03975								
13:2114693-2117246							-4.61729	0.04655				
13:2262149-2263481											-5.33824	0.0359
13:2436086-2439836			4.81331	0.03065	4.70266	0.01645						
13:2478804-2479629			5.02497	0.02665								
13:2484189-2485808											4.41135	0.05005
13:2486718-2488170									-5.00373	0.0254		
13:2650268-2659502			-26.8865	0.02025	-23.724	0.02715	-43.1608	0.03475	-46.8331	0.0191		
13:2702832-2708170					3.92115	0.02645						
13:2817720-2818697			-4.50808	0.0397								
13:2826753-2827778	4.35736	0.0127										
13:2856995-2864550									-6.40173	0.02	-4.73659	0.03445
13:1779170-1779994									4.62184	0.0498		
13:2363749-2366367	4.48944	0.01085										
13:2783498-2784116									-5.60451	0.0265		
14:46833-47745	5.26939	0.0048										
14:83980-89183									15.6164	0.04515		

14:100846-105624	2.76394	0.03125										
14:181954-185032											5.398	0.0487
14:279175-280673											5.34893	0.02345
14:293470-294832	3.13778	0.02095										
14:297467-298823											-4.16041	0.04885
14:381998-385430	4.47119	0.0038										
14:420441-422790			5.7636	0.03265								
14:575494-577054			4.61485	0.03125	5.40038	0.0264						
14:793750-796876							-4.90294	0.0416				
14:922875-924153											4.76384	0.0427
14:936112-941583											4.41613	0.0384
14:942136-946935									4.15567	0.0472		
14:1056103-1060066					-5.56411	0.02355						
14:1091870-1094945	3.01669	0.0467										
14:1118504-1126431	-2.96296	0.04685										
14:1184851-1186984							6.66462	0.03085	5.21233	0.024	5.10535	0.04245
14:1263767-1268953					5.30769	0.02165						
14:1419325-1420877									5.57083	0.02425		
14:1433156-1450802									246.91	0.0414		
14:1458073-1459813									-4.32368	0.04535		
14:1665405-1667486	3.28218	0.02335										
14:1690935-1692456	3.32026	0.0173										
14:2008373-2012437					4.81454	0.0404						
14:2014482-2015663			5.92983	0.0197								
14:2016900-2019503							-4.12332	0.04575			-4.59724	0.0291
14:2123396-2126535					4.21014	0.04715						
14:2134233-2140803					5.9344	0.01295						
14:2167154-2169476	4.2283	0.00785										
14:2259400-2261233					4.69061	0.0417						
14:2406603-2410095											-4.72629	0.04255
14:2619613-2624122							-4.77945	0.0383				
14:2642331-2648098					-3.57807	0.0495						
14:2809413-2812002					4.67485	0.0469						

14:2929759-2931484									5.38543	0.0178		
14:2981536-2982901	3.96165	0.01815			4.34736	0.0489						
14:2989513-2992159											5.02845	0.02585
14:3007520-3018777					-3.79575	0.0408						
14:3137015-3141618	-3.28857	0.0317			3.77671	0.03525						
14:3156910-3157972									6.78737	0.0235		
14:3188822-3190483							-4.49745	0.0494	-5.12474	0.02685	-4.54944	0.04705
14:3213177-3216279	-4.2377	0.01535										
14:3232573-3236470	3.66873	0.0279										
14:3266070-3267312	3.57356	0.0384										
14:40519-43734			-5.03235	0.01575								
14:74187-76793									5.28787	0.03165		
14:120068-121134							5.20526	0.0466	5.35633	0.02465	5.73746	0.0471
14:163989-166884					-6.80727	0.0187						
14:200492-204143					-5.97593	0.03655						
14:246379-247721			5.98976	0.035	7.43211	0.0296	6.26861	0.02045			4.52218	0.04155
14:300724-305833					-4.30286	0.0445	-4.33232	0.0451				
14:323252-346701							-11.5183	0.02645				
14:365158-368180			5.03365	0.0299								
14:726977-728378									-4.70773	0.0334		
14:782103-782304	3.93372	0.00875			4.62236	0.04985						
14:783076-785818					-5.35443	0.05055						
14:867747-868866									6.12851	0.0268		
14:953566-965858							-63.4677	0.04045	-73.5343	0.0203		
14:1049991-1050258							4.51312	0.04145				
14:1144459-1148020											5.57483	0.04385
14:1150416-1153329							6.38044	0.02855			5.60651	0.0337
14:1216594-1217113							5.96778	0.0318				
14:1238792-1241537											4.28581	0.04235
14:1289687-1291668			4.10846	0.0382								
14:1421843-1431164					4.4444	0.0322						
14:1490596-1494157					5.03773	0.03035						
14:1494720-1499763			3.3029	0.0416								

14:1600875-1601943			4.71688	0.03055	5.1632	0.03375						
14:1851598-1854997							-5.39749	0.0305			-4.57083	0.04165
14:1959975-1963905					-4.97434	0.04715	-4.61348	0.04565				
14:1983493-1995517					3.14705	0.04535						
14:1998916-2006305									6.02589	0.027		
14:2020151-2024080											5.74536	0.03155
14:2075522-2077286											-4.8223	0.0481
14:2245035-2245500											5.94545	0.0225
14:2531366-2541335					-4.68126	0.0373						
14:2553648-2555039			4.27927	0.0432	4.85106	0.03335						
14:2630492-2632613									-4.74133	0.03175		
14:2658189-2660846									7.17805	0.0397		
14:2686159-2688239									5.79653	0.0149		
14:2724949-2728122							3.99195	0.03795				
14:2889614-2896193			3.5414	0.05005								
14:2905100-2906503	3.92444	0.0283										
14:2918140-2920925			4.26242	0.0358								
14:2921651-2924494							4.64958	0.01505				
14:927131-929353									5.53763	0.0222		
2:25231-31168			-5.35993	0.0495					-6.85544	0.0318		
2:120523-124102	2.62448	0.03845							-4.59477	0.0339		
2:533222-555294	2.88196	0.0465										
2:566120-567894									-4.18471	0.0483		
2:594195-602101	-2.97522	0.01825										
2:605044-609325									-3.98389	0.05035		
2:610624-615372					5.4459	0.02695						
2:626975-627710											4.72436	0.04265
2:645928-650085							-4.57856	0.04445				
2:720407-722799			4.70232	0.0343	6.31612	0.0065					3.46017	0.04435
2:738858-739851											5.32427	0.0224
2:741012-742458									5.73583	0.0292		
2:755416-762559	3.3176	0.01055										
2:804517-807739					-5.0825	0.04745	-4.62854	0.0492	-4.75736	0.03025		

2:833802-834686									-6.09078	0.0309	-4.57795	0.043
2:838840-844114					4.6717	0.04165						
2:91317-98838	-3.66062	0.01485										
2:103384-105796	-5.12848	0.00525	-7.03441	0.00865								
2:171562-174912									4.94616	0.0286		
2:277493-278456			5.21803	0.0177								
2:294272-297616			5.4036	0.01685	5.10192	0.0285						
2:298896-302564			5.5387	0.03765	5.90227	0.02055						
2:303592-307027			6.58866	0.0115								
2:308846-312155			6.23157	0.0252								
2:313448-316741			4.39037	0.0385	5.81679	0.0224						
2:317595-321310			4.54269	0.03485								
2:432040-434086											4.50552	0.04075
2:628641-631308									-5.51788	0.04175		
2:783720-784826	-4.0042	0.0086										
2:813824-814652	4.51057	0.0111										
2:916351-923648			-5.79365	0.01125	-5.10567	0.0267	-6.02721	0.0184				
3:33640-41158	-4.07356	0.0097										
3:116137-121415			4.35021	0.04425	6.457	0.0281						
3:166193-173574					3.76371	0.0359						
3:315478-317011											4.64055	0.04395
3:366599-367464	3.67558	0.0322										
3:369306-370905									-5.26698	0.0231		
3:399606-402114							-4.10046	0.04935				
3:416759-418896									6.81174	0.02215		
3:645308-646205									5.50337	0.0169		
3:692350-693244			6.45452	0.04475								
3:833976-834941							5.77692	0.0275				
3:839208-840960											5.87332	0.0266
3:842382-843990									-4.30894	0.0409		
3:858916-866310							4.78978	0.02285				
3:867033-869827			4.50523	0.0342					3.90571	0.0183		
3:74872-77071	-4.10671	0.01365										

3:87792-90141							-4.58198	0.04775	-4.60258	0.0346		
3:92104-95765	4.84655	0.02985										
3:122264-127171			-4.60199	0.0203								
3:187920-189267	3.28014	0.02275										
3:325461-331944	3.56352	0.0349	5.46518	0.0318								
3:410907-416347											4.63833	0.03725
3:476797-480129			4.35103	0.03395	5.0121	0.03265						
3:642594-643518									4.47059	0.04765		
3:686826-691488					4.48714	0.03455						
3:965346-965874	3.18931	0.02385										
3:1027491-1034924									-4.72187	0.03265		
4:35152-44124									-4.81222	0.04175		
4:144097-153112							-4.73059	0.0497				
4:285540-288452					5.01916	0.03585						
4:463818-464917									5.23732	0.0358		
4:504793-512436					-5.63268	0.02015						
4:543550-546933			4.64954	0.01375								
4:629076-629919									4.76491	0.039		
4:672955-675130					5.74977	0.0238						
4:850541-852149							4.95943	0.04585			5.9847	0.02685
4:856772-859375							4.6009	0.0485				
4:1049820-1051136									6.80538	0.04005		
4:1062343-1066037			5.30938	0.0143								
4:1085397-1086628	3.37958	0.04185										
4:1113688-1115922	-4.86197	0.00685	-5.95164	0.04535					-5.85206	0.0294		
4:1122058-1126043	-3.01672	0.01875										
4:1131405-1132429	-5.70785	0.00515	-6.18157	0.0396								
4:69550-70923	4.04955	0.00985										
4:99022-100256	2.94456	0.02055										
4:114829-117745	-3.4987	0.04625										
4:121051-122509	4.83593	0.00535										
4:165736-166772	3.31424	0.04385										
4:204304-207667			4.44022	0.04065	6.64934	0.0254					5.01541	0.0279

4:240477-243386									-4.76246	0.0292		
4:273708-274845					5.09356	0.0272						
4:323011-325650					6.29062	0.02175						
4:415575-417594					4.48072	0.045						
4:504793-512436									-20.2777	0.0326		
4:535088-541382			4.51788	0.0311	4.74964	0.03585						
4:552883-560707	-5.48914	0.00935	-8.08563	0.00195								
4:568563-576239			-5.72095	0.0271								
4:784050-786405	3.48681	0.0392			7.37963	0.0324						
4:973469-981030			-5.60247	0.04875	-5.06649	0.0448			-4.7632	0.03325		
4:1073648-1077539			5.45642	0.04055								
4:1116973-1118115	-4.85178	0.0092					-5.14343	0.0466				
4:1176674-1183848	-3.53084	0.0169							-6.15105	0.03645	-4.91885	0.04685
5:20928-28456									-4.69027	0.0456		
5:55840-56872	3.51139	0.00955							-4.44682	0.04825		
5:64064-65489	4.48378	0.00075							-6.16354	0.0436		
5:74508-79842	-3.94749	0.02625										
5:125115-126514							4.70413	0.0459			5.80833	0.02715
5:167648-170590					5.80683	0.02345						
5:177409-178555					5.06034	0.03525						
5:280885-283345					4.3264	0.0272						
5:330970-339990			4.87654	0.01545								
5:344930-346744											5.45466	0.0263
5:459298-460342									6.28005	0.03115		
5:536270-536975											5.18792	0.03655
5:555261-556071									4.48861	0.0402		
5:637351-640379											5.04904	0.0282
5:956339-965158							-5.70315	0.0304				
5:1036223-1039621	4.45637	0.0036										
5:1071020-1073282			4.71069	0.03065								
5:1079773-1081003									5.18867	0.0226		
5:1088417-1089503									5.45261	0.02515		
5:1153669-1155667									-5.35668	0.03035		

5:1255213-1257247							5.68982	0.0243	5.49501	0.018		
5:1308272-1309935					-5.99634	0.03415	-6.37245	0.02815			-4.85266	0.03185
5:1312465-1314149							-5.89796	0.0352	-7.18159	0.0293		
5:1333464-1342959									-5.84191	0.02195		
5:43326-47761	4.85153	0.02465										
5:58036-59726							-6.25221	0.02165			-5.64636	0.0401
5:81133-82336					6.35579	0.0311						
5:84040-85237					5.07929	0.03645						
5:108604-112241			-4.38043	0.04075								
5:151421-152384	3.60229	0.03145										
5:569179-569917									-4.54689	0.04865		
5:652697-656545			5.57088	0.0188								
5:720740-722639					-5.06198	0.03375	-5.28608	0.0308				
5:791716-792934	2.62453	0.04075										
5:801446-809124	2.84938	0.02685										
5:856567-861520					-17.258	0.02285						
5:1190992-1191675									5.12776	0.0379		
5:1257715-1265743					4.42164	0.04185						
5:1279077-1280870									4.24143	0.02525		
5:1318525-1319224	3.87658	0.0289										
5:47843-48240	3.74978	0.0153										
5:63659-63870	3.3708	0.04305										
6:3502-12835									-5.52529	0.04345		
6:247093-250058	-3.33695	0.02535										
6:257371-278536							18.385	0.05025	18.2202	0.037		
6:320394-331721					4.14974	0.01275						
6:362630-363875	4.95401	0.01255										
6:444741-445152			6.67314	0.0096								
6:452610-461167							-4.95531	0.0414				
6:617756-621531											4.31489	0.04945
6:644047-646165			4.21636	0.04535								
6:669855-671628	2.93273	0.0411										
6:674113-676372											4.62504	0.0383

6:758605-760993					4.59868	0.04735						
6:784350-788024									4.43026	0.0374		
6:862497-870457	5.33187	0.02745										
6:881960-882992											-0.212759	0.03585
6:956765-958757									-4.50939	0.032		
6:993674-995276					4.72001	0.01985						
6:1078995-1081321					0.041506 2	0.02755						
6:1209027-1213794	3.69764	0.01055										
6:171297-177237					4.31652	0.04485						
6:313851-316878			4.10481	0.0491	4.79107	0.04415						
6:512279-520068					4.59683	0.0156					3.71729	0.041
6:530594-536310			32.5068	0.03845			32.0426	0.03715				
6:579429-585891					4.4086	0.0312						
6:614570-616954	3.36366	0.03095										
6:676884-677568											-0.21061	0.03585
6:723113-731487									-5.90419	0.0403	-4.88637	0.036
6:747281-749272			5.60537	0.01925								
6:815618-818244					3.6946	0.04775						
6:870531-873142	4.002	0.02965										
6:893447-895778											-4.1642	0.04895
6:1032536-1033775							-5.01795	0.04105	-4.56462	0.0361	-4.93825	0.03865
6:1252054-1253988									6.14526	0.01615	5.83456	0.03175
6:1272365-1273535									5.22155	0.0266		
6:1353947-1366431	-2.95059	0.03645							-5.75532	0.0223	-4.90357	0.0365
6:1374797-1382628	-3.53559	0.0302	-4.66965	0.03115	-6.08476	0.03745	-5.42134	0.02785				
7:30672-38328	-3.65888	0.03115			-5.09439	0.04485			-7.28021	0.03535		
7:106223-109753									-10.6772	0.0052		
7:140172-141050	-4.61159	0.01205										
7:144050-146252	-4.57387	0.0122							-5.44883	0.04525		
7:152126-154096	-3.07342	0.0295										
7:157059-157651	5.05352	0.0089										
7:165281-165545											6.68262	0.0321

7:393724-398616			4.13436	0.02505	5.69715	0.0296						
7:399290-401282	3.5087	0.03825										
7:661030-661818	3.35755	0.0455										
7:754614-756285			4.75094	0.0294								
7:772656-774344					4.9661	0.0272						
7:807252-809508							-4.66752	0.04665	-4.36175	0.04435		
7:896483-898206									4.88177	0.0393		
7:1001738-1003988			4.16809	0.0475	6.02219	0.0169						
7:1031141-1033784					5.88925	0.0172						
7:1140273-1141223	-3.5742	0.0173										
7:1145254-1145733	3.73142	0.02295										
7:1376681-1377743					-6.34265	0.0173						
7:1380750-1387189	4.18556	0.04055					-5.89606	0.04755	-6.05041	0.02185		
7:1394334-1397166					-4.21619	0.04165						
7:1413430-1418305					4.05688	0.04165			-4.50085	0.0456		
7:1423536-1425972									-6.39637	0.0385		
7:135265-138319	-4.92508	0.0047	-5.87273	0.0499			-5.36818	0.03505	-6.24242	0.023		
7:293538-297432					4.96638	0.0335						
7:320353-322819			4.85438	0.0255	4.6949	0.0328						
7:345480-346601					4.55448	0.04275						
7:567327-574803	-4.29125	0.0105	-6.99378	0.02915	-7.39891	0.043						
7:582715-590509			-4.38129	0.0377					-4.72426	0.0362		
7:598283-605898	-4.53336	0.00615							-5.14733	0.0266	-4.38082	0.0497
7:607535-614456	-4.88194	0.00565									-5.43986	0.0408
7:622103-629686	-3.87827	0.02085	-5.66546	0.0444	-5.73355	0.0478	-4.69462	0.0406	-5.36355	0.04895		
7:645703-653111											-5.31595	0.03645
7:743266-744224									7.05701	0.02885		
7:749398-752823			4.35019	0.0381								
7:762330-766824					-4.27098	0.0197						
7:1034295-1038564							-4.2231	0.04165				
7:1103087-1104738											5.09698	0.0312
7:1147855-1151777	-4.32051	0.0244	-5.72464	0.01175								
7:1472963-1481610									-6.55709	0.02925		

7:96906-98542			-6.08078	0.0351			-6.27133	0.028				
8:22368-29661									-5.14909	0.03195		
8:80946-81766	4.2812	0.02165										
8:166535-167198	4.04858	0.00945									4.48311	0.04415
8:181682-189233					5.4221	0.0158						
8:193095-197235					6.01874	0.01125						
8:208498-224335					5.08526	0.0436						
8:243222-245895					-6.08517	0.0447						
8:432286-440173	-3.41543	0.01895							-4.54579	0.04275		
8:441529-449184	-4.13398	0.01735							-5.16562	0.0215		
8:454923-456040	2.8317	0.042										
8:460432-468464	-3.57309	0.01485			-5.34288	0.0182						
8:676976-679939									32.324	0.02765		
8:933368-937474			4.13182	0.0464								
8:1094636-1097230							3.9052	0.04075				
8:1107119-1109033											4.40309	0.047
8:1147238-1151230			3.66576	0.0435								
8:1178381-1181396											5.22593	0.029
8:1185666-1186323											4.59326	0.03715
8:1269808-1271495									4.02019	0.04405		
8:1273788-1276527					4.33465	0.0498						
8:1276695-1279866					6.45389	0.0434						
8:1304393-1305884					-4.92197	0.04775	-4.97615	0.0372				
8:85691-89546					-5.09445	0.02815	-6.98152	0.04755				
8:152928-160809					4.73963	0.0086						
8:323494-324805									6.22879	0.031		
8:526178-528725			4.02672	0.04685	4.52898	0.04245						
8:552319-553549					6.72752	0.04725	5.47937	0.0294				
8:704622-705804											5.22475	0.0333
8:736800-742302					4.22076	0.03775						
8:795973-814591					6.50977	0.0493						
8:1063375-1065133	4.40438	0.01165										
8:1211413-1216651									-27.2144	0.0032		

8:1227156-1229731			4.4343	0.0304								
8:1238356-1240573											5.37608	0.02145
8:1297204-1299399	-3.18124	0.01435			-4.63167	0.02775						
8:1345475-1346558	4.75941	0.0064										
8:529374-531700							5.14981	0.0483	4.39929	0.04965		
8:790904-792936	3.07957	0.03275										
9:20079-27885									-6.55837	0.0354		
9:74408-75746	3.94594	0.02735										
9:162767-166922									-4.47068	0.0381		
9:168196-178893			5.34956	0.0034	6.57479	0.02915						
9:179157-184889									4.12931	0.0464		
9:414418-419887									4.43097	0.01445		
9:499508-509086									2.77215	0.0469		
9:509924-510518									6.09446	0.0329		
9:617953-619826							6.23639	0.0309			5.52455	0.03655
9:905987-908347											5.32947	0.0315
9:1008810-1016602							5.92322	0.03575	5.63939	0.02025		
9:1024245-1031995			21.0277	0.0453			21.9932	0.04065				
9:1113947-1115993					6.01464	0.02805						
9:1171328-1173473											4.41728	0.04935
9:1175192-1182158			5.3575	0.0346								
9:1187124-1189459									5.01714	0.00945		
9:1201801-1206964			6.01252	0.01565								
9:1286997-1288362			5.63292	0.01775			5.57544	0.02935				
9:1357399-1359251					4.39895	0.03915						
9:1361923-1363837					5.53774	0.02965						
9:1404921-1407968	-5.50951	0.0011	-5.01879	0.01945								
9:1413828-1419743					5.65604	0.0148						
9:89359-91294									-5.09771	0.04685		
9:109332-111328	-4.16468	0.0077										
9:119259-120748	3.26687	0.0478										
9:121619-125005			4.46071	0.03725	4.92749	0.04015						
9:133760-135586	3.2812	0.0455										

9:270737-274787			5.31055	0.0364								
9:548451-551182	-3.06257	0.0295										
9:579775-584352											3.69342	0.0248
9:601066-604567					4.55685	0.0434					4.3912	0.04345
9:729663-732126							-5.20802	0.0481	-5.54524	0.02545		
9:772635-774369									-4.66985	0.0364		
9:809648-819794					4.90605	0.02025						
9:852514-854047							-4.62594	0.03935			-4.7072	0.03135
9:862770-870481											-6.10721	0.0326
9:961922-963776					4.4523	0.04105						
9:1182676-1183685									4.79121	0.0348		
9:1420471-1422796			-6.44124	0.0154								
9:1426712-1428738			-5.13191	0.013			-8.68315	0.01405	-10.3347	0.0184		
9:1435590-1437970			-5.3475	0.0187								
9:1495566-1503324	-3.11974	0.02935										
9:1394359-1397560							4.95857	0.0422				
9:1438038-1438348	-3.50307	0.03825										

Appendix II

Table A2: The nsSNPs that predicted to affect protein function by at least two programs (SIFT, PROVEAN and PredictSNP) in *Plasmodium falciparum*

Table A2 (a).nsSNPs of PF3D7_0324900 (UniProt ID: O97312) that were predicted to be affect protein function by SIFT, PROVEAN and PredictSNP. From the table it can be observed that 28, 40 and 16 nsSNPs were reported to affect protein function by SIFT, PROVEAN and PredictSNP, respectively. Overall 25 nsSNPs in PF3D7_0324900 gene were found to be deleterious by these three tools. The nsSNPs that were predicted to affect protein function by at least two programs are shown in red color.

Uniprot ID / Gene ID	**SNP ID**	**AA change**	**SIFT Score**	**SIFT (Cutoff= < .05)**	**PROVEAN Score**	**PROVEAN (Cutoff= -2.5)**	**PredictSNP**	**Confidence**
O97312 / PF3D7_0324900	Pf3D7_03_v3.1038252	M1N	0	Affect protein function	-1.467	Neutral	Deleterious	51%
	Pf3D7_03_v3.1037913	A114S	0	Affect protein function	-2.75	Deleterious	Neutral	60%
	Pf3D7_03_v3.1037738	S173F			-3.6	Deleterious	Deleterious	51%
	Pf3D7_03_v3.1037087	K390Q	0	Affect protein function	-3.433	Deleterious	Neutral	73%
	Pf3D7_03_v3.1036410	N615K	0	Affect protein function	-4.067	Deleterious	Deleterious	51%
	Pf3D7_03_v3.1033921	G1445D	0	Affect protein function	-3.9	Deleterious	Neutral	60%
	Pf3D7_03_v3.1033916	D1447H			-2.533	Deleterious	Deleterious	61%
	Pf3D7_03_v3.1033848	R1469S	0.03	Affect protein function	-4.4	Deleterious	Deleterious	61%
	Pf3D7_03_v3.1033842	L1471F	0	Affect protein function	-3.667	Deleterious	Neutral	60%
	Pf3D7_03_v3.1032097	T1745P	0	Affect protein function	-5.867	Deleterious	Neutral	60%
	Pf3D7_03_v3.1032093	S1747R			-3.267	Deleterious	Deleterious	51%
	Pf3D7_03_v3.1032069	I1755F	0.03	Affect protein function	-3	Deleterious	Deleterious	64%
	Pf3D7_03_v3.1032054	S1760G	0	Affect protein function	-2.9	Deleterious	Neutral	83%
	Pf3D7_03_v3.1031477	Y1952F			-3.367	Deleterious	Neutral	60%

	Pf3D7_03_v3.1031478	Y1952N	0	Affect protein function	-7.5	Deleterious	Deleterious	65%
	Pf3D7_03_v3.1031346	V1996F	0.04	Affect protein function	-3	Deleterious	Deleterious	51%
	Pf3D7_03_v3.1031310	N2008D			-4.167	Deleterious	Deleterious	51%
	Pf3D7_03_v3.1031293	T2013I	0	Affect protein function	-5.8	Deleterious	Deleterious	62%
	Pf3D7_03_v3.1031278	V2018L			-2.6	Deleterious	Deleterious	62%
	Pf3D7_03_v3.1031262	S2024C	0.03	Affect protein function	-4.067	Deleterious	Deleterious	65%
	Pf3D7_03_v3.1031260	S2024R			-3.6	Deleterious	Deleterious	87%
	Pf3D7_03_v3.1031254	P2026S	0	Affect protein function	-7.567	Deleterious	Deleterious	51%
	Pf3D7_03_v3.1031239	L2031F			-3.6	Deleterious	Deleterious	52%
	Pf3D7_03_v3.1031189	W2048L	0	Affect protein function	-12.333	Deleterious	Neutral	83%
	Pf3D7_03_v3.1031139	E2065K	0	Affect protein function	-3.267	Deleterious	Neutral	75%

Table A2 (b).nsSNPs of PF3D7_1306000 (UniProt ID: Q8IEQ5) that were predicted to be affect protein function by SIFT, PROVEAN and PredictSNP. From the table it can be observed that 52, 4 and 18nsSNPs were reported to affect protein function by SIFT, PROVEAN and PredictSNP, respectively. Overall 20nsSNPs in PF3D7_1306000 gene were found to be deleterious by these three tools. The nsSNPs that were predicted to affect protein function by at least two programs are shown in red color.

Uniprot ID / Gene ID	**SNP ID**	**AA change**	**SIFT Score**	**SIFT (Cutoff= < .05)**	**PROVEAN Score**	**PROVEAN (Cutoff= -2.5)**	**PredictSNP**	**Confidence**
Q8IEQ5 / PF3D7_1306000	Pf3D7_13_v3.283953	L160H	0	Affect protein function	-2.167	Neutral	Deleterious	61%
	Pf3D7_13_v3.283983	N170I	0	Affect protein function	-1.167	Neutral	Deleterious	52%
	Pf3D7_13_v3.284025	E184G	0	Affect protein function	-1.167	Neutral	Deleterious	55%
	Pf3D7_13_v3.284553	W360L	0	Affect protein function	-0.5	Neutral	Deleterious	61%
	Pf3D7_13_v3.284826	I451N	0	Affect protein function	-0.333	Neutral	Deleterious	61%
	Pf3D7_13_v3.284877	Y468S	0	Affect protein function	-1	Neutral	Deleterious	87%
	Pf3D7_13_v3.284882	N470H	0	Affect protein function	-2.667	Deleterious	Neutral	63%
	Pf3D7_13_v3.285171	T566K	0	Affect protein function	-1.167	Neutral	Deleterious	61%
	Pf3D7_13_v3.285219	G582E	0	Affect protein function	0.167	Neutral	Deleterious	61%

	Pf3D7_13_v3.285340	W622C	0	Affect protein function	-2	Neutral	Deleterious	87%
	Pf3D7_13_v3.285345	Y624C	0	Affect protein function	-1.5	Neutral	Deleterious	87%
	Pf3D7_13_v3.285495	I674K	0	Affect protein function	0.333	Neutral	Deleterious	87%
	Pf3D7_13_v3.285588	K705I	0	Affect protein function	-2	Neutral	Deleterious	52%
	Pf3D7_13_v3.285655	R727S	0	Affect protein function	-1.167	Neutral	Deleterious	61%
	Pf3D7_13_v3.285674	N734D	0	Affect protein function	-1	Neutral	Deleterious	61%
	Pf3D7_13_v3.285809	Y779D	0	Affect protein function	-0.833	Neutral	Deleterious	87%
	Pf3D7_13_v3.285932	N820Y	0	Affect protein function	-3	Deleterious	Deleterious	61%
	Pf3D7_13_v3.285944	D824N	0	Affect protein function	-0.5	Neutral	Deleterious	61%
	Pf3D7_13_v3.286058	Y862N	0	Affect protein function	0.667	Neutral	Deleterious	61%
	Pf3D7_13_v3.286811	D1113Y	0.03	Affect protein function	-3.167	Deleterious	Deleterious	52%

Table A2 (c).nsSNPs of PF3D7_1439500 (UniProt ID: Q8IL74) that were predicted to be affect protein function by SIFT, PROVEAN and PredictSNP. From the table it can be observed that 60, 0 and 11nsSNPs were reported to affect protein function by SIFT, PROVEAN and PredictSNP, respectively. Overall 11nsSNPs in PF3D7_1439500 gene were found to be deleterious by these three tools. The nsSNPs that were predicted to affect protein function by at least two programs are shown in red color.

Uniprot ID / Gene ID	**SNP ID**	**AA change**	**SIFT Score**	**SIFT (Cutoff= < .05)**	**PROVEAN Score**	**PROVEAN (Cutoff= -2.5)**	**PredictSNP**	**Confidence**
Q8IL74 / PF3D7_1439500	Pf3D7_14_v3.1609074	N186H	0	Affect protein function	-0.021	Neutral	Deleterious	61%
	Pf3D7_14_v3.1609453	S312N	0	Affect protein function	-0.488	Neutral	Deleterious	51%
	Pf3D7_14_v3.1609456	S313N	0	Affect protein function	0.043	Neutral	Deleterious	51%
	Pf3D7_14_v3.1609519	N334S	0	Affect protein function	-0.774	Neutral	Deleterious	51%
	Pf3D7_14_v3.1609530	Y338N	0	Affect protein function	-0.151	Neutral	Deleterious	61%
	Pf3D7_14_v3.1609560	S348C	0	Affect protein function	-0.485	Neutral	Deleterious	55%
	Pf3D7_14_v3.1609726	Y403S	0	Affect protein function	-0.555	Neutral	Deleterious	61%
	Pf3D7_14_v3.1609750	T411I	0	Affect protein function	0.087	Neutral	Deleterious	55%
	Pf3D7_14_v3.1609852	S445L	0	Affect protein function	-0.51	Neutral	Deleterious	61%

	Pf3D7_14_v3.1610226	R570G	0	Affect protein function	-0.5	Neutral	Deleterious	51%
	Pf3D7_14_v3.1610756	N746K	0	Affect protein function	-0.311	Neutral	Deleterious	61%

Table A2 (d).nsSNPs of PF3D7_0705600 (UniProt ID: C0H4K7) that were predicted to be affect protein function by SIFT, PROVEAN and PredictSNP. From the table it can be observed that 33, 1 and 9nsSNPs were reported to affect protein function by SIFT, PROVEAN and PredictSNP, respectively. Overall 10nsSNPs in PF3D7_0705600 gene were found to be deleterious by these three tools. The nsSNPs that were predicted to affect protein function by at least two programs are shown in red color.

Uniprot ID / Gene ID	**SNP ID**	**AA change**	**SIFT Score**	**SIFT (Cutoff= < .05)**	**PROVEAN Score**	**PROVEAN (Cutoff= -2.5)**	**PredictSNP**	**Confidence**
C0H4K7 / PF3D7_0705600	Pf3D7_07_v3.284385	E17K	0	Affect protein function	-0.2	Neutral	Deleterious	61%
	Pf3D7_07_v3.284064	N124Y	0	Affect protein function	-0.267	Neutral	Deleterious	51%
	Pf3D7_07_v3.284003	G144D	0	Affect protein function	-0.217	Neutral	Deleterious	87%
	Pf3D7_07_v3.282569	H622V	0	Affect protein function	0.477	Neutral	Deleterious	55%
	Pf3D7_07_v3.282565	E623V	0	Affect protein function	-0.18	Neutral	Deleterious	55%
	Pf3D7_07_v3.282556	A626D	0	Affect protein function	-0.992	Neutral	Deleterious	61%
	Pf3D7_07_v3.282501	D645Y	0	Affect protein function	0.635	Neutral	Deleterious	61%
	Pf3D7_07_v3.282371	T688R	0	Affect protein function	0.226	Neutral	Deleterious	61%
	Pf3D7_07_v3.282311	G708D	0	Affect protein function	-0.633	Neutral	Deleterious	55%
	Pf3D7_07_v3.281879	T852S	0	Affect protein function	-3.389	Deleterious	Deleterious	51%

Table A2 (e).nsSNPs of PF3D7_1207100 (UniProt ID: Q8I5X4) that were predicted to be affect protein function by SIFT, PROVEAN and PredictSNP. From the table it can be observed that 19, 2 and 6nsSNPs were reported to affect protein function by SIFT, PROVEAN and PredictSNP, respectively. Overall 6nsSNPs in PF3D7_1207100 gene were found to be deleterious by these three tools. The nsSNPs that were predicted to affect protein function by at least two programs are shown in red color.

Uniprot ID / Gene ID	**SNP ID**	**AA change**	**SIFT Score**	**SIFT (Cutoff= < .05)**	**PROVEAN Score**	**PROVEAN (Cutoff= -2.5)**	**PredictSNP**	**Confidence**
Q8I5X4 / PF3D7_1207100	Pf3D7_12_v3.326154	N235K	0	Affect protein function	0.244	Neutral	Deleterious	61%
	Pf3D7_12_v3.326030	R277W	0	Affect protein function	-1.5	Neutral	Deleterious	65%
	Pf3D7_12_v3.325730	D377H	0	Affect protein function	-1.142	Neutral	Deleterious	52%
	Pf3D7_12_v3.325277	P528S	0.06	Tolerated	-5.367	Deleterious	Deleterious	61%
	Pf3D7_12_v3.325169	D564Y	0	Affect protein function	-3.773	Deleterious	Deleterious	61%
	Pf3D7_12_v3.324878	N661Y	0	Affect protein function	-0.764	Neutral	Deleterious	61%

Table A2 (f).nsSNPs of PF3D7_0508100 (UniProt ID: Q8I422) that were predicted to be affect protein function by SIFT, PROVEAN and PredictSNP. From the table it can be observed that 38, 0 and 4nsSNPs were reported to affect protein function by SIFT, PROVEAN and PredictSNP, respectively. Overall 4nsSNPs in PF3D7_0508100 gene were found to be deleterious by these three tools. The nsSNPs that were predicted to affect protein function by at least two programs are shown in red color.

Uniprot ID / Gene ID	**SNP ID**	**AA change**	**SIFT Score**	**SIFT (Cutoff= < .05)**	**PROVEAN Score**	**PROVEAN (Cutoff= -2.5)**	**PredictSNP**	**Confidence**
Q8I422 / PF3D7_0508100	Pf3D7_05_v3.334078	I800T	0	Affect protein function	-1.389	Neutral	Deleterious	55%
	Pf3D7_05_v3.334988	I1103K	0	Affect protein function	-0.648	Neutral	Deleterious	51%
	Pf3D7_05_v3.335910	S1411P	0.01	Affect protein function	-1.178	Neutral	Deleterious	61%
	Pf3D7_05_v3.336099	Y1474H	0.02	Affect protein function	-1.1	Neutral	Deleterious	55%

Table A2 (g).nsSNPs of PF3D7_1126700 (UniProt ID: Q8II97) that were predicted to be affect protein function by SIFT, PROVEAN and PredictSNP. From the table it can be observed that 7, 0 and 2nsSNPs were reported to affect protein function by SIFT, PROVEAN and PredictSNP, respectively. Overall 2nsSNPs in PF3D7_1126700 gene were found to be deleterious by these three tools. The nsSNPs that were predicted to affect protein function by at least two programs are shown in red color.

Uniprot ID / Gene ID	**SNP ID**	**AA change**	**SIFT Score**	**SIFT (Cutoff= < .05)**	**PROVEAN Score**	**PROVEAN (Cutoff= -2.5)**	**PredictSNP**	**Confidence**
Q8II97 / PF3D7_1126700	Pf3D7_11_v3.1041641	T644P	0.03	Affect protein function	-1.104	Neutral	Deleterious	61%
	Pf3D7_11_v3.1041262	S770I	0	Affect protein function	-1.11	Neutral	Deleterious	61%

Table A2 (h).nsSNPs of PF3D7_1234300 (UniProt ID: Q8I579) that were predicted to be affect protein function by SIFT, PROVEAN and PredictSNP. From the table it can be observed that 2, 1 and 1nsSNPs were reported to affect protein function by SIFT, PROVEAN and PredictSNP, respectively. Overall 1nsSNPs in PF3D7_1234300 gene were found to be deleterious by these three tools. The nsSNPs that were predicted to affect protein function by at least two programs are shown in red color.

Uniprot ID / Gene ID	**SNP ID**	**AA change**	**SIFT Score**	**SIFT (Cutoff= < .05)**	**PROVEAN Score**	**PROVEAN (Cutoff= -2.5)**	**PredictSNP**	**Confidence**
Q8I579 / PF3D7_1234300	Pf3D7_12_v3.1434394	S434Y	0.02	Affect protein function	-2.923	Deleterious	Deleterious	61%

Appendix III

Table A3: NetSurfP result

Residues that show a change in ASA from buried to exposed state and vice versa with change in ASA change of ≥ 10 $Å^2$ and also show change in their secondary structure.

