

**Towards discovering novel aspects of nuclear biology in
the malaria parasite *Plasmodium falciparum***

Inauguraldissertation

zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von

Sophie Clara Oehring
aus Deutschland

Basel, 2012

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von
Dr. Till S. Voss und Dr. Volker Heussler

Basel, den 26. Juni 2012

Prof. Dr. Martin Spiess
Dekan

Table of contents

Acknowledgments	3
Summary	5
Chapter 1: Introduction	7
1.1 Malaria – an overview.....	7
1.2. Prevention and treatment of malaria.....	9
1.2.1. Insecticides and chemotherapy.....	9
1.2.2. Vaccines.....	10
1.3. The <i>Plasmodium</i> nucleus.....	11
1.3.1. The genome of <i>P. falciparum</i>	13
1.3.2. The transcriptome of <i>P. falciparum</i>	14
1.3.3. Transcription factors in <i>P. falciparum</i>	17
1.3.4. The epigenome of <i>P. falciparum</i>	20
1.3.4.1. Histone modifications in <i>P. falciparum</i>	21
1.3.4.2. Chromatin-interacting factors in <i>P. falciparum</i>	22
1.4. The proteome of <i>P. falciparum</i>	24
1.5. Specific objectives of the thesis.....	27
1.6 References.....	28
Chapter 2: Organellar proteomics reveals hundreds of novel nuclear proteins in the malaria parasite <i>P. falciparum</i>	35
Abstract.....	36
Background.....	37
Results.....	39
Isolation and fractionation of parasite nuclei.....	39
Nuclear proteome determination by MudPIT.....	40
The preparation of isolated nuclei is enriched in nuclear proteins.....	41
Application of bioinformatic reduction techniques to generate a high-confidence core nuclear proteome dataset.....	42
Experimental validation by subcellular localisation of selected nuclear candidates.....	43
Two novel nucleolar proteins in <i>P. falciparum</i>	45
Identification of novel domains associated with <i>P. falciparum</i> nuclear proteins.....	45
Classical nuclear localisation signals are only marginally over-represented in <i>P. falciparum</i> nuclear proteins.....	46
Stage-specific patterns of nuclear proteins.....	47
Discussion.....	48
Functional classification of the core nuclear proteome.....	48
Fraction-specific aspects of the core nuclear proteome.....	51
Nuclear import in <i>P. falciparum</i>	52

Table of contents

Lineage-specific nuclear proteins	52
Conclusion	53
Materials and Methods	54
References.....	57
Chapter 3: Complex formation of three bromo-domain proteins of <i>P. falciparum</i>	141
Introduction	141
Results	144
Nuclear extraction properties and putative interaction partner candidates of the bromo-domain protein PF10_0328 in <i>P. falciparum</i>	144
Validation of the complex formation of three <i>P. falciparum</i> bromo-domain containing proteins	146
Conclusion and Outlook	148
Materials and Methods	151
References.....	154
Chapter 4: Visualizing the inner life of the nuclear compartment of <i>P. falciparum</i>	163
.....	163
Introduction	163
Results	167
Visualization of the chromosome ends in live and fixed <i>P. falciparum</i> parasites	167
Visualization of the nuclear pores across the asexual life cycle of <i>P. falciparum</i> using 3D7/PFI0170w-GFP and 3D7/PF14_0442-GFP cell lines.....	168
Size increase of the nucleolar compartment at the trophozoite stage of <i>P. falciparum</i> using 3D7/PF11_0250-3xHA.....	169
Interrelated visualization of the silenced chromosome ends, the nuclear pores and the nucleolus in <i>P. falciparum</i>	169
Conclusion and Outlook	170
Materials and Methods	172
References.....	174
Chapter 5: General discussion and Outlook	187
Aspects underlying recent findings in <i>P. falciparum</i> transcriptional regulation	187
Conserved 'reader' proteins of epigenetic histone marks in <i>P. falciparum</i>	190
Nuclear landmark structures in <i>P. falciparum</i>	193
References.....	199
Curriculum vitae.....	203

Acknowledgments

Special thanks go to my supervisor Prof. Dr. Till Voss for guarding me through my thesis. Furthermore, I would like to thank the members of the GR group and many great people from the Swiss Tropical and Public Health Institute, in particular Caroline Kulangara, Annette Gaida and Anna Perchuc for their support during the last four years and Céline Chassot for correcting my thesis.

Furthermore, I would like to thank my collaborators: Paul Jenö and Suzette Moes for successful mass spectrometry analysis. Stuart Ralph, Ben Woodcroft and Chaitali Dekiwadia for great bioinformatics and EM pictures. Tim Gilberger, Mike Duffy and Michael Terns for kindly providing antibodies. And finally, Chris Tonkin for giving nice advices concerning IFAs.

Finally, I would like to thank my parents Hannelore and Eckart, my sister Lisa and her husband Jorge including their cute children/my nephews Noé and Laia.

Summary

The apicomplexan parasite *P. falciparum* continues to cause morbidity and mortality imposing a significant health and economic burden on human society. In light of antimalarial drug resistance and the lack of an effective vaccine there is an urgent need to understand the basic biology of *Plasmodium* parasites in much greater detail. In particular, basic nuclear processes such as those remain surprisingly unsought despite their importance in parasite survival and life cycle progression. Thus, identification and localisation of novel parasite proteins to areas of the nucleus is an important first step towards giving new insights into nuclear architecture and function. The main aim of this thesis was to compile an inventory of the nuclear proteome across the intra-erythrocytic cell cycle using high accuracy mass spectrometry coupled with bioinformatics and *in vivo* localisation experiments. The dataset was analysed for accuracy and retention of true nuclear proteins revealing a final list of 802 potential nuclear proteins with an estimated precision of 76%. Interestingly, the informational pool of this study was able to identify a large number of novel nuclear components including novel protein domains possibly involved in gene regulation, members of the nuclear pores, the nucleolus and the proteasome (chapter 2). Several transgenic parasite lines used for the experimental validation part of the nuclear core proteome were further investigated in more detail. One of these transgenic cell lines expresses the C-terminally tagged bromo-domain protein PF10_0328 and was investigated by co-immunoprecipitation experiments followed by LC-MS/MS to identify interacting proteins. Bromodomain proteins bind specifically to acetylated lysine residues in histone tails and are important regulators of transcription. Our results suggest that PF10_0328 acts in concert with two additional bromo-domain proteins in regulating transcription in *P. falciparum* (chapter 3). Further characterisation on the functional level of these three important regulators is currently ongoing in a collaborative effort. Characterisation of bromo-domain proteins could establish new intervention strategies against malaria as the recognition of acetylated histone tails by bromo-domains can be selectively prevented by small molecules. Furthermore, several proteins residing in the nuclear pores and the nucleolus of *P. falciparum* were used to visualise these structures in relation to chromosome end clusters based on fluorescence microscopy. We show that both structures, involved in nuclear-cytoplasmic trafficking and ribosomal biogenesis, respectively, do not appear to 'cross-talk' with silenced chromosome ends at the nuclear periphery of *P. falciparum* (chapter 4). In conclusion, I

Summary

believe that my work about several aspects of gene regulation and nuclear architecture increases the understanding of the biology of this medically important pathogen and could have potential to identify new avenues for interventions against malaria.

Chapter 1: Introduction

1.1 Malaria – an overview

The malaria disease is caused by the *apicomplexan* parasite *Plasmodium*. These protozoan parasites include over 100 species and infect land-living mammals, reptiles and birds. Sanguivorous mosquitoes act as vectors transmitting the *Plasmodium* species to vertebrates. In mammals, several different species can function as hosts including primates, monkeys, rodents and humans. The four classic *Plasmodium* species which naturally infect humans are *P. vivax*, *P. ovale*, *P. malariae* and *P. falciparum* [1]. *P. knowlesi*, originally the causative agent for malaria in macaques of South-East Asia, showed recently an increased transmission rate to humans [2]. *P. ovale* and *P. malariae* malaria cases are rare and substantially less dangerous compared to *P. falciparum* and *P. vivax* cases. *P. falciparum* is responsible for the most severe form of human malaria causing the majority of malaria morbidity and mortality, mainly in Sub-Saharan Africa. *P. vivax* infections are dominant in the remaining malaria-endemic regions of the world (South-East Asia, Latin- and South America).

The parasite *P. falciparum* requires two different hosts to complete its complex life cycle, the human and the mosquito host (Figure 1). The female *Anopheles* mosquito transmits haploid sporozoites to the human host by taking a blood meal. The sporozoites reach the blood stream within seconds and migrate to the liver where they invade hepatocytes. Schizogony takes place after an incubation time of five to twenty days and thousands of merozoites are released into the blood stream [3]. Erythrocytes are infected independently of their maturity stage, and parasite develop intracellularly within a parasitophorous vacuolar membrane (PVM), The intra-erythrocytic developmental cycle (IDC) starts with the ring stage (0-22 hours post invasion (hpi)), continues with the trophozoite stage (22-32 hpi) and finishes with the schizont stage (32-48 hpi). The mature parasite finally bursts the red blood cell and produces 24-32 new merozoites, which initiate new rounds of intra-erythrocytic schizogony. Some of the asexually produced parasites develop into sexual forms called gametocytes. Transmission of gametocytes back to the mosquito happens again through blood meal. The gametocytes develop in the mosquito gut into distinct sexual forms, either flagellated microgametes (male) or macrogametes (female). Both gametes form the diploid zygote, which further

develops into the ookinete. The motile ookinete penetrates the gut wall and matures to an oocyst during ten days. Finally, the oocyst ruptures and releases haploid sporozoites into the hemolymph system of the mosquito. The sporozoites finally reach the salivary glands and are transmitted back to the human host through the next blood meal.

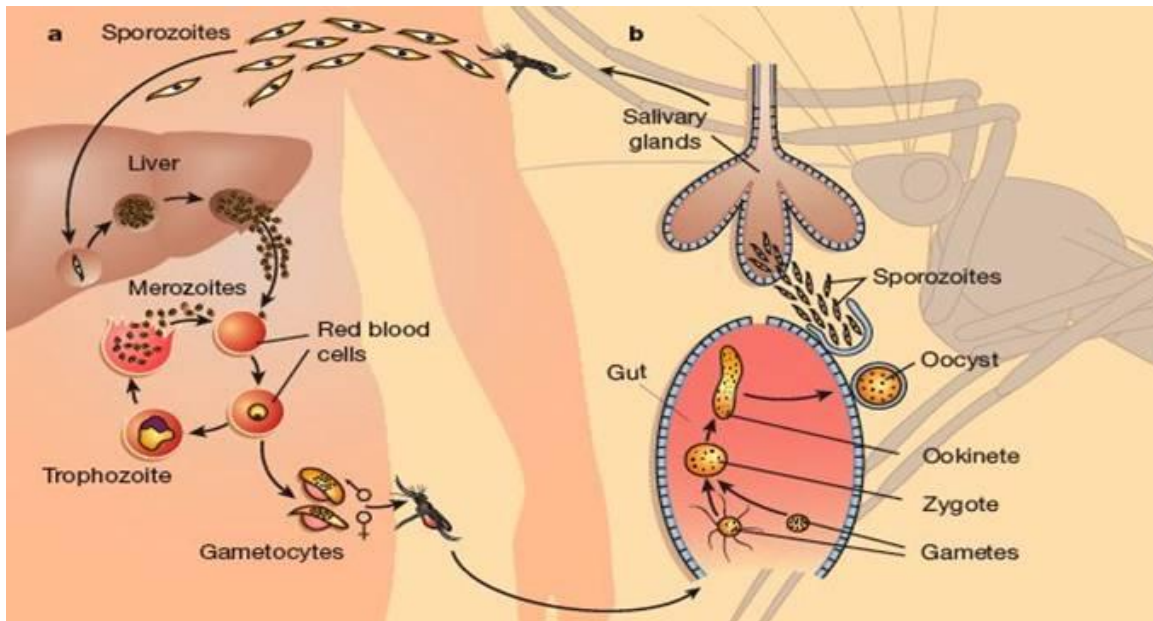


Figure 1: Life cycle of *Plasmodium falciparum* [4].

The replication of the parasite within red blood cell (RBC) causes anemia and clinical symptoms like fever, shivering, arthralgia and vomiting. The common knowledge about malaria displays the cyclical occurrence of sudden coldness followed by fever lasting four to six hours. In *P. vivax* and *P. ovale*-infected patients, the recurrent fever appears every 2 days and for *P. malariae* every 3 days. *P. falciparum* causes disease with fever every 36-48 hours or a less pronounced and continuous fever course. Sequestration of *P. falciparum*-infected RBCs in the capillaries of internal organs like the heart, placenta or liver can progress to organ failure, in severe cases to coma or death. In cerebral malaria, cytoadherence by *P. falciparum* infected erythrocytes occurs in the microvasculature system of the brain and appears to be the principal cause of the fatal outcome of the disease.

1.2. Prevention and treatment of malaria

Published data in the WHO World Malaria Report 2011 inform about the progress of recommended implementations to prevent and treat malaria in 106 endemic countries [5]. Main goals of the prevention measures are the improvement of vector control through the use of insecticide-treated nets (ITNs) as well as chemoprevention for the most vulnerable populations. Intermittent preventive treatment of pregnant women and infants is recommended in areas of high transmission in Sub-Saharan African countries. Main targets for efficient treatment are patients with confirmed malaria diagnosis by microscopy and rapid diagnostic tests, followed by application of appropriate malaria medicines. In 2010, 3.3 billion people have been reported to be at risk of acquiring malaria, which is equivalent to nearly half of the world's population. The highest risk is carried by people living in Sub-Saharan Africa where 81 % of cases and 91% of deaths occur, mainly amongst pregnant women and children under five years. International funding for malaria control increased continuously, reaching up to 2 billion US\$ in the year 2011 [5].

1.2.1. Insecticides and chemotherapy

Eradication of malaria is hampered by the evolution of insecticide resistance in the mosquito vectors, and drug resistance in the parasite during their residence in the human host. Resistance in the mosquito occurs against the single class of pyrethroid insecticides, which are widely used for ITNs in Sub-Saharan Africa. Monitoring of insecticide resistance is necessary to evaluate the effectiveness of vector control. Regarding drug treatment, the WHO recommends artemisinin-based combination therapy (ACT) for uncomplicated *P. falciparum* malaria and explicitly discouraged artemisinin-based monotherapies. ACTs combine artemisinin with lumefantrine, mefloquine, amodiaquine, sulfadoxine/pyrimethamine, piperaquine and chlorproguanil/dapsone. Chloroquine should be used for *P. vivax* infections in combination with a 14 day course with primaquine to prevent relapses. Alternatively, ACTs can be used in areas where *P. vivax* is resistant against chloroquine [5].

Anti-malarial drugs can be classified by their mode of action. One class, the folate antagonists, include sulfonamides, sulfones or pyrimethamines, inhibit specific enzymes of the folate synthesis pathway and ultimately pyrimidine synthesis [6]. Atovaquone, a hydroxyl-naphthoquinone, is always applied in combination with proguanil and seems to

inhibit the dihydroorotate dehydrogenase, by blocking the electron transport chain and pyrimidine synthesis [7][6]. The mode of action of a second class, the quinoline-containing drugs like chloroquine (4-aminoquinoline) and mefloquine/halofantrine (acrylamino alcohols quinine), affects the haem detoxification in the food vacuole (FV) of the parasite [8]. Finally, artemisinin and derivatives (artemether and artesunate) are hypothesized to kill the parasites by the production of free radicals in the food vacuole, where high iron concentrations reduce its peroxide structure [6].

1.2.2. Vaccines

In addition to vector control and antimalarial drugs, an effective vaccination is urgently needed to better control or even eradicate the malaria disease. However, no effective malaria vaccine is available to date mainly due to the complexity of the parasite life cycle. The parasite *P. falciparum* proceeds through several stages in the human host, with stage-specific protein expression, high levels of antigen polymorphism and redundancy of essential invasion pathways [9]. Moreover, the understanding of interactions between the human immune system and the parasite is poorly understood, including the different immune evasion mechanisms utilised by the parasite. Malaria candidate vaccines are categorised according to the life-cycle stage at which the parasite expresses the target antigen and are divided into pre-erythrocytic, asexual blood stage and transmission blocking vaccines. The pre-erythrocytic vaccines aim to prevent the development of the asexual blood stages. Vaccines targeting the asexual blood stages should control the level of parasitemia, and transmission-blocking vaccines aim to inhibit the transmission of the parasites from the human host to the vector.

The most advanced vaccine in development is named RTS,S and consists of parts of the hepatitis B surface antigen (HBsAg) fused to its central repeat and the thrombospondin domain of the circumsporozoite surface protein (CSP) [10][11]. CSP is the major surface protein of the sporozoite stage and the protein part selected for the vaccine is responsible for the development of sporozoite. Phase III trials in Mozambique, Tanzania and Kenya reported 30% to 50% protection from clinical malaria in infants and children which were immunised with RTS,S [11][12]. One possibility to explain this partial protection is that the pre-erythrocytic vaccination with RTS,S reduces the incidences and numbers of infected hepatocytes and thereby the number of merozoites starting blood

stage infection. However, efforts to improve efficacy and prolongation of protection time are still required to optimise RTS,S vaccination.

Multistage vaccine strategies targeting additionally the blood stages could increase the developmental delay of the parasite and decrease the probability to reach parasitaemias for disease threshold. In a recent study, a randomised phase 1b trial, virosome-formulated *P. falciparum* apical membrane antigen 1 (AMA1) and CSP-derived peptides were used to vaccinate semi-immune adults and children [13]. The merozoite antigen AMA1 spreads onto the surface of merozoites and is involved in re-infecting new red blood cells [14]. Incidence rates of clinical malaria in children between day 30 and day 365 were reported to be half of the rate of the control [13]. Besides combining stage specific immunisation targets, an alternative successful vaccination method is to use live-attenuated sporozoites [15][16]. γ -radiation of sporozoites does not affect their survivability or ability to invade hepatocytes, but leads to a defective development into liver stages and results in growth arrest [16]. However, major challenges of producing metabolic active and non-replicating sporozoites are the safety and production costs.

1.3. The *Plasmodium* nucleus

Since the year 1942, numerous studies using emerging microscopic techniques expanded the knowledge about the nuclear fine structure of different *Plasmodium* species [17][18]. Despite morphological changes of the parasite across the asexual stages, the nucleus appears always round to elongated in shape (approximately 500-800nm in diameter) containing electron-sparse and electron-dense material. Regions of condensed genetic material are well distinguishable from relaxed genetic material consistent with heterochromatin (silent) and euchromatin (active), respectively. Interestingly, a recent study describing the 3D nuclear architecture of *P. falciparum* uncovered dramatic changes in chromatin organization showing characteristic patterns at each developmental stage during the IDC [19]. Ring stage parasites contain nuclei where compact heterochromatin is diffusely and equally distributed within the compartment. At the trophozoite stage, small islands of condensed genetic material can be observed distributed within the nucleoplasm compared to the schizont stage at which a 'sub-compartmentalization' of eu- and heterochromatin is visible [19]. Moreover, the inner side of the nuclear membrane in general creates a functional environment of silent

heterochromatin whereas the central core of the nucleus is suggested to be transcriptionally active in *P. falciparum* [20][21][19]. Interestingly, recent observations revealed that both nuclear landmark structures, the chromosome ends and the centromeres, localise to the nuclear periphery of *P. falciparum* [22]. FISH experiments revealed that the chromosome ends are organized as two to seven perinuclear clusters [23], which was suggested to increase parasite virulence by generating new antigenic variants through recombination between heterologous chromosomes [24][25]. The centromeres of *P. falciparum* are observed to occupy a single nuclear location spacially separated from the chromosome end domains by using IFA experiments [22]. These specialized chromosomal regions are essential in all eukaryotes as they allow the equal segregation of the chromosomes in preparation for cell division during mitosis. This process is a critical step in the life cycle of *P. falciparum* as the parasite undergoes three to four rounds of DNA synthesis/mitosis followed by orchestrated cytokinesis to rapidly produce the progeny cells (16 to 32 nuclei per schizont parasite) [26][27]. The process of mitosis in *P. falciparum* is characterised by several unusual aspects compared to its traditional view and many questions remain open to date [28][29]. The creation of multinuclear parasites proceeds asynchronously [17][26][30][31][32], which makes it difficult to observe the transition phase between DNA synthesis and mitosis [33].

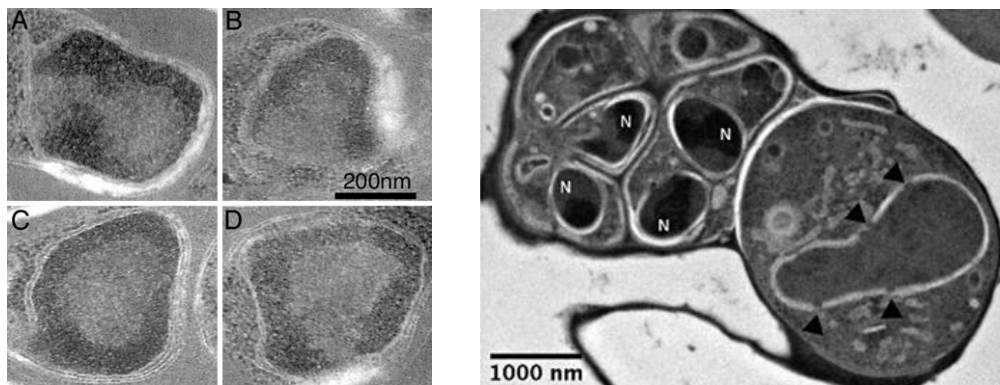


Figure 2: Ultrastructure of *P. falciparum* nuclei.

Left: Segmented schizont nuclei containing regions of condensed heterochromatin additional to gaps of loose euchromatin at the nuclear periphery of *P. falciparum* [20]. Right: RBC infected with a late schizont and a trophozoite parasite where nuclei show heterochromatin patterns of well-defined, compact regions or diffuse distributed patches, respectively (N, nucleus, arrowheads indicate nuclear pores) [19].

1.3.1. The genome of *P. falciparum*

One unchallenged milestone in malaria research is the publication of the complete genome of *P. falciparum* in the year 2002, providing unprecedented possibilities to understand the complex biology of the parasite [34]. The 22.8 Mb haploid genome of the 3D7 reference strain consists of 14 linear chromosomes and encodes 5400 open reading frames (ORFs) characterised by a unique AT-richness [34]. Remarkably, more than 50% of all genes are annotated as hypothetical proteins since they encode proteins that show no sufficient similarities to proteins in other organisms to make functional predictions possible. The overall AT- content lies around 80% and pervades throughout the complete genome, with a higher percentage in intergenic regions and a lower percentage in protein encoding regions simply due to the universal genetic code [34]. The average gene length of *P. falciparum* is 2.3 kb, which is larger compared to values ranking from 1.3 to 1.6kb in other eukaryotes [34]. For example, the genome size of the fission yeast *Schizosaccharomyces pombe* is half of *P. falciparum* but comprehends the same number of ORFs.

Besides the chromosomal DNA within the nucleus, the parasite *P. falciparum* possesses a 6 kb linear mitochondrial genome [35] and a 35 kb circular plastid genome in the apicoplast (or plastid) [36]. The apicoplast represents a homologue of the chloroplast of plants [37][38] that was acquired through secondary endosymbiosis [39][40][41][42], and functions in the essential anabolic synthesis of fatty acids [43], isoprenoids [44] and haem [45][46]. The nuclear and plastid genomes encode known transfer RNAs (tRNA) genes, in contrast to the mitochondrial organelle which contains no tRNAs [47][35]. The chromosomal length of the nuclear genome varies considerably between approximately 0.643 and 3.29 Mb and variation occur mainly in the subtelomeric regions [34][48]. Chromosomes of *P. falciparum* contain a single stretch of short tandem repeats with AT-contents of over 97%, which were identified initially by sequence comparisons as putative centromeres [34]. Recently, centromere identity has been confirmed on the functional and structural levels [49][50][22]. Interestingly, the configuration of rRNA genes in the genome of *P. falciparum* is not consistent with the common eukaryotic formation of tandem repetitive clusters of rRNA genes [51][52]. *P. falciparum* encodes several single 18S-5.8S-28S units located on different chromosomes with sequence differences between the same rRNA genes of different units [53]. Furthermore, two types of rDNA genes are differentially transcribed during the life cycle, the A-type rDNA genes

are transcribed primarily in the human host and the S-types in the mosquito vector [54][55]. Thus, the parasite is able to regulate developmentally its different copies and types of rRNA units, resulting in the expression of different sets of rRNAs at different life cycle stages [53].

1.3.2. The transcriptome of *P. falciparum*

The complex life cycle of *P. falciparum* requires the tight activation or repression of thousands of genes with highly differentiated functions at the right time. Le Roch and collaborators used a high density oligonucleotide array to generate expression profiles of the human and mosquito stages of *P. falciparum* to correlate genes of the same temporal expression to predict possible functional clusters [56]. As expected, the most abundant transcripts encoded proteins of the ribosomal machinery, the glycolysis pathway, histones and actin. Half of the expressed genes were constitutively transcribed including many housekeeping and hypothetical genes, in contrast to specific expression patterns of genes encoding cell cycle-dependent proteins. A total of 15 co-regulated clusters were defined and proposed to be required to survive in the rapid changing environment of the human and mosquito host. These results suggested that groups of functionally related genes that participate in highly differentiated processes during the life cycle share common expression profiles. Detailed analysis of the asexual life cycle by Bozdech *et al.* even improved the picture of the concerted expression describing the transcriptional profile of *P. falciparum* as a continuous cascade during the 48 hour period [57]. Expression patterns of 60 % of genes, many of which are involved in very specific processes, show a “just-in-time” profile, expressed only once per cycle at the required time. The different phases of the cascade show characteristic smooth transitions and key cellular functions proceed in a distinct chronological order. Besides the concerted expression of functionally related genes, Le Roch *et al.* hypothesised that transcriptional regulation in *P. falciparum* could be partly controlled through localisation on the same chromosome [56]. Their results showed that several genes involved in the erythrocyte refurbishment share subtelomeric localisation in contrast to genes involved in general functions like parasite maintenance and growth, which are centrally located. However, Bozdech *et al.* explained the putative ‘co-regulation’ of genes sharing the same chromosomal positioning due to the similar promoter forces [57].

Several examples extracted from Bozdech and colleagues [57] are mentioned here to illustrate this peculiar mode of transcriptional regulation in *P. falciparum*. Profiling data showed that some nuclear-encoded genes with functions in transcription and translation were expressed at early ring stages, and a distinct set of genes of the same functional class were transcribed during later stages, including those of mitochondrion/plastid origin. Thus, protein synthesis of both the mitochondrion and the apicoplast compartments appears to correlate with organellar maturation. No transcripts encoding proteins linking the glycolysis pathway with the TCA cycle were detected by microarray, consistent with the hypothesis that both biochemical processes act separately. The TCA 'cycle' in *P. falciparum* does not seem to operate cyclical and to regulate the degenerate oxidative degradation of organic compounds [58]. The parasite rather seems to favour the glycolytic energy metabolism during the asexual stages resulting in low mitochondrial electron transport chain complexity [58] and activity [59]. During schizogony the parasite restructures its cellular program from generic metabolic processes to highly parasitic functions synchronised with the expression of the proteasome. This special modus starts with the concomitant expression of most of the plastid-encoded genes which regulation is thought to be polycistronic [60]. At the end of the schizont stage, the parasite expresses genes with protein functions in merozoite invasion like parasite attachment, apical re-orientation and erythrocyte penetration. The group of merozoite invasion proteins contains well-known antigens that contribute to the naturally acquired immunity and hence are considered good vaccine candidates [61], like MSP1, AMA1, EBA175 and RAP1.

Llinas *et al.* raised a further question in their comprehensive transcriptome study of three different *Plasmodium* strains about whether transcriptional differences related to phenotypic properties such as drug sensitivity could deliver insights into the worldwide spreading of drug resistance [62]. The three *Plasmodium* strains analysed show the following differences: 3D7 was isolated in Europe (Netherlands) and shows sensitivity to chloroquine and resistance to sulfadoxine [63]. The HB3 strain was found in Latin America (Honduras) and is sensitive to chloroquine but resistant to pyrimethamine [63]. The Dd2 strain was isolated in Southeast Asia and is resistant against chloroquine and mefloquine [63]. Comparison of expression profiles suggested that gene cascades are highly conserved between *Plasmodium* strains and only little phase differences were observed. Cellular functions of glycolysis, transcription, translation, the TCA cycle, the

proteasome and plastid biosynthesis were nicely synchronised, whereas gene groups involved in cell adhesion, immunity and host-pathogen interactions showed significant differences. Although Llinas *et al.* anticipated changes in expression for genes known to be involved in drug resistance, they detected no such changes in the comprehensive transcriptome analysis of three *Plasmodium falciparum* strains 3D7, HB3 and Dd2 [62].

The availability of the complete genome sequence coupled to information about concerted mRNA profiles also allowed the prediction of *cis*-regulatory elements in *P. falciparum*. A new algorithm called Gene Enrichment Motif Searching (GEMS) overcame the *Plasmodium*-specific challenge of a high AT-content and identified upstream promoter motifs of co-expressed gene clusters [64]. The *in silico* search tool identified several putative *cis*-regulatory elements with high confidence composed of 5-8mers located within 1000bp upstream of the start codon. Additionally, *cis*-regulatory motifs were found downstream of the start codons within the intron sequences occurring in around 50% of genes of *P. falciparum*. A wide range of functional processes such as sexual development, antigenic variation, cell invasion, sporozoite development, ribosomal function and DNA replication seem to enclose *cis*-acting elements; however, functional confirmation of these motifs is required to make any firm conclusions about their role in gene expression.

Recently, our knowledge about the transcriptome in blood stage parasites was improved by a RNA-based high-throughput sequencing method resulting in the confirmation and correction of the gene models annotated so far across the IDC of *P. falciparum* [65]. RNA-sequencing was further able to upgrade the existing splice sites leading to the identification of novel alternative splicing forms. Four different types were found: exon-skipping, intron retention/creation, 3' and 5' alternative splicing and 3' and 5' splicing that create a new start or stop codon [65]. Interestingly, novel mRNA transcripts were detected, which correspond in sequence neither to annotated genes nor to corresponding untranslated regions (UTRs). These non-protein coding RNAs (npcRNAs) are 57 to 8931bp in length and seem to be conserved in *Plasmodium* [65].

Another group reported on npcRNAs and classified them into small nucleolar (snoRNAs) and small Cajal body-specific RNAs (scaRNAs) based on their sequence and structural motifs [66]. Several snoRNAs seem to be transcribed from telomeric and subtelomeric

repetitive regions and show similarities to yeast npcRNAs. Thus, it can be assumed that some snoRNAs of *P. falciparum* participate in the maintenance of telomeres and in subtelomeric silencing. Also, *cis*-encoded antisense npcRNAs were detected to be stably and complementary transcribed to various virulence-associated genes including *var* genes [66]. RNA-based regulatory elements could functionally interact with chromatin factors and could participate in so-called post-transcriptional gene silencing or RNA interference (RNAi) pathways. However, in *P. falciparum* genes encoding 'bona fide' components of the RNAi machinery have not been identified [67][68]. Nevertheless, this topic is controversially discussed as several studies reported down-regulation of gene transcription through interference with exogenously added RNA [69][70][71][72].

In conclusion, the transcriptional profile of *P. falciparum* describes a striking periodic cascade of gene expression during the intra-erythrocytic life cycle. Possible *cis*-regulatory elements were detected and suggested to contribute in the concerted expression of functionally related genes but remain underrepresented. As functional determination of *cis*-regulatory elements is based on the time-consuming and difficult development of a transfection system, only a few have been investigated to date in *P. falciparum* [73]. Furthermore, different types of non-coding RNAs were detected but their possible interference with gene expression mechanisms are still elusive.

1.3.3. Transcription factors in *P. falciparum*

Conserved components of the transcriptional machinery within the eukaryotic crown group are the basal transcription factors, whereas specific transcription factors possess an immense structural diversity [73]. However, several eukaryotic sequence-specific DNA-binding domains like the homeodomain, MADS, C2H2 zinc fingers, bZip, bHLH, GATA finger and FKH domains are well represented in animals, plants and fungi [74]. The process of transcription in *P. falciparum* is of special interest because lineage-specific evolution resulted in an apparent lack of specific transcription factors, of which only a small number have been found to date. As expected, the general transcription factors (GTFs) associated with the RNA-polymerase II subunits and co-activators of initiation and elongation were identified in *P. falciparum* [75][76][77]. However, a comparison of all transcription-associated proteins (TAPs) of a reference set of free-living eukaryotes with representatives found in *P. falciparum* revealed that the parasite *P. falciparum* possesses only one third of the usual numbers of such proteins [75]. A

recent publication by Bischoff and Vaquero nicely concentrated and depicted the available information concerning the transcriptional apparatus across the IDC of *P. falciparum* [78]. In summary, the family of specific transcription factors in *P. falciparum* include the plant-related ApiAP2 family, the zinc-coordinating DNA-binding domain protein family (C2H2, C3H4 and CCCH), the helix-turn-helix family with Myb domain-containing members and one protein of the 4SC beta scaffold factor family.

The first specific transcription factor investigated in detail was the PfMyb1 protein, which was shown to directly control the transcription of several differentially expressed genes involved in signal transduction and cell cycle regulation during the IDC [70]. Downregulation of *pfmyb1* transcription resulted in levels of Myb1 of only 20% of the wild-type level in the trophozoite stage, which caused growth inhibition and finally parasite death at the transition to the schizont stage [70].

The family of specific transcription factors with a lineage-specific version of the AP2 (Apetala2)-integrase DNA binding domain was quite recently discovered using sensitive sequence profile searches [79]. Balaji and colleagues screened the predicted proteins for known DNA binding motifs, resulting in the identification of a small DNA-binding motif, the AT hook, which is N-terminal coupled to a conserved globular domain of *Plasmodium*. Expanded PSI-BLAST searches for this large globular domain found significant hits in plant developmental transcription factors, bacterial endonucleases and integrases. The reciprocal search for DNA-binding domains based on the enlarged information led to the remarkable discovery of the ApiAP2 family in *P. falciparum*. The family belongs to the AP2-integrase DNA-binding domain superfamily and consists of 27 members in *P. falciparum*. The domain architecture of ApiAP2 proteins is composed of one to four globular AP2 domain units, and in only one member the AP2 domain occurs in association with another DNA-binding motif, the above mentioned AT-hook.

The ApiAP2 family can be divided into four groups due to stage-specific expression with four to six members each corresponding to the ring, trophozoite, early schizont and late schizont-merozoite stages [79]. ApiAP2 proteins are likely to recognise DNA in a sequence-specific manner due to analogy to plant AP2s and the differentiated expression profile across the IDC. Indeed, sequence-specific binding was proven for two members of the ApiAP2 family that differ in domain architecture (PF14_0633 with a

single AP2 domain and an additional AT-hook, PFF0200c with two AP2 domains in tandem) [80]. The unique characteristics of the DNA motifs were narrowed down by protein-binding microarrays (PBMs). The identified sequences are both palindromic, with the consensus sequence TGCATGCA bound by PF14_0633, and GTGCAC bound by PFF0200c, PBM experiments verified that the single AP2 binding domain of the AT-hook-containing protein is sufficient for specific binding as well as the first AP2 domain of the tandem repeated protein. Subsequently, the Finding Informative Regulatory Elements (FIRE) algorithm was used to compile target genes that had to be enriched with this particular motif and show a peak of transcription within the same phase of the asexual life cycle. FIRE identified the activation of a co-regulated set of genes which participate in host cell rupture and reinvasion. A different computational test, the ScanACE (Scans for Nucleic Acid Conserved Elements), was used to screen the 5' upstream regions of the *P. falciparum* genome in parallel to the related *Cryptosporidium parvum* genome in order to identify putative target proteins. The *in silico* results led to the hypothesis that PF14_0633 functions in *var* gene regulation affecting cytoadherence properties and that both proteins, PFF0200c and PF14_0633, influence each other by an ApiAP2 regulatory cascade. Furthermore, DNA-binding specificities were determined by PBMs for most of the 27 members of the ApiAP2 protein family including individual AP2 protein domains and tandem arrangements. Interestingly, in addition to a high affinity primary motif interactions the majority of these proteins also show binding to secondary motifs, which may increase the number of target genes that could be regulated by each factor. To predict potential ApiAP2 target genes, the newly identified DNA sequence motifs were linked with previously reported *cis*-regulatory elements of *Plasmodium* genes [81]. The results provide evidence that ApiAP2 proteins are involved in a comprehensive regulatory network of transcription which modulates key steps at all stages of *P. falciparum* development [82].

However, another publication about PFF0200c (or PfSIP2) suggested a different functional implication of this ApiAP2 protein in the recruitment and organisation of subtelomeric heterochromatin during schizogony [83]. Interaction characteristics were determined for PfSIP2 and its corresponding conserved DNA-binding motif (SPE2). In this study the SPE2/PfSIP2 interaction required both AP2 domains to efficiently bind the bipartite SPE2 consensus motif. Genome-wide *in silico* mapping identified several hundreds of putative binding sites located upstream of subtelomeric *var* genes and the

specific binding predictions were confirmed *in vivo* by genome-wide ChIP. Co-localisation of PfSIP2 and heterochromatin protein 1 (PfHP1) was detected, suggesting a potential role for this ApiAP2 factor in chromosome end organisation. Moreover, over-expression data of PfSIP2 contradicted the proposed function of PFF0200c in global gene activation as published by De Silva *et al.* [83].

Additionally, further publications confirmed on the functional level that ApiAP2 proteins can function as specific transcription factors by characterising two members in *P. berghei* ookinetes and sporozoites, respectively. The AP2-O transcription factor was found to regulate stage-specific genes in ookinetes by binding to promoters of genes functionally involved in mosquito midgut invasion [84]. The other ApiAP2 transcription factor AP2-Sp was characterised by the same group. AP2-Sp is expressed during sporozoite development in oocysts and salivary glands where gene disruption led to a loss of sporozoite formation. EMSA experiments showed that AP2-Sp specifically recognises a sequence of eight base pairs found in a subset of sporozoite-specific genes. Further experiments like domain swapping assumed that the AP2 domain of AP2-Sp acts as a transcription activator of genes in the sporozoite stage [85].

1.3.4. The epigenome of *P. falciparum*

The chromatin structure of eukaryotes could be described as spools of nucleosomes densely packaged with DNA, histone and non-histone proteins, which represents the underlying nature of epigenetic regulation. Alterations of this higher order structure of chromatin lead to differences in the transcriptional “on-off” states reflected by euchromatin (on) or heterochromatin (off) [86]. Post-translational modifications affect multiple residues of the N-terminal histone tails extending outward, which perform the actual dynamics of chromatin structure and function. Histone modifications can either directly stabilise/destabilise the physical interactions between DNA and histones or can recruit other proteins by recognition via specific domains [86]. Post-transcriptional modifications have shown to act in a concerted way by influencing each other in either a synergistic or an antagonistic manner in order to define the epigenetic environment of a chromosomal region. [86].

1.3.4.1. Histone modifications in *P. falciparum*

In silico data identified the four core histones H2A, H2B, H3 and H4 as well as the variant versions H3.3, CenH3, H2AZ and H2Bv in the genome of *P. falciparum* [87]. Interestingly, no linker histone H1 has been identified so far in *P. falciparum* which generally coordinates the higher-order chromatin formation in eukaryotes. Western blot analysis using antibodies specifically recognising particular histone modifications and mass spectrometry analysis compiled a broad range of epigenetic marks in *P. falciparum* [87][88]. Furthermore, several published 'ChIP-on-chip' studies expanded our knowledge of chromatin modifications as they mapped epigenetic marks on a genome-wide scale to corresponding gene expression profiles in *P. falciparum* [89][90]. All eight histones of *P. falciparum* are affected by methylation and/or acetylation events, which are summarised in the review of Horrocks *et al.* 'control of gene expression in *Plasmodium falciparum* - ten years on' [73]. A high correlation of histone modifications between *P. falciparum* and eukaryotes is observed including the temporal occurrence and associated gene expression [89][91][92][93].

The best studied epigenetic marks in *P. falciparum* are involved in *var* gene regulation where they show dynamic associations. Two marks of euchromatin, H3K4me_{2/3} and H3K9ac, are enriched at the 5' ends of the active *var* gene in ring stage parasites, whereas the remaining silent *var* genes are occupied by the heterochromatic mark H3K9me₃ [90][89][23][94]. Interestingly, the single *var* gene activated at the ring stage changes its epigenetic marks from H3K4me₃/H3K9ac to H3K4me₂ to ensure epigenetic memory during the trophozoite stage [95]. Another study characterised the highly dynamic H3 epigenome of *P. falciparum* as mostly euchromatic with well defined heterochromatic islands [90]. Besides active *var* genes, H3K4me₃ and H3K9ac are evenly distributed across the genome of ring stage parasites and are uniformly found at active and repressed genes, in contrast to the schizont stage where they are distinctly enriched at 5' ends of active genes [90].

1.3.4.2. Chromatin-interacting factors in *P. falciparum*

Like in other eukaryotes, histone acetylation of *P. falciparum* correlates with the transition to an active transcriptional state termed as euchromatin and acetyl-marks are found on K9, K14, K18, K27 of histone H3 and K5, K8, K12, K16 of histone 4 [87][88]. Proteins which mediate acetylation are classified as histone acetyltransferases (HATs) and the only described member in *P. falciparum* is PfGCN5, a homologue to the yeast transcriptional coactivator GCN5 [96]. PfGCN5 participates to the formation of a large multimeric complex and targets mainly K9 of histone 3, resulting in the transcriptional activation of a subset of genes [89]. Histone deacetylases (HDAC) perform the inverse function as HATs. They thus remove acetyl groups from the histone residues leading to a tighter chromatin density and a transcriptional repression.

The *P. falciparum* genome encodes two type III HDACs (sirtuins) termed PfSIR2A and PfSIR2B, which in other eukaryotes were shown to associate into multimeric protein complexes spreading a gradient of heterochromatin along the chromosome [97][98][99]. *In vitro* deacetylation activity for PfSIR2A was shown at residues H3K9, H3K14 and H4K12 [99] and deletion studies of both paralogs led to the loss of silencing of a subclass of *var* genes [98][100]. The type II class HDACs are known to be good drug targets of cancer research and possess additionally anti-parasitic activity. It was shown that HDACs inhibitors prevent parasite growth *in vitro* and *in vivo* leading to hyperacetylated histones [101][102][103]. The crystal structure of several plasmodial HDACs was predicted to identify the docking specificities which could lead to efficacy optimisation of the compound class [104][105].

The event of post-translational methylation is more variable than acetylation due to the association with actively transcribed as well as silenced states. Methylated residues are found on K4, K9, K14, R17, K36 of histone 3 and K5, K16, R17, K20 of histone 4 in *P. falciparum* [87][88]. The type of methylated amino acid reflects the nomenclature of the modifying enzyme class and distinguishes between lysine methyltransferases (HKMT) and arginine methyltransferases (PRMT). The PfHKMT family contains a characteristic SET methyltransferase subgroup of nine members, which display dynamic changes in expression during the intra-erythrocytic life cycle of *P. falciparum* [106]. Preferential methylation sites are proposed for several SET-domain HKMTs but actual functions remain to be demonstrated. Notably, the tenth SET-domain containing protein (PfSET10;

PFL11010c) was identified quite recently and linked to antigenic variation due to perinuclear localisation with the active *var* gene expression site [107]. This H3K4 methyltransferase seems to be responsible for the preservation of the 'poised' gene state during parasitic division and growth across the IDC of *P. falciparum* [107].

The recruitment process of regulatory complexes to chromatin is based on the communication between the histone modifications and specific interaction domains that are often chromo- or bromodomains [108]. A famous member of the chromodomain containing class is heterochromatin 1 protein (HP1), which has specific functions in *var* gene silencing [109][23]. Flueck and colleagues showed that PfHP1 specifically binds to H3K9me3 but not to other histone marks using pull down assays [23]. ChIP-on-chip determined its genome-wide occupancy uncovering a striking association between PfHP1 and virulence genes clusters in subtelomeric and chromosome-internal regions, which was highly correlated with the mapped H3K9me3 modifications. Finally, PfHP1 was observed to be representative for heterochromatin of chromosome end clusters at the nuclear periphery of *P. falciparum* by IFA. In addition, the genome of *P. falciparum* encodes seven proteins containing a putative bromo-domain [78]. The bromo-domain is the only domain known to interact with acetylated lysine residues of histones and mediates gene activation. Unfortunately, little is known about their important function in transcriptional regulation in *P. falciparum*.

Besides the array of post-translational modifications, changes in gene expression can be caused by direct activity of chromatin remodelling complexes on nucleosomes, producing dramatic changes in chromatin structure [86]. These ATP-driven processes can relocate particular nucleosomes, or can exchange core histones with their variant forms. The complex pool of remodellers includes beside the ATPase family also nucleosomal assembly proteins and high mobility group B (HMGB) proteins which are found in *P. falciparum* [110][111][112].

For the first time, Volz *et al.* visualised the nuclear architecture of *P. falciparum* by dividing the compartment into distinct functional areas [21]. The localisation and the functional prediction of a broad range of nuclear proteins were verified including putative transcription factors, histone modifying proteins, chromatin structuring proteins and chromatin remodeling proteins (containing SET-, PHD-, chromo- or bromo-domains).

The nucleus was dissected from the center towards the nuclear periphery in three distinct areas following PfNUP100 (nucleoporin 100), PfCENP-A (centromeric H3) and DAPI stained areas. PfNUP100 defines the nuclear envelope. PfCENP-A, which is required for nucleation and maintenance of centromeres, marks the ‘transition’ zone of the nuclear periphery. The center of the nucleus is stained by DAPI. In general, the most transcriptionally active region was assigned to the core center of the nucleus, in contrast to the nuclear periphery of *P. falciparum* where gene silencing is mainly occurring [20][21][19]. In another study, the localisation of PfCENP-A was observed as a single nuclear focus prior to and during mitosis, clearly separated from chromosome end clusters [22]. Chromatin-IP coupled to deep sequencing revealed that PfCENP-A covers a 4-4.5kb region on each chromosome of *P. falciparum* and thus defines the exact position of the centromeres [113].

To complement information about epigenetic gene regulation, it is worth mentioning that DNA methylation does not play a role in epigenetic mechanisms of *P. falciparum*. Indeed, the modification of 5-methyl-2'-deoxycytosine commonly characterising heterochromatin in eukaryotes is absent in *P. falciparum* [114][115]. In summary, epigenetic processes are proven to play an important role in the regulation of gene expression in *P. falciparum*, which possesses a complete set of histone modifying and chromatin remodelling proteins. Unfortunately, most of the proteins responsible for long term structuring of chromatin by recognising different histone modifications via specific motifs are somewhat underrepresented and poorly characterised to date in *P. falciparum*.

1.4. The proteome of *P. falciparum*

Mass spectrometry-based proteomics approaches permit to improve the gene information of an organism in correlation to temporal and spatial appearance and can be successfully complemented by approaches based on bioinformatics and microarray platforms. The sensitivity of peptide separation and detection systems allows the detection of individual proteins in complex protein mixtures and provides high accuracy tandem mass spectra (MS/MS). Large scale proteomics research in *P. falciparum* was triggered by two global proteome studies covering the complex life cycle (ring, trophozoite, schizont, gametocyte, sporozoite, and gamete stages) [116][117]. Lasonder

et al. identified by LC-MS/MS a total of 1289 proteins of which 714 proteins were detected in asexual blood stages, 931 in gametocytes and 645 in gametes [117]. The second study published by Florens *et al.* characterised four stages of the parasite life cycle (sporozoites, merozoites, trophozoites and gametocytes) by using MudPIT coupled to MS/MS analysis [116]. This group identified 2415 proteins in all four stages and remarked striking differences in the protein profiles between sporozoites and blood stages [116]. These studies were followed by scores of proteomic studies focusing on subcellular compartments or particular developmental stages of *P. falciparum* and *P. berghei*. Novel proteins located on the surface of infected red blood cells and members of the Maurer's clefts were identified to better understand the biological role of the exported protein network [118][119]. Food vacuole proteomics led to more insights about FV biogenesis and protein trafficking with possible involvement in haemoglobin ingestion [120]. The protein content of merozoite rhoptries from three different *Plasmodium* rodent species was measured and bioinformatically analysed for the presence of transmembrane domains, secretory signals, cell adhesion and rhoptry-specific motifs [121]. Proteomics of the invasive parasite stage, the ookinete, described the microneme organisation in *P. berghei* [122]. Novel proteins were identified to be involved in vesicle trafficking and signalling, in line with the hypothesis that micronemes participate in a secretory pathway important for invasion.