Here Residue_N and Residue_SNP = residues in normal and SNP sequence; Class_N and Class_SNP = Class of residues in normal and SNP sequence [Class=buried (B) or exposed (E)]; SS_N and SS_SNP = secondary structure of residue in normal and SNP sequence

Table A3 (a). PF3D7_0324900:

SNP	Position	Residue_N	Class_N	SS_N	Residue_SNP	Class_SNP	SS_SNP
N615K	480	Y	B	C	Y	E	E
T1745P	1717	D	E	C	D	B	H
S1747R	1317	Q	E	C	Q	B	H
S2024R	1317	Q	E	C	Q	B	H
P2026S	1317	Q	E	C	Q	B	H
E2065K	1112	A	E	E	A	B	C
E2065K	1155	Y	E	C	Y	B	E

Table A3 (b). PF3D7_1306000:

SNP	Position	Residue_N	Class_N	SS_N	Residue_SNP	Class_SNP	SS_SNP
Y779D	779	Y	B	H	D	E	C
Y862N	865	T	B	E	T	E	C
D1113Y	1107	I	E	H	I	B	C

Table A3 (c). PF3D7_1439500:

SNP	Position	Residue_N	Class_N	SS_N	Residue_SNP	Class_SNP	SS_SNP
N186H	188	Q	E	C	Q	B	E
N186H	381	E	B	C	E	E	E
N186H	515	I	B	C	I	E	E
N186H	516	M	B	C	M	E	E
N186H	646	N	E	C	N	B	E
N186H	831	Q	B	C	Q	E	H
N186H	877	M	B	C	M	E	H
N186H	888	N	B	C	N	E	H
N186H	892	M	E	C	M	B	H
N186H	898	M	E	C	M	B	H

N186H	929	T	E	C	T	B	H
N186H	972	I	E	C	I	B	H
N186H	1001	V	E	C	V	B	H
N186H	681	T	B	E	T	E	C
N186H	683	T	B	E	T	E	C
N186H	984	T	B	H	T	E	C
N186H	986	K	B	H	K	E	C
S312N	320	F	B	E	F	E	C
S312N	596	V	E	C	V	B	E
S312N	667	F	E	C	F	B	E
S312N	681	T	B	E	T	E	C
S312N	683	T	B	E	T	E	C
S312N	789	K	B	C	K	E	E
S312N	831	Q	B	C	Q	E	H
S312N	972	I	E	C	I	B	H
S312N	984	T	B	H	T	E	C
S312N	986	K	B	H	K	E	C
S312N	992	I	B	C	I	E	H
S312N	1017	V	E	C	V	B	H
S312N	1030	Y	B	C	Y	E	E
S313N	320	F	B	E	F	E	C
S313N	381	E	B	C	E	E	E
S313N	449	Q	E	C	Q	B	E
S313N	542	Y	E	C	Y	B	E
S313N	594	S	E	C	S	B	E
S313N	642	Q	E	C	Q	B	E
S313N	644	L	E	C	L	B	E
S313N	683	T	B	E	T	E	C
S313N	789	K	B	C	K	E	H
S313N	829	M	E	C	M	B	H
S313N	830	N	B	C	N	E	H
S313N	831	Q	B	C	Q	E	H
S313N	984	T	B	H	T	E	E
S313N	986	K	B	H	K	E	C
S313N	992	I	B	C	I	E	H
S313N	1004	N	E	C	N	B	H
S313N	1030	Y	B	C	Y	E	E
N334S	447	I	E	C	I	B	E
N334S	623	N	B	C	N	E	E
N334S	681	T	B	E	T	E	C
N334S	683	T	B	E	T	E	C
N334S	846	S	E	C	S	B	H

N334S	877	M	B	C	M	E	H
N334S	879	S	B	C	S	E	H
N334S	880	M	E	C	M	B	H
N334S	889	M	E	C	M	B	H
N334S	975	I	B	C	I	E	H
N334S	983	I	B	H	I	E	C
N334S	984	T	B	H	T	E	C
N334S	986	K	B	H	K	E	C
Y338N	294	T	E	C	T	B	E
Y338N	295	Q	E	C	Q	B	E
Y338N	475	C	E	C	C	B	E
Y338N	491	I	E	E	I	B	C
Y338N	625	I	E	C	I	B	E
Y338N	681	T	B	E	T	E	C
Y338N	716	N	E	C	N	B	E
Y338N	925	M	E	C	M	B	H
Y338N	930	M	E	C	M	B	E
Y338N	984	T	B	H	T	E	C
Y338N	986	K	B	H	K	E	C
Y338N	1030	Y	B	C	Y	E	E
S348C	381	E	B	C	E	E	E
S348C	423	M	B	C	M	E	E
S348C	681	T	B	E	T	E	C
S348C	828	P	E	C	P	B	H
S348C	830	N	B	C	N	E	H
S348C	831	Q	B	C	Q	E	H
S348C	836	L	E	C	L	B	H
S348C	879	S	B	C	S	E	H
S348C	920	M	B	C	M	E	E
S348C	956	I	B	C	I	E	E
S348C	974	D	E	C	D	B	H
S348C	976	I	E	C	I	B	H
S348C	984	T	B	H	T	E	C
S348C	986	K	B	H	K	E	C
Y403S	320	F	B	E	F	E	C
Y403S	515	I	B	C	I	E	E
Y403S	516	M	B	C	M	E	E
Y403S	681	T	B	E	T	E	C
Y403S	836	L	E	C	L	B	H
Y403S	920	M	B	C	M	E	H
Y403S	925	M	E	C	M	B	H
Y403S	930	M	E	C	M	B	H

Y403S	940	M	E	C	M	B	H
Y403S	984	T	B	H	T	E	C
Y403S	986	K	B	H	K	E	C
Y403S	992	I	B	C	I	E	H
Y403S	993	I	E	C	I	B	H
Y403S	1002	V	E	C	V	B	H
Y403S	1012	H	B	C	H	E	H
Y403S	1015	N	B	C	N	E	H
Y403S	1016	N	E	C	N	B	H
Y403S	1017	V	E	C	V	B	H
T411I	320	F	B	E	F	E	C
T411I	372	N	B	C	N	E	E
T411I	381	E	B	C	E	E	E
T411I	449	Q	E	C	Q	B	E
T411I	491	I	E	E	I	B	C
T411I	625	I	E	C	I	B	E
T411I	644	L	E	C	L	B	E
T411I	681	T	B	E	T	E	C
T411I	789	K	B	C	K	E	H
T411I	791	L	B	C	L	E	H
T411I	830	N	B	C	N	E	H
T411I	831	Q	B	C	Q	E	H
T411I	874	M	B	C	M	E	H
T411I	910	I	E	C	I	B	H
T411I	937	N	E	C	N	B	E
T411I	972	I	E	C	I	B	H
T411I	984	T	B	H	T	E	C
T411I	986	K	B	H	K	E	C
T411I	1002	V	E	C	V	B	E
S445L	189	I	E	E	I	B	C
S445L	320	F	B	E	F	E	C
S445L	372	N	B	C	N	E	E
S445L	381	E	B	C	E	E	E
S445L	491	I	E	E	I	B	C
S445L	625	I	E	C	I	B	E
S445L	645	I	E	E	I	B	C
S445L	681	T	B	E	T	E	C
S445L	829	M	E	C	M	B	H
S445L	830	N	B	C	N	E	H
S445L	831	Q	B	C	Q	E	H
S445L	845	I	E	C	I	B	H
S445L	873	N	B	C	N	E	H

S445L	874	M	B	C	M	E	H
S445L	920	M	B	C	M	E	H
S445L	984	T	B	H	T	E	C
S445L	986	K	B	H	K	E	C
S445L	1001	V	E	C	V	B	H
S445L	1002	V	E	C	V	B	H
S445L	1023	L	E	C	L	B	E
R570G	189	I	E	E	I	B	C
R570G	320	F	B	E	F	E	C
R570G	381	E	B	C	E	E	E
R570G	491	I	E	E	I	B	C
R570G	625	I	E	C	I	B	E
R570G	645	I	E	E	I	B	C
R570G	681	T	B	E	T	E	C
R570G	742	N	E	C	N	B	E
R570G	756	N	E	C	N	B	E
R570G	767	M	B	C	M	E	E
R570G	829	M	E	C	M	B	H
R570G	830	N	B	C	N	E	H
R570G	831	Q	B	C	Q	E	H
R570G	873	N	B	C	N	E	H
R570G	874	M	B	C	M	E	H
R570G	920	M	B	C	M	E	H
R570G	984	T	B	H	T	E	C
R570G	986	K	B	H	K	E	C
R570G	1001	V	E	C	V	B	H
R570G	1002	V	E	C	V	B	H
N746K	84	E	E	H	E	B	H
N746K	87	K	E	C	K	B	H
N746K	301	N	E	C	N	B	E
N746K	320	F	B	E	F	E	C
N746K	372	N	B	C	N	E	E
N746K	374	N	B	C	N	E	E
N746K	381	E	B	C	E	E	E
N746K	491	I	E	E	I	B	C
N746K	547	N	E	C	N	B	E
N746K	556	I	B	C	I	E	E
N746K	594	S	E	C	S	B	E
N746K	596	V	E	C	V	B	E
N746K	767	M	B	C	M	E	E
N746K	829	M	E	C	M	B	H
N746K	830	N	B	C	N	E	H

N746K	831	Q	B	C	Q	E	H
N746K	939	T	E	C	T	B	H
N746K	940	M	E	C	M	B	H
N746K	983	I	B	H	I	E	C
N746K	984	T	B	H	T	E	C
N746K	986	K	B	H	K	E	C
N746K	1001	V	E	C	V	B	H
N746K	1002	V	E	C	V	B	H

Table A3 (d). PF3D7_0705600:

SNP	Position	Residue_N	Class_N	SS_N	Residue_SNP	Class_SNP	SS_SNP
N124Y	817	I	B	C	I	E	C
N124Y	687	N	B	C	N	E	C
N124Y	1042	L	E	C	L	B	C
N124Y	124	N	E	C	Y	B	E
N124Y	968	K	B	C	K	E	H
E623V	675	L	E	C	L	B	E
E623V	676	G	E	C	G	B	E
E623V	571	T	E	C	T	B	H
E623V	570	F	B	C	F	E	H
E623V	681	K	E	E	K	B	C
E623V	672	K	E	H	K	B	C
E623V	701	D	E	H	D	B	C
E623V	673	D	E	H	D	B	C
E623V	502	K	B	H	K	E	C
E623V	624	M	B	H	M	E	C
E623V	546	L	B	H	L	E	C
E623V	658	Y	B	H	Y	E	C
E623V	670	N	B	H	N	E	C
E623V	700	D	E	H	D	B	E
E623V	674	R	E	H	R	B	E
E623V	697	Q	E	H	Q	B	E
E623V	508	N	E	H	N	B	E
A626D	675	L	E	C	L	B	E
A626D	676	G	E	C	G	B	E
A626D	570	F	B	C	F	E	H
A626D	571	T	E	C	T	B	H
A626D	681	K	E	E	K	B	C
A626D	502	K	B	H	K	E	C
A626D	546	L	B	H	L	E	C

A626D	658	Y	B	H	Y	E	C
A626D	670	N	B	H	N	E	C
A626D	672	K	E	H	K	B	C
A626D	701	D	E	H	D	B	C
A626D	673	D	E	H	D	B	C
A626D	700	D	E	H	D	B	E
A626D	674	R	E	H	R	B	E
A626D	697	Q	E	H	Q	B	E
A626D	508	N	E	H	N	B	E
D645Y	675	L	E	C	L	B	E
D645Y	676	G	E	C	G	B	E
D645Y	636	D	E	C	D	B	H
D645Y	571	T	E	C	T	B	H
D645Y	570	F	B	C	F	E	H
D645Y	643	I	B	C	I	E	H
D645Y	641	H	B	C	H	E	H
D645Y	681	K	E	E	K	B	C
D645Y	672	K	E	H	K	B	C
D645Y	701	D	E	H	D	B	C
D645Y	673	D	E	H	D	B	C
D645Y	502	K	B	H	K	E	C
D645Y	546	L	B	H	L	E	C
D645Y	670	N	B	H	N	E	C
D645Y	700	D	E	H	D	B	E
D645Y	674	R	E	H	R	B	E
D645Y	697	Q	E	H	Q	B	E
D645Y	508	N	E	H	N	B	E
D645Y	658	Y	B	H	Y	E	E
T688R	675	L	E	C	L	B	E
T688R	676	G	E	C	G	B	E
T688R	571	T	E	C	T	B	H
T688R	653	S	B	C	S	E	H
T688R	638	M	B	C	M	E	H
T688R	687	N	B	C	N	E	H
T688R	570	F	B	C	F	E	H
T688R	641	H	B	C	H	E	H
T688R	681	K	E	E	K	B	C
T688R	701	D	E	H	D	B	C
T688R	672	K	E	H	K	B	C
T688R	718	N	E	H	N	B	C
T688R	673	D	E	H	D	B	C
T688R	502	K	B	H	K	E	C
T688R	546	L	B	H	L	E	C

T688R	658	Y	B	H	Y	E	C
T688R	670	N	B	H	N	E	C
T688R	700	D	E	H	D	B	E
T688R	674	R	E	H	R	B	E
T688R	697	Q	E	H	Q	B	E
T688R	621	E	E	H	E	B	E
T688R	508	N	E	H	N	B	E
T852S	401	V	B	C	V	E	E
T852S	497	V	B	C	V	E	H
T852S	454	N	E	E	N	B	C
T852S	308	N	B	E	N	E	C
T852S	334	S	E	H	S	B	C
T852S	489	E	E	H	E	B	E

Table A3 (e). PF3D7_1207100:

SNP	Position	Residue_N	Class_N	SS_N	Residue_SNP	Class_SNP	SS_SNP
N235K	64	I	E	C	I	B	E
N235K	47	I	E	C	I	B	H
N235K	337	L	E	C	L	B	H
N235K	187	I	E	H	I	B	C
N235K	158	G	B	C	G	E	H
R277W	274	F	E	C	F	B	H
R277W	275	E	E	C	E	B	H
D377H	367	Y	E	H	Y	B	C
D377H	410	I	B	H	I	E	C
D377H	512	S	B	E	S	E	H
P528S	337	L	E	C	L	B	H
P528S	340	Y	E	C	Y	B	H
N661Y	337	L	E	C	L	B	H
N661Y	410	I	B	H	I	E	C

Table A3 (f). PF3D7_0508100:

SNP	Position	Residue_N	Class_N	SS_N	Residue_SNP	Class_SNP	SS_SNP
I800T	558	E	E	C	E	B	H

I800T	570	G	E	H	G	B	C
I800T	576	N	E	C	N	B	H
I800T	634	M	B	H	M	E	E
I800T	638	I	B	H	I	E	E
I800T	754	E	E	H	E	B	C
I800T	757	Y	E	H	Y	B	C
I800T	769	G	B	H	G	E	C
I800T	969	F	B	H	F	E	C
I800T	976	Y	E	C	Y	B	H
I800T	1032	I	E	H	I	B	C
I800T	1423	I	E	H	I	B	C
I800T	1586	E	E	C	E	B	H
I1103K	343	Q	B	C	Q	E	H
I1103K	399	K	B	E	K	E	H
I1103K	528	L	E	C	L	B	E
I1103K	530	R	B	C	R	E	E
I1103K	531	F	E	E	F	B	C
I1103K	532	D	B	E	D	E	C
I1103K	561	Q	B	C	Q	E	H
I1103K	568	Y	E	H	Y	B	C
I1103K	569	N	B	H	N	E	C
I1103K	634	M	B	H	M	E	E
I1103K	638	I	B	H	I	E	E
I1103K	754	E	E	H	E	B	C
I1103K	757	Y	E	H	Y	B	C
I1103K	769	G	B	H	G	E	C
I1103K	969	F	B	H	F	E	C
I1103K	976	Y	E	C	Y	B	H
I1103K	1105	N	B	H	N	E	C
I1103K	1119	K	B	H	K	E	C
I1103K	1423	I	E	H	I	B	C
I1103K	1610	M	B	C	M	E	H
I1103K	1635	L	E	C	L	B	H
I1103K	1653	I	E	C	I	B	E
I1103K	1662	K	E	H	K	B	C
S1411P	521	A	E	C	A	B	H
S1411P	523	I	B	C	I	E	H
S1411P	558	E	E	C	E	B	H
S1411P	576	N	E	C	N	B	H
S1411P	634	M	B	H	M	E	E
S1411P	638	I	B	H	I	E	E
S1411P	754	E	E	H	E	B	C

S1411P	757	Y	E	H	Y	B	C
S1411P	769	G	B	H	G	E	C
S1411P	967	E	B	H	E	E	E
S1411P	969	F	B	H	F	E	C
S1411P	976	Y	E	C	Y	B	H
S1411P	1032	I	E	H	I	B	C
S1411P	1395	Q	B	H	Q	E	C
S1411P	1485	C	B	E	C	E	H
S1411P	1489	M	B	E	M	E	H
S1411P	1501	S	B	H	S	E	C
S1411P	1502	F	B	H	F	E	C
S1411P	1503	N	B	H	N	E	C
S1411P	1504	H	B	H	H	E	C
S1411P	1505	Q	B	H	Q	E	C
S1411P	1506	Q	B	H	Q	E	C
S1411P	1507	R	E	H	R	B	C
S1411P	1586	E	E	C	E	B	H
S1411P	1607	Q	E	C	Q	B	H
S1411P	1635	L	E	C	L	B	H
S1411P	1650	D	E	C	D	B	E
S1411P	1653	I	E	C	I	B	E
S1411P	1663	T	B	H	T	E	C
S1411P	1665	F	B	H	F	E	C
Y1474H	558	E	E	C	E	B	H
Y1474H	570	G	E	H	G	B	C
Y1474H	576	N	E	C	N	B	H
Y1474H	634	M	B	H	M	E	E
Y1474H	638	I	B	H	I	E	E
Y1474H	757	Y	E	H	Y	B	C
Y1474H	769	G	B	H	G	E	C
Y1474H	774	P	E	C	P	B	H
Y1474H	967	E	B	H	E	E	E
Y1474H	969	F	B	H	F	E	C
Y1474H	976	Y	E	C	Y	B	H
Y1474H	1119	K	B	H	K	E	C
Y1474H	1136	S	E	C	S	B	H
Y1474H	1493	S	E	C	S	B	H
Y1474H	1501	S	B	H	S	E	C
Y1474H	1502	F	B	H	F	E	C
Y1474H	1503	N	B	H	N	E	C
Y1474H	1504	H	B	H	H	E	C
Y1474H	1505	Q	B	H	Q	E	C

Y1474H	1506	Q	B	H	Q	E	C
Y1474H	1507	R	E	H	R	B	C
Y1474H	1508	N	E	H	N	B	C
Y1474H	1586	E	E	C	E	B	H
Y1474H	1587	G	E	C	G	B	H
Y1474H	1610	M	B	C	M	E	H
Y1474H	1635	L	E	C	L	B	H
Y1474H	1650	D	E	C	D	B	E
Y1474H	1653	I	E	C	I	B	E

Table A3 (g). PF3D7_1234300:

SNP	Position	Residue_N	Class_N	SS_N	Residue_SNP	Class_SNP	SS_SNP
S434Y	137	I	B	C	I	E	H

Appendix IV

Table A4: List of KEGG pathways of Ring verses other stages (R-ET, R-LT, R-Sc, R-GII, R-GV and R-Oo) of *P. falciparum*

A. The KEGG pathways of R-ET

KEGG Orthology (KO)	Pathway Name
ko01100	Metabolic pathways (102)
ko01110	Biosynthesis of secondary metabolites (44)
ko03010	Ribosome (43)
ko01130	Biosynthesis of antibiotics (32)
ko01120	Microbial metabolism in diverse environments (20)
ko00230	Purine metabolism (17)
ko01200	Carbon metabolism (17)
ko00240	Pyrimidine metabolism (14)
ko05144	Malaria (13)
ko04141	Protein processing in endoplasmic reticulum (12)
ko05010	Alzheimer's disease (12)
ko05016	Huntington's disease (11)
ko00190	Oxidative phosphorylation (11)
ko03030	DNA replication (10)
ko04110	Cell cycle (10)
ko05200	Pathways in cancer (10)
ko04114	Oocyte meiosis (10)
ko05012	Parkinson's disease (10)
ko00010	Glycolysis / Gluconeogenesis (10)
ko00970	Aminoacyl-tRNA biosynthesis (10)
ko01230	Biosynthesis of amino acids (10)
ko01212	Fatty acid metabolism (9)
ko00564	Glycerophospholipid metabolism (9)
ko05166	HTLV-I infection (9)
ko04120	Ubiquitin mediated proteolysis (9)
ko03018	RNA degradation (9)
ko04145	Phagosome (8)
ko04310	Wnt signaling pathway (8)
ko03440	Homologous recombination (8)
ko03430	Mismatch repair (8)
ko00620	Pyruvate metabolism (8)
ko05152	Tuberculosis (8)
ko04922	Glucagon signaling pathway (8)
ko04111	Cell cycle - yeast (8)
ko03420	Nucleotide excision repair (8)
ko05169	Epstein-Barr virus infection (8)
ko04932	Non-alcoholic fatty liver disease (NAFLD) (8)

ko04138	Autophagy - yeast (7)
ko00860	Porphyrin and chlorophyll metabolism (7)
ko04144	Endocytosis (7)
ko03013	RNA transport (7)
ko03040	Spliceosome (7)
ko04217	Necroptosis (6)
ko00520	Amino sugar and nucleotide sugar metabolism (6)
ko04260	Cardiac muscle contraction (6)
ko00280	Valine, leucine and isoleucine degradation (6)
ko03410	Base excision repair (6)
ko04140	Autophagy - animal (6)
ko04728	Dopaminergic synapse (6)
ko05165	Human papillomavirus infection (6)
ko05167	Kaposi's sarcoma-associated herpesvirus infection (6)
ko04962	Vasopressin-regulated water reabsorption (6)
ko00640	Propanoate metabolism (6)
ko00061	Fatty acid biosynthesis (6)
ko05168	Herpes simplex infection (5)
ko04130	SNARE interactions in vesicular transport (5)
ko04020	Calcium signaling pathway (5)
ko04910	Insulin signaling pathway (5)
ko04142	Lysosome (5)
ko00630	Glyoxylate and dicarboxylate metabolism (5)
ko04921	Oxytocin signaling pathway (5)
ko05203	Viral carcinogenesis (5)
ko00020	Citrate cycle (TCA cycle) (5)
ko05132	Salmonella infection (5)
ko04151	PI3K-Akt signaling pathway (5)
ko00900	Terpenoid backbone biosynthesis (5)
ko00051	Fructose and mannose metabolism (5)
ko03460	Fanconi anemia pathway (5)
ko04218	Cellular senescence (5)
ko04022	cGMP-PKG signaling pathway (5)
ko02020	Two-component system (5)
ko00260	Glycine, serine and threonine metabolism (5)
ko03015	mRNA surveillance pathway (5)
ko04113	Meiosis - yeast (5)
ko04212	Longevity regulating pathway - worm (5)
ko03060	Protein export (4)
ko04270	Vascular smooth muscle contraction (4)
ko01524	Platinum drug resistance (4)
ko04530	Tight junction (4)
ko05230	Central carbon metabolism in cancer (4)
ko05034	Alcoholism (4)
ko04360	Axon guidance (4)
ko05014	Amyotrophic lateral sclerosis (ALS) (4)

ko04261	Adrenergic signaling in cardiomyocytes (4)
ko00330	Arginine and proline metabolism (4)
ko04914	Progesterone-mediated oocyte maturation (4)
ko03050	Proteasome (4)
ko05031	Amphetamine addiction (4)
ko03008	Ribosome biogenesis in eukaryotes (4)
ko04540	Gap junction (4)
ko00250	Alanine, aspartate and glutamate metabolism (4)
ko04371	Apelin signaling pathway (4)
ko04066	HIF-1 signaling pathway (4)
ko04626	Plant-pathogen interaction (4)
ko04341	Hedgehog signaling pathway - fly (4)
ko04724	Glutamatergic synapse (4)
ko05145	Toxoplasmosis (4)
ko00710	Carbon fixation in photosynthetic organisms (4)
ko04721	Synaptic vesicle cycle (4)
ko05134	Legionellosis (4)
ko04924	Renin secretion (4)
ko00480	Glutathione metabolism (4)
ko04720	Long-term potentiation (4)
ko04918	Thyroid hormone synthesis (3)
ko04210	Apoptosis (3)
ko05215	Prostate cancer (3)
ko00450	Selenocompound metabolism (3)
ko05322	Systemic lupus erythematosus (3)
ko05161	Hepatitis B (3)
ko04810	Regulation of actin cytoskeleton (3)
ko05160	Hepatitis C (3)
ko00680	Methane metabolism (3)
ko00983	Drug metabolism - other enzymes (3)
ko05130	Pathogenic Escherichia coli infection (3)
ko04213	Longevity regulating pathway - multiple species (3)
ko05210	Colorectal cancer (3)
ko04380	Osteoclast differentiation (3)
ko05222	Small cell lung cancer (3)
ko00333	Prodigiosin biosynthesis (3)
ko04152	AMPK signaling pathway (3)
ko01040	Biosynthesis of unsaturated fatty acids (3)
ko04214	Apoptosis - fly (3)
ko04740	Olfactory transduction (3)
ko03320	PPAR signaling pathway (3)
ko00030	Pentose phosphate pathway (3)
ko04068	FoxO signaling pathway (3)
ko04916	Melanogenesis (3)
ko00780	Biotin metabolism (3)
ko04024	cAMP signaling pathway (3)

ko00510	N-Glycan biosynthesis (3)
ko00220	Arginine biosynthesis (3)
ko04713	Circadian entrainment (3)
ko05146	Amoebiasis (3)
ko05110	Vibrio cholerae infection (3)
ko04919	Thyroid hormone signaling pathway (3)
ko04915	Estrogen signaling pathway (3)
ko04611	Platelet activation (3)
ko00513	Various types of N-glycan biosynthesis (3)
ko04010	MAPK signaling pathway (3)
ko04013	MAPK signaling pathway - fly (3)
ko04660	T cell receptor signaling pathway (3)
ko04350	TGF-beta signaling pathway (3)
ko04970	Salivary secretion (3)
ko05162	Measles (3)
ko04115	p53 signaling pathway (3)
ko04340	Hedgehog signaling pathway (3)
ko04137	Mitophagy - animal (3)
ko05323	Rheumatoid arthritis (3)
ko00790	Folate biosynthesis (3)
ko04971	Gastric acid secretion (3)
ko05164	Influenza A (3)
ko04727	GABAergic synapse (3)
ko04662	B cell receptor signaling pathway (3)
ko04657	IL-17 signaling pathway (3)
ko00910	Nitrogen metabolism (3)
ko05225	Hepatocellular carcinoma (3)
ko00770	Pantothenate and CoA biosynthesis (2)
ko05032	Morphine addiction (2)
ko00071	Fatty acid degradation (2)
ko02024	Quorum sensing (2)
ko04139	Mitophagy - yeast (2)
ko05133	Pertussis (2)
ko04923	Regulation of lipolysis in adipocytes (2)
ko00500	Starch and sucrose metabolism (2)
ko04666	Fc gamma R-mediated phagocytosis (2)
ko04211	Longevity regulating pathway (2)
ko04510	Focal adhesion (2)
ko04014	Ras signaling pathway (2)
ko04668	TNF signaling pathway (2)
ko00750	Vitamin B6 metabolism (2)
ko04070	Phosphatidylinositol signaling system (2)
ko05418	Fluid shear stress and atherosclerosis (2)
ko04930	Type II diabetes mellitus (2)
ko04011	MAPK signaling pathway - yeast (2)
ko04146	Peroxisome (2)

ko04730	Long-term depression (2)
ko04650	Natural killer cell mediated cytotoxicity (2)
ko05120	Epithelial cell signaling in Helicobacter pylori infection (2)
ko00760	Nicotinate and nicotinamide metabolism (2)
ko05020	Prion diseases (2)
ko04925	Aldosterone synthesis and secretion (2)
ko04370	VEGF signaling pathway (2)
ko00720	Carbon fixation pathways in prokaryotes (2)
ko00670	One carbon pool by folate (2)
ko00730	Thiamine metabolism (2)
ko00410	beta-Alanine metabolism (2)
ko00310	Lysine degradation (2)
ko00563	Glycosylphosphatidylinositol (GPI)-anchor biosynthesis (2)
ko04912	GnRH signaling pathway (2)
ko00785	Lipoic acid metabolism (2)
ko05226	Gastric cancer (2)
ko04710	Circadian rhythm (2)
ko04062	Chemokine signaling pathway (2)
ko00270	Cysteine and methionine metabolism (2)
ko04390	Hippo signaling pathway (2)
ko04658	Th1 and Th2 cell differentiation (2)
ko04136	Autophagy - other (2)
ko04750	Inflammatory mediator regulation of TRP channels (2)
ko03020	RNA polymerase (2)
ko04071	Sphingolipid signaling pathway (2)
ko04722	Neurotrophin signaling pathway (2)
ko00561	Glycerolipid metabolism (2)
ko04016	MAPK signaling pathway - plant (2)
ko04659	Th17 cell differentiation (2)
ko04621	NOD-like receptor signaling pathway (2)
ko00062	Fatty acid elongation (1)
ko00920	Sulfur metabolism (1)
ko05202	Transcriptional misregulation in cancer (1)
ko04974	Protein digestion and absorption (1)
ko04976	Bile secretion (1)
ko05214	Glioma (1)
ko00140	Steroid hormone biosynthesis (1)
ko04913	Ovarian steroidogenesis (1)
ko04550	Signaling pathways regulating pluripotency of stem cells (1)
ko00514	Other types of O-glycan biosynthesis (1)
ko05030	Cocaine addiction (1)
ko04711	Circadian rhythm - fly (1)
ko05213	Endometrial cancer (1)
ko00072	Synthesis and degradation of ketone bodies (1)
ko04122	Sulfur relay system (1)
ko04920	Adipocytokine signaling pathway (1)

ko00565	Ether lipid metabolism (1)
ko05416	Viral myocarditis (1)
ko04725	Cholinergic synapse (1)
ko04926	Relaxin signaling pathway (1)
ko05224	Breast cancer (1)
ko04742	Taste transduction (1)
ko04744	Phototransduction (1)
ko05414	Dilated cardiomyopathy (DCM) (1)
ko05142	Chagas disease (American trypanosomiasis) (1)
ko05206	MicroRNAs in cancer (1)
ko04150	mTOR signaling pathway (1)
ko04940	Type I diabetes mellitus (1)
ko00052	Galactose metabolism (1)
ko03070	Bacterial secretion system (1)
ko00627	Aminobenzoate degradation (1)
ko00524	Neomycin, kanamycin and gentamicin biosynthesis (1)
ko05205	Proteoglycans in cancer (1)
ko04215	Apoptosis - multiple species (1)
ko00380	Tryptophan metabolism (1)
ko01522	Endocrine resistance (1)
ko00521	Streptomycin biosynthesis (1)
ko04745	Phototransduction - fly (1)
ko05223	Non-small cell lung cancer (1)
ko04112	Cell cycle - Caulobacter (1)
ko00040	Pentose and glucuronate interconversions (1)
ko00940	Phenylpropanoid biosynthesis (1)
ko04966	Collecting duct acid secretion (1)
ko05211	Renal cell carcinoma (1)
ko04623	Cytosolic DNA-sensing pathway (1)
ko00440	Phosphonate and phosphinate metabolism (1)
ko04624	Toll and Imd signaling pathway (1)
ko04723	Retrograde endocannabinoid signaling (1)
ko00562	Inositol phosphate metabolism (1)
ko05100	Bacterial invasion of epithelial cells (1)
ko00908	Zeatin biosynthesis (1)
ko03022	Basal transcription factors (1)
ko00740	Riboflavin metabolism (1)
ko05217	Basal cell carcinoma (1)
ko04917	Prolactin signaling pathway (1)
ko00195	Photosynthesis (1)
ko00362	Benzoate degradation (1)
ko04012	ErbB signaling pathway (1)
ko04911	Insulin secretion (1)
ko05131	Shigellosis (1)
ko04961	Endocrine and other factor-regulated calcium reabsorption (1)
ko04216	Ferroptosis (1)

ko03450	Non-homologous end-joining (1)
ko04391	Hippo signaling pathway - fly (1)
ko04392	Hippo signaling pathway - multiple species (1)
ko04330	Notch signaling pathway (1)
ko04931	Insulin resistance (1)
ko01521	EGFR tyrosine kinase inhibitor resistance (1)
ko04726	Serotonergic synapse (1)
ko04015	Rap1 signaling pathway (1)
ko00650	Butanoate metabolism (1)
ko04973	Carbohydrate digestion and absorption (1)