Besides these mostly organellar proteomics approaches, several studies concentrated on particular developmental stages of *P. falciparum*. Comparison of proteome data of midgut and matured salivary gland sporozoites led to interesting differences responsible for parasite development and infectivity [123]. Proteome analysis of separated male and female gametocytes accentuated the diverged sex-specific features for transition between the human host and mosquito vector [124]. Another study showed that the preliminary sexual differentiation of gametocytes in *P. falciparum* is marked by intense protein export probably contributing to host cell remodelling [125].

The integration of transcriptomic and proteomic data enables to differentiate the mechanism of gene regulation at the post-transcriptional level by comparing mRNA and protein levels. Although recent transcriptome studies revealed that gene expression according to morphological and metabolic changes during the IDC is managed 'just in time', there is evidences for the involvement of post-transcriptional regulatory

mechanisms. Le Roch and collaborators performed abundance correlations for thousands of mRNA transcripts and corresponding protein levels across seven developmental stages of *P. falciparum* [126][127]. Non-shifted abundance peaks of mRNAs and proteins were observed for a representative large part of each stage, in addition to a significant part showing a delay in protein production. This 'shifting' phenomenon could be shown for several stages of the *P. falciparum* life cycle suggesting possible stabilising effects at the mRNA level controlling protein translation. The authors also asked whether 5' and/or 3' UTRs are involved in this mechanism of gene regulation as additional regulatory elements. They identified a few consensus motifs in the 5' UTRs correlated to the observed time shifts, suggesting the involvement of mRNA binding proteins. Recently, a novel RNA-binding family was identified with two members in *P. falciparum*, PfPUF1 and PfPUF2, which regulate translation and RNA stability by binding to specific sequences in the 3'-untranslated region of target mRNAs [128][129]. PUF members can be found in other *Plasmodium* species suggesting an evolutionary conserved function during parasite development [128]. Furthermore, it was shown in *P. berghei* that DOZI (development of zygote inhibited), a multiprotein complex of mRNAs and the DDX6 class RNA helicase, translationally represses various genes until gamete fertilisation [130][131].

Unfortunately, the most relevant proteome studies of *P. falciparum* employed non-quantitative or semi-quantitative mass spectrometry techniques and did not address post-translational modifications of proteins. Although individual examples of post-transcriptional modified proteins are known, widespread detection is needed to get an extended overview of post-translational modifications (PTMs) and their role in timing protein activities. The study of Foth *et al.* presented a quantitative proteomics approach, which revealed extensive PTMs in the schizont stage of *P. falciparum* [132]. The two-dimensional differential gel electrophoresis separation technique facilitated the identification of 54 protein spots, 58% of which have different isoforms with distinct expression patterns. The enolase protein had the most striking number of seven different isoforms present during the schizont stage. These results are consistent with the identification of five different post-transcriptionally modified isoforms of *P. yoelii* enolase, which immuno-localised in the cytosol and was found to be associated with membranes, nuclei and cytoskeletal fractions [133].

A recent phosphoproteome study found unusual phosphorylation events probably triggered by the AT-richness of the genome, revealing a lineage-specific evolution within and beyond parasites boundaries [134]. Treeck and colleagues identified 7835 phosphopeptides with 8463 phosphorylation sites matching to 1673 proteins were identified in the schizont stage of *P. falciparum* [134]. Interestingly, they also detected phosphorylated proteins occur outside the boundaries of the parasite suggesting that they are subject to regulation after release. In total, Surprisingly, phosphorylated tyrosines were identified despite the apparent lack of genes encoding tyrosine-specific kinases or phosphatases in the genome of *P. falciparum* [135][136][137]. Solyakov *et al.* performed global kinomic and phospho-proteomic analysis and identified 37 kinases as likely essential for asexual schizogony by using reverse genetics strategies [138][139].

1.5. Specific objectives of the thesis

Although basic biological processes within the nucleus of *P. falciparum* are key to parasite survival, life cycle progression and immune evasion, they remain surprisingly unsought to date. Thus, the overall objective of this thesis was to compile an inventory of the *P. falciparum* nuclear proteome across the IDC by complementary methods to identify and characterise novel nuclear proteins.

Specific aims:

- Global analysis of the nuclear proteome by high accuracy mass spectrometry coupled with bioinformatics and *in vivo* localisation experiments to study the biology of the *P. falciparum* nucleus in detail. Submission of a nuclear core proteome of *P. falciparum* to the public malaria database PlasmoDB (*see Chapter 2*).
- Detailed analysis of selected nuclear proteins on the functional level using various approaches including reverse genetics, co-immunoprecipitation, genome-wide transcriptional profiling and chromatin immuno-precipitation (*see Chapter 3*).
- Description of the nuclear compartment by visualization of landmark structures based on fluorescence microscopy to gain more insights into interrelations and functional implications in parasite nuclear biology (*see Chapter 4*).

1.6 References

1. Tuteja R (2007) Malaria - an overview. *FEBS J* 274: 4670–4679. doi:10.1111/j.1742-4658.2007.05997.x.
2. Putaporntip C, Hongrimumuang T, Seethamchai S, Kobasa T, Limkittikul K, et al. (2009) Differential prevalence of Plasmodium infections and cryptic Plasmodium knowlesi malaria in humans in Thailand. *J Infect Dis* 199: 1143–1150. doi:10.1086/597414.
3. Mazier D, Beaudoin RL, Mellouk S, Druilhe P, Texier B, et al. (1985) Complete development of hepatic stages of Plasmodium falciparum in vitro. *Science* 227: 440–442.
4. Wirth DF (2002) The parasite genome: Biological revelations. *Nature* 419: 495–496. doi:10.1038/419495a.
5. WHO | World Malaria Report 2011 (n.d.). WHO. Available: http://www.who.int/malaria/world_malaria_report_2011/en/. Accessed 29 May 2012.
6. Olliaro P (2001) Mode of action and mechanisms of resistance for antimalarial drugs. *Pharmacol Ther* 89: 207–219.
7. Srivastava IK, Rottenberg H, Vaidya AB (1997) Atovaquone, a broad spectrum antiparasitic drug, collapses mitochondrial membrane potential in a malarial parasite. *J Biol Chem* 272: 3961–3966.
8. Bray PG, Janneh O, Ward SA (1999) Chloroquine uptake and activity is determined by binding to ferriprotoporphyrin IX in Plasmodium falciparum. *Novartis Found Symp* 226: 252–260; discussion 260–264.
9. Pierce SK, Miller LH (2009) World Malaria Day 2009: what malaria knows about the immune system that immunologists still do not. *J Immunol* 182: 5171–5177. doi:10.4049/jimmunol.0804153.
10. Gordon DM, McGovern TW, Krzych U, Cohen JC, Schneider I, et al. (1995) Safety, immunogenicity, and efficacy of a recombinantly produced Plasmodium falciparum circumsporozoite protein-hepatitis B surface antigen subunit vaccine. *J Infect Dis* 171: 1576–1585.
11. Casares S, Brumeau T-D, Richie TL (2010) The RTS,S malaria vaccine. *Vaccine* 28: 4880–4894. doi:10.1016/j.vaccine.2010.05.033.
12. Olotu A, Lusingu J, Leach A, Lievens M, Vekemans J, et al. (2011) Efficacy of RTS,S/AS01E malaria vaccine and exploratory analysis on anti-circumsporozoite antibody titres and protection in children aged 5–17 months in Kenya and Tanzania: a randomised controlled trial. *Lancet Infect Dis* 11: 102–109. doi:10.1016/S1473-3099(10)70262-0.
13. Cech PG, Aebi T, Abdallah MS, Mpina M, Machunda EB, et al. (2011) Virosome-formulated Plasmodium falciparum AMA-1 & CSP derived peptides as malaria vaccine: randomized phase 1b trial in semi-immune adults & children. *PLoS ONE* 6: e22273. doi:10.1371/journal.pone.0022273.
14. Cowman AF, Crabb BS (2006) Invasion of red blood cells by malaria parasites. *Cell* 124: 755–766. doi:10.1016/j.cell.2006.02.006.
15. Plotkin SA (2005) Vaccines: past, present and future. *Nat Med* 11: S5–11. doi:10.1038/nm1209.
16. Chattopadhyay R, Conteh S, Li M, James ER, Epstein JE, et al. (2009) The Effects of radiation on the safety and protective efficacy of an attenuated Plasmodium yoelii sporozoite malaria vaccine. *Vaccine* 27: 3675–3680. doi:10.1016/j.vaccine.2008.11.073.
17. Aikawa M (1966) The fine structure of the erythrocytic stages of three avian malarial parasites, Plasmodium fallax, P. lophurae, and P. cathemerium. *Am J Trop Med Hyg* 15: 449–471.
18. Aikawa M, Beaudoin RL (1968) Studies on nuclear division of a malarial parasite under pyrimethamine treatment. *J Cell Biol* 39: 749–754.
19. Weiner A, Dahan-Pasternak N, Shimoni E, Shinder V, von Huth P, et al. (2011) 3D nuclear architecture reveals coupled cell cycle dynamics of chromatin and nuclear pores in the malaria parasite Plasmodium falciparum. *Cell Microbiol* 13: 967–977. doi:10.1111/j.1462-5822.2011.01592.x.
20. Ralph SA, Scheidig-Benatar C, Scherf A (2005) Antigenic variation in Plasmodium falciparum is associated with movement of var loci between subnuclear locations. *Proc Natl Acad Sci USA* 102: 5414–5419. doi:10.1073/pnas.0408883102.
21. Volz J, Carvalho TG, Ralph SA, Gilson P, Thompson J, et al. (2010) Potential epigenetic regulatory proteins localise to distinct nuclear sub-compartments in Plasmodium falciparum. *Int J Parasitol* 40: 109–121. doi:10.1016/j.ijpara.2009.09.002.
22. Hoeijmakers WAM, Flueck C, François K-J, Smits AH, Wetzel J, et al. (2012) Plasmodium falciparum centromeres display a unique epigenetic makeup and cluster prior to and during schizogony. *Cellular microbiology*. Available: <http://www.ncbi.nlm.nih.gov/pubmed/22507744>. Accessed 23 May 2012.
23. Flueck C, Bartfai R, Volz J, Niederwieser I, Salcedo-Amaya AM, et al. (2009) Plasmodium falciparum heterochromatin protein 1 marks genomic loci linked to phenotypic variation of exported virulence factors. *PLoS Pathog* 5: e1000569. doi:10.1371/journal.ppat.1000569.
24. Marty AJ, Thompson JK, Duffy MF, Voss TS, Cowman AF, et al. (2006) Evidence that Plasmodium falciparum chromosome end clusters are cross-linked by protein and are the sites of both virulence gene silencing and activation. *Mol Microbiol* 62: 72–83. doi:10.1111/j.1365-2958.2006.05364.x.
25. Hernandez-Rivas R, Mattei D, Sterkers Y, Peterson DS, Wellems TE, et al. (1997) Expressed var genes are found in Plasmodium falciparum subtelomeric regions. *Mol Cell Biol* 17: 604–611.

26. Bannister LH, Hopkins JM, Fowler RE, Krishna S, Mitchell GH (2000) A brief illustrated guide to the ultrastructure of *Plasmodium falciparum* asexual blood stages. *Parasitol Today (Regul Ed)* 16: 427–433.
27. Margos G, Bannister LH, Dluzewski AR, Hopkins J, Williams IT, et al. (2004) Correlation of structural development and differential expression of invasion-related molecules in schizonts of *Plasmodium falciparum*. *Parasitology* 129: 273–287.
28. Striepen B, Jordan CN, Reiff S, van Dooren GG (2007) Building the perfect parasite: cell division in apicomplexa. *PLoS Pathog* 3: e78. doi:10.1371/journal.ppat.0030078.
29. Gerald N, Mahajan B, Kumar S (2011) Mitosis in the Human Malaria Parasite *Plasmodium Falciparum*. *Eukaryotic Cell* 10: 474–482. doi:10.1128/EC.00314-10.
30. Azimzadeh J, Bornens M (2007) Structure and duplication of the centrosome. *J Cell Sci* 120: 2139–2142. doi:10.1242/jcs.005231.
31. Read M, Sherwin T, Holloway SP, Gull K, Hyde JE (1993) Microtubular organization visualized by immunofluorescence microscopy during erythrocytic schizogony in *Plasmodium falciparum* and investigation of post-translational modifications of parasite tubulin. *Parasitology* 106 (Pt 3): 223–232.
32. Arnot DE, Gull K (1998) The *Plasmodium* cell-cycle: facts and questions. *Ann Trop Med Parasitol* 92: 361–365.
33. Naughton JA, Bell A (2007) Studies on cell-cycle synchronization in the asexual erythrocytic stages of *Plasmodium falciparum*. *Parasitology* 134: 331–337. doi:10.1017/S0031182006001466.
34. Gardner MJ, Hall N, Fung E, White O, Berriman M, et al. (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* 419: 498–511. doi:10.1038/nature01097.
35. Vaidya AB, Lashgari MS, Pologe LG, Morrissey J (1993) Structural features of *Plasmodium* cytochrome b that may underlie susceptibility to 8-aminoquinolines and hydroxynaphthoquinones. *Mol Biochem Parasitol* 58: 33–42.
36. Wilson RJ, Denny PW, Preiser PR, Rangachari K, Roberts K, et al. (1996) Complete gene map of the plastid-like DNA of the malaria parasite *Plasmodium falciparum*. *J Mol Biol* 261: 155–172.
37. Köhler S, Delwiche CF, Denny PW, Tilney LG, Webster P, et al. (1997) A plastid of probable green algal origin in Apicomplexan parasites. *Science* 275: 1485–1489.
38. McFadden GI, Reith ME, Munholland J, Lang-Unnasch N (1996) Plastid in human parasites. *Nature* 381: 482. doi:10.1038/381482a0.
39. Palmer JD, Delwiche CF (1996) Second-hand chloroplasts and the case of the disappearing nucleus. *Proc Natl Acad Sci USA* 93: 7432–7435.
40. Wilson RJMI (2002) Progress with parasite plastids. *J Mol Biol* 319: 257–274. doi:10.1016/S0022-2836(02)00303-0.
41. Fast NM, Kissinger JC, Roos DS, Keeling PJ (2001) Nuclear-encoded, plastid-targeted genes suggest a single common origin for apicomplexan and dinoflagellate plastids. *Mol Biol Evol* 18: 418–426.
42. Roos DS, Crawford MJ, Donald RG, Kissinger JC, Klimczak LJ, et al. (1999) Origin, targeting, and function of the apicomplexan plastid. *Curr Opin Microbiol* 2: 426–432. doi:10.1016/S1369-5274(99)80075-7.
43. Surolia N, Surolia A (2001) Triclosan offers protection against blood stages of malaria by inhibiting enoyl-ACP reductase of *Plasmodium falciparum*. *Nat Med* 7: 167–173. doi:10.1038/84612.
44. Jomaa H, Wiesner J, Sanderbrand S, Altincicek B, Weidemeyer C, et al. (1999) Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs. *Science* 285: 1573–1576.
45. Sato S, Wilson RJM (2002) The genome of *Plasmodium falciparum* encodes an active delta-aminolevulinic acid dehydratase. *Curr Genet* 40: 391–398. doi:10.1007/s00294-002-0273-3.
46. van Dooren GG, Su V, D’Ombrain MC, McFadden GI (2002) Processing of an apicoplast leader sequence in *Plasmodium falciparum* and the identification of a putative leader cleavage enzyme. *J Biol Chem* 277: 23612–23619. doi:10.1074/jbc.M201748200.
47. Vaidya AB, Akella R, Suplick K (1989) Sequences similar to genes for two mitochondrial proteins and portions of ribosomal RNA in tandemly arrayed 6-kilobase-pair DNA of a malarial parasite. *Mol Biochem Parasitol* 35: 97–107.
48. Hinterberg K, Mattei D, Wellems TE, Scherf A (1994) Interchromosomal exchange of a large subtelomeric segment in a *Plasmodium falciparum* cross. *EMBO J* 13: 4174–4180.
49. Kelly JM, McRobert L, Baker DA (2006) Evidence on the chromosomal location of centromeric DNA in *Plasmodium falciparum* from etoposide-mediated topoisomerase-II cleavage. *Proc Natl Acad Sci USA* 103: 6706–6711. doi:10.1073/pnas.0510363103.
50. Iwanaga S, Khan SM, Kaneko I, Christodoulou Z, Newbold C, et al. (2010) Functional identification of the *Plasmodium* centromere and generation of a *Plasmodium* artificial chromosome. *Cell Host Microbe* 7: 245–255. doi:10.1016/j.chom.2010.02.010.
51. Carmo-Fonseca M, Mendes-Soares L, Campos I (2000) To be or not to be in the nucleolus. *Nat Cell Biol* 2: E107–112. doi:10.1038/35014078.
52. McStay B, Grummt I (2008) The epigenetics of rRNA genes: from molecular to chromosome biology. *Annu Rev Cell Dev Biol* 24: 131–157. doi:10.1146/annurev.cellbio.24.110707.175259.
53. Waters AP (1994) The ribosomal RNA genes of *Plasmodium*. *Adv Parasitol* 34: 33–79.

54. Fang J, McCutchan TF (2002) Thermoregulation in a parasite's life cycle. *Nature* 418: 742. doi:10.1038/418742a.
55. Fang J, Sullivan M, McCutchan TF (2004) The effects of glucose concentration on the reciprocal regulation of rRNA promoters in *Plasmodium falciparum*. *J Biol Chem* 279: 720–725. doi:10.1074/jbc.M308284200.
56. Le Roch KG, Zhou Y, Blair PL, Grainger M, Moch JK, et al. (2003) Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science* 301: 1503–1508. doi:10.1126/science.1087025.
57. Bozdech Z, Llinás M, Pulliam BL, Wong ED, Zhu J, et al. (2003) The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS Biol* 1: E5. doi:10.1371/journal.pbio.0000005.
58. Mather MW, Henry KW, Vaidya AB (2007) Mitochondrial drug targets in apicomplexan parasites. *Curr Drug Targets* 8: 49–60.
59. Vander Jagt DL, Hunsaker LA, Campos NM, Baack BR (1990) D-lactate production in erythrocytes infected with *Plasmodium falciparum*. *Mol Biochem Parasitol* 42: 277–284.
60. Wilson RJ, Denny PW, Preiser PR, Rangachari K, Roberts K, et al. (1996) Complete gene map of the plastid-like DNA of the malaria parasite *Plasmodium falciparum*. *J Mol Biol* 261: 155–172.
61. Cowman AF, Baldi DL, Healer J, Mills KE, O'Donnell RA, et al. (2000) Functional analysis of proteins involved in *Plasmodium falciparum* merozoite invasion of red blood cells. *FEBS Lett* 476: 84–88.
62. Llinás M, Bozdech Z, Wong ED, Adai AT, DeRisi JL (2006) Comparative whole genome transcriptome analysis of three *Plasmodium falciparum* strains. *Nucleic Acids Res* 34: 1166–1173. doi:10.1093/nar/gkj517.
63. Rathod PK, McErlean T, Lee P-C (1997) Variations in frequencies of drug resistance in *Plasmodium falciparum*. *Proc Natl Acad Sci U S A* 94: 9389–9393.
64. Young JA, Johnson JR, Benner C, Yan SF, Chen K, et al. (2008) In silico discovery of transcription regulatory elements in *Plasmodium falciparum*. *BMC Genomics* 9: 70. doi:10.1186/1471-2164-9-70.
65. Otto TD, Wilinski D, Assefa S, Keane TM, Sarry LR, et al. (2010) New insights into the blood-stage transcriptome of *Plasmodium falciparum* using RNA-Seq. *Mol Microbiol* 76: 12–24. doi:10.1111/j.1365-2958.2009.07026.x.
66. Raabe CA, Sanchez CP, Randau G, Robeck T, Skryabin BV, et al. (2010) A global view of the nonprotein-coding transcriptome in *Plasmodium falciparum*. *Nucleic Acids Res* 38: 608–617. doi:10.1093/nar/gkp895.
67. Ullu E, Tschudi C, Chakraborty T (2004) RNA interference in protozoan parasites. *Cellular Microbiology* 6: 509–519. doi:10.1111/j.1462-5822.2004.00399.x.
68. Baum J, Papenfuss AT, Mair GR, Janse CJ, Vlachou D, et al. (2009) Molecular genetics and comparative genomics reveal RNAi is not functional in malaria parasites. *Nucleic Acids Res* 37: 3788–3798. doi:10.1093/nar/gkp239.
69. McRobert L, McConkey GA (2002) RNA interference (RNAi) inhibits growth of *Plasmodium falciparum*. *Mol Biochem Parasitol* 119: 273–278.
70. Gissot M, Briquet S, Refour P, Boschet C, Vaquero C (2005) PfMyb1, a *Plasmodium falciparum* transcription factor, is required for intra-erythrocytic growth and controls key genes for cell cycle regulation. *J Mol Biol* 346: 29–42. doi:10.1016/j.jmb.2004.11.045.
71. Kumar R, Adams B, Oldenburg A, Musiyenko A, Barik S (2002) Characterisation and expression of a PP1 serine/threonine protein phosphatase (PfPP1) from the malaria parasite, *Plasmodium falciparum*: demonstration of its essential role using RNA interference. *Malar J* 1: 5.
72. Mohammed A, Dasaradhi PVN, Bhatnagar RK, Chauhan VS, Malhotra P (2003) In vivo gene silencing in *Plasmodium berghei*--a mouse malaria model. *Biochem Biophys Res Commun* 309: 506–511.
73. Horrocks P, Wong E, Russell K, Emes RD (2009) Control of gene expression in *Plasmodium falciparum* - ten years on. *Mol Biochem Parasitol* 164: 9–25. doi:10.1016/j.molbiopara.2008.11.010.
74. Lespinet O, Wolf YI, Koonin EV, Aravind L (2002) The role of lineage-specific gene family expansion in the evolution of eukaryotes. *Genome Res* 12: 1048–1059. doi:10.1101/gr.174302.
75. Coulson RMR, Hall N, Ouzounis CA (2004) Comparative genomics of transcriptional control in the human malaria parasite *Plasmodium falciparum*. *Genome Res* 14: 1548–1554. doi:10.1101/gr.2218604.
76. Coulson RMR, Ouzounis CA (2003) The phylogenetic diversity of eukaryotic transcription. *Nucleic Acids Res* 31: 653–660.
77. Aravind L, Iyer LM, Wellems TE, Miller LH (2003) *Plasmodium* biology: genomic gleanings. *Cell* 115: 771–785.
78. Bischoff E, Vaquero C (2010) In silico and biological survey of transcription-associated proteins implicated in the transcriptional machinery during the erythrocytic development of *Plasmodium falciparum*. *BMC Genomics* 11: 34. doi:10.1186/1471-2164-11-34.
79. Balaji S, Babu MM, Iyer LM, Aravind L (2005) Discovery of the principal specific transcription factors of Apicomplexa and their implication for the evolution of the AP2-integrase DNA binding domains. *Nucleic Acids Res* 33: 3994–4006. doi:10.1093/nar/gki709.
80. De Silva EK, Gehrke AR, Olszewski K, León I, Chahal JS, et al. (2008) Specific DNA-binding by apicomplexan AP2 transcription factors. *Proc Natl Acad Sci USA* 105: 8393–8398. doi:10.1073/pnas.0801993105.

81. Militello KT, Dodge M, Bethke L, Wirth DF (2004) Identification of regulatory elements in the *Plasmodium falciparum* genome. *Mol Biochem Parasitol* 134: 75–88.
82. Campbell TL, De Silva EK, Olszewski KL, Elemento O, Llinás M (2010) Identification and Genome-Wide Prediction of DNA Binding Specificities for the ApiAP2 Family of Regulators from the Malaria Parasite. *PLoS Pathog* 6. doi:10.1371/journal.ppat.1001165.
83. Flueck C, Bartfai R, Niederwieser I, Witmer K, Alako BT, et al. (2010) A major role for the *Plasmodium falciparum* ApiAP2 protein PfSIP2 in chromosome end biology. *PLoS Pathog* 6: e1000784. doi:10.1371/journal.ppat.1000784.
84. Yuda M, Iwanaga S, Shigenobu S, Mair GR, Janse CJ, et al. (2009) Identification of a transcription factor in the mosquito-invasive stage of malaria parasites. *Mol Microbiol* 71: 1402–1414. doi:10.1111/j.1365-2958.2009.06609.x.
85. Yuda M, Iwanaga S, Shigenobu S, Kato T, Kaneko I (2010) Transcription factor AP2-Sp and its target genes in malarial sporozoites. *Mol Microbiol* 75: 854–863. doi:10.1111/j.1365-2958.2009.07005.x.
86. Jenuwein T, Allis CD (2001) Translating the Histone Code. *Science* 293: 1074–1080. doi:10.1126/science.1063127.
87. Miao J, Fan Q, Cui L, Li J, Li J, et al. (2006) The malaria parasite *Plasmodium falciparum* histones: organization, expression, and acetylation. *Gene* 369: 53–65. doi:10.1016/j.gene.2005.10.022.
88. Trelle MB, Salcedo-Amaya AM, Cohen AM, Stunnenberg HG, Jensen ON (2009) Global histone analysis by mass spectrometry reveals a high content of acetylated lysine residues in the malaria parasite *Plasmodium falciparum*. *J Proteome Res* 8: 3439–3450. doi:10.1021/pr9000898.
89. Cui L, Miao J, Furuya T, Li X, Su X, et al. (2007) PfGCN5-mediated histone H3 acetylation plays a key role in gene expression in *Plasmodium falciparum*. *Eukaryotic Cell* 6: 1219–1227. doi:10.1128/EC.00062-07.
90. Salcedo-Amaya AM, van Driel MA, Alako BT, Trelle MB, van den Elzen AMG, et al. (2009) Dynamic histone H3 epigenome marking during the intraerythrocytic cycle of *Plasmodium falciparum*. *Proc Natl Acad Sci USA* 106: 9655–9660. doi:10.1073/pnas.0902515106.
91. Gissot M, Kelly KA, Ajioka JW, Grealley JM, Kim K (2007) Epigenomic modifications predict active promoters and gene structure in *Toxoplasma gondii*. *PLoS Pathog* 3: e77. doi:10.1371/journal.ppat.0030077.
92. Cui L, Miao J, Cui L (2007) Cytotoxic effect of curcumin on malaria parasite *Plasmodium falciparum*: inhibition of histone acetylation and generation of reactive oxygen species. *Antimicrob Agents Chemother* 51: 488–494. doi:10.1128/AAC.01238-06.
93. Cui L, Miao J, Furuya T, Fan Q, Li X, et al. (2008) Histone acetyltransferase inhibitor anacardic acid causes changes in global gene expression during in vitro *Plasmodium falciparum* development. *Eukaryotic Cell* 7: 1200–1210. doi:10.1128/EC.00063-08.
94. Lopez-Rubio JJ, Gontijo AM, Nunes MC, Issar N, Hernandez Rivas R, et al. (2007) 5' flanking region of var genes nucleate histone modification patterns linked to phenotypic inheritance of virulence traits in malaria parasites. *Mol Microbiol* 66: 1296–1305. doi:10.1111/j.1365-2958.2007.06009.x.
95. Dzikowski R, Deitsch KW (2008) Active transcription is required for maintenance of epigenetic memory in the malaria parasite *Plasmodium falciparum*. *J Mol Biol* 382: 288–297. doi:10.1016/j.jmb.2008.07.015.
96. Fan Q, An L, Cui L (2004) *Plasmodium falciparum* histone acetyltransferase, a yeast GCN5 homologue involved in chromatin remodeling. *Eukaryotic Cell* 3: 264–276.
97. Duraisingh MT, Voss TS, Marty AJ, Duffy MF, Good RT, et al. (2005) Heterochromatin silencing and locus repositioning linked to regulation of virulence genes in *Plasmodium falciparum*. *Cell* 121: 13–24. doi:10.1016/j.cell.2005.01.036.
98. Freitas-Junior LH, Hernandez-Rivas R, Ralph SA, Montiel-Condado D, Ruvalcaba-Salazar OK, et al. (2005) Telomeric heterochromatin propagation and histone acetylation control mutually exclusive expression of antigenic variation genes in malaria parasites. *Cell* 121: 25–36. doi:10.1016/j.cell.2005.01.037.
99. Merrick CJ, Duraisingh MT (2007) *Plasmodium falciparum* Sir2: an unusual sirtuin with dual histone deacetylase and ADP-ribosyltransferase activity. *Eukaryotic Cell* 6: 2081–2091. doi:10.1128/EC.00114-07.
100. Duraisingh MT, Voss TS, Marty AJ, Duffy MF, Good RT, et al. (2005) Heterochromatin silencing and locus repositioning linked to regulation of virulence genes in *Plasmodium falciparum*. *Cell* 121: 13–24. doi:10.1016/j.cell.2005.01.036.
101. Andrews KT, Walduck A, Kelso MJ, Fairlie DP, Saul A, et al. (2000) Anti-malarial effect of histone deacetylation inhibitors and mammalian tumour cytodifferentiating agents. *Int J Parasitol* 30: 761–768.
102. Darkin-Rattray SJ, Gurnett AM, Myers RW, Dulski PM, Crumley TM, et al. (1996) Apicidin: a novel antiprotozoal agent that inhibits parasite histone deacetylase. *Proc Natl Acad Sci USA* 93: 13143–13147.
103. Mai A, Cerbara I, Valente S, Massa S, Walker LA, et al. (2004) Antimalarial and antileishmanial activities of aroyl-pyrrolyl-hydroxyamides, a new class of histone deacetylase inhibitors. *Antimicrob Agents Chemother* 48: 1435–1436.
104. Andrews KT, Tran TN, Lucke AJ, Kahnberg P, Le GT, et al. (2008) Potent antimalarial activity of histone deacetylase inhibitor analogues. *Antimicrob Agents Chemother* 52: 1454–1461. doi:10.1128/AAC.00757-07.

105. Mukherjee P, Pradhan A, Shah F, Tekwani BL, Avery MA (2008) Structural insights into the *Plasmodium falciparum* histone deacetylase 1 (PfHDAC-1): A novel target for the development of antimalarial therapy. *Bioorg Med Chem* 16: 5254–5265. doi:10.1016/j.bmc.2008.03.005.
106. Cui L, Fan Q, Cui L, Miao J (2008) Histone lysine methyltransferases and demethylases in *Plasmodium falciparum*. *Int J Parasitol* 38: 1083–1097. doi:10.1016/j.ijpara.2008.01.002.
107. Volz JC, Bártfai R, Petter M, Langer C, Josling GA, et al. (2012) PfSET10, a *Plasmodium falciparum* methyltransferase, maintains the active var gene in a poised state during parasite division. *Cell Host Microbe* 11: 7–18. doi:10.1016/j.chom.2011.11.011.
108. Kouzarides T (2007) Chromatin modifications and their function. *Cell* 128: 693–705. doi:10.1016/j.cell.2007.02.005.
109. Pérez-Toledo K, Rojas-Meza AP, Mancio-Silva L, Hernández-Cuevas NA, Delgadillo DM, et al. (2009) *Plasmodium falciparum* heterochromatin protein 1 binds to tri-methylated histone 3 lysine 9 and is linked to mutually exclusive expression of var genes. *Nucleic Acids Res* 37: 2596–2606. doi:10.1093/nar/gkp115.
110. Chandra BR, Olivieri A, Silvestrini F, Alano P, Sharma A (2005) Biochemical characterization of the two nucleosome assembly proteins from *Plasmodium falciparum*. *Mol Biochem Parasitol* 142: 237–247. doi:10.1016/j.molbiopara.2005.04.006.
111. Briquet S, Boschet C, Gissot M, Tissandé E, Sevilla E, et al. (2006) High-mobility-group box nuclear factors of *Plasmodium falciparum*. *Eukaryotic Cell* 5: 672–682. doi:10.1128/EC.5.4.672-682.2006.
112. Gissot M, Ting L-M, Daly TM, Bergman LW, Sinnis P, et al. (2008) High mobility group protein HMGB2 is a critical regulator of *Plasmodium* oocyst development. *J Biol Chem* 283: 17030–17038. doi:10.1074/jbc.M801637200.
113. Hoeijmakers WAM, Flueck C, François K-J, Smits AH, Wetzel J, et al. (2012) *Plasmodium falciparum* centromeres display a unique epigenetic makeup and cluster prior to and during schizogony. *Cellular microbiology*. Available: <http://www.ncbi.nlm.nih.gov/pubmed/22507744>. Accessed 30 May 2012.
114. Choi S-W, Keyes MK, Horrocks P (2006) LC/ESI-MS demonstrates the absence of 5-methyl-2'-deoxycytosine in *Plasmodium falciparum* genomic DNA. *Mol Biochem Parasitol* 150: 350–352. doi:10.1016/j.molbiopara.2006.07.003.
115. Gissot M, Choi S-W, Thompson RF, Grealley JM, Kim K (2008) *Toxoplasma gondii* and *Cryptosporidium parvum* lack detectable DNA cytosine methylation. *Eukaryotic Cell* 7: 537–540. doi:10.1128/EC.00448-07.
116. Florens L, Washburn MP, Raine JD, Anthony RM, Grainger M, et al. (2002) A proteomic view of the *Plasmodium falciparum* life cycle. *Nature* 419: 520–526. doi:10.1038/nature01107.
117. Lasonder E, Ishihama Y, Andersen JS, Vermunt AMW, Pain A, et al. (2002) Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry. *Nature* 419: 537–542. doi:10.1038/nature01111.
118. Florens L, Liu X, Wang Y, Yang S, Schwartz O, et al. (2004) Proteomics approach reveals novel proteins on the surface of malaria-infected erythrocytes. *Mol Biochem Parasitol* 135: 1–11.
119. Vincensini L, Richert S, Blisnick T, Van Dorsselaer A, Leize-Wagner E, et al. (2005) Proteomic analysis identifies novel proteins of the Maurer's clefts, a secretory compartment delivering *Plasmodium falciparum* proteins to the surface of its host cell. *Mol Cell Proteomics* 4: 582–593. doi:10.1074/mcp.M400176-MCP200.
120. Lamarque M, Tastet C, Poncet J, Demetree E, Jouin P, et al. (2008) Food vacuole proteome of the malarial parasite *Plasmodium falciparum*. *Proteomics Clin Appl* 2: 1361–1374. doi:10.1002/prca.200700112.
121. Sam-Yellowe TY, Florens L, Wang T, Raine JD, Carucci DJ, et al. (2004) Proteome analysis of rhoptry-enriched fractions isolated from *Plasmodium* merozoites. *J Proteome Res* 3: 995–1001. doi:10.1021/pr049926m.
122. Lal K, Prieto JH, Bromley E, Sanderson SJ, Yates JR 3rd, et al. (2009) Characterisation of *Plasmodium* invasive organelles; an ookinete microneme proteome. *Proteomics* 9: 1142–1151. doi:10.1002/pmic.200800404.
123. Lasonder E, Janse CJ, van Gemert G-J, Mair GR, Vermunt AMW, et al. (2008) Proteomic profiling of *Plasmodium* sporozoite maturation identifies new proteins essential for parasite development and infectivity. *PLoS Pathog* 4: e1000195. doi:10.1371/journal.ppat.1000195.
124. Khan SM, Franke-Fayard B, Mair GR, Lasonder E, Janse CJ, et al. (2005) Proteome analysis of separated male and female gametocytes reveals novel sex-specific *Plasmodium* biology. *Cell* 121: 675–687. doi:10.1016/j.cell.2005.03.027.
125. Silvestrini F, Lasonder E, Olivieri A, Camarda G, van Schaijk B, et al. (2010) Protein export marks the early phase of gametocytogenesis of the human malaria parasite *Plasmodium falciparum*. *Mol Cell Proteomics* 9: 1437–1448. doi:10.1074/mcp.M900479-MCP200.
126. Le Roch KG, Johnson JR, Florens L, Zhou Y, Santrosyan A, et al. (2004) Global analysis of transcript and protein levels across the *Plasmodium falciparum* life cycle. *Genome Res* 14: 2308–2318. doi:10.1101/gr.2523904.
127. Hall N, Karras M, Raine JD, Carlton JM, Kooij TWA, et al. (2005) A comprehensive survey of the *Plasmodium* life cycle by genomic, transcriptomic, and proteomic analyses. *Science* 307: 82–86. doi:10.1126/science.1103717.

128. Cui L, Fan Q, Li J (2002) The malaria parasite *Plasmodium falciparum* encodes members of the Puf RNA-binding protein family with conserved RNA binding activity. *Nucleic Acids Res* 30: 4607–4617.
129. Fan Q, Li J, Kariuki M, Cui L (2004) Characterization of PfPuf2, member of the Puf family RNA-binding proteins from the malaria parasite *Plasmodium falciparum*. *DNA Cell Biol* 23: 753–760. doi:10.1089/1044549042531413.
130. Mair GR, Braks JAM, Garver LS, Wiegant JCAG, Hall N, et al. (2006) Regulation of sexual development of *Plasmodium* by translational repression. *Science* 313: 667–669. doi:10.1126/science.1125129.
131. Mair GR, Braks JAM, Garver LS, Dimopoulos G, Hall N, et al. (2006) Translational Repression is essential for *Plasmodium* sexual development and mediated by a DDX6-type RNA helicase. *Science* 313: 667–669. doi:10.1126/science.1125129.
132. Foth BJ, Zhang N, Mok S, Preiser PR, Bozdech Z (2008) Quantitative protein expression profiling reveals extensive post-transcriptional regulation and post-translational modifications in schizont-stage malaria parasites. *Genome Biol* 9: R177. doi:10.1186/gb-2008-9-12-r177.
133. Pal-Bhowmick I, Vora HK, Jarori GK (2007) Sub-cellular localization and post-translational modifications of the *Plasmodium yoelii* enolase suggest moonlighting functions. *Malar J* 6: 45. doi:10.1186/1475-2875-6-45.
134. Treeck M, Sanders JL, Elias JE, Boothroyd JC (2011) The phosphoproteomes of *Plasmodium falciparum* and *Toxoplasma gondii* reveal unusual adaptations within and beyond the parasites' boundaries. *Cell Host Microbe* 10: 410–419. doi:10.1016/j.chom.2011.09.004.
135. Tewari R, Straschil U, Bateman A, Böhme U, Cherevach I, et al. (2010) The systematic functional analysis of *Plasmodium* protein kinases identifies essential regulators of mosquito transmission. *Cell Host Microbe* 8: 377–387. doi:10.1016/j.chom.2010.09.006.
136. Peixoto L, Chen F, Harb OS, Davis PH, Beiting DP, et al. (2010) Integrative genomic approaches highlight a family of parasite-specific kinases that regulate host responses. *Cell Host Microbe* 8: 208–218. doi:10.1016/j.chom.2010.07.004.
137. Wilkes JM, Doerig C (2008) The protein-phosphatome of the human malaria parasite *Plasmodium falciparum*. *BMC Genomics* 9: 412. doi:10.1186/1471-2164-9-412.
138. Solyakov L, Halbert J, Alam MM, Semblat J-P, Dorin-Semblat D, et al. (2011) Global kinomic and phosphoproteomic analyses of the human malaria parasite *Plasmodium falciparum*. *Nat Commun* 2: 565. doi:10.1038/ncomms1558.
139. Miranda-Saavedra D, Gabaldón T, Barton GJ, Langsley G, Doerig C (n.d.) The kinomes of apicomplexan parasites. *Microbes and Infection*. Available:<http://www.sciencedirect.com/science/article/pii/S1286457912000974>. Accessed 30 May 2012.

Chapter 2: Organellar proteomics reveals hundreds of novel nuclear proteins in the malaria parasite *P. falciparum* (manuscript in review, Genome Biology)

Sophie C. Oehring^{1,2,#}, Ben J. Woodcroft^{3,#}, Suzette Moes⁴, Johanna Wetzel^{1,2}, Olivier Dietz^{1,2}, Andreas Pulfer^{1,2}, Chaitali Dekiwadia³, Pascal Maeser^{1,2}, Christian Flueck^{1,2}, Kathrin Witmer^{1,2}, Nicolas M. B. Brancucci^{1,2}, Igor Niederwieser^{1,2}, Paul Jenoe⁴, Stuart A. Ralph³ and Till S. Voss^{1,2,*}

¹Department of Medical Parasitology and Infection Biology, Swiss Tropical and Public Health Institute, 4002 Basel, Switzerland. ²University of Basel, 4003 Basel, Switzerland. ³Department of Biochemistry and Molecular Biology, Bio21 Molecular Science and Biotechnology Institute, The University of Melbourne, Victoria 3010, Australia. ⁴Biozentrum, University of Basel, 4056 Basel, Switzerland.

[#]These authors contributed equally to this work.

*Corresponding author: till.voss@unibas.ch

Author's email addresses (in the order of appearance):

sophie.oehring@unibas.ch, woodibe@gmail.com, suzette.moes@unibas.ch, wetzel.johanna@gmail.com, olivier.dietz@unibas.ch, pulfer.a@gmail.com, dcd@unimelb.edu.au, pascal.maeser@unibas.ch, chrigu@gmail.com, kwkiwi@gmail.com, nicolas.brancucci@unibas.ch, igor.niederwieser@unibas.ch, paul.jenoe@unibas.ch, saralph@unimelb.edu.au, till.voss@unibas.ch

Abstract

Background: The post-genomic era of malaria research provided unprecedented insights into the biology of *Plasmodium* parasites. Due to the large evolutionary distance to model eukaryotes, however, we lack a profound understanding of many processes in *Plasmodium* biology. One example is the cell nucleus, which controls the parasite genome in a development- and cell cycle-specific manner through mostly unknown mechanisms. To study this important organelle in detail, we conducted an integrative analysis of the *P. falciparum* nuclear proteome.

Results: We combined high accuracy mass spectrometry and bioinformatic approaches to present for the first time an experimentally determined core nuclear proteome for *P. falciparum*. Besides a large number of factors implicated in known nuclear processes, one third of all detected proteins carry no functional annotation, including many phylum- or genus-specific factors. Importantly, extensive experimental validation using 30 transgenic cell lines confirmed the high specificity of this inventory, and revealed distinct nuclear localisation patterns of hitherto uncharacterised proteins. Further, our detailed analysis of the core nuclear proteome identified novel protein domains potentially implicated in gene transcription pathways, and sheds important new light on nuclear compartments and processes including regulatory complexes, the nucleolus, nuclear pores, and nuclear import pathways.

Conclusion: Our study provides comprehensive new insight into the biology of the *Plasmodium* nucleus and will serve as an important platform for dissecting general and parasite-specific nuclear processes in malaria parasites. Moreover, as the first nuclear proteome characterised in any protist organism, it will provide an important resource for studying evolutionary aspects of nuclear biology.

Keywords

Malaria, *Plasmodium falciparum*, nucleus, proteomics, bioinformatics, IFA, transcription, nucleolus, nuclear pore, transfection

Background

As one of the most deadly infectious diseases in the world, malaria causes close to 500 million clinical cases and one million deaths every year [1,2]. Most of this burden is due to infections with *Plasmodium falciparum*, one of six *Plasmodium* species known to elicit malaria in humans [3,4]. Malaria-related morbidity and mortality is exclusively associated with the erythrocytic stage of infection where repeated rounds of intracellular parasite development and re-invasion into red blood cells (RBCs) lead to exponential parasite proliferation. The entire parasite life cycle is much more complex involving several morphologically and functionally distinct extra- and intracellular stages, and obligate transmission between two hosts, female *Anopheles* spp. and humans.

The key to this amazing biological complexity lies within the parasite nucleus that, in case of *P. falciparum*, encloses and regulates a 23Mb genome encoding 5400 genes on 14 linear chromosomes [5]. However, albeit many nuclear processes such as transcription, splicing, DNA replication/repair, mitosis, and the temporal and spatial organisation of the nucleus have been studied in detail in model eukaryotes our understanding of nuclear biology in *P. falciparum* is very limited. This is not surprising given that more than 50% of all genes code for proteins with no known or even inferred function [5-7]. While many seminal studies in the post-genomic era of malaria research provided unprecedented insights into the biology of *P. falciparum*, they also highlighted our profound lack of understanding of basic biological processes in this parasite. In light of spreading drug resistance and the eager expectation for an effective vaccine, acquisition of such knowledge is urgently needed.

During the pre-replicative phase of the intra-erythrocytic developmental cycle (IDC), parasites develop into morphologically distinct ring and trophozoite stages. Schizogony is characterised by multiple rounds of genome replication and closed mitosis before cytokinesis produces new daughter merozoites from multinucleated schizonts [8,9]. At the ultrastructural level, the parasite nucleus appears spherical and contains a mixture of electron-sparse and electron-dense material probably reflecting euchromatic and heterochromatic zones, respectively. The distribution of this material appears to be sensitive to its precise fixation and preparation [10-12].

Several high-throughput transcriptome and proteome profiling studies revealed that in addition to a core set of genes expressed in multiple/all life cycle stages, a large number of genes are specifically expressed in only a single stage, many of which are involved in highly specialised processes and pathways [13-19]. Differential gene expression is also

strikingly observed during the 48 hour IDC. Detailed microarray experiments performed at high temporal resolution identified a surprisingly structured cascade of gene transcription during this stage [20-22]. About 80% of the genes expressed during the IDC display temporal variation in transcript abundance where genes appear to be activated only when their encoded protein functions are required [21]. Notably, the timely expression of variant protein families involved in immune evasion and RBC invasion is directly related to the pronounced virulence of *P. falciparum* [23]. However, despite the fact that many of these processes are likely governed by transcriptional control little detail on the underlying mechanisms has yet been elucidated.

General transcription factors (TFs) and chromatin remodelling activities are well conserved in the *P. falciparum* proteome [24-27]. Several factors of the latter class have recently been localised to different subcompartments within the parasite nucleus [28]. Functional studies identified important roles for the histone deacetylase PfGCN5, silent information regulator 2 (PfSIR2) and heterochromatin protein 1 (PfHP1) in parasite development, heterochromatin formation and virulence gene expression [29-37]. In contrast, however, the striking under-representation of identified sequence-specific TFs in the *P. falciparum* proteome compared to those of fungi, plants and metazoans has hampered targeted research to understand gene-specific control [24,38]. Until recently, only a single TF, PfMYB1, had been analysed to any extent *in vivo* [39]. Fortunately, the discovery of the apicomplexan-specific ApiAP2 family of DNA-binding factors and functional analysis of some members sparked new interest in this field [40-46].

Most proteomic studies on *Plasmodium* parasites have focused on elucidating whole cell proteomes, which generated valuable insight into the overall structure of, and differences between, the active proteomes in different parasite life cycle stages [13,14,16,17,47], or in response to perturbations such as drug treatment [48-51]. However, while these approaches typically detect large numbers of different proteins they fail to provide information on subcellular protein localisation. Organelar proteomics is an important tool to gain new insight into cellular structures and functions since proteins localising to distinct subcellular compartments are usually associated with the function of these compartments. To date, only a few mass spectrometry-based studies aimed at identifying protein compositions of *Plasmodium* cellular compartments. These include the analysis of fractions enriched for *P. falciparum* food vacuoles (116 proteins) [52], Maurer's clefts (78 proteins) [53], and the infected RBC membrane (36 proteins) [54].

Rodent *Plasmodium* spp. have also been analysed, searching for rhoptry (36 proteins) [55] and micronemal proteins (345 proteins) [56].

Here, we performed a detailed proteomic analysis of *P. falciparum* nuclei during intra-erythrocytic development. Our approach combined multidimensional protein identification technology (MudPIT) of purified and fractionated nuclei with validation by both bioinformatic appraisal and *in vivo* localisation experiments. We present a robust core nuclear proteome consisting of 802 proteins. Our comprehensive analysis of this inventory provides unprecedented insight into the parasite nucleus and will be of great benefit to future studies investigating nuclear biology in this important pathogen.