B. The KEGG pathways of R-LT

ko01100	Metabolic pathways (146)
ko01110	Biosynthesis of secondary metabolites (63)
ko01130	Biosynthesis of antibiotics (43)
ko00230	Purine metabolism (33)
ko01120	Microbial metabolism in diverse environments (27)
ko05169	Epstein-Barr virus infection (26)
ko00240	Pyrimidine metabolism (26)
ko01200	Carbon metabolism (26)
ko03030	DNA replication (24)
ko00190	Oxidative phosphorylation (23)
ko03050	Proteasome (23)
ko05016	Huntington's disease (22)
ko03008	Ribosome biogenesis in eukaryotes (21)
ko05010	Alzheimer's disease (21)
ko05012	Parkinson's disease (20)
ko04110	Cell cycle (19)
ko04141	Protein processing in endoplasmic reticulum (19)
ko04111	Cell cycle - yeast (18)
ko04120	Ubiquitin mediated proteolysis (16)
ko03420	Nucleotide excision repair (15)
ko05144	Malaria (15)
ko04144	Endocytosis (15)
ko03010	Ribosome (15)
ko04113	Meiosis - yeast (13)
ko05166	HTLV-I infection (13)
ko04114	Oocyte meiosis (13)
ko00970	Aminoacyl-tRNA biosynthesis (13)
ko01230	Biosynthesis of amino acids (13)
ko03430	Mismatch repair (13)
ko03410	Base excision repair (12)
ko00020	Citrate cycle (TCA cycle) (12)
ko03440	Homologous recombination (12)
ko03013	RNA transport (11)

ko05200	Pathways in cancer (11)
ko04138	Autophagy - yeast (10)
ko03060	Protein export (10)
ko00620	Pyruvate metabolism (10)
ko04932	Non-alcoholic fatty liver disease (NAFLD) (10)
ko00010	Glycolysis / Gluconeogenesis (9)
ko03020	RNA polymerase (9)
ko03460	Fanconi anemia pathway (9)
ko05152	Tuberculosis (8)
ko04217	Necroptosis (8)
ko01212	Fatty acid metabolism (8)
ko04310	Wnt signaling pathway (8)
ko00564	Glycerophospholipid metabolism (8)
ko04145	Phagosome (8)
ko04922	Glucagon signaling pathway (8)
ko03018	RNA degradation (8)
ko04140	Autophagy - animal (7)
ko00510	N-Glycan biosynthesis (7)
ko00860	Porphyrin and chlorophyll metabolism (7)
ko05165	Human papillomavirus infection (7)
ko00630	Glyoxylate and dicarboxylate metabolism (7)
ko00640	Propanoate metabolism (7)
ko00900	Terpenoid backbone biosynthesis (7)
ko04218	Cellular senescence (7)
ko04260	Cardiac muscle contraction (7)
ko04728	Dopaminergic synapse (6)
ko05203	Viral carcinogenesis (6)
ko04721	Synaptic vesicle cycle (6)
ko04623	Cytosolic DNA-sensing pathway (6)
ko03015	mRNA surveillance pathway (6)
ko00480	Glutathione metabolism (6)
ko00280	Valine, leucine and isoleucine degradation (6)
ko03040	Spliceosome (6)
ko04022	cGMP-PKG signaling pathway (6)
ko05134	Legionellosis (6)
ko00520	Amino sugar and nucleotide sugar metabolism (6)
ko05167	Kaposi's sarcoma-associated herpesvirus infection (6)
ko04020	Calcium signaling pathway (6)
ko04151	PI3K-Akt signaling pathway (6)
ko00730	Thiamine metabolism (5)
ko00710	Carbon fixation in photosynthetic organisms (5)
ko04921	Oxytocin signaling pathway (5)
ko04914	Progesterone-mediated oocyte maturation (5)
ko05110	Vibrio cholerae infection (5)
ko00061	Fatty acid biosynthesis (5)
ko04070	Phosphatidylinositol signaling system (5)

ko00051	Fructose and mannose metabolism (5)
ko04066	HIF-1 signaling pathway (5)
ko04371	Apelin signaling pathway (5)
ko00260	Glycine, serine and threonine metabolism (5)
ko04142	Lysosome (5)
ko05034	Alcoholism (5)
ko00720	Carbon fixation pathways in prokaryotes (5)
ko00030	Pentose phosphate pathway (5)
ko00513	Various types of N-glycan biosynthesis (5)
ko00250	Alanine, aspartate and glutamate metabolism (5)
ko02020	Two-component system (5)
ko04962	Vasopressin-regulated water reabsorption (4)
ko04130	SNARE interactions in vesicular transport (4)
ko04621	NOD-like receptor signaling pathway (4)
ko00670	One carbon pool by folate (4)
ko04270	Vascular smooth muscle contraction (4)
ko04152	AMPK signaling pathway (4)
ko04918	Thyroid hormone synthesis (4)
ko04540	Gap junction (4)
ko04924	Renin secretion (4)
ko04212	Longevity regulating pathway - worm (4)
ko04137	Mitophagy - animal (4)
ko05225	Hepatocellular carcinoma (4)
ko00983	Drug metabolism - other enzymes (4)
ko05031	Amphetamine addiction (4)
ko04530	Tight junction (4)
ko04360	Axon guidance (4)
ko04810	Regulation of actin cytoskeleton (4)
ko04341	Hedgehog signaling pathway - fly (4)
ko04139	Mitophagy - yeast (4)
ko03320	PPAR signaling pathway (4)
ko01524	Platinum drug resistance (4)
ko04261	Adrenergic signaling in cardiomyocytes (4)
ko04910	Insulin signaling pathway (4)
ko05145	Toxoplasmosis (4)
ko00450	Selenocompound metabolism (4)
ko04011	MAPK signaling pathway - yeast (4)
ko05168	Herpes simplex infection (4)
ko04136	Autophagy - other (4)
ko05322	Systemic lupus erythematosus (4)
ko04626	Plant-pathogen interaction (4)
ko04010	MAPK signaling pathway (4)
ko04720	Long-term potentiation (4)
ko05230	Central carbon metabolism in cancer (4)
ko04024	cAMP signaling pathway (3)
ko04713	Circadian entrainment (3)

ko04390	Hippo signaling pathway (3)
ko02024	Quorum sensing (3)
ko00910	Nitrogen metabolism (3)
ko05210	Colorectal cancer (3)
ko00130	Ubiquinone and other terpenoid-quinone biosynthesis (3)
ko04380	Osteoclast differentiation (3)
ko05418	Fluid shear stress and atherosclerosis (3)
ko04722	Neurotrophin signaling pathway (3)
ko04662	B cell receptor signaling pathway (3)
ko05132	Salmonella infection (3)
ko01040	Biosynthesis of unsaturated fatty acids (3)
ko04740	Olfactory transduction (3)
ko05130	Pathogenic Escherichia coli infection (3)
ko04970	Salivary secretion (3)
ko04068	FoxO signaling pathway (3)
ko05226	Gastric cancer (3)
ko04666	Fc gamma R-mediated phagocytosis (3)
ko04072	Phospholipase D signaling pathway (3)
ko04611	Platelet activation (3)
ko00790	Folate biosynthesis (3)
ko00562	Inositol phosphate metabolism (3)
ko04915	Estrogen signaling pathway (3)
ko04916	Melanogenesis (3)
ko04350	TGF-beta signaling pathway (3)
ko04931	Insulin resistance (3)
ko00561	Glycerolipid metabolism (3)
ko04340	Hedgehog signaling pathway (3)
ko05323	Rheumatoid arthritis (3)
ko05162	Measles (3)
ko00680	Methane metabolism (3)
ko04122	Sulfur relay system (3)
ko04657	IL-17 signaling pathway (3)
ko00310	Lysine degradation (3)
ko04724	Glutamatergic synapse (3)
ko04146	Peroxisome (3)
ko05215	Prostate cancer (3)
ko00780	Biotin metabolism (3)
ko01210	2-Oxocarboxylic acid metabolism (3)
ko04971	Gastric acid secretion (3)
ko04660	T cell receptor signaling pathway (3)
ko05161	Hepatitis B (2)
ko04214	Apoptosis - fly (2)
ko04016	MAPK signaling pathway - plant (2)
ko05133	Pertussis (2)
ko04961	Endocrine and other factor-regulated calcium reabsorption (2)
ko00052	Galactose metabolism (2)

ko04370	VEGF signaling pathway (2)
ko04510	Focal adhesion (2)
ko04062	Chemokine signaling pathway (2)
ko00740	Riboflavin metabolism (2)
ko04391	Hippo signaling pathway - fly (2)
ko04925	Aldosterone synthesis and secretion (2)
ko05014	Amyotrophic lateral sclerosis (ALS) (2)
ko04919	Thyroid hormone signaling pathway (2)
ko04150	mTOR signaling pathway (2)
ko04213	Longevity regulating pathway - multiple species (2)
ko05146	Amoebiasis (2)
ko04750	Inflammatory mediator regulation of TRP channels (2)
ko04013	MAPK signaling pathway - fly (2)
ko00380	Tryptophan metabolism (2)
ko00521	Streptomycin biosynthesis (2)
ko04658	Th1 and Th2 cell differentiation (2)
ko04071	Sphingolipid signaling pathway (2)
ko00410	beta-Alanine metabolism (2)
ko04668	TNF signaling pathway (2)
ko05020	Prion diseases (2)
ko05222	Small cell lung cancer (2)
ko05100	Bacterial invasion of epithelial cells (2)
ko03070	Bacterial secretion system (2)
ko00333	Prodigiosin biosynthesis (2)
ko04710	Circadian rhythm (2)
ko03450	Non-homologous end-joining (2)
ko04659	Th17 cell differentiation (2)
ko00785	Lipoic acid metabolism (2)
ko04624	Toll and Imd signaling pathway (2)
ko04210	Apoptosis (2)
ko05120	Epithelial cell signaling in Helicobacter pylori infection (2)
ko05032	Morphine addiction (2)
ko04650	Natural killer cell mediated cytotoxicity (2)
ko05202	Transcriptional misregulation in cancer (2)
ko00760	Nicotinate and nicotinamide metabolism (2)
ko00220	Arginine biosynthesis (2)
ko00270	Cysteine and methionine metabolism (2)
ko00071	Fatty acid degradation (2)
ko05213	Endometrial cancer (2)
ko00500	Starch and sucrose metabolism (2)
ko04730	Long-term depression (2)
ko04064	NF-kappa B signaling pathway (2)
ko05160	Hepatitis C (2)
ko04912	GnRH signaling pathway (2)
ko00330	Arginine and proline metabolism (2)
ko04115	p53 signaling pathway (2)

ko04930	Type II diabetes mellitus (2)
ko04923	Regulation of lipolysis in adipocytes (2)
ko04727	GABAergic synapse (2)
ko05205	Proteoglycans in cancer (2)
ko04014	Ras signaling pathway (2)
ko05231	Choline metabolism in cancer (2)
ko04725	Cholinergic synapse (1)
ko00563	Glycosylphosphatidylinositol (GPI)-anchor biosynthesis (1)
ko00195	Photosynthesis (1)
ko05223	Non-small cell lung cancer (1)
ko04216	Ferroptosis (1)
ko04520	Adherens junction (1)
ko00750	Vitamin B6 metabolism (1)
ko04112	Cell cycle - Caulobacter (1)
ko00524	Neomycin, kanamycin and gentamicin biosynthesis (1)
ko04976	Bile secretion (1)
ko04744	Phototransduction (1)
ko04726	Serotonergic synapse (1)
ko05211	Renal cell carcinoma (1)
ko05212	Pancreatic cancer (1)
ko00040	Pentose and glucuronate interconversions (1)
ko04926	Relaxin signaling pathway (1)
ko04973	Carbohydrate digestion and absorption (1)
ko04015	Rap1 signaling pathway (1)
ko05142	Chagas disease (American trypanosomiasis) (1)
ko00650	Butanoate metabolism (1)
ko00362	Benzoate degradation (1)
ko01521	EGFR tyrosine kinase inhibitor resistance (1)
ko00072	Synthesis and degradation of ketone bodies (1)
ko04742	Taste transduction (1)
ko01522	Endocrine resistance (1)
ko05340	Primary immunodeficiency (1)
ko05224	Breast cancer (1)
ko04913	Ovarian steroidogenesis (1)
ko05217	Basal cell carcinoma (1)
ko00430	Taurine and hypotaurine metabolism (1)
ko00460	Cyanoamino acid metabolism (1)
ko04012	ErbB signaling pathway (1)
ko04917	Prolactin signaling pathway (1)
ko04920	Adipocytokine signaling pathway (1)
ko00590	Arachidonic acid metabolism (1)
ko00565	Ether lipid metabolism (1)
ko04550	Signaling pathways regulating pluripotency of stem cells (1)
ko04723	Retrograde endocannabinoid signaling (1)
ko04966	Collecting duct acid secretion (1)
ko02010	ABC transporters (1)

ko05214	Glioma (1)
ko04911	Insulin secretion (1)
ko04622	RIG-I-like receptor signaling pathway (1)
ko04974	Protein digestion and absorption (1)
ko05206	MicroRNAs in cancer (1)
ko05164	Influenza A (1)
ko05414	Dilated cardiomyopathy (DCM) (1)
ko04711	Circadian rhythm - fly (1)
ko00440	Phosphonate and phosphinate metabolism (1)
ko00062	Fatty acid elongation (1)
ko00940	Phenylpropanoid biosynthesis (1)
ko04745	Phototransduction - fly (1)
ko00400	Phenylalanine, tyrosine and tryptophan biosynthesis (1)
ko00770	Pantothenate and CoA biosynthesis (1)
ko04211	Longevity regulating pathway (1)
ko04712	Circadian rhythm - plant (1)
ko05030	Cocaine addiction (1)
ko01523	Antifolate resistance (1)
ko05131	Shigellosis (1)
ko04940	Type I diabetes mellitus (1)

C. The KEGG pathways of R-Sc

ko01100	Metabolic pathways (129)
ko03010	Ribosome (56)
ko03040	Spliceosome (55)
ko01110	Biosynthesis of secondary metabolites (53)
ko01130	Biosynthesis of antibiotics (35)
ko00230	Purine metabolism (32)
ko00240	Pyrimidine metabolism (30)
ko03008	Ribosome biogenesis in eukaryotes (29)
ko01120	Microbial metabolism in diverse environments (29)
ko03013	RNA transport (27)
ko05016	Huntington's disease (27)
ko05169	Epstein-Barr virus infection (26)
ko01200	Carbon metabolism (23)
ko03018	RNA degradation (23)
ko03030	DNA replication (21)
ko00190	Oxidative phosphorylation (19)
ko04110	Cell cycle (18)
ko05012	Parkinson's disease (17)
ko05010	Alzheimer's disease (17)
ko00970	Aminoacyl-tRNA biosynthesis (16)
ko04141	Protein processing in endoplasmic reticulum (16)
ko01230	Biosynthesis of amino acids (16)
ko04111	Cell cycle - yeast (16)

ko05144	Malaria (16)
ko03420	Nucleotide excision repair (16)
ko04113	Meiosis - yeast (14)
ko03020	RNA polymerase (13)
ko03050	Proteasome (13)
ko04120	Ubiquitin mediated proteolysis (13)
ko03015	mRNA surveillance pathway (12)
ko00020	Citrate cycle (TCA cycle) (11)
ko04144	Endocytosis (11)
ko04138	Autophagy - yeast (11)
ko05166	HTLV-I infection (11)
ko03430	Mismatch repair (11)
ko05203	Viral carcinogenesis (11)
ko04932	Non-alcoholic fatty liver disease (NAFLD) (10)
ko03060	Protein export (10)
ko00620	Pyruvate metabolism (10)
ko04114	Oocyte meiosis (9)
ko00010	Glycolysis / Gluconeogenesis (9)
ko05200	Pathways in cancer (9)
ko00480	Glutathione metabolism (8)
ko00564	Glycerophospholipid metabolism (8)
ko03440	Homologous recombination (8)
ko03410	Base excision repair (8)
ko04217	Necroptosis (8)
ko04145	Phagosome (8)
ko03022	Basal transcription factors (8)
ko04922	Glucagon signaling pathway (8)
ko04139	Mitophagy - yeast (7)
ko00640	Propanoate metabolism (7)
ko04623	Cytosolic DNA-sensing pathway (7)
ko02020	Two-component system (7)
ko04140	Autophagy - animal (7)
ko05134	Legionellosis (6)
ko05322	Systemic lupus erythematosus (6)
ko00860	Porphyrin and chlorophyll metabolism (6)
ko05152	Tuberculosis (6)
ko00280	Valine, leucine and isoleucine degradation (6)
ko04212	Longevity regulating pathway - worm (6)
ko05168	Herpes simplex infection (6)
ko04310	Wnt signaling pathway (6)
ko04260	Cardiac muscle contraction (6)
ko05167	Kaposi's sarcoma-associated herpesvirus infection (5)
ko00630	Glyoxylate and dicarboxylate metabolism (5)
ko05165	Human papillomavirus infection (5)
ko00710	Carbon fixation in photosynthetic organisms (5)
ko00720	Carbon fixation pathways in prokaryotes (5)

ko04962	Vasopressin-regulated water reabsorption (5)
ko05110	Vibrio cholerae infection (5)
ko05145	Toxoplasmosis (5)
ko04137	Mitophagy - animal (5)
ko04151	PI3K-Akt signaling pathway (5)
ko04721	Synaptic vesicle cycle (5)
ko04728	Dopaminergic synapse (4)
ko04727	GABAergic synapse (4)
ko00900	Terpenoid backbone biosynthesis (4)
ko00030	Pentose phosphate pathway (4)
ko00330	Arginine and proline metabolism (4)
ko05034	Alcoholism (4)
ko05132	Salmonella infection (4)
ko04122	Sulfur relay system (4)
ko04150	mTOR signaling pathway (4)
ko03320	PPAR signaling pathway (4)
ko05225	Hepatocellular carcinoma (4)
ko00680	Methane metabolism (4)
ko04136	Autophagy - other (4)
ko05230	Central carbon metabolism in cancer (4)
ko04214	Apoptosis - fly (4)
ko00510	N-Glycan biosynthesis (4)
ko01212	Fatty acid metabolism (4)
ko04530	Tight junction (4)
ko03460	Fanconi anemia pathway (4)
ko00270	Cysteine and methionine metabolism (4)
ko04066	HIF-1 signaling pathway (4)
ko00730	Thiamine metabolism (4)
ko05160	Hepatitis C (4)
ko01524	Platinum drug resistance (4)
ko05164	Influenza A (4)
ko04540	Gap junction (4)
ko05162	Measles (4)
ko00513	Various types of N-glycan biosynthesis (3)
ko04914	Progesterone-mediated oocyte maturation (3)
ko04921	Oxytocin signaling pathway (3)
ko00983	Drug metabolism - other enzymes (3)
ko04216	Ferroptosis (3)
ko05161	Hepatitis B (3)
ko04020	Calcium signaling pathway (3)
ko05323	Rheumatoid arthritis (3)
ko04130	SNARE interactions in vesicular transport (3)
ko04919	Thyroid hormone signaling pathway (3)
ko00130	Ubiquinone and other terpenoid-quinone biosynthesis (3)
ko05130	Pathogenic Escherichia coli infection (3)
ko04341	Hedgehog signaling pathway - fly (3)

ko00562	Inositol phosphate metabolism (3)
ko02024	Quorum sensing (3)
ko00790	Folate biosynthesis (3)
ko05146	Amoebiasis (3)
ko04371	Apelin signaling pathway (3)
ko04626	Plant-pathogen interaction (3)
ko00250	Alanine, aspartate and glutamate metabolism (3)
ko04390	Hippo signaling pathway (3)
ko04724	Glutamatergic synapse (3)
ko04810	Regulation of actin cytoskeleton (3)
ko04918	Thyroid hormone synthesis (3)
ko00450	Selenocompound metabolism (3)
ko04072	Phospholipase D signaling pathway (3)
ko04360	Axon guidance (3)
ko04340	Hedgehog signaling pathway (3)
ko05014	Amyotrophic lateral sclerosis (ALS) (3)
ko00670	One carbon pool by folate (3)
ko00260	Glycine, serine and threonine metabolism (3)
ko04657	IL-17 signaling pathway (3)
ko04152	AMPK signaling pathway (3)
ko00740	Riboflavin metabolism (3)
ko04621	NOD-like receptor signaling pathway (3)
ko04010	MAPK signaling pathway (3)
ko05210	Colorectal cancer (3)
ko04070	Phosphatidylinositol signaling system (3)
ko04210	Apoptosis (3)
ko05418	Fluid shear stress and atherosclerosis (3)
ko04011	MAPK signaling pathway - yeast (3)
ko00520	Amino sugar and nucleotide sugar metabolism (3)
ko04142	Lysosome (3)
ko04022	cGMP-PKG signaling pathway (3)
ko00561	Glycerolipid metabolism (3)
ko04910	Insulin signaling pathway (3)
ko04666	Fc gamma R-mediated phagocytosis (3)
ko00310	Lysine degradation (3)
ko04218	Cellular senescence (3)
ko04660	T cell receptor signaling pathway (2)
ko00071	Fatty acid degradation (2)
ko00760	Nicotinate and nicotinamide metabolism (2)
ko04668	TNF signaling pathway (2)
ko04270	Vascular smooth muscle contraction (2)
ko04740	Olfactory transduction (2)
ko05031	Amphetamine addiction (2)
ko04213	Longevity regulating pathway - multiple species (2)
ko04068	FoxO signaling pathway (2)
ko00410	beta-Alanine metabolism (2)

ko04261	Adrenergic signaling in cardiomyocytes (2)
ko05120	Epithelial cell signaling in Helicobacter pylori infection (2)
ko04062	Chemokine signaling pathway (2)
ko04730	Long-term depression (2)
ko05215	Prostate cancer (2)
ko04722	Neurotrophin signaling pathway (2)
ko01521	EGFR tyrosine kinase inhibitor resistance (2)
ko04391	Hippo signaling pathway - fly (2)
ko04611	Platelet activation (2)
ko04924	Renin secretion (2)
ko00220	Arginine biosynthesis (2)
ko03070	Bacterial secretion system (2)
ko04970	Salivary secretion (2)
ko05202	Transcriptional misregulation in cancer (2)
ko05231	Choline metabolism in cancer (2)
ko04931	Insulin resistance (2)
ko04380	Osteoclast differentiation (2)
ko04350	TGF-beta signaling pathway (2)
ko01210	2-Oxocarboxylic acid metabolism (2)
ko00380	Tryptophan metabolism (2)
ko04662	B cell receptor signaling pathway (2)
ko04211	Longevity regulating pathway (2)
ko04115	p53 signaling pathway (2)
ko00563	Glycosylphosphatidylinositol (GPI)-anchor biosynthesis (2)
ko04966	Collecting duct acid secretion (2)
ko00627	Aminobenzoate degradation (2)
ko04923	Regulation of lipolysis in adipocytes (2)
ko05205	Proteoglycans in cancer (2)
ko04713	Circadian entrainment (2)
ko05032	Morphine addiction (2)
ko04915	Estrogen signaling pathway (2)
ko05222	Small cell lung cancer (2)
ko04024	cAMP signaling pathway (2)
ko04916	Melanogenesis (2)
ko05020	Prion diseases (2)
ko04720	Long-term potentiation (2)
ko01040	Biosynthesis of unsaturated fatty acids (2)
ko00910	Nitrogen metabolism (2)
ko04146	Peroxisome (2)
ko00500	Starch and sucrose metabolism (1)
ko01522	Endocrine resistance (1)
ko00362	Benzoate degradation (1)
ko04650	Natural killer cell mediated cytotoxicity (1)
ko04971	Gastric acid secretion (1)
ko00514	Other types of O-glycan biosynthesis (1)
ko04550	Signaling pathways regulating pluripotency of stem cells (1)

ko05133	Pertussis (1)
ko05131	Shigellosis (1)
ko00590	Arachidonic acid metabolism (1)
ko04976	Bile secretion (1)
ko00770	Pantothenate and CoA biosynthesis (1)
ko05223	Non-small cell lung cancer (1)
ko04064	NF-kappa B signaling pathway (1)
ko05416	Viral myocarditis (1)
ko04920	Adipocytokine signaling pathway (1)
ko00980	Metabolism of xenobiotics by cytochrome P450 (1)
ko00062	Fatty acid elongation (1)
ko05414	Dilated cardiomyopathy (DCM) (1)
ko00400	Phenylalanine, tyrosine and tryptophan biosynthesis (1)
ko05100	Bacterial invasion of epithelial cells (1)
ko04330	Notch signaling pathway (1)
ko00195	Photosynthesis (1)
ko04926	Relaxin signaling pathway (1)
ko04925	Aldosterone synthesis and secretion (1)
ko04712	Circadian rhythm - plant (1)
ko04520	Adherens junction (1)
ko01523	Antifolate resistance (1)
ko05212	Pancreatic cancer (1)
ko04726	Serotonergic synapse (1)
ko04016	MAPK signaling pathway - plant (1)
ko00051	Fructose and mannose metabolism (1)
ko00750	Vitamin B6 metabolism (1)
ko04913	Ovarian steroidogenesis (1)
ko04725	Cholinergic synapse (1)
ko05030	Cocaine addiction (1)
ko04013	MAPK signaling pathway - fly (1)
ko04974	Protein digestion and absorption (1)
ko00440	Phosphonate and phosphinate metabolism (1)
ko04071	Sphingolipid signaling pathway (1)
ko01051	Biosynthesis of ansamycins (1)
ko04012	ErbB signaling pathway (1)
ko00565	Ether lipid metabolism (1)
ko04370	VEGF signaling pathway (1)
ko00140	Steroid hormone biosynthesis (1)
ko05224	Breast cancer (1)
ko02010	ABC transporters (1)
ko05204	Chemical carcinogenesis (1)
ko04917	Prolactin signaling pathway (1)
ko04014	Ras signaling pathway (1)
ko04912	GnRH signaling pathway (1)
ko04710	Circadian rhythm (1)
ko04510	Focal adhesion (1)

ko00982	Drug metabolism - cytochrome P450 (1)
ko00650	Butanoate metabolism (1)
ko04711	Circadian rhythm - fly (1)
ko00061	Fatty acid biosynthesis (1)
ko04658	Th1 and Th2 cell differentiation (1)
ko04659	Th17 cell differentiation (1)
ko05217	Basal cell carcinoma (1)
ko05213	Endometrial cancer (1)
ko04622	RIG-I-like receptor signaling pathway (1)
ko04940	Type I diabetes mellitus (1)
ko04742	Taste transduction (1)
ko03450	Non-homologous end-joining (1)
ko00460	Cyanoamino acid metabolism (1)
ko04911	Insulin secretion (1)
ko00908	Zeatin biosynthesis (1)
ko05142	Chagas disease (American trypanosomiasis) (1)
ko00940	Phenylpropanoid biosynthesis (1)
ko04112	Cell cycle - Caulobacter (1)
ko04723	Retrograde endocannabinoid signaling (1)
ko00920	Sulfur metabolism (1)
ko05226	Gastric cancer (1)
ko00072	Synthesis and degradation of ketone bodies (1)
ko04215	Apoptosis - multiple species (1)
ko04961	Endocrine and other factor-regulated calcium reabsorption (1)
ko04750	Inflammatory mediator regulation of TRP channels (1)
ko04624	Toll and Imd signaling pathway (1)
ko04930	Type II diabetes mellitus (1)

D. The KEGG pathways of R-GII

ko01100	Metabolic pathways (137)
ko01110	Biosynthesis of secondary metabolites (56)
ko03040	Spliceosome (51)
ko01130	Biosynthesis of antibiotics (39)
ko03010	Ribosome (38)
ko01120	Microbial metabolism in diverse environments (29)
ko03008	Ribosome biogenesis in eukaryotes (27)
ko05169	Epstein-Barr virus infection (27)
ko03013	RNA transport (27)
ko01200	Carbon metabolism (25)
ko00190	Oxidative phosphorylation (24)
ko04141	Protein processing in endoplasmic reticulum (23)
ko00230	Purine metabolism (23)
ko05016	Huntington's disease (22)
ko05010	Alzheimer's disease (20)
ko00970	Aminoacyl-tRNA biosynthesis (20)

ko00240	Pyrimidine metabolism (19)
ko05012	Parkinson's disease (18)
ko04120	Ubiquitin mediated proteolysis (17)
ko03018	RNA degradation (16)
ko01230	Biosynthesis of amino acids (15)
ko04145	Phagosome (14)
ko04138	Autophagy - yeast (13)
ko03050	Proteasome (13)
ko00010	Glycolysis / Gluconeogenesis (13)
ko05144	Malaria (13)
ko03420	Nucleotide excision repair (13)
ko04144	Endocytosis (11)
ko05110	Vibrio cholerae infection (11)
ko05203	Viral carcinogenesis (11)
ko04721	Synaptic vesicle cycle (11)
ko03430	Mismatch repair (10)
ko03020	RNA polymerase (10)
ko05152	Tuberculosis (10)
ko03015	mRNA surveillance pathway (10)
ko04932	Non-alcoholic fatty liver disease (NAFLD) (10)
ko05200	Pathways in cancer (9)
ko03030	DNA replication (9)
ko04110	Cell cycle (9)
ko03060	Protein export (9)
ko04922	Glucagon signaling pathway (9)
ko04114	Oocyte meiosis (9)
ko05166	HTLV-I infection (9)
ko00620	Pyruvate metabolism (9)
ko00860	Porphyrin and chlorophyll metabolism (8)
ko05323	Rheumatoid arthritis (8)
ko00520	Amino sugar and nucleotide sugar metabolism (8)
ko04142	Lysosome (8)
ko00640	Propanoate metabolism (8)
ko03440	Homologous recombination (7)
ko05120	Epithelial cell signaling in Helicobacter pylori infection (7)
ko03410	Base excision repair (7)
ko04260	Cardiac muscle contraction (7)
ko00900	Terpenoid backbone biosynthesis (7)
ko05167	Kaposi's sarcoma-associated herpesvirus infection (7)
ko02020	Two-component system (7)
ko04310	Wnt signaling pathway (7)
ko05134	Legionellosis (7)
ko00710	Carbon fixation in photosynthetic organisms (7)
ko05165	Human papillomavirus infection (7)
ko00250	Alanine, aspartate and glutamate metabolism (7)
ko00030	Pentose phosphate pathway (7)

ko00020	Citrate cycle (TCA cycle) (7)
ko04140	Autophagy - animal (7)
ko05164	Influenza A (6)
ko04921	Oxytocin signaling pathway (6)
ko00630	Glyoxylate and dicarboxylate metabolism (6)
ko00280	Valine, leucine and isoleucine degradation (6)
ko04966	Collecting duct acid secretion (6)
ko04150	mTOR signaling pathway (6)
ko04111	Cell cycle - yeast (6)
ko04623	Cytosolic DNA-sensing pathway (6)
ko05168	Herpes simplex infection (6)
ko05034	Alcoholism (6)
ko00680	Methane metabolism (6)
ko00051	Fructose and mannose metabolism (5)
ko04962	Vasopressin-regulated water reabsorption (5)
ko04010	MAPK signaling pathway (5)
ko04218	Cellular senescence (5)
ko05162	Measles (5)
ko04113	Meiosis - yeast (5)
ko00720	Carbon fixation pathways in prokaryotes (5)
ko04066	HIF-1 signaling pathway (5)
ko00260	Glycine, serine and threonine metabolism (5)
ko04371	Apelin signaling pathway (5)
ko04070	Phosphatidylinositol signaling system (5)
ko04728	Dopaminergic synapse (5)
ko00480	Glutathione metabolism (5)
ko00510	N-Glycan biosynthesis (5)
ko04910	Insulin signaling pathway (5)
ko04020	Calcium signaling pathway (5)
ko05322	Systemic lupus erythematosus (5)
ko04720	Long-term potentiation (4)
ko04217	Necroptosis (4)
ko01212	Fatty acid metabolism (4)
ko04213	Longevity regulating pathway - multiple species (4)
ko04013	MAPK signaling pathway - fly (4)
ko00270	Cysteine and methionine metabolism (4)
ko01524	Platinum drug resistance (4)
ko03460	Fanconi anemia pathway (4)
ko04022	cGMP-PKG signaling pathway (4)
ko04530	Tight junction (4)
ko04011	MAPK signaling pathway - yeast (4)
ko05145	Toxoplasmosis (4)
ko04139	Mitophagy - yeast (4)
ko05160	Hepatitis C (4)
ko04136	Autophagy - other (4)
ko00513	Various types of N-glycan biosynthesis (4)

ko04724	Glutamatergic synapse (4)
ko04130	SNARE interactions in vesicular transport (4)
ko04924	Renin secretion (4)
ko04360	Axon guidance (4)
ko05210	Colorectal cancer (4)
ko04919	Thyroid hormone signaling pathway (4)
ko04810	Regulation of actin cytoskeleton (4)
ko05031	Amphetamine addiction (4)
ko00730	Thiamine metabolism (4)
ko05132	Salmonella infection (4)
ko05014	Amyotrophic lateral sclerosis (ALS) (4)
ko04727	GABAergic synapse (4)
ko05230	Central carbon metabolism in cancer (4)
ko00562	Inositol phosphate metabolism (4)
ko04212	Longevity regulating pathway - worm (3)
ko05020	Prion diseases (3)
ko04915	Estrogen signaling pathway (3)
ko03022	Basal transcription factors (3)
ko04660	T cell receptor signaling pathway (3)
ko05418	Fluid shear stress and atherosclerosis (3)
ko04722	Neurotrophin signaling pathway (3)
ko04216	Ferroptosis (3)
ko05202	Transcriptional misregulation in cancer (3)
ko04540	Gap junction (3)
ko04214	Apoptosis - fly (3)
ko00740	Riboflavin metabolism (3)
ko04261	Adrenergic signaling in cardiomyocytes (3)
ko04210	Apoptosis (3)
ko04137	Mitophagy - animal (3)
ko04626	Plant-pathogen interaction (3)
ko04662	B cell receptor signaling pathway (3)
ko00450	Selenocompound metabolism (3)
ko05146	Amoebiasis (3)
ko00500	Starch and sucrose metabolism (3)
ko05161	Hepatitis B (3)
ko04270	Vascular smooth muscle contraction (3)
ko04341	Hedgehog signaling pathway - fly (3)
ko00760	Nicotinate and nicotinamide metabolism (3)
ko00564	Glycerophospholipid metabolism (3)
ko00983	Drug metabolism - other enzymes (3)
ko00220	Arginine biosynthesis (3)
ko00910	Nitrogen metabolism (3)
ko04971	Gastric acid secretion (3)
ko00790	Folate biosynthesis (3)
ko04918	Thyroid hormone synthesis (3)
ko05130	Pathogenic Escherichia coli infection (3)

ko04380	Osteoclast differentiation (3)
ko04666	Fc gamma R-mediated phagocytosis (3)
ko04390	Hippo signaling pathway (3)
ko04916	Melanogenesis (3)
ko04151	PI3K-Akt signaling pathway (3)
ko00330	Arginine and proline metabolism (2)
ko04912	GnRH signaling pathway (2)
ko04391	Hippo signaling pathway - fly (2)
ko04510	Focal adhesion (2)
ko04072	Phospholipase D signaling pathway (2)
ko00770	Pantothenate and CoA biosynthesis (2)
ko05131	Shigellosis (2)
ko02024	Quorum sensing (2)
ko04750	Inflammatory mediator regulation of TRP channels (2)
ko04930	Type II diabetes mellitus (2)
ko04350	TGF-beta signaling pathway (2)
ko04713	Circadian entrainment (2)
ko05100	Bacterial invasion of epithelial cells (2)
ko04650	Natural killer cell mediated cytotoxicity (2)
ko03320	PPAR signaling pathway (2)
ko04370	VEGF signaling pathway (2)
ko04068	FoxO signaling pathway (2)
ko05226	Gastric cancer (2)
ko04658	Th1 and Th2 cell differentiation (2)
ko04330	Notch signaling pathway (2)
ko00071	Fatty acid degradation (2)
ko04024	cAMP signaling pathway (2)
ko04115	p53 signaling pathway (2)
ko00130	Ubiquinone and other terpenoid-quinone biosynthesis (2)
ko04624	Toll and Imd signaling pathway (2)
ko00670	One carbon pool by folate (2)
ko04152	AMPK signaling pathway (2)
ko04014	Ras signaling pathway (2)
ko04657	IL-17 signaling pathway (2)
ko04122	Sulfur relay system (2)
ko04659	Th17 cell differentiation (2)
ko00061	Fatty acid biosynthesis (2)
ko04621	NOD-like receptor signaling pathway (2)
ko05206	MicroRNAs in cancer (2)
ko05213	Endometrial cancer (2)
ko05225	Hepatocellular carcinoma (2)
ko01040	Biosynthesis of unsaturated fatty acids (2)
ko04970	Salivary secretion (2)
ko04064	NF-kappa B signaling pathway (2)
ko00521	Streptomycin biosynthesis (2)
ko04340	Hedgehog signaling pathway (2)

ko04062	Chemokine signaling pathway (2)
ko05133	Pertussis (2)
ko04611	Platelet activation (2)
ko04961	Endocrine and other factor-regulated calcium reabsorption (2)
ko04925	Aldosterone synthesis and secretion (2)
ko00563	Glycosylphosphatidylinositol (GPI)-anchor biosynthesis (2)
ko05205	Proteoglycans in cancer (2)
ko04740	Olfactory transduction (2)
ko04914	Progesterone-mediated oocyte maturation (2)
ko00052	Galactose metabolism (2)
ko04725	Cholinergic synapse (1)
ko03070	Bacterial secretion system (1)
ko05142	Chagas disease (American trypanosomiasis) (1)
ko00195	Photosynthesis (1)
ko00650	Butanoate metabolism (1)
ko04146	Peroxisome (1)
ko04926	Relaxin signaling pathway (1)
ko04622	RIG-I-like receptor signaling pathway (1)
ko01523	Antifolate resistance (1)
ko04612	Antigen processing and presentation (1)
ko04211	Longevity regulating pathway (1)
ko04712	Circadian rhythm - plant (1)
ko04973	Carbohydrate digestion and absorption (1)
ko05215	Prostate cancer (1)
ko04976	Bile secretion (1)
ko05212	Pancreatic cancer (1)
ko00785	Lipoic acid metabolism (1)
ko04917	Prolactin signaling pathway (1)
ko04071	Sphingolipid signaling pathway (1)
ko00514	Other types of O-glycan biosynthesis (1)
ko00627	Aminobenzoate degradation (1)
ko05222	Small cell lung cancer (1)
ko00310	Lysine degradation (1)
ko00072	Synthesis and degradation of ketone bodies (1)
ko04923	Regulation of lipolysis in adipocytes (1)
ko00362	Benzoate degradation (1)
ko01521	EGFR tyrosine kinase inhibitor resistance (1)
ko04016	MAPK signaling pathway - plant (1)
ko04723	Retrograde endocannabinoid signaling (1)
ko04742	Taste transduction (1)
ko00333	Prodigiosin biosynthesis (1)
ko04931	Insulin resistance (1)
ko04392	Hippo signaling pathway - multiple species (1)
ko04730	Long-term depression (1)
ko05217	Basal cell carcinoma (1)
ko05214	Glioma (1)

ko04520	Adherens junction (1)
ko05416	Viral myocarditis (1)
ko01210	2-Oxocarboxylic acid metabolism (1)
ko05220	Chronic myeloid leukemia (1)
ko00590	Arachidonic acid metabolism (1)
ko04911	Insulin secretion (1)
ko04920	Adipocytokine signaling pathway (1)
ko03450	Non-homologous end-joining (1)
ko05340	Primary immunodeficiency (1)
ko00920	Sulfur metabolism (1)
ko00780	Biotin metabolism (1)
ko04215	Apoptosis - multiple species (1)
ko04711	Circadian rhythm - fly (1)
ko00524	Neomycin, kanamycin and gentamicin biosynthesis (1)
ko05032	Morphine addiction (1)
ko05414	Dilated cardiomyopathy (DCM) (1)
ko04012	ErbB signaling pathway (1)
ko04745	Phototransduction - fly (1)
ko04550	Signaling pathways regulating pluripotency of stem cells (1)
ko05030	Cocaine addiction (1)
ko00062	Fatty acid elongation (1)
ko05231	Choline metabolism in cancer (1)
ko01051	Biosynthesis of ansamycins (1)
ko01522	Endocrine resistance (1)
ko04913	Ovarian steroidogenesis (1)
ko00460	Cyanoamino acid metabolism (1)
ko00908	Zeatin biosynthesis (1)
ko04015	Rap1 signaling pathway (1)
ko00430	Taurine and hypotaurine metabolism (1)
ko00940	Phenylpropanoid biosynthesis (1)
ko00380	Tryptophan metabolism (1)
ko00040	Pentose and glucuronate interconversions (1)
ko00410	beta-Alanine metabolism (1)
ko05211	Renal cell carcinoma (1)
ko04726	Serotonergic synapse (1)
ko05224	Breast cancer (1)
ko00561	Glycerolipid metabolism (1)
ko04744	Phototransduction (1)
ko04710	Circadian rhythm (1)