Results

Isolation and fractionation of parasite nuclei

To obtain a broad overview of proteins localised to the parasite nucleus we identified the protein content of crude nuclear preparations followed by further biochemical fractionation from ring, trophozoite and schizont stage parasites by high accuracy mass spectrometry. An overview of the experimental procedure is illustrated in Additional file 1. Isolated nuclei were significantly increased in size compared to intact parasites, a phenomenon regularly seen in preparations of nuclei after hypotonic lysis [57,58] (Figure 1a). Despite extensive washing of the nuclear pellet contamination with free hemozoin crystals was obvious. Analysis by transmission electron microscopy (TEM) showed that the nuclear fraction mainly consisted of rounded and enlarged nuclei with varying degrees of intactness (Figure 1b). No consistent organellar impurities were apparent, although some haemozoin crystals were also visible throughout these samples. Importantly, immunolabeling with antibodies specific for histone 3 lysine 4 tri-methylation (H3K4me3) identified these structures as parasite nuclear material (Figure 1c).

We next performed protein fractionation to reduce sample complexity. Nuclei were serially extracted with 0.1M KCl (fraction 2), DNaseI (fraction 3), 1M KCl (fraction 4) and 2% SDS (fraction 5). Each fraction displayed a distinct 1D-SDS-PAGE protein pattern indicative of differential protein extraction (Figure 1d and Additional file 1). The cytosolic enzyme glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and the mitochondrial heat shock protein 60 (HSP60) were exclusively detected in the NP40-soluble cytoplasmic extract demonstrating efficient lysis of the parasite plasma membrane and the double membrane-bound mitochondrion (Figure 1e). In contrast, H4 and PfHP1 were

soluble only after extraction of nuclear pellets with high salt and SDS, consistent with their intimate association with chromatin. In summary, the microscopy-, SDS-PAGE- and immunodetection-based assessments show that the protocol applied here efficiently separated the cytoplasmic and nuclear compartments and yielded distinct subnuclear protein fractions suitable for mass spectrometry analysis.

Nuclear proteome determination by MudPIT

Endoproteinase LysC- and trypsin-digested total proteins were analysed by two-dimensional capillary liquid chromatography and tandem mass spectrometry (MudPIT). MS/MS spectra were searched against a combined *P. falciparum*/human proteome database using the SEQUEST algorithm [59] (Additional file 2). We identified 1518 different parasite proteins that were represented by at least one tryptic peptide in any of the 30 samples analysed, and the numbers of unique peptides in individual samples ranged from 406 to 3852 (Additional file 3). A decoy search strategy querying a reversed sequence database resulted in very low mean false discovery rates (FDR) of 0.28 +/- 0.24 sd per sample. Furthermore, random inspection of single peptide-based protein identifications confirmed the automated protein identification of the search engine. The SEQUEST output data for ring, trophozoite and schizont stages (Additional files 4-6) and a comprehensive summary table (Additional file 7) are provided as supplementary information.

Similar numbers of proteins were identified for ring stages (1050), trophozoites (1017) and schizonts (1092). 649 proteins were shared by all three stages, and similar numbers of proteins were either unique to one stage or shared between any two of the three stages (Figure 2a). In the cytoplasmic and combined nuclear fractions 870 and 1273 proteins were detected, respectively, and 625 were shared between both compartments (Figure 2b). Figure 2c shows the distribution of proteins in individual nuclear fractions (DNaseI- and high salt-soluble fractions were combined into one category potentially enriched in DNA/chromatin-associated proteins). 20.1%, 25.5% and 43.6% of proteins measured in the low salt, DNaseI/high salt and SDS fractions, respectively, were unique to the corresponding fraction.

The preparation of isolated nuclei is enriched in nuclear proteins

Functional enrichment analyses using the David and the GOstat tools [60,61] revealed that in all cases the combined nuclear fractions were statistically enriched in annotations consistent with nuclear functions (Additional files 8 and 9). As shown in Table 1, all annotated histones were detected and their distribution was marked by a clear enrichment in the nuclear fractions. We measured ten out of twelve subunits of the DNA-directed RNA polymerase II (RNAPolII) complex, all of which were detected in fraction 4. Peptides derived from ApiAP2 factors (13 out of 27 detected) were exclusively present in nuclear fractions. Furthermore, we detected 111 out of 247 *in silico* predicted *P. falciparum* transcription-associated proteins (TAPs) [38,62], 108 of which (97.3%) were present in any of the four nuclear fractions (Additional file 7).

To further assess the enrichment in nuclear proteins, we carried out comparisons to curated lists of proteins with known or likely nuclear localisation. First, lists were derived from ApiLoc (apiloc.biochem.unimelb.edu.au), a database of apicomplexan proteins previously localised by microscopy (unpublished). At the time of analysis, it held information on 424 *P. falciparum* proteins, 60 of which were nuclear-localised at any stage during the life cycle (14.3%). A simple algorithm, designed to decide whether the annotation recorded in ApiLoc is consistent with a protein being found in the nucleus (Additional file 9), showed that 37 of these 60 nuclear proteins were detected in the nuclear preparation (Additional file 7). Proteins in the cytoplasmic and any of the four nuclear fractions were enriched in nuclear-localised proteins, whereas those detected exclusively in the cytoplasmic fraction were depleted (one nuclear protein detected) (Figure 3a). These enrichments were statistically significant under the assumption that the proteins with recorded localisation annotation were representative (p -value $< 10^{-6}$; Fisher's exact test). Similarly, accuracy was tested by comparison to a curated list of proteins whose localisation has been inferred in peer-reviewed articles, largely through their having an experimentally localised predicted orthologue in model organisms such as *Saccharomyces cerevisiae* or *Homo sapiens* (Additional file 10). Of 465 proteins in this list, 296 (63.7%) are linked to nuclear localisation, 136 of which were detected in the nuclear preparation (enrichment p -value $< 10^{-4}$), and only eight were present exclusively in the cytoplasmic fraction (Figure 3a and Additional file 7). In summary, this overall assessment of the proteomic output highlights the enrichment of known and probable nuclear proteins in the nuclear fractions.

Next, we applied two measures, positive predictive value (PPV) and the number of proteins remaining in each set, to assess differences in content between the nuclear fractions. The pool of detected proteins annotated as nuclear in the ApiLoc database and/or by literature review was used as reference set (159 proteins). As expected, the sets of proteins found in individual fractions showed varying PPVs with the least and most predictive observed in the cytoplasmic lysate and high salt nuclear extract, respectively (Figure 3b). The sum of proteins found in the combined nuclear fractions gave the largest set at a high PPV. Considering only those proteins in this set that were not also detected in the cytoplasm increased the PPV by 8% but came at too steep a cost by removing 49% of all proteins including many of nuclear and unknown localisation. This shows that the cytoplasmic fraction is not especially deficient in nuclear proteins, consistent with the presence of many nuclear proteins in both compartments. In contrast, proteins found exclusively in the cytoplasmic fraction were markedly deficient in predicted nuclear proteins.

Application of bioinformatic reduction techniques to generate a high-confidence core nuclear proteome dataset

A variety of bioinformatic filters were applied to remove components inconsistent with nuclear localisation. Each filter applied was motivated by a hypothesis grounded in current knowledge of the properties of nuclear proteins. We accepted filters to remove proteins (1) carrying a predicted signal peptide (SP); (2) carrying a predicted transmembrane (TM) domain; (3) predicted to be exported to the RBC based on the presence of a PEXEL motif [63,64] and/or the association of the encoding genes with PfHP1 [35]; and (4,5) found in two previously published *P. falciparum* organelle proteomic studies of Maurer's clefts [53] and the food vacuole [52]. We rejected filters to remove proteins (1) predicted to localise to the mitochondrion by PlasMit [65]; (2) detected by a single peptide only; and (3) detected in the cytoplasmic fraction. Benchmarked results for each filter are shown in Figure 4, and full rationales for accepting or rejecting filters are available in Additional file 11.

Applying the five filters resulted in a set of 841 proteins with a PPV of 70% (Figure 4b). After removal of 55 proteins experimentally localised to non-nuclear compartments (Additional file 12), re-addition of 16 known nuclear proteins (Additional file 13), and adjustments due to altered SP annotations (Additional files 14 and 15) we present a set of 802 proteins, whereof we estimate that 76% represent true nuclear proteins

(compared to 46% in the unfiltered dataset) (Figure 4b). This set is henceforth referred to as the “core nuclear proteome” and all further analyses were carried out on this set (Additional file 16).

An overview of the contents of the core nuclear proteome is shown in Figure 4c where all proteins are categorised into broad generic classes based on functional annotation. Apart from proteins engaged in typical nuclear processes such as transcription, chromatin remodelling, DNA replication/repair, splicing and nucleolar functions, the core nuclear proteome contains many putative RNA-binding proteins and factors involved in RNA processing and metabolism, protein folding, modification and degradation, and cytoskeletal organisation. Notably, one third of the core nuclear proteome consists of proteins with no annotated function, and only 5% of all proteins are reliably allocated to non-nuclear compartments.

Experimental validation by subcellular localisation of selected nuclear candidates

To test if the core nuclear proteome will be useful in identifying true nuclear proteins we experimentally validated the subcellular localisation of 28 proteins by indirect immunofluorescence assays (IFA). We generated transgenic parasite lines expressing full-length C-terminally tagged proteins from episomally maintained plasmids. We chose 22 nuclear protein candidates (NuProCs1-22) that were included in the core nuclear proteome, and six non-nuclear protein candidates (n-NuProCs1-6) that were excluded from the list based on the presence of predicted TM domains and/or SPs (Additional file 17). Of the 22 NuProCs, seven carried annotations indicative for nuclear localisation and 15 were annotated as hypothetical proteins at the time of selection (PlasmoDB releases 5.0-6.4). Amongst the negative control set, n-NuProC1 (PF11_0099) is annotated as heat shock protein DnaJ homolog Pfj2 whereas all other n-NuProCs are hypothetical proteins with unknown function.

16 of the 22 NuProCs co-localised exclusively with the DAPI-stained area of the nucleus in trophozoites (Figure 5a). Detailed imaging of each cell line throughout the IDC is available as supplementary material (Additional files 18-39). Within this group of confirmed nuclear proteins are six of the seven candidates annotated as putative nuclear proteins. A diffuse pattern was observed for NuProC1 (putative regulator of chromosome condensation) (Additional file 18) and NuProC5 (putative fork head domain protein) (Additional file 22). NuProC7 (structure-specific recognition protein) displayed a more condensed appearance throughout the IDC (Additional file 24). NuProC2 (putative

nucleolar preribosomal assembly protein) and NuProC3 (putative splicing factor) both localised to a distinct subnuclear region (Additional files 19 and 20). NuProC6 encodes a putative bromodomain protein with a more restricted and peripheral localisation in ring stages and a diffuse nuclear pattern in trophozoites and schizonts (Additional file 23).

Ten candidates annotated as hypothetical proteins (NuProC9, 10, 11, 12, 14, 15, 17-20) were also unambiguously localised within the parasite nucleus. A diffuse pattern was evident for NuProC12 (Additional file 29), NuProC15 (Additional file 32), NuProC17 (Additional file 34) and NuProC20 (Additional file 37). NuProC18 showed a cytosolic staining in late schizont and ring stages but was clearly nuclear in trophozoites (Additional file 35). NuProC9 changed from a rather ubiquitous to a more condensed nuclear localisation upon transition from the early to late schizont stage (Additional file 26). Both NuProC10 (Additional file 27) and NuProC11 (Additional file 28) are localised to unknown nuclear subcompartments. NuProC14 (Additional file 31) and NuProC19 (Additional file 36) were preferentially associated with the nuclear periphery in ring stages but showed a diffuse pattern in trophozoite nuclei.

Two additional proteins, NuProC4 (putative multiprotein bridging factor type 1 MBF1) (Additional file 21) and NuProC13 (Additional file 30) localised outside but in close proximity to the DAPI signal (Figure 5b). These patterns were clearly distinct from the cytosolic GAPDH signal and reminiscent of that obtained for the nuclear porin PfNUP100 [28] suggesting that these two factors may be associated with the nuclear periphery. In addition, NuProC4 appeared to accumulate in the cytosol in mature trophozoites (Additional file 21).

The remaining four of the 22 tagged NuProCs showed a fluorescence pattern inconsistent with nuclear localisation (Figure 6a). NuProC21 (nucleosome assembly protein PfNAPL; PFL0185c) (Additional file 38), NuProC8 (Additional file 25) and NuProC22 (Additional file 39) co-localised to various degrees with GAPDH throughout the IDC suggesting a cytosolic/cytoplasmic localisation. Interestingly, NuProC22 occurred in close proximity to the nuclear periphery in late and segmented schizonts indicating that this protein may be transiently associated with the parasite nucleus. NuProC16, currently annotated as putative PFMNL-1 C1SD1-like iron-sulphur protein, localised to a cytoplasmic structure distinct from the cytosolic and ER compartments, which likely represents the mitochondrion (Additional file 33). This conclusion is supported by the presence of an iron-binding zinc finger domain CDGSH [66] found in the outer mitochondrial membrane protein MitoNEET in vertebrates [67].

None of the six non-nuclear protein candidates localised to the nucleus. Five co-localised with PfBIP throughout the IDC, and n-NuProC3 was associated with the ER in early stages and exported into the host cell in trophozoites and schizonts (Figure 6b and Additional files 40-45). The strong representation of ER-localised proteins in the negative control set indicates that the crude nuclear preparation was enriched in ER-associated proteins. However, we can not exclude the possibility that ER association of some of these proteins may be artefactual due to over-expression of the tagged proteins.

In summary, our experimental validation of the predicted core nuclear proteome tested 16 out of 22 nuclear protein candidates (72.7%) positive for true nuclear localisation, plus another two that are associated with the nuclear periphery. This figure is consistent with the estimated precision of the bioinformatically reduced core set of nuclear proteins and validates this inventory as an important platform for the identification and characterisation of novel parasite nuclear proteins.

Two novel nucleolar proteins in *P. falciparum*

To test if the two distinct subnuclear domains delineated by NuProC2 and NuProC3 represented the same compartment we generated a parasite line expressing both proteins simultaneously. IFA analysis revealed that both proteins co-localised within the same intra-nuclear compartment that was preferentially located towards the nuclear periphery in a DAPI-negative area (Figure 7a). A monoclonal antibody directed against the human nucleolar marker fibrillarin [68] that is cross-reactive with *Toxoplasma gondii* fibrillarin [69], specifically recognised the same compartment as NuProC2 or NuProC3 (Figure 7b). Hence, these co-localisation experiments identified two novel nucleolar proteins in *P. falciparum*.

Identification of novel domains associated with *P. falciparum* nuclear proteins

Using the core nuclear proteome as an input collection, we sought to detect novel functional domains enriched in nuclear proteins using an all-versus-all BLAST search followed by manual inspection (Additional file 46). Six novel domains comprised of 17 proteins warranted further inspection, all of which were conserved across Apicomplexa providing further evidence that these modules are functional.

One of these domains (approx. 80aa) was found exclusively in ApiAP2 proteins [40,70,71]. It occurred in nine *P. falciparum*, 15 *T. gondii* and two *Babesia bovis* proteins, and in one protein from each of the sequenced *Cryptosporidium* spp. (Additional files 47-

49). Unlike AP2 domains, this novel domain was only ever found once in each protein. Interestingly, this domain was located at the C-terminus in all but one of the *P. falciparum* ApiAP2s where it is found at the N-terminus (Figure 8). This relative position appears to be at least moderately conserved evolutionarily with 42 of 56 proteins having this domain located less than 50 amino acids away from the C-terminus. Hence, we termed this novel domain ACDC (AP2-coincident domain mostly at the C-terminus). The noticeable co-occurrence of these domains in a family of apicomplexan DNA-binding proteins invites testable hypotheses about potential roles of the ACDC domain in gene expression, chromatin structure and/or other aspects of chromosome biology.

Four of the five other novel domains are associated with either cleavage of mRNA 3' UTRs (partial cleavage stimulation factor (CSTF) domain), transcriptional regulation (extended ELM2 domain, MYND domain), or the cytoskeleton (alveolin domain) (Additional files 11 and 50-55). A further domain identified in our search was encoded by two proteins annotated as PfNUP100 (PFI0250c) and a hypothetical protein (PF14_0442). The localisation of NUP100 to the nuclear membrane [28], and the fact that much of the conservation between the two proteins lies in the phenylalanine-glycine (FG) pairs of amino acids, suggested that both are FG-repeat nuclear pore components [72] (Additional file 56). To verify if PF14_0442 encodes a nuclear pore protein we generated a transgenic cell line expressing endogenously tagged PF14_0442-GFP (Additional file 57). Indeed, IFA analysis revealed that this protein localised to the nuclear rim, internal to the ER, with several focal points of higher intensity, a staining pattern reminiscent of nuclear pores (Figure 9).

Classical nuclear localisation signals are only marginally over-represented in *P. falciparum* nuclear proteins

Classical nuclear localisation signals (cNLSs) have been the subject of several bioinformatic prediction algorithms, and our core nuclear proteome provided an opportunity to test their utility in identifying *P. falciparum* nuclear proteins. We used three different bioinformatic tools to predict cNLSs: NLStradamus [73], predictNLS [74] and cNLS mapper [75]. All three algorithms attempt to predict nuclear localisation based solely on the presence of cNLSs and not other protein domains.

NLStradamus, PredictNLS and cNLS mapper suggest 51%, 37%, and 22% of proteins found in the core nuclear proteome contain a cNLS (Additional file 16), respectively, compared to 45%, 31% and 17% of the entire *P. falciparum* proteome ($p < 0.01$ for all

three predictors; Fisher's exact test). In contrast, only 47%, 32% and 18% of proteins found in the whole cell trophozoite proteome [17] were predicted by these algorithms (p-values 0.17, 0.4 and 0.18, respectively, relative to all *P. falciparum* proteins; Fisher's exact test). While this result adds some evidence to the hypothesis that cNLS-mediated nuclear import may operate in *P. falciparum* similarly to model eukaryotes, the difference in the percentage of cNLS-containing proteins in the core nuclear proteome compared to the set of all *P. falciparum* proteins is only marginal. Likewise, the same three predictors identify only 50% (NLSStradamus), 37% (PredictNLS) and 28% (cNLS mapper) of proteins in the combined ApiLoc and literature review reference set of 317 *P. falciparum* nuclear proteins to contain a cNLS (p-values 0.08, 0.07 and 0.0004, respectively, relative to all *P. falciparum* proteins; Fisher's exact test). Together, this suggests that either *P. falciparum* cNLSs may be too divergent for current bioinformatic predictors, or that the major mode of nuclear import may not occur through classical import signals, or a combination of both.

Stage-specific patterns of nuclear proteins

58, 90 and 105 proteins were found exclusively in ring, trophozoite and schizont stages, respectively. In trophozoites and schizonts, expression of these proteins occurred roughly in line with transcription of the encoding genes (Additional files 9, 58 and 59). mRNA expression of ring stage-specific nuclear proteins was somewhat surprising, with a collective profile similar to that observed in schizonts, and an additional smaller peak at 15-20 hpi. This suggests that some proteins were newly synthesized in the ring stage while some remained from the preceding schizont stage.

159 of the 253 stage-specifically detected proteins carry no functional annotation and several others carry annotated domains that indicate little about their function. Hence, a large number of novel proteins have been assigned a potential stage-specific role in nuclear biology where previously no informative annotation was available. Further, 13 predicted TFs [62] were detected specifically in a single IDC stage only, indicating a role for these factors as cell cycle-specific regulators of transcription and genome regulation (Additional file 60).

Discussion

To deepen our insight into nuclear biology in *P. falciparum* we performed a comprehensive proteomic analysis of the parasite nucleus and detected a total of 1273 proteins in sequential extracts of crude nuclei. The proportion of predicted nuclear proteins in this set was estimated at 46%, which compares well with two similar studies in *S. cerevisiae*. Mosley et al. detected 2674 yeast proteins in crude nuclei, 46% of which are annotated as nuclear proteins in the *Saccharomyces* genome database SGD [76], and a second study based on sucrose gradient purification of nuclei detected 1889 proteins, 55% of which were annotated as nuclear in SGD [77].

The major sources of contamination in nuclear preparations are membrane fractions of non-nuclear organelles, particularly the ER. To eliminate such likely contaminants we applied an informed bioinformatic filtering approach that resulted in a markedly improved positive predictive value (76%) for true nuclear proteins in the core nuclear proteome. Importantly, we confirmed the specificity of this set by *in vivo* experimental validation. Out of 22 candidates 18 localised in different patterns to the nucleus. Some proteins displayed a rather ubiquitous distribution within the nucleus, whereas others localised to more restricted, undefined subnuclear regions, to the nucleolus or to the nuclear periphery. These alternative localisations are a likely consequence of the different functions these factors carry out in the nucleus. Four candidates localised primarily to the cytoplasm, however, some of them may also be nuclear. Firstly, we determined localisations using episomally expressed epitope-tagged proteins, which may cause overexpression and cytosolic accumulation of some candidates. Secondly, proteins primarily located in the cytosol may still be present in the nucleus in lower concentrations. For instance, PfNAPL (NuProC21), which localised to the cytoplasm here and in a previous study [78], is a putative orthologue of ScNAP1 that shuttles between the nucleus and cytosol [79]. Another example is PFL0450c (NuProc22), a protein with unknown function that localised to the cytoplasm in ring and trophozoite stages but was found at the nuclear periphery in late schizonts.

Functional classification of the core nuclear proteome

We detected 299 proteins (37.3%) implicated in nuclear processes such as transcription, chromatin remodelling, DNA replication/repair, RNA-binding and processing, ribosome biogenesis, but also in more general processes such as protein folding and modification, protein degradation, translation, cytoskeleton organisation and metabolism. This

classification is based on direct experimental evidence, literature review and/or sequence similarity to known *S. cerevisiae* nuclear proteins (Additional file 16). 118 proteins (14.7%), for which such evidence is missing, carry annotations similarly consistent with roles in DNA/chromatin interactions, cell cycle control, mitosis, RNA-binding and processing, protein folding and modification, protein degradation, cytoskeleton organisation, and metabolic processes. 95 proteins (11.8%) represent ribosomal subunits or translation-associated factors, and 35 proteins (4.4%) are predicted to localise to other compartments such as the mitochondrion, ER/Golgi, protein sorting vesicles or the vacuole. Finally, the largest fraction of the core nuclear proteome consists of 255 proteins of unknown function (31.8%).

To date, only three proteins have been localised to the parasite nucleolus (RNA poll, PFE0465c; fibrillarin/PfNOP1, Pf14_0068; PfNOP5, PF10_0085) [80,81]. Further, only six *P. falciparum* proteins are annotated with the GO term “nucleolus”, and 36 are annotated by GeneDB (www.genedb.org) as putative nucleolar proteins. Here, we detected twelve of these proteins including fibrillarin, PfNOP5, RNA methyltransferase (PF11_0305), putative ribonucleoprotein (RNP) components such as LSM homologs (PFL0460w, PF08_0049), and several predicted pre-ribosomal assembly proteins, as well as another 19 putative homologs of *S. cerevisiae* nucleolar proteins (Additional file 16). Notably, two of these candidates, NuProC2 (PF10_0278) and NuProC3 (PF11_0250), co-localised with fibrillarin to the parasite nucleolus. These novel identifications expand our current knowledge of the *P. falciparum* nucleolus and provide a basis for detailed analyses of this essential nuclear compartment.

Plasmodium protein kinases (PK) play central roles in growth, development and differentiation throughout the life cycle, and are intensely studied as a class of promising antimalarial targets [82,83]. We identified kinases or their accessory factors of seven PK systems, six of which have been implicated in nuclear roles in *Plasmodium* or other eukaryotes: PfMAT1 (PFE0610c) [84]; the α - and β -subunits of casein kinase 2 (PF11_0096, PF11_0048) [85-87]; casein kinase 1 (PF11_0377) [88,89]; the catalytic and regulatory subunits of cAMP-dependent protein kinase (PFI1685w, PFL1110c) [90]; NIMA-related protein kinase PfNEK-1 (PFL1370w) [91,92] and mitogen-activated protein kinase 2 PfMAP2 (PF11_0147) [93,94]. PfMAP2 is a substrate of PfNEK-1 [92] and both kinases are essential for completion of the IDC in *P. falciparum* [91,95], whereas in *P. berghei* PbMAP2 is implicated specifically in the development of male gametes [93]. Interestingly, while the targets of these and other nuclear kinases remain largely

unknown, we find that 70% of proteins in the core nuclear proteome (562/802) were recently shown to be phosphorylated in intra-erythrocytic parasites [96] [Additional file 16] suggesting an important role for kinase signalling in the regulation of nuclear processes in *P. falciparum*.

We detected 32 of the 33 annotated *P. falciparum* proteasome components and all, except two of the 19S regulatory particle (RP) subunits, were represented in the cytoplasmic fraction. Intriguingly, all subunits of the RP were also associated with the insoluble nuclear fraction in ring stage parasites (Additional file 16). In other species, the RP interacts with chromatin and is involved in transcriptional regulation [97]. For example, the *S. cerevisiae* the RP participates in SAGA histone acetyltransferase complex recruitment [98], interacts with the FACT (facilitates chromatin transcription) complex [99], and influences H3 methylation and gene silencing [100]. Hence, we speculate that the 19S RP may have similar non-proteolytic roles in *P. falciparum* transcriptional regulation.

The nuclear proteome also contained several orthologues of the mRNP complex implicated in translational repression during *P. berghei* gametocytogenesis [101,102], such as the RNA helicase PfDOZI (PFC0915w), PfCITH (PF14_0717), the RNA-binding proteins PfHOBO (PF14_0096) and PfHOMU (PFI0820c), and a homolog of a yeast poly(A)-binding protein (PFL1170w). Although DOZI and CITH were described as cytoplasmic proteins in *P. berghei*, their localisations do not appear to exclude the nucleus. Furthermore, homologs of PfHOMU and PFL1170w shuttle between the cytosol and nucleus in mammalian cells and *S. cerevisiae*, respectively [103,104]. The detection of these proteins in the asexual nuclear proteome suggests the presence of translational repression machinery during the IDC, although *P. berghei* DOZI and CITH loss-of-function mutants have no apparent phenotype in asexual parasites [101,102].

We also discovered novel domains that are likely linked to previously unrecognised nuclear functions of parasite proteins, and we envisage their roles will be more thoroughly recognised in future studies. Of particular interest is the identification of the ACDC domain, which we identified exclusively in members of the ApiAP2 family. Moreover, we experimentally validated the identification of PF14_0442 as a novel subunit of *P. falciparum* nuclear pores. Testing this in more detail will be an important future task given that *Plasmodium* parasites lack identifiable orthologues of most nuclear pore components [105].

Fraction-specific aspects of the core nuclear proteome

Out of 145 proteins predicted to interact with DNA and/or chromatin, 137 (94.5%) were identified in the chromatin-containing fractions (DNaseI-, high salt- and/or SDS-soluble fractions) (Additional file 16). Some of these factors, such as histones, SNF2 helicase (PFF1185w), chromodomain-helicase-DNA-binding protein 1 (CHD1) (PF10_0232), putative chromosome assembly factor 1 (PFE0090w), and the nuclear peroxiredoxin PfnPRX (PF10_0268) [106] were detected in all three chromatin-associated fractions. In contrast, nine out of the ten RNA pol II subunits identified were exclusively detected in the high salt nuclear extract. Both FACT components (PFE0870w, PF14_0393) extracted almost identically only after high salt and SDS extraction. The two recently described high mobility group box proteins (PFL0145c, MAL8P1.72) [107] were identified in DNaseI- and high salt-soluble extracts but not in the SDS fraction. These examples suggest that at least some regulatory complexes were extracted as interacting entities. Interestingly, all ApiAP2 factors showed a noticeable association with the insoluble nuclear fraction. While the reason for this remains unknown our observation hints at possible functions of these DNA-binding proteins. Most ApiAP2 proteins are large and, apart from the short and well-defined DNA-binding AP2 domains, consist of extensive uncharacterised regions. It is possible that these non-AP2 regions may mediate the formation of regulatory complexes involved in diverse processes such as DNA replication, transcriptional regulation or functional organisation of the genome. Such complexes are often resistant to extraction with DNaseI and high salt buffers, as observed for many RNA- or DNA-binding proteins associated with the nuclear matrix [108]. In case of PfSIP2, the only *P. falciparum* ApiAP2 characterised *in vivo*, the detection of PfSIP2-derived peptides in the insoluble nuclear fraction is consistent with the association of this factor with condensed heterochromatin [46].

A large number of proteins in the core nuclear proteome (379) were also detected in the cytoplasmic fraction. This pool of proteins contained members of all functional classes but was clearly enriched in distinct pathways associated with the various functions of the nucleolus in other eukaryotes [109-111]. These include the majority of ribosomal subunits (90.5%), RNA-binding proteins (78.9%), factors involved in protein degradation (69.0%), and protein folding and modification (65.1%). Furthermore, 84.4% of translation-related factors were identified in the cytoplasmic and nuclear fractions, a finding consistent with the existence of nuclear translation [112,113]. We also noticed that 42% (62 proteins) of confirmed or likely nuclear proteins in the core nuclear

proteome were detected in the cytoplasmic fraction. Moreover, six proteins experimentally localised to the nucleus in this study had peptides detected in the cytoplasmic fraction. Hence, while some of the dually detected proteins may represent cytoplasmic contaminants such as abundant cytosolic proteins or macromolecular complexes, our results show that many nuclear proteins in *P. falciparum* shuttle between the nucleus and cytosol and/or perform their tasks at multiple destinations.

Nuclear import in *P. falciparum*

Transport of proteins into the nucleus remains poorly understood in apicomplexan parasites [114]. In yeast and mammalian model systems short, arginine- and lysine-rich cNLSs are thought to be the major mediators of nuclear import, though alternative and redundant mechanisms have been described [115]. The poor enrichments of cNLSs in both the core nuclear proteome and the set of 317 curated nuclear proteins show that bioinformatic discrimination of nuclear vs. non-nuclear *P. falciparum* proteins via prediction of cNLSs remains impractical. Notably, however, cNLS predictors are also problematic in the reliable identification of nuclear proteins in the model systems they were designed for [116]. Nevertheless, differences between *P. falciparum* nuclear proteins and these current computational models of cNLSs appear prevalent but remain unclear. It has not been determined whether the majority of *P. falciparum* nuclear proteins are imported independently of importin α , the protein that binds cNLSs [114], or alternately by similar but unrecognized cNLSs that do mediate importin α -dependent translocation. The fact that only 22% to 51% of nuclear proteins are predicted to contain cNLSs (depending on the predictor used) suggests that in its current definitional form the cNLS is not the major mode of nuclear import in *P. falciparum*. Notably, the important insight that most verifiable *Plasmodium* nuclear proteins lack a recognizable nuclear localization sequence, and need thus be identified through empirical strategies, reinforces the value of an experimentally robust nuclear proteome for understanding *Plasmodium* nuclear biology.

Lineage-specific nuclear proteins

Using a combination of OrthoMCL and synteny analyses (Additional files 61 and 62) only around ten proteins are genuinely *falciparum*-specific. Unsurprisingly, nearly all of them have no functional annotation, though one is the gametocyte-specific protein Pfg27 (PF13_0011). Pfg27 binds RNA and has previously been localised to the cytoplasm and

nucleus of gametocytes [117]. Interestingly, we detected another nuclear *P. falciparum*-specific protein of unknown function (PFB0115w) that is expressed during the IDC and in gametocytes and contains C-terminal homology to the Pfg27-specific fold [118] suggesting a functional relation between these two proteins.

The pool of around 100 genus-specific proteins represents a promising group from which to characterise features of nuclear biology peculiar to *Plasmodium*. Several are potentially worthy of prioritised treatment; these include some of the previously characterised ApiAP2s [40], as well as a large number of uncharacterised proteins with nucleic acid-binding and chromatin-interacting domains. Several others in this group have kinase or phosphatase domains and may be involved in expression regulation cascades. One of the *Plasmodium*-specific nuclear proteins is a putative metacaspase (PF14_0363). Interestingly, unrelated trypanosomatid parasites also possess nuclear localised metacaspases that are required for proliferation, possibly through modulation of cell cycle progression [119].

Conclusion

To our knowledge, this is the most detailed organelle proteome analysis in apicomplexan parasites and, importantly, the first study describing a large scale nuclear proteome of any protist species. During the past decade several intriguing features of nuclear biology in *P. falciparum* have been discovered, including the periodical behaviour of gene transcription during the IDC [21,22], the discovery of the ApiAP2 family [40], the epigenetic control of genes related to virulence [29,30,33-35,120], or the fascinating dynamic surface distribution of nuclear pores throughout the IDC [12]. These examples highlight the existence of nuclear processes distinct from those in the host and underscore the importance of understanding the underlying mechanisms. For the first time, our study provides a comprehensive and unprecedented insight into the biology of the *P. falciparum* nucleus, and the nuclear proteome presented herein will assist greatly in dissecting conserved as well as evolutionary specialised nuclear processes in *Plasmodium* and related parasites.

Materials and Methods

Parasite culture and transfection

P. falciparum 3D7 parasites were cultured as described previously [121]. Growth synchronisation was achieved by repeated sorbitol lysis [122]. Transfections were performed as described [123] and selected on 5µg/ml blasticidin-S-HCl and/or 4nM WR99210. Parasites transfected with pH_0442-GFP were cycled on/off WR three times to select for integration of the plasmid into the endogenous locus.

Isolation of nuclei and nuclear fractionation

Nuclei from ring stages, trophozoites and schizonts were isolated and fractionated as previously described with minor modifications [35]. A detailed protocol is provided in Additional file 2. For Western blot analysis protein fractions were separated on 10% SDS-PAGE gels and transferred to nitrocellulose membranes (Schleicher&Schuell). Primary antibody dilutions were: anti-H4 (Abcam, ab10158) 1:10,000; anti-HSP60 (kind gift from Geoff McFadden, University of Melbourne) 1:2,000; mAb anti-PfGAPDH 1:2,000 [124]; rabbit anti-PfHP1 (kind gift from Mike Duffy, University of Melbourne) 1:2,000; anti-GFP 1 (Roche Diagnostics, 11814460001) 1:1000.

Microscopic assessment of isolated nuclei

Nuclei were incubated with DAPI (10µg/ml) in CLB for 10 min on ice and directly observed by light and fluorescence microscopy (100x magnification). Images were obtained using a Leica DM 5000B microscope with a Leica DFC 300 FX camera and acquired via the Leica IM 1000 software and processed and overlaid using Adobe Photoshop CS2. For TEM analysis isolated nuclei or whole asynchronous 3D7 cultures were fixed, prepared for microscopy and imaged as described in [125]. For immunolabelling, nuclei were incubated with rabbit anti-H3K4Me3 (Abcam ab8580) diluted 1:300, then detected with 18nm colloidal gold-conjugated goat anti-rabbit secondary antibody (Jackson ImmunoResearch, Baltimore, USA) diluted 1:20 in the blocking buffer.

Sample preparation, MudPIT liquid chromatography tandem mass spectrometry and protein identification

A detailed protocol for sample preparation, MudPIT analysis and protein identification is provided in Additional file 2. All proteomics data are publicly available on PlasmoDB (www.plasmoDB.org) and in Additional files 4-6.

Bioinformatic data analysis

Bioinformatic methods used to assess and analyse the nuclear proteome are presented in detail in Additional files 9, 46 and 62.

Transfection constructs

To express epitope-tagged candidate proteins in *P. falciparum*, full-length genes were amplified from gDNA or cDNA and cloned into pBcam-3xHA [35] using *Bam*HI (or *Bgl*II) and *Nco*I (or *Nhe*I) restriction sites. The PFL0635c gene (NuProC6) was cloned into *Bam*HI/*Nhe*I-digested pHcam-2xTy [46]. To generate the double transfectant co-expressing NuProC2-3xHA/NuProC3-2xTy the PF11_0250 gene (NuProC3) was cloned into *Bam*HI/*Nco*I-digested pHcam-2xTy and transfected into BSD-resistant 3D7/NuProC2-3xHA parasites. All forward primers included five wild-type nucleotides directly upstream of the natural ATG start codon. pH_0442-GFP was obtained by cloning the 3' end of PF14_0442 (551 bp) into *Pst*I/*Not*I-digested pHcam-GFP (unpublished). Successful 3' replacement integration of pH_0442-GFP into the endogenous locus was verified by PCR and Western blot (Additional file 57). Primer details are listed in Additional file 63.

Indirect immunofluorescence assays

IFAs were performed with iRBCs fixed in 4% formaldehyde/0.01% glutaraldehyde as described elsewhere [126]. Primary antibody dilutions were: rat mAb anti-HA 3F10 (Roche Diagnostics) 1:100; mouse mAb anti-Ty BB2 1:2,000 (kind gift from Keith Gull); mouse mAb anti-GAPDH 1:500 [124]; rabbit anti-PfBIP 1:500 [127,128] (kind gift from Tim Gilberger); mouse mAb anti-fibrillarin 1:500 [68] (kind gift from Michael Terns, University of Georgia). Secondary antibody dilutions were: Alexa-Fluor® 568-conjugated anti-rat IgG (Molecular Probes) 1:500; Alexa-Fluor® 488-conjugated anti-rabbit IgG (Molecular Probes) 1:500; FITC-conjugated anti-mouse IgG (Kirkegaard Perry Laboratories) 1:250. Images were taken on a Leica DM 5000B microscope with a Leica

DFC 300 FX camera and acquired via the Leica IM 1000 software and processed and overlaid using Adobe Photoshop CS2. Observed protein localisations have been deposited in ApiLoc.

Competing interest

The authors declare that they have no competing interests.

Authors' contributions

SCO performed experiments related to nuclear isolation and fractionation, generation of transgenic cell lines and IFA analysis, and participated in the design and coordination of the study and in writing the manuscript. BJW performed bioinformatic data analysis, and participated in the design of the study and in writing the manuscript. SM prepared and analysed samples by MudPIT. JW, OD, and AP generated transgenic cell lines and performed IFA analyses. CD performed the electron microscopy experiments. PM participated in data analysis. CF, KW, NMBB and IN performed experiments and participated in the coordination of this study. PJ and SAR participated in the design, coordination and interpretation of the study and critically revised the manuscript. TSV performed experiments related to nuclear isolation and fractionation, conceived of the study, participated in the design, coordination and interpretation of the study and in data analysis and wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgments

The authors would like to thank Terry Speed and James Bailey for helpful comments. We are grateful to Tim-Wolf Gilberger, Mike Duffy, Michael Terns and Geoff McFadden for providing antibodies. BJW is funded by a University of Melbourne MRS scholarship (<http://cms.services.unimelb.edu.au/scholarships/pgrad/local/available/mrs>). SAR is funded by an Australian Research Council Future Fellowship (FT0990350; www.arc.gov.au). NMBB received a Boehringer Ingelheim PhD fellowship (www.bifonds.de). This work was supported by the Swiss National Science Foundation (PP00A-110835; PP00P3_130203; www.snf.ch), the Novartis Foundation for Medicine and Biology (08C46; www.stiftungmedbiol.novartis.com), the Emilia-Guggenheim-Schnurr Foundation, and the Rudolf Geigy Foundation. The funders had no role in study design, data collection, analysis, and interpretation, the decision to publish, or preparation of the manuscript.

References

1. 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.
2. World Health Organisation: World Malaria Report 2010. Geneva: WHO Press; 2010.
3. White NJ: Plasmodium knowlesi: the fifth human malaria parasite. Clin Infect Dis 2008, 46: 172-173.
4. Sutherland CJ, Tanomsing N, Nolder D, Oguike M, Jennison C, Pukrittayakamee S et al.: Two nonrecombining sympatric forms of the human malaria parasite Plasmodium ovale occur globally. J Infect Dis 2010, 201: 1544-1550.
5. Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW et al.: Genome sequence of the human malaria parasite Plasmodium falciparum. Nature 2002, 419: 498-511.
6. Hall N, Karras M, Raine JD, Carlton JM, Kooij TW, Berriman M et al.: A comprehensive survey of the Plasmodium life cycle by genomic, transcriptomic, and proteomic analyses. Science 2005, 307: 82-86.
7. Waters AP: Genome-informed contributions to malaria therapies: feeding somewhere down the (pipe)line. Cell Host Microbe 2008, 3: 280-283.
8. Bannister LH, Hopkins JM, Fowler RE, Krishna S, Mitchell GH: A brief illustrated guide to the ultrastructure of Plasmodium falciparum asexual blood stages. Parasitol Today 2000, 16: 427-433.
9. Striepen B, Jordan CN, Reiff S, van Dooren GG: Building the perfect parasite: cell division in apicomplexa. PLoS Pathog 2007, 3: e78.
10. Aikawa M: The fine structure of the erythrocytic stages of three avian malarial parasites, Plasmodium fallax, P. lophurae, and P. cathemerium. Am J Trop Med Hyg 1966, 15: 449-471.
11. Ralph SA, Scheidig-Benatar C, Scherf A: Antigenic variation in Plasmodium falciparum is associated with movement of var loci between subnuclear locations. Proc Natl Acad Sci U S A 2005, 102: 5414-5419.
12. Weiner A, han-Pasternak N, Shimoni E, Shinder V, von HP, Elbaum M et al.: 3D nuclear architecture reveals coupled cell cycle dynamics of chromatin and nuclear pores in the malaria parasite Plasmodium falciparum. Cell Microbiol 2011, 13: 967-977.
13. Florens L, Washburn MP, Raine JD, Anthony RM, Grainger M, Haynes JD et al.: A proteomic view of the Plasmodium falciparum life cycle. Nature 2002, 419: 520-526.
14. Silvestrini F, Lasonder E, Olivieri A, Camarda G, van SB, Sanchez M et al.: Protein export marks the early phase of gametocytogenesis of the human malaria parasite plasmodium falciparum. Mol Cell Proteomics 2010, 9: 1437-1448.
15. Lasonder E, Janse CJ, van Gemert GJ, Mair GR, Vermunt AM, Douradinha BG et al.: Proteomic profiling of Plasmodium sporozoite maturation identifies new proteins essential for parasite development and infectivity. PLoS Pathog 2008, 4: e1000195.
16. Khan SM, Franke-Fayard B, Mair GR, Lasonder E, Janse CJ, Mann M et al.: Proteome analysis of separated male and female gametocytes reveals novel sex-specific Plasmodium biology. Cell 2005, 121: 675-687.
17. Lasonder E, Ishihama Y, Andersen JS, Vermunt AM, Pain A, Sauerwein RW et al.: Analysis of the Plasmodium falciparum proteome by high-accuracy mass spectrometry. Nature 2002, 419: 537-542.
18. Le Roch KG, Johnson JR, Florens L, Zhou Y, Santosyan A, Grainger M et al.: Global analysis of transcript and protein levels across the Plasmodium falciparum life cycle. Genome Res 2004, 14: 2308-2318.
19. Le Roch KG, Zhou Y, Blair PL, Grainger M, Moch JK, Haynes JD et al.: Discovery of gene function by expression profiling of the malaria parasite life cycle. Science 2003, 301: 1503-1508.
20. Otto TD, Wilinski D, Assefa S, Keane TM, Sarry LR, Bohme U et al.: New insights into the blood-stage transcriptome of Plasmodium falciparum using RNA-Seq. Mol Microbiol 2010, 76: 12-24.
21. Bozdech Z, Llinas M, Pulliam BL, Wong ED, Zhu J, DeRisi JL: The transcriptome of the intraerythrocytic developmental cycle of Plasmodium falciparum. PLoS Biol 2003, 1: E5.
22. Llinas M, Bozdech Z, Wong ED, Adai AT, DeRisi JL: Comparative whole genome transcriptome analysis of three Plasmodium falciparum strains. Nucleic Acids Res 2006, 34: 1166-1173.
23. Coleman BI, Duraisingh MT: Transcriptional control and gene silencing in Plasmodium falciparum. Cell Microbiol 2008, 10: 1935-1946.
24. Aravind L, Iyer LM, Wellem TE, Miller LH: Plasmodium biology: genomic gleanings. Cell 2003, 115: 771-785.
25. Meissner M, Soldati D: The transcription machinery and the molecular toolbox to control gene expression in Toxoplasma gondii and other protozoan parasites. Microbes Infect 2005, 7: 1376-1384.
26. Horrocks P, Wong E, Russell K, Emes RD: Control of gene expression in Plasmodium falciparum - Ten years on. Mol Biochem Parasitol 2009, 164: 9-25.
27. Cui L, Miao J: Chromatin-mediated Epigenetic Regulation in the Malaria Parasite Plasmodium falciparum. Eukaryot Cell 2010, 9: 1138-1149.
28. Volz J, Carvalho TG, Ralph SA, Gilson P, Thompson J, Tonkin CJ et al.: Potential epigenetic regulatory proteins localise to distinct nuclear sub-compartments in Plasmodium falciparum. Int J Parasitol 2010, 40: 109-121.

29. Lopez-Rubio JJ, Mancio-Silva L, Scherf A: Genome-wide analysis of heterochromatin associates clonally variant gene regulation with perinuclear repressive centers in malaria parasites. *Cell Host Microbe* 2009, 5: 179-190.
30. Tonkin CJ, Carret CK, Duraisingh MT, Voss TS, Ralph SA, Hommel M et al.: Sir2 paralogs cooperate to regulate virulence genes and antigenic variation in *Plasmodium falciparum*. *PLoS Biol* 2009, 7: e84.
31. Perez-Toledo K, Rojas-Meza AP, Mancio-Silva L, Hernandez-Cuevas NA, Delgadillo DM, Vargas M et al.: *Plasmodium falciparum* heterochromatin protein 1 binds to tri-methylated histone 3 lysine 9 and is linked to mutually exclusive expression of var genes. *Nucleic Acids Res* 2009, 37: 2596-2606.
32. Freitas-Junior LH, Hernandez-Rivas R, Ralph SA, Montiel-Condado D, Ruvalcaba-Salazar OK, Rojas-Meza AP et al.: Telomeric heterochromatin propagation and histone acetylation control mutually exclusive expression of antigenic variation genes in malaria parasites. *Cell* 2005, 121: 25-36.
33. Duraisingh MT, Voss TS, Marty AJ, Duffy MF, Good RT, Thompson JK et al.: Heterochromatin silencing and locus repositioning linked to regulation of virulence genes in *Plasmodium falciparum*. *Cell* 2005, 121: 13-24.
34. Salcedo-Amaya AM, van Driel MA, Alako BT, Trelle MB, van den Elzen AM, Cohen AM et al.: Dynamic histone H3 epigenome marking during the intraerythrocytic cycle of *Plasmodium falciparum*. *Proc Natl Acad Sci U S A* 2009, 106: 9655-9660.
35. Flueck C, Bartfai R, Volz J, Niederwieser I, Salcedo-Amaya AM, Alako BT et al.: *Plasmodium falciparum* heterochromatin protein 1 marks genomic loci linked to phenotypic variation of exported virulence factors. *PLoS Pathog* 2009, 5: e1000569.
36. Cui L, Miao J, Furuya T, Fan Q, Li X, Rathod PK et al.: Histone acetyltransferase inhibitor anacardic acid causes changes in global gene expression during in vitro *Plasmodium falciparum* development. *Eukaryot Cell* 2008, 7: 1200-1210.
37. Cui L, Miao J, Furuya T, Li X, Su XZ, Cui L: PfGCN5-mediated histone H3 acetylation plays a key role in gene expression in *Plasmodium falciparum*. *Eukaryot Cell* 2007, 6: 1219-1227.
38. Coulson RM, Hall N, Ouzounis CA: Comparative genomics of transcriptional control in the human malaria parasite *Plasmodium falciparum*. *Genome Res* 2004, 14: 1548-1554.
39. Gissot M, Briquet S, Refour P, Boschet C, Vaquero C: PfMyb1, a *Plasmodium falciparum* transcription factor, is required for intra-erythrocytic growth and controls key genes for cell cycle regulation. *J Mol Biol* 2005, 346: 29-42.
40. Balaji S, Babu MM, Iyer LM, Aravind L: Discovery of the principal specific transcription factors of Apicomplexa and their implication for the evolution of the AP2-integrase DNA binding domains. *Nucleic Acids Res* 2005, 33: 3994-4006.
41. Yuda M, Iwanaga S, Shigenobu S, Mair GR, Janse CJ, Waters AP et al.: Identification of a transcription factor in the mosquito-invasive stage of malaria parasites. *Mol Microbiol* 2009, 71: 1402-1414.
42. Yuda M, Iwanaga S, Shigenobu S, Kato T, Kaneko I: Transcription factor AP2-Sp and its target genes in malarial sporozoites. *Mol Microbiol* 2010, 75: 854-863.
43. Lindner SE, De Silva EK, Keck JL, Llinas M: Structural determinants of DNA binding by a *P. falciparum* ApiAP2 transcriptional regulator. *J Mol Biol* 2010, 395: 558-567.
44. Campbell TL, De Silva EK, Olszewski KL, Elemento O, Llinas M: Identification and genome-wide prediction of DNA binding specificities for the ApiAP2 family of regulators from the malaria parasite. *PLoS Pathog* 2010, 6: e1001165.
45. De Silva EK, Gehrke AR, Olszewski K, Leon I, Chahal JS, Bulyk ML et al.: Specific DNA-binding by apicomplexan AP2 transcription factors. *Proc Natl Acad Sci U S A* 2008, 105: 8393-8398.
46. Flueck C, Bartfai R, Niederwieser I, Witmer K, Alako BT, Moes S et al.: A major role for the *Plasmodium falciparum* ApiAP2 protein PfSIP2 in chromosome end biology. *PLoS Pathog* 2010, 6: e1000784.
47. Roobsoong W, Roytrakul S, Sattabongkot J, Li J, Udomsangpetch R, Cui L: Determination of the *Plasmodium vivax* schizont stage proteome. *J Proteomics* 2011, 74: 1701-1710.
48. Radfar A, Diez A, Bautista JM: Chloroquine mediates specific proteome oxidative damage across the erythrocytic cycle of resistant *Plasmodium falciparum*. *Free Radic Biol Med* 2008, 44: 2034-2042.
49. van Brummelen AC, Olszewski KL, Wilinski D, Llinas M, Louw AI, Birkholtz LM: Co-inhibition of *Plasmodium falciparum* S-Adenosylmethionine Decarboxylase/Ornithine Decarboxylase Reveals Perturbation-specific Compensatory Mechanisms by Transcriptome, Proteome, and Metabolome Analyses. *J Biol Chem* 2009, 284: 4635-4646.
50. Aly NS, Hiramoto A, Sanai H, Hiraoka O, Hiramoto K, Kataoka H et al.: Proteome analysis of new antimalarial endoperoxide against *Plasmodium falciparum*. *Parasitol Res* 2007, 100: 1119-1124.
51. Briolant S, Almeras L, Belghazi M, Boucomont-Chapeaublanc E, Wurtz N, Fontaine A et al.: *Plasmodium falciparum* proteome changes in response to doxycycline treatment. *Malar J* 2010, 9: 141.
52. Lamarque M, Tastet C, Poncet J, Demette E, Jouin P, Vial H et al.: Food vacuole proteome of the malarial parasite *Plasmodium falciparum*. *Proteomics Clin Appl* 2008, 2: 1361-1374.
53. Vincensini L, Richert S, Blisnick T, Van DA, Leize-Wagner E, Rabilloud T et al.: Proteomic analysis identifies novel proteins of the Maurer's clefts, a secretory compartment delivering *Plasmodium falciparum* proteins to the surface of its host cell. *Mol Cell Proteomics* 2005, 4: 582-593.