E. The KEGG pathways of R-GV

ko01100	Metabolic pathways (96)
ko01110	Biosynthesis of secondary metabolites (41)
ko03010	Ribosome (35)
ko03040	Spliceosome (34)

ko01130	Biosynthesis of antibiotics (25)
ko03008	Ribosome biogenesis in eukaryotes (20)
ko00230	Purine metabolism (20)
ko01120	Microbial metabolism in diverse environments (19)
ko03013	RNA transport (17)
ko05016	Huntington's disease (17)
ko00240	Pyrimidine metabolism (17)
ko05169	Epstein-Barr virus infection (16)
ko01200	Carbon metabolism (16)
ko00190	Oxidative phosphorylation (15)
ko00970	Aminoacyl-tRNA biosynthesis (14)
ko03018	RNA degradation (13)
ko05010	Alzheimer's disease (12)
ko05012	Parkinson's disease (12)
ko03030	DNA replication (11)
ko03420	Nucleotide excision repair (11)
ko01230	Biosynthesis of amino acids (10)
ko03015	mRNA surveillance pathway (10)
ko04141	Protein processing in endoplasmic reticulum (10)
ko00010	Glycolysis / Gluconeogenesis (9)
ko04120	Ubiquitin mediated proteolysis (9)
ko05144	Malaria (8)
ko05203	Viral carcinogenesis (8)
ko04110	Cell cycle (8)
ko04144	Endocytosis (7)
ko04111	Cell cycle - yeast (7)
ko04138	Autophagy - yeast (7)
ko00860	Porphyrin and chlorophyll metabolism (7)
ko03020	RNA polymerase (7)
ko03430	Mismatch repair (7)
ko03440	Homologous recombination (7)
ko00620	Pyruvate metabolism (6)
ko04113	Meiosis - yeast (6)
ko05166	HTLV-I infection (6)
ko05134	Legionellosis (6)
ko00640	Propanoate metabolism (6)
ko04217	Necroptosis (6)
ko03050	Proteasome (6)
ko00520	Amino sugar and nucleotide sugar metabolism (6)
ko05034	Alcoholism (6)
ko05165	Human papillomavirus infection (6)
ko00250	Alanine, aspartate and glutamate metabolism (5)
ko05200	Pathways in cancer (5)
ko04721	Synaptic vesicle cycle (5)
ko04260	Cardiac muscle contraction (5)
ko00680	Methane metabolism (5)

ko04145	Phagosome (5)
ko04932	Non-alcoholic fatty liver disease (NAFLD) (5)
ko00730	Thiamine metabolism (5)
ko00480	Glutathione metabolism (5)
ko03060	Protein export (5)
ko03022	Basal transcription factors (5)
ko03410	Base excision repair (5)
ko00983	Drug metabolism - other enzymes (4)
ko04139	Mitophagy - yeast (4)
ko05110	Vibrio cholerae infection (4)
ko00260	Glycine, serine and threonine metabolism (4)
ko05322	Systemic lupus erythematosus (4)
ko04114	Oocyte meiosis (4)
ko04142	Lysosome (4)
ko00030	Pentose phosphate pathway (4)
ko04140	Autophagy - animal (4)
ko00280	Valine, leucine and isoleucine degradation (4)
ko00630	Glyoxylate and dicarboxylate metabolism (4)
ko05168	Herpes simplex infection (4)
ko00051	Fructose and mannose metabolism (4)
ko00020	Citrate cycle (TCA cycle) (4)
ko00900	Terpenoid backbone biosynthesis (4)
ko04066	HIF-1 signaling pathway (4)
ko04727	GABAergic synapse (3)
ko04922	Glucagon signaling pathway (3)
ko04013	MAPK signaling pathway - fly (3)
ko05202	Transcriptional misregulation in cancer (3)
ko05162	Measles (3)
ko05230	Central carbon metabolism in cancer (3)
ko04150	mTOR signaling pathway (3)
ko00220	Arginine biosynthesis (3)
ko04623	Cytosolic DNA-sensing pathway (3)
ko04966	Collecting duct acid secretion (3)
ko05323	Rheumatoid arthritis (3)
ko05145	Toxoplasmosis (3)
ko02020	Two-component system (3)
ko04070	Phosphatidylinositol signaling system (3)
ko00450	Selenocompound metabolism (3)
ko04213	Longevity regulating pathway - multiple species (3)
ko03460	Fanconi anemia pathway (3)
ko05164	Influenza A (3)
ko00790	Folate biosynthesis (3)
ko04136	Autophagy - other (3)
ko00510	N-Glycan biosynthesis (3)
ko05120	Epithelial cell signaling in Helicobacter pylori infection (3)
ko04137	Mitophagy - animal (3)

ko00513	Various types of N-glycan biosynthesis (3)
ko04918	Thyroid hormone synthesis (3)
ko04621	NOD-like receptor signaling pathway (3)
ko00910	Nitrogen metabolism (3)
ko04310	Wnt signaling pathway (2)
ko04214	Apoptosis - fly (2)
ko04216	Ferroptosis (2)
ko00670	One carbon pool by folate (2)
ko00720	Carbon fixation pathways in prokaryotes (2)
ko00760	Nicotinate and nicotinamide metabolism (2)
ko05167	Kaposi's sarcoma-associated herpesvirus infection (2)
ko04152	AMPK signaling pathway (2)
ko00561	Glycerolipid metabolism (2)
ko04919	Thyroid hormone signaling pathway (2)
ko04212	Longevity regulating pathway - worm (2)
ko04371	Apelin signaling pathway (2)
ko05152	Tuberculosis (2)
ko00500	Starch and sucrose metabolism (2)
ko04668	TNF signaling pathway (2)
ko04921	Oxytocin signaling pathway (2)
ko04657	IL-17 signaling pathway (2)
ko04330	Notch signaling pathway (2)
ko00270	Cysteine and methionine metabolism (2)
ko04728	Dopaminergic synapse (2)
ko04122	Sulfur relay system (2)
ko00785	Lipoic acid metabolism (2)
ko04930	Type II diabetes mellitus (2)
ko04218	Cellular senescence (2)
ko04350	TGF-beta signaling pathway (2)
ko00710	Carbon fixation in photosynthetic organisms (2)
ko04011	MAPK signaling pathway - yeast (2)
ko00564	Glycerophospholipid metabolism (2)
ko05160	Hepatitis C (2)
ko05418	Fluid shear stress and atherosclerosis (2)
ko01212	Fatty acid metabolism (2)
ko00562	Inositol phosphate metabolism (2)
ko04626	Plant-pathogen interaction (2)
ko04530	Tight junction (2)
ko02024	Quorum sensing (2)
ko04261	Adrenergic signaling in cardiomyocytes (2)
ko04910	Insulin signaling pathway (2)
ko04962	Vasopressin-regulated water reabsorption (2)
ko04915	Estrogen signaling pathway (2)
ko05161	Hepatitis B (2)
ko04750	Inflammatory mediator regulation of TRP channels (1)
ko05014	Amyotrophic lateral sclerosis (ALS) (1)

ko00524	Neomycin, kanamycin and gentamicin biosynthesis (1)
ko05133	Pertussis (1)
ko04912	GnRH signaling pathway (1)
ko03070	Bacterial secretion system (1)
ko04341	Hedgehog signaling pathway - fly (1)
ko05231	Choline metabolism in cancer (1)
ko05210	Colorectal cancer (1)
ko04612	Antigen processing and presentation (1)
ko04744	Phototransduction (1)
ko01210	2-Oxocarboxylic acid metabolism (1)
ko00330	Arginine and proline metabolism (1)
ko04730	Long-term depression (1)
ko04064	NF-kappa B signaling pathway (1)
ko05142	Chagas disease (American trypanosomiasis) (1)
ko04112	Cell cycle - Caulobacter (1)
ko04022	cGMP-PKG signaling pathway (1)
ko04390	Hippo signaling pathway (1)
ko04024	cAMP signaling pathway (1)
ko05220	Chronic myeloid leukemia (1)
ko00908	Zeatin biosynthesis (1)
ko05205	Proteoglycans in cancer (1)
ko04973	Carbohydrate digestion and absorption (1)
ko00430	Taurine and hypotaurine metabolism (1)
ko04520	Adherens junction (1)
ko00130	Ubiquinone and other terpenoid-quinone biosynthesis (1)
ko00062	Fatty acid elongation (1)
ko00627	Aminobenzoate degradation (1)
ko04010	MAPK signaling pathway (1)
ko04740	Olfactory transduction (1)
ko04015	Rap1 signaling pathway (1)
ko04071	Sphingolipid signaling pathway (1)
ko04391	Hippo signaling pathway - fly (1)
ko05020	Prion diseases (1)
ko04624	Toll and Imd signaling pathway (1)
ko04916	Melanogenesis (1)
ko01524	Platinum drug resistance (1)
ko02010	ABC transporters (1)
ko04020	Calcium signaling pathway (1)
ko05340	Primary immunodeficiency (1)
ko04540	Gap junction (1)
ko00590	Arachidonic acid metabolism (1)
ko04151	PI3K-Akt signaling pathway (1)
ko04016	MAPK signaling pathway - plant (1)
ko05130	Pathogenic Escherichia coli infection (1)
ko05031	Amphetamine addiction (1)
ko04925	Aldosterone synthesis and secretion (1)

ko04931	Insulin resistance (1)
ko04072	Phospholipase D signaling pathway (1)
ko05214	Glioma (1)
ko04724	Glutamatergic synapse (1)
ko04068	FoxO signaling pathway (1)
ko04720	Long-term potentiation (1)
ko05131	Shigellosis (1)
ko00410	beta-Alanine metabolism (1)
ko04710	Circadian rhythm (1)
ko03320	PPAR signaling pathway (1)
ko01523	Antifolate resistance (1)
ko04014	Ras signaling pathway (1)
ko04666	Fc gamma R-mediated phagocytosis (1)
ko00052	Galactose metabolism (1)
ko04115	p53 signaling pathway (1)
ko04971	Gastric acid secretion (1)
ko00460	Cyanoamino acid metabolism (1)
ko04270	Vascular smooth muscle contraction (1)
ko04210	Apoptosis (1)
ko00071	Fatty acid degradation (1)
ko04920	Adipocytokine signaling pathway (1)
ko04924	Renin secretion (1)
ko00940	Phenylpropanoid biosynthesis (1)
ko04970	Salivary secretion (1)
ko04713	Circadian entrainment (1)
ko04745	Phototransduction - fly (1)
ko05225	Hepatocellular carcinoma (1)
ko00740	Riboflavin metabolism (1)
ko00061	Fatty acid biosynthesis (1)
ko00195	Photosynthesis (1)
ko01040	Biosynthesis of unsaturated fatty acids (1)
ko00521	Streptomycin biosynthesis (1)
ko04722	Neurotrophin signaling pathway (1)
ko04712	Circadian rhythm - plant (1)
ko05211	Renal cell carcinoma (1)
ko04622	RIG-I-like receptor signaling pathway (1)
ko04130	SNARE interactions in vesicular transport (1)
ko00770	Pantothenate and CoA biosynthesis (1)
ko04146	Peroxisome (1)
ko04810	Regulation of actin cytoskeleton (1)
ko03450	Non-homologous end-joining (1)

F. The KEGG pathways of R-Oo

ko01100	Metabolic pathways (132)
ko01110	Biosynthesis of secondary metabolites (60)

ko03010	Ribosome (47)
ko01130	Biosynthesis of antibiotics (44)
ko03040	Spliceosome (36)
ko01120	Microbial metabolism in diverse environments (31)
ko00230	Purine metabolism (31)
ko01200	Carbon metabolism (28)
ko03013	RNA transport (27)
ko00240	Pyrimidine metabolism (24)
ko03030	DNA replication (23)
ko05016	Huntington's disease (23)
ko03008	Ribosome biogenesis in eukaryotes (22)
ko03018	RNA degradation (22)
ko03420	Nucleotide excision repair (20)
ko04141	Protein processing in endoplasmic reticulum (19)
ko00970	Aminoacyl-tRNA biosynthesis (18)
ko00190	Oxidative phosphorylation (18)
ko04110	Cell cycle (18)
ko05010	Alzheimer's disease (17)
ko04111	Cell cycle - yeast (17)
ko05169	Epstein-Barr virus infection (16)
ko04120	Ubiquitin mediated proteolysis (16)
ko05012	Parkinson's disease (15)
ko05144	Malaria (14)
ko04113	Meiosis - yeast (13)
ko01230	Biosynthesis of amino acids (13)
ko00010	Glycolysis / Gluconeogenesis (13)
ko05166	HTLV-I infection (13)
ko00620	Pyruvate metabolism (13)
ko03430	Mismatch repair (12)
ko04144	Endocytosis (12)
ko03060	Protein export (11)
ko03015	mRNA surveillance pathway (11)
ko03410	Base excision repair (10)
ko00020	Citrate cycle (TCA cycle) (10)
ko05200	Pathways in cancer (10)
ko04114	Oocyte meiosis (9)
ko03020	RNA polymerase (9)
ko03440	Homologous recombination (9)
ko04217	Necroptosis (8)
ko04932	Non-alcoholic fatty liver disease (NAFLD) (8)
ko00640	Propanoate metabolism (8)
ko04145	Phagosome (8)
ko00564	Glycerophospholipid metabolism (8)
ko03460	Fanconi anemia pathway (8)
ko05134	Legionellosis (8)
ko03050	Proteasome (8)

ko00860	Porphyrin and chlorophyll metabolism (7)
ko00250	Alanine, aspartate and glutamate metabolism (7)
ko00710	Carbon fixation in photosynthetic organisms (7)
ko04310	Wnt signaling pathway (7)
ko04922	Glucagon signaling pathway (7)
ko01212	Fatty acid metabolism (7)
ko05167	Kaposi's sarcoma-associated herpesvirus infection (6)
ko04138	Autophagy - yeast (6)
ko00260	Glycine, serine and threonine metabolism (6)
ko04066	HIF-1 signaling pathway (6)
ko05152	Tuberculosis (6)
ko05203	Viral carcinogenesis (6)
ko00280	Valine, leucine and isoleucine degradation (6)
ko00630	Glyoxylate and dicarboxylate metabolism (6)
ko00520	Amino sugar and nucleotide sugar metabolism (6)
ko05145	Toxoplasmosis (6)
ko00680	Methane metabolism (6)
ko04013	MAPK signaling pathway - fly (5)
ko04721	Synaptic vesicle cycle (5)
ko01524	Platinum drug resistance (5)
ko05164	Influenza A (5)
ko04142	Lysosome (5)
ko00480	Glutathione metabolism (5)
ko04022	cGMP-PKG signaling pathway (5)
ko04212	Longevity regulating pathway - worm (5)
ko04218	Cellular senescence (5)
ko03022	Basal transcription factors (5)
ko04260	Cardiac muscle contraction (5)
ko00900	Terpenoid backbone biosynthesis (5)
ko05230	Central carbon metabolism in cancer (5)
ko00061	Fatty acid biosynthesis (5)
ko05168	Herpes simplex infection (5)
ko00720	Carbon fixation pathways in prokaryotes (5)
ko05165	Human papillomavirus infection (5)
ko00030	Pentose phosphate pathway (5)
ko05034	Alcoholism (5)
ko00051	Fructose and mannose metabolism (4)
ko04657	IL-17 signaling pathway (4)
ko04728	Dopaminergic synapse (4)
ko04360	Axon guidance (4)
ko04020	Calcium signaling pathway (4)
ko04010	MAPK signaling pathway (4)
ko00670	One carbon pool by folate (4)
ko04130	SNARE interactions in vesicular transport (4)
ko04962	Vasopressin-regulated water reabsorption (4)
ko05110	Vibrio cholerae infection (4)

ko05014	Amyotrophic lateral sclerosis (ALS) (4)
ko04910	Insulin signaling pathway (4)
ko04918	Thyroid hormone synthesis (4)
ko05418	Fluid shear stress and atherosclerosis (4)
ko05225	Hepatocellular carcinoma (4)
ko05162	Measles (4)
ko00510	N-Glycan biosynthesis (4)
ko04626	Plant-pathogen interaction (4)
ko04921	Oxytocin signaling pathway (4)
ko02020	Two-component system (4)
ko04213	Longevity regulating pathway - multiple species (4)
ko04070	Phosphatidylinositol signaling system (4)
ko04914	Progesterone-mediated oocyte maturation (3)
ko04139	Mitophagy - yeast (3)
ko00740	Riboflavin metabolism (3)
ko00563	Glycosylphosphatidylinositol (GPI)-anchor biosynthesis (3)
ko04350	TGF-beta signaling pathway (3)
ko05160	Hepatitis C (3)
ko05132	Salmonella infection (3)
ko04660	T cell receptor signaling pathway (3)
ko04919	Thyroid hormone signaling pathway (3)
ko00333	Prodigiosin biosynthesis (3)
ko04530	Tight junction (3)
ko05031	Amphetamine addiction (3)
ko00760	Nicotinate and nicotinamide metabolism (3)
ko04150	mTOR signaling pathway (3)
ko04136	Autophagy - other (3)
ko00730	Thiamine metabolism (3)
ko04915	Estrogen signaling pathway (3)
ko00513	Various types of N-glycan biosynthesis (3)
ko04623	Cytosolic DNA-sensing pathway (3)
ko04214	Apoptosis - fly (3)
ko04540	Gap junction (3)
ko05210	Colorectal cancer (3)
ko05202	Transcriptional misregulation in cancer (3)
ko00983	Drug metabolism - other enzymes (3)
ko04140	Autophagy - animal (3)
ko04924	Renin secretion (3)
ko04341	Hedgehog signaling pathway - fly (3)
ko00561	Glycerolipid metabolism (3)
ko05322	Systemic lupus erythematosus (3)
ko05323	Rheumatoid arthritis (3)
ko00790	Folate biosynthesis (3)
ko04146	Peroxisome (3)
ko04662	B cell receptor signaling pathway (3)
ko04210	Apoptosis (3)

ko04621	NOD-like receptor signaling pathway (3)
ko04931	Insulin resistance (3)
ko04151	PI3K-Akt signaling pathway (3)
ko00310	Lysine degradation (3)
ko04137	Mitophagy - animal (3)
ko04380	Osteoclast differentiation (3)
ko00910	Nitrogen metabolism (3)
ko00450	Selenocompound metabolism (3)
ko04720	Long-term potentiation (3)
ko02024	Quorum sensing (2)
ko04390	Hippo signaling pathway (2)
ko05120	Epithelial cell signaling in Helicobacter pylori infection (2)
ko04016	MAPK signaling pathway - plant (2)
ko05133	Pertussis (2)
ko04740	Olfactory transduction (2)
ko00627	Aminobenzoate degradation (2)
ko04216	Ferroptosis (2)
ko03320	PPAR signaling pathway (2)
ko05215	Prostate cancer (2)
ko01040	Biosynthesis of unsaturated fatty acids (2)
ko00770	Pantothenate and CoA biosynthesis (2)
ko04730	Long-term depression (2)
ko05020	Prion diseases (2)
ko04340	Hedgehog signaling pathway (2)
ko05130	Pathogenic Escherichia coli infection (2)
ko04650	Natural killer cell mediated cytotoxicity (2)
ko04668	TNF signaling pathway (2)
ko04115	p53 signaling pathway (2)
ko04930	Type II diabetes mellitus (2)
ko04011	MAPK signaling pathway - yeast (2)
ko00780	Biotin metabolism (2)
ko04370	VEGF signaling pathway (2)
ko04713	Circadian entrainment (2)
ko00750	Vitamin B6 metabolism (2)
ko04068	FoxO signaling pathway (2)
ko04624	Toll and Imd signaling pathway (2)
ko04970	Salivary secretion (2)
ko04724	Glutamatergic synapse (2)
ko05161	Hepatitis B (2)
ko04371	Apelin signaling pathway (2)
ko04270	Vascular smooth muscle contraction (2)
ko00500	Starch and sucrose metabolism (2)
ko00220	Arginine biosynthesis (2)
ko04152	AMPK signaling pathway (2)
ko00380	Tryptophan metabolism (2)
ko04916	Melanogenesis (2)

ko04722	Neurotrophin signaling pathway (2)
ko04658	Th1 and Th2 cell differentiation (2)
ko04261	Adrenergic signaling in cardiomyocytes (2)
ko05222	Small cell lung cancer (2)
ko00071	Fatty acid degradation (2)
ko04330	Notch signaling pathway (2)
ko04710	Circadian rhythm (2)
ko04659	Th17 cell differentiation (2)
ko04711	Circadian rhythm - fly (1)
ko00062	Fatty acid elongation (1)
ko04520	Adherens junction (1)
ko04923	Regulation of lipolysis in adipocytes (1)
ko00362	Benzoate degradation (1)
ko05223	Non-small cell lung cancer (1)
ko00460	Cyanoamino acid metabolism (1)
ko03450	Non-homologous end-joining (1)
ko00410	beta-Alanine metabolism (1)
ko05142	Chagas disease (American trypanosomiasis) (1)
ko04611	Platelet activation (1)
ko04122	Sulfur relay system (1)
ko04666	Fc gamma R-mediated phagocytosis (1)
ko01521	EGFR tyrosine kinase inhibitor resistance (1)
ko04012	ErbB signaling pathway (1)
ko05211	Renal cell carcinoma (1)
ko04744	Phototransduction (1)
ko00590	Arachidonic acid metabolism (1)
ko05224	Breast cancer (1)
ko04973	Carbohydrate digestion and absorption (1)
ko04072	Phospholipase D signaling pathway (1)
ko00072	Synthesis and degradation of ketone bodies (1)
ko05212	Pancreatic cancer (1)
ko04211	Longevity regulating pathway (1)
ko05205	Proteoglycans in cancer (1)
ko04550	Signaling pathways regulating pluripotency of stem cells (1)
ko05340	Primary immunodeficiency (1)
ko04064	NF-kappa B signaling pathway (1)
ko04810	Regulation of actin cytoskeleton (1)
ko04112	Cell cycle - Caulobacter (1)
ko04966	Collecting duct acid secretion (1)
ko04750	Inflammatory mediator regulation of TRP channels (1)
ko05220	Chronic myeloid leukemia (1)
ko04745	Phototransduction - fly (1)
ko04612	Antigen processing and presentation (1)
ko00052	Galactose metabolism (1)
ko05214	Glioma (1)
ko00920	Sulfur metabolism (1)

ko04024	cAMP signaling pathway (1)
ko01051	Biosynthesis of ansamycins (1)
ko04920	Adipocytokine signaling pathway (1)
ko00040	Pentose and glucuronate interconversions (1)
ko04917	Prolactin signaling pathway (1)
ko04912	GnRH signaling pathway (1)
ko00270	Cysteine and methionine metabolism (1)
ko04940	Type I diabetes mellitus (1)
ko00521	Streptomycin biosynthesis (1)
ko05226	Gastric cancer (1)
ko00330	Arginine and proline metabolism (1)
ko00140	Steroid hormone biosynthesis (1)
ko05217	Basal cell carcinoma (1)
ko04215	Apoptosis - multiple species (1)
ko04510	Focal adhesion (1)
ko01210	2-Oxocarboxylic acid metabolism (1)
ko00562	Inositol phosphate metabolism (1)
ko04971	Gastric acid secretion (1)
ko00430	Taurine and hypotaurine metabolism (1)
ko04062	Chemokine signaling pathway (1)
ko00400	Phenylalanine, tyrosine and tryptophan biosynthesis (1)
ko04727	GABAergic synapse (1)
ko04071	Sphingolipid signaling pathway (1)
ko04015	Rap1 signaling pathway (1)
ko01523	Antifolate resistance (1)
ko05231	Choline metabolism in cancer (1)
ko00130	Ubiquinone and other terpenoid-quinone biosynthesis (1)
ko04974	Protein digestion and absorption (1)
ko04925	Aldosterone synthesis and secretion (1)
ko04712	Circadian rhythm - plant (1)
ko00524	Neomycin, kanamycin and gentamicin biosynthesis (1)
ko02010	ABC transporters (1)
ko05213	Endometrial cancer (1)
ko03070	Bacterial secretion system (1)
ko00195	Photosynthesis (1)
ko04014	Ras signaling pathway (1)
ko00650	Butanoate metabolism (1)
ko00940	Phenylpropanoid biosynthesis (1)
ko05416	Viral myocarditis (1)
ko04391	Hippo signaling pathway - fly (1)

Appendix V

Table A5: Compounds which follows the Lipinski's rule of five

Compounds	Formula	Molecular weight	H-bond acceptors	H-bond donors	TPSA (Topological surface area)	iLOGP (n-octanol/water partition coefficient)
ZINC000022910880	C21H20N6O	372.42	3	4	118.95	2.39
ZINC000000001963	C14H14ClN3O2S	323.8	4	2	100.41	2.18
ZINC000000000740	C16H13ClN2O2	300.74	3	1	52.9	2.13
ZINC000040863182	C16H15N7O	321.34	5	2	121.67	2.33
ZINC000019804668	C22H24ClN3OS	413.96	3	0	52.09	4.16
ZINC000001999515	C21H25NO2	323.43	3	1	40.54	2.88
ZINC000115619865	C21H24FN5O4S	461.51	8	4	150.85	2.73
ZINC000000004594	C19H20N2O3	324.37	4	1	64.35	2.03
ZINC000008584337	C15H13N3O3S	315.35	4	2	106.52	2.61
ZINC000003628643	C17H19NO3	285.34	4	2	52.93	2.51
ZINC000002036738	C18H24O2	272.38	2	2	40.46	2.63
ZINC000013982572	C29H27N5O	461.56	4	2	85.83	3.34
ZINC000000006694	C16H14N2O3S	314.36	5	1	94.57	1.78
ZINC000002019958	C16H18N2	238.33	1	1	15.27	2.57
ZINC000000607731	C14H14ClN3O4S2	387.86	6	3	135.12	0.51
ZINC000003830351	C26H37N5O2	451.6	4	2	71.68	3.54
ZINC000000001282	C12H8Cl2N2O2	283.11	3	2	62.22	2.05
ZINC000022942298	C30H35FN2O3	490.61	6	0	42.01	5.07
ZINC000000001181	C14H9Cl2N3O3	338.15	4	1	74.91	1.99
ZINC000095936819	C20H19F5N2O	398.37	7	2	37.05	3.04
ZINC000000000365	C16H23NO2	261.36	3	1	38.33	3.54
ZINC000001493878	C21H16ClF3N4O3	464.82	7	3	92.35	3.45
ZINC000000968274	C20H31NO	301.47	2	1	23.47	3.72
ZINC000000537891	C18H20FN3O4	361.37	6	1	75.01	2.49
ZINC000039341568	C24H36O5	404.54	5	1	72.83	4.15
ZINC000000002193	C10H12ClN3	209.68	1	2	36.42	1.91
ZINC000000608261	C20H24ClN3O2	373.88	3	2	67.59	3.14
ZINC000001693537	C9H9N3O5	239.18	6	0	100.86	0.98
ZINC000000001656	C15H11ClN2O	270.71	2	0	34.89	2.73
ZINC000000000587	C18H24ClN3O2	349.86	3	2	67.59	2.96
ZINC000000000133	C11H12ClNO3S	273.74	3	0	62.83	1.73
ZINC000003784384	C21H30N4O4	402.49	5	2	113.41	3.03
ZINC000011616841	C23H29NO3	367.48	4	1	49.77	3.82
ZINC000000000806	C20H23NO3	325.4	4	1	49.77	3.17
ZINC000004214702	C20H19N3O4S	397.45	5	1	107.75	2.53
ZINC000000004503	C16H21N3	255.36	1	1	27.63	2.63
ZINC000002017397	C22H25NO3	351.44	4	0	38.77	3.69

ZINC000004215333	C23H29ClN4O3	444.95	4	2	73.91	2.89
ZINC000000004354	C15H22N2O2	262.35	3	3	57.28	2.64
ZINC000002008866	C16H20FN3O4	337.35	5	1	71.11	2.31
ZINC000000601288	C19H18ClN3O3	371.82	4	0	62.21	3.12
ZINC000100030989	C20H15F3N4O3	416.35	8	2	101.45	2.37
ZINC000001544908	C16H21N3O3	303.36	4	1	73.22	3.12
ZINC000033903720	C18H29NO	275.43	2	1	23.47	3.41
ZINC000003806063	C25H32F3N3O4	495.53	8	3	97.05	3.46
ZINC000038418475	C16H22O2	246.34	2	1	37.3	2.53
ZINC000001551732	C11H16N2O3S	256.32	4	2	78.02	1.13
ZINC000001842900	C22H32NO3	358.49	3	1	46.53	0.06
ZINC000000001498	C12H11N3O2	229.23	4	1	67.49	2.16
ZINC000000001095	C10H10N2O3	206.2	3	1	72.63	1.49
ZINC000006037116	C14H14O2S	246.32	2	0	54.54	3.01
ZINC000003786299	C17H18F3N3O3	369.34	7	1	65.78	2.27
ZINC000004214151	C22H27NO2	337.46	3	1	38.33	3.74
ZINC000002038967	C11H16ClN3O4S2	353.85	6	3	135.12	0.39
ZINC000034035805	C13H11F6N3O5	403.23	11	3	124.25	1.32
ZINC000000006126	C14H20ClN3O2	297.78	5	1	57.95	2.73
ZINC000000049153	C12H17N4OS	265.35	3	2	104.15	-1.6
ZINC000000000507	C12H18N2O4	254.28	5	3	93.81	1.19
ZINC000000538202	C21H27N3O3	369.46	5	0	64.55	3.92
ZINC000018191874	C11H15BrN2O3	303.15	3	2	75.27	1.7
ZINC000000000394	C10H16N4O4	256.26	5	2	109.69	1.27
ZINC000000000808	C18H22N2S	298.45	1	0	31.78	3.39
ZINC000003776651	C19H15F3N2OS	376.4	5	0	52.35	3.17
ZINC000000000413	C18H31NO2	293.44	3	1	32.7	4.01
ZINC000001686103	C20H25NO	295.42	2	1	21.26	3.47
ZINC000038197764	C19H22FN3O4	375.39	6	2	83.8	2.44
ZINC000004213023	C23H38ClN3O	408.02	3	1	49.57	3.95
ZINC000000001380	C11H15NO4S	257.31	5	1	83.06	1.7
ZINC000001842903	C8H21NO6P	258.23	6	3	106.03	-1.87
ZINC000000001128	C14H18ClN3S	295.83	2	0	47.61	3.2
ZINC000002018421	C17H28N2O2	292.42	3	1	41.57	3.98
ZINC000001530569	C18H31NO4	325.44	5	2	59.95	4.14
ZINC000058438005	C8H10N2OS	182.24	2	2	69.37	1.66
ZINC000019942898	C14H32N2O4	292.41	6	4	87.4	2.69
ZINC000000895559	C6H9N3O	139.16	3	2	72.03	1.31
ZINC000001529994	C6H10N3O4P	219.14	6	3	128.37	0.19
ZINC000008219605	C6H11N3O7P2	299.11	9	4	184.71	-0.31
ZINC000263607103	C17H18N2O4	314.34	5	2	98.64	2.31
ZINC000000160790	C6H9NOS	143.21	2	1	61.36	1.65
ZINC000001578333	C22H20ClN5O2	421.88	5	1	84.15	4.32
ZINC000003869379	C2H5O5P	140.03	5	2	93.64	-0.3
ZINC000003870145	C3H5O6P	168.04	6	3	113.87	-0.64
ZINC000003869774	C4H10N3O5P	211.11	6	4	146.26	-0.8

ZINC000001532839	C12H18N4O4PS	345.33	6	3	160.49	-3.47
NSC-1614	C27H42O5	446.62	5	2	91.67	3.02
NSC-5476	C18H26N2O2S	334.48	3	1	57.79	3.06
NSC-7578	C21H13NO4	343.33	4	2	83.47	2.01
NSC-19061	C18H14Cl2N6	385.25	3	2	67.66	3.48
NSC-37168	C17H12N2O4	308.29	4	2	95.15	1.78
ZINC000004720969	C21H17NO5	363.36	5	2	80.93	2.93
NSC-96996	C22H17N3O	339.39	3	0	47.78	3.23
NSC-112541	C18H16N2O2	292.33	3	2	65.98	2.49
NSC-116709	C24H19NO2	353.41	2	0	37.38	3.12
NSC-156565	C22H17NO4	359.37	4	3	86.63	2.94
NSC-201631	C21H16N4O5S	436.44	6	2	179.27	2.93
ZINC000000615883	C17H12N2O4	308.29	4	2	95.15	1.89
NSC-338963	C22H20ClN5O2	421.88	5	1	84.15	4.32
NSC-359472	C20H19N5O2	361.4	5	2	96.87	2.86

PUBLICATION

[ISSN 0253-7613]
Volume 51 | Issue 6 | November-December 2019
Impact Factor ® as reported in the 2018 Journal Citation Reports ® (Clarivate Analytics, 2019): 1.040

IJP | INDIAN JOURNAL OF PHARMACOLOGY

Official Publication of The Indian Pharmacological Society (IPS)

www.indianpharmacology.org
website: http://www.ijp-online.com

Wolters Kluwer

Medknow

Research Article

Access this article online

Quick Response Code:

Website: www.ijp-online.com

DOI: 10.4103/ijp.IJP_535_19

Investigation of hub genes and their nonsynonymous single nucleotide polymorphism analysis in *Plasmodium falciparum* for designing therapeutic methodologies using next-generation sequencing approach

Sanjay Kumar Singh, Sudhakara M. Reddy

Abstract:

BACKGROUND: Incidences of resistance to current drugs by *Plasmodium* is increasing, hence, it is necessary to investigate and explore new drug targets to combat malarial disease.

OBJECTIVE: Analysis of the transcriptome sequence information to characterize hub genes and their nonsynonymous single nucleotide polymorphisms (nsSNPs) to derive therapeutic objectives for *Plasmodium falciparum*.

MATERIALS AND METHODS: Differentially expressed genes between Ring and other stages of *P. falciparum* were identified using Cufflinks tool. Using DAVID and KAAS programs, the gene ontology and pathway analysis were performed. The networks of protein-protein interaction (PPI) were developed by Search Tool for the Retrieval of Interacting Genes/Proteins and Cytoscape, and the node degree in the network was calculated by using Network Analyzer, and MCODE plugins of Cytoscape. SIFT, PROVEAN, and PredictSNP programs were used to study the genetic variations, which affect protein functions.

RESULTS: A list of 4196 nonredundant genes was used for functional annotation cluster analysis, and 8 significant hub genes have been picked from the PPI network using MCODE plugins of Cytoscape. Various nsSNPs were identified in these 8 hub genes and were investigated both for its native and mutant stage for solvent accessibility and alteration in secondary structure protein residues.

CONCLUSION: Hub genes identified in this study serve as potential targets to develop therapy to suppress the pathogenic action of *P. falciprum* through experimental techniques.

Keywords:

Bioinformatics analysis, deleterious mutation, hub genes, malaria, next-generation sequencing approach, non-synonymous single nucleotide polymorphisms, *Plasmodium falciparum*

Department of Biotechnology, Thapar Institute of Engineering and Technology, Patiala, Punjab, India

Address for correspondence:
Dr. Sudhakara M. Reddy, Department of Biotechnology, Thapar Institute of Engineering and Technology, Patiala - 147 004, Punjab, India.
E-mail: msreddy@thapar.edu

Received: 18 September 2019
Revised: 18 October 2019
Accepted: 23 December 2019
Published: 16 January 2020

Introduction

Malaria, a deadly infectious disease, is caused by intracellular single-celled parasites belongs to the genus *Plasmodium*. According to World Malaria Report 2018 by World Health Organization (WHO), about 219 million malaria cases and 435,000 deaths were reported in 87 nations in 2017. In the human genome, it is recognized as one of the most powerful evolutionary selection. *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium knowlesi*, *Plasmodium malariae*,

How to cite this article: Singh SK, Reddy SM. Investigation of hub genes and their nonsynonymous single nucleotide polymorphism analysis in *Plasmodium falciparum* for designing therapeutic methodologies using next-generation sequencing approach. Indian J Pharmacol 2019;51:389-99.

and *Plasmodium ovale* are different species that cause disease in humans. The most severe malignant malaria particularly in children below 5 years is *P. falciparum*.[1] In India, almost 95% population lives in endemic areas of malaria, and 80% of recorded malaria in India is limited to 20% of the population areas who lives in rural, hilly, remote, or difficult-to-reach areas.[2] Traditional first-line therapies such as chloroquine and pyrimethamine/sulfadoxine have lost their efficacy in most countries, which resulted in the development of new and more effective anti-malarial medicines, like artemisinin-based combination therapy.[3] Although efforts were made to develop vaccines and medicines to fight malaria, there is still a problem with vaccine escape and drug resistance.[4] What remains in dearth is how plasmodium genetic variation can result in drug resistance or can provide new drug targets. It has been shown that genetic variation and recombination accelerate antigen heterogeneity, immune escape, the production of anti-malarial drug resistance, and comparison of entire genomes may aid in these efforts.[5] RNA-seq is a method that can use next-generation sequencing to analyze the quantity and sequences of RNA in a sample. Knowing the transcriptome is essential to linking the genome data to its functional protein expression. RNA-seq technology enables accurate gene isoform identification, translocation events, nucleotide mutations, and posttranscriptional modifications.

In *P. falciparum*, the genetic diversity exists in the form of single nucleotide polymorphism (SNPs), microsatellite repeats, insertions, deletions, and a variety of gene duplication. Several studies have been reported to investigate the SNPs of *P. falciparum*.[6,7] Nevertheless, to the best of our acquaintance, no research to date has used the impact of deleterious SNPs and solvent accessibility and secondary structure change of protein residues at the native and mutant level in *P. falciparum*.

López-Barragán *et al*. 2011 have sequenced seven bidirectional and four strand specific libraries for the analysis of gene expression and antisense transcripts. RNA-Seq data sets of seven bidirectional libraries were used in the study. Therefore, network analysis was performed to recognize the main hubs from the Ring (R), Schizont (Sc), gametocyte stage V (GV), gametocyte stage II (GII), early trophozoite (ET), Late trophozoite (LT), and Ookinete (Oo) stages of *P. falciprum*. In these prospective hubs, we have examined SNPs which can be proposed as key areas for vulnerability to affect the function of protein. The areas projected in this research can be further targeted to develop therapy to suppress the pathogenic action of *P. falciprum* through experimental therapeutic techniques such as gene knockout method and gene targeting.

Materials and Methods

Datasets

The RNA-Seq dataset was obtained from NCBI website, https://www.ncbi.nlm.nih.gov/sra/for all stages with accession number SRP009370[8] for analysis. Seven bidirectional reads from the 3D7 parasite have been taken for further investigation. The non-synonymous SNPs (nsSNPs) information of the *P. falciparum* genes were collected from PlasmoDB database (https://plasmodb.org/plasmo/). The protein sequence was obtained from the UniProtKB database (https://www.uniprot.org/) in the FASTA format.

Sequence quality control and preprocessing

The quality of sequence data obtained from high throughput sequencing pipelines were checked using FASTQC tool (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) was followed to clean the reads with low base-call quality using a quality filter tool.