54. Florens L, Liu X, Wang Y, Yang S, Schwartz O, Peglar M et al.: Proteomics approach reveals novel proteins on the surface of malaria-infected erythrocytes. *Mol Biochem Parasitol* 2004, 135: 1-11.
55. Sam-Yellowe TY, Florens L, Wang T, Raine JD, Carucci DJ, Sinden R et al.: Proteome analysis of rhoptry-enriched fractions isolated from *Plasmodium* merozoites. *J Proteome Res* 2004, 3: 995-1001.
56. Lal K, Prieto JH, Bromley E, Sanderson SJ, Yates JR, III, Wastling JM et al.: Characterisation of *Plasmodium* invasive organelles; an ookinete microneme proteome. *Proteomics* 2009, 9: 1142-1151.
57. Cohn ZA, Hirsch JG: The isolation and properties of the specific cytoplasmic granules of rabbit polymorphonuclear leucocytes. *J Exp Med* 1960, 112: 983-1004.
58. Crawley JC, Harris H: The Fine Structure of Isolated HeLa Cell Nuclei. *Exp Cell Res* 1963, 31: 70-81.
59. Gatlin CL, Eng JK, Cross ST, Detter JC, Yates JR, III: Automated identification of amino acid sequence variations in proteins by HPLC/microspray tandem mass spectrometry. *Anal Chem* 2000, 72: 757-763.
60. Huang da W, Sherman BT, Lempicki RA: Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009, 4: 44-57.
61. Beissbarth T, Speed TP: GOstat: find statistically overrepresented Gene Ontologies within a group of genes. *Bioinformatics* 2004, 20: 1464-1465.
62. Bischoff E, Vaquero C: In silico and biological survey of transcription-associated proteins implicated in the transcriptional machinery during the erythrocytic development of *Plasmodium falciparum*. *BMC Genomics* 2010, 15: 34.
63. Marti M, Good RT, Rug M, Knuepfer E, Cowman AF: Targeting malaria virulence and remodeling proteins to the host erythrocyte. *Science* 2004, 306: 1930-1933.
64. Hiller NL, Bhattacharjee S, van OC, Liolios K, Harrison T, Lopez-Estrano C et al.: A host-targeting signal in virulence proteins reveals a secretome in malarial infection. *Science* 2004, 306: 1934-1937.
65. Bender A, van Dooren GG, Ralph SA, McFadden GI, Schneider G: Properties and prediction of mitochondrial transit peptides from *Plasmodium falciparum*. *Mol Biochem Parasitol* 2003, 132: 59-66.
66. Finn RD, Mistry J, Tate J, Coghill P, Heger A, Pollington JE et al.: The Pfam protein families database. *Nucleic Acids Res* 2010, 38: D211-D222.
67. Wiley SE, Murphy AN, Ross SA, van der GP, Dixon JE: MitoNEET is an iron-containing outer mitochondrial membrane protein that regulates oxidative capacity. *Proc Natl Acad Sci U S A* 2007, 104: 5318-5323.
68. Yang JM, Baserga SJ, Turley SJ, Pollard KM: Fibrillarin and other snoRNP proteins are targets of autoantibodies in xenobiotic-induced autoimmunity. *Clin Immunol* 2001, 101: 38-50.
69. Gubbels MJ, Wieffer M, Striepen B: Fluorescent protein tagging in *Toxoplasma gondii*: identification of a novel inner membrane complex component conserved among Apicomplexa. *Mol Biochem Parasitol* 2004, 137: 99-110.
70. Altschul SF, Wootton JC, Zaslavsky E, Yu YK: The construction and use of log-odds substitution scores for multiple sequence alignment. *PLoS Comput Biol* 2010, 6: e1000852.
71. Iyer LM, Anantharaman V, Wolf MY, Aravind L: Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes. *Int J Parasitol* 2008, 38: 1-31.
72. Peleg O, Lim RY: Converging on the function of intrinsically disordered nucleoporins in the nuclear pore complex. *Biol Chem* 2010, 391: 719-730.
73. Nguyen Ba AN, Pogoutse A, Provart N, Moses AM: NLStradamus: a simple Hidden Markov Model for nuclear localization signal prediction. *BMC Bioinformatics* 2009, 10: 202.
74. Cokol M, Nair R, Rost B: Finding nuclear localization signals. *EMBO Rep* 2000, 1: 411-415.
75. Kosugi S, Hasebe M, Tomita M, Yanagawa H: Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. *Proc Natl Acad Sci U S A* 2009, 106: 10171-10176.
76. Mosley AL, Pattenden SG, Carey M, Venkatesh S, Gilmore JM, Florens L et al.: Rtr1 is a CTD phosphatase that regulates RNA polymerase II during the transition from serine 5 to serine 2 phosphorylation. *Mol Cell* 2009, 34: 168-178.
77. Gauci S, Veenhoff LM, Heck AJ, Krijgsveld J: Orthogonal separation techniques for the characterization of the yeast nuclear proteome. *J Proteome Res* 2009, 8: 3451-3463.
78. Chandra BR, Olivieri A, Silvestrini F, Alano P, Sharma A: Biochemical characterization of the two nucleosome assembly proteins from *Plasmodium falciparum*. *Mol Biochem Parasitol* 2005, 142: 237-247.
79. Calvert ME, Keck KM, Ptak C, Shabanowitz J, Hunt DF, Pemberton LF: Phosphorylation by casein kinase 2 regulates Nap1 localization and function. *Mol Cell Biol* 2008, 28: 1313-1325.
80. Mancio-Silva L, Zhang Q, Scheidig-Benatar C, Scherf A: Clustering of dispersed ribosomal DNA and its role in gene regulation and chromosome-end associations in malaria parasites. *Proc Natl Acad Sci U S A* 2010, 107: 15117-15122.
81. Mancio-Silva L, Rojas-Meza AP, Vargas M, Scherf A, Hernandez-Rivas R: Differential association of Orc1 and Sir2 proteins to telomeric domains in *Plasmodium falciparum*. *J Cell Sci* 2008, 121: 2046-2053.
82. Doerig C, Meijer L: Antimalarial drug discovery: targeting protein kinases. *Expert Opin Ther Targets* 2007, 11: 279-290.

83. Jirage D, Keenan SM, Waters NC: Exploring novel targets for antimalarial drug discovery: plasmodial protein kinases. *Infect Disord Drug Targets* 2010, 10: 134-146.
84. Jirage D, Chen Y, Caridha D, O'Neil MT, Eyase F, Witola WH et al.: The malarial CDK Pfmrk and its effector PfMAT1 phosphorylate DNA replication proteins and co-localize in the nucleus. *Mol Biochem Parasitol* 2010, 172: 9-18.
85. Dastidar EG, Dayer G, Holland ZM, Dorin-Semblat D, Claes A, Chene A et al.: Involvement of Plasmodium falciparum protein kinase CK2 in the chromatin assembly pathway. *BMC Biol* 2012, 10: 5.
86. Holland Z, Prudent R, Reiser JB, Cochet C, Doerig C: Functional analysis of protein kinase CK2 of the human malaria parasite Plasmodium falciparum. *Eukaryot Cell* 2009, 8: 388-397.
87. Filhol O, Cochet C: Protein kinase CK2 in health and disease: Cellular functions of protein kinase CK2: a dynamic affair. *Cell Mol Life Sci* 2009, 66: 1830-1839.
88. Milne DM, Looby P, Meek DW: Catalytic activity of protein kinase CK1 delta (casein kinase 1delta) is essential for its normal subcellular localization. *Exp Cell Res* 2001, 263: 43-54.
89. Cheong JK, Virshup DM: Casein kinase 1: Complexity in the family. *Int J Biochem Cell Biol* 2011, 43: 465-469.
90. Tudisca V, Recouvreur V, Moreno S, Boy-Marcotte E, Jacquet M, Portela P: Differential localization to cytoplasm, nucleus or P-bodies of yeast PKA subunits under different growth conditions. *Eur J Cell Biol* 2010, 89: 339-348.
91. Dorin-Semblat D, Schmitt S, Semblat JP, Sicard A, Reininger L, Goldring D et al.: Plasmodium falciparum NIMA-related kinase Pfnk-1: sex-specificity and assessment of essentiality for the erythrocytic asexual cycle. *Microbiology* 2011, epub ahead of print.
92. Dorin D, Le RK, Sallicandro P, Alano P, Parzy D, Pouillet P et al.: Pfnk-1, a NIMA-related kinase from the human malaria parasite Plasmodium falciparum Biochemical properties and possible involvement in MAPK regulation. *Eur J Biochem* 2001, 268: 2600-2608.
93. Rangarajan R, Bei AK, Jethwaney D, Maldonado P, Dorin D, Sultan AA et al.: A mitogen-activated protein kinase regulates male gametogenesis and transmission of the malaria parasite Plasmodium berghei. *EMBO Rep* 2005, 6: 464-469.
94. Dorin D, Alano P, Boccaccio I, Ciceron L, Doerig C, Sulpice R et al.: An atypical mitogen-activated protein kinase (MAPK) homologue expressed in gametocytes of the human malaria parasite Plasmodium falciparum. Identification of a MAPK signature. *J Biol Chem* 1999, 274: 29912-29920.
95. Dorin-Semblat D, Quashie N, Halbert J, Sicard A, Doerig C, Peat E et al.: Functional characterization of both MAP kinases of the human malaria parasite Plasmodium falciparum by reverse genetics. *Mol Microbiol* 2007, 65: 1170-1180.
96. Treeck M, Sanders JL, Elias JE, Boothroyd JC: The phosphoproteomes of Plasmodium falciparum and Toxoplasma gondii reveal unusual adaptations within and beyond the parasites' boundaries. *Cell Host Microbe* 2011, 10: 410-419.
97. Collins GA, Tansey WP: The proteasome: a utility tool for transcription? *Curr Opin Genet Dev* 2006, 16: 197-202.
98. Lee D, Ezhkova E, Li B, Pattenden SG, Tansey WP, Workman JL: The proteasome regulatory particle alters the SAGA coactivator to enhance its interactions with transcriptional activators. *Cell* 2005, 123: 423-436.
99. Ferdous A, Gonzalez F, Sun L, Kodadek T, Johnston SA: The 19S regulatory particle of the proteasome is required for efficient transcription elongation by RNA polymerase II. *Mol Cell* 2001, 7: 981-991.
100. Ezhkova E, Tansey WP: Proteasomal ATPases link ubiquitylation of histone H2B to methylation of histone H3. *Mol Cell* 2004, 13: 435-442.
101. Mair GR, Lasonder E, Garver LS, Franke-Fayard BM, Carret CK, Wiegant JC et al.: Universal features of post-transcriptional gene regulation are critical for Plasmodium zygote development. *PLoS Pathog* 2010, 6: e1000767.
102. Mair GR, Braks JA, Garver LS, Wiegant JC, Hall N, Dirks RW et al.: Regulation of sexual development of Plasmodium by translational repression. *Science* 2006, 313: 667-669.
103. Brune C, Munchel SE, Fischer N, Podtelejnikov AV, Weis K: Yeast poly(A)-binding protein Pab1 shuttles between the nucleus and the cytoplasm and functions in mRNA export. *RNA* 2005, 11: 517-531.
104. MacNicol AM, Wilczynska A, MacNicol MC: Function and regulation of the mammalian Musashi mRNA translational regulator. *Biochem Soc Trans* 2008, 36: 528-530.
105. Neumann N, Lundin D, Poole AM: Comparative genomic evidence for a complete nuclear pore complex in the last eukaryotic common ancestor. *PLoS One* 2010, 5: e13241.
106. Richard D, Bartfai R, Volz J, Ralph SA, Muller S, Stunnenberg HG et al.: A genome-wide chromatin-associated nuclear peroxiredoxin from the malaria parasite Plasmodium falciparum. *J Biol Chem* 2011, 286: 11746-11755.
107. Briquet S, Boschet C, Gissot M, Tissandie E, Sevilla E, Franetich JF et al.: High-mobility-group box nuclear factors of Plasmodium falciparum. *Eukaryot Cell* 2006, 5: 672-682.
108. Albrethsen J, Knol JC, Jimenez CR: Unravelling the nuclear matrix proteome. *J Proteomics* 2009, 72: 71-81.

109. Boisvert FM, van KS, Navascues J, Lamond AI: The multifunctional nucleolus. *Nat Rev Mol Cell Biol* 2007, 8: 574-585.
110. Andersen JS, Lam YW, Leung AK, Ong SE, Lyon CE, Lamond AI et al.: Nucleolar proteome dynamics. *Nature* 2005, 433: 77-83.
111. Shaw P, Doonan J: The nucleolus. Playing by different rules? *Cell Cycle* 2005, 4: 102-105.
112. Iborra FJ, Jackson DA, Cook PR: Coupled transcription and translation within nuclei of mammalian cells. *Science* 2001, 293: 1139-1142.
113. Iborra FJ, Jackson DA, Cook PR: The case for nuclear translation. *J Cell Sci* 2004, 117: 5713-5720.
114. Frankel MB, Knoll LJ: The ins and outs of nuclear trafficking: unusual aspects in apicomplexan parasites. *DNA Cell Biol* 2009, 28: 277-284.
115. Chook YM, Suel KE: Nuclear import by karyopherin-betas: recognition and inhibition. *Biochim Biophys Acta* 2011, 1813: 1593-1606.
116. Marfori M, Mynott A, Ellis JJ, Mehdi AM, Saunders NF, Curmi PM et al.: Molecular basis for specificity of nuclear import and prediction of nuclear localization. *Biochim Biophys Acta* 2011, 1813: 1562-1577.
117. Camarda G, Bertuccini L, Singh SK, Salzano AM, Lanfrancotti A, Olivieri A et al.: Regulated oligomerisation and molecular interactions of the early gametocyte protein Pfg27 in *Plasmodium falciparum* sexual differentiation. *Int J Parasitol* 2010, 40: 663-673.
118. Sharma A, Sharma I, Kogkasuriyachai D, Kumar N: Structure of a gametocyte protein essential for sexual development in *Plasmodium falciparum*. *Nat Struct Biol* 2003, 10: 197-203.
119. Szallies A, Kubata BK, Duzenko M: A metacaspase of *Trypanosoma brucei* causes loss of respiration competence and clonal death in the yeast *Saccharomyces cerevisiae*. *FEBS Lett* 2002, 517: 144-150.
120. Cortes A, Carret C, Kaneko O, Yim Lim BY, Ivens A, Holder AA: Epigenetic silencing of *Plasmodium falciparum* genes linked to erythrocyte invasion. *PLoS Pathog* 2007, 3: e107.
121. Trager W, Jensen JB: Cultivation of malarial parasites. *Nature* 1978, 273: 621-622.
122. Lambros C, Vanderberg JP: Synchronization of *Plasmodium falciparum* erythrocytic stages in culture. *J Parasitol* 1979, 65: 418-420.
123. Voss TS, Healer J, Marty AJ, Duffy MF, Thompson JK, Beeson JG et al.: A var gene promoter controls allelic exclusion of virulence genes in *Plasmodium falciparum* malaria. *Nature* 2006, 439: 1004-1008.
124. Daubenberger CA, Tisdale EJ, Curcic M, Diaz D, Silvie O, Mazier D et al.: The N'-terminal domain of glyceraldehyde-3-phosphate dehydrogenase of the apicomplexan *Plasmodium falciparum* mediates GTPase Rab2-dependent recruitment to membranes. *Biol Chem* 2003, 384: 1227-1237.
125. Riglar DT, Richard D, Wilson DW, Boyle MJ, Dekiwadia C, Turnbull L et al.: Super-resolution dissection of coordinated events during malaria parasite invasion of the human erythrocyte. *Cell Host Microbe* 2011, 9: 9-20.
126. Tonkin CJ, van Dooren GG, Spurck TP, Struck NS, Good RT, Handman E et al.: Localization of organelle proteins in *Plasmodium falciparum* using a novel set of transfection vectors and a new immunofluorescence fixation method. *Mol Biochem Parasitol* 2004, 137: 13-21.
127. Kumar N, Koski G, Harada M, Aikawa M, Zheng H: Induction and localization of *Plasmodium falciparum* stress proteins related to the heat shock protein 70 family. *Mol Biochem Parasitol* 1991, 48: 47-58.
128. Struck NS, Herrmann S, Schmuck-Barkmann I, de Souza DS, Haase S, Cabrera AL et al.: Spatial dissection of the cis- and trans-Golgi compartments in the malaria parasite *Plasmodium falciparum*. *Mol Microbiol* 2008, 67: 1320-1330.
129. Fink JL, Hamilton N: DomainDraw: a macromolecular feature drawing program. *In Silico Biol* 2007, 7: 145-150.

Figure legends

Figure 1 Preparation of crude *P. falciparum* nuclei for proteomic analysis. (a)

Visual assessment of the crude nuclear preparation. Intact parasites (control) and isolated nuclei were stained with DAPI and analysed by DIC and fluorescence microscopy. Scale bar 2µm. (b) TEM analysis of isolated nuclei shows intact as well as damaged nuclei and very little contamination from other organelles. Arrows indicate individual nuclei. (c) Immunoelectron microscopy verifies the identity of nuclei. Anti-H3K4me3 antibodies were used to label nuclei in intact cells (left panel) and in the

nuclear preparation (right panel) (arrows). **(d)** 1D-SDS PAGE analysis of all protein fractions obtained from trophozoites, visualised by silver staining. **(e)** Analysis of trophozoite cytoplasmic and nuclear fractions by Western blot.

Figure 2 Distribution of the 1518 detected proteins in developmental stages, subcellular compartments and nuclear fractions. Detection of proteins in the various pools is depicted by Venn diagrams. Absolute numbers of proteins in each pool are indicated in brackets. **(a)** Protein distribution in ring stages, trophozoites and schizonts. **(b)** Protein distribution in the cytoplasmic versus the combined nuclear fractions. **(c)** Protein distribution across the four nuclear fractions (DNaseI and high salt fractions were combined).

Figure 3 Enrichment of nuclear proteins in nuclear fractions by comparison to gold standards. **(a)** Enrichment was defined as the percentage of true nuclear proteins found in each fraction, divided by the number of nuclear proteins expected in a random set of equivalent size. The horizontal line indicates the theoretical enrichment in a random selection of proteins. **(b)** Nuclear protein content in various protein fractions. The vertical dashed line indicates the PPV for proteins taken at random from the entire parasite proteome. 1-5, fractions 1 to 5; 2-5, all proteins found in any of the four nuclear fractions; 1 excl., proteins found exclusively in fraction 1; 2-5 excl., all proteins found exclusively in any of the four nuclear fractions.

Figure 4 Bioinformatic filtering of the nuclear proteomics data. **(a)** Post-processing enrichments in the set of proteins found in fractions 2 to 5. **(b)** Nuclear protein content in fractions 2 to 5 after bioinformatic filtering. exported, removal of proteins predicted to be exported; organellar proteomics, removal of proteins detected in the Maurer's cleft (MC) [53] and/or the food vacuole (FV) proteome [52]; SP, removal of proteins carrying a predicted SP; TM, removal of proteins carrying predicted TM domains; combined, the five accepted filters (SP, TM, FV, MC, EXP) combined; mitochondrial transit peptides, removal of proteins predicted by PlasMit [65]; at least two peptides, removal of proteins detected by a single peptide only; fractions 2-5 excl., removal of proteins detected in the cytoplasmic fraction. **(c)** Classification of the core nuclear proteome into broad generic functional classes was done by manual inspection of the dataset and is based on

functional annotation available on PlasmoDB (www.plasmodb.org) and literature review. Each protein was assigned to a single class only.

Figure 5 Experimental validation of the core nuclear proteome by IFA - nuclear localisation. All panels show protein localisation in trophozoites. Protein acronym and PlasmoDB gene accession numbers are indicated to the left. Peptide detection or absence in individual fractions is indicated by orange or white boxes, respectively. Epitope-tagged candidates were detected using anti-HA or anti-Ty antibodies (red). The reference compartments cytosol and ER were visualised using antibodies against GAPDH and PfBIP, respectively (green). White frames refer to the magnified view of the nuclear area shown in the rightmost images. **(a)** 16 NuProCs co-localise exclusively with the DAPI-stained area of the parasite nucleus. **(b)** NuProC4 and NuProC13 are associated with the nuclear periphery.

Figure 6 Experimental validation of the core nuclear proteome by IFA – non-nuclear localisation. All panels show protein localisation in trophozoites. Protein acronym and PlasmoDB gene accession numbers are indicated to the left. Peptide detection or absence in individual fractions is indicated by orange or white boxes, respectively. Epitope-tagged candidates were detected using anti-HA antibodies (red). The reference compartments cytosol and ER were visualised using antibodies against GAPDH and PfBIP, respectively (green). White frames refer to the magnified view of the nuclear area shown in the rightmost images. **(a)** Four NuProCs do not co-localise exclusively with the DAPI-stained area of the parasite nucleus. **(b)** None of the six non-nuclear protein candidates (n-NuProCs) localise to the parasite nucleus. N-NuProCs 1, 2, 4, 5 and 6 co-localise with the ER marker PfBIP.

Figure 7 Identification of two novel *P. falciparum* nucleolar proteins. **(a)** IFA co-localisation of NuProC2-3xHA and NuProC3-2xTy in a double transgenic line. Epitope-tagged proteins were detected using anti-HA (red) and anti-Ty (green) antibodies simultaneously. **(b)** NuProC2-3xHA and NuProC3-3xHA co-localise with nucleolar fibrillarin. HA-tagged proteins were detected in single transgenic lines using anti-HA antibodies (red). Fibrillarin (PfFIB) was visualised using an anti-human fibrillarin antibody [68,69].

Figure 8 Schematic of the novel ACDC domain in *P. falciparum* proteins. (a) The novel ACDC domain was found in nine *P. falciparum* ApiAP2 proteins. The figure was generated with DomainDraw [129] and modified with Inkscape and Adobe Illustrator. **(b)** Multiple sequence alignment of the nine *P. falciparum* ACDC domains is shown. The vertical gap signifies a short, less conserved region omitted for clarity.

Figure 9 Identification of a novel nuclear pore component in *P. falciparum*. PF14_0442 represents a novel *P. falciparum* nuclear pore candidate. Endogenous GFP-tagged PF14_0442 and PfBIP were visualised in 3D7/pH_0442-GFPint using anti-GFP (green) and anti-PfBIP (red) antibodies, respectively. The white frame refers to the magnified view shown in the rightmost image.

Figure 1_Oehring et al.

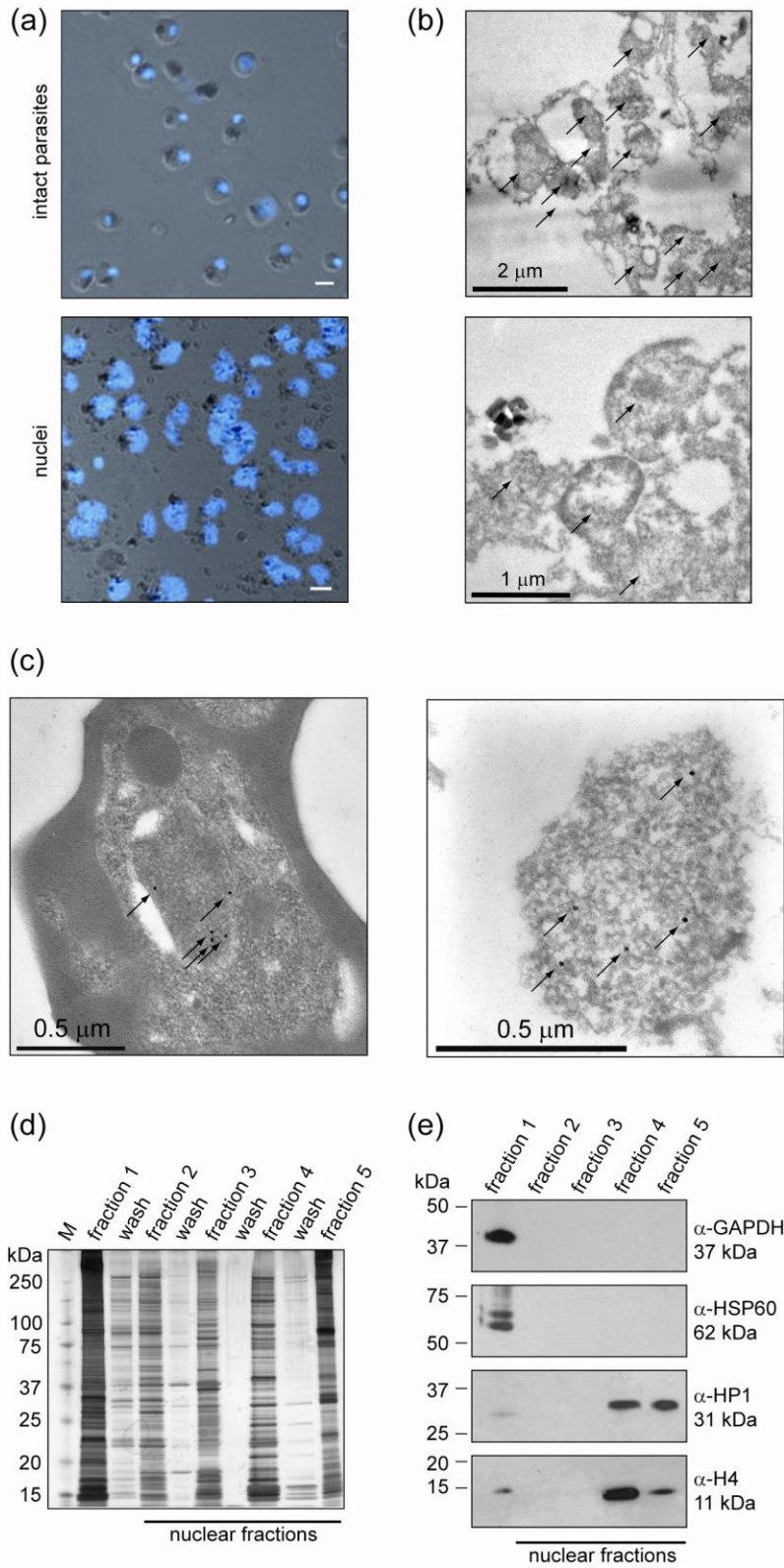


Figure 2_Oehring et al.

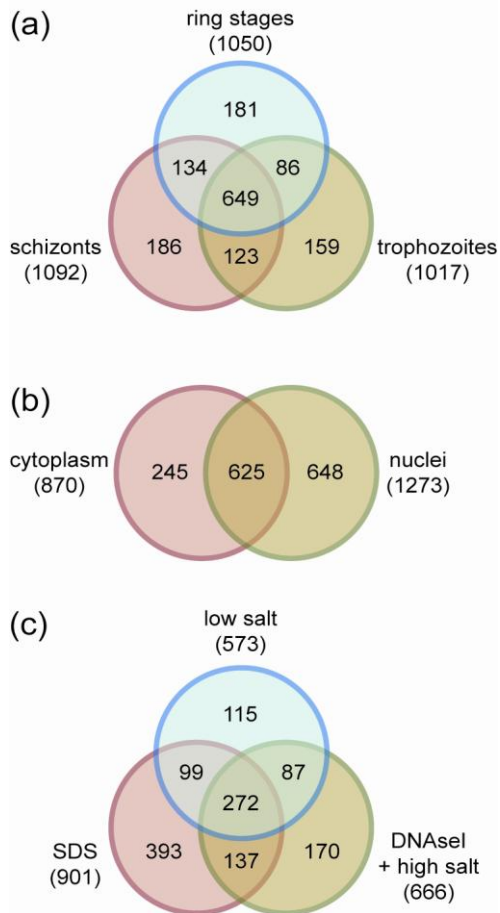


Figure 3_Oehring et al.

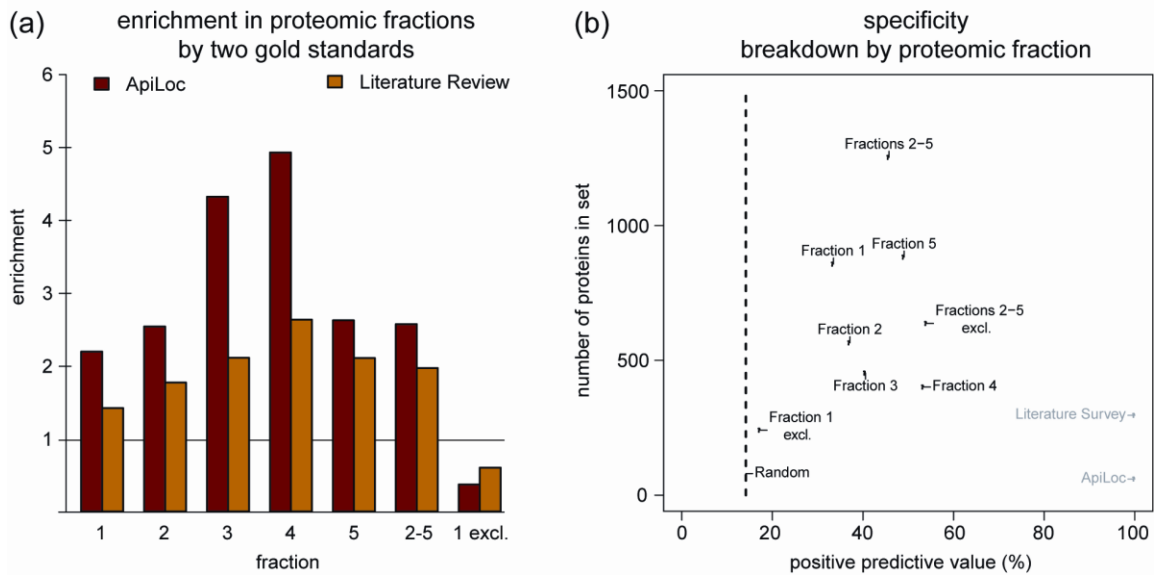


Figure 4_Oehring et al.

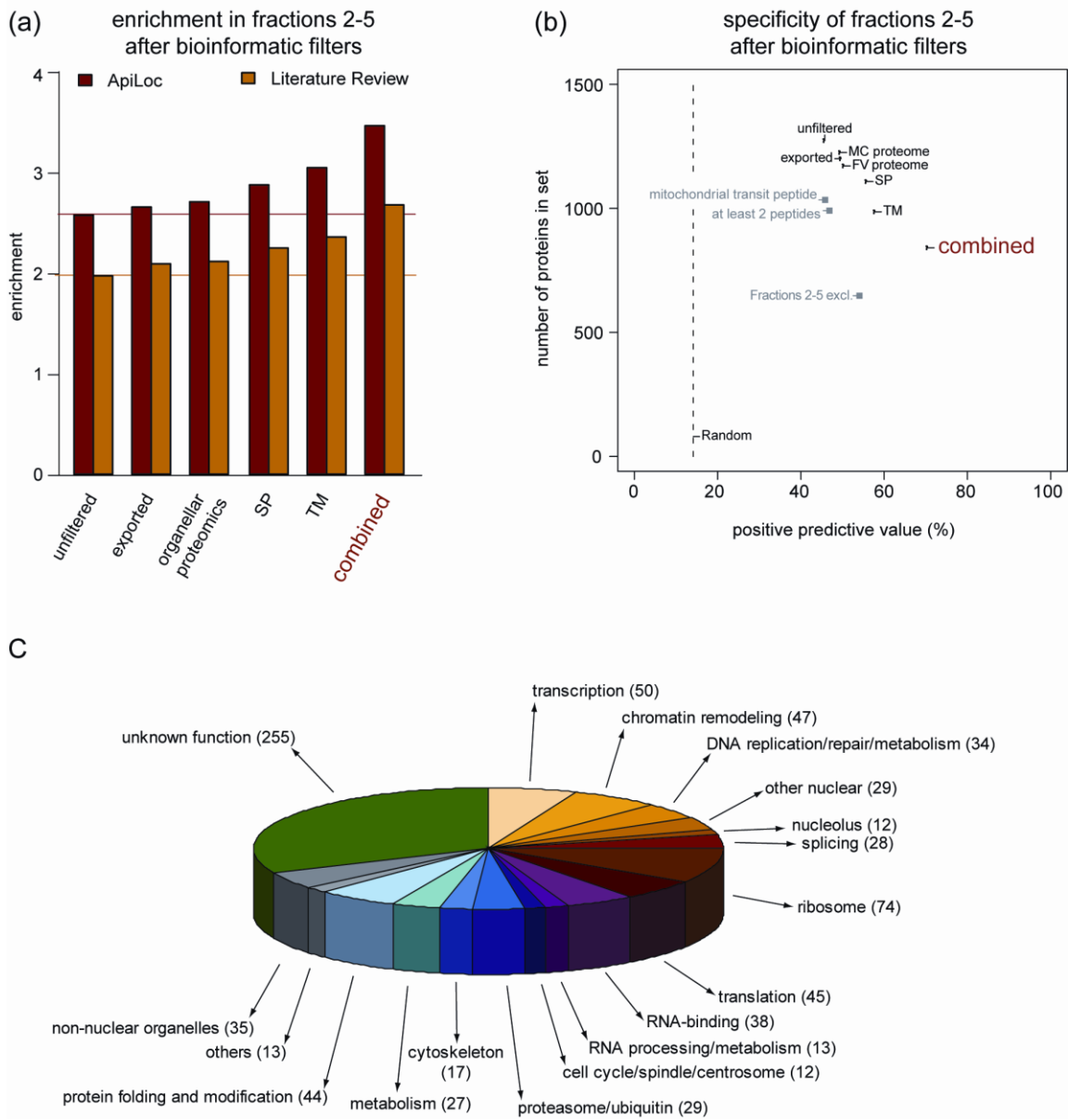


Figure 5_Oehring et al.

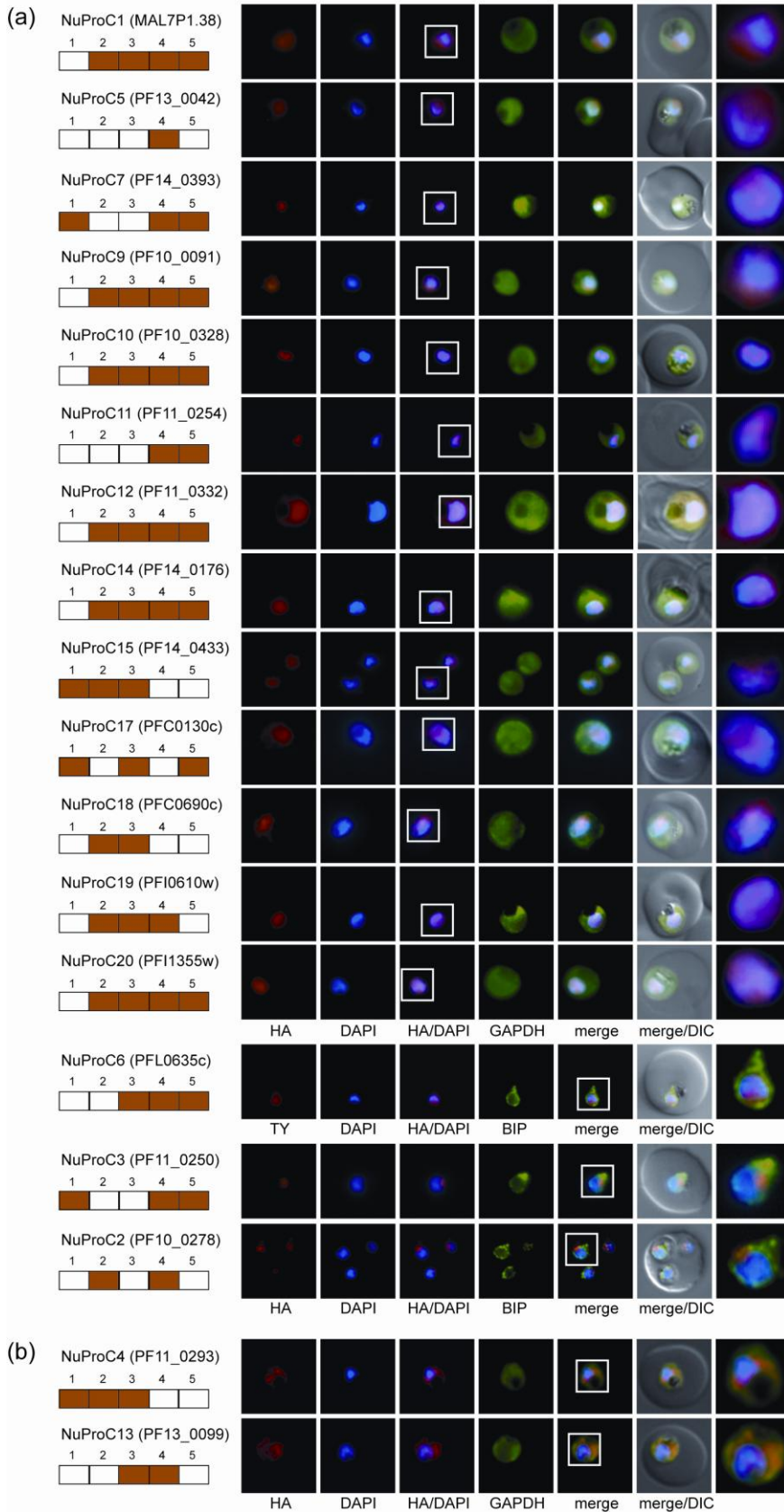


Figure 6_Oehring et al.

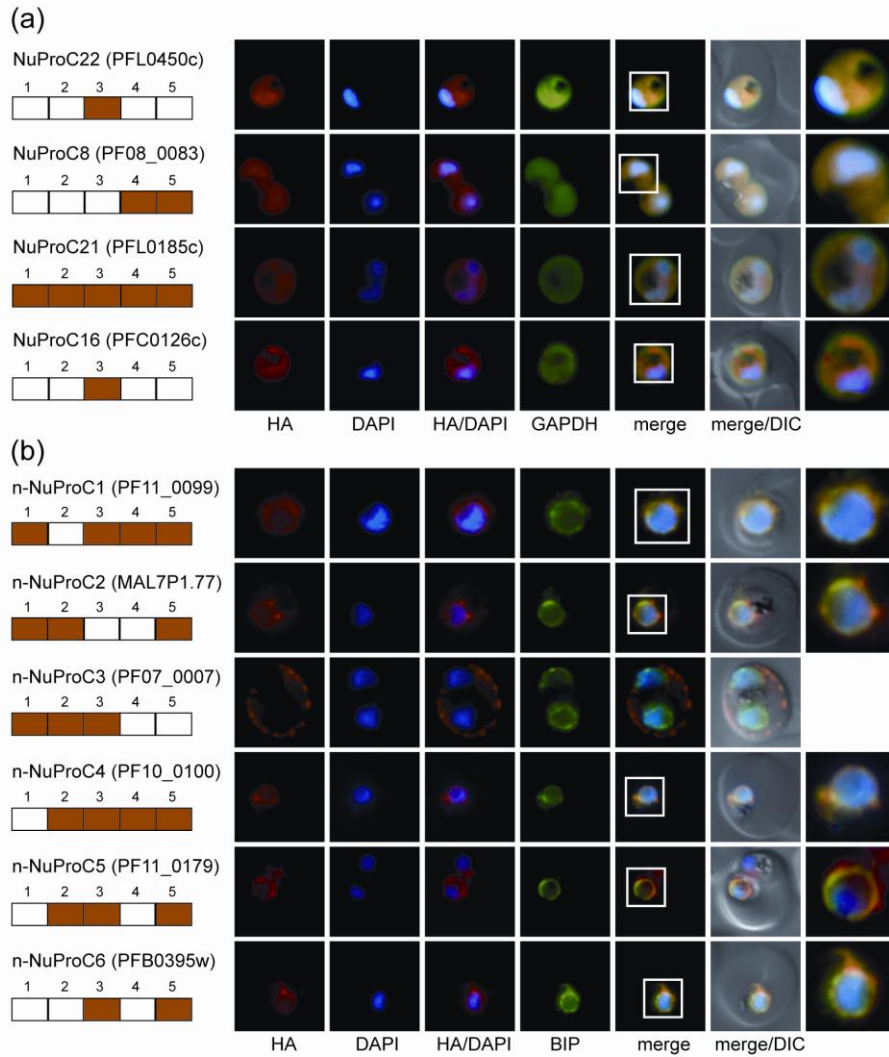


Figure 7_Oehring et al.

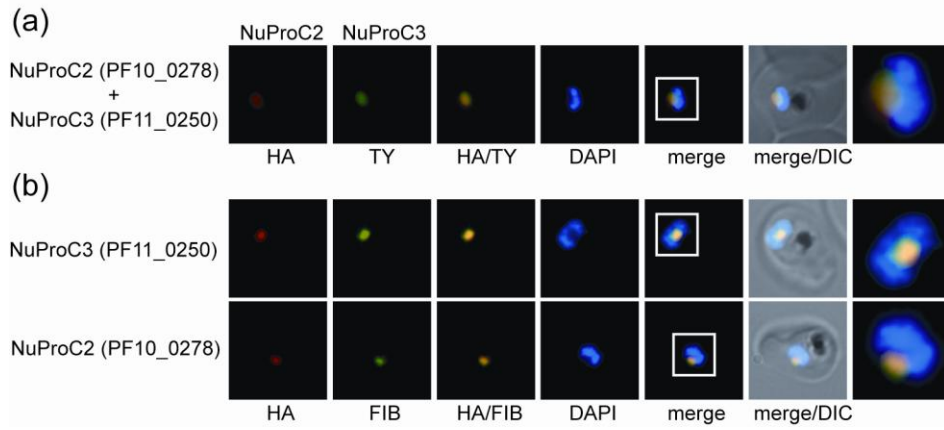


Figure 8_Oehring et al.

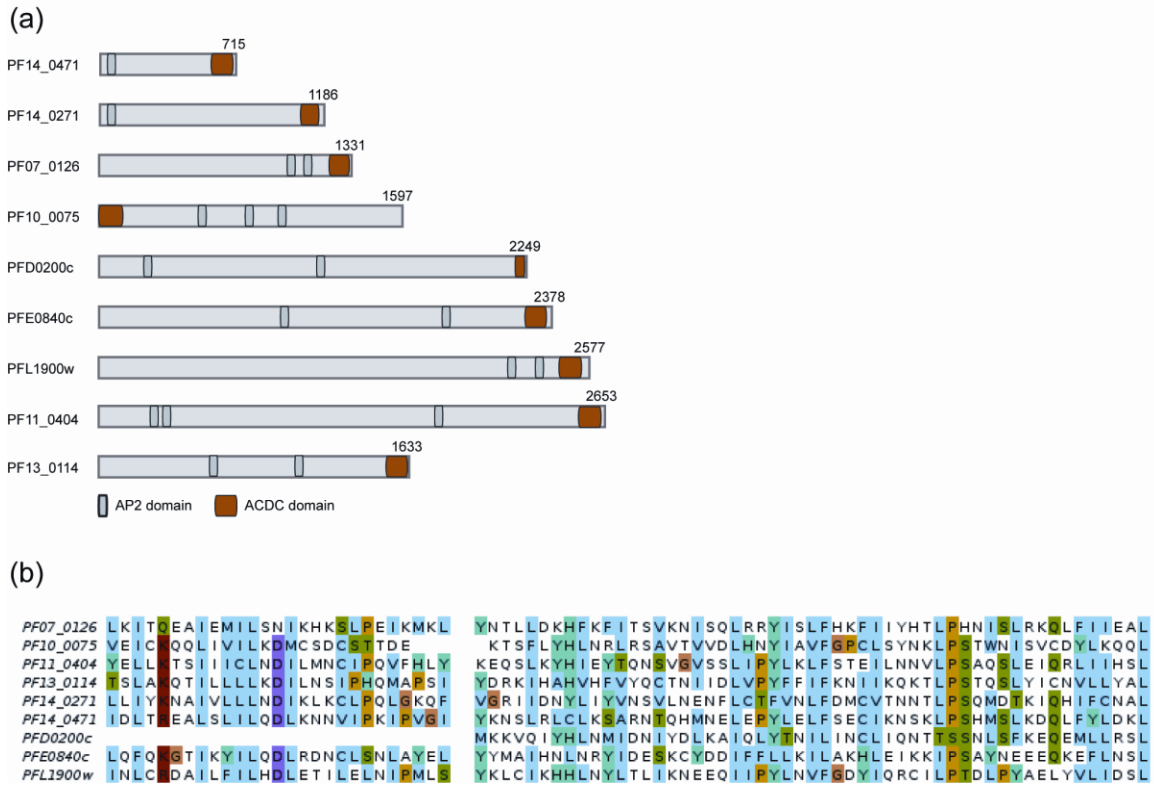


Figure 9_Oehring et al.

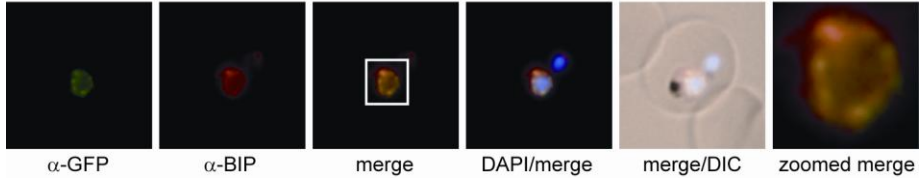


Table 1 Distribution of peptides derived from a representative selection of known nuclear proteins across the three IDC stages.

Indicated are the numbers of unique tryptic peptides measured for each protein in each fraction and IDC stage, derived from either replicate A or B, whichever displayed the higher number of unique peptides. R, ring stages; T, trophozoites; S, schizonts.

Protein ID	Product Description	Fraction 1			Fraction 2			Fraction 3			Fraction 4			Fraction 5		
		R	T	S	R	T	S	R	T	S	R	T	S	R	T	S
Histones																
PFF0860c	histone H2A	1	1	1	7	1	3	12	3	5	7	4	8	1	1	3
PFC0920w	histone H2A.Z				2		2	9	4	6	8	3	5	4	3	4
PF11_0062	histone H2B	2	1		7	2	9	36	7	10	12	14	22	9	7	12
PF07_0054	histone H2B variant	2	1		1		3	17	9	4	12	7	15	8	5	7
PFF0510w	histone H3				1			4			8	4	7	2		
PFF0865w	histone H3 variant							1					3			
PF13_0185	histone H3, CENP-A variant										3		2			
PF11_0061	histone H4	1	3	2	3	1	7	10	3	4	15	13	22	8	6	13
DNA-directed RNA polymerase II complex																
PFC0805w	DNA-directed RNA pol II, putative	4									8	31	5	4	4	
PF11130c	DNA-directed RNA pol II, putative										1	6	2			
PF10_0269	DNA-directed RNA pol II, putative							1				2				
PFB0715w	DNA-directed RNA pol II 2nd largest subunit, putative										1	18	4		2	
PFL0665c	RNA polymerase subunit 8c, putative											1				
PFB0245c	DNA-directed RNA pol II 16 kDa subunit, putative											2				
PF13_0341	DNA-directed RNA pol II, putative											3	1			
PFA0505c	DNA-directed RNA pol II, putative										1	1				
PF13_0023	DNA-directed RNA pol II, putative											3	2			
PF07_0027	DNA-directed RNA pol II 8.2 kDa polypeptide, putative											1				
ApiAP2 transcription factors																
PFF0200c	PiSIP2									1						7
PFE0840c	transcription factor with AP2 domain(s), putative					3									1	
PFD0985w	transcription factor with AP2 domain(s), putative													1	1	
PF11_0404	transcription factor with AP2 domain(s), putative								1						6	
PF10_0075	transcription factor with AP2 domain(s), putative				3	9				4	3		2	8	10	
PF14_0471	transcription factor with AP2 domain(s), putative					4								1	4	5
PF14_0633	transcription factor with AP2 domain(s), putative					1			2	2				2	6	6
PF14_0533	transcription factor with AP2 domain(s), putative														1	3
PF11_0091	transcription factor with AP2 domain(s), putative				2	9	1			2					1	5
PFL1900w	transcription factor with AP2 domain(s), putative					29			2						14	21
PFF0670w	transcription factor with AP2 domain(s), putative									4				2	3	35
PF14_0079	transcription factor with AP2 domain(s), putative														1	
MAL8P1.153	transcription factor with AP2 domain(s), putative															4

Additional material

Additional files with the numbers 4, 5, 6, 7, 10, 12, 16, 48, 49, 53, 54, 55, 60 and 61 are not included in the thesis due to exceeding sizes.

Additional file 1 (.pdf). Schematic of the experimental approach for the purification and fractionation of parasite nuclei and 1D-SDS PAGE analysis of all protein fractions visualised by silver staining.

Additional file 2 (.pdf). Detailed protocols used for nuclear isolation and fractionation, sample preparation for MudPIT analysis, tandem mass spectrometry and peptide identification.

Additional file 3 (.pdf). Summary table of tryptic peptides measured by MudPIT.

Additional file 4 (.xls). Raw SEQUEST output data for peptide tandem mass spectra and proteins detected in ring stage parasites.

Additional file 5 (.xls). Raw SEQUEST output data for peptide tandem mass spectra and proteins detected in trophozoites.

Additional file 6 (.xls). Raw SEQUEST output data for peptide tandem mass spectra and proteins detected in schizonts.

Additional file 7 (.xls). Extensive summary table for all proteins detected in this study.

Additional file 8 (.pdf). Enrichment analyses for KEGG pathways and GO terms in the combined nuclear protein fractions.

Additional file 9 (.pdf). Detailed protocols used for bioinformatic analyses and validation of the core nuclear proteome.

Additional file 10 (.xls). *P. falciparum* protein localisations derived from published sources through homology.

Additional file 11 (.pdf). Additional text detailing the rationales for acceptance or rejection of bioinformatics filters, and discussing the identification of novel nuclear protein domains.

Additional file 12 (.xls). Proteins removed from the bioinformatically filtered nuclear proteome.

Additional file 13 (.xls). Proteins re-added to the nuclear proteome after they were removed by bioinformatic filtering.

Additional file 14 (.xls). Proteins with recently added signal peptide annotation.

Additional file 15 (.xls). Proteins with recently removed signal peptide annotation.

Additional file 16 (.xls). Extensive summary table for all proteins in the core nuclear proteome.

Additional file 17 (.pdf). List of proteins tagged and localised in this study.

Additional file 18 (.png). Localisation of NuProC1-3xHA (MAL7P1.38) during the IDC.

Additional file 19 (.png). Localisation of NuProC2-3xHA (PF10_0278) during the IDC.

Additional file 20 (.png). Localisation of NuProC3-3xHA (PF11_0250) during the IDC.

Additional file 21 (.png). Localisation of NuProC4-3xHA (PF11_0293) during the IDC.

Additional file 22 (.png). Localisation of NuProC5-3xHA (PF13_0042) during the IDC.

Additional file 23 (.png). Localisation of NuProC6-2xTy (PFL0625c) during the IDC.

Additional file 24 (.png). Localisation of NuProC7-3xHA (PF14_0393) during the IDC.

Additional file 25 (.png). Localisation of NuProC8-3xHA (PF08_0083) during the IDC.

Additional file 26 (.png). Localisation of NuProC9-3xHA (PF10_0091) during the IDC.

Additional file 27 (.png). Localisation of NuProC10-3xHA (PF10_0328) during the IDC.

Additional file 28 (.png). Localisation of NuProC11-3xHA (PF11_0254) during the IDC.

Additional file 29 (.png). Localisation of NuProC12-3xHA (PF11_0332) during the IDC.

Additional file 30 (.png). Localisation of NuProC13-3xHA (PF13_0099) during the IDC.

Additional file 31 (.png). Localisation of NuProC14-3xHA (PF14_0176) during the IDC.

Additional file 32 (.png). Localisation of NuProC15-3xHA (PF14_0433) during the IDC.

Additional file 33 (.png). Localisation of NuProC16-3xHA (PFC0126c) during the IDC.

Additional file 34 (.png). Localisation of NuProC17-3xHA (PFC0130c) during the IDC.

Additional file 35 (.png). Localisation of NuProC18-3xHA (PFC0690c) during the IDC.

Additional file 36 (.png). Localisation of NuProC19-3xHA (PFI0610w) during the IDC.

Additional file 37 (.png). Localisation of NuProC20-3xHA (PFI1355w) during the IDC.

Additional file 38 (.png). Localisation of NuProC21-3xHA (PFL0185c) during the IDC.

Additional file 39 (.png). Localisation of NuProC22-3xHA (PFL0450c) during the IDC.

Additional file 40 (.png). Localisation of n-NuProC1-3xHA (PF11_0099) during the IDC.

Additional file 41 (.png). Localisation of n-NuProC2-3xHA (MAL7P1.77) during the IDC.

Additional file 42 (.png). Localisation of n-NuProC3-3xHA (PF07_0007) during the IDC.

Additional file 43 (.png). Localisation of n-NuProC4-3xHA (PF10_0100) during the IDC.

Additional file 44 (.png). Localisation of n-NuProC5-3xHA (PF11_0179) during the IDC.

Additional file 45 (.png). Localisation of n-NuProC6-3xHA (PFB0395w) during the IDC.

Additional file 46 (.pdf). Protocol for the identification of novel protein domains.

Additional file 47 (.pdf). A multiple sequence alignment of the ACDC domains in apicomplexan proteins in graphical format.

Additional file 48 (.fasta). A multiple sequence alignment of the apicomplexan proteins encoding the ACDC domain is shown in aligned FASTA format. This file can be viewed using JalView (www.jalview.org).

Additional file 49 (.hmm). HMM used to query the *P. falciparum* and other alveolate whole proteomes to detect ACDC domains, in stockholm format. Note that this is not an HMM built from all putative ACDC domains, but simply an intermediate file used during the methodology of this publication. This file can be viewed using LogoMat-P (www.sanger.ac.uk/cgi-bin/software/analysis/logomat-p.cgi).

Additional file 50 (.pdf). A multiple sequence alignment of the apicomplexan proteins encoding the partial CSTF domain in graphical format.

Additional file 51 (.pdf). A pairwise sequence alignment of the *P. falciparum* proteins encoding the extended ELM2 domain in graphical format.

Additional file 52 (.pdf). A multiple sequence alignment of *P. falciparum* proteins encoding the MYND domain in graphical format.

Additional file 53 (.fasta). A multiple sequence alignment of the apicomplexan proteins encoding the partial CSTF domain in aligned FASTA format. This file can be viewed using JalView (www.jalview.org).

Additional file 54 (.fasta). A multiple sequence alignment of the apicomplexan proteins encoding the extended ELM2 domain in aligned FASTA format. This file can be viewed using JalView (www.jalview.org).

Additional file 55 (.fasta). A multiple sequence alignment of the apicomplexan proteins encoding the MYND domain in aligned FASTA format. This file can be viewed using JalView (www.jalview.org).

Additional file 56 (.pdf). Pairwise sequence alignment of the two putative FG-repeat proteins PFI0250c and PF14_0442.

Additional file 57 (.pdf). PCR and Western blot analysis to confirm C-terminal tagging of endogenous PF14_0442 by 3' replacement.

Additional file 58 (.pdf). mRNA expression profiles of proteins found stage-specifically in the nuclear proteome.

Additional file 59 (.pdf). A density plot displays maximal hour of mRNA expression of stage-specific proteins in the nuclear proteome.

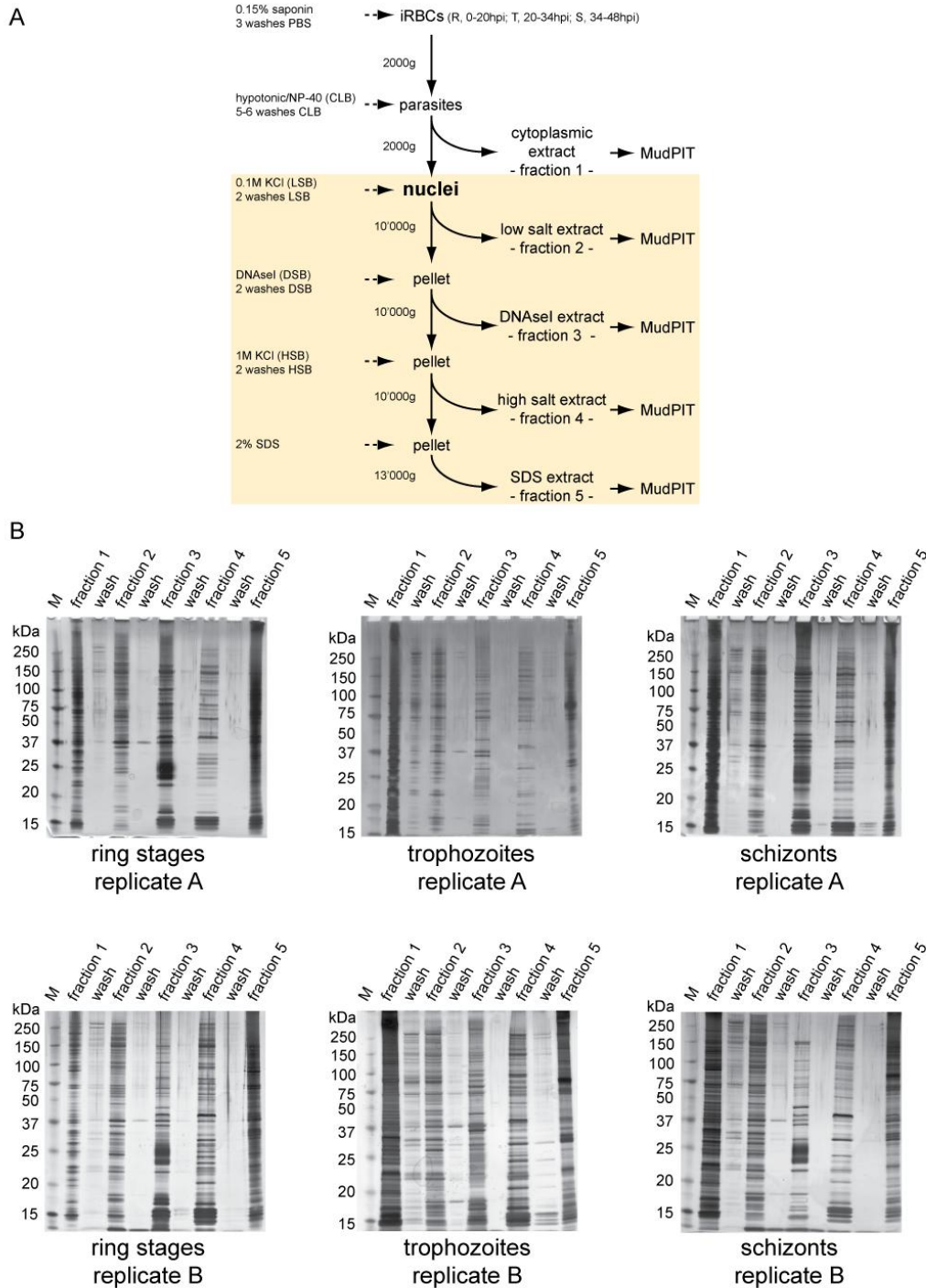
Additional file 60 (.xls). List of stage-specifically detected TFs.

Additional file 61 (.xls). Summary table for lineage-specific proteins in the core nuclear proteome.

Additional file 62 (.pdf). Protocol for identification of lineage-specific proteins in the core nuclear proteome.

Additional file 63 (.pdf). List of all primers used in this study.

Additional file 1_Oehring et al.



Additional file 1. Cytoplasmic and nuclear protein fractions. (A) The flowchart illustrates the experimental approach for the purification and fractionation of parasite nuclei. Fraction 1 corresponds to the cytoplasmic extract. The shaded box highlights the differential extraction of nuclear proteins into fractions 2 to 5. Isolation and fractionation has been performed twice each for ring stages, trophozoites and schizonts. (B) 1D-SDS PAGE analysis of all protein fractions obtained from ring stages, trophozoites and schizonts (two replicates each), visualised by silver staining. Fractions 1 to 5 are labelled as in (A). The last wash fraction prior to each subsequent extraction is also shown.

Additional file 2_Oehring et al.**Isolation of nuclei and nuclear fractionation.**

Parasites (2×10^{10} ring stages (2-20 hpi); 10^{10} trophozoites (20-34 hpi); 5×10^9 schizonts (34-48 hpi) were released from RBCs by saponin lysis and washed three times in PBS. Parasites were lysed in a hypotonic cytoplasmic lysis buffer CLB (20mM HEPES (pH7.9), 10mM KCl, 1mM EDTA, 1mM EGTA, 0.65% NP-40, 1mM DTT, Complete TM protease inhibitors (Roche Diagnostics)) for 5 min on ice. Nuclei were pelleted at 2,000g for 5 min and the supernatant saved (cytoplasmic extract; fraction 1). After four to seven washes in CLB nuclei were extracted in 1.5ml (ring stages and trophozoites) or 3ml (schizonts) low salt buffer LSB (20mM HEPES (pH7.9), 0.1M KCl, 1mM EDTA, 1mM EGTA, 1mM DTT, protease inhibitors) for 20 min at 4°C under constant agitation. Insoluble material was pelleted for 3 min at 13,000rpm and the soluble fraction saved (low salt extract; fraction 2). Pellets were washed twice in 1.5ml (3ml) LSB and chromatin was solubilised in 1.5ml (3ml) DNaseI-digestion buffer DDB (20mM Tris-HCl (pH7.5); 15mM NaCl, 60mM KCl, 1mM CaCl₂, 5mM MgCl₂, 5mM MnCl₂, 300mM sucrose, 0.4% NP-40, 1mM DTT, protease inhibitors) containing 100-500U DNaseI for 20 min under constant agitation at RT and 37°C. Samples were centrifuged at 13,000rpm for 3 min and the soluble fraction was saved (DNaseI extract; fraction 3). After two washes in 1.5ml (3ml) DDB pellets were extracted with high salt buffer HSB (20mM HEPES (pH7.9), 1M KCl, 1mM EDTA, 1mM EGTA, 1mM DTT, protease inhibitors) for 20 min at 4°C under constant agitation. The soluble fraction was recovered by centrifugation at 13,000rpm for 3 min and saved (high salt extract; fraction 4). The insoluble nuclear matrix fraction was washed twice in 1.5ml (3ml) HSB, solubilised in 1.5ml (3ml) SDS extraction buffer SEB (2%SDS, 10mM Tris-HCl (pH 7.5)) for 20 min under constant agitation at room temperature, cleared by centrifugation at 13,000rpm for 15min and saved (SDS extract; fraction 5). The quality of all samples was analysed by SDS-PAGE and silver staining and by Western blot.

Sample preparation for MudPIT LC-MS/MS analysis

Within each set of corresponding fractions from each of the three different parasite stages we attempted to analyse equal amounts of total protein extracted. We therefore estimated the protein concentration in each fraction by inspection of silver-stained SDS-PAGE gels and precipitated adequate volumes in 10% TCA. Protein pellets were

washed twice in 1ml ice-cold 100% acetone, air-dried and re-dissolved in 10µl 100mM Tris-HCl, pH 8.0 containing 6M urea and diluted with 20µl 100mM Tris-HCl, pH8.0 to lower the urea concentration to 2M. Digestion was done with 0.25µg endoproteinase LysC (ELC, Roche Diagnostics) for one hour at 37°C followed by a second aliquot of ELC and incubation at 37°C. Subsequently, trypsin digestion was done with 0.25µg trypsin (Promega) at 37°C for 2 hrs followed by a second digestion with trypsin (0.25µg) and incubation for 18 hrs at 37°C. The digest was stopped by adding 3µl 10% trifluoro acetic acid (TFA). The digest was diluted with 100µl 0.1% TFA/1% acetonitrile and desalted on a MacroSpin column packed with Vydac C18 reverse-phase material (The Nest Group, Southborough, MA) according to the manufacturer's recommendations. The peptides were eluted with 300µl of 80% acetonitrile/0.1% TFA, dried in a speed vacuum system and re-dissolved in 100µl 0.1% formic acid/2% acetonitrile.

MudPIT liquid chromatography tandem mass spectrometry

The desalted peptides were analyzed by two-dimensional capillary liquid chromatography and tandem mass spectrometry using a Polysulfoethyl A ion-exchange column (0.15 x 50 mm, PolyLC, Columbia, MD) connected in series to a C18 trap column (Zorbax 300SB, 0.3 x 50 mm, Agilent Technologies) and a Magic C18 separation column (0.1 x 100 mm, Thermo Scientific). 10 µl of the digest were injected first onto the cation-exchange column. Unadsorbed peptides were trapped on the Zorbax column and eluted onto the separation column with a linear 75 min gradient from 2 to 75% B (0.1% acetic acid in 80% acetonitrile) in solvent A (0.1% acetic acid in 2% acetonitrile). Next, peptides retained by the cation-exchange column were sequentially eluted and trapped onto the C18 trap column with 10µL pulses of 25, 50, 100, 150, 200, 250 and 500mM ammonium acetate, pH 3.3. Peptides eluted by each individual salt pulse were separated by the acetonitrile gradient as described above. The flow was delivered with a Rheos 2200 HPLC system (Thermo Scientific) at 150µL/min. A pre-column splitter reduced the flow to approximately 500nl/min. The eluting peptides were ionized by a Finnigan nanospray ionization source (Thermo Scientific).

Mass spectrometric analysis was carried out on an LTQ Orbitrap hybrid instrument (Thermo Finnigan, San José, CA). The instrument was operated in a data-dependent mode. A survey scan was performed in the Orbitrap between m/z 400-1600 Da at 60,000 resolution. The five most abundant ions detected in the survey scan were fragmented and mass analysed in the LTQ part of the instrument. Singly charged ions

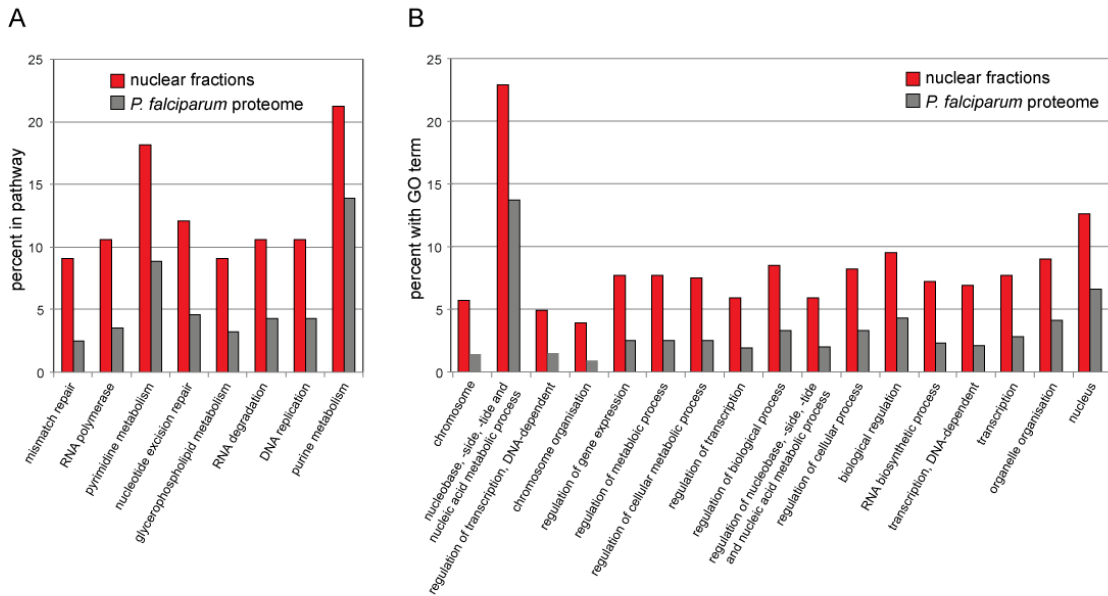
were omitted from fragmentation. The normalized collision energy was set to 35%. For peptide identification, the MS/MS spectra from each individual salt step-off were searched against a combined *P. falciparum* (www.plasmoDB.org; release 5.5)/human (NCBI, version August 2008) database using the TurboSEQUEST software [1]. False discovery rates (FDR) were estimated by searching against the reversed merged database. The SEQUEST filter parameters were as follows: Xcorr versus charge state was 2.00 for doubly, and 2.50 for triply and 3.00 for quadruply charged ions, respectively; the Δ CN was 0.1, peptide probability was 0.5, and the protein probability was set to 0.01. Single peptide hits were allowed if they matched the above search criteria. The search results from each individual salt step-off were merged with the multiconsensus option of TurboSEQUEST (Additional files 4-6).

1. Gatlin CL, Eng JK, Cross ST, Detter JC, Yates JR, III: Automated identification of amino acid sequence variations in proteins by HPLC/microspray tandem mass spectrometry. *Anal Chem* 2000, 72: 757-763.

Additional file 3. Summary table of tryptic peptides measured by MudPIT. Column A: parasite stage. Column B: cellular compartment. Column C: protein fractions 1 to 5. Column D replicates A and B. Column E: number of unique peptides detected. Column F: false discovery rate. Column G: number of different proteins detected. Column H: total number of proteins detected per developmental stage. Column I: total number of proteins detected exclusively in one developmental stage. Column J: total number of proteins detected in the nuclear fractions 2 to 5 per developmental stage.

A	B	C	D	E	F	G	H	I	J	
stage	compartment	fraction	replicate	# unique peptides	FDR (%)	# proteins	total # proteins	# unique proteins	# proteins in nuclear fractions	
rings	cytoplasm	1	A	1797	0.11	303	1050	181	896	
	cytoplasm	1	B	3852	0.13	538				
	nuclear		2	A	649	0.15	167			
			2	B	1003	0.10	164			
			3	A	679	0.59	117			
			3	B	456	0.00	49			
			4	A	657	0.30	63			
			4	B	2042	0.29	253			
			5	A	1816	0.28	318			
			5	B	3421	0.06	565			
trophozoites	cytoplasm	1	A	2024	0.05	347	1017	159	901	
	cytoplasm	1	B	3611	0.00	493				
	nuclear		2	A	1416	0.71	315			
			2	B	2093	0.19	309			
			3	A	811	0.86	156			
			3	B	2821	0.18	309			
			4	A	1198	0.42	125			
			4	B	2583	0.12	227			
			5	A	1620	0.19	149			
			5	B	3590	0.14	350			
schizonts	cytoplasm	1	A	2331	0.47	401	1092	186	948	
	cytoplasm	1	B	3675	0.22	513				
	nuclear		2	A	576	0.00	162			
			2	B	406	0.25	117			
			3	A	865	0.12	176			
			3	B	1480	0.27	197			
			4	A	931	0.86	91			
			4	B	1497	0.47	199			
			5	A	1811	0.39	267			
			5	B	2945	0.37	435			
TOTAL						1525	1518	526	1273	

Additional file 8_Oehring et al.



Additional file 8. Enrichment analyses for KEGG pathways and GO terms. Proteins found only in the nuclear preparation (fractions 2 to 5) were subject to enrichment analyses. Percentages of nuclear proteins with relevant pathways and terms are shown in red, the percentages of the whole genome with the corresponding pathways and terms are shown in grey. (A) DAVID [1] was used to test for KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway enrichment. Enriched metabolic pathways included DNA replication, mismatch repair, RNA polymerase, purine and pyrimidine metabolism, and glycerophospholipid metabolism. (B) Gostat [2] was used to test for GO term enrichment. Enriched GO terms included nucleus, chromosome, and transcription.

1. Huang da W, Sherman BT, Lempicki RA: Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009, 4: 44-57.
2. Beissbarth T, Speed TP: Gostat: find statistically overrepresented Gene Ontologies within a group of genes. *Bioinformatics* 2004, 20: 1464-1465.

Additional file 9_Oehring et al.**Over-Representation of Functional Classes**

To examine for enrichments in functional classes associated with nuclear or other processes, we used the tools GStat [1] and DAVID [2] to check for overrepresented GO terms or KEGG pathways in the nuclear fractions. Statistically over-represented GO terms (p-value < 10⁻⁵) and KEGG pathways that were enriched more than 1.5-fold in the nuclear proteome were inspected and manually curated to verify assignments.

Gold Standard Definition for Accuracy Appraisal

Two sources of information about protein localisation were first independently and then together used as a gold standard: ApiLoc version 2 (apiloc.biochem.unimelb.edu.au), and a literature-linked survey of protein localisations (Additional file 10). For each gold standard a protein was classified as (1) nuclear if it was nuclear-localised during any life cycle stage, even if it was dually localised (whether concurrently or not); (2) non-nuclear if it was localised to another location; and (3) unknown if no localisation information was recorded. Proteins annotated only with “cytoplasm” were considered non-nuclear. PF10_0155, PF10_0395, PFE0360c and PFE0285c were excluded due to conflicting information regarding localisation. PF10_0177b was treated as having unknown localisation since it is unclear whether it localises to the outer regions of the nucleus or immediately outside of the [3]. Enrichment (Figures 3A and 4A) and positive predictive value (Figures 3B and 4B) were calculated as per Equations 1 and 2, respectively.

Determination of Enrichment (Equation 1):

$$Enrichment = \frac{n/N}{t/T}$$

n refers to the number of nuclear proteins in the set being appraised. N refers to the number of nuclear proteins annotated in the gold standard, t refers to the total number of proteins annotated as nuclear or non-nuclear in the gold standard and T refers to the total number of proteins in the *P. falciparum* proteome (5446).

Determination of Positive Predictive Value (Equation 2):

$$PPV = \frac{n}{n+c}$$

PPV was the positive predictive value, n was the number of nuclear proteins in the set being appraised, and c was the number of non-nuclear proteins in the set being appraised. The proteins that were counted as nuclear, non-nuclear and excluded varied according to the gold standard being used.

Determination of Statistical Over-representation of Nuclear Proteins

To determine statistical significance of over-representation of nuclear proteins, Fisher's exact test was applied using R version 2.13.1 (<http://r-project.org>), with 37 positive, 130 negative, 18 excluded (those above as well as those classified as ER) and 1088 unknown proteins as classified using ApiLoc, and assuming that the positive predictive value was maintained in the proteins of unknown localisation that were detected here by mass spectrometry.

A similar result obtained using the literature survey was found using 136 positive, 55 negative, 7 excluded and 1075 unknown. All further calculations were carried out using the combined ApiLoc and literature review gold standards.

Bioinformatic Contaminant Removal Techniques

Exported proteins were predicted using ExportPred [4], taken from PlasmoDB v6.4 using the default cut-off of 4.3, and PfHP1-association [5]. SPs were firstly predicted using SignalP [6] using default settings at PlasmoDB v6.4. TM domains were predicted using a locally installed version of TMHMM 2.0c [7] and a wrapper script. The wrapper script was subsequently made into a Biogem (<http://biogems.info>), named bio-tm_hmm. To predict the percentage of *S. cerevisiae* nuclear TM domain proteins, proteins annotated with the GO term [8] "nucleus" (GO:0005634) were downloaded (Oct 11, 2010) from Amigo version 1.7 (www.geneontology.org). Corresponding protein sequences were taken from the SGD [9] on 11 Oct, 2011. Signal peptides were removed prior to TM domain prediction using a locally installed version of SignalP v3.0 and a custom BioRuby [10] wrapper script git version 22593c (github.com/wwood/bbbin/blob/master/signalp.rb) so proteins with a predicted signal peptide but no transmembrane domain predicted were not filtered out when only the TM filter was applied. To predict the number of proteins in the entire *S. cerevisiae* proteome that had TMs, the "orf_trans.fasta" file was downloaded from SGD on Jan 6, 2010 and underwent cleavage of signal peptides and transmembrane domain prediction as above. Proteins found in the food vacuole [11] and Maurer's cleft [12] proteomes were removed by taking the gene IDs from the

supplementary data of each publication. Mitochondrial transit peptides were predicted using PlasMit [13], <http://gecco.org.chemie.uni-frankfurt.de/plasmit/index.html>, accessed June 10, 2010) using PlasmoDB v6.4.

The final core nuclear proteome was defined as the proteins that remained after applying the combined filter (exported; SP positive; TM positive; present in the Maurer's cleft proteome; present in the food vacuole proteome), with four amendments. Firstly, proteins predicted to be non-nuclear according to the ApiLoc or literature review gold standards were removed (Additional file 12). Secondly, proteins predicted as true nuclear proteins by the gold standards but ruled out by the combined bioinformatic filter were re-added (Additional file 13). Thirdly, two proteins recently demonstrated to be associated with the nucleus were added (PF14_0443 [14] and PF10_0268 [15]). Fourthly, proteins recently annotated as having a SP were removed from the core nuclear proteome (Additional file 14), and those where a previous SP annotation was recently removed were re-added (Additional file 15) (PlasmoDB version 8.0).

Accuracy of the Core Nuclear Proteome

According to the gold standards, the core nuclear proteome contained 148 known nuclear proteins, 652 proteins of unknown location, and two with conflicting localisation information. Under the assumption that 70% of these unknown proteins are nuclear (since this was the estimate after the combined filter was applied), it was estimated that 473 of these unknown proteins were in fact nuclear-localised, giving an overall estimate of positive predictive value of 76% ($(70\% \text{ of } 652 + 148) / 802$).

Classical Nuclear Localisation Signal Predictors

PredictNLS [16] version 1.0.15 was used with default parameters after installing the debian package from http://roslab.org/debian/pool/main/predictnls_1.0.15_all.deb. To predict from multiple genes, a custom BioPerl script was created, available from <https://github.com/wwood/bbbin/blob/master/predictmnl>, and git commit e4730033 was used. NLStradamus [17] version 1.4 was used with default parameters after download of the standalone program from <http://www.moseslab.csb.utoronto.ca/NLStradamus>. Prediction using cNLS was carried out by the creation of a custom bioruby [10] plugin (<https://github.com/helios/bioruby-gem>) `cnls_screenscraper` (rubygem version 0.1.0, available from http://rubygems.org/gems/bio-cnls_screenscraper), which queried the cNLS [18] webpage (<http://nls-mapper.iab.keio.ac.jp/>) automatically for each protein

in the *P. falciparum* proteome (PlasmoDB v7.0). Positive predictions cutoffs were taken as 8.0 and 7.0 for monopartite and bipartite signals, respectively. Predictions were obtained for 5452 proteins, with the remaining minority of proteins not predicted because they were too short (<19 residues), too long (>5000 residues), or contained 'X' characters in their protein sequence. Fisher's exact tests were applied using R 2.12.0.

Stage-Specific Patterns

3D7 glass slide microarray information was taken from a previous study [19] and linked to the proteomic data by direct matching of PlasmoDB gene IDs. In case of multiple matches, one was taken at random as representative. Plots were generated with ggplot2 [20]. Proteins found stage-specifically were defined as those found in only one of ring, trophozoite or schizont stages. Predicted TFs [21] were intersected with the core nuclear proteome list. The stages where they were found were tabulated and annotated with Affymetrix microarray annotation [22] using PlasmoDB version 6.4.

Similarity between *P. falciparum* and *S. cerevisiae* nuclear proteins

For each protein in the predicted *P. falciparum* proteome (PlasmoDB version 6.3, 5446 proteins), the best match in the *S. cerevisiae* proteome was determined with BLASTP using default parameters but low-frequency filtering off [23]. *P. falciparum* proteins were classified as 'likely nuclear' if they had as a best match (E-value < 10⁻⁵) a *S. cerevisiae* protein associated with the GO term nucleus in SGD and/or localised to the yeast nucleus [24].

1. Beissbarth T, Speed TP: GStat: find statistically overrepresented Gene Ontologies within a group of genes. *Bioinformatics* 2004, 20: 1464-1465.
2. Huang da W, Sherman BT, Lempicki RA: Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009, 4: 44-57.
3. van Ooij C., Tamez P, Bhattacharjee S, Hiller NL, Harrison T, Liolios K et al.: The malaria secretome: from algorithms to essential function in blood stage infection. *PLoS Pathog* 2008, 4: e1000084.
4. Sargeant TJ, Marti M, Caler E, Carlton JM, Simpson K, Speed TP et al.: Lineage-specific expansion of proteins exported to erythrocytes in malaria parasites. *Genome Biol* 2006, 7: R12.
5. Flueck C, Bartfai R, Volz J, Niederwieser I, Salcedo-Amaya AM, Alako BT et al.: Plasmodium falciparum heterochromatin protein 1 marks genomic loci linked to phenotypic variation of exported virulence factors. *PLoS Pathog* 2009, 5: e1000569.
6. Bendtsen JDv, Nielsen H, von Heijne G, Brunak SÅ: Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 2004, 340: 783-795.
7. Krogh A, Larsson B, von Heijne G, Sonnhammer EL: Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 2001, 305: 567-580.
8. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM et al.: Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics* 2000, 25: 25-29.
9. Cherry JM, Adler C, Ball C, Chervitz SA, Dwight SS, Hester ET et al.: SGD: Saccharomyces Genome Database. *Nucleic Acids Res* 1998, 26: 73-79.

10. Goto N, Prins P, Nakao M, Bonnal R, Aerts J, Katayama T: BioRuby: bioinformatics software for the Ruby programming language. *Bioinformatics* 2010, 26: 2617-2619.
11. Lamarque M, Tastet C, Poncet J, Demetree E, Jouin P, Vial H et al.: Food vacuole proteome of the malarial parasite *Plasmodium falciparum*. *Proteomics Clin Appl* 2008, 2: 1361-1374.
12. Vincensini L, Richert S, Blisnick T, Van DA, Leize-Wagner E, Rabilloud T et al.: Proteomic analysis identifies novel proteins of the Maurer's clefts, a secretory compartment delivering *Plasmodium falciparum* proteins to the surface of its host cell. *Mol Cell Proteomics* 2005, 4: 582-593.
13. Bender A, van Dooren GG, Ralph SA, McFadden GI, Schneider G: Properties and prediction of mitochondrial transit peptides from *Plasmodium falciparum*. *Mol Biochem Parasitol* 2003, 132: 59-66.
14. Mahajan B, Selvapandian A, Gerald NJ, Majam V, Zheng H, Wickramarachchi T et al.: Centrins, cell cycle regulation proteins in human malaria parasite *Plasmodium falciparum*. *J Biol Chem* 2008, 283: 31871-31883.
15. Richard D, Bartfai R, Volz J, Ralph SA, Muller S, Stunnenberg HG et al.: A genome-wide chromatin-associated nuclear peroxiredoxin from the malaria parasite *Plasmodium falciparum*. *J Biol Chem* 2011, 286: 11746-11755.
16. Cokol M, Nair R, Rost B: Finding nuclear localization signals. *EMBO Rep* 2000, 1: 411-415.
17. Nguyen Ba AN, Pogoutse A, Provart N, Moses AM: NLStradamus: a simple Hidden Markov Model for nuclear localization signal prediction. *BMC Bioinformatics* 2009, 10: 202.
18. Kosugi S, Hasebe M, Tomita M, Yanagawa H: Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. *Proc Natl Acad Sci U S A* 2009, 106: 10171-10176.
19. Llinas M, Bozdech Z, Wong ED, Adai AT, DeRisi JL: Comparative whole genome transcriptome analysis of three *Plasmodium falciparum* strains. *Nucleic Acids Res* 2006, 34: 1166-1173.
20. Wickham H: *ggplot2: elegant graphics for data analysis*. Springer-Verlag New York Inc; 2009.
21. Bischoff E, Vaquero C: In silico and biological survey of transcription-associated proteins implicated in the transcriptional machinery during the erythrocytic development of *Plasmodium falciparum*. *BMC Genomics* 2010, 15: 34.
22. Le Roch KG, Zhou Y, Blair PL, Grainger M, Moch JK, Haynes JD et al.: Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science* 2003, 301: 1503-1508.
23. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W et al.: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997, 25: 3389-3402.
24. Huh WK, Falvo JV, Gerke LC, Carroll AS, Howson RW, Weissman JS et al.: Global analysis of protein localization in budding yeast. *Nature* 2003, 425: 686-691.

Additional file 11_Oehring et al.**Accepted bioinformatic filters**

Removal of proteins carrying putative SPs and/or predicted to be exported to the host cell removed many of the proteins directed to destinations in the endomembrane system and RBC [1-3]. These filters were consistent with the hypothesis that proteins with amino acid sequences encoding putative signals for trafficking to non-nuclear compartments were unlikely to represent true nuclear proteins. Proteins carrying predicted TM were considered less likely to represent nuclear proteins, despite the fact that integral membrane proteins are present in the nuclear envelope. Analysis of the *S. cerevisiae* proteome undertaken here showed that 17% (1014/5885) of all proteins contain a predicted TM compared to only 3% (57/2132) of proteins annotated as nuclear in the gene ontology [4]. Lastly, proteins found in proteomic studies of non-nuclear organelles were considered less likely to be nuclear-localised, either because they are truly associated with that organelle or because they may represent possible contaminants in organelle proteomics studies in general.

Rejected bioinformatics filters

In terms of numbers of distinct proteins, the largest subcellular destinations outside the endomembrane system apart from the nucleus are the cytosol and mitochondrion. While no commonly used predictors attempting to specifically predict cytosolic proteins in *P. falciparum* have yet emerged, *P. falciparum* mitochondrial proteins have been the target of at least four predictors [5-8]. Removal of proteins predicted to localise to the mitochondrion using PlasMit [5] resulted in a very slight reduction of positive predictive value (0.1%) of the set, combined with a large loss in nuclear proteins (unfiltered set size reduced by 239 proteins or 19%). Thus, from this empirical point of view, choosing to exclude proteins based on PlasMit prediction would result in a substantially reduced set of proteins that is not further enriched in nuclear proteins.

However, the decision not to use a mitochondrial transit peptide predictor as bioinformatic filter leaves open the question of whether significant mitochondrial contamination is present in the core nuclear proteome. To address this, we looked specifically for proteins recorded as being mitochondrial in ApiLoc or as part of our literature review. 28 of these 110 proteins were found in the nuclear proteome, and 16 of these remained in the set after bioinformatic filtering. Based on this we predict 7% of the

core nuclear proteome to be mitochondrial (55 proteins), and that approximately one-third of contaminating proteins are mitochondrial.

Other potential filters found to be of limited utility were based on the distribution of detected peptides across the five fractions, developmental stages, and biological replicates. Removal of proteins detected by a single peptide increased the positive predictive value by 1% only, while incurring a loss of 22% of proteins from the unfiltered set. Finally, removal of proteins that were also detected in the cytoplasmic fraction gives an 8% increase in positive predictive value at a cost of 49% of the proteins in the set. Applying this criterion in addition to those already accepted leaves a set of 440 proteins with 81% being nuclear.

Identification of novel domains associated with nuclear proteins

PF10_0279 and PFI1600w, annotated as hypothetical and mRNA processing protein, respectively, were found to share a 56aa C terminal domain (Additional files 52 and 53). According to OrthoMCL v4 [9] PFI1600w is orthologous to the opisthokont cleavage stimulation factor (CSTF) 64kDa subunit, which is known to be involved in cleavage of 3' untranslated regions (UTRs) [10]. PFI1600w was found in all sequenced Apicomplexans except *Theileria* spp. PF10_0279 lacks primary sequence similarity apart from the partial CSTF domain with proteins outside *Plasmodium* spp., suggesting it may be a genus-specific innovation.

A region of sequence similarity related to the ELM2 domain was found in PF11_0429 and PFE0995c (Additional files 54 and 56). ELM2 domains, first found in *Caenorhabditis elegans* [11] were subsequently found in a number of *Drosophila melanogaster* and vertebrate proteins involved in transcriptional repression through association with histone deacetylases and methyltransferases, and have been localised to the nucleus [12-14]. This domain had recently been annotated as an ELM2 domain in PlasmoDB version 7.0. However, the regions of similarity between the two proteins extended approx. six and 50 residues beyond the N and C terminal boundaries, respectively (PROSITE, domain PS51156) [15]. As such, we refer to this region of sequence similarity as an “extended ELM2 domain”. PF11_0429 and some apicomplexan homologues also carry annotated PHD-finger domains that are involved in chromatin remodelling [16].

The algorithm also found a domain conserved between PF14_0310 (conserved *Plasmodium* protein) and PF10_0150 (methionine aminopeptidase 1b, putative)

(Additional files 55 and 57). Subsequent characterisation with PSI-BLAST [17] and the HMM-HMM comparison tool HHSearch [18] revealed homology to the MYND zinc finger domain (PFAM ID PF01753), which is found in many eukaryote proteomes. Further searching on full amino acid sequences (no longer masked by EuPathDB-annotated domains) from all *P. falciparum* proteins identified three additional proteins (PFF0350w, PFF0105w, PF13_0293) each of which carried a previously annotated MYND domain. PF13_0293 encodes a putative histone-lysine N-methyltransferase and has previously been localised to the nucleus [19].

In addition, our search also identified similarity between five alveolin proteins [20]. Two of these have been localised to the cytoskeleton [20,21] in *P. falciparum*, and all five have been localised to the parasite cytoskeleton in *T. gondii* [22]. It is possible that these proteins are peripherally involved in nuclear biology as part of the cytoskeleton, which has been implicated in the import of proteins into the nucleus [23], though it seems most probable that they are simply contaminants in the nuclear fractions.

1. Tonkin CJ, Pearce JA, McFadden GI, Cowman AF: Protein targeting to destinations of the secretory pathway in the malaria parasite *Plasmodium falciparum*. *Curr Opin Microbiol* 2006, 9: 381-387.
2. Sargeant TJ, Marti M, Caler E, Carlton JM, Simpson K, Speed TP et al.: Lineage-specific expansion of proteins exported to erythrocytes in malaria parasites. *Genome Biol* 2006, 7: R12.
3. Flueck C, Bartfai R, Volz J, Niederwieser I, Salcedo-Amaya AM, Alako BT et al.: *Plasmodium falciparum* heterochromatin protein 1 marks genomic loci linked to phenotypic variation of exported virulence factors. *PLoS Pathog* 2009, 5: e1000569.
4. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM et al.: Gene ontology: tool for the unification of biology. *The Gene Ontology Consortium. Nature genetics* 2000, 25: 25-29.
5. Bender A, van Dooren GG, Ralph SA, McFadden GI, Schneider G: Properties and prediction of mitochondrial transit peptides from *Plasmodium falciparum*. *Mol Biochem Parasitol* 2003, 132: 59-66.
6. Verma R, Varshney GC, Raghava GP: Prediction of mitochondrial proteins of malaria parasite using split amino acid composition and PSSM profile. *Amino Acids* 2010, 39: 101-110.
7. Chen YL, Li QZ, Zhang LQ: Using increment of diversity to predict mitochondrial proteins of malaria parasite: integrating pseudo-amino acid composition and structural alphabet. *Amino Acids* 2010, [epub ahead of print].
8. Jia C, Liu T, Chang AK, Zhai Y: Prediction of mitochondrial proteins of malaria parasite using bi-profile Bayes feature extraction. *Biochimie* 2011, 93: 778-782.
9. Li L, Stoeckert CJ, Jr., Roos DS: OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* 2003, 13: 2178-2189.
10. Takagaki Y, Manley JL, MacDonald CC, Wilusz J, Shenk T: A multisubunit factor, CstF, is required for polyadenylation of mammalian pre-mRNAs. *Genes Dev* 1990, 4: 2112-2120.
11. Solari F, Bateman A, Ahringer J: The *Caenorhabditis elegans* genes *egl-27* and *egr-1* are similar to MTA1, a member of a chromatin regulatory complex, and are redundantly required for embryonic patterning. *Development* 1999, 126: 2483-2494.
12. Wang L, Charroux B, Kerridge S, Tsai CC: Atrophin recruits HDAC1/2 and G9a to modify histone H3K9 and to determine cell fates. *EMBO Rep* 2008, 9: 555-562.
13. Wang L, Rajan H, Pitman JL, McKeown M, Tsai CC: Histone deacetylase-associating Atrophin proteins are nuclear receptor corepressors. *Genes Dev* 2006, 20: 525-530.
14. Ding Z, Gillespie LL, Paterno GD: Human MI-ER1 alpha and beta function as transcriptional repressors by recruitment of histone deacetylase 1 to their conserved ELM2 domain. *Mol Cell Biol* 2003, 23: 250-258.
15. Hulo N, Bairoch A, Bulliard V, Cerutti L, Cuče BA, de Castro E et al.: The 20 years of PROSITE. *Nucleic Acids Res* 2008, 36: D245-D249.

16. Musselman CA, Kutateladze TG: PHD fingers: epigenetic effectors and potential drug targets. *Mol Interv* 2009, 9: 314-323.
17. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W et al.: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997, 25: 3389-3402.
18. Soding J, Biegert A, Lupas AN: The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* 2005, 33: W244-W248.
19. Volz J, Carvalho TG, Ralph SA, Gilson P, Thompson J, Tonkin CJ et al.: Potential epigenetic regulatory proteins localise to distinct nuclear sub-compartments in *Plasmodium falciparum*. *Int J Parasitol* 2010, 40: 109-121.
20. Gould SB, Tham WH, Cowman AF, McFadden GI, Waller RF: Alveolins, a new family of cortical proteins that define the protist infrakingdom Alveolata. 2008, 25: 1219.
21. Hu G, Cabrera A, Kono M, Mok S, Chaal BK, Haase S et al.: Transcriptional profiling of growth perturbations of the human malaria parasite *Plasmodium falciparum*. *Nat Biotechnol* 2010, 28: 91-98.
22. Anderson-White BR, Ivey FD, Cheng K, Szatanek T, Lorestani A, Beckers CJ et al.: A family of intermediate filament-like proteins is sequentially assembled into the cytoskeleton of *Toxoplasma gondii*. *Cell Microbiol* 2011, 13: 18-31.
23. Wagstaff KM, Jans DA: Importins and beyond: non-conventional nuclear transport mechanisms. *Traffic* 2009, 10: 1188-1198.