Sequence reads alignment and transcript assembly

The reference genome of *P. falciparum* 3D7 (PlasmoDB version 7.1) was used to map the sequence reads with TopHat tool version: 2.0.14 (https://ccb.jhu.edu/software/tophat/).[9] The minimum anchor length was seven base pairs for reads present at each side and a maximum size of intron 800 bp. The output of TopHat was filtered to keep only reads mapped from Ring to gametocyte stages with 0 mismatches and up to 1 mismatch in Oo stage to maximize the accuracy. The Cufflinks (http://cole-trapnell-lab.github.io/cufflinks/)[9] tools were used to further evaluate the matched reads using a multifasta files (*plasmodium_falciparum*.fa), which improves the reliability of abundance of transcript by detecting bias and an algorithm for correction. The relative affluence of transcripts has been depicted as fragments/kilobase of exon/million fragments mapped (FPKM) and the fold change of the FPKM value between Ring (R) and other stage was reported. Poorly expressed genes were removed from the dataset by eliminating genes with FPKM value <2 in all the stages. A list of nonredundant genes which differentially expressed was created from the RNA-Seq data by combining all identified genes of Ring and other stages and the duplicate gene were removed. Fold change ratio was calculated for R versus ET, R versus LT, R versus Sc, R versus GII, R versus GV and R vs Oo groups. Further, genes which are expressed differentially were manually checked in the PlasmoDB[10] and UniProt database.[11]

Functional analysis and pathway mapping

DAVID web server[12] was used to identify and select significantly enriched gene ontology (GO) terms and

pathways. The functional annotation was determined with DAVID program (http://david.abcc.ncifcrf.gov/home.jsp). Those terms with count number of ≥5 genes, and $P \leq 0.05$ was chosen for analysis. In DAVID, the terms GO cellular component (CC), biological process (BP), and molecular function (MF) were used to classify improved biological topics in lists of genes which differentially expressed. The KEGG Automatic Annotation Server (KAAS) has been used to map the pathway.[13] Using the best single-directional hit method for orthology assignment, the amino acid sequences of genes which were up regulated and under regulated were submitted as input to the KAAS server. KAAS offers functional gene annotation in the database KEGG GENES through a similarity search tool of BLAST for a manually curated orthology group sets. For datasets mapped to one of reference pathways of KEGG, KAAS assigned a KEGG Orthology (KO) number to the genes.

Investigation of protein–protein interactions

The Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) program was used to study the functional relationships between differentially expressed genes (DEGs).[14] The GO terms – MF, CC, BP, and pathways of KEGG were used to create the interaction network using reference organism *P. falciparum*. The physical and functional interactions, coexpression, colocation, pathways, predicted interactions, and protein domain similarity are included in the network relationship between genes. The network has been filtered by eliminating all interactions with weights below 0.1.

Identification of hub genes

Cytoscape 3.4.0,[15] a data integration and network visualization bioinformatics package, has been used to identify hub genes by measuring distribution of node degrees via Network Analyzer plugin. Clusters were created using Cytoscape molecular complex detection (MCODE) plug-in. It detects regions, which are highly interconnected in the network.[16] The statistical criteria for MCODE are as follows: K score = 2, Cutoff degree = 2, Cutoff node score = 0.2 as the default. In the current network, top 5 genes with the largest distribution have been regarded as hubs.

Prediction of deleterious nonsynonymous single nucleotide polymorphisms of hub genes

PlasmoDB database was used to extract the information of SNPs for hub genes identified for *P. falciparum*. Deleterious coding nsSNPs were predicted by using PROVEAN, SIFT, and PredictSNP. SIFT and PROVEAN are tools to determine how amino acid substitution affects protein function in case the score is lower than threshold value. The PROVEAN tool was used to perform BLAST hits clustering. For each supporting sequence, a delta alignment score was calculated. To calculate the final PROVEAN score, average scores are generated across clusters.[17] Deleterious is considered a score of −2.5 or higher, when anything below this cut-off rating, which has an effect neutral. SIFT tool based on the physical properties of amino acids and homology of sequences determines whether substitution of an amino acid can influence the function of protein.[18] The sequence tool SIFT makes SIFT predictions for a particular sequence of proteins in FASTA format. The sequence of protein queries and interest substitutions of nsSNPs and hub genes with default parameters have been submitted to http://sift.bii.a-star.edu.sg/www/SIFT_seq_submit2.html. SIFT program predicts substitutions with values <0.05 to be detrimental. PredictSNP[19] has been explicitly designed to combine the projected outcomes of several tools to form a prediction of consensus. For the predictions, prediction tools use a list of variants in the protein sequence as input.

Protein solvent accessibility and prediction of secondary structure

To predict protein accessible surface area (ASA), surface accessibility and secondary structure, NetSurfP program, http://www.cbs.dtu.dk/services/NetSurfP was used. The NetSurfP simultaneously predicts accuracy for each prediction by calculating the Z-score. It uses two types of neural networks, the first type networks are based on secondary structure predictions and sequence profiles, with two outputs with respect to buried or exposed and in combination with sequence profiles, the other networks use these outputs as inputs and are trained to assess the relative surface exposure of each amino acid residues.[20] The normal and predicted SNP sequences in FASTA format have been submitted for prediction to the NetSurfP. For prediction of secondary structure and ASA, protein encoding gene in normal and its predicted SNP substitutions have been uploaded individually to NetSurfP web server. Microsoft Excel® 2016 has converted the most predicted secondary structure probabilities from NetSurfP to a single letter code representing helical (H), β-strand (E), and coil (C).

Results

The RNA-Seq dataset for all the stages of *P. falciprum* such as Ring SRR364849, ET SRR364848, LT SRR364847, Schizont SRR364843, GII SRR364840, GV SRR364838, and Oo SRR364834 has downloaded from NCBI SRA for analysis. SRA Toolkit modules, vdb-validate and fastq-dump were used to validate the integrity of downloaded SRA data and to convert SRA data into fastq format, respectively. FASTX Toolkit, a quality filter tool, was used to clean the reads with low base-call quality. Bowtie index was built from a set of DNA sequences of

P. falciparum chromosomes. RNA-seq reads for every stage were mapped for *P. falciparum* genome using TopHat program. The output of TopHat was filtered to keep only reads, which were mapped with 0 mismatches from ring to gametocyte stages, and up to 1 mismatch in Oo stage to maximize the accuracy.

Prediction of differentially expressed genes

Transcripts were assembled and analyzed in FPKM for their relative expression levels by Cufflinks tool after sequencing reads mapped to the TopHat reference genome. Between Ring and other stages, fold change analysis was done to compare the gene expression. The data with FPKM values equivalent to zero were removed, and remaining values were subjected to further analysis. As a result, 7517 genes from ring (R), 6799 genes from ET, 7482 genes from LT, 5102 genes from schizont (Sc), 8731 genes from gametocyte stages (GII), 8831 genes from gametocyte stages (GV), and 5155 genes from Oo stages were identified. Then, six groups – RvET, RvLT, RvSc, RvGII, RvGV, and RvOo were created, and the common gene in each group has been identified [Table 1].

The common genes between RvET-4402; RvLT-4354; RvSc-4022; RvGII-3917; RvGV-2988; and RvOo– 4009 have been identified. The fold change in each group was estimated using these common genes, which was defined as the FPKM value ratio of RvET, RvLT, RvSc, RvGII, RvGV, and RvOo groups. To classify the DEGs, less expressed genes with FPKM value <2 have been removed from the dataset in all the stages. On the basis of above criteria, there are 2442 DEGs between RvET, 2796 between RvLT, 2935 between RvSc, 2807 between RvGII, 2180 between RvGV, and 2895 between RvOo groups were sorted out for the analysis [Table 1]. The complete workflow for the analysis is depicted in Figure 1.

Functional annotations and pathway analyses

The functional annotation cluster analysis was conducted by DAVID tool on the DEGs. The GO terms – BP, MF, and CC were used for interpretation and only those terms with count number of ≥ 5 genes and $P \leq 0.05$ were selected. Five top GO terms with significant P values for each group from functional analysis are presented in Table 2. Only those gene IDs which were mapped via DAVID tool is used for further study. From this data, it is clear that the GO terms are enriched in RvET2209, RvLT2546, and RvGV1990 genes, which represent the functions necessary for host cell plasma membrane, antigenic variation, and pathogenesis [Table 2]. The enhanced GO terms in RvSc2735 including functions related to pathogenesis, single organismal cell–cell adhesion, receptor activity, and cell adhesion molecule binding [Table 2], and RvGII2594 include functions related to single organismal cell-cell adhesion, pathogenesis, receptor activity, and cell adhesion molecule binding. Similarly, functions-related single organismal cell–cell adhesion, pathogenesis, and cell adhesion molecule binding are enriched in RvOo2726 samples [Table 2].

Table 1: Genes which are differentially expressed in different stages of *Plasmodium falciparum*

Stages	Common genes	DEGs (≥2 fold)		
		Upregulated	Downregulated	Total
RvET	4402	731	1711	2442
RvLT	4354	1895	901	2796
RvSc	4022	1249	1686	2935
RvGII	3917	897	1910	2807
RvGV	2988	1116	1064	2180
RvOo	4009	2070	825	2895

DEGs=Differentially expressed genes

The KAAS has been used to map the pathways of DEGs in all six stages. DEGs amino acid sequences in FASTA format were submitted to the KAAS server. As a result, 277 pathways were predicted for RvET2209, 283 pathways for RvLT2546, 280 pathways for RvSc2735, 285 pathways for RvGII2594, 229 pathways for RvGV1990, and 272 pathways for RvOo2726 were recognized. The top 10 KEGG pathways for each six categories are shown in Table 3, and Supplementary Table 1 provides a complete list of pathways. It was observed from Table 3 that most DEGs have been linked with important biological processes, many of which are classified as metabolic pathways, secondary metabolite production pathways, ribosome, or being involved in biosynthesis of antibiotics.

Identification of hub genes

STRING program has been used to investigate DEGs interaction. Only those genes showing significant interactions with weights higher than 0.4 were selected for network analysis. A network between DEGs was constructed for all the six groups, namely RvET2209,

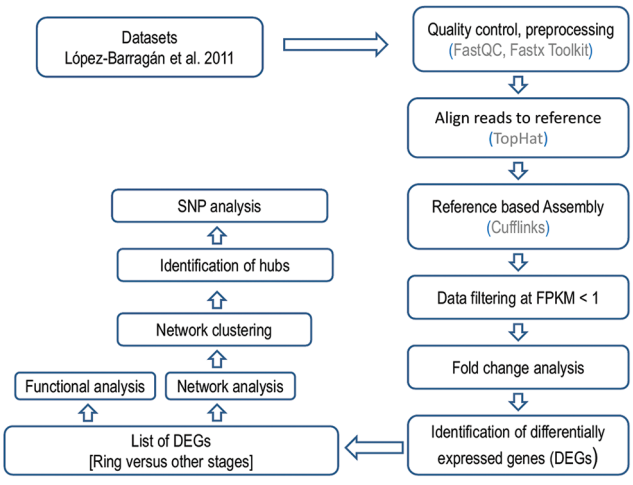


Figure 1: The flow chart depicting the overall methodology adopted in this study

Table 2: Five top improved gene ontology terms detected by DAVID in genes expressed differentially between a) RvET2209, b) RvLT2546, c) RvSc2735, d) RvGII2594, e) RvGV1990, and f) RvOo2726

Gene Ontology (GO) term	Total number of genes	*P* value
a) RvET2209		
GO: 0009405~pathogenesis	60	1.50E-11
GO: 0004872~receptor activity	55	2.49E-10
GO: 0020033~antigenic variation	112	2.61E-10
GO: 0020002~host cell plasma membrane	122	7.27E-09
GO: 0020013~modulation by symbiont of host erythrocyte aggregation	55	1.09E-07
b) RvLT2546		
GO: 0016337~single organismal cell-cell adhesion	55	1.25E-08
GO: 0009405~pathogenesis	57	8.41E-06
GO: 0004872~receptor activity	53	1.69E-05
GO: 0020033~antigenic variation	111	2.22E-04
GO: 0020002~host cell plasma membrane	117	0.00279306
c) RvSc2735		
GO: 0016337~single organismal cell-cell adhesion	51	1.41E-04
GO: 0020035~cytoadherence to microvasculature, mediated by symbiont protein	44	4.92E-04
GO: 0050839~cell adhesion molecule binding	46	0.001862072
GO: 0009405~pathogenesis	53	0.008046292
GO: 0004872~receptor activity	49	0.050202845
d) RvGII2594		
GO: 0016337~single organismal cell-cell adhesion	50	7.85E-05
GO: 0020035~cytoadherence to microvasculature, mediated by symbiont protein	44	1.16E-04
GO: 0050839~cell adhesion molecule binding	45	6.05E-04
GO: 0004872~receptor activity	49	0.00977134
GO: 0009405~pathogenesis	50	0.015754517
e) RvGV1990		
GO: 0009405~pathogenesis	56	3.05E-09
GO: 0004872~receptor activity	53	3.57E-09
GO: 0020002~host cell plasma membrane	114	3.83E-09
GO: 0020033~antigenic variation	106	2.00E-08
GO: 0020013~modulation by symbiont of host erythrocyte aggregation	53	5.99E-07
f) RvOo2726		
GO: 0016337~single organismal cell-cell adhesion	47	9.00E-04
GO: 0020035~cytoadherence to microvasculature, mediated by symbiont protein	39	0.008244114
GO: 0020030~infected host cell surface knob	38	0.008860222
GO: 0050839~cell adhesion molecule binding	40	0.04538152
GO: 0009405~pathogenesis	47	0.056291162

GO=Gene ontology

RvLT2546, RvSc2735, RvGII2594, RvGV1990, and RvOo2726. The six interaction networks resulted from STRING were then subjected to Cytoscape. The network consists of 1647 nodes and 23970 edges for RvET2209, 1969 nodes and 36152 edges for RvLT2209, 2236 nodes and 48514 edges for RvSc2735, 2010 nodes and 30113 edges for RvGII2594, 1473 nodes and 17986 edges for RvGV1990, and 2095 nodes and 37531 edges for RvOo2726. All genes in the network are represented in circles and their interactions represent edges. The interaction networks of all six groups are given in Supplementary Figure 1.

Further, all six networks have been analyzed using Network Analyzer and MCODE modules available in Cytoscape. Network Analyzer is calculating the node degrees in the network, whereas the MCODE module is creating the clusters in the network. The higher node degrees were regarded to be more significant genes and were referred as hub genes. The top 8 MCODE seed are PF3D7_1126700, PF3D7_0508100, PF3D7_1306000, PF3D7_1439500, PF3D7_0324900, PF3D7_1234300, PF3D7_1207100, and PF3D7_0705600. The interaction between these hubs and their first neighbors was presented as Figure 2.

Single nucleotide polymorphism analysis

A SNP is a significant source of variance in a genome. SNPs may result in affecting protein function by decreasing protein solubility or by destabilizing structure of protein.[21] The PlasmoDB database has been used to retrieve the nsSNPs and SNPs for the identified hub genes.

Table 3: Ten top KEGG pathways in each of the six categories

	Pathway name	Number of mapped Genes
	a) RvET	
ko01100	Metabolic pathways	102
ko01110	Biosynthesis of secondary metabolites	44
ko03010	Ribosome	43
ko01130	Biosynthesis of antibiotics	32
ko01120	Microbial metabolism in diverse environments	20
ko00230	Purine metabolism	17
ko01200	Carbon metabolism	17
ko00240	Pyrimidine metabolism	14
ko05144	Malaria	13
ko04141	Protein processing in endoplasmic reticulum	12
	b) RvLT	
ko01100	Metabolic pathways	146
ko01110	Biosynthesis of secondary metabolites	63
ko01130	Biosynthesis of antibiotics	43
ko00230	Purine metabolism	33
ko01120	Microbial metabolism in diverse environments	27
ko05169	Epstein-Barr virus infection	26
ko00240	Pyrimidine metabolism	26
ko01200	Carbon metabolism	26
ko03030	DNA replication	24
ko00190	Oxidative phosphorylation	23
	c) RvSc	
ko01100	Metabolic pathways	129
ko03010	Ribosome	56
ko03040	Spliceosome	55
ko01110	Biosynthesis of secondary metabolites	53
ko01130	Biosynthesis of antibiotics	35
ko00230	Purine metabolism	32
ko00240	Pyrimidine metabolism	30
ko03008	Ribosome biogenesis in eukaryotes	29
ko01120	Microbial metabolism in diverse environments	29
ko03013	RNA transport	27
	d) RvGII	
ko01100	Metabolic pathways	137
ko01110	Biosynthesis of secondary metabolites	56
ko03040	Spliceosome	51
ko01130	Biosynthesis of antibiotics	39
ko03010	Ribosome	38
ko01120	Microbial metabolism in diverse environments	29
ko03008	Ribosome biogenesis in eukaryotes	27
ko05169	Epstein-Barr virus infection	27
ko03013	RNA transport	27
ko01200	Carbon metabolism	25
	e) RvGV	
ko01100	Metabolic pathways	96
ko01110	Biosynthesis of secondary metabolites	41
ko03010	Ribosome	35
ko03040	Spliceosome	34
ko01130	Biosynthesis of antibiotics	25
ko03008	Ribosome biogenesis in eukaryotes	20
ko00230	Purine metabolism	20

Contd...

Table 3: Contd...

	Pathway name	Number of mapped Genes
	e) RvGV	
ko01120	Microbial metabolism in diverse environments	19
ko03013	RNA transport	17
ko05016	Huntington's disease	17
	f) RvOo	
ko01100	Metabolic pathways	132
ko01110	Biosynthesis of secondary metabolites	60
ko03010	Ribosome	47
ko01130	Biosynthesis of antibiotics	44
ko03040	Spliceosome	36
ko01120	Microbial metabolism in diverse environments	31
ko00230	Purine metabolism	31
ko01200	Carbon metabolism	28
ko03013	RNA transport	27
ko00240	Pyrimidine metabolism	24

The PF3D7_0324900 have highest SNPs and nsSNPs information, i.e., 2416 SNPs and 1606 nsSNPs from the database among all hub genes; whereas PF3D7_1439500, PF3D7_0508100, PF3D7_1306000, PF3D7_0705600, PF3D7_1207100, PF3D7_1126700, and PF3D7_1234300 have 171, 128, 101, 62, 38, 28, and 11 nsSNPs, respectively. The total number of nsSNPs and SNPs recognized for the hub genes are illustrated in Figure 3.

Analysis of nsSNPs for protein function

Deleterious coding nsSNPs were predicted using SIFT, PROVEAN, and PredictSNP tools. SIFT is a sequence homology-based tool used to classify substitutions for amino acids. The SIFT tool predicts whether substitution of an amino acid can affect protein function for a given FASTA protein sequence or not. The SIFT predicts substitutions with values <0.05 to be deleterious. The SIFT sequence tool predicted 60, 52, 38, 33, 28, 19, 7, and 2 positions to be affect protein function for PF3D7_1439500, PF3D7_1306000, PF3D7_0508100, PF3D7_0705600, PF3D7_0324900, PF3D7_1207100, PF3D7_1126700, and PF3D7_1234300, respectively. To predict the final PROVEAN score, a delta alignment score is calculated for every supporting sequence and mean value across the clusters. By default, a score of −2.5 or above of it is taken as deleterious, while anything short of this cutoff rating has been considered as neutral. In the hub genes, the PROVEAN protein tool predicted 40, 4, and 2 positions to be deleterious for PF3D7_0324900, PF3D7_1306000, PF3D7_1207100 and 1 position for PF3D7_0705600 and PF3D7_1234300, respectively. However, PROVEAN tool not found any deleterious mutation in PF3D7_1439500, PF3D7_0508100, and PF3D7_112670. PredictSNP unambiguously designed to combine outcomes of several methods, mostly to annotate disease-variant relationships. According to PredictSNP tool, there are 18, 16, 11, 9, 6, 4, 2, and

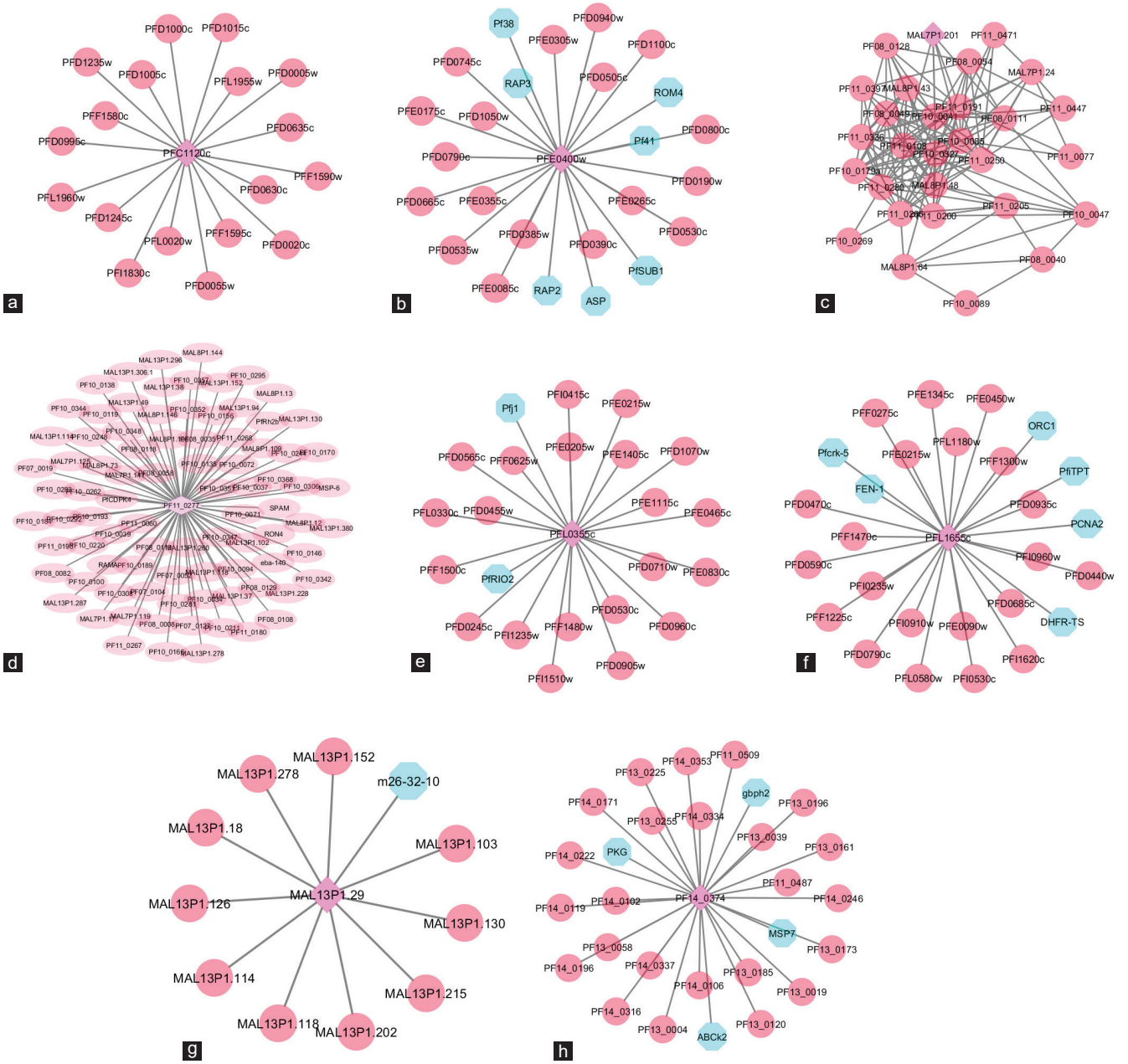


Figure 2: (a) PF3D7_0324900: Interaction between hub gene and their first neighbors. A red node indicates query genes and the genes predicted by Search Tool for the Retrieval of Interacting Genes/Proteins are shown in cyan. The hub is shown as purple diamond in each cluster. (b) PF3D7_0508100: Interaction between hub gene and their first neighbors. A red node indicates query genes and the genes predicted by Search Tool for the Retrieval of Interacting Genes/Proteins are shown in cyan. The hub is shown as purple diamond in each cluster. (c) PF3D7_0705600: Interaction between hub gene and their first neighbors. A red node indicates query genes and the genes predicted by Search Tool for the Retrieval of Interacting Genes/Proteins are shown in cyan. The hub is shown as purple diamond in each cluster. (d) PF3D7_1126700: Interaction between hub gene and their first neighbors. A red node indicates query genes and the genes predicted by Search Tool for the Retrieval of Interacting Genes/Proteins are shown in cyan. The hub is shown as purple diamond in each cluster. (e) PF3D7_1207100: Interaction between hub gene and their first neighbors. A red node indicates query genes and the genes predicted by Search Tool for the Retrieval of Interacting Genes/Proteins are shown in cyan. The hub is shown as purple diamond in each cluster. (f) PF3D7_1234300: Interaction between hub gene and their first neighbors. A red node indicates query genes and the genes predicted by Search Tool for the Retrieval of Interacting Genes/Proteins are shown in cyan. The hub is shown as purple diamond in each cluster. (g) PF3D7_1306000: Interaction between hub gene and their first neighbors. A red node indicates query genes and the genes predicted by Search Tool for the Retrieval of Interacting Genes/Proteins are shown in cyan. The hub is shown as purple diamond in each cluster. (h) PF3D7_1439500: Interaction between hub gene and their first neighbors. A red node indicates query genes and the genes predicted by Search Tool for the Retrieval of Interacting Genes/Proteins are shown in cyan. The hub is shown as purple diamond in each cluster

1 mutations to be deleterious for PF3D7_1306000, PF3D7_0324900, PF3D7_1439500, PF3D7_0705600, PF3D7_1207100, PF3D7_0508100, PF3D7_1126700, and PF3D7_1234300, respectively.

It was confirmed and verified at least by two tools used above in the study that 25 positions for PF3D7_0324900, 20 for PF3D7_1306000, 11 for PF3D7_1439500, 10 for PF3D7_0705600, 6 for PF3D7_1207100, 4 for

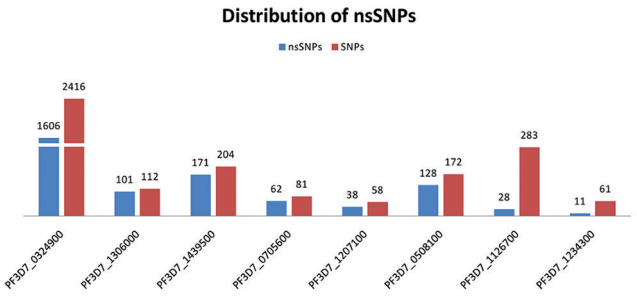


Figure 3: Distribution of total number of single nucleotide polymorphisms and nonsynonymous single nucleotide polymorphisms identified for *Plasmodium falciparum* hub genes

PF3D7_0508100, 2 for PF3D7_1126700, and 1 for PF3D7_1234300 were identified to influence the function of protein [Supplementary Table 2].

Solvent accessibility and secondary structure prediction

Further, the secondary structures and solvent accessibility of the hub genes was investigated using NetSurfP-1.1. The SNPs which were predicted by at least two tools to have a negative effect on the function of protein were used for the analysis of secondary structure and solvent accessibility. Moreover, the data were screened by selecting the residues, which showed change in ASA ≥10 Å^2 from buried state to exposed state and also exposed state to buried state and its secondary structure change. The information related to ASA and secondary structure are shown individually in Supplementary Table 3.

In PF3D7_0324900, N615K mutation showed an ASA change to exposed state from buried but T1745P, S1747R, S2024R, P2026S and E2065K mutations showed the opposite change to buried state from exposed state. There was a change in most of the conformations from Coil to Alpha-Helix, due to these mutations [Table 4]. Due to mutations Y779D and Y862N, an ASA change to exposed state from buried was shown in PF3D7_1306000 whereas D1113Y showed an inverse change. Conformations in Y779D and D1113Y were changed from Alpha-Helix to Coil, while in Y862N, Beta-strand was changed to Coil. Mutations in the residue position N186H, S312N, N334S, Y338N, S348C, Y403S, S445L, and R570G in PF3D7_1439500 show a change to exposed state from the buried state and mutations S313N, T411I, and N746 K show a change from buried state in the majority of the cases. In most cases, it shows changes in secondary structure from coil to helix and coil to beta-strand. In PF3D7_0705600, mutations show almost the same change to exposed state from buried state and vice versa. It shows changes in secondary structure from helix to coil and helix to beta-strand in most cases and also from coil to helix in some cases. Due to mutation D377H, an ASA change primarily from buried state to exposed state was shown in PF3D7_1207100 whereas N235K, R277W, P528S, and N661Y show an inverse change. In R277W mutation, both F274 and E275 show change in C to H conformation. D377H shows change in conformation mainly from H to C. In P528S mutation, both L337 and Y340 show change from C to H conformation. In N661Y mutation, L337 show change from C to H and I410

Table 4: NetSurfP results showing change in accessible surface area from buried to exposed state and vice versa and also show change in their secondary structure

Mutation	Class change		Conformation change					
	B-E	E-B	C-E	C-H	E-C	E-H	H-C	H-E
PF3D7_0324900								
N615K	5	3	1	-	-	-	-	-
T1745P	4	4	-	1	-	-	-	-
S1747R	1	2	-	1	-	-	-	-
S2024R	1	2	-	1	-	-	-	-
P2026S	-	3	-	1	-	-	-	-
E2065K	1	2	1	-	1	-	-	-
PF3D7_1306000								
Y779D	1	-	-	-	-	-	1	-
Y862N	1	-	-	-	1	-	-	-
D1113Y	2	1	-	-	-	-	1	-
PF3D7_1439500								
N186H	83	25	5	8	2	-	2	-
S312N	61	54	4	4	3	-	2	-
S313N	64	69	7	6	2	-	1	1
N334S	85	25	2	6	2	-	3	-
Y338N	69	54	7	1	2	-	2	-
S348C	78	33	4	7	1	-	2	-
Y403S	70	68	2	12	2	-	2	-
T411I	59	66	7	7	3	-	2	-
S445L	69	59	4	9	5	-	2	-
R570G	67	58	5	8	5	-	2	-
N746K	69	73	9	8	2	-	3	-
PF3D7_0705600								
N124Y	3	2	1	1	-	-	-	-
E623V	24	24	2	2	1	-	8	4
A626D	20	26	2	2	1	-	7	4
D645Y	22	27	2	5	1	-	6	5
T688R	20	22	2	6	1	-	8	5
T852S	14	13	1	1	2	-	1	1
PF3D7_1207100								
N235K	6	15	1	3	-	-	1	-
R277W	3	9	-	2	-	-	-	-
D377H	16	8	-	-	-	1	2	-
P528S	4	3	-	2	-	-	-	-
N661Y	5	8	-	1	-	-	1	-
PF3D7_0508100								
I800T	30	45	-	4	-	-	7	2
I1103K	43	51	3	5	2	1	10	2
S1411P	47	58	2	8	-	2	15	3
Y1474H	55	64	2	10	-	-	13	3
PF3D7_1234300								
S434Y	3	4	-	1	-	-	-	-

B=Buried, E=Exposed, H=Alpha-Helix, E=Beta-strand, C=Coil

show an opposite change in conformation. Mutations in the residues I800T, I1103K, S1411P and Y1474H, in PF3D7_0508100 show mostly from exposed to buried state. In most of the cases, it shows changes in secondary structure from helix to coil and vice versa. Some also change from helix to beta-strand conformations. In PF3D7_1234300, S434Y mutation, I137 show change in conformation from coil to helix. In PF3D7_1126700, no significant change was detected [Table 4].

Further, a database has been developed to show the analysis done by PROVEAN, SIFT, PredictSNP, and NetSurfP software, which is available online at URL www.cdkd.org/pfsnp/.

Database was developed using PHP and JavaScript with user-friendly search environment using a range of options, such as simple searches and advanced searches. Users can search easily with any of the three fields such as gene name, UniProt ID, or SNP ID. Gene name with corresponding residues is available in advance search. Users can search for particular residues in any gene or genes available in this database (currently the hub genes). For displaying the SNP data, two-step approach is exploited by these two search methods. The first step is to display Gene name, UniProt ID, SNP ID, AA change, PredictSNP cuttoff, Confidence, PROVEAN Score, SIFT Score, and cutoff. The second step will reveal the details of the secondary structure, SNP ID, Residue N, Location, ASA N, Class N, SS N, Residue SNP, SNP, ASA SNP, Class SNP, and SS SNP. Users can also customize their search results by choosing any field like Class change, SS change, ASA change or Residue N or any combination for a particular gene. The relevant information is displayed dynamically. In the help section, all headers are defined and hyperlinked to this web portal from the search pages. In addition, every SNP ID is directly linked to its relevant entries in the PlasmoDB database.

Discussion

The most virulent pathogen of malaria and malaria mortality worldwide is *P. falciparum*. To control the disease in parasites of malaria, the study of genetic variation is of practical importance. A nonredundant list of 4196 genes was used for the functional annotation cluster analysis by the DAVID tool. The most important enriched GO terms identified in these genes by DAVID functional cluster analysis consist of functions required for the host cell plasma membrane, antigenic variation, and pathogenesis [Table 2]. The node degree was calculated for each gene in the network to explore the functional roles of genes involved in different processes and interaction networks were created using Network Analyzer and MCODE plugins of Cytoscape. PF3D7_0324900, PF3D7_1306000, PF3D7_1439500, PF3D7_0705600, PF3D7_1207100, PF3D7_0508100, PF3D7_1126700, and PF3D7_1234300 genes were considered as hubs genes in the network constructed from 4196 *P. falciparum* genes by Network Analyzer and MCODE plugins of Cytoscape. Hub genes are considered functionally important because these are highly interconnected with nodes in a system. Therefore, these can act as putative targets for drug designing.

In this study, a number of nsSNPs have been identified for nearly all hubs, but few have had an impact on protein functions. PF3D7_0324900, PF3D7_1306000, PF3D7_1439500, PF3D7_0705600, PF3D7_1207100, PF3D7_0508100, PF3D7_1126700, and PF3D7_1234300 genes were annotated by at least two tools to affect protein function. The NetSurfP web server has been used to predict the protein secondary structure and surface accessibility in normal for each gene and its predicted SNP substitutions. The erythrocyte membrane protein 1 of *P. falciparum*, PfEMP1 (PF3D7_0324900) is considered as potential hub in this study, which mediates the attachment of infected erythrocytes to a range of host cells in the vascular lining during the blood stage of the infection with malaria. PfEMP1 is inserted into the RBC membrane and then laterally transferred to the preformed knob structures in the central region. Knob assembly may result in new ways of inhibiting PfEMP1 presentation on infected RBCs.[22] Strategies for overcoming PfEMP1 antigenic diversity would provide an exciting new opportunity for the development of malaria vaccines.[23] PF3D7_1306000 is a conserved protein of *Plasmodium* with unknown function and considered as hub gene. This gene was found as essential along with druggability index 0.5 in Tropical Disease Research (TDR) Targets database (https://tdrtargets.org/). Another hub CCAAT-binding transcription factor or oocyst rupture protein 2 (ORP2) (PF3D7_1439500) was involved in sporozoite egress. Sporozoite egress of the oocyst can be blocked by removing the N-terminal histone fold domain of ORP2.[24]

Hub gene RNA helicase, (PF3D7_0705600) play various roles, including the cell growth and development. Helicases are significant unwinding enzymes that are needed in the malaria parasite for nearly all the nucleic acid metabolism. RNA helicases could be used as reasonable targets to develop new antiparasite therapies and solve the problem of drug resistance.[25] Small subunit rRNA processing factor, (PF3D7_1207100) involved in the maturation of SSU-rRNA from tricistronic rRNA transcript was also found as a hub gene. It is a protease group of enzymes, which play key roles in the development and invasion of parasites. The ability to design particular protease inhibitors makes them promising objectives for drugs.[26] SET domain protein, (PF3D7_0508100), which enables histone-lysine

N-methyltransferase activity was also observed as hub gene. Two types of protein methyltransferase enzymes (PMTs) are present in eukaryotic cells: lysine specific and arginine specific. They were both linked with a variety of diseases including neurodegenerative and inflammatory diseases, and cancer. PMT enzymes have emerged as a target class against human disease for drug discovery.[27]

The gene PF3D7_1126700 that was also observed as a hub, codes for autophagy-related protein 23. The ATG18 autophagy-related protein controls the biogenesis of apicoplast in apicomplex parasites and decline of ATG18 in *P. falciparum* showed in delayed death.[28] DNA polymerase epsilon subunit B (PF3D7_1234300) is involved in the DNA-dependent DNA replication and enables DNA-directed DNA polymerase activity. The main function of Pol epsilon is to extend the leading strand synthesis during replication.[29]

A total of 6 mutations (N615K, T1745P, S1747R, S2024R, P2026S, and E2065K) were identified for PF3D7_0324900, 3 mutations (Y779D, Y862N and D1113Y) for PF3D7_1306000, 11 mutations (N186H, S312N, S313N, N334S, Y338N, S348C, Y403S, T411I, S445 L, R570G and N746K) for PF3D7_1439500, 6 mutations (N124Y, E623V, A626D, D645Y, T688R and T852S) for PF3D7_0705600, 5 mutations (N235K, R277W, D377H, P528S and N661Y) for PF3D7_1207100, 4 mutations (I800T, I1103K, S1411P and Y1474H) for PF3D7_0508100, and only one mutation for PF3D7_1234300. These mutations brought a change in accessible surface area from buried to exposed state and vice versa and also change in their secondary structure as observed through NetSurfP tool.

A computational approach to systematically analyze nsSNPs was undertaken in this study to predict deleterious mutations. SNP within the proteins significantly affects stability of structure of a protein and its function. The role of nsSNPs in understanding the functional effects of mutations that may cause changes in amino acids in hub proteins was also examined. Further, determination of nsSNPs that interfere with the function of protein and cause a disease needs to be determined.[30]

Conclusion

In this study, a comprehensive approach has been used to derive potential therapeutic targets for *P. falciparum* using RNA-seq dataset. The differential expression of genes, functional, and pathway enrichment analysis of *P. falciparum* was apprized in detail. The present study results suggested that PF3D7_0705600, PF3D7_1207100, PF3D7_0508100, PF3D7_1126700, and PF3D7_1234300 hub genes serves as putative targets for drug designing. These hub genes are showing less mutation and no similarity with human proteins. These genes act as lysine methyltransferases, transcription, translation, RNA splicing, and other important cellular pathways in *P. falciparum*. Moreover, the study also provides clusters of hub genes and their network pathways analysis which can be used further studies to devise a therapeutic target that stabilizes their gene expression. In addition, nsSNPs and their functional impact of these hub genes were also calculated.

Financial support and sponsorship
Nil.

Conflicts of interest
There are no conflicts of interest.