Additional file 13. Proteins re-added to the nuclear proteome after they were removed by bioinformatic filtering. Some proteins known to be nuclear by experimental localisation (recorded in Apilooc), or suggested to be localised to the nucleus (as part of Additional file 10), were removed by the bioinformatic filtering steps. These proteins were re-added and were part of the core nuclear proteome. The most prevalent reason for exclusion of these proteins was the presence of putative TM domains.

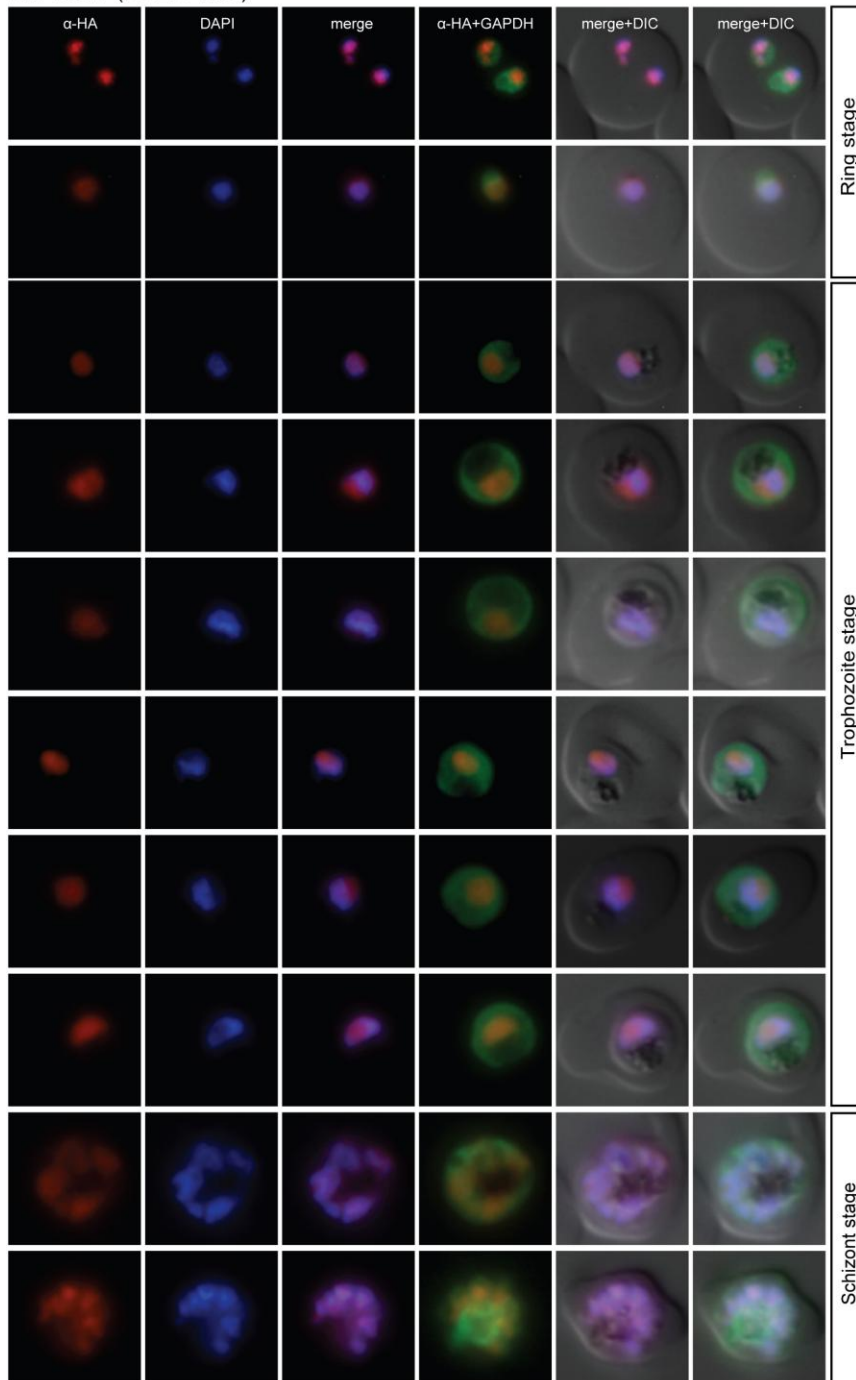
Protein ID	MC proteome	FV proteome	SP	exported	# TM domains	Apilooc literature review: nuclear	Annotation	Reason for re-addition
MAL13P1.19	false	false	false	false	2	positive	peptidase, putative	Known nuclear protein
MAL13P1.56	false	true	false	false	1	positive	m1-family aminopeptidase	Known nuclear protein
MAL13P1.61	false	false	true	true	3	positive	Plasmodium exported protein (hyp8), unknown function	Known nuclear protein
MALBP1.69	true	true	false	false		positive	14-3-3 protein, putative	Known nuclear protein
PF08_0054	true	true	false	false		positive	heat shock 70 kDa protein	Known nuclear protein
PF10_0069	false	true	false	false	1	positive	RNA binding protein, putative	Known nuclear protein
PF10_0143	false	false	false	false		positive	transcriptional activator ADA2, putative	Known nuclear protein
PF11_0049	false	false	false	false	4	positive	NOT family protein, putative	Known nuclear protein
PF14_0315	false	false	false	false	12	positive	conserved Plasmodium membrane protein, unknown function	Known nuclear protein
PF14_0360	false	false	false	false	2	positive	N-acetyltransferase, putative	Known nuclear protein
PF14_0393	false	true	false	false		positive	structure specific recognition protein	Known nuclear protein
PF14_0690	false	false	false	false	3	positive	histone deacetylase, putative	Known nuclear protein
PF00425w	false	false	false	false	1	positive	conserved Plasmodium protein, unknown function	Known nuclear protein
PF0250c	false	false	false	false	4	positive	conserved Plasmodium membrane protein, unknown function	Known nuclear protein
PFL1345c	false	false	false	false	1	positive	histone acetyltransferase, putative	Known nuclear protein
PFL2390c	false	false	false	false	2	positive	conserved Plasmodium protein, unknown function	Known nuclear protein

Additional file 14. Proteins with recently added signal peptide annotation. Proteins in this set had a SP annotated in PlasmoDB version 8.0 and were removed from the core nuclear proteome.	
PlasmoDB ID	Annotation
PFL1760w	conserved Plasmodium protein, unknown function
PF14_0198	glycine-tRNA ligase, putative
PF14_0186	conserved Plasmodium protein, unknown function
PF11_0511	Plasmodium exported protein, unknown function
PF11_0349	conserved Plasmodium protein, unknown function
PF10_0149	cysteinyI-tRNA synthetase, putative
MAL7P1.141	conserved Plasmodium protein, unknown function

Additional file 15. Proteins with recently removed signal peptide annotation. Proteins originally annotated with a SP had this motif removed in PlasmoDB version8.0 and were added to the core nuclear proteome.	
PlasmoDB ID	Annotation
PFL1745c	clustered-asparagine-rich protein
MAL13P1.79	conserved Plasmodium protein, unknown function
PF08_0081	conserved Plasmodium protein, unknown function
PF10_0177a	serine/threonine protein phosphatase, putative
PF11_0065	40S ribosomal protein S4, putative
PF14_0205	40S ribosomal protein S25, putative
PF14_0363	metacaspase-like protein
PFF0700c	60S ribosomal protein L19, putative

Additional file 18_Oehring et al.

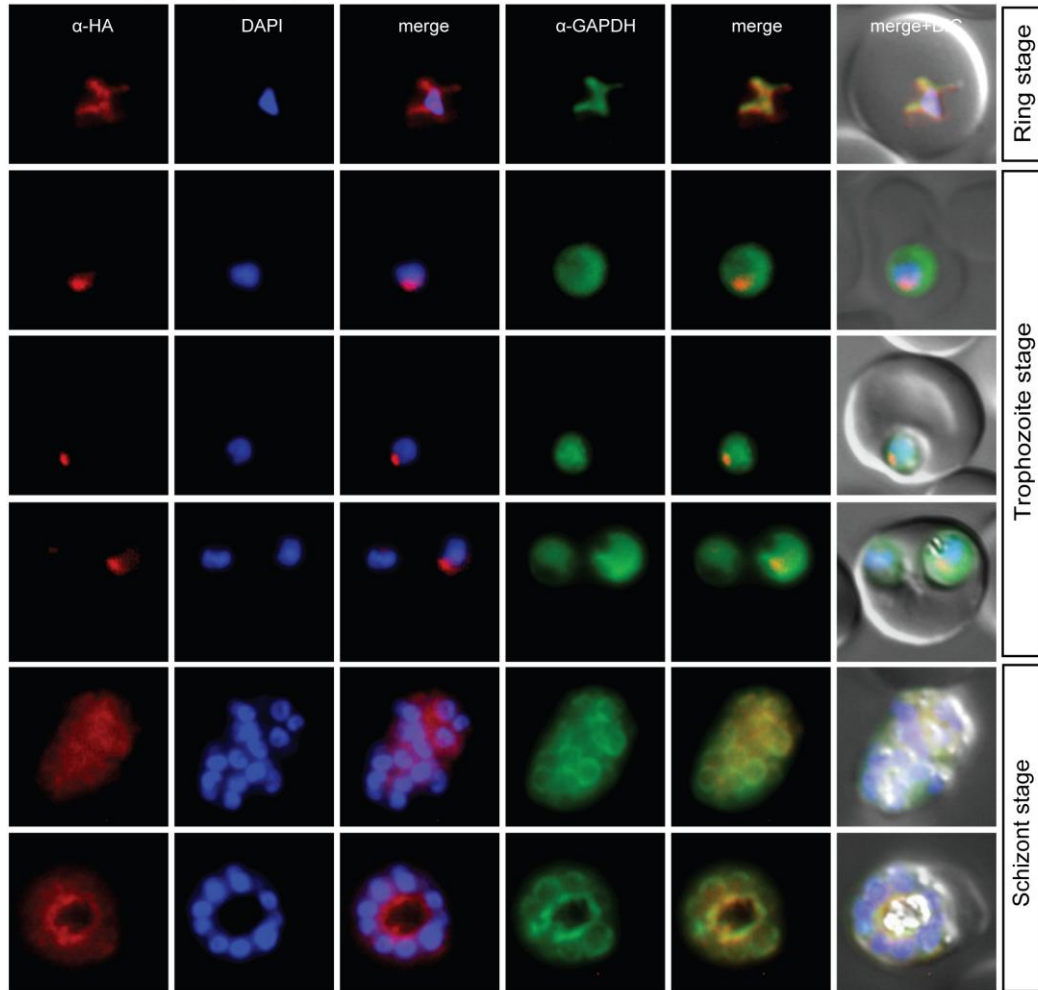
NuProC1 (MAL7P1.38)



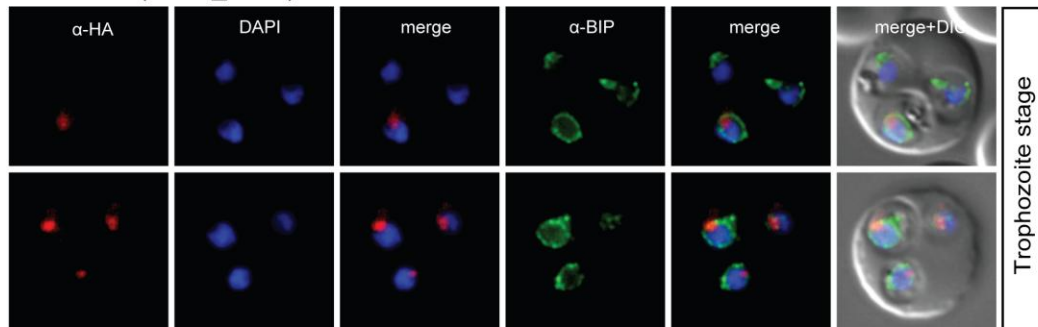
Additional file 18. Localisation of NuProC1-3xHA (MAL7P1.38) during the IDC. Localisation of the tagged protein was visualised using anti-HA antibodies (red). Antibodies against GAPDH were used to visualise the cytosolic compartment. DAPI was used to visualise the nucleus. DIC images are shown as reference.

Additional file 19_Oehring et al.

NuProcC2 (PF10_0278)



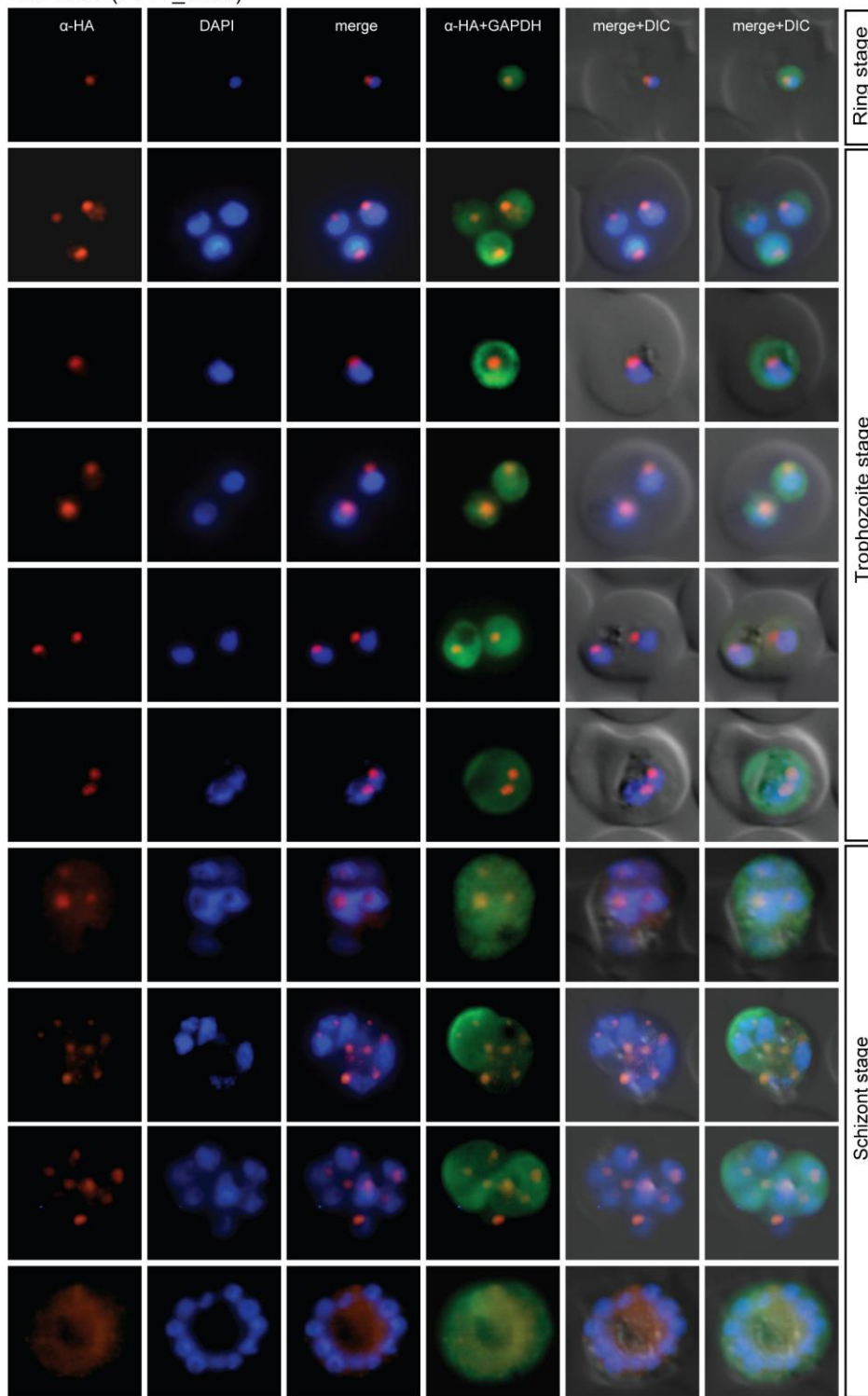
NuProcC2 (PF10_0278)



Additional file 19. Localisation of NuProcC2-3xHA (PF10_0278) during the IDC. Localisation of the tagged protein was visualised using anti-HA antibodies (red). Antibodies against GAPDH and PfBIP were used to visualise the cytosolic (top panel) and ER (bottom panel) compartments, respectively. DAPI was used to visualise the nucleus. DIC images are shown as reference.

Additional file 20_Oehring et al.

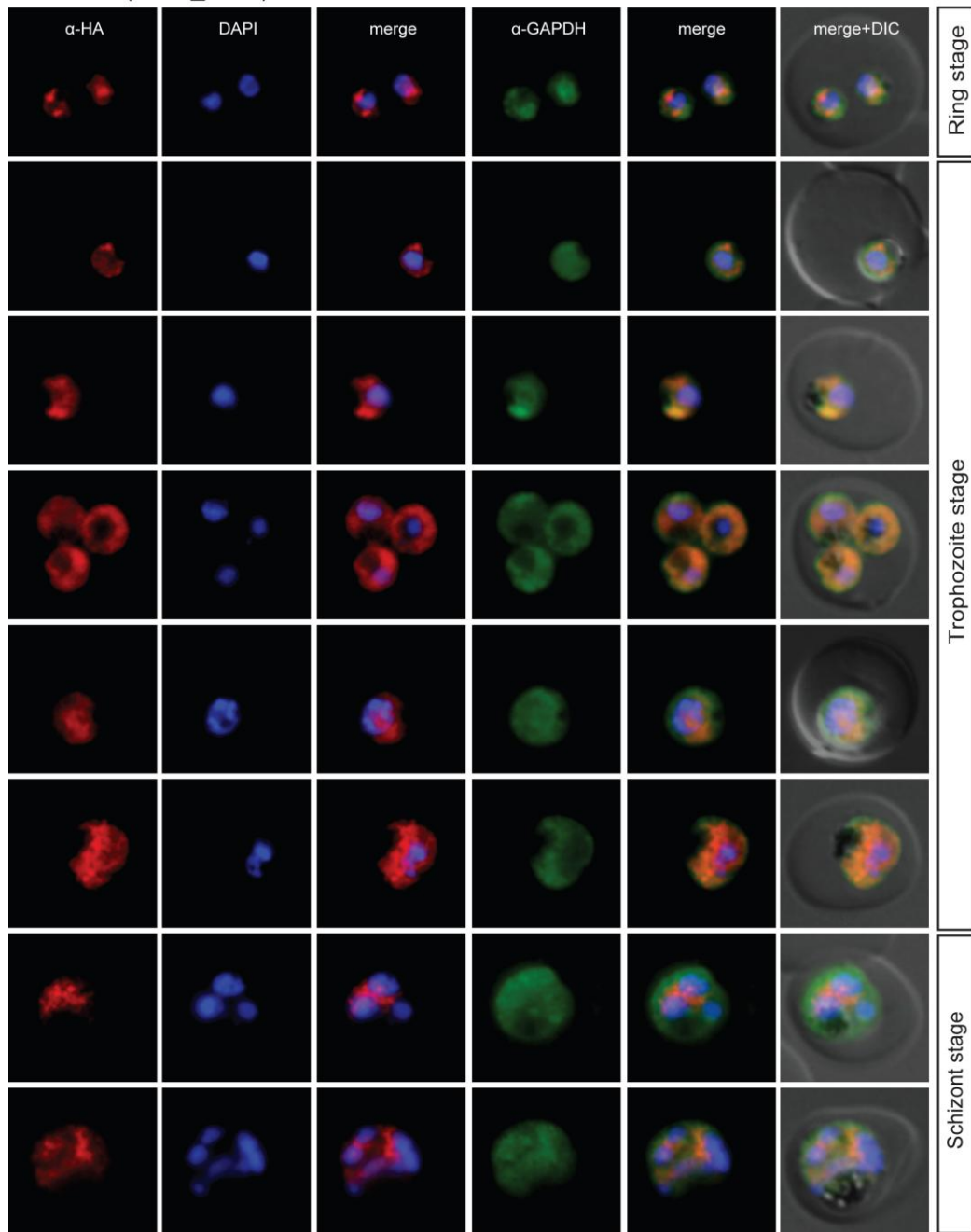
NuProC3 (PF11_0250)



Additional file 20. Localisation of NuProC3-3xHA (PF11_0250) during the IDC. Localisation of the tagged protein was visualised using anti-HA antibodies (red). Antibodies against GAPDH were used to visualise the cytosolic compartment. DAPI was used to visualise the nucleus. DIC images are shown as reference.

Additional file 21_Oehring et al.

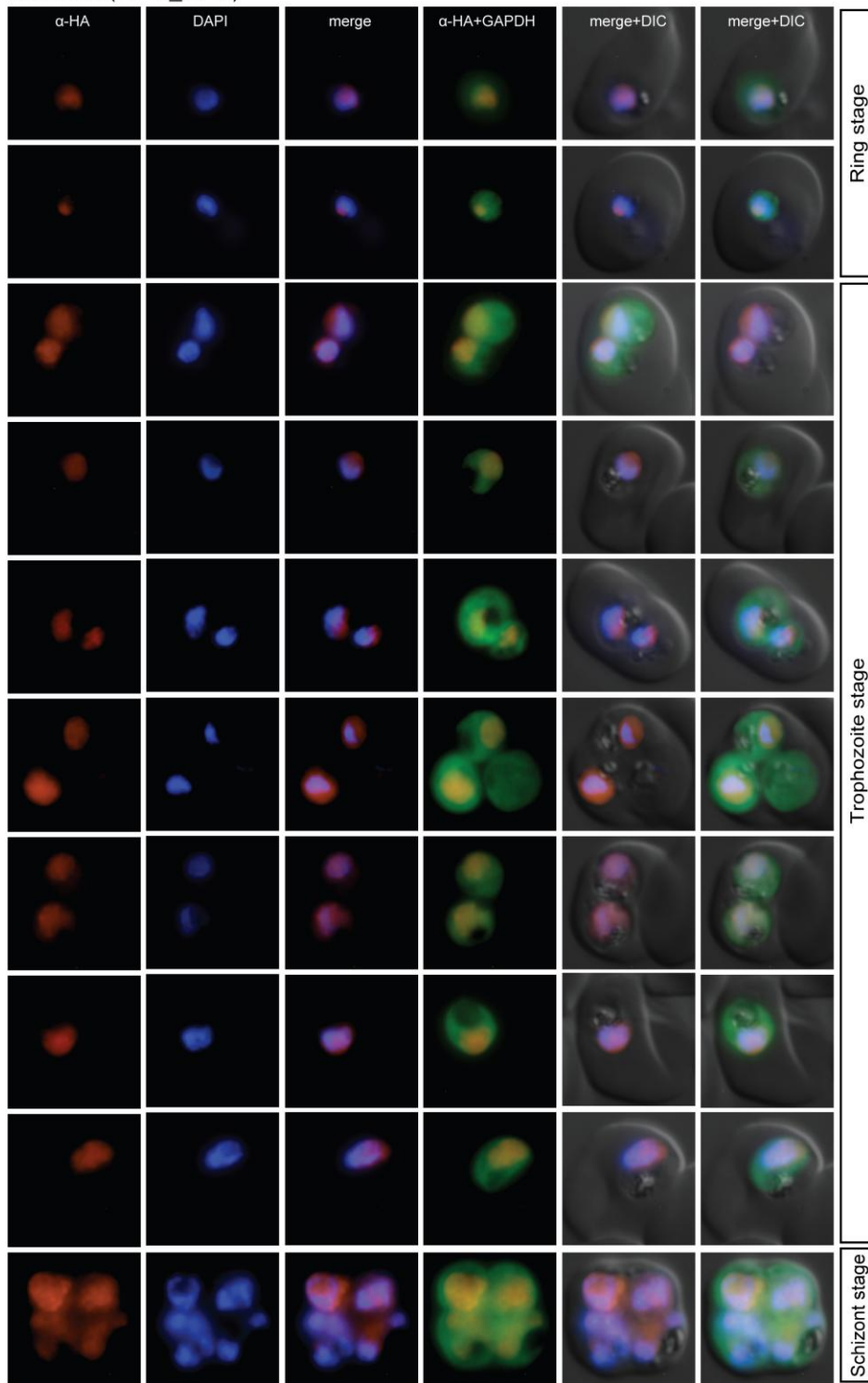
NuProC4 (PF11_0293)



Additional file 21. Localisation of NuProC4-3xHA (PF11_0293) during the IDC. Localisation of the tagged protein was visualised using anti-HA antibodies (red). Antibodies against GAPDH were used to visualise the cytosolic compartment. DAPI was used to visualise the nucleus. DIC images are shown as reference.

Additional file 22_Oehring et al.

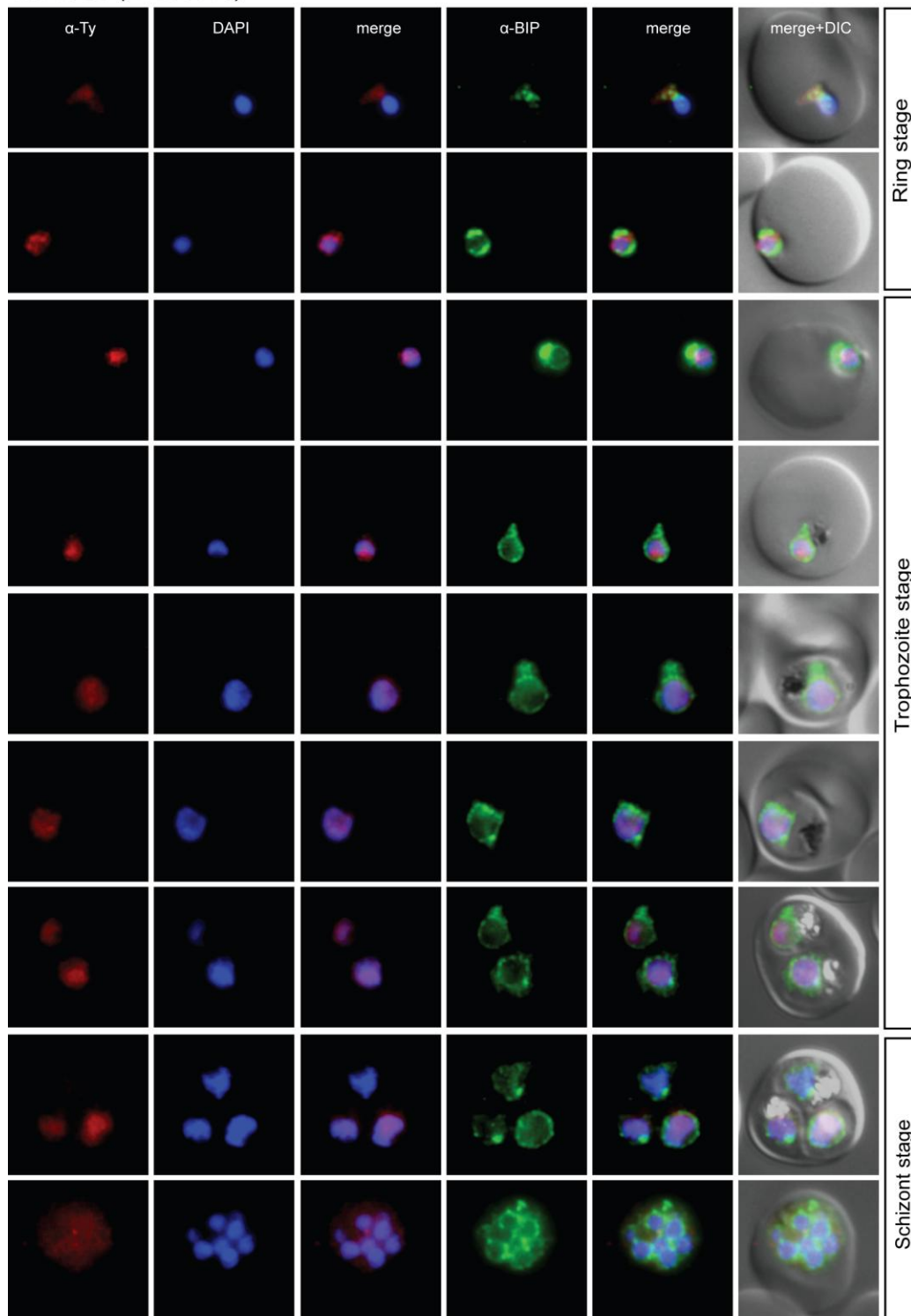
NuProC5 (PF13_0042)



Additional file 22. Localisation of NuProC5-3xHA (PF13_0042) during the IDC. Localisation of the tagged protein was visualised using anti-HA antibodies (red). Antibodies against GAPDH were used to visualise the cytosolic compartment. DAPI was used to visualise the nucleus. DIC images are shown as reference.

Additional file 23_Oehring et al.

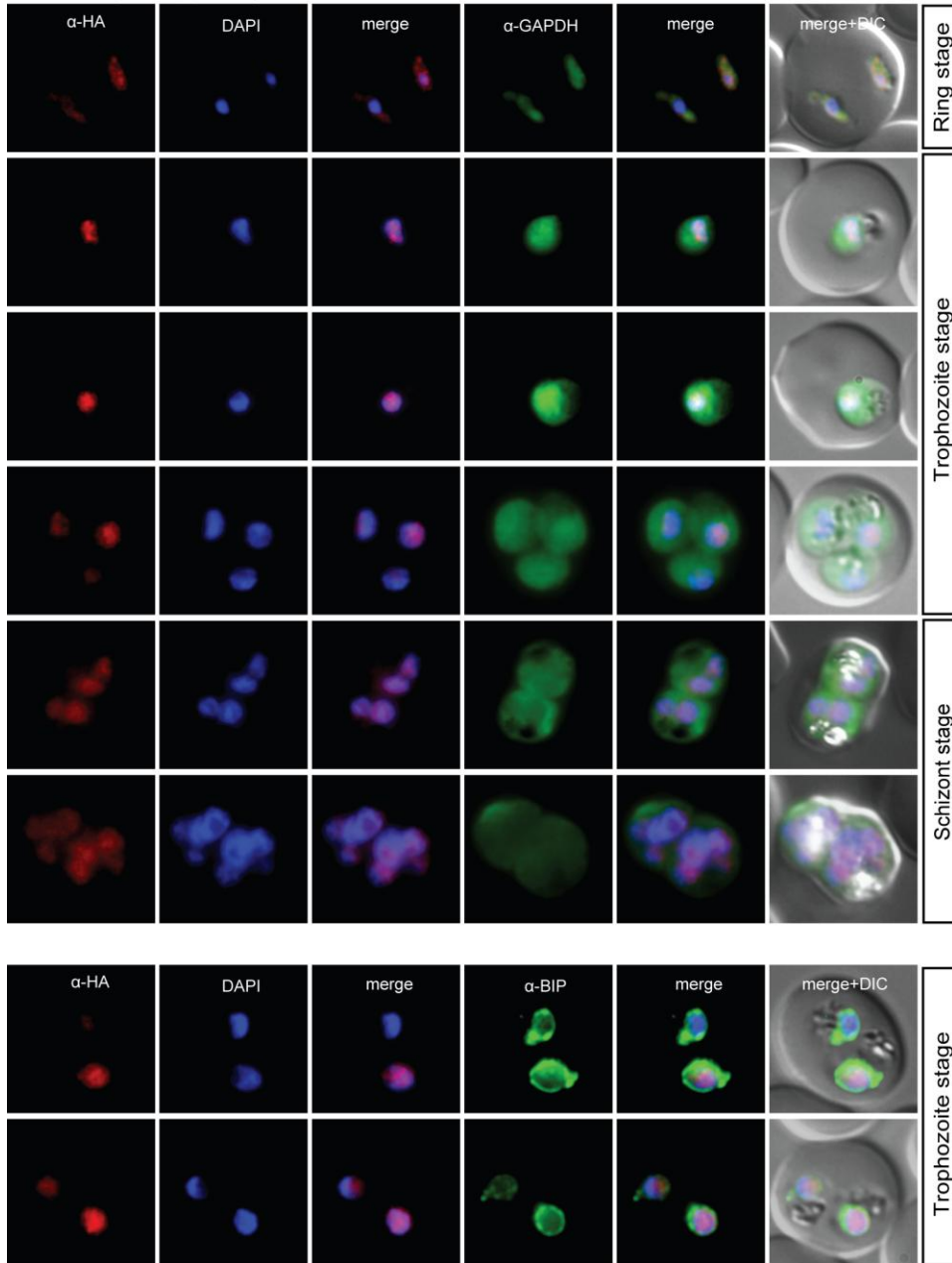
NuProC6 (PFL0635c)



Additional file 23. Localisation of NuProC6-2xTy (PFL0625c) during the IDC. Localisation of the tagged protein was visualised using anti-Ty antibodies (red). Antibodies against PFBIP were used to visualise the ER. DAPI was used to visualise the nucleus. DIC images are shown as reference.

Additional file 24_Oehring et al.

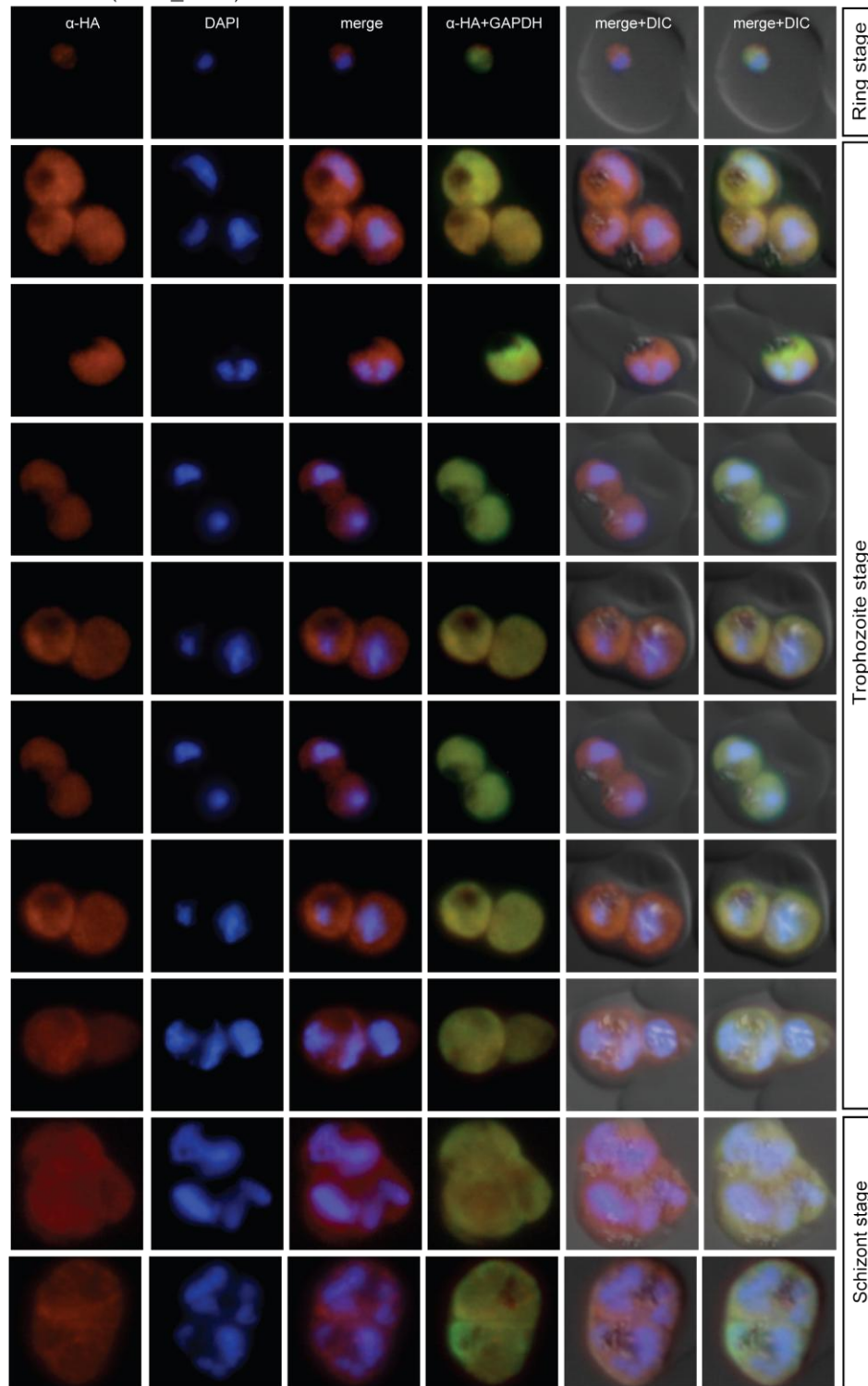
NuProc7 (PF14_0393)



Additional file 24. Localisation of NuProc7-3xHA (PF14_0393) during the IDC. Localisation of the tagged protein was visualised using anti-HA antibodies (red). Antibodies against GAPDH were used to visualise the cytosolic compartment. Antibodies against P/BIP were used to visualise the ER (bottom). DAPI was used to visualise the nucleus. DIC images are shown as reference.

Additional file 25_Oehring et al.

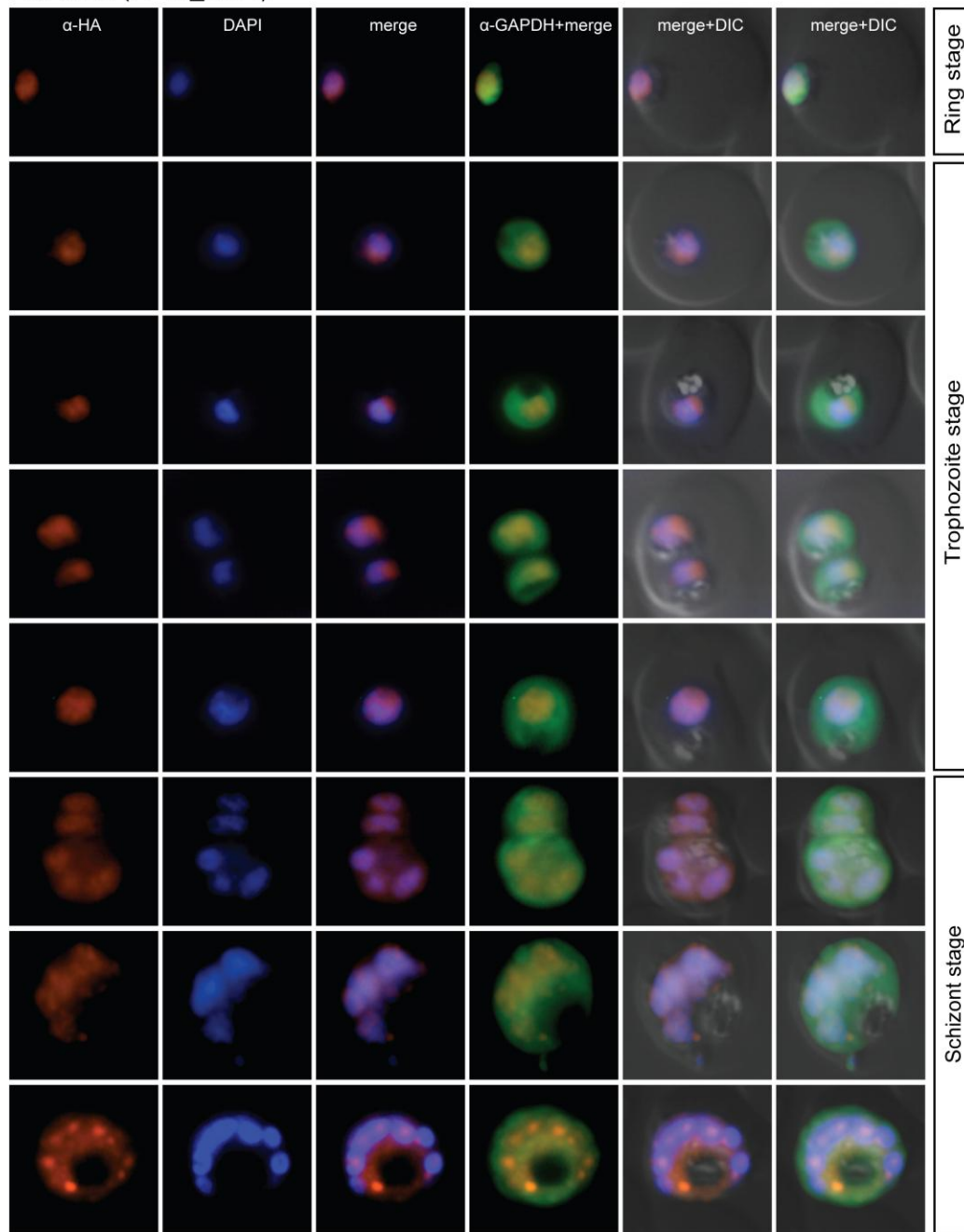
NuProC8 (PF08_0083)



Additional file 25. Localisation of NuProC8-3xHA (PF08_0083) during the IDC. Localisation of the tagged protein was visualised using anti-HA antibodies (red). Antibodies against GAPDH were used to visualise the cytosolic compartment. DAPI was used to visualise the nucleus. DIC images are shown as reference.

Additional file 26_Oehring et al.

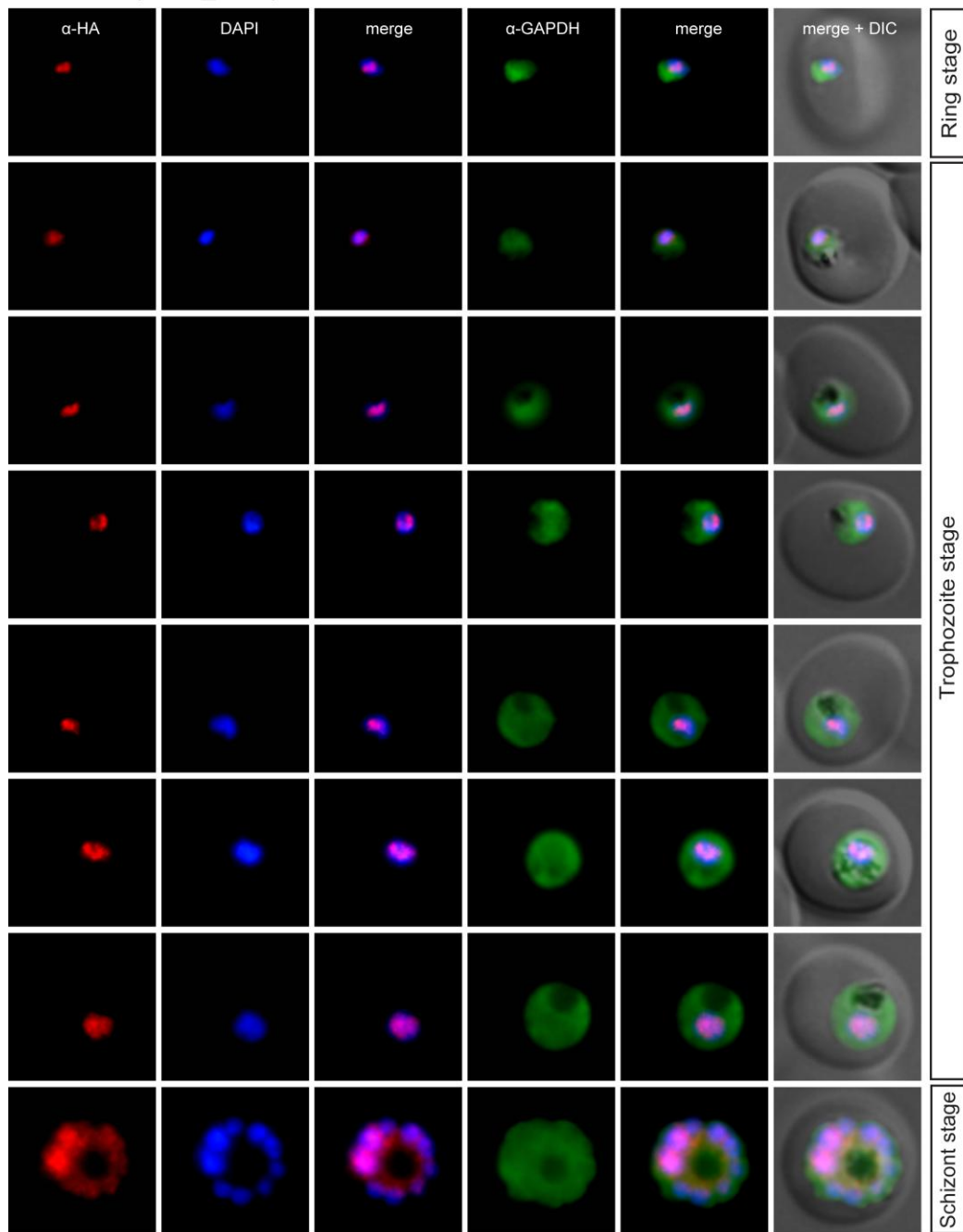
NuProC9 (PF10_0091)



Additional file 26. Localisation of NuProC9-3xHA (PF10_0091) during the IDC. Localisation of the tagged protein was visualised using anti-HA antibodies (red). Antibodies against GAPDH were used to visualise the cytosolic compartment. DAPI was used to visualise the nucleus. DIC images are shown as reference.

Additional file 27_Oehring et al.

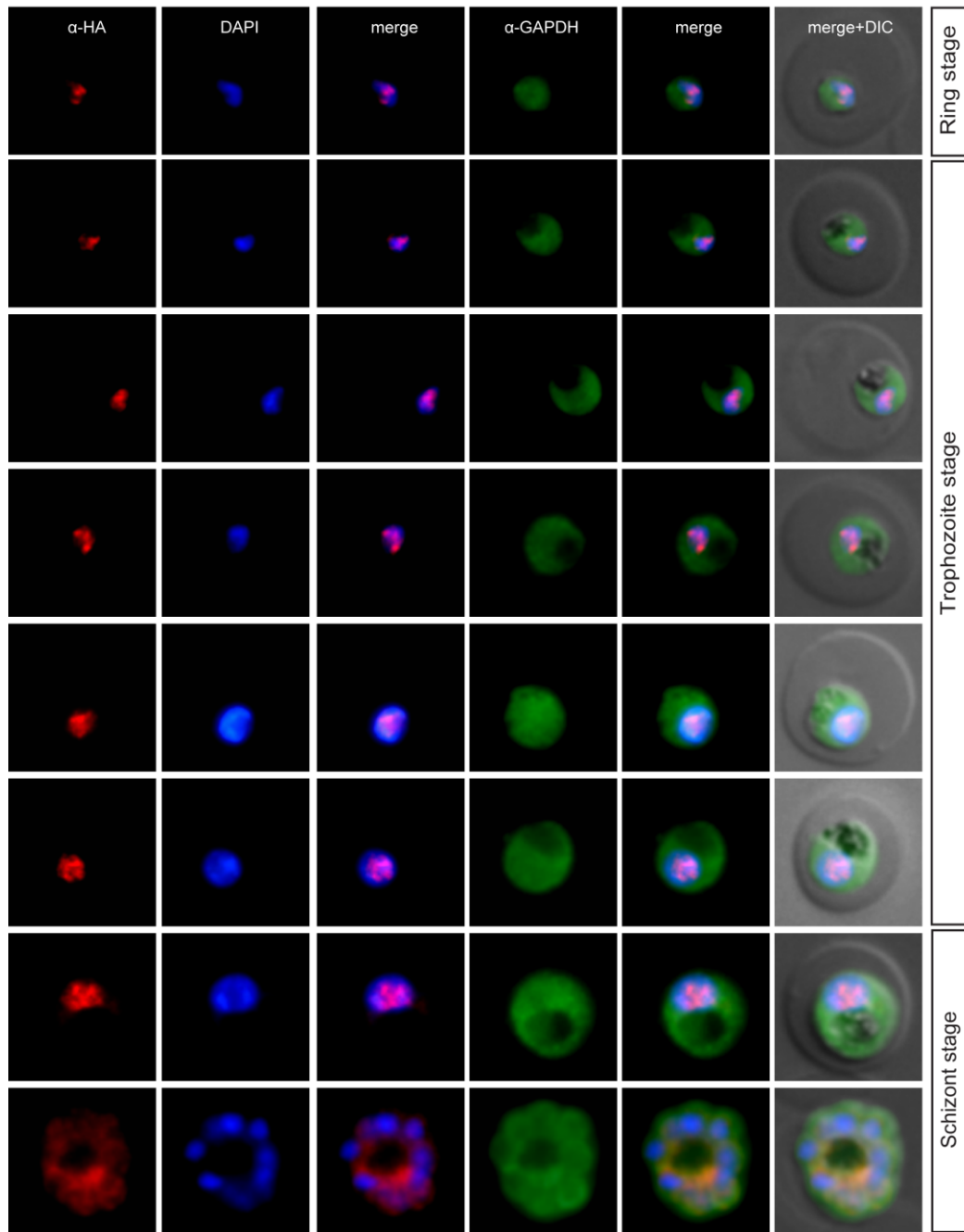
NuProC10 (PF10_0328)



Additional file 27. Localisation of NuProC10-3xHA (PF10_0328) during the IDC. Localisation of the tagged protein was visualised using anti-HA antibodies (red). Antibodies against GAPDH were used to visualise the cytosolic compartment. DAPI was used to visualise the nucleus. DIC images are shown as reference.