References

1. Le Roch KG, Chung DW, Ponts N. Genomics and integrated systems biology in *Plasmodium falciparum*: A path to malaria control and eradication. Parasite Immunol 2012;34:50-60.
2. Sharma VP. Continuing challenge of malaria in India. Curr Sci 2012;102:678-82.
3. Goswami D, Baruah I, Dhiman S, Rabha B, Veer V, Singh L, *et al*. Chemotherapy and drug resistance status of malaria parasite in Northeast India. Asian Pac J Trop Med 2013;6:583-8.
4. Hay SI, Okiro EA, Gething PW, Patil AP, Tatem AJ, Guerra CA, *et al*. Estimating the global clinical burden of *Plasmodium falciparum* malaria in 2007. PLoS Med 2010;7:e1000290.
5. Imwong M, Dondorp AM, Nosten F, Yi P, Mungthin M, Hanchana S, *et al*. Exploring the contribution of candidate genes to artemisinin resistance in *Plasmodium falciparum*. Antimicrob Agents Chemother 2010;54:2886-92.
6. Breglio KF, Amato R, Eastman R, Lim P, Sa JM, Guha R, *et al*. A single nucleotide polymorphism in the *Plasmodium falciparum* atg18 gene associates with artemisinin resistance and confers enhanced parasite survival under nutrient deprivation. Malar J 2018;17:391.
7. Subudhi AK, Boopathi PA, Pandey I, Kaur R, Middha S, Acharya J, *et al*. Disease specific modules and hub genes for intervention strategies: A co-expression network based approach for *Plasmodium falciparum* clinical isolates. Infect Genet Evol 2015;35:96-108.
8. López-Barragán MJ, Lemieux J, Quiñones M, Williamson KC, Molina-Cruz A, Cui K, *et al*. Directional gene expression and antisense transcripts in sexual and asexual stages of *Plasmodium falciparum*. BMC Genomics 2011;12:587.
9. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, *et al*. Differential gene and transcript expression analysis of RNA-seq experiments with tophat and cufflinks. Nat Protoc 2012;7:562-78.
10. Aurrecoechea C, Brestelli J, Brunk BP, Dommer J, Fischer S, Gajria B, *et al*. PlasmoDB: A functional genomic database for malaria parasites. Nucleic Acids Res 2009;37:D539-43.
11. Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, *et al*. UniProt: The universal protein knowledgebase. Nucleic Acids Res 2004;32:D115-9.
12. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 2009;37:1-3.
13. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: An automatic genome annotation and pathway reconstruction server. Nucleic Acids Res 2007;35:W182-5.
14. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S,

Simonovic M, *et al*. The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res 2017;45:D362-8.
15. Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T. Cytoscape 2.8: New features for data integration and network visualization. Bioinformatics 2011;27:431-2.
16. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics 2003;4:2.
17. Choi Y, Chan AP. PROVEAN web server: A tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics 2015;31:2745-7.
18. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 2009;4:1073-81.
19. Bendl J, Stourac J, Salanda O, Pavelka A, Wieben ED, Zendulka J, *et al*. PredictSNP: Robust and accurate consensus classifier for prediction of disease-related mutations. PLoS Comput Biol 2014;10:e1003440.
20. Petersen B, Petersen TN, Andersen P, Nielsen M, Lundegaard C. A generic method for assignment of reliability scores applied to solvent accessibility predictions. BMC Struct Biol 2009;9:51.
21. Kucukkal TG, Petukh M, Li L, Alexov E. Structural and physico-chemical effects of disease and non-disease nsSNPs on proteins. Curr Opin Struct Biol 2015;32:18-24.
22. Looker O, Blanch AJ, Liu B, Nunez-Iglesias J, McMillan PJ, Tilley L, *et al*. The knob protein KAHRP assembles into a ring-shaped structure that underpins virulence complex assembly. PLoS Pathog 2019;15:e1007761.
23. Chan JA, Fowkes FJ, Beeson JG. Surface antigens of *Plasmodium falciparum*-infected erythrocytes as immune targets and malaria vaccine candidates. Cell Mol Life Sci 2014;71:3633-57.
24. Siden-Kiamos I, Pace T, Klonizakis A, Nardini M, Garcia CR, Currà C. Identification of *Plasmodium* berghei Oocyst Rupture Protein 2 (ORP2) domains involved in sporozoite egress from the oocyst. Int J Parasitol 2018;48:1127-36.
25. Marchat LA, Arzola-Rodríguez SI, Hernandez-de la Cruz O, Lopez-Rosas I, Lopez-Camarillo C. DEAD/DExH-Box RNA helicases in selected human parasites. Korean J Parasitol 2015;53:583-95.
26. Lilburn TG, Cai H, Zhou Z, Wang Y. Protease-associated cellular networks in malaria parasite *Plasmodium falciparum*. BMC Genomics 2011;12 Suppl 5:S9.
27. Copeland RA, Solomon ME, Richon VM. Protein methyltransferases as a target class for drug discovery. Nat Rev Drug Discov 2009;8:724-32.
28. Bansal P, Tripathi A, Thakur V, Mohmmed A, Sharma P. Autophagy-related protein ATG18 regulates apicoplast biogenesis in *Apicomplexan* Parasites. MBio 2017;8. pii: E01468-17.
29. Pursell ZF, Isoz I, Lundström EB, Johansson E, Kunkel TA. Yeast DNA polymerase epsilon participates in leading-strand DNA replication. Science 2007;317:127-30.
30. Karchin R, Diekhans M, Kelly L, Thomas DJ, Pieper U, Eswar N, *et al*. LS-SNP: Large-scale annotation of coding non-synonymous SNPs based on multiple information sources. Bioinformatics 2005;21:2814-20.

Research Paper

Homology Modeling, Docking, Absorption, Distribution, Metabolism, Excretion and Toxicity Studies and Prediction of Deleterious Non-Synonymous Single Nucleotide Polymorphisms (nsSNPs) of Thiamin Phosphate Synthase: A Potential Drug Target in *Plasmodium Falciparum*

S. K. SINGH* AND M.S. REDDY*

Department of Biotechnology, Thapar Institute of Engineering and Technology, Patiala 147004, Punjab, India

Singh *et al.*: Thiamin phosphate synthase a potential drug target

The drug resistance in malarial parasites is increasingly emerging, hence it is essential to discover and develop alternative anti-malarial agents against both new and established drug targets. One of such possible drug targets is thiamine phosphate synthase because of its role and essentialness in the thiamine biosynthesis pathway. The present study aims to model the three-dimensional (3D) structure of thiamine phosphate synthase and to predict the potential inhibitors to derive therapeutic objectives for *Plasmodium falciparum*. The 3D structure was constructed using SWISS-MODEL and several computer-aided approaches were used for screening of drug-like compounds. In PyRx 0.8, molecular docking was conducted using AutoDock Vina. The absorption, distribution, metabolism and excretion properties were predicted using admetSAR. Post-docking results were analyzed using LigPlot+ program. The 3D model of thiamine phosphate synthase was generated using thiamine phosphate pyrophosphorylase from *Pyrococcus furiosus* as a template. Out of 156 compounds screened, only those 98 compounds which followed the Lipinski's rule of five were used for molecular docking. The best 25 docked ligands were further subjected to admetSAR for evaluation of absorption, distribution, metabolism, excretion and toxicity properties. Among these, 3 compounds, 5b (ZINC000003953801), 5m (ZINC000001686969), and 5u (ZINC000002036738) showed good absorption, distribution, metabolism, excretion and toxicity properties. Impact of 14 nsSNPs on the PfThiE protein structure or function was also investigated. The predicted inhibitors in this study may be further oriented to the development of treatment through experimental therapeutic methods to suppress pathogenic action of *P. falciparum*.

Key words: Thiamine phosphate synthase, PfThiE protein; malaria, *plasmodium falciparum*, nsSNP docking, parasitic diseases, protozoan infections, vector borne diseases, drug targets

Malaria is caused by the genus *Plasmodium* parasite and is transmitted by infected female Anopheles mosquitoes through the bites. An estimated 228 million malaria cases and 4,05,000 deaths were reported in 87 nations in 2018 as stated by the World Health Organization in World Malaria Report of 2019[1]. Malaria remains a major public health issue in the world. Drug resistance is increasingly emerging in malaria parasites, so it is important to identify and develop alternative anti-malarial agents against both new and existing drug targets. In apicomplexan parasites, thiamine biosynthesis offers a potential and exciting chance to achieve such goals, as the pathway is found only in prokaryotes, fungi, and plants, but is not present in mammals[2,3]. Thiamine pyrophosphate (Thi-PP) is the active form of vitamin B1, which is a co-factor for various enzymes

***Address for correspondence**
E-mail: sanjaybiosoft@gmail.com

Accepted 05 July 2020
Revised 20 June 2020
Received 11 January 2020
Indian J Pharm Sci 2020;82(4):665-676

primarily involved in the metabolism of carbohydrates such as 2-oxoglutarate dehydrogenase, transketolase or pyruvate dehydrogenase. For a few days, the culturing of *P. falciparum* in a thiamine deficient medium showed no adverse effect but a substantial need for 4-amino-5-hydroxymethyl-2-methylpyrimidine (HMP) or thiamine itself for parasite growth was reported after ten days[1]. In thiamine biosynthesis pathway (fig. 1), thiazole THZ-P (5-methyl-4-(beta-hydroxyethyl)thiazole phosphate) and pyrimidine HMP-PP (2-methyl-4-amino-5-hydroxymethylpyrimidine pyrophosphate) moieties are combined to yield thiamine phosphate by PfThiE[3,4]. For several enzymes, thiamine is metabolized as an essential cofactor[5]. So, a novel drug target thiamine phosphate synthase (PfThiE) of *P. falciparum* which is essential enzyme in thiamine biosynthesis was chosen to screen potent anti-malarial drugs. The human host's lack of vitamin biosynthesis signifies that inhibition of the parasite pathways can be a way to particularly interfere with the development of parasites[6].

The present study aimed to investigate the possible effects of nsSNPs in PfThiE and their effects on its structure and function, 3D structure formation and prediction of inhibitors for the modelled structure. Till date no reports are available on the effect of deleterious SNPs and docking studies experimentally or computationally on PfThiE of *P. falciparum*. Non-synonymous single nucleotide polymorphisms (nsSNPs) lead to variations in the amino acid sequence, as they influence the primary polypeptide directly. These changes are not only associated with their primary sequence modification but can also alter or impair the structure and function of protein in the amino acid sequence. Numerous studies have predicted most deleterious nsSNPs among recorded polymorphisms and understood their effect on protein function, structure, and stability[7,8]. Many researchers have been studied SNPs of *P. falciparum*[9,10]. Functional analysis, stability analysis and conservation analysis were performed for PfThiE protein. The 3D structure of PfThiE was developed and validated. Several potent ligand molecules were identified by virtual screening method and evaluated through ADMET (Absorption, Distribution, Metabolism, Excretion and Toxicity) properties. The interactions between proteins and ligands were studied in this study using molecular docking. In the lack of the molecular structure, the proposed 3D model will be useful in providing a novel target against malaria for structure-based drug design.

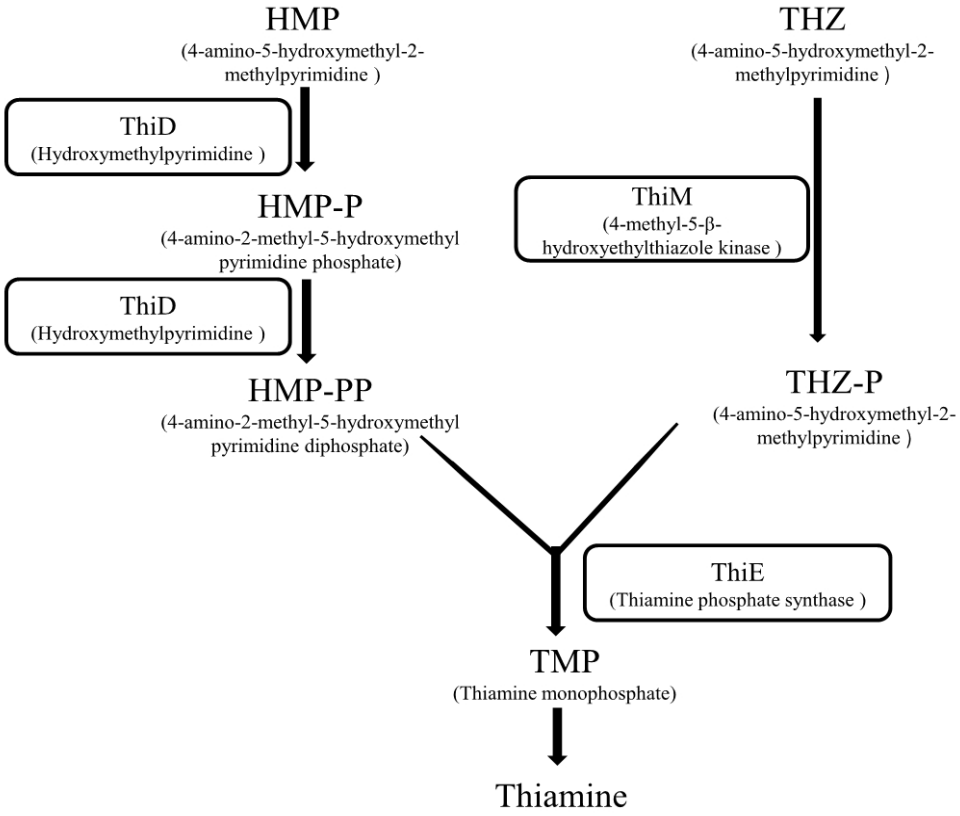


Fig. 1: The flow diagram of Thiamine biosynthesis pathway

MATERIALS AND METHODS

Identification of nonsynonymous SNPs:

The protein sequence of PfThiE protein was retrieved from the NCBI Gene Bank (XP_966127.2) and UniProtKB (C6KSY2) databases. The Polymorphism information for the PfThiE (PF3D7_0614000) was retrieved from PlasmoDB database (https://plasmodb.org/plasmo/). In this study, only nsSNPs were used for analysis because non-synonymous mutations can change the protein sequences, which ultimately change structure and function of protein.

Functional analysis of nonsynonymous SNPs:

SIFT - Sorting Intolerant from Tolerant[11] (http://sift.bii.a-star.edu.sg/www/SIFT_seq_submit2.html), PROVEAN - Protein Variation Effect Analyzer[12] (http://provean.jcvi.org/), PredictSNP - Consensus classifiers for prediction of disease-related mutations[13] (https://loschmidt.chemi.muni.cz/predictsnp/) and SNAP2 - Screening for Non-acceptable Polymorphisms[14] (https://rostlab.org/services/snap2web/) web-based tools were used to predict whether a substitution of amino acid affects the biological function of a protein. SIFT is a multi-step algorithm that uses homology sequences to distinguish amino acid substitutions. SIFT expects deleterious substitutions with values < 0.05. PROVEAN is a sequence-based method that utilizes clustering of BLAST hits. For each supporting sequence, a delta alignment score is calculated, and then combined in and through clusters for PROVEAN score calculation. PROVEAN scores below -2.5 are known to be having a deleterious effect for individual SNPs. PredictSNP was specifically designed to combine the predicted outcomes of a number of tools to create a consensus forecast. By using various biophysical features, evolutionary knowledge and several features of structure, SNAP2 determines whether or not an SNP is likely to alter the function of proteins. It provides prediction results in the form of effect or neutral and a score ranging from -100 to 100.

Protein stability analysis:

I-Mutant 2.0[15] (http://folding.biofold.org/i-mutant/i-mutant2.0.html) has been used for the analysis of protein stability and alterations by taking into account the SNPs. Protein sequence, temperature (25°C), pH (7.0), and detailed SNP data are the input parameters for this tool. It provides prediction in the form of either Reliability Index (RI) or Free Energy change value (DDG).

ConSurf server:

Using a Bayesian algorithm ConSurf[16] (consurf.tau.ac.il/) was used to evaluate the evolutionary stability of amino acid positions in the protein. Conserved regions were predicted using conservation scores and a colour scheme and further divided into different nine-degree scales. The score of conservation is 1 - 4 for variable, 5 - 6 for intermediate and 7 - 9 for conserved regions.

Prediction of 3D structure and validation of modelled protein:

The Protein Data Bank (https://www.rcsb.org/) lacks the 3D structure of thiamine phosphate synthase (PfThiE). Therefore, SWISS-MODEL (https://swissmodel.expasy.org/)[17] was used to build 3D structure of protein by submitting FASTA protein sequence. SWISS-MODEL is a fully automated server that uses the crystal structure of similar protein as a template to predict 3D protein structures. Depending on global model quality estimation (GMQE) and qualitative model energy analysis (QMEAN) values, the most reliable 3D structure has been selected. Further, Verify3D, ERRAT and PROCHECK tools available on SAVES v5.0 (https://servicesn.mbi.ucla.edu/SAVES/) were used to validate the predicted 3D model of protein.

Alignment of model and the template structure:

The Dali[18] (http://ekhidna2.biocenter.helsinki.fi/dali/) web server was used for comparing 3D structure of proteins. For structure comparisons, this offers four choices, PDB search, PDB25, Pairwise and All against all. Pairwise structure comparison was used to compare template structure and modelled structure. It also provides secondary amino acid structure of a protein by means of DSSP.

Screening of compounds:

Several thiamine phosphate synthase (EC 2.5.1.3) inhibitors and their analogs were taken from BRENDA (BRaunschweig ENzyme DAtabase), PubChem (https://pubchem.ncbi.nlm.nih.gov/), ZINC[19] (https://zinc.docking.org/) and DrugBank (https://www.drugbank.ca/) compound databases and also 3D structure of protein modelled by SWISS-MODEL was uploaded to MTiOpenScreen[20] (http://bioserv.rpbs.univ-paris-diderot.fr/services/MTiOpenScreen/) for

screening of drug-like compounds. MTiOpenScreen conducts automatic virtual ligand screening, based on AutoDock Vina docking. This enables a curated library of small compounds to be screened in order to identify compounds that are likely to bind to a given protein receptor. This comprises five in-house prepared libraries, containing drug-like molecules. There are many compound library filters available for screening customization.

Molecular Docking:

AutoDock Vina 1.1.2[21] in PyRx 0.8[22] was used to do molecular docking. ZINC database was used for retrieval of compounds in Structure Data File (SDF) format. Open Babel (http://openbabel.org) tool was used to convert various file formats. PyRx was initially used to minimize compounds energy and convert all molecules to AutoDock Ligand (PDBQT) format. The compounds without any predefined binding sites were docked against the entire surface of protein. The outcomes of docking results were reported in the form of binding energy. LigPlot+[23] program was used for the analysis of post-docking results. Using PyMOL, the docked complexes which showing lowest binding affinity values were further analyzed hydrogen and hydrophobic bond interaction analysis.

Molecular features analyses:

The ADME and drug-likeness predictions of compounds were carried out using SwissADME[24] (http://www.swissadme.ch/). The SMILES of compounds have been used in SwissADME web tool as input. Further, ADMET and the pharmacokinetic properties were evaluated using admetSAR[25] (http://lmmd.ecust.edu.cn/admetsar2) web server to ensure the druggability potential of compounds.

RESULTS

There are two key parts of this study: sequence and structure. Sequence part used the protein sequence to identify SNP, functional analysis, stability analysis, conservation analysis, and 3D structure modelling. The *P. falciparum* thiamine phosphate synthase (PfThiE) consists of 1617 bp, with a total of 538 amino acids in its protein. The Thiamin phosphate pyrophosphorylase gene studied had a total of 14 nsSNPs in PlasmoDB database. For further study, all these non-synonymous coding SNPs have been chosen. Second part starts with the modelled 3D structure of protein. The 3D structure of protein was used for the screening of compounds, ADMET analysis and molecular docking studies. The schematic representation of the work flow for the analysis is depicted in fig. 2.

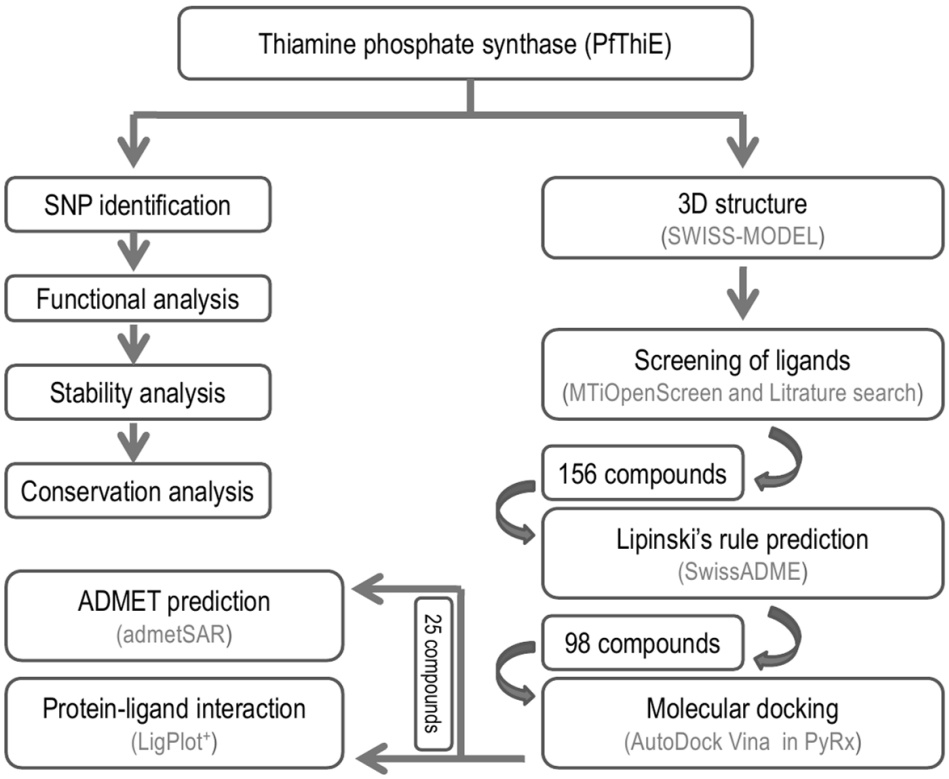


Fig. 2: The schematic representation of the work flow for the analysis

Functional analysis of nsSNPs:

To examine whether these SNPs have any impact on protein structure or function of PfThiE, we subjected all nsSNPs to five separate damaging prediction tools. The protein amino acid sequence in FASTA format and list of mutation positions and mutations were submitted to prediction tools for predicting harmful effects. The SIFT sequence tool predicted all of 14 variants that had an effect on protein function in PfThiE. Two nsSNPs (G165D and H190P) with a score of 0.01 and one nsSNP (I220V) with a score of 0.02 were deleterious, while the remaining 11 nsSNPs exhibited a deleterious score of 0.00. Two nsSNPs (A145V and G165D) were considered to be deleterious with the PROVEAN method, with a PROVEAN score below -2.5, and the remaining nsSNPs (12) were recognized as neutral. The PROVEAN method utilizes -2.5 as a cut-off value for all predictions. According to PredictSNP, in PF3D7_0614000, 5 nsSNPs (G165D, S330C, G401E, S427W and C456F) were predicted to be deleterious while 9 were found to be in neutral. The results from the SNAP2 server anticipated that two variants (G165D and S427W) would be effective, while the remaining 12 nsSNPs were intended to be neutral. Only one nsSNP (G165D) was estimated to affect protein function by all prediction tools (SIFT, PROVEAN, PredictSNP and SNAP2) (Table 1).

Analysis of mutation effects on the protein stability:

I-Mutant further evaluated all 14 SNPs for their effect on the stability of proteins. Reliability index (RI) was predicted for each mutation. Among the 14 SNPs proposed to predict stability, 13 predicted a decrease in the stability of the protein while one was found to increase the stability. With the exception of D494E, all nsSNPs decreased the stability of proteins with a range in the RI of 0 - 9 after mutation. Analysis of mutation effects on the protein stability of 14 nsSNPs is provided in Table 2. This result suggests that these mutations of PfThiE may directly or indirectly destabilize the amino acid interactions, leading to functional protein deviations.

Conservation analysis of deleterious nsSNPs:

To further explore the possible effects of deleterious nsSNPs, the evolutionary conservation of amino acid residues of PfThiE protein was calculated using ConSurf web server. The ConSurf tool predictions consist of a structural protein representation which includes the colorimetric conservation score (fig. 3). ConSurf predicted I433, A145 and C456 with conservative score 9, 8 and 7 respectively. Conservation score 6 was projected for G165, L310 and D311 while score 4 was for N239, S330 and D494. However, the remaining 5 amino acids (H190, I220, N355, G401 and S427) with conservative score 1 were predicted in variable region. Positions I433, A145 and C456 were expected in highly conserved regions so mutation in these amino acids suggests more possibility of altering the protein structure. The residues that are highly conserved are sometimes important for biological function. The ConSurf findings are presented in Table 3.

Protein 3D modeling and structural analysis:

Total three 3D structure of PfThiE were generated by SWISS-MODEL. The best model with GMQE scores

TABLE 1: THE NSSNPS THAT PREDICTED TO AFFECT PROTEIN FUNCTION BY SIFT, PROVEAN, PREDICTSNP AND SNAP2 TOOLS IN PFTHIE

Amino acid change	SIFT	PROVEAN	PredictSNP	SNAP2
A145V	Affect protein function	Deleterious	Neutral	Neutral
G165D	Affect protein function	Deleterious	Deleterious	Effect
H190P	Affect protein function	Neutral	Neutral	Neutral
I220V	Affect protein function	Neutral	Neutral	Neutral
N239H	Affect protein function	Neutral	Neutral	Neutral
L310V	Affect protein function	Neutral	Neutral	Neutral
D311E	Affect protein function	Neutral	Neutral	Neutral
S330C	Affect protein function	Neutral	Deleterious	Neutral
N355S	Affect protein function	Neutral	Neutral	Neutral
G401E	Affect protein function	Neutral	Deleterious	Neutral
S427W	Affect protein function	Neutral	Deleterious	Effect
I433L	Affect protein function	Neutral	Neutral	Neutral
C456F	Affect protein function	Neutral	Deleterious	Neutral
D494E	Affect protein function	Neutral	Neutral	Neutral

0.22 and QMEAN score −4.71 was generated using thiamine phosphate pyrophosphorylase from *P. furiosus* (pdb ID: 1xi3.1.A) as a template for this purpose. The structure for PfThiE was predicted as homo-dimer. Ramachandran plot of the 3D model showed 82.7% of its residues in the core while 13.3% in allowed, 2.0% in generously allowed and 2.0% in disallowed regions (fig. 4A). Overall ERRAT quality score of 80.8 suggested that the structure could be regarded as a good model (fig. 4B). Verify 3D result passed the model with 81.25% of the residues have averaged 3D-1D score > = 0.2 (fig. 4C).

Target and template alignment:

The Dali web server was used for alignment of target and template 3D structure of proteins. Alignment of template and modelled structure and sequence was performed in Dali (fig. 5A & fig 5B). Alignment score was predicted in the form of Z-score 32.9 and 32.4 for chain A and chain B, respectively. The identical amino acids are labelled with vertical bars. It also provides secondary structure of amino acid of protein by DSSP.

TABLE 2: I-MUTANT 2.0 OUTCOMES FOR 14 nsSNPs IN THE PROTEIN PfThiE

			I-Mutant 2.0	
Position	WT	NEW	Stability	RI
145	A	V	Decrease	1
165	G	D	Decrease	6
190	H	P	Decrease	5
220	I	V	Decrease	8
239	N	H	Decrease	9
310	L	V	Decrease	9
311	D	E	Decrease	2
330	S	C	Decrease	2
355	N	S	Decrease	7
401	G	E	Decrease	0
427	S	W	Decrease	0
433	I	L	Decrease	2
456	C	F	Decrease	5
494	D	E	Increase	4

TABLE 3: CONSERVATION PROFILE OF AMINO ACIDS IN PfThiE

Position	Amino Acid	Conservation Score	ConSurf Prediction
145	A145	8	Buried
165	G165	6	Exposed
190	H190	1	Exposed
220	I220	1	Exposed
239	N239	4	Exposed
310	L310	6	Buried
311	D311	6	Exposed
330	S330	4	Exposed
355	N355	1	Exposed
401	G401	1	Exposed
427	S427	1	Buried
433	I433	9	Highly conserved and buried (s)
456	C456	7	Buried
494	D494	4	Exposed

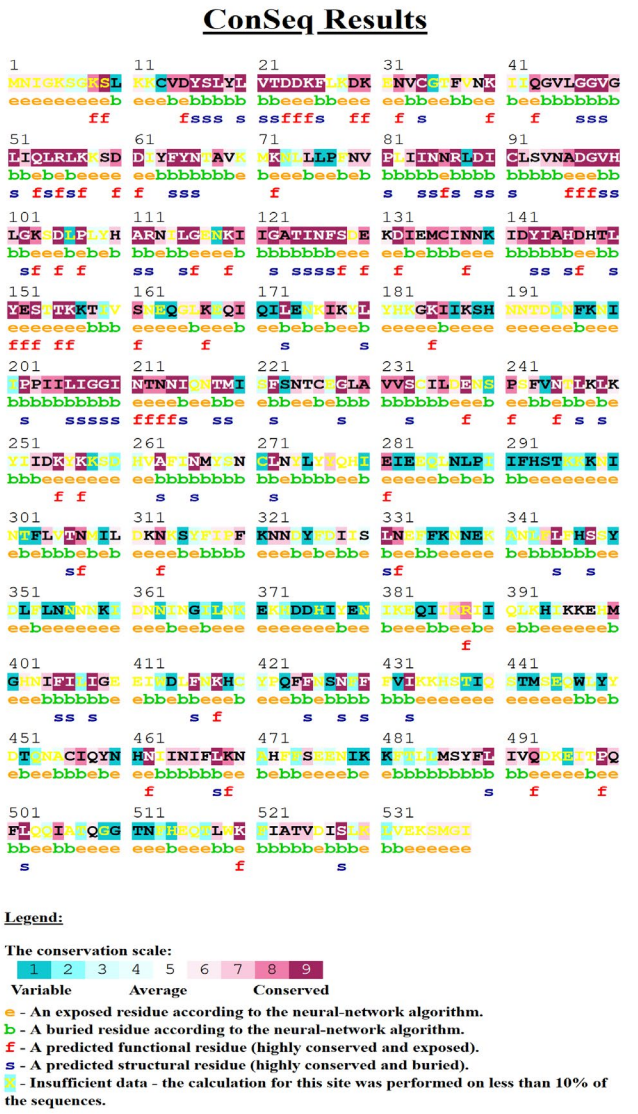
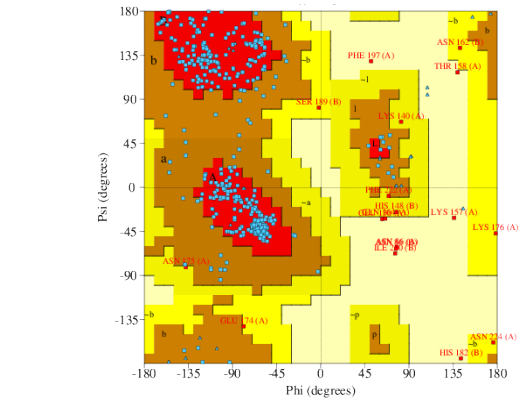


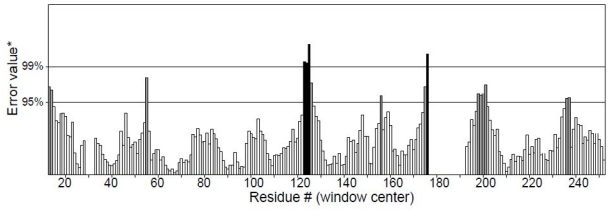
Fig. 3: Evolutionary stability of amino acid positions in PfThiE

Screening of compounds:

Several compound databases were searched for thiamine phosphate synthase inhibitors, and 12 from BRENDA and 6 from Drug Bank were taken. Also 39 top Inhibitors against *Mycobacterium tuberculosis* thiamin phosphate synthase from Khare et al.[24] study was retrieved from PubChem database. MTiOpenScreen screened top 100 drug-like compounds from 10,000 compound libraries that may be inhibitors for thiamine phosphate synthase. 156 compounds were screened from all searches. These 156 drug-like compounds were retrieved in SDF format.



(A)



(B)

(C)

Fig. 4: Structure validation of the PfThiE

Molecular properties analysis:

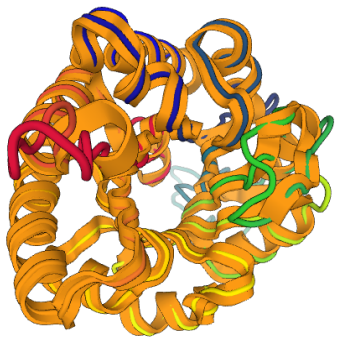
The ADME and drug-likeness predictions of 156 compounds were carried out using SwissADME. Out of 156 compounds, 98 follows the Lipinski's rule of five is provided as supplementary information (Table S1).

Molecular Docking and interaction analysis:

Molecular Docking was performed by using AutoDock Vina in PyRx 0.8 with 98 compounds which followed the Lipinski's rule. PyRx was initially used to minimize compounds energy and convert all molecules to AutoDock Ligand (PDBQT) format. The value of the grid box was set to center_x = 17.7657, center_y = 21.3541, center_z = 69.9140 while size_x = 65.1899, size_y = 58.2979, and size_z = 83.6737. The default exhaustiveness value was 8. All 98 compounds without any specified binding sites were docked against whole surface of the protein. Table 4 lists the outcomes of docking results that were shown as binding affinity less than -8.0 in at least one pose. The graphical representation of the 25 best screened compounds is

```
DSSP  LLLLllLLLEEEELLLEeehhhleeeELHHHHHHHHHHLLLLEEEELLLLLLHHHHHHHH
Query SLKKcvDYSLYLVTDDKflkdkenvcGTFVNKIIQGVLGGVGLIQLRLKKSDDIYFYNTA   60
ident |       ||  ||            |        ||   || | |       |
Sbjct NLRN--KLKLYVITDRR--------lKPEVESVREALEGGATAIQMRIKNAPTREMYEIG   50
DSSP  LHHH--HLLEEEELLL--------lLLHHHHHHHHHHLLLEEEELLLLLLHHHHHHHH

DSSP  HHHHHHHHHHLLLEEEELLHHHHHHLLLEEEELLLLLLHHHHHHHHlLLLEEEEEELl
Query VKMKNLLLPFNVPLIINNRLDICLSVNADGVHLGKSDLPLYHARNILqENKIIGATINFs  120
ident      |           | |  | | |||| ||  | |   |   |   | ||||
Sbjct KTLRQLTREYDALFFVDDRVDVALAVDADGVQLGPEDMPIEVAKEIA-PNLIIGASVYS-  108
DSSP  HHHHHHHHHLLLEEEEELLHHHHHHLLLEEEELLLLLLHHHHHHHL-LLLEEEEEELL-

DSSP  lLHHHHHHHHLLLLEEEELlLLLLLllllLLLLL-LHHHHHHHLLlllllllllllll
Query dEKDIEMCINNKIDYIAHDhTLYESttkkTIVSNE-QGLKEQIQILenkikylyhkgkii  179
ident            ||                 ||
Sbjct -LEEALEAEKKGADYLGAG-SVFPT----DARVIGLEGLRKIVESV--------------  148
DSSP  -HHHHHHHHHHLLLEEEEE-LLLLL----LLLLLHhHHHHHHHHL--------------

DSSP  lllllllllllllLLEEEELLLLLLLHHHHHLLlllLLLLEEELHHHHLLLLLLLlHH
Query kshnntddnfkniIPPIILIGGINTNNIQNTMISfsnTCEGLAVVSCILDENSPSFvNTL  239
ident      |  |||||  |                  | || |  |          |
Sbjct -------------KIPVVAIGGINKDNAREVLKT---GVDGIAVISAVMGAEDVRK-ATE  191
DSSP  -------------LLLEEEELLLLLLLHHHHHLL---LLLEEEELHHHHLLLLHHH-HHH

DSSP  HHHHHHHHL--
Query KLKYIIDKY--  248
ident |  | |
Sbjct ELRKIVEEVlg  202
DSSP  HHHHHHHHHhl
```

(A)



(B)

Fig. 5: Sequence and structure alignment of target and templet

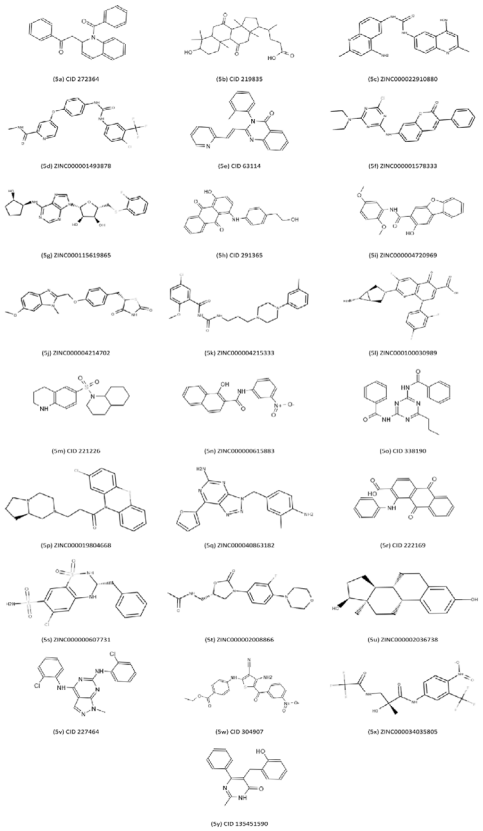


Fig. 6: Graphical representations of the best screened compounds

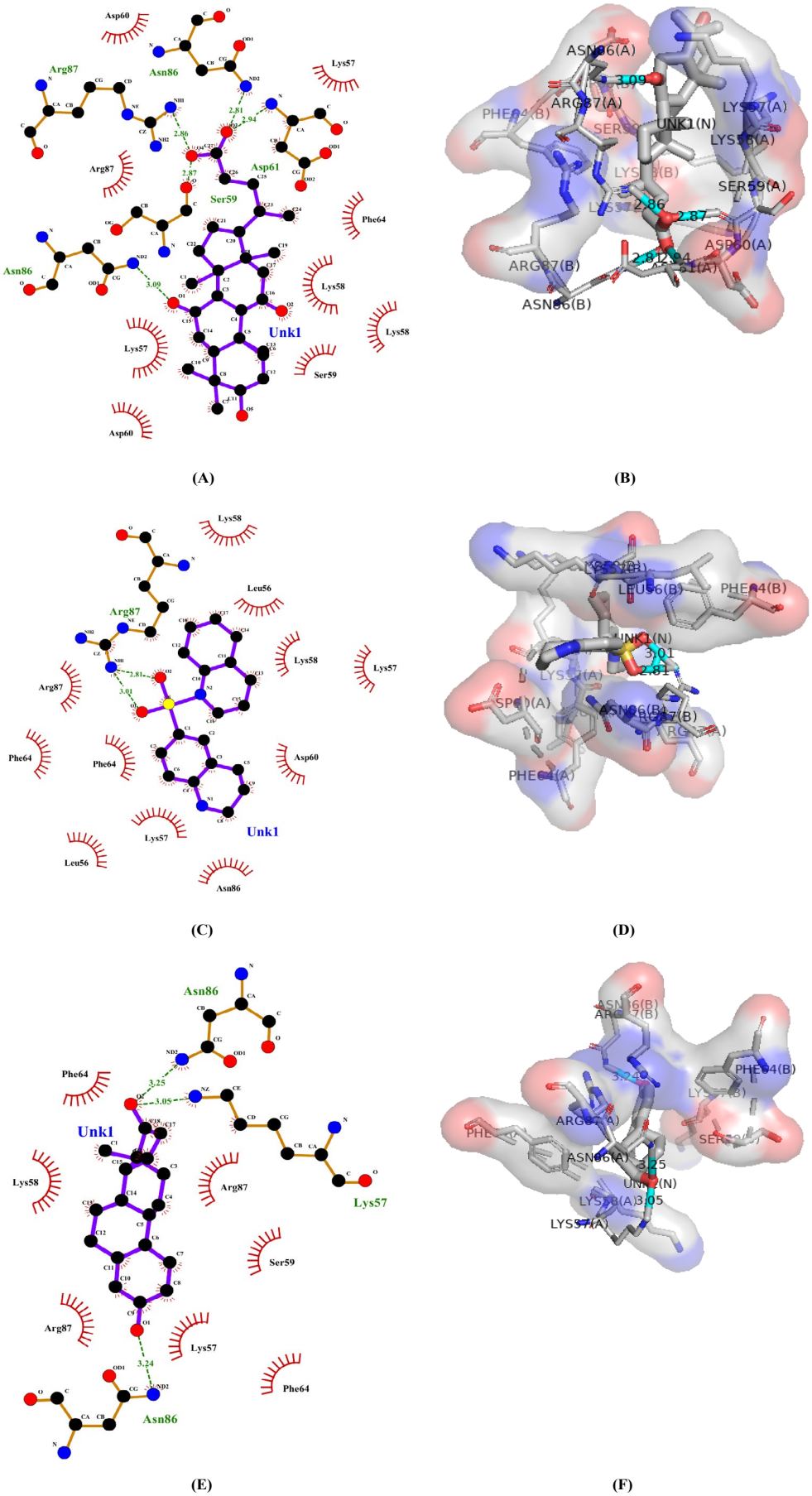


Fig. 7: Molecular docking interactions between compound and the binding sites of PfThiE

TABLE 4: DOCKING RESULTS OF DIFFERENT POSES SHOWING BINDING AFFINITY WITH LESS THAN -8.0 IN AT LEAST ONE POSE.

Compounds	Binding affinity (kcal/mol)								
	Pose 1	Pose 2	Pose 3	Pose 4	Pose 5	Pose 6	Pose 7	Pose 8	Pose 9
CID 272364	-11.2	-11.2	-11.1	-10.8	-10.7	-10.7	-10.7	-10.6	-10.4
CID 219835	-9.6	-8.1	-7.9	-7.7	-7.6	-7.6	-7.6	-7.5	-7.3
ZINC000022910880	-9.3	-9.1	-9	-8.9	-8.9	-8.8	-8.8	-8.8	-8.3
ZINC000001493878	-9	-8.7	-8.7	-8.5	-8.4	-8.4	-8.3	-8	-8
CID 63114	-9	-8.6	-8.4	-8.3	-8.3	-7.4	-7.3	-7.3	-7.3
ZINC000001578333	-8.9	-8.3	-8.3	-8.2	-8	-7.9	-7.8	-7.7	-7.7
ZINC000115619865	-8.8	-8.5	-8.3	-8.2	-8	-7.9	-7.8	-7.7	-7.7
CID 291365	-8.8	-8.4	-8.3	-8.2	-7.9	-7.8	-7.8	-7.5	-7.2
ZINC000004720969	-8.7	-8.5	-8.4	-8.1	-8.1	-7.8	-7.6	-7.4	-7.3
ZINC000004214702	-8.6	-8.5	-8.2	-8.2	-8.1	-8.1	-8	-7.8	-7.6
ZINC000004215333	-8.6	-8.5	-8.4	-8.3	-8.2	-7.6	-7.5	-7.4	-7.4
ZINC000100030989	-8.6	-8	-7.9	-7.6	-7.4	-7.3	-7.3	-7.3	-7.1
CID 221226	-8.6	-8.2	-8.1	-8	-7.6	-7.5	-7.4	-6.8	-6.8
ZINC000000615883	-8.6	-8.5	-8.2	-7.9	-7.7	-7.6	-7.6	-7.5	-7.4
CID 338190	-8.5	-7.8	-7.8	-7.8	-7.8	-7.7	-7.3	-7.3	-7.3
ZINC000019804668	-8.4	-7.4	-7.3	-7.3	-7.2	-7.2	-7	-7	-7
ZINC000040863182	-8.4	-8.2	-8.1	-7.8	-7.6	-7.6	-7.6	-7.6	-7.3
CID 222169	-8.4	-8.1	-7.8	-7.5	-7.4	-7.4	-7.2	-7.1	-7.1
ZINC000000607731	-8.3	-7.7	-7.6	-7.5	-7.3	-7	-6.9	-6.9	-6.9
ZINC000002008866	-8.2	-7.9	-7.8	-7.8	-7.6	-7.3	-7.3	-7.1	-7.1
ZINC000002036738	-8.2	-7.7	-7.7	-7.3	-7.2	-7.2	-7	-6.9	-6.7
CID 227464	-8.2	-8.2	-7.7	-7.7	-7.6	-7.5	-7.3	-7.2	-7.2
CID 304907	-8.2	-8.2	-8.1	-7.9	-7.7	-7.7	-7.6	-7.5	-7.5
ZINC000034035805	-8.1	-8	-7.9	-7.7	-7.7	-7.7	-7.6	-7.6	-7.6
CID 135451590	-8.1	-7.3	-7	-6.9	-6.7	-6.6	-6.5	-6.5	-6.5

TABLE 5: ADMET AND PHARMACOKINETIC PROPERTIES OF 3 BEST COMPOUNDS

Model Name	5b	5m	5u
Ames mutagenesis	-	-	-
Acute Oral Toxicity (c)	III	III	III
Androgen receptor binding	+	+	+
Aromatase binding	+	-	+
Avian toxicity	-	-	-
Blood Brain Barrier	-	+	-
BRCP inhibitior	-	-	+
Biodegradation	-	-	-
BSEP inhibitior	+	-	-
Caco-2	+	+	+
Carcinogenicity (binary)	-	-	-
Carcinogenicity (trinary)	Non-required	Non-required	Danger
crustacea aquatic toxicity	-	-	+
CYP1A2 inhibition	-	-	+
CYP2C19 inhibition	-	+	-
CYP2C9 inhibition	-	-	-
CYP2C9 substrate	-	-	+
CYP2D6 inhibition	-	-	-
CYP2D6 substrate	-	+	+
CYP3A4 inhibition	-	+	-
CYP3A4 substrate	+	+	+
CYP inhibitory promiscuity	-	+	-
Eye corrosion	-	-	-

Eye irritation	-	-	-
Estrogen receptor binding	+	-	+
Fish aquatic toxicity	+	+	+
Glucocorticoid receptor binding	+	-	+
Honey bee toxicity	+	-	+
Hepatotoxicity	-	-	-
Human either-a-go-go inhibition	+	-	+
Human Intestinal Absorption	+	+	+
Human oral bioavailability	-	+	-
MATE1 inhibitior	-	-	-
micronuclear	-	+	-
Acute Oral Toxicity	3.02	3.50	2.26
OATP1B1 inhibitior	+	+	+
OATP1B3 inhibitior	+	+	+
OATP2B1 inhibitior	-	-	-
OCT1 inhibitior	+	-	+
OCT2 inhibitior	-	-	-
P-glycoprotein inhibitior	-	-	-
P-glycoprotein substrate	-	-	-
PPAR gamma	+	-	-
Plasma protein binding	0.84	1.19	0.95
Subcellular localzation	Mitochondria	Lysosomes	Mitochondria
Tetrahymena pyriformis	0.65	2.17	0.73
Thyroid receptor binding	+	+	+
UGT catelyzed	-	-	+
Water solubility	-3.98	-3.70	-4.78

depicted in the fig 6. LigPlot+ software was used to predict all residues that interact with the cofactors of protein. Compound 5b had hydrogen bonds with amino acids SER59, ASP61, ASN86 AND ARG87, whereas hydrophobic interaction with LYS57, LYS58, SER59, ASP60, PHE64 and ARG87 residues (fig 7A and 7B). Compound 5m had only one hydrogen bond with amino acid ARG87, while LEU56, LYS57, LYS58, ASP60, PHE64, ARG86 and ARG87 residues had hydrophobic interaction (fig 7C and 7D). Amino acids LYS57 and ASN86 were connected to compound 5u with hydrogen bonds, while residues LYS58, SER59, PHE64 and ARG87 had hydrophobic interactions (fig 7E and 7F).

Evaluation of ADMET and pharmacokinetic properties:

Twenty-five best docked ligands were further subjected to admetSAR for evaluation of ADMET properties (Table 5). Out of 25 compounds, 14 compounds not exhibited toxicity to AMES. Blood Brain Barrier penetration was shown by all except 5b, 5s and 5u. There was no hepatotoxicity shown with only three compounds. The water solubility of all docked compounds is greater than -3.05. Compounds 5a, 5b, 5m, and 5v demonstrated Caco2 permeability. Intestinal absorption (human) was observed in all compounds. The acute oral toxicity of the maximum compounds was estimated as class III while 5d and 5s were indicated as II and IV, respectively. Eye corrosion and Eye irritation was not observed in any compounds.

DISCUSSION

The pathway of thiamine biosynthesis is present only in prokaryotes, fungi, and plants, but is not found in mammals[2]. THZ-P and HMP-PP moieties are combined in the thiamine biosynthesis pathway to produce thiamine phosphate by thiamine phosphate synthase (PfThiE)[3]. PfThiE is an important enzyme in the thiamine biosynthesis[5]. Thiamine phosphate synthase inhibitors have been studied in many species including *M. tuberculosis*[26], *Pyrobaculum calidifontis*[27], *Escherichia coli*[28] and *Zea mays*[29] but comprehensive research has not been done in *P. falciparum*. PfThiE is an essential enzyme in thiamine biosynthesis[30]. The thiamine biosynthesis pathway of the parasite has been proposed as a novel and indispensable antimalarial target[4]. So, current research attempted to explore the possible effects of nsSNPs in PfThiE and their effects on its structure and function, 3D structure formation and prediction of inhibitors for the modelled structure.

Rapid adaptation to changes in the environment due to high mutation rate in *P. falciparum* may result in

drug resistance to standard medicines[2]. Thus, new drug targets are needed to develop potential inhibitors against the disease. In the present study, the impact of nsSNPs of PfThiE was investigated using various bioinformatics tools. All of the 14 nsSNPs obtained were submitted to functional analysis. SIFT tool predicted that all 14 variants had an effect on protein function. All four tools predicted G165D to affect protein function. 5 nsSNPs (G165D, S330C, G401E, S427W and C456F) by PredictSNP, 2 nsSNPs (A145V and G165D) by PROVEAN and also 2 nsSNPs (G165D and S427W) by SNAP2 considered deleterious. Furthermore, after mutation as a consequence of these nsSNPs, I-Mutant 2.0 displayed a decrease in stability except for D494E, indicating to some degree that the protein is directly or indirectly destabilized. Phylogenetic analysis using ConSurf predicted that only two nsSNPs were found in highly conserved regions. This result shows that the majority of highly conserved residues are stable.

The pharmaceutical industry increasingly utilizes computational techniques to minimize time and financial costs in the drug discovery and development process. In this study, a computational approach was used to systematically evaluate the nsSNPs to predict deleterious mutations and after that 3D model of *P. falciparum* thiamine phosphate synthase was developed using the X-ray crystal structure of *P. furiosus* thiamine phosphate pyrophosphorylase as the crystal structure of PfThiE was not available. Various validation approaches found the overall structure to be an excellent model.

Among the 156 potential inhibitor compounds screened by different computational tools, 98 compounds followed Lipinski's rule of five and these were chosen for molecular docking studies. Top 25 compounds exhibited dock score values between -8.1 and -11.2 kcal/mol. 14 of the 25 compounds did not exhibit AMES toxicity. Only three compounds (5b, 5 m, and 5u) had all the good ADMET properties. Compounds (5a, 5b, 5e, 5h, 5i, 5 m, 5n, 5o, 5r, 5v, 5w, and 5y) investigated by Khare et al [24] as inhibitors against *M. tuberculosis* thiamine phosphate synthase had demonstrated strong binding affinity to PfThiE. All the three compounds (5b, 5 m, and 5u) had hydrogen bonds and hydrophobic interaction with amino acids. Further, these compounds have good ADMET properties.

P. falciparum thiamine phosphate synthase (PfThiE) is an essential enzyme in the thiamine biosynthesis which is not present in humans that can be considered as potential drug target to combat malaria menace. It is one of the few untouched targets for developing anti-malaria drugs. The PfThiE molecular model was developed in the present study by using crystal structure of *P. furiosus* thiamine phosphate pyrophosphorylase as a template. Potential ligands were tried to be identified through docking-based virtual screening with drug-likeness and ADMET analysis. In this analysis, various bioinformatics approaches were also used to examine the effect of non-synonymous SNPs of PfThiE. ConSurf prediction suggests that no mutation was present in the binding site of this protein. Among the compounds being screened, 5b (ZINC000003953801), 5m (ZINC000001686969), and 5u (ZINC000002036738) had the high binding affinity and good ADMET properties. Also, for further confirmation of the protein target and potential ligands, experimental characterization is also required.

Acknowledgments:

Authors thankful to Department of Biotechnology, Thapar Institute of Engineering & Technology, Patiala, India, for providing required facilities to carry out this research work.

REFERENCES

1. WHO (2020) World malaria report 2019. World Health Organization.
2. Wrenger C, Eschbach ML, Müller IB, Laun NP, Begley TP, Walter RD. Vitamin B1 de novo synthesis in the human malaria parasite *Plasmodium falciparum* depends on external provision of 4-amino-5-hydroxymethyl-2-methylpyrimidine. Biol Chem 2006;387:41-51.
3. Wrenger C, Knöckel J, Walter RD, Müller IB. Vitamin B1 and B6 in the malaria parasite: requisite or dispensable? Braz J Med Biol Res 2008;41:82-8.
4. Zhang Y, Taylor SV, Chiu HJ, Begley TP. Characterization of the Bacillus subtilis thiC operon involved in thiamine biosynthesis. J Bacteriol 1997;179:3030-5.
5. Chan XW, Wrenger C, Stahl K, Bergmann B, Winterberg M, Müller IB, *et al.* Chemical and genetic validation of thiamine utilization as an antimalarial drug target. Nat Commun 2013;4:2060.
6. Müller S, Kappes B. Vitamin and cofactor biosynthesis pathways in Plasmodium and other apicomplexan parasites. Trends Parasitol 2007;23(3):112-21.
7. Desai M, Chauhan JB (2017) Computational analysis for the determination of deleterious nsSNPs in human MTHFD1 gene. Comput Biol Chem 70:7-14.
8. Solayman M, Saleh MA, Paul S, Khalil MI, Gan SH. In silico analysis of nonsynonymous single nucleotide polymorphisms of the human adiponectin receptor 2 (ADIPOR2) gene. Comput Biol Chem 2017;68:175-85.
9. Subudhi AK, Boopathi PA, Pandey I, Kaur R, Middha S, Acharya J, *et al.* Disease specific modules and hub genes for intervention strategies: A co-expression network based approach for *Plasmodium falciparum* clinical isolates. Infect Genet Evol 2015;35:96-108.

10. Singh SK, Reddy SM. Investigation of hub genes and their nonsynonymous single nucleotide polymorphism analysis in *Plasmodium falciparum* for designing therapeutic methodologies using next−generation sequencing approach. Indian J Pharmacol 2019;51:389-99.
11. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 2009;4:1073-81.
12. Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics 2015;31:2745-7.
13. Bendl J, Stourac J, Salanda O, Pavelka A, Wieben ED, Zendulka J *et al.* PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. PLoS Comput Biol 2014;10: e1003440.
14. Hecht M, Bromberg Y, Rost B. Better prediction of functional effects for sequence variants. BMC Genomics 2015;16:S1.
15. Capriotti E, Fariselli P, Casadio R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res 2005;33:W306-10.
16. Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T *et al.* ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res 2016;44:W344-50.
17. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R *et al*. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res 2018;46:W296-W303.
18. Holm L, Rosenström P. Dali server: conservation mapping in 3D. Nucleic Acids Res 2010;38:W545-9.
19. Sterling T, Irwin JJ. ZINC 15--Ligand Discovery for Everyone. J Chem Inf Model 2015; 55:2324-37.
20. Labbé CM, Rey J, Lagorce D, Vavruša M, Becot J, Sperandio O *et al*. MTiOpenScreen: a web server for structure-based virtual screening. Nucleic Acids Res 2015;43:W448-54.
21. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 2010;31:455-61.
22. Dallakyan S, Olson AJ. Small-molecule library screening by docking with PyRx. Methods Mol Biol 2015;1263:243-50.
23. Laskowski RA, Swindells MB. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. J Chem Inf Model 2011;51:2778-86.
24. Daina A, Michielin O, Zoete V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Sci Rep 2017;7:42717.
25. Yang H, Lou C, Sun L, Li J, Cai Y, Wang Z *et al.* admetSAR 2.0: web-service for prediction and optimization of chemical ADMET properties. Bioinformatics 2019;35:1067-9.
26. Khare G, Kar R, Tyagi AK. Identification of inhibitors against *Mycobacterium tuberculosis* thiamin phosphate synthase, an important target for the development of anti-TB drugs. PLoS One 2011;6:e22441.
27. Hayashi M, Kobayashi K, Esaki H, Konno H, Akaji K, Tazuya K *et al.* Enzymatic and structural characterization of an archaeal thiamin phosphate synthase. Biochim Biophys Acta2014;1844:803-9.
28. Kawasaki T. Thiamine phosphate pyrophosphorylase. Methods Enzymol 1979;62:69-73.
29. Rapala-Kozik M, Olczak M, Ostrowska K, Starosta A, Kozik A. Molecular characterization of the thi3 gene involved in thiamine biosynthesis in *Zea mays*: cDNA sequence and enzymatic and structural properties of the recombinant bifunctional protein with 4-amino-5-hydroxymethyl-2-methylpyrimidine (phosphate) kinase and thiamine monophosphate synthase activities. Biochem J 2007;408:149-59.
30. Liu X, Wang Y, Liang J, Wang L, Qin N, Zhao Y, *et al.* In-depth comparative analysis of malaria parasite genomes reveals protein-coding genes linked to human disease in *Plasmodium falciparum* genome. BMC Genomics 2018;19:312.

Computational Biology and Chemistry 92 (2021) 107461

Contents lists available at ScienceDirect

Computational Biology and Chemistry

journal homepage: www.elsevier.com/locate/cbac

Computational prediction of the effects of non-synonymous single nucleotide polymorphisms on the GPI-anchor transamidase subunit GPI8p of *Plasmodium falciparum*

Sanjay Kumar Singh, M Sudhakara Reddy *

Department of Biotechnology, Thapar Institute of Engineering and Technology, Patiala, 147004, Punjab, India

ARTICLE INFO

Keywords:
Malaria
Plasmodium falciparum
GPI anchor transamidase
nsSNPs
Ligand binding sites

ABSTRACT

Drug resistance is increasingly evolving in malaria parasites; hence, it is important to discover and establish alternative drug targets. In this context, GPI-anchor transamidase (GPI-T) is a potential drug target primarily of its crucial role in the development and survival of the parasite in the GPI anchor biosynthesis pathway. The present investigation was undertaken to explore the plausible effects of nsSNP on the structure and functions of GPI-T subunit GPI8p of *Plasmodium falciparum*. The GPI8p (PF3D7_1128700) was analyzed using various sequence-based and structure-based computational tools such as SIFT, PROVEAN, PredictSNP, SNAP2, I-Mutant, MuPro, ConSurf, NetSurfP, MUSTER, COACH server and STRING server. Of the 34 nsSNPs submitted for functional analysis, 18 nsSNPs (R124 L, N143 K, Y145 F, V157I, T195S, K379E, I392 K, I437 T, Y438H, N439D, Y441H, N442D, N448D, N451D, D457A, D457Y, I458 L and N460 K) were predicted to have deleterious effects on the protein GPI8p. Additionally, I-Mutant 2.0 and MuPro both showed a decrease in stability after mutation as a result of these nsSNPs, suggesting the destabilization of protein. ConSurf findings suggest that most of the regions were highly conserved. In addition, COACH server was used to predict the ligand binding sites. It was found that no mutation was present at the predicted ligand binding site. The results of the STRING database showed that the protein GPI8p interacts with those proteins which either involve the biosynthetic process of attaching GPI anchor to protein or GPI anchor. The present study suggested that the GPI8p could be a novel target for anti-malarial drugs, which provides significant details for further experimentation.

1. Introduction

Malaria is caused by single-cell intracellular parasites belonging to the genus *Plasmodium*, one of the most damaging infectious diseases in the world. According to the World Health Organization (WHO) Malaria Report of 2018, in 2017, about 219 million malaria cases and 435,000 deaths were reported in 87 nations. In most countries, conventional first-line therapies such as chloroquine and pyrimethamine / sulphadoxine have lost their effectiveness, resulting in the use of new and more effective anti-malarial drugs, such as artemisinin-based combination therapy (ACT) (Goswami et al., 2013). While efforts have been made to improve malaria control measures including vaccines and medicines, vaccine escape and drug resistance still present a challenge (Hay et al., 2010).

Glycosylphosphatidylinositol (GPI) anchoring in eukaryotic organisms is a widespread mode of posttranslational modification (McConville and Menon, 2000). A complex enzyme GPI-T mediates the binding of GPI anchors to proteins (McConville and Menon, 2000; Ikezawa, 2002). The complex of GPI-T consists of five distinct subunits and is conserved in different species (Nagamune et al., 2003). At their C termini, proteins intended to be GPI-anchored have a signal sequence, which directs GPI anchoring: GPI transamidase cleaves the signal sequence and replaces it with pre-assembled GPI anchoring (Udenfriend and Kodukula, 1995; Liu et al., 2018). At least two proteins, GAA-1 and GPI8, are essential for the transamidation reaction (Ohishi et al., 2000; Benghezal et al., 1996; Hamburger et al., 1995). GPI8p is most likely the catalytic subunit that generates a carbonyl intermediate with a substrate protein (Ohishi et al., 2001; Spurway et al., 2001; Vidugiriene et al., 2001). GPI biosynthesis is a multi-step process that ultimately results in the addition of a precursor protein to the assembled GPI. This last step in transferring the GPI to a protein is catalysed by the putative transamidase complex using GPI8. GPI8 acts dually to conduct the proteolytic

* Corresponding author.
E-mail addresses: sanjaybiosoft@gmail.com (S.K. Singh), msreddy@thapar.edu (M.S. Reddy).

https://doi.org/10.1016/j.compbiolchem.2021.107461
Received 18 September 2019; Received in revised form 3 November 2020; Accepted 15 February 2021
Available online 17 February 2021

cleavage of the precursor protein's C-terminal signal sequence, accompanied by the creation of an amide bond between the GPI protein and ethanolamine phosphate (Ohishi et al., 2001).

In parasitic protozoa, GPI-APs are particularly abundant (Ferguson, 1999). A GPI-anchored protein in *P. falciparum* called MSP1 is the most abundant merozoite surface component (Das et al., 2015). Merozoites or parasite-infected red blood cells of *P. falciparum* release GPIs that contributes to severe symptoms by causing production of cytokines including TNFα (GPI toxin) (Schofield et al., 2002). The surfaces of all stages of malaria parasite, including the merozoite, gamete, ookinete, and sporozoite are covered with various proteins known or assumed to be GPI-anchored.

There is a dearth of evidence in how the genetic variation in *Plasmodium* can contribute to drug resistance or can provide new drug targets. Genetic variation and recombination have been shown to increase antigen heterogeneity, immune escape, the development of parasite resistance to drugs (Imwong et al., 2010). The genetic variation in *P. falciparum* occurs in the form of single nucleotide polymorphism (SNPs), microsatellite repeats, insertions, deletions and a number of gene duplication. SNP-based barcodes have been useful in identifying malaria parasites from various geographical regions (Preston et al., 2014), indicating that they may be helpful in pre-elimination settings to determine the source of imported infections. Studies have been conducted to predict deleterious nonsynonymous SNPs (nsSNPs) among reported polymorphisms and to understand their impact on protein function, structure and stability (Desai and Chauhan, 2017; Solayman et al., 2017; Joshi et al., 2015; Chitranshi et al., 2017; Firoz et al., 2015). The relevance of SNPs on *P. falciparum* biology has been assessed by multiple studies (Singh and Reddy, 2019; Amambua-Ngwa et al., 2012; Breglio et al., 2018; Campino et al., 2011; Daniels et al., 2008; Subudhi et al., 2015). Nevertheless, to our best knowledge, no research to date has studied the effect of deleterious SNPs experimentally or computationally on *P. falciparum* GPI8p. Hence, the present study aimed to investigate the possible effects of nsSNPs on *P. falciparum* GPI8p and their effects on its structure and function, 3D structure formation and prediction of ligand binding sites in the modelled structure. Since these analytical methods are machine-based algorithms, further validation is required with laboratory testing and clinical evidence to supplement the findings of this study.

2. Materials and methods

2.1. Datasets

Primary data on the *P. falciparum* GPI8p (PF3D7_1128700) gene was retrieved from the NCBI Gene Bank, PlasmoDB and UniProtKB databases. SNPs data of *P. falciparum* GPI8p used in this computational analysis was collected from PlasmoDB database (https://plasmodb.org/plasmo/) while protein sequence in FASTA format was retrieved from the UniProtKB database (http://www.uniprot.org/uniprot/). Analytical processes used for GPI-anchor transamidase protein is depicted in Fig. 1.

2.2. SIFT (Sorting Intolerant from Tolerant)

SIFT predicts tolerated and deleterious SNPs and determines the effect of amino acid substitution on protein function and phenotype alterations. SIFT is a multi-step algorithm using a sequence homology to classify amino acids substitutions. SIFT conducts analysis based on various algorithms and uses Swiss-Prot and TrEMBL to view the homologous sequences. The SIFT predicts substitutions with values less than 0.05 being deleterious (Kumar et al., 2009). The SIFT sequence tool gives SIFT predictions for a given FASTA protein sequence. The sequence of protein queries and interest substitutions of nsSNPs and genes with default parameters have been submitted to http://sift.bii.a-star.edu.sg/www/SIFT_seq_submit2.html.

2.3. PROVEAN (Protein Variation Effect Analyzer)

PROVEAN is a web-based tool that predicts whether a substitution of amino acid affects the biological function of a protein. PROVEAN tool was used to perform BLAST hits clustering to generate the final PROVEAN score, a delta alignment score is calculated for each supporting sequence and then averaged in and across clusters (Choi and Chan, 2015). A default score of -2.5 or higher is considered deleterious, whereas all other scores are neutral.

2.4. PredictSNP web server

PredictSNP (Bendl et al., 2014) is optimized to combine different methods, mostly to annotate disease-variant relationships. For the

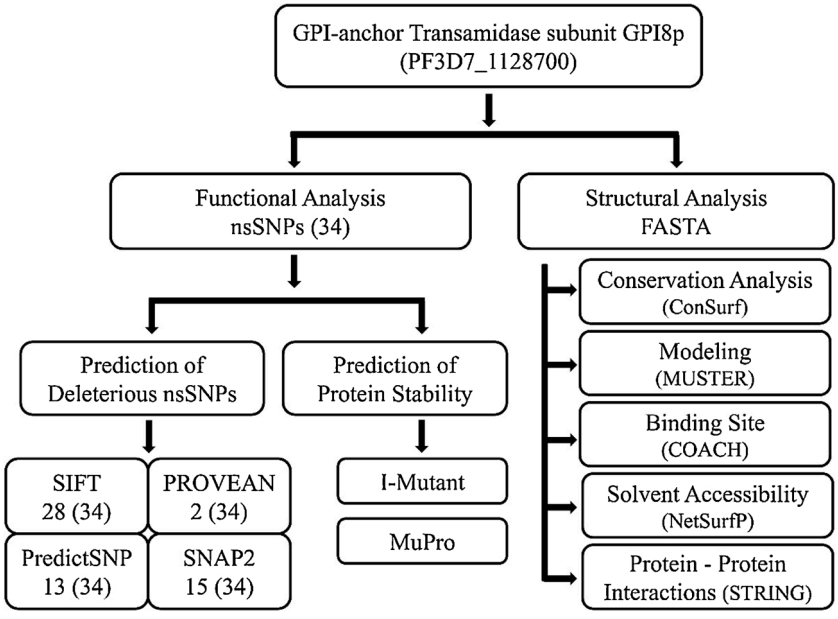


Fig. 1. The schematic representation of the work flow for the analysis performed for GPI8p.

predictions, a FASTA format of amino acid sequence of query protein, mutation positions and desired mutations were submitted using the input page of PredictSNP web server.

2.5. SNAP2 (Screening for Non-acceptable Polymorphisms)

The functional impact of single amino acid substitutions in the GPI8p protein was evaluated using a neural-network-based tool known as SNAP2 (https://rostlab.org/services/snap2web/). The server predicts whether or not a SNP is likely to alter protein function by using different SNPs biophysical characteristics, evolutionary information, and several structural properties. The results include a prediction (effect or neutral) and a score (ranging from-100 to 100). Score between -100 and 0 indicates a neutral prediction while 1–100 indicates an effect (Hecht et al., 2015).

2.6. I-Mutant 2.0

The protein stability changes due to the SNPs effect were predicted by I-Mutant 2.0 (http://folding.biofold.org/cgi-bin/i-mutant2.0.cgi), a SVM-constructed server. The input parameters of this tool are protein sequence, room temperature (25 °C), neutral pH (7.0) and detailed SNP data. The I-Mutant 2.0 tool predicts whether a point mutation stabilizes or destabilizes the native protein structure based on free energy change (detal delta G) (Capriotti et al., 2005).

2.7. MuPro and ConSurf server

MuPro (http://mupro.proteomics.ics.uci.edu/) is a set of machine learning programs used for the prediction of change in protein stability upon amino acid variation (Cheng et al., 2006). ConSurf (consurf.tau.ac.il/) use Bayesian algorithm to analyze the evolutionary conservation of amino acid positions in the protein (Ashkenazy et al., 2016). Conserved regions are identified by conservation scores with a colour scheme and subsequently divided into separate nine-grade scales. Conservation score between 1 and 4 are classified as variable, 5 and 6 intermediate and between 7 and 9 as conserved.

2.8. NetSurfP

The web server NetSurfP-2.0 (http://www.cbs.dtu.dk/services/NetSurfP) is used to predict protein solvent accessibility and secondary structure. NetSurfP also predicts reliability as a Z-score for each prediction, in addition to the ASA and secondary structure prediction (Klausen et al., 2019). The normal and predicted SNP sequences in FASTA format were submitted for prediction to the NetSurfP.

2.9. MUSTER

MUSTER (MUlti-Sources ThreadER) is a useful threading tool for the prediction of protein structure (http://zhanglab.ccmb.med.umich.edu/MUSTER/). MODELLER v8.2, generate a Z-score and completed full-length models. If a Z-score is > 7.5, the corresponding template is classified as good template otherwise a bad template. The six distinct sources used by MUSTER are sequence-derived profiles, secondary structures, structured derived profiles, solvent accessibility, backbone torsion angles and hydrophobic scoring matrix (Wu and Zhang, 2008).

2.10. COACH server

The ligand-binding site (LBS) of the modelled proteins was identified using the COACH program (Yang et al., 2013, 2012). This tool generates structures from submitted protein sequence by using MUSTER tool and then, it matches Protein Data Bank (PDB) format structures with already recognized template structures or known binding sites of experimental protein-ligand structures present in the PDB database. This tool combines the prediction results of various web-based binding sites identification tools like TM-SITE, S-SITE, COFACTOR, FINDSITE and ConCavity to provide a consensus output.

2.11. STRING server

Protein-protein interactions were predicted using STRING (Search Tool for the Retrieval of Interacting Genes/Proteins; https://string-db.org/) (Szklarczyk et al., 2014). Co-expression, physical and functional interactions, pathways, co-localization, protein domain similarity, and predicted interactions are included in the network connection between the genes. The network was filtered by removing all weight interactions below 0.1.

3. Results

The *P. falciparum* GPI8p (PF3D7_1128700) consist of 1482 bp and 493 amino acids. The GPI8p investigated in this study had a total of 40 SNPs, 34 of which were nsSNPs. Only non-synonymous SNPs have been selected for further analysis as non-synonymous mutations may change the protein sequences, which ultimately change structure and function of protein.

3.1. Prediction of deleterious coding nsSNPs

The SIFT sequence tool predicted a total of 28 variants that had an effect on protein function while 4 variants had no effect on GPI8p. Overall, 4 nsSNPs (T81S, Q121 K, R158 L and T195S) were found to be tolerated with a score of greater than 0.05. Three nsSNPs were recognized as deleterious with a score of 0.01 while the remaining 25 nsSNPs showed a highly deleterious score of 0.00. Two nsSNPs (out of 32) were expected to be deleterious with the PROVEAN tool, with a PROVEAN score below -2.5, and the remaining nsSNPs (30) showed scores above the limit recognizing them as neutral. The PROVEAN tool uses -2.5 for all predictions as a cut-off score. The amino acid sequence of the query protein, mutation positions and desired mutations were submitted using the PredictSNP input page in the FASTA format. According to PredictSNP, 13 mutations were expected to be deleterious while 19 were found to be in neutral in GPI8p. The findings from the SNAP2 server showed 15 effective variants, while the remaining 17 nsSNPs were neutral.

By combining the observations of four prediction tools (SIFT, PROVEAN, PredictSNP and SNAP2), 18 nsSNPs (R124 L, N143 K, Y145 F, V157I, T195S, K379E, I392 K, I437 T, Y438H, N439D, Y441H, N442D, N448D, N451D, D457A, D457Y, I458 L and N460 K) have been shown to influence protein function by at least two software tools (Table 1). These nsSNPs were used for further analyses.

3.2. Prediction of mutation impacts on the stability of proteins

In order to predict the free energy change value (DDG) and reliability index (RI) upon mutation, the selected variants were submitted to the I-Mutant 2.0 web server. According to I-Mutant 2.0, the results on amino acid substitutions predicted either an increase or a decrease in the free energy. All of the modified nsSNPs following mutation resulted in a decrease in protein stability, with a range in the reliability index of 1 to 9. Similarly, all the mutations predicted by I-Mutant 2.0 web server were also predicted by MuPro server. The prediction of changes in stability by I-Mutant 2.0 and MuPro of the 18 selected nsSNPs is provided in Table 2. This finding indicates that the amino acid interactions could be directly or indirectly destabilized by GPI8p mutations, leading to functional deviations of the protein.

3.3. Conservation of amino acids

The results of the ConSurf tool consist of a structural representation

Table 1
nsSNPs prediction in GPI8p of *P. falciparum* by using SIFT, PROVEAN, PredictSNP and SNAP2 programs and indicating deleterious by at least two programs at least two programs.

Amino acid change	SIFT	PROVEAN	PredictSNP	SNAP2
R124L	Affect protein function	Neutral	Neutral	effect
N143K	Affect protein function	Neutral	Deleterious	effect
Y145F	Affect protein function	Deleterious	Deleterious	effect
V157I	Affect protein function	Neutral	Neutral	effect
T195S	Tolerated	Deleterious	Deleterious	neutral
K379E	Affect protein function	Neutral	Neutral	effect
I392K	Affect protein function	Neutral	Neutral	effect
I437T	Affect protein function	Neutral	Deleterious	neutral
Y438H	Affect protein function	Neutral	Deleterious	neutral
N439D	Affect protein function	Neutral	Deleterious	effect
Y441H	Affect protein function	Neutral	Neutral	effect
N442D	Affect protein function	Neutral	Deleterious	effect
N448D	Affect protein function	Neutral	Deleterious	effect
N451D	Affect protein function	Neutral	Deleterious	effect
D457Y	Affect protein function	Neutral	Deleterious	effect
D457A	Affect protein function	Neutral	Deleterious	effect
I458L	Affect protein function	Neutral	Deleterious	neutral
N460K	Affect protein function	Neutral	Deleterious	effect

SIFT - Sorting Intolerant from Tolerant, PROVEAN - Protein Variation Effect Analyzer, PredictSNP - Consensus classifiers for prediction of disease-related mutations, SNAP2 - Screening for Non-acceptable Polymorphisms.

Table 2
Prediction of GPI8p protein stability due to mutations analysed through I-Mutant and MuPro programs.