Additional file 28_Oehring et al.

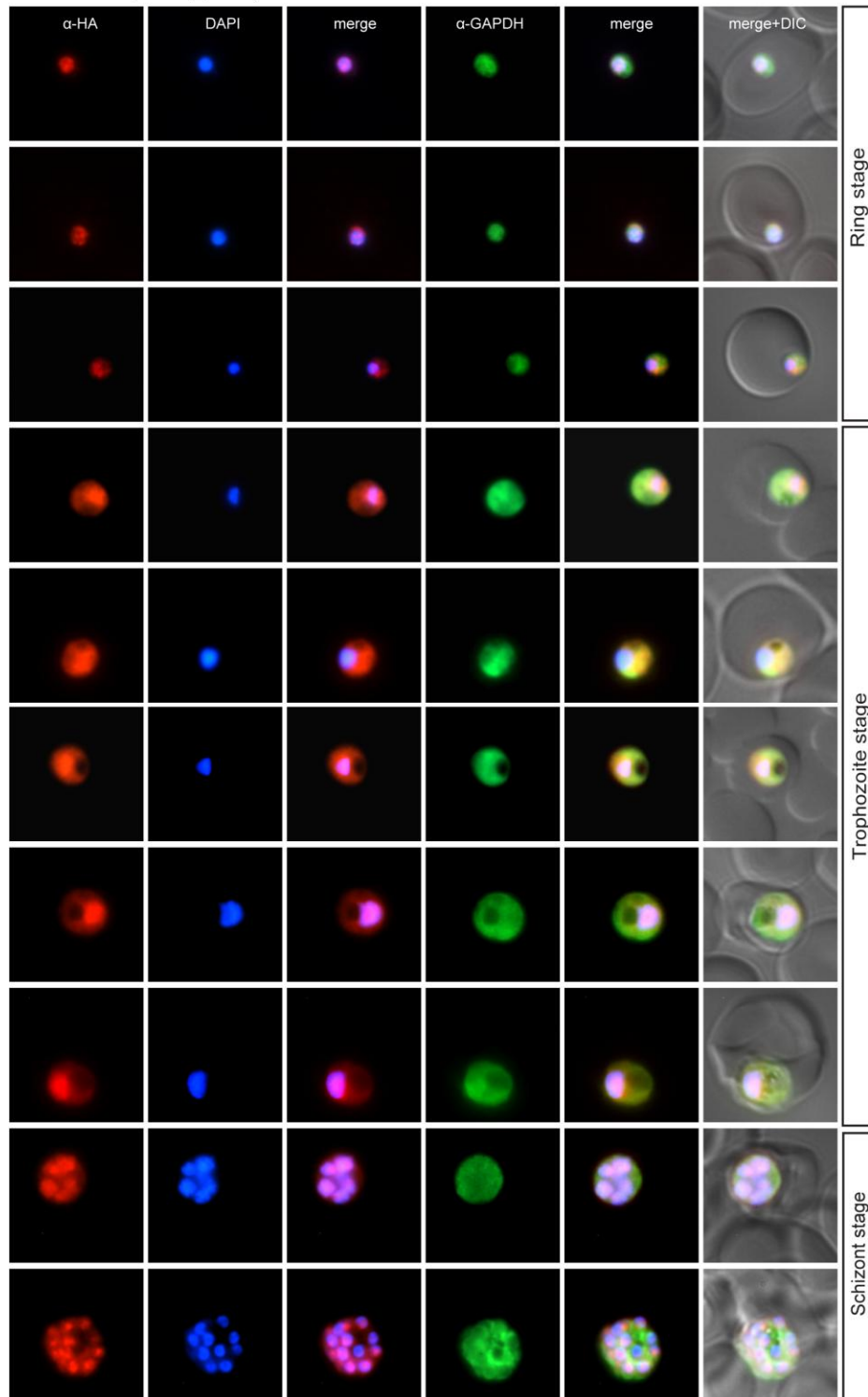
NuProC11 (PF11_0254)



Additional file 28. Localisation of NuProC11-3xHA (PF11_0254) during the IDC. Localisation of the tagged protein was visualised using anti-HA antibodies (red). Antibodies against GAPDH were used to visualise the cytosolic compartment. DAPI was used to visualise the nucleus. DIC images are shown as reference.

Additional file 29_Oehring et al.

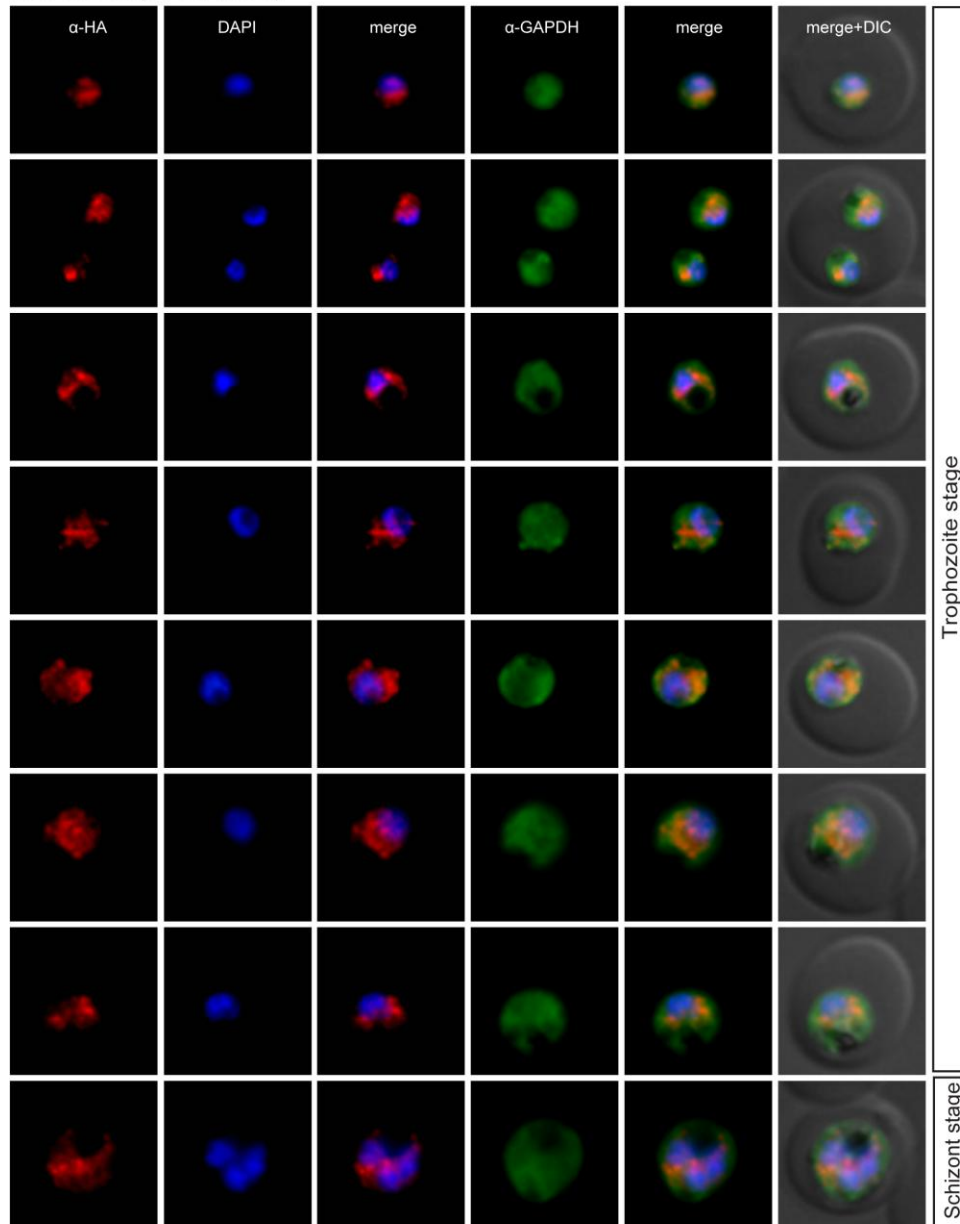
NuProC 12 (PF11_0332)



Additional file 29. Localisation of NuProC12-3xHA (PF11_0332) during the IDC. Localisation of the tagged protein was visualised using anti-HA antibodies (red). Antibodies against GAPDH were used to visualise the cytosolic compartment. DAPI was used to visualise the nucleus. DIC images are shown as reference.

Additional file 30_Oehring et al.

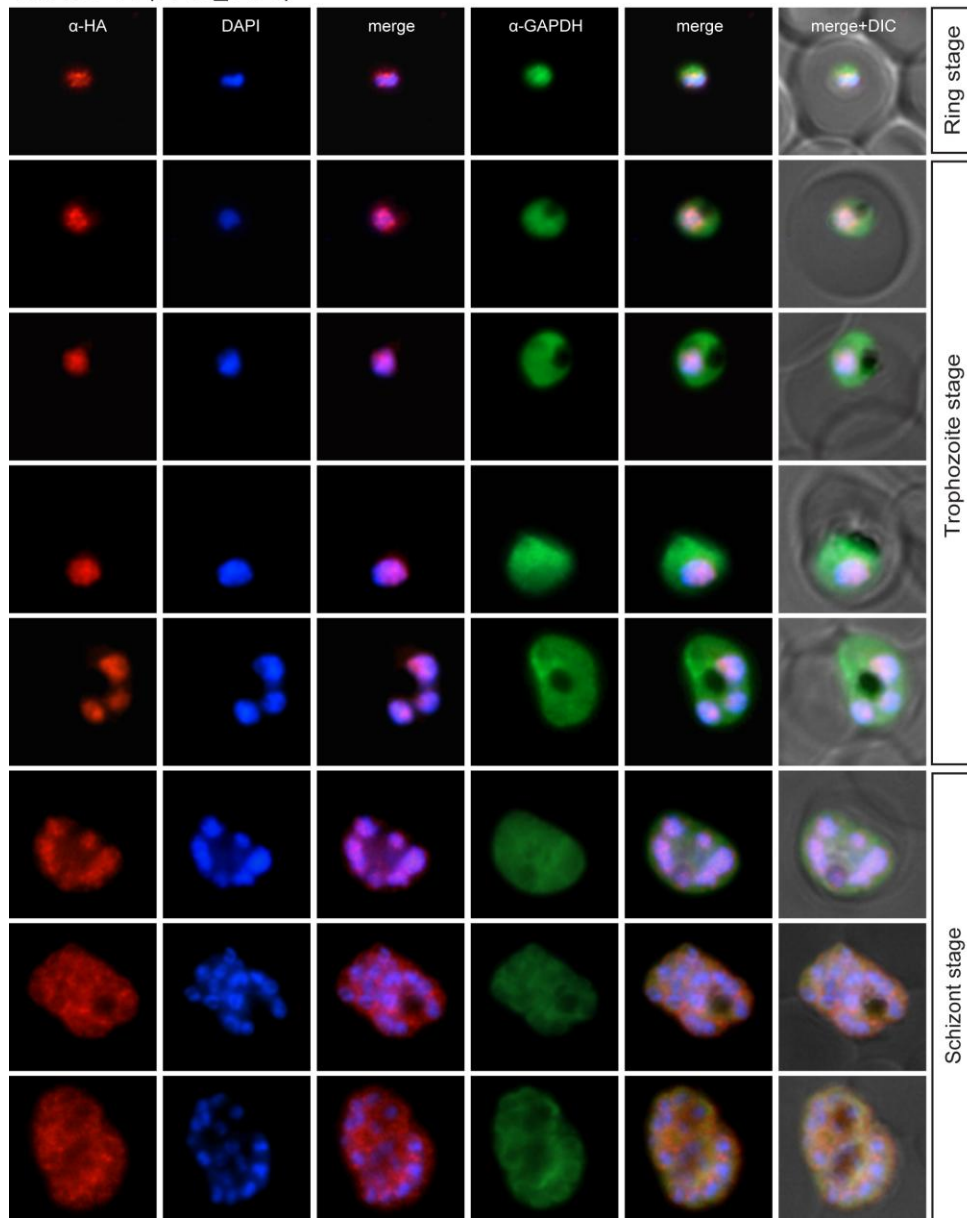
NuProC13 (PF13_0099)



Additional file 30. Localisation of NuProC13-3xHA (PF13_0099) during the IDC. Localisation of the tagged protein was visualised using anti-HA antibodies (red). Antibodies against GAPDH were used to visualise the cytosolic compartment. DAPI was used to visualise the nucleus. DIC images are shown as reference.

Additional file 31_Oehring et al.

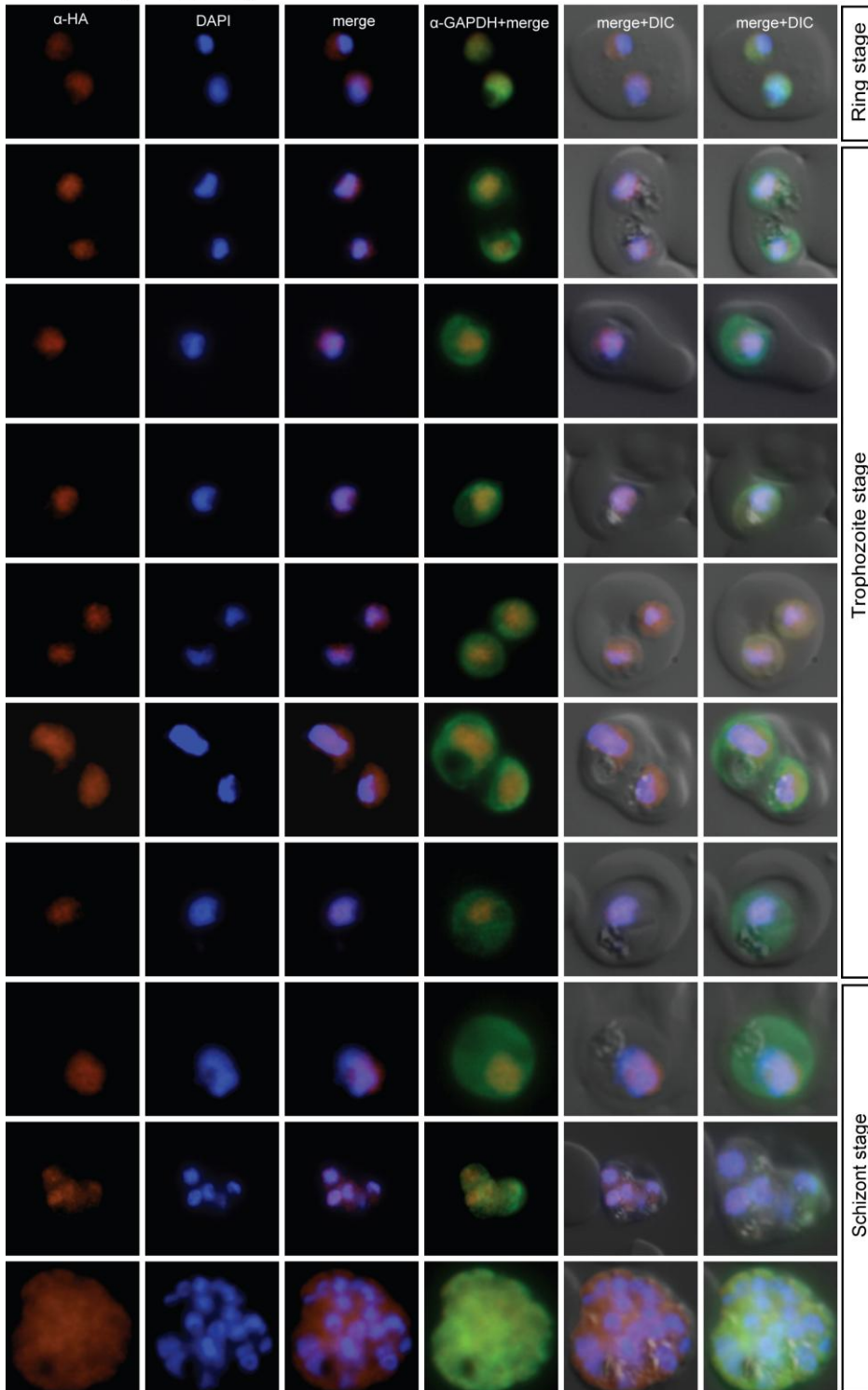
NuProC 14 (PF14_0176)



Additional file 31. Localisation of NuProC14-3xHA (PF14_0176) during the IDC. Localisation of the tagged protein was visualised using anti-HA antibodies (red). Antibodies against GAPDH were used to visualise the cytosolic compartment. DAPI was used to visualise the nucleus. DIC images are shown as reference.

Additional file 32

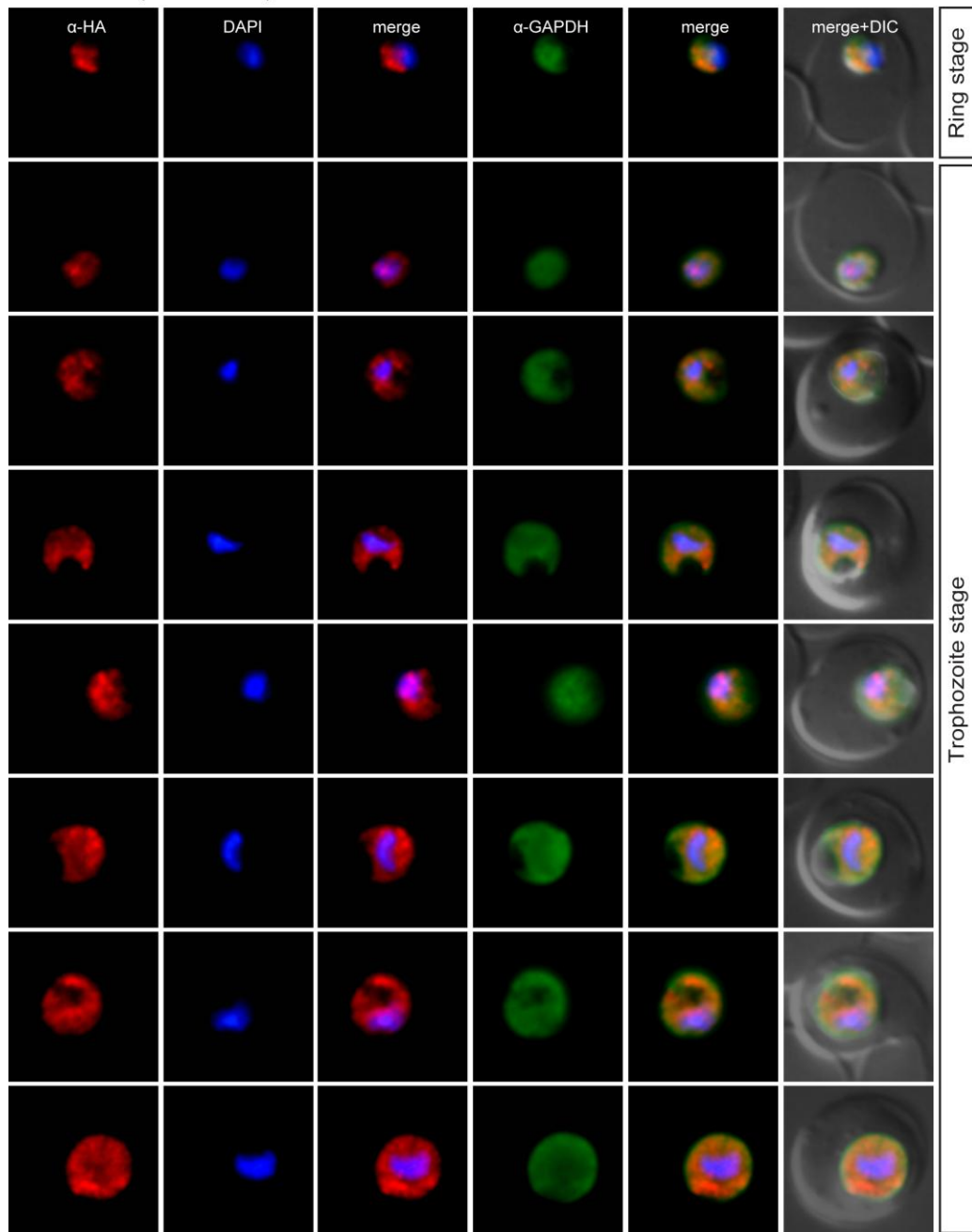
NuProC15 (PF14_0433)



Additional file 32. Localisation of NuProC15-3xHA (PF14_0433) during the IDC. Localisation of the tagged protein was visualised using anti-HA antibodies (red). Antibodies against GAPDH were used to visualise the cytosolic compartment. DAPI was used to visualise the nucleus. DIC images are shown as reference.

Additional file 33

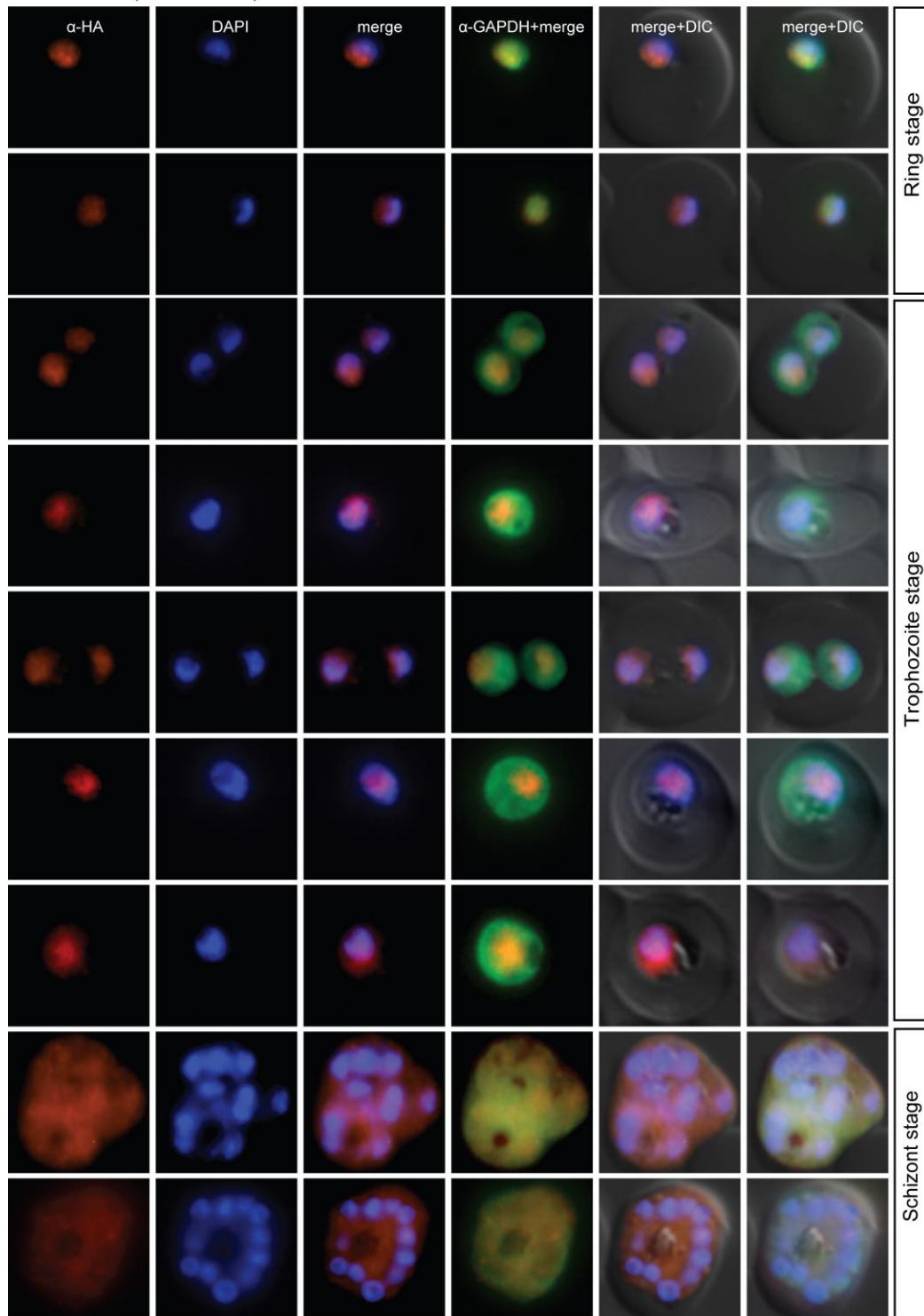
NuProC16 (PFC0126c)



Additional file 33. Localisation of NuProC16-3xHA (PFC0126c) during the IDC. Localisation of the tagged protein was visualised using anti-HA antibodies (red). Antibodies against GAPDH were used to visualise the cytosolic compartment. DAPI was used to visualise the nucleus. DIC images are shown as reference.

Additional file 34_Oehring et al.

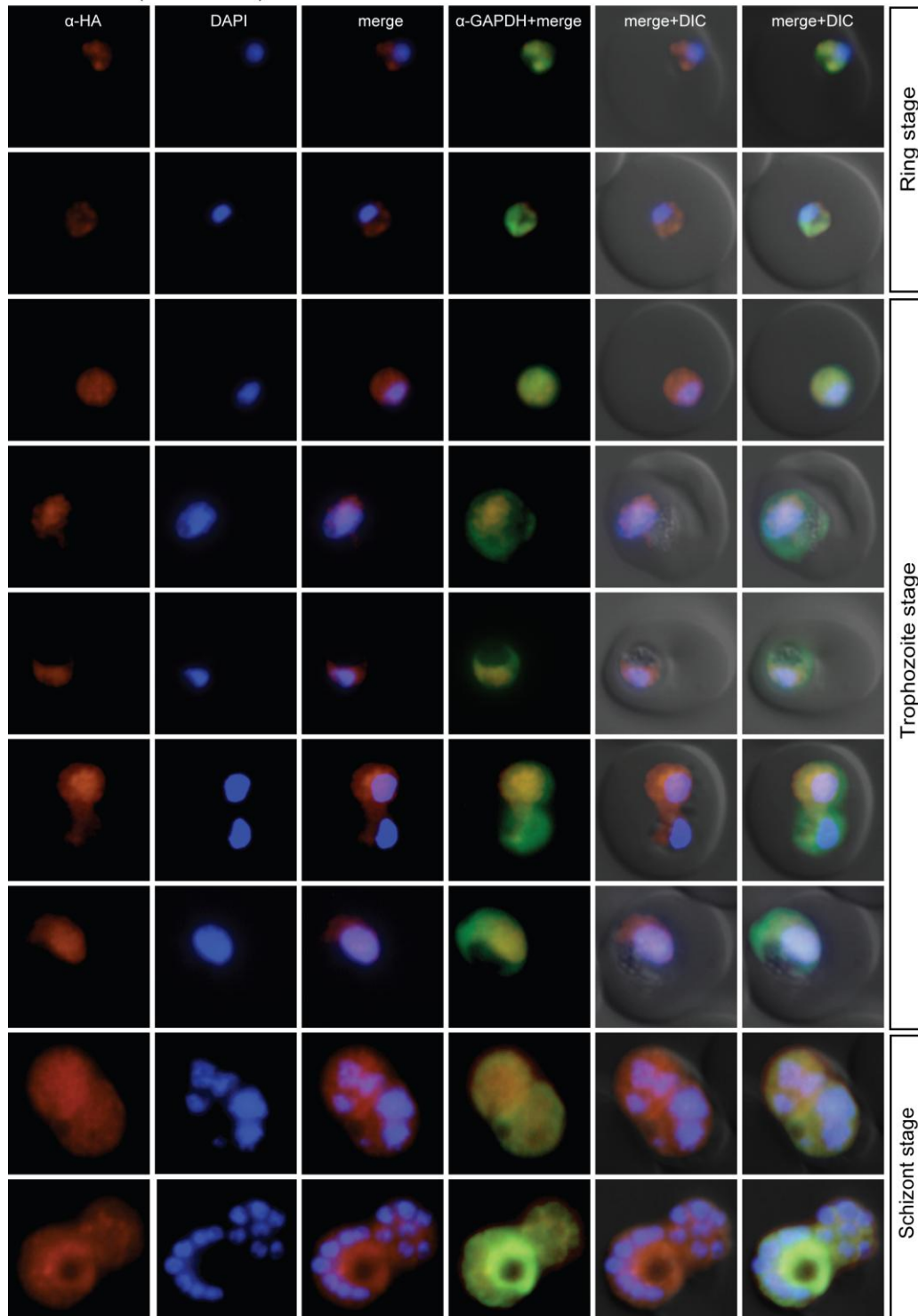
NuProC17 (PFC0130c)



Additional file 34. Localisation of NuProC17-3xHA (PFC0130c) during the IDC. Localisation of the tagged protein was visualised using anti-HA antibodies (red). Antibodies against GAPDH were used to visualise the cytosolic compartment. DAPI was used to visualise the nucleus. DIC images are shown as reference.

Additional file 35_Oehring et al.

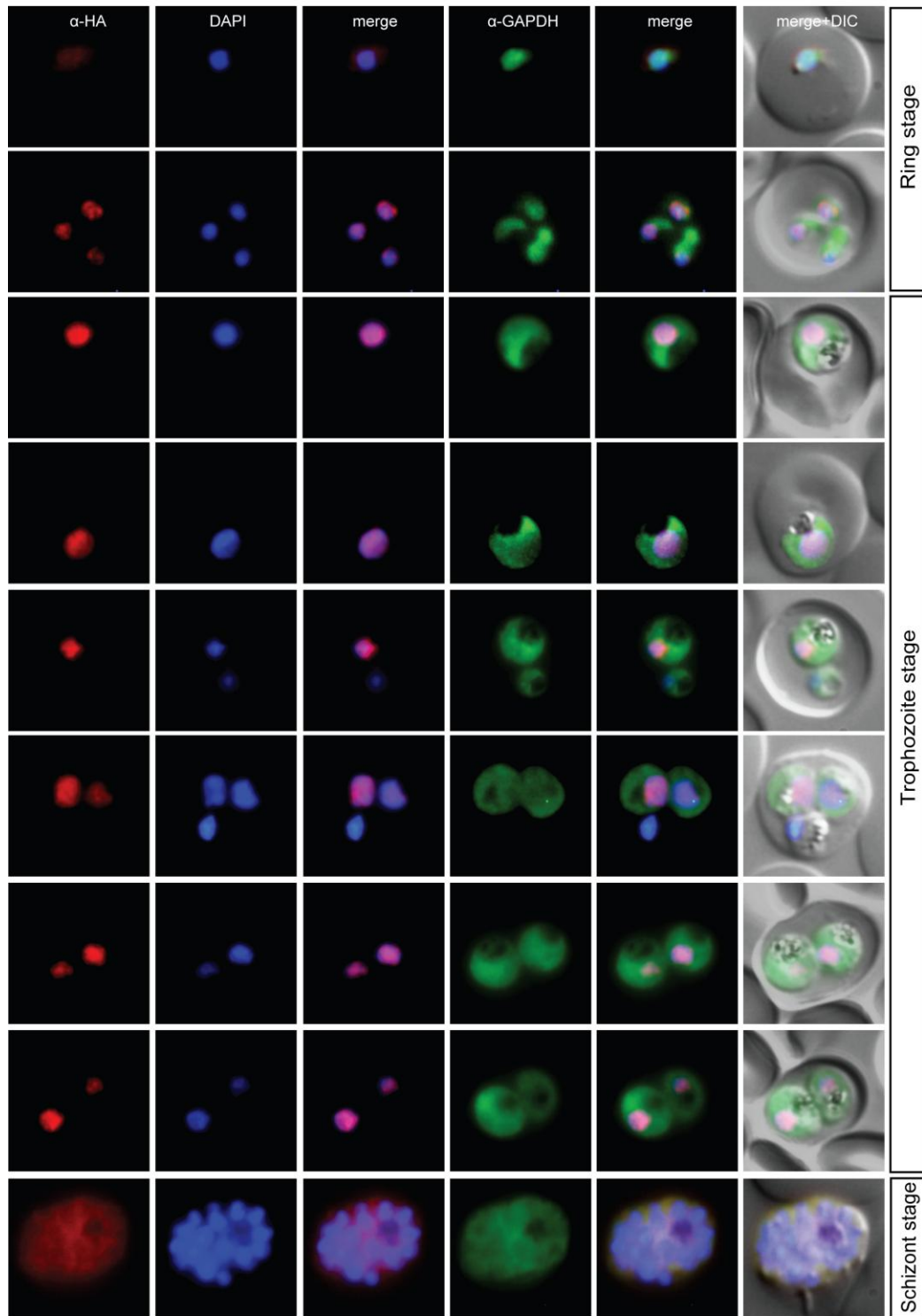
NuProC18 (PFC0690c)



Additional file 35. Localisation of NuProC18-3xHA (PFC0690c) during the IDC. Localisation of the tagged protein was visualised using anti-HA antibodies (red). Antibodies against GAPDH were used to visualise the cytosolic compartment. DAPI was used to visualise the nucleus. DIC images are shown as reference.

Additional file 36_Oehring et al.

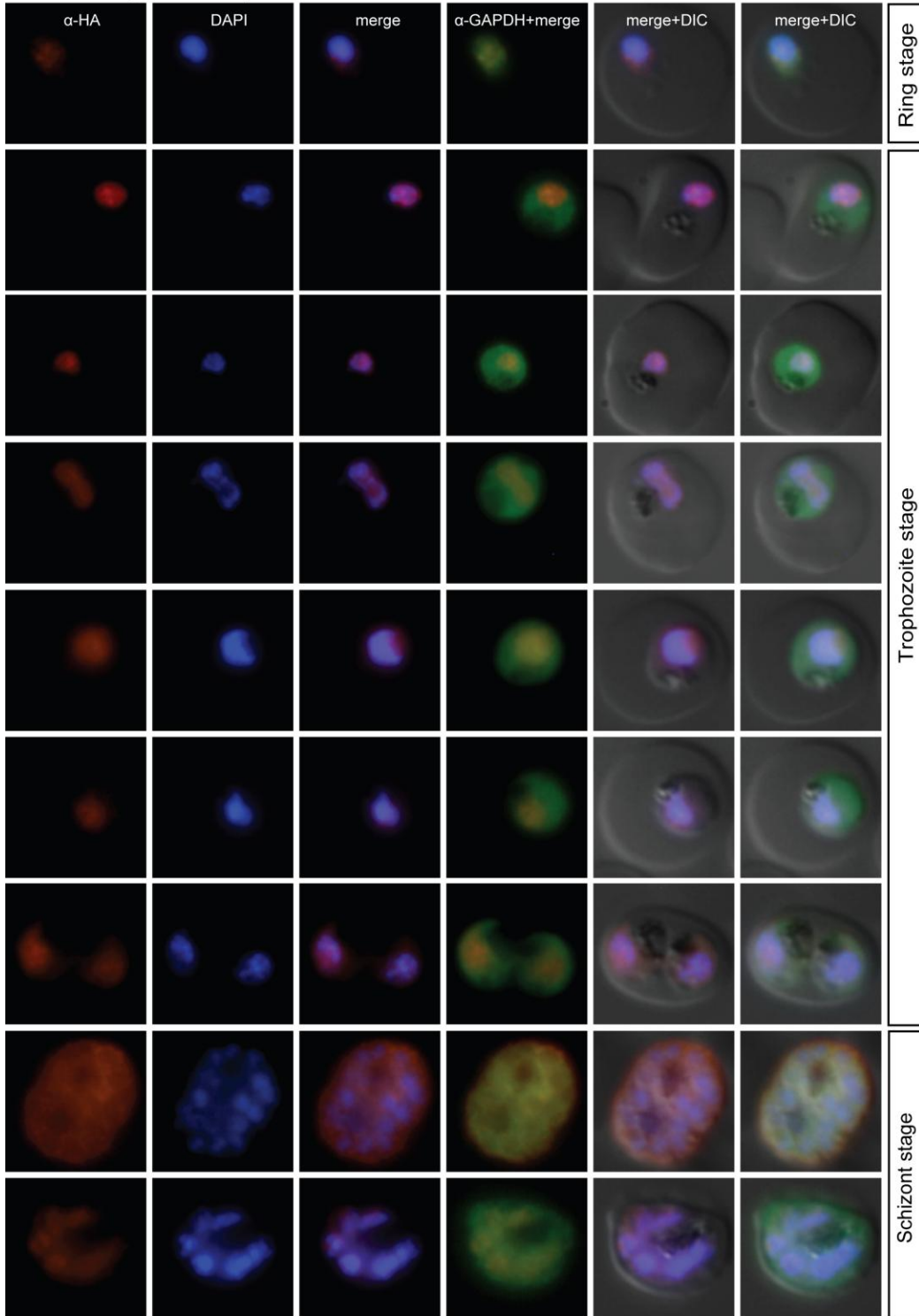
NuProC19 (PFI0610w)



Additional file 36. Localisation of NuProC19-3xHA (PFI0610w) during the IDC. Localisation of the tagged protein was visualised using anti-HA antibodies (red). Antibodies against GAPDH were used to visualise the cytosolic compartment. DAPI was used to visualise the nucleus. DIC images are shown as reference.

Additional file 37_Oehring et al.

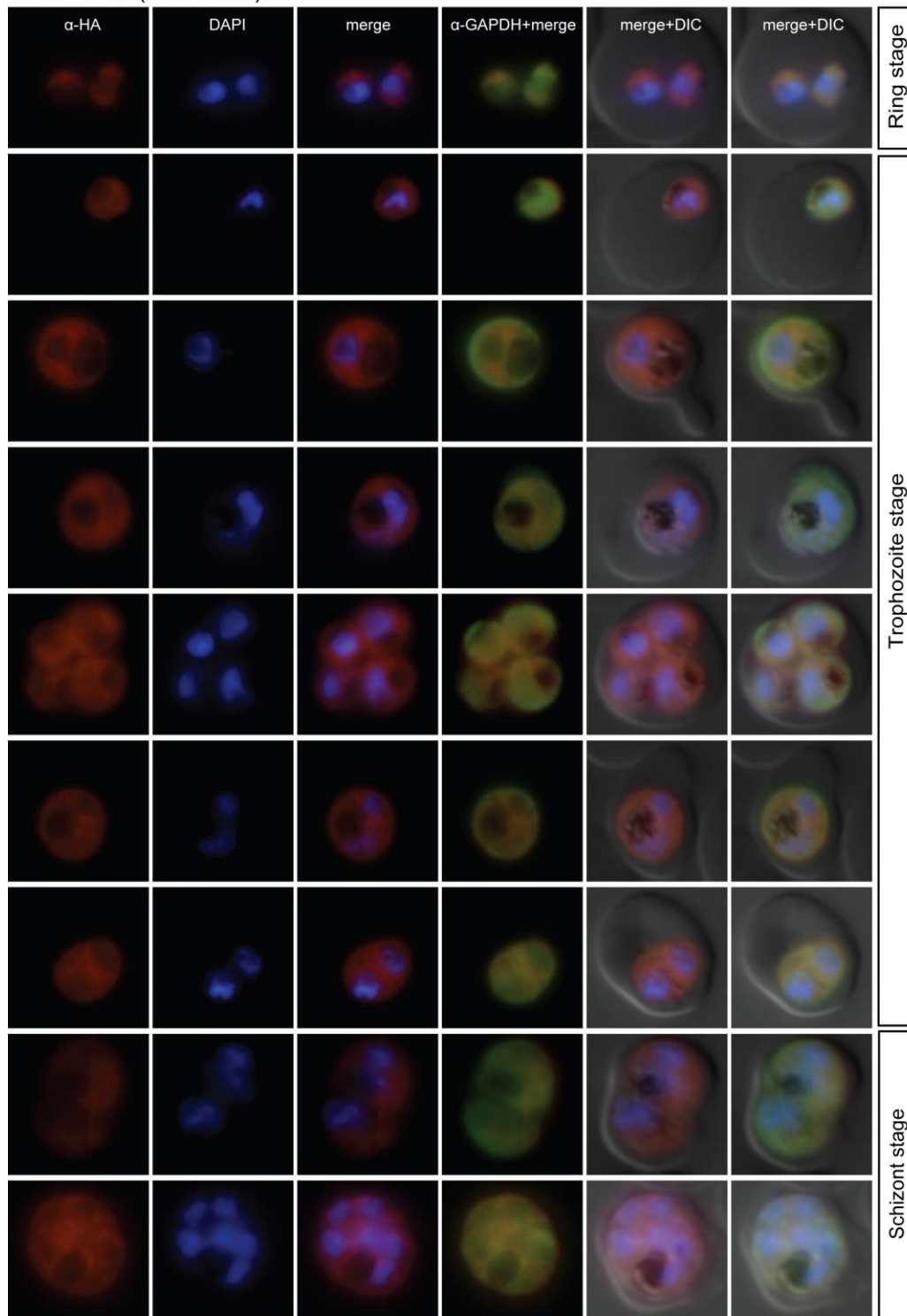
NuProC20 (PF11355w)



Additional file 37. Localisation of NuProC20-3xHA (PF11355w) during the IDC. Localisation of the tagged protein was visualised using anti-HA antibodies (red). Antibodies against GAPDH were used to visualise the cytosolic compartment. DAPI was used to visualise the nucleus. DIC images are shown as reference.

Additional file 38_Oehring et al.

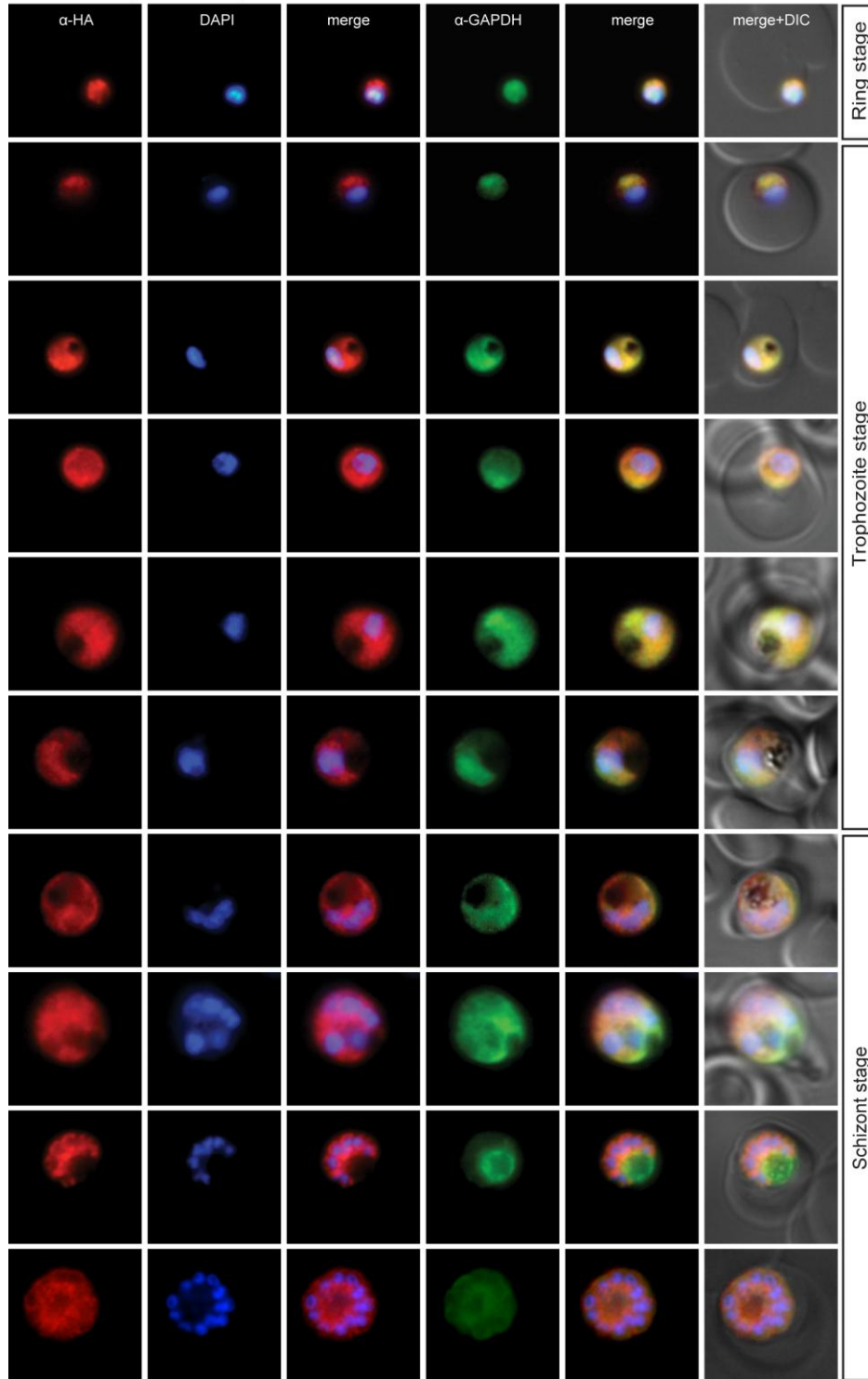
NuProC21 (PFL0185c)



Additional file 38. Localisation of NuProC21-3xHA (PFL0185c) during the IDC. Localisation of the tagged protein was visualised using anti-HA antibodies (red). Antibodies against GAPDH were used to visualise the cytosolic compartment. DAPI was used to visualise the nucleus. DIC images are shown as reference.

Additional file 39_Oehring et al.

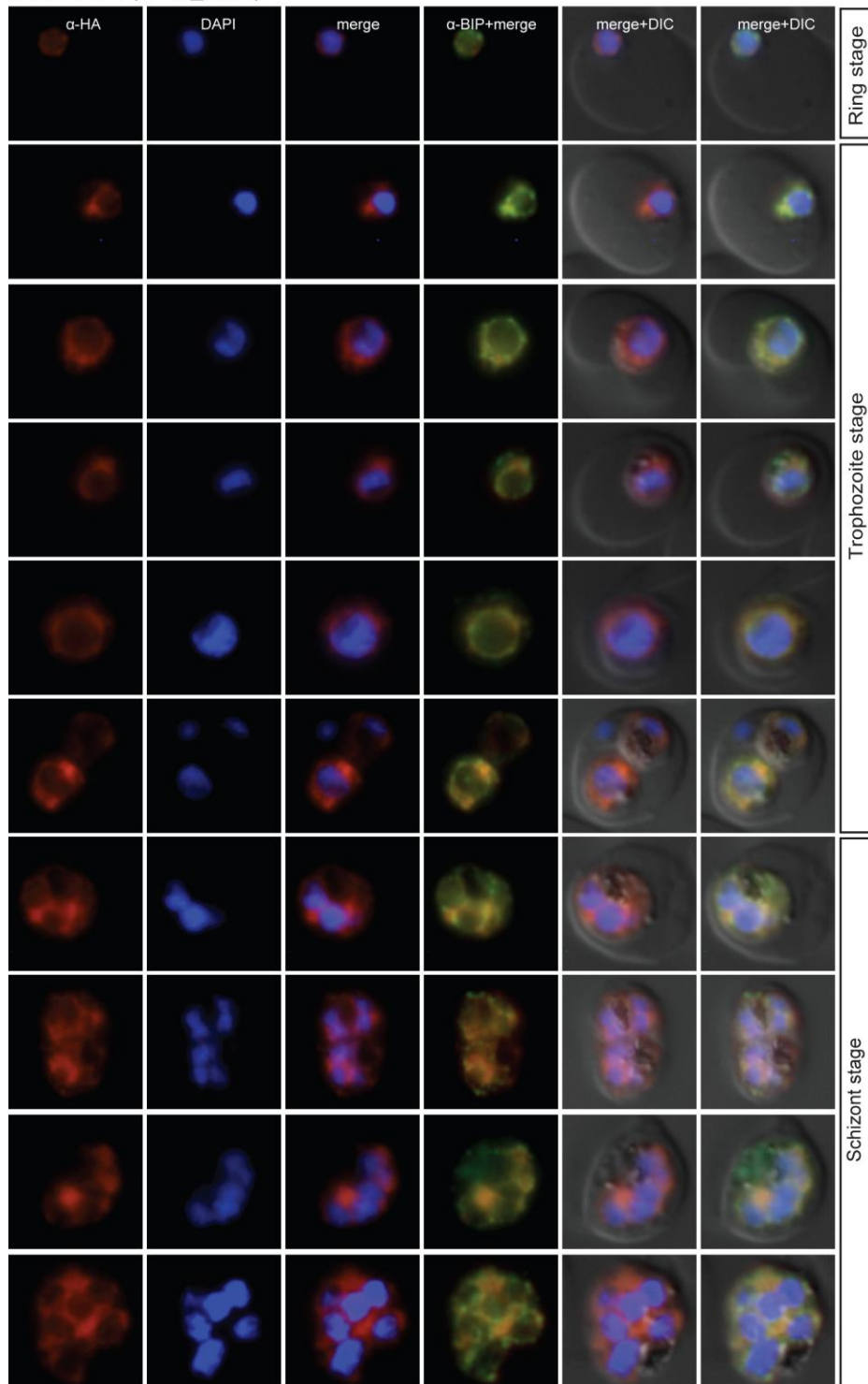
NuProC 22 (PFL0450c)



Additional file 39. Localisation of NuProC22-3xHA (PFL0450c) during the IDC. Localisation of the tagged protein was visualised using anti-HA antibodies (red). Antibodies against GAPDH were used to visualise the cytosolic compartment. DAPI was used to visualise the nucleus. DIC images are shown as reference.