			I-Mutant 2.0			MuPro	
Position	WT	NEW	Stability	RI	DDG	Stability	DDG
124	R	L	Decrease	6	−0.25	Decrease	−0.231
143	N	K	Decrease	1	0.36	Decrease	−1.335
145	Y	F	Decrease	2	−0.05	Decrease	−0.569
157	V	I	Decrease	6	−0.53	Decrease	−0.983
195	T	S	Decrease	7	−0.54	Decrease	−1.178
379	K	E	Decrease	2	−0.71	Decrease	−0.602
392	I	K	Decrease	9	−1.19	Decrease	−1.672
437	I	T	Decrease	7	−1.93	Decrease	−2.639
438	Y	H	Decrease	5	−0.39	Decrease	−1.557
439	N	D	Decrease	2	−0.19	Decrease	−0.458
441	Y	H	Decrease	5	−0.09	Decrease	−1.628
442	N	D	Decrease	1	0.15	Decrease	−0.517
448	N	D	Decrease	2	−0.19	Decrease	−0.517
451	N	D	Decrease	2	−0.19	Decrease	−0.734
457	D	Y	Decrease	5	−0.42	Decrease	−0.434
457	D	A	Decrease	8	−0.94	Decrease	−1.187
458	I	L	Decrease	7	−0.75	Decrease	−1.024
460	N	K	Decrease	5	−0.81	Decrease	−0.800

"WT" indicates the amino acid in native protein, "NEW" is mutant amino acid, "RI" is the reliability index and DDG is the stability (DDG<0: Decrease Stability, DDG>0: Increase Stability). All the mutations in GPI8p were predicted to decrease in protein stability.

of the protein containing a colorimetric conservation score (Fig. 2). Out of the 18 most deleterious SNPs, 2 amino acids with conservation score 9, one with conservation score 8 and 8 with conservation score 6 were predicted by ConSurf. The single amino acid was predicted as average conserved region and 6 residues as variable (Table 3). Positions N143, Y145 and T195 were predicted in highly conserved regions, hence showing more chances to alter the protein structure. The highly conserved residues are often essential for biological function.

3.4. Prediction of solvent accessibility and secondary structure of protein

Secondary structures and solvent accessibility of protein were investigated using NetSurfP-2.0 server. The data was filtered by selecting only those residues that showed an ASA change from buried to exposed state and vice versa and change of its secondary structure. Due to mutations in N143K, T195S, I392K, I392K, N448D, N451D, D457Y, D457A and N460K residues showed a change in class from exposed state to buried state and conformation change from coil to helix. On the other hand, residue I329 in the mutation R124L showed the opposite change to the exposed state from the buried state and also showed an opposite change in conformation from helix to coil (Table 4).

3.5. Protein 3D modeling and structural analysis

The 3D structure of GPI8p was generated using the available protein sequence with homology-based modelling. All the predicted models of GPI8p by MUSTER had a Z score above 8.75. The best predicted model with deleterious residues is shown in Fig. 3. Our findings suggested that all of the related templates could be considered as good models (Table 5).

3.6. Prediction of protein ligand binding site and protein - protein interactions

The structure predicted by MUSTER with template 4fguA (z-score: 11.29) was used as input for COACH server. The best ranked active site of the GPI8p by COACH had a C-score of 0.14. It was predicted R79, H80, G196, H197, G198, D237, C239, S272, Y273, S274, S284, D287 and R288 as consensus binding residues by using 4aw9A PDB hit. The BioLiP server (Yang et al., 2012) shows the protein ligand interaction with residues involved in ligand binding sites as shown in Fig. 4.

Additionally, STRING web server was used to investigate the protein–protein interaction of GPI8p (Fig. 5). STRING results predicted the functional association partner of PF11_0298 protein with PF11_0229 (Conserved *Plasmodium* protein), MAL13P1.165 (GPI transamidase subunit PIG-U, putative), MAL13P1.348 (Uncharacterized protein), PFL0685w (Phosphatidylinositol-glycan biosynthesis class O protein, putative), PFL2270w (GPI mannosyltransferase 2), PIG-M (GPI mannosyltransferase I), PFF0915w (N-acetylglucosamine transferase), PF10_0316 (N-acetylglucosaminyl-phosphatidylinositol biosynthetic protein, putative), PF11_0361 (Uncharacterized protein) and Alg9 (Mannosyltransferase-III). Out of these proteins, PF11_0229 and MAL13P1.165 were involved in the attachment of GPI anchor to protein and PFL0685w, PFL2270w, PIG-M, PFF0915w, PF10_0316 and Alg9 are involve in GPI anchor biosynthetic process. These two MAL13P1.348 and PF11_0361 are uncharacterized proteins.

4. Discussion

Glycosylphosphatidylinositols (GPIs) attach a diverse group of macromolecules to the plasma membrane of eukaryotes (Ferguson, 1994). In parasitic protozoans, the majority of the cell surface proteins are anchored by GPIs and play an important roles in infectivity, survival, virulence and immune evasion (Zacks and Garg, 2006). GPIs and GPI anchors are transported to the cell surface via a process called GPI transamidation, which involves the GPI transamidase (GPI-T) complex

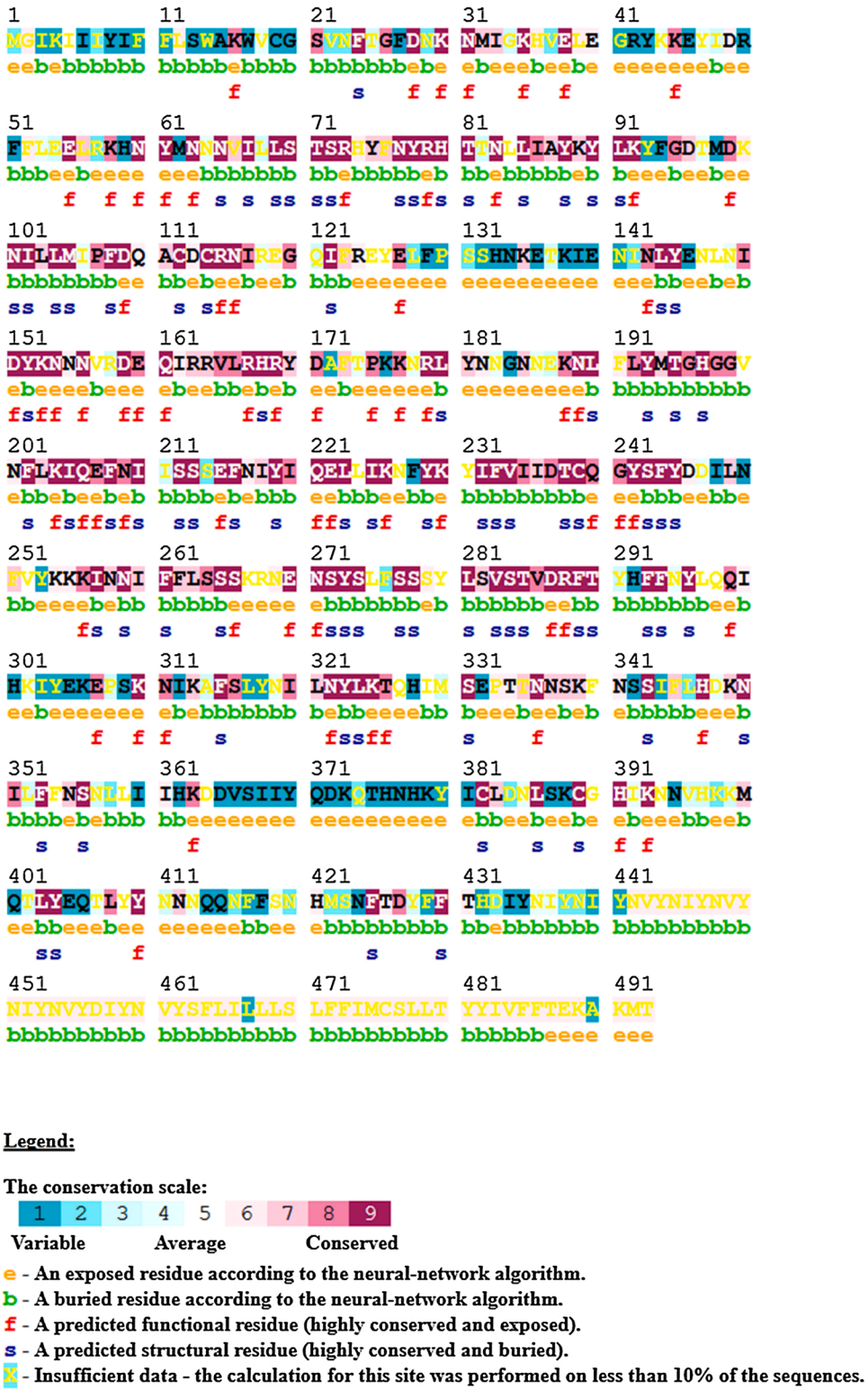


Fig. 2. Unique and conserved regions in the GPI8p protein determined using ConSurf. The color coding bar shows the color scheme representation of the conservation score. Score of conservation is 1–4 for variable, 5–6 for intermediate and 7–9 for conserved.

(Liu et al., 2018). GPI-T mediates the anchoring of GPI in the endoplasmic reticulum by replacing the C-terminal GPI attachment signal peptide of a protein with a fully assembled glycosylphosphatidylinositol (GPI). When infected with *P. falciparum*, glycoproteins play a major role in determining the magnitude and virulence of the parasite. GPI anchoring is the primary form of glycosylation in *P. falciparum* proteins, and the biosynthesis of GPI anchors seems to require much of the ability of the glycosylation machinery (Heng et al., 2010). The differences between mammalian and parasite GPI modifications and the proteins involved in assembling these structural components can reveal potential targets for drugs (Smith, 2009).

About 30 GPI-anchored proteins are express during the asexual stage. In addition to the documented asexual-stage GPI-APs, several GPI-anchored proteins are also found in the *P. falciparum* sexual stage

Table 3
Evolutionary stability of amino acid positions in GPI8p predicted through ConSurf program.

Position	Amino Acid	Conservation Score	ConSurf Prediction
124	R124	6	Exposed
143	N143	8	Highly conserved and exposed (f)
145	Y145	9	Highly conserved and buried (s)
157	V157	5	Buried
195	T195	9	Highly conserved and buried (s)
379	K379	1	Exposed
392	I392	4	Buried
437	I437	4	Buried
438	Y438	2	Buried
439	N439	2	Buried
441	Y441	1	Buried
442	N442	6	Buried
448	N448	6	Buried
451	N451	6	Buried
457	D457	6	Buried
457	D457	6	Buried
458	I458	6	Buried
460	N460	6	Buried

Conservation score is 1–4 for variable, 5–6 for intermediate and 7–9 for conserved. Positions N143, Y145 and T195 were predicted in highly conserved regions, hence showing more chances to alter the protein structure.

(Kurucz et al., 2013). In protozoa parasitic, GPI anchor proteins are particularly abundant (Ferguson, 1999). Sporozoites initiate infection process by GPI-AP coated circumsporozoite proteins (CSP) after releasing from the salivary glands of the vector (Wang et al., 2005). Several GPI-APs such as merozoite surface protein 1 (MSP1) expressed by merozoites cause significant symptoms (Das et al., 2015). Merozoites

Table 5
Z score value of different templates of GPI8p predicted by MUSTER program.

Rank	Template	Align_length	Coverage	Z score	Seq_id	Type
1	4fguA	394	0.799	11.297	0.155	Good
2	5h0iA	377	0.764	11.121	0.175	Good
3	5zbiA	390	0.791	11.003	0.138	Good
4	5nijA	396	0.803	10.846	0.167	Good
5	6idvA	392	0.795	10.407	0.156	Good
6	6dhiA	388	0.787	10.265	0.17	Good
7	5nijA1	293	0.594	9.305	0.181	Good
8	5zbiA1	274	0.555	8.85	0.179	Good
9	6dhiA1	282	0.572	8.781	0.213	Good
10	4fguA1	277	0.561	8.753	0.191	Good

Rank: Top 1 to Top 10 models; Template: the template identified by MUSTER; Align_length: The length of aligned region in the threading results of MUSTER; Coverage: Align_length/target length; Z score = (raw_score-mean_score)/std_score; Seq_id: Sequence identity by threading results; Type: If Z-score >7.5, the corresponding template is a 'Good' template. Otherwise, it is a 'Bad' template.

Table 4
Surface accessibility and secondary structure of wild type and mutant variants of GPI8p analysed through NetSurfP program.

Residue N	Position	ASA N	Class N	SS N	Residue SNP	SNP	ASA SNP	Class SNP	SS SNP	ASA diff
I	329	45.121	B	H	I	R124L	47.459	E	C	−2.338
V	366	40.7	E	C	V	N143K	35.672	B	H	5.028
V	366	40.7	E	C	V	T195S	36.55	B	H	4.15
F	345	57.367	E	C	F	I392K	38.945	B	H	18.422
L	352	48.69	E	C	L	I392K	44.894	B	H	3.796
F	345	57.367	E	C	F	N448D	41.279	B	H	16.088
F	345	57.367	E	C	F	N451D	41.621	B	H	15.746
F	345	57.367	E	C	F	D457Y	40.489	B	H	16.878
N	451	51.017	E	C	N	D457Y	32.574	B	H	18.443
Y	453	53.696	E	C	Y	D457Y	48.285	B	H	5.411
F	345	57.367	E	C	F	D457A	39.302	B	H	18.065
N	451	51.017	E	C	N	D457A	29.643	B	H	21.374
Y	453	53.696	E	C	Y	D457A	43.361	B	H	10.335
F	345	57.367	E	C	F	N460K	42.659	B	H	14.708

Residues that show a change in ASA from buried to exposed state or vice versa, and also show secondary structure change in GPI8p due to SNPs predicted by NetSurfP. Here Residue N and Residue SNP = residues in normal and SNP sequence; ASA N and ASA SNP = ASA of residues in normal and SNP sequence; Class N and Class SNP = Class of residues in normal and SNP sequence [Class = buried (B) or exposed (E)]; SS N and SS SNP = secondary structure of residue in normal and SNP sequence.

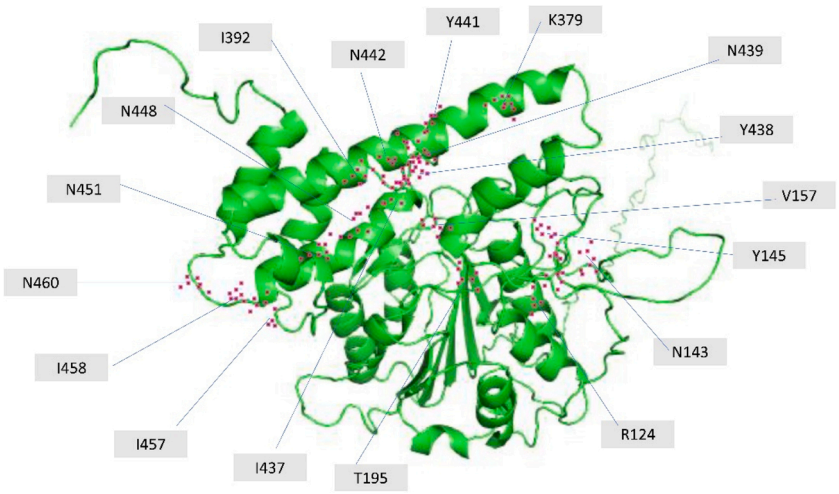


Fig. 3. Modelled 3D structure of GPI-anchor transamidase (GPI8p). The positions of mutation are indicated in red.

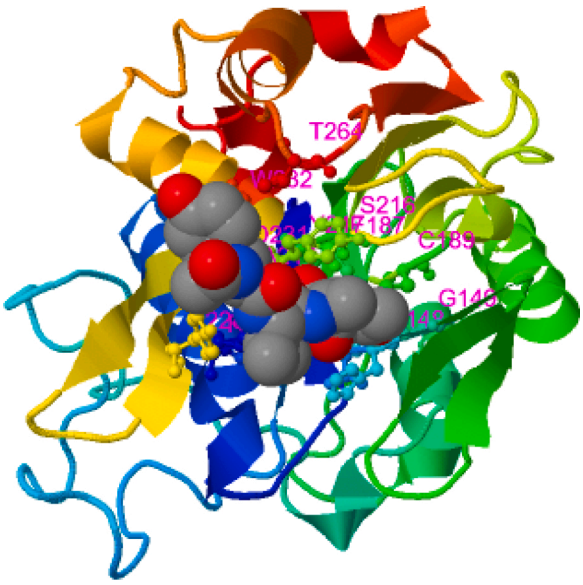


Fig. 4. Ligand-protein interaction with amino acid residues involved in ligand binding sites in the GPI8p template 4aw9A as shown by BioLip.

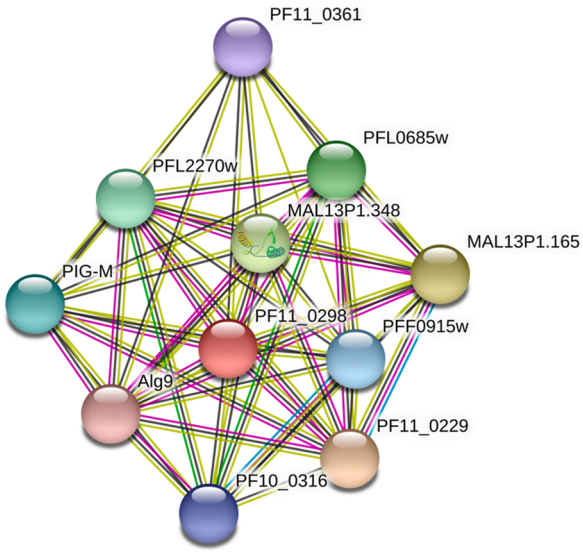


Fig. 5. Protein–protein interaction network of GPI-anchor transamidase (PF11_0298).
STRING interaction network showed that the protein GPI8p interacts with certain proteins that either involve in attachment of GPI anchor to protein or GPI anchor biosynthetic process.

or parasite-infected red blood cells of *P. falciparum* release GPIs that contributes to severe malaria symptoms by causing the production of cytokines like TNFα (GPI toxin) (Schofield et al., 2002).

The complete mechanisms by which a nucleotide variant can trigger a phenotypic shift remain largely unknown to date. However, in silico research using computational methods, including the prediction of the phenotypic effect of non-synonymous SNPs on the physico-chemical properties of the proteins concerned could facilitate the task. Such research is vital for associations between genotype and phenotype, and for understanding the genetics of diseases.

In this study we attempted to identify SNPs that can alter the structure, function and expression of the GPI8p. As described in the methodology section, various bioinformatics tools were used systematically predict deleterious mutations using the SIFT, PROVEAN, PredictSNP and SNAP2 servers. Out of 34 nsSNPs subjected for functional analysis, 18 nsSNPs (Arg124Leu, Asn143Lys, Tyr145Phe, Val157Ile, Thr195Ser, Lys379Glu, Ile392Lys, Ile437Thr, Tyr438His, Asn439Asp, Tyr441His, Asn442Asp, Asn448Asp, Asn451Asp, Asp457Tyr, Asp457Ala, Ile458Leu, and Asn460Lys) were predicted by at least two or more analytical tools. Additionally, I-Mutant 2.0 and MuPro both showed a decrease in stability after mutation as a result of these nsSNPs, suggesting to some extent that the protein was directly or indirectly destabilized. Furthermore, NetSurfP tool revealed that solvent accessibility and secondary structures due to these SNPs were changed mainly from exposed state to buried state and conformation change from coil to helix. Only three nsSNPs found in highly conserved regions were predicted by phylogenetic analysis using ConSurf. This finding suggests that most of highly conserved regions are intact. Additionally, the 3D structure of the protein sequence was generated using MUSTER tool and the best model was used as input by COACH server to predict protein ligand binding site. No mutation was present at the predicted ligand binding site. STRING database interaction network also showed that the protein GPI8p interacts with certain proteins that either involve in attachment of GPI anchor to protein or GPI anchor biosynthetic process.

GPI8s are categorised among the CD clan of cysteine proteases on the basis of amino acid homology and motifs, and are subdivided further based on their peptide bond hydrolysis mechanism into the C13 family (Rawlings and Barrett, 1994; Rawlings et al., 2004). Clan CD has drawn parasitologists interest in part because this clan contains the enzyme that adds a lipid moiety of glycosylphosphatidylinositol (GPI) to the plasma membrane of a variety of parasitic organisms; a secondary modification has been related to *T. brucei* virulence (Lillico et al., 2003). Abe et al. (1993) showed that if mutation occurs in the conserved Cys and His residues in legumain resulted in a loss of protease activity because these are essential to the active site of this protein. The role of GPI8 in protein-GPI anchoring was examined for yGPI8 (Meyer et al., 2000), hGPI8 (Ohishi et al., 2000) and LmGPI8 (Ellis et al., 2002), based on mutation of putative active-site residues. Two histidine and two cysteine residues were conserved in a majority of C13 family members, including the GPI8s, among the potential active-site residues present in yGPI8. Mutation of His54 in yGPI8 (Meyer et al., 2000) and Cys92 in hGPI8 (Ohishi et al., 2000) also resulted in a partial loss of protein-GPI anchoring function, indicating that protein-protein interactions are directly or indirectly affected by these residues (Zacks and Garg, 2006).

Our analysis demonstrated that active site residues H80, C112, H197 and C239 were conserved and no mutation was found. Our finding suggest that GPI8p enzyme could be a good drug candidate. By inhibiting the GPI8p, we can disrupt the GPI anchor biosynthesis pathways and prevents GPI anchoring of protein necessary for parasite virulence. No such experimental research for functional SNPs of GPI8p had been published to date. Thus, a computational approach has been taken to systematically predict deleterious mutations. Studying genetic variation is of practical importance in order to control the disease in endemic regions. Since these computational methods are machine-based algorithms, further evaluation with laboratory testing and clinical evidence is underway in order to complement the results of this report.

5. Conclusion

In this study, many predicted deleterious SNPs were identified in GPI8p and evaluated for their possible deleterious effect on the function and stability of the protein. Of the 34 nsSNPs, 18 nsSNPs were predicted by at least two software as affecting protein function. Hence, solvent accessibility was primarily altered from exposed to buried state and the conformation of secondary structures were changed from coil to helix. Interestingly, ConSurf prediction suggests that most regions of this protein were highly conserved. Additionally, COACH server found no mutation was at predicted ligand binding sites. Inhibition of GPI8p may disrupt the GPI anchor biosynthesis pathways which in turns, prevents GPI anchoring of protein. Therefore, this study provides data on the

functional and structural impact of nsSNPs and conservation of amino acid positions in the protein. These findings can be used to develop therapy to suppress the pathogenic action of *P. falciparum*.

CRediT authorship contribution statement

Sanjay Kumar Singh: Conceptualization, Methodology, Software, Data curation, Writing - original draft. **M Sudhakara Reddy:** Investigation, Conceptualization, Methodology, Supervision, Writing - review & editing.

Declaration of Competing Interest

Authors declare no conflict of interest

Acknowledgment

Authors acknowledge the Department of Biotechnology, Thapar Institute of Engineering & Technology, Patiala, India.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.compbiolchem.2021.107461.

References

Abe, Y., Shirane, K., Yokosawa, H., Matsushita, H., Mitta, M., Kato, I., Ishii, S.I., 1993. Asparaginyl endopeptidase of jack bean seeds. Purification, characterization, and high utility in protein sequence analysis. J. Biol. Chem. 268, 3525–3529.

Amambua-Ngwa, A., Park, D.J., Volkman, S.K., Barnes, K.G., Bei, A.K., Lukens, A.K., Sene, P., Van Tyne, D., Ndiaye, D., Wirth, D.F., Conway, D.J., 2012. SNP genotyping identifies new signatures of selection in a deep sample of West African *Plasmodium falciparum* malaria parasites. Mol. Biol. Evol. 29, 3249–3253. https://doi.org/10.1093/molbev/mss151.

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., Ben-Tal, N., 2016. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res. 44, 344–350. https://doi.org/10.1093/nar/gkw408.

Bendl, J., Stourac, J., Salanda, O., Pavelka, A., Wieben, E.D., Zendulka, J., Brezovsky, J., Damborsky, J., 2014. PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. PLoS Comput. Biol. 10, e1003440 https://doi.org/10.1371/journal.pcbi.1003440.

Benghezal, M., Benachour, A., Rusconi, S., Aebi, M., Conzelmann, A., 1996. Yeast Gpi8p is essential for GPI anchor attachment onto proteins. EMBO J. 15, 6575–6583.

Breglio, K.F., Amato, R., Eastman, R., Lim, P., Sa, J.M., Guha, R., Ganesan, S., Dorward, D.W., Klumpp-Thomas, C., McKnight, C., Fairhurst, R.M., 2018. A single nucleotide polymorphism in the *Plasmodium falciparum* atg18 gene associates with artemisinin resistance and confers enhanced parasite survival under nutrient deprivation. Malar. J. 17, 391. https://doi.org/10.1186/s12936-018-2532-x.

Campino, S., Auburn, S., Kivinen, K., Zongo, I., Ouedraogo, J.B., Mangano, V., Djimde, A., Doumbo, O.K., Kiara, S.M., Nzila, A., Borrmann, S., 2011. Population genetic analysis of *Plasmodium falciparum* parasites using a customized Illumina GoldenGate genotyping assay. PLoS One 6, e20251. https://doi.org/10.1371/journal.pone.0020251.

Capriotti, E., Fariselli, P., Casadio, R., 2005. I-Mutant2. 0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res. 33, 306–310. https://doi.org/10.1093/nar/gki375.

Cheng, J., Randall, A., Baldi, P., 2006. Prediction of protein stability changes for single-site mutations using support vector machines. Proteins Struct. Funct. Bioinform. 62, 1125–1132. https://doi.org/10.1002/prot.20810.

Chitranshi, N., Dheer, Y., Vander Wall, R., Gupta, V., Abbasi, M., Graham, S.L., Gupta, V., 2017. Computational analysis unravels novel destructive single nucleotide polymorphisms in the non-synonymous region of human *Caveolin* gene. Gene Rep. 6, 142–157. https://doi.org/10.1016/j.genrep.2016.08.008.

Choi, Y., Chan, A.P., 2015. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics 31, 2745–2747. https://doi.org/10.1093/bioinformatics/btv195.

Daniels, R., Volkman, S.K., Milner, D.A., Mahesh, N., Neafsey, D.E., Park, D.J., Rosen, D., Angelino, E., Sabeti, P.C., Wirth, D.F., Wiegand, R.C., 2008. A general SNP-based molecular barcode for *Plasmodium falciparum* identification and tracking. Malar. J. 7, 223. https://doi.org/10.1186/1475-2875-7-223.

Das, S., Hertrich, N., Perrin, A.J., Withers-Martinez, C., Collins, C.R., Jones, M.L., Watermeyer, J.M., Fobes, E.T., Martin, S.R., Saibil, H.R., Wright, G.J., 2015. Processing of *Plasmodium falciparum* merozoite surface protein MSP1 activates a spectrin-binding function enabling parasite egress from RBCs. Cell Host Microbe 18, 433–444. https://doi.org/10.1016/j.chom.2015.09.007.

Desai, M., Chauhan, J.B., 2017. Computational analysis for the determination of deleterious nsSNPs in human *MTHFD1* gene. Comput. Biol. Chem. 70, 7–14. https://doi.org/10.1016/j.compbiolchem.2017.07.001.

Ellis, M., Sharma, D.K., Hilley, J.D., Coombs, G.H., Mottram, J.C., 2002. Processing and trafficking of *Leishmania mexicana* GP63. Analysis using GP18 mutants deficient in glycosylphosphatidylinositol protein anchoring. J. Biol. Chem. 277, 27968–27974. https://doi.org/10.1074/jbc.M202047200.

Ferguson, M.A.J., 1994. What can GPI do for you? Parasitol. Today. 10, 48–52. https://doi.org/10.1016/0169-4758(94)90392-1.

Ferguson, M.A., 1999. The structure, biosynthesis and functions of glycosylphosphatidylinositol anchors, and the contributions of trypanosome research. J. Cell. Sci. 112, 2799–2809.

Firoz, A., Malik, A., Singh, S.K., Jha, V., Ali, A., 2015. Identification of hub glycogenes and their nsSNP analysis from mouse RNA-Seq data. Gene 574, 235–246. https://doi.org/10.1016/j.gene.2015.08.012.

Goswami, D., Baruah, I., Dhiman, S., Rabha, B., Veer, V., Singh, L., Sharma, D.K., 2013. Chemotherapy and drug resistance status of malaria parasite in northeast India. Asian Pac. J. Trop. Med. 6, 583–588. https://doi.org/10.1016/S1995-7645(13)60101-7.

Hamburger, D., Egerton, M., Riezman, H., 1995. Yeast Gaa1p is required for attachment of a completed GPI anchor onto proteins. J. Cell Biol. 129, 629–639. https://doi.org/10.1083/jcb.129.3.629.

Hay, S.I., Okiro, E.A., Gething, P.W., Patil, A.P., Tatem, A.J., Guerra, C.A., Snow, R.W., 2010. Estimating the global clinical burden of *Plasmodium falciparum* malaria in 2007. PLoS Med. 7, e1000290 https://doi.org/10.1371/journal.pmed.1000290.

Hecht, M., Bromberg, Y., Rost, B., 2015. Better prediction of functional effects for sequence variants. BMC Genom. 16, S1. https://doi.org/10.1186/1471-2164-16-S8-S1.

Heng, J., Naderer, T., Ralph, S.A., McConville, M.J., 2010. Glycosylated compounds of parasitic protozoa. Microbial Glycobiology, 1st edn. Academic Press, UK, pp. 203–231. https://doi.org/10.1016/B978-0-12-374546-0.00012-2.

Ikezawa, H., 2002. Glycosylphosphatidylinositol (GPI)-anchored proteins. Biol. Pharm. Bull. 25, 409–417. https://doi.org/10.1248/bpb.25.409.

Imwong, M., Dondorp, A.M., Nosten, F., Yi, P., Mungthin, M., Hanchana, S., Das, D., Phyo, A.P., Lwin, K.M., Pukrittayakamee, S., Lee, S.J., 2010. Exploring the contribution of candidate genes to artemisinin resistance in *Plasmodium falciparum*. Antimicrob. Agents Chemother. 54, 2886–2892. https://doi.org/10.1128/AAC.00032-10.

Joshi, B.B., Koringa, P.G., Mistry, K.N., Patel, A.K., Gang, S., Joshi, C.G., 2015. In silico analysis of functional nsSNPs in human *TRPC6* gene associated with steroid resistant nephrotic syndrome. Gene 572, 8–16. https://doi.org/10.1016/j.gene.2015.06.069.

Klausen, M.S., Jespersen, M.C., Nielsen, H., Jensen, K.K., Jurtz, V.I., Soenderby, C.K., Sommer, M.O.A., Winther, O., Nielsen, M., Petersen, B., Marcatili, P., 2019. NetSurfP-2.0: improved prediction of protein structural features by integrated deep learning. Proteins Struct. Funct. Bioinform. 87, 520–527. https://doi.org/10.1002/prot.25674.

Kumar, P., Henikoff, S., Ng, P.C., 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat. Protoc. 4, 1073–1081. https://doi.org/10.1038/nprot.2009.86.

Kurucz, R., Seeberger, P.H., Varón Silva, D., 2013. Glycosylphosphatidylinositols in malaria: GPI biosynthesis and GPI-derived proteins. In: Hommel, M., Kremsner, P. (Eds.), Encyclopedia of Malaria. Springer, New York, NY. https://doi.org/10.1007/978-1-4614-8757-9_22-1.

Lillico, S., Field, M.C., Blundell, P., Coombs, G.H., Mottram, J.C., 2003. Essential roles for GPI-anchored proteins in African trypanosomes revealed using mutants deficient in GPI8. Mol. Biol. Cell 14, 1182–1194. https://doi.org/10.1091/mbc.e02-03-0167.

Liu, Q., Zhao, Y., Zheng, L., Zhu, X., Cui, L., Cao, Y., 2018. The glycosylphosphatidylinositol transamidase complex subunit PbGPI16 of *Plasmodium berghei* is important for inducing experimental cerebral malaria. Infect. Immun. 86, e00929–17. https://doi.org/10.1128/IAI.00929-17.

McConville, M.J., Menon, A.K., 2000. Recent developments in the cell biology and biochemistry of glycosylphosphatidylinositol lipids (review). Mol. Membr. Biol. 17, 1–16. https://doi.org/10.1080/096876800294443.

Meyer, U., Benghezal, M., Imhof, I., Conzelmann, A., 2000. Active site determination of Gpi8p, a caspase-related enzyme required for glycosylphosphatidylinositol anchor addition to proteins. Biochemistry 39, 3461–3471. https://doi.org/10.1021/bi9921860.

Nagamune, K., Ohishi, K., Ashida, H., Hong, Y., Hino, J., Kangawa, K., Inoue, N., Maeda, Y., Kinoshita, T., 2003. GPI transamidase of Trypanosoma brucei has two previously uncharacterized (trypanosomatid transamidase 1 and 2) and three common subunits. Proc. Natl. Acad. Sci. U. S. A. 100, 10682–10687. https://doi.org/10.1073/pnas.1833260100.

Ohishi, K., Inoue, N., Maeda, Y., Takeda, J., Riezman, H., Kinoshita, T., 2000. Gaa1p and gpi8p are components of a glycosylphosphatidylinositol (GPI) transamidase that mediates attachment of GPI to proteins. Mol. Biol. Cell 11, 1523–1533. https://doi.org/10.1091/mbc.11.5.1523.

Ohishi, K., Inoue, N., Kinoshita, T., 2001. PIG-S and PIG-T, essential for GPI anchor attachment to proteins, form a complex with GAA1 and GPI8. EMBO J. 20, 4088–4098. https://doi.org/10.1093/emboj/20.15.4088.

Preston, M.D., Campino, S., Assefa, S.A., Echeverry, D.F., Ocholla, H., Amambua-Ngwa, A., Stewart, L.B., Conway, D.J., Borrmann, S., Michon, P., Zongo, I., 2014. A barcode of organellar genome polymorphisms identifies the geographic origin of *Plasmodium falciparum* strains. Nat. Commun. 5, 1–7. https://doi.org/10.1038/ncomms5052.

Rawlings, N.D., Barrett, A.J., 1994. Families of cysteine peptidases. Methods Enzymol 244, 461–486. https://doi.org/10.1016/0076-6879(94)44034-4. Academic Press.
Rawlings, N.D., Tolle, D.P., Barrett, A.J., 2004. MEROPS: the peptidase database. Acids Res. 32, 160–164. https://doi.org/10.1093/nar/gkp971.
Schofield, L., Hewitt, M.C., Evans, K., Siomos, M.A., Seeberger, P.H., 2002. Synthetic GPI as a candidate anti-toxic vaccine in a model of malaria. Nature 418, 785–789. https://doi.org/10.1038/nature00937.
Singh, S.K., Reddy, S.M., 2019. Investigation of hub genes and their nonsynonymous single nucleotide polymorphism analysis in *Plasmodium falciparum* for designing therapeutic methodologies using next-generation sequencing approach. Indian J. Pharmacol. 51, 389–399. https://doi.org/10.4103/ijp.IJP_535_19.
Smith, T.K., 2009. Inhibitors of GPI biosynthesis. Enzymes 26, 247–267. https://doi.org/10.1016/S1874-6047(09)26012-4.
Solayman, M., Saleh, M.A., Paul, S., Khalil, M.I., Gan, S.H., 2017. In silico analysis of nonsynonymous single nucleotide polymorphisms of the human adiponectin receptor 2 (ADIPOR2) gene. Comput. Biol. Chem. 68, 175–185. https://doi.org/10.1016/j.compbiolchem.2017.03.005.
Spurway, T.D., Dalley, J.A., High, S., Bulleid, N.J., 2001. Early events in glycosylphosphatidylinositol anchor addition. Substrate proteins associate with the transamidase subunit gpi8p. J. Biol. Chem. 276, 15975–15982. https://doi.org/10.1074/jbc.M010128200.
Subudhi, A.K., Boopathi, P.A., Pandey, I., Kaur, R., Middha, S., Acharya, J., Kochar, S.K., Kochar, D.K., Das, A., 2015. Disease specific modules and hub genes for intervention strategies: a co-expression network based approach for *Plasmodium falciparum* clinical isolates. Infect. Genet. Evol. 35, 96–108. https://doi.org/10.1016/j.meegid.2015.08.007.
Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., Kuhn, M., 2014. STRING v10: protein–protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 43, 447–452. https://doi.org/10.1093/nar/gku1003.
Udenfriend, S., Kodukula, K., 1995. How glycosyl-phosphatidylinositol-anchored membrane proteins are made. Annu. Rev. Biochem. 64, 563–591. https://doi.org/10.1146/annurev.bi.64.070195.003023.
Vidugiriene, J., Vainauskas, S., Johnson, A.E., Menon, A.K., 2001. Endoplasmic reticulum proteins involved in glycosylphosphatidylinositol-anchor attachment: photocrosslinking studies in a cell-free system. Eur. J. Biochem. 268, 2290–2300. https://doi.org/10.1046/j.1432-1327.2001.02106.x.
Wang, Q., Fujioka, H., Nussenzweig, V., 2005. Mutational analysis of the GPI-anchor addition sequence from the circumsporozoite protein of *Plasmodium*. Cell. Microbiol. 7, 1616–1626. https://doi.org/10.1111/j.1462-5822.2005.00579.x.
Wu, S., Zhang, Y., 2008. MUSTER: improving protein sequence profile–profile alignments by using multiple sources of structure information. Proteins Struct. Funct. Bioinform. 72, 547–556. https://doi.org/10.1002/prot.21945.
Yang, J., Roy, A., Zhang, Y., 2012. BioLiP: a semi-manually curated database for biologically relevant ligand–protein interactions. Nucleic Acids Res. 41, 1096–1103. https://doi.org/10.1093/nar/gks966.
Yang, J., Roy, A., Zhang, Y., 2013. Protein–ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. Bioinformatics 29, 2588–2595. https://doi.org/10.1093/bioinformatics/btt447.
Zacks, M.A., Garg, N., 2006. Recent developments in the molecular, biochemical and functional characterization of GPI8 and the GPI-anchoring mechanism. Mol. Membr. Biol. 23, 209–225. https://doi.org/10.1080/09687860600601494.