Additional file 40_Oehring et al.

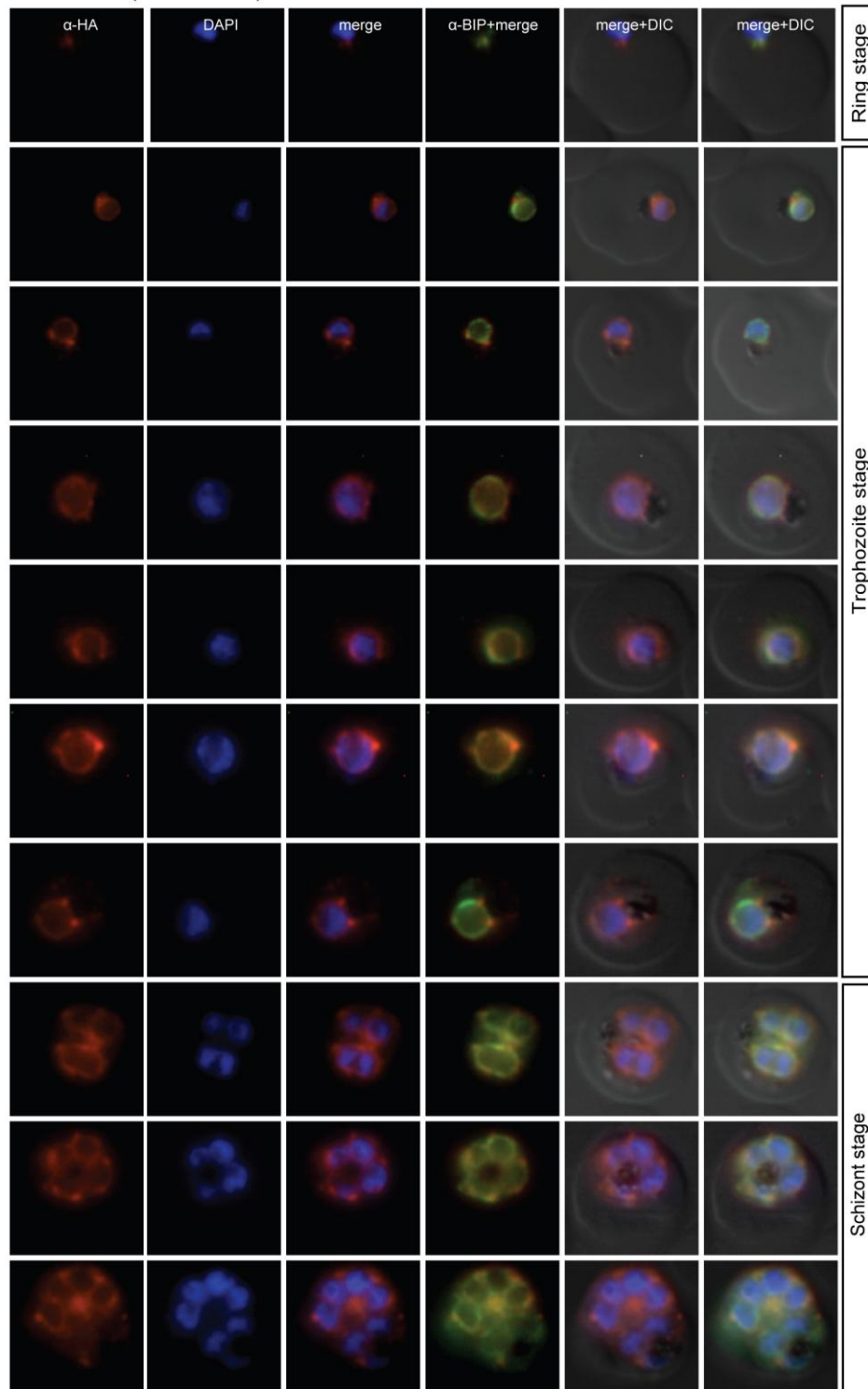
n-NuProC1 (PF11_0099)



Additional file 40. Localisation of n-NuProC1-3xHA (PF11_0099) during the IDC. Localisation of the tagged protein was visualised using anti-HA antibodies (red). Antibodies against PfBIP were used to visualise the ER. DAPI was used to visualise the nucleus. DIC images are shown as reference.

Additional file 41_Oehring et al.

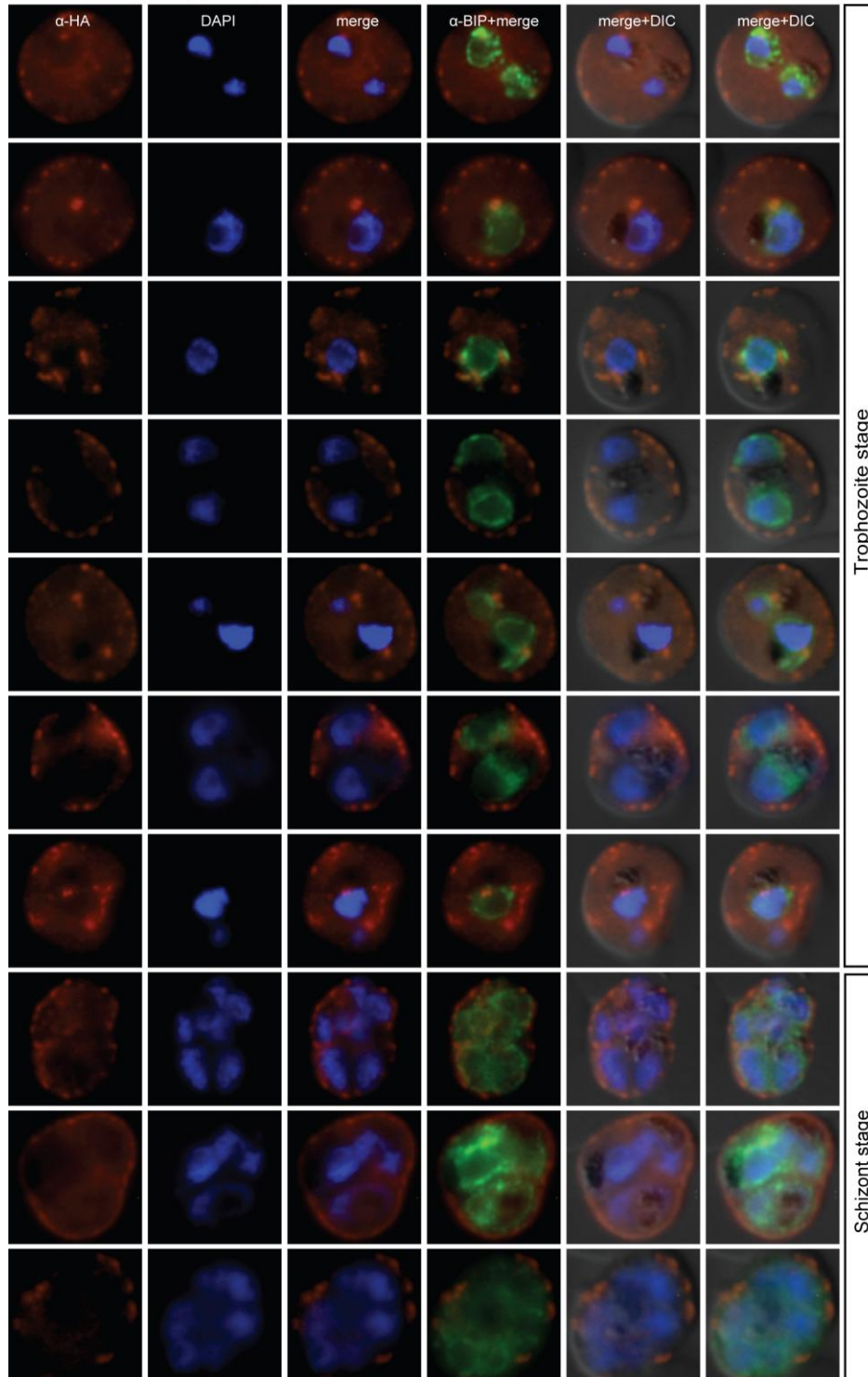
n-NuProC2 (MAL7P1.77)



Additional file 41. Localisation of n-NuProC2-3xHA (MAL7P1.77) during the IDC. Localisation of the tagged protein was visualised using anti-HA antibodies (red). Antibodies against PfBIP were used to visualise the ER. DAPI was used to visualise the nucleus. DIC images are shown as reference.

Additional file 42_Oehring et al.

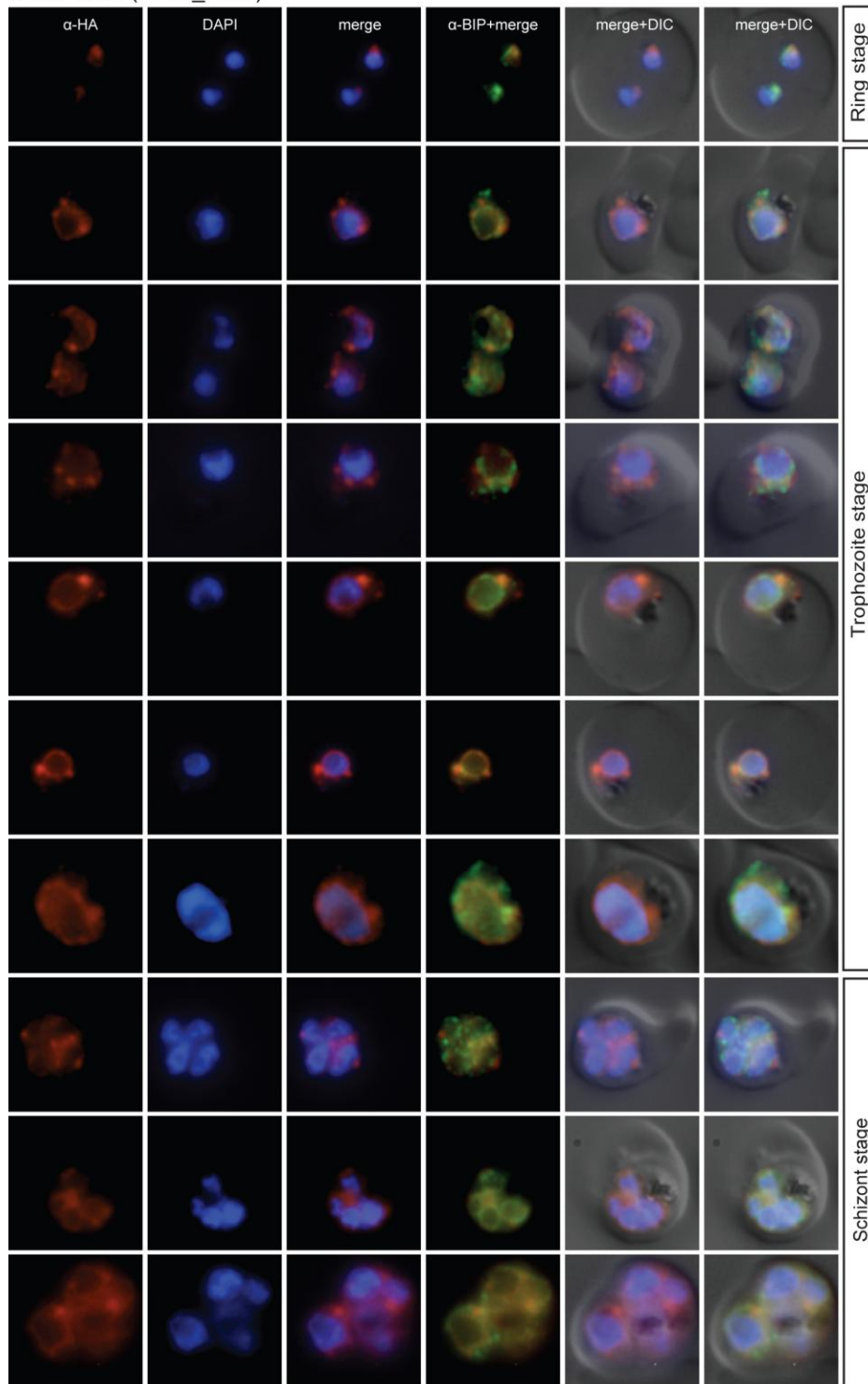
n-NuProC3 (PF07_0007)



Additional file 42. Localisation of n-NuProC3-3xHA (PF07_0007) during the IDC. Localisation of the tagged protein was visualised using anti-HA antibodies (red). Antibodies against PfBIP were used to visualise the ER. DAPI was used to visualise the nucleus. DIC images are shown as reference.

Additional file 43_Oehring et al.

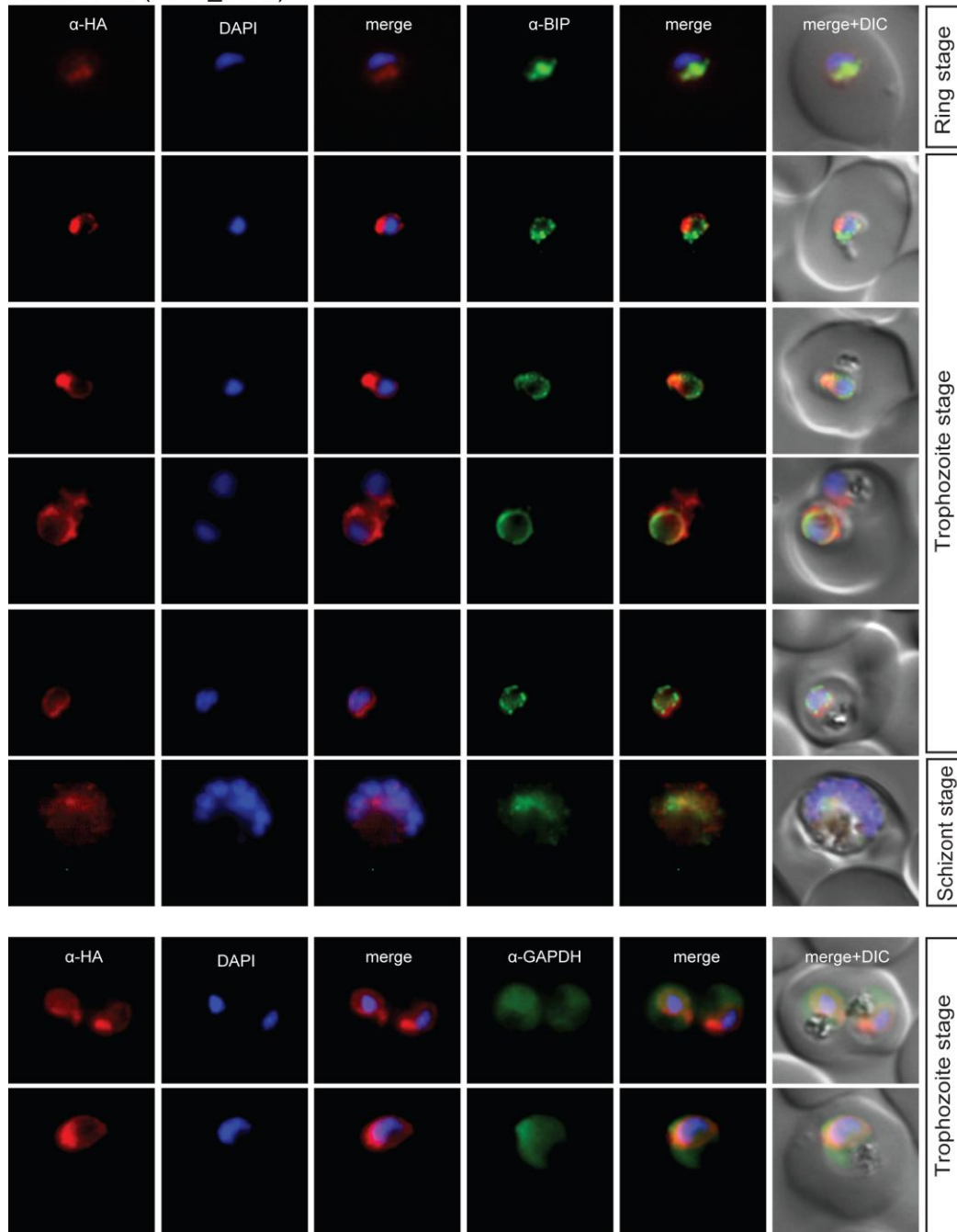
n-NuProC4 (PF10_0100)



Additional file 43. Localisation of n-NuProC4-3xHA (PF10_0100) during the IDC. Localisation of the tagged protein was visualised using anti-HA antibodies (red). Antibodies against PfbIP were used to visualise the ER. DAPI was used to visualise the nucleus. DIC images are shown as reference.

Additional file 44_Oehring et al.

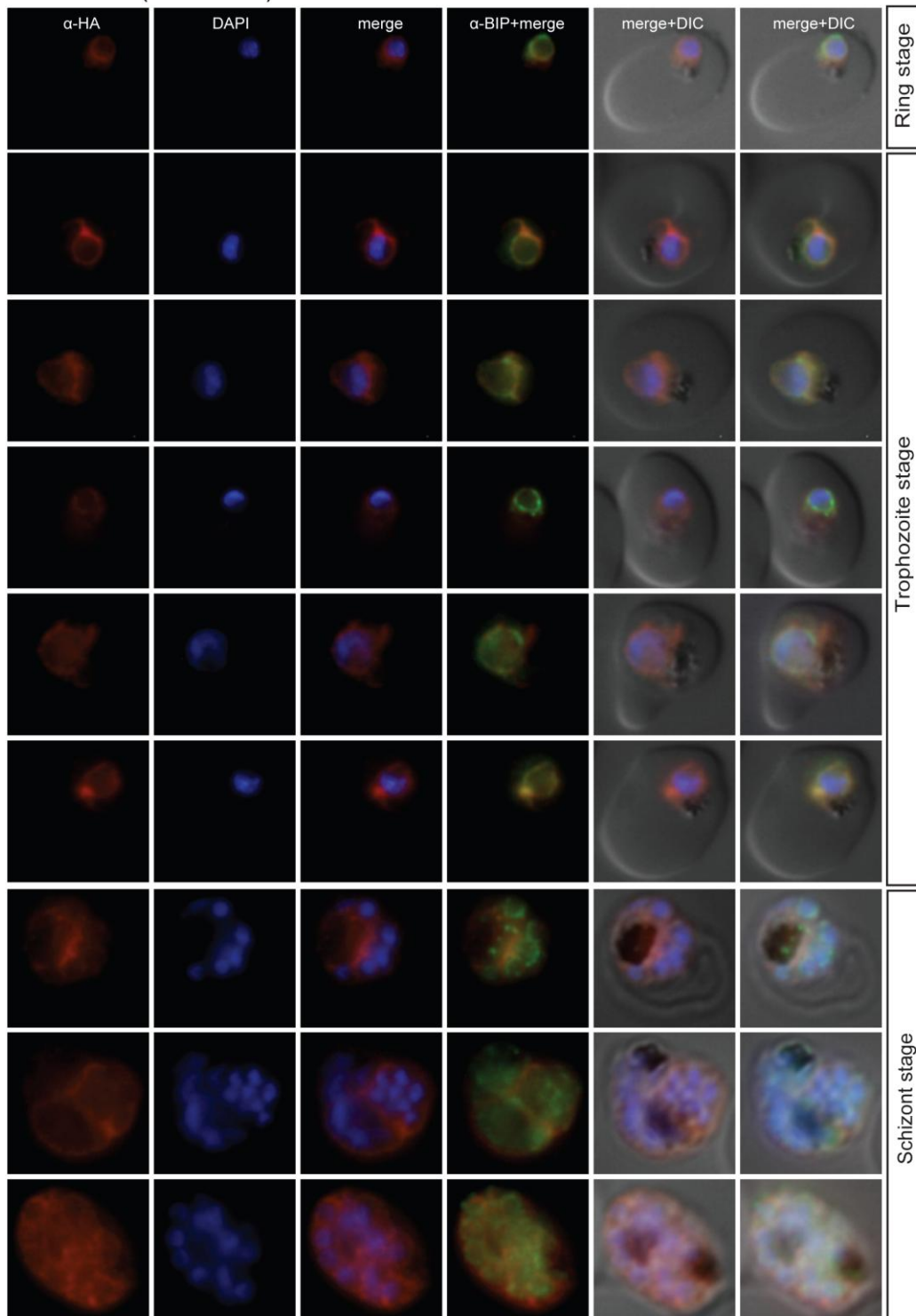
n-NuProC5 (PF11_0179)



Additional file 44. Localisation of n-NuProC5-3xHA (PF11_0179) during the IDC. Localisation of the tagged protein was visualised using anti-HA antibodies (red). Antibodies against PfBIP were used to visualise the ER. Antibodies against GAPDH were used to visualise the cytosolic compartment (bottom). DAPI was used to visualise the nucleus. DIC images are shown as reference.

Additional file 45_Oehring et al.

n-NuProC6 (PFB0395w)



Additional file 45. Localisation of n-NuProC6-3xHA (PFB0395w) during the IDC. Localisation of the tagged protein was visualised using anti-HA antibodies (red). Antibodies against PfbIP were used to visualise the ER. DAPI was used to visualise the nucleus. DIC images are shown as reference.

Additional file 46_Oehring et al.**Identification of Novel Domains**

To identify novel domains in the core nuclear proteome, all proteins were assembled into a FASTA file taken from PlasmoDB v6.4. InterPro domains taken from the PlasmoDB InterPro annotation file (http://plasmodb.org/common/downloads/release6.4/Pfalciparum/PfalciparumInterpro_PlasmoDB-6.4.txt) were used to mask previously defined domains using a custom-built script (http://github.com/wwood/bbbin/blob/master/mask_eupathdb_domains.rb commit 0bbb3eb). NCBI BLAST+ 2.2.19 [1] was then used to find homology between these masked protein sequences, using the BLASTP program, a permissive E-value cutoff ($1E-2$) and without a low complexity filter. One-way BLAST clustering (i.e. proteins with a hit to another protein were clustered together) was implemented with a custom-built script (http://github.com/wwood/bbbin/blob/master/blast_to_blastclust.rb commit ef9278b), and inspected by interrogation of the pairwise BLAST alignment file. The majority of alignments appeared spurious, consisting mainly of the common amino acids asparagine, lysine, glutamine, glutamic acid, isoleucine and serine, and the others were inspected further. The amino acid sequences of these proteins were more rigorously confirmed not to encode any domains by using the InterProScan [2] website (<http://www.ebi.ac.uk/Tools/InterProScan/>) and PlasmoDB IDs at the EuPathDomains website (<http://www.atgc-montpellier.fr/EuPathDomains/>) that claims to be more sensitive at finding PFAM domains based on co-occurrence of particular domains [3]. However, no additional domains were found. To further characterise the AP2-coincident domain, PSI-BLAST [4] at the NCBI website (<http://blast.ncbi.nlm.nih.gov/>) was used to query the nr database with default parameters starting with the conserved region found in PF11_0404 (residues 2543-2644, YVELLKTSIIICLNDILMNCIPQVFHLYKNINTSNDIKLEDILYTERKRKEQSLKYHIEYTQN SVGVSSLIPYLKLFSTEILNNVLPQAQSLEIQRLLIHSLS). Successive iterations showed that the majority of hits were apicomplexan proteins annotated in the one line description as AP2 domain-containing proteins. A more rigorous workup of whether hits contained AP2 domains was carried out later in the analysis. Matching sequences were downloaded from the NCBI website in FASTA format. Sequences were then mapped to EuPathDB identifiers using a custom BioPerl [5] script

(http://github.com/wwood/bbbin/blob/master/extract_from_fasta.pl commit fc1295e) by exact matching of amino acid sequences, and those not matching exactly (often rodent malaria proteins that have been the subject of recent genome-wide re-annotation) were included by using BLASTP and trusting the corresponding PlasmoDB sequence. A locally downloaded version of PSI-BLAST (2.2.24+) was then used to create a multiple sequence alignment of the domain, using residues 2668-2869 of the *P. vivax* protein PVX_092570 (PlasmoDB version 7.0) (FAEDMFENAPDANNNGTTQVDSNREDSLGETKKIASENNSSFPLNIDDKYYELLKTAIIICLNDILMNSIPKLFHIYKDISTTTNVKMEDLLNDEKKRREQSVKYHIAYTQNSIGVSSLIPYLRLFSMEILNNVLPSTQSLEIQRKIIYSLDLQAYNTSY) against the proteins downloaded from NCBI. Output was then parsed and hit sequences together with 20 amino acids either side were extracted using another custom BioPerl script (<http://github.com/wwood/bbbin/blob/master/sequenceChop.pl> commit 8c27d33). Hit sequences were aligned using MAFFT 6.704 using the MAFFT L-INS-i algorithm with default parameters. The resulting alignment was checked by manual inspection using JalView 2.4. The beginning and end of the alignment corresponding to likely spurious matching were removed by manual inspection. Highly similar sequences were removed since they corresponded to direct orthologs and their inclusion would bias the alignment toward these sequences unnecessarily. These highly similar sequences were from Plasmodium, Babesia, Theileria or Cryptosporidium spp. The alignment was saved in aligned FASTA format and converted to STOCKHOLM format using sreformat (part of the Ubuntu biosquid package version 1.9g+cv520050121-2). An HMM was created using the HMMER (<http://hmmmer.org/>) program hmmbuild version 3.0 with default parameters (Additional file 49). A protein sequence database comprising of amino acid sequences from all sequenced alveolate species was compiled. This included sequences from PlasmoDB 7.0 [6-9], ToxoDB 6.2 (strain ME49) [10], CryptoDB 4.3 [11,12], ParameciumDB v1.49 [13,14], TGD [15,16] Aug_2004 gene predictions, and Perkinsus marinus genome-derived GenBank [17] accession AAXJ000000000 proteins downloaded Nov 11 2010, as well as Babesia bovis [18], Theileria annulata [19] and Theileria parva [20] genome-derived RefSeq proteins downloaded May, 2008. The HMMER program hmmsearch was then used to search this database with default parameters, except using the -A flag to output a multiple sequence alignment that comprised of all the significant (E-value < 0.001) hits. The output STOCKHOLM format was first converted to MSF format using sreformat, and then aligned in FASTA format using JalView. This

aligned FASTA format file is given as Additional file 48. To search for the AP2 domain in PF13_0114, the HMM file for PF00847 was downloaded from PFAM (v24) and searched against the PlasmoDB 7.1 version of the protein sequence using HMMER v3.0 with default parameters.

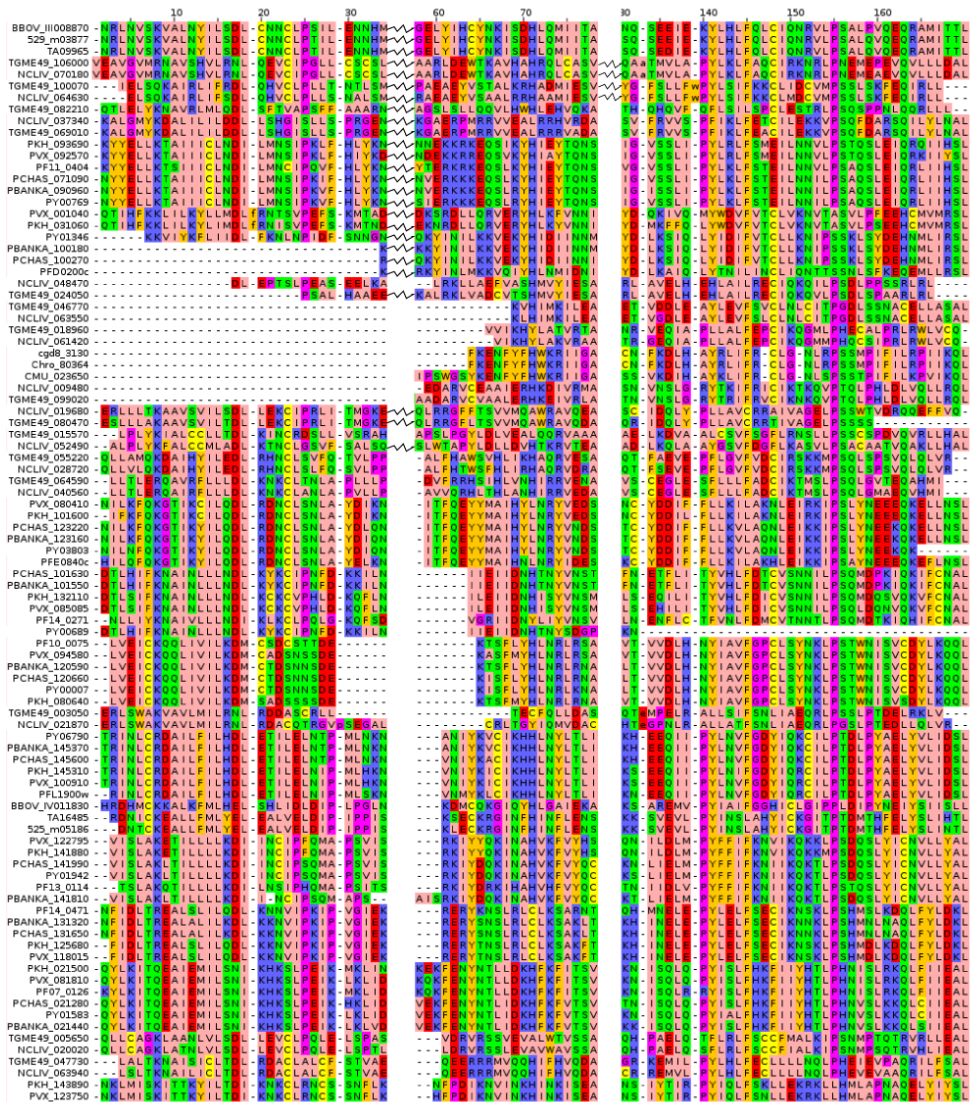
Analysis of each characterised protein domain was similar to the method used for the AP2-coincident domain, with slight differences. For partial CSTF, extended ELM2, and MYND domains, MAFFT version 6.833b, sequenceChop.pl commit 6a9ffda were used, except sequenceChop.pl commit 5abf5d was used for the MYND domain. For each of these three, highly similar sequences were not removed from the alignment prior to HMM creation. For the partial CSTF domain PSI-BLAST was initiated with the domain from PF10_0279, KDIPYAEEDLVKEIINEKSILQNILISKYVDMLNWTSEQVLRVLSIRKSLKRIGYNI. For the extended ELM2 domain PSI-BLAST was initiated with the domain from PF11_0429, KKKDKVKQKSEGSKYTNQINVGENYQVSNVSTFFLNHSEKYDETSKSELVYSPYLLE RMKENYLSEGQYELVIKNDYELAIFIKELAKNWKQCLGWHPFTPEYAFKILHHVDYNPK KAIELLKSSEFNFLICDPPIRKYENKWRPRDKRGQSDSPYPSSSELLQSYLKR, and 50 amino acids either side of the domain were aligned. The final alignment was created using MAFFT L-INS-i instead of hmmsearch. For the MYND domain, PSI-BLAST was initiated with the domain from PF10_0150, LCSGCKKVLIKLSCPICLNKIFSYFCNQECFKGSWKEHQKIHENMKNENNEKEDH.

Newly defined protein domains will be submitted to PFAM upon this article's acceptance into a journal (Additional files 48 and 53-55).

1. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K et al.: BLAST+: architecture and applications. BMC Bioinformatics 2009, 10: 421.
2. Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R et al.: InterProScan: protein domains identifier. Nucleic Acids Res 2005, 33: W116.
3. Ghouila A, Terrapon N, Gascuel O, Guerfali FZ, Laouini D, Marechal E et al.: EuPathDomains: The divergent domain database for eukaryotic pathogens. Infect Genet Evol 2010, 11: 698-707.
4. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W et al.: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997, 25: 3389-3402.
5. Stajich JE, Block D, Boulez K, Brenner SE, Chervitz SA, Dagdigian C et al.: The Bioperl toolkit: Perl modules for the life sciences. Genome Res 2002, 12: 1611-1618.
6. Carlton JM, Adams JH, Silva JC, Bidwell SL, Lorenzi H, Caler E et al.: Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*. Nature 2008, 455: 757-763.
7. Carlton JM, Angiuoli SV, Suh BB, Kooij TW, Perteau M, Silva JC et al.: Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii*. Nature 2002, 419: 512-519.
8. Hall N, Karras M, Raine JD, Carlton JM, Kooij TW, Berriman M et al.: A comprehensive survey of the *Plasmodium* life cycle by genomic, transcriptomic, and proteomic analyses. Science 2005, 307: 82-86.
9. Pain A, Bohme U, Berry AE, Mungall K, Finn RD, Jackson AP et al.: The genome of the simian and human malaria parasite *Plasmodium knowlesi*. Nature 2008, 455: 799-803.

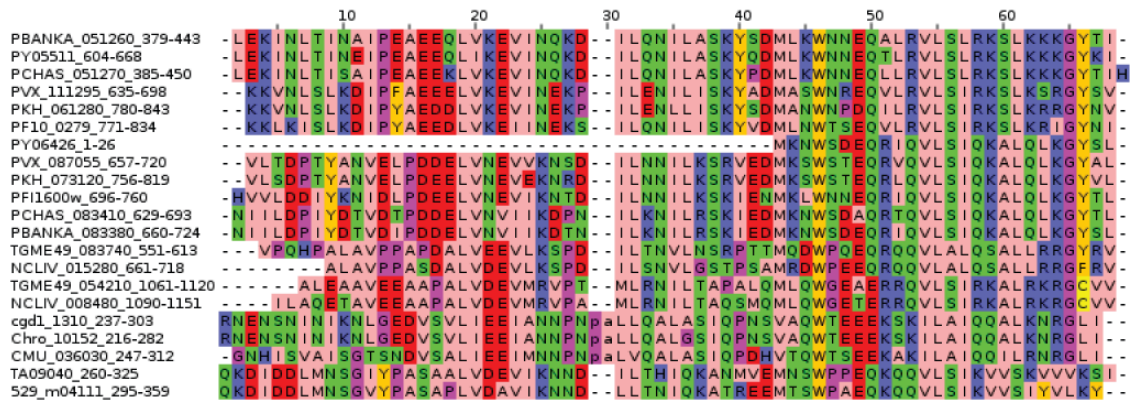
10. Kissinger JC, Gajria B, Li L, Paulsen IT, Roos DS: ToxoDB: accessing the *Toxoplasma gondii* genome. *Nucleic Acids Res* 2003, 31: 234-236.
11. Heiges M, Wang H, Robinson E, Aurrecochea C, Gao X, Kaluskar N et al.: CryptoDB: a *Cryptosporidium* bioinformatics resource update. *Nucleic Acids Res* 2006, 34: D419.
12. Abrahamsen MS, Templeton TJ, Enomoto S, Abrahante JE, Zhu G, Lancto CA et al.: Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science* 2004, 304: 441-445.
13. Arnaiz O, Cain S, Cohen J, Sperling L: ParameciumDB: a community resource that integrates the *Paramecium tetraurelia* genome sequence with genetic data. *Nucleic Acids Res* 2006, 35: D439.
14. Aury JM, Jaillon O, Duret L, Noel B, Jubin C, Porcel BM et al.: Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature* 2006, 444: 171-178.
15. Stover NA, Krieger CJ, Binkley G, Dong Q, Fisk DG, Nash R et al.: *Tetrahymena* Genome Database (TGD): a new genomic resource for *Tetrahymena thermophila* research. *Nucleic Acids Res* 2006, 34: D500.
16. Eisen JA, Coyne RS, Wu M, Wu D, Thiagarajan M, Wortman JR et al.: Macronuclear genome sequence of the ciliate *Tetrahymena thermophila*, a model eukaryote. *PLoS Biol* 2006, 4: e286.
17. Benson DA, Boguski MS, Lipman DJ, Ostell J, Ouellette BF: GenBank. *Nucleic Acids Res* 1998, 26: 1.
18. Brayton KA, Lau AO, Herndon DR, Hannick L, Kappmeyer LS, Berens SJ et al.: Genome sequence of *Babesia bovis* and comparative analysis of apicomplexan hemoprotozoa. *PLoS pathogens* 2007, 3: 1401-1413.
19. Pain A, Renauld H, Berriman M, Murphy L, Yeats CA, Weir W et al.: Genome of the host-cell transforming parasite *Theileria annulata* compared with *T. parva*. *Science* 2005, 309: 131-133.
20. Gardner MJ, Bishop R, Shah T, De Villiers EP, Carlton JM, Hall N et al.: Genome sequence of *Theileria parva*, a bovine pathogen that transforms lymphocytes. *Science* 2005, 309: 134-137.

Additional file 47_Oehring et al.



Additional file 47. ACDC domain alignment graphic. A multiple sequence alignment of the ACDC domains in apicomplexan proteins. The raw data used to generate this graphic and the location of the domain in each protein's equence can be found in Additional file 48. Sequence identifiers refer to EuPathDB gene identifiers. Jagged lines ndicate unconserved insertions in the domain.

Additional file 50_Oehring et al.



Additional file 50. Partial CSTF domain alignment graphic. A multiple sequence alignment of the apicomplexan proteins encoding the domain is shown in graphical format. The raw data used to generate this graphic and the location of the domain in each protein's sequence can be found in Additional file 53. Sequence identifiers refer to EuPathDB gene identifiers.

Additional file 52_Oehring et al.

```

PFF0350w  1 C N N C K E L A E - - L Q C S Q C - - - - K K T Y Y C S K E C Q M K D W I N H R D V C   37
PFF0105w  1 C E H C Y K E Q N - L K R C G R C - - - - K K V Y Y C S V E C Q K S D Y V F H K R I C   38
PF14_0310  1 C S G C G V V T E V T I C C P I C I k n n K K I F Y C S Q E C F E Q N Y K E H K K I H   43
PF13_0293  1 C F Y C L E K F N K C I C C P N C - - - - K Y V V Y C S D M C L E R A W K S H R E E C   39
PF10_0150  1 C S G C K K V L I K K L S C P I C I k n k I F S Y F C N Q E C F K G S W K E H Q K I H   43

```

Additional file 52. MYND domain alignment graphic. A multiple sequence alignment of *P. falciparum* proteins encoding the domain is shown in graphical format. The raw data used to generate this graphic and the location of the domain in each protein's sequence can be found in Additional file 55.

Additional file 56-Oehring et al.

Score = 62.0 bits (149), Expect = 1e-12, Method: Compositional matrix adjust.
 Identities = 71/283 (26%), Positives = 115/283 (41%), Gaps = 38/283 (13%)

```

PFI0250c  218  NSI FGGLGTSTNQSTGGGLFGNTGATSQNKTGGIFGGLSSTNQASTSSTSMFGGLSSNQA  277
+S+FG  + + +  LFG+T  +FG L  N+  S+FG  S+
PF14_0442 1328  DSLFGS-SINNDKSKINLFGSTMNDGDKNKTNLFGSL---NKDDKDKPSIFGSPSNKDD  1383

PFI0250c  278  KPTNSLFGGLSS-----GATSNTGTQQSGNLFGSASGIGQSKTVGGIFGNLSSTN  327
K  ++FG  S+  G +SN  +  +FGS S  K  IFG  S +N
PF14_0442 1384  KDKATIFGSPSNKDDKDKATIFGFSSNKDDKDKAPIFGSPSN-KDDKDKATIFG--SPSN  1440

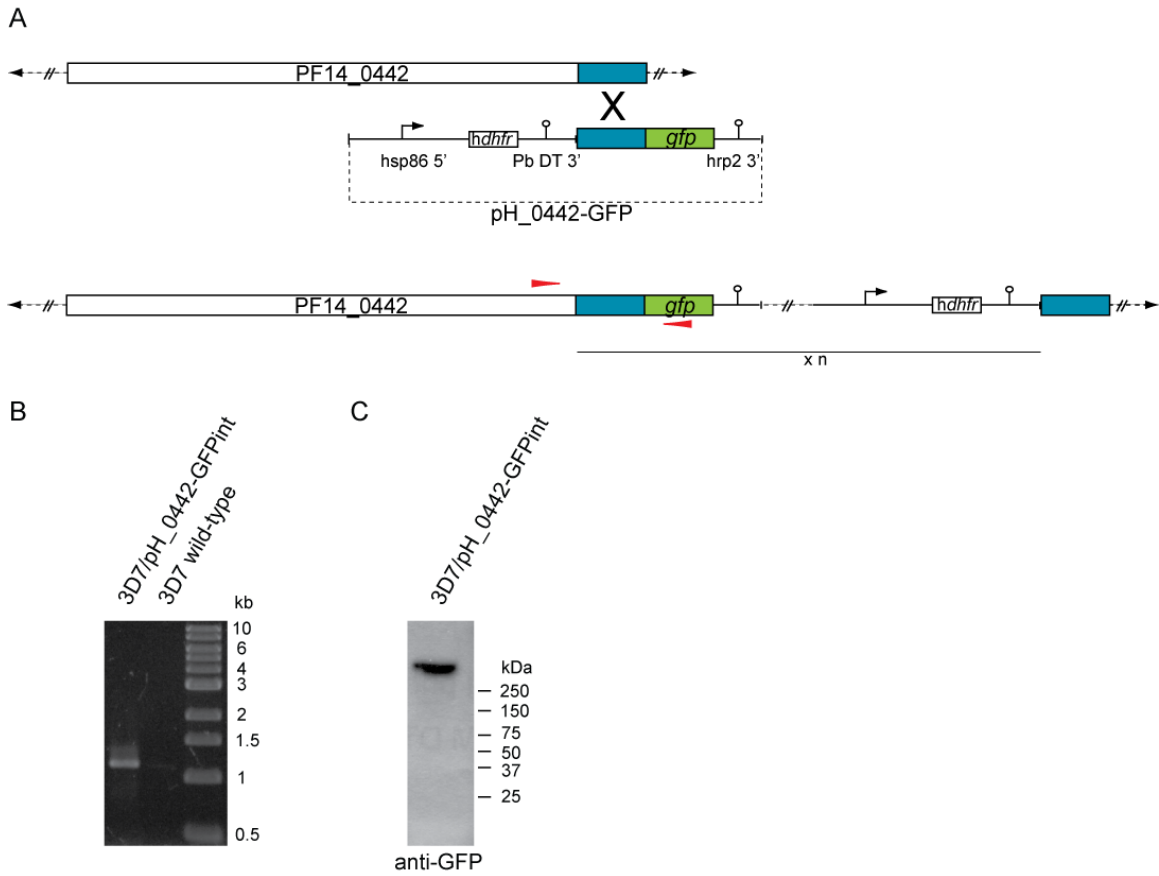
PFI0250c  328  QASTSSNMFGGLSSNQAKPTSSLFGGLSS-----GTTTNTSTQQSGNLFGSATG  377
+  + +FG  S+  K  + +FG  S+  G+ +N  +  +FGS  +
PF14_0442 1441  KDDKDKATIFGFSSNKDDKDKAPIFGFSSNKDDKDKAPIFGSPSNKDDKDKAPIFGSPSN  1500

PFI0250c  378  LGQNKTGGGIFGTLPSANQTSTTSSNMFGGLSTNQAKPTSSLFGGMSSGT-----TGI  430
+K  IFG+ PS N+  + +FG  S  K  +++FGG + G  GI
PF14_0442 1501  -KDDKDKTPIFGS-PS-NKDDKDKTPIFGSPSNKDDKDKTAIFGGSTFGNNKSPMFGQGI  1557

PFI0250c  431  TTNTTAQSGNLFGGTGTSQNKTGNLFGALPG-ANQTSTTSNIF  472
S  +FG T TS +  +  G  ++  T  N+F
PF14_0442 1558  LNKGDNVSVPVFGNTVTSTDNDKSKLKLFGNGNKDKEETKENVF  1600
    
```

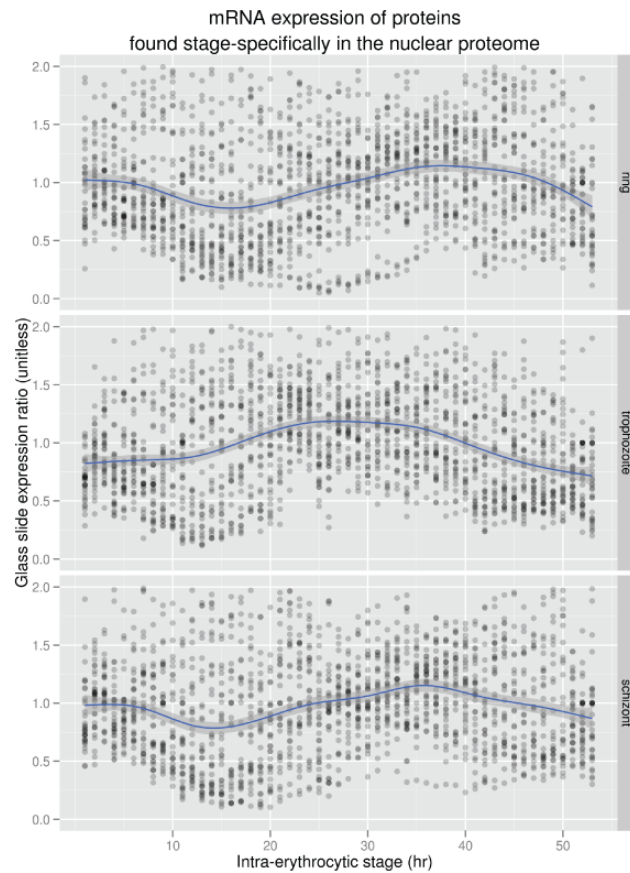
Additional file 56.FG-repeat alignment graphic. A pairwise alignment of the two putative FG-repeat proteins, PFI0250c (PFNUP100) and PF14_0442, is shown. The display is modified from the output of a bl2seq alignment between the two protein sequences. While the alignment of specific FG pairs is unlikely to be the evolutionarily-motivated true alignment because of the repetition in these proteins, the number of FG motifs and detection of both proteins in the nuclear proteome suggest that these proteins are functional FG-repeat proteins. Amino acid positions refer to the respective PlasmoDB version 7.0 protein sequences.

Additional file 57_Oehring et al.



Additional file 57. C-terminal tagging of endogenous PF14_0442 by 3' replacement. (A) The schematic illustrates integration of pH_0442-GFP into the endogenous PF14_0442 locus by single cross-over recombination. Red arrowheads indicate the position of PCR primers used to verify the single cross-over event. (B) PCR using primers on gDNA isolated from 3D7/pH0442_GFPint and 3D7 wild-type parasites. (C) Western blot using anti-GFP antibodies on a nuclear extract derived from 3D7/pH0442-GFPint parasites.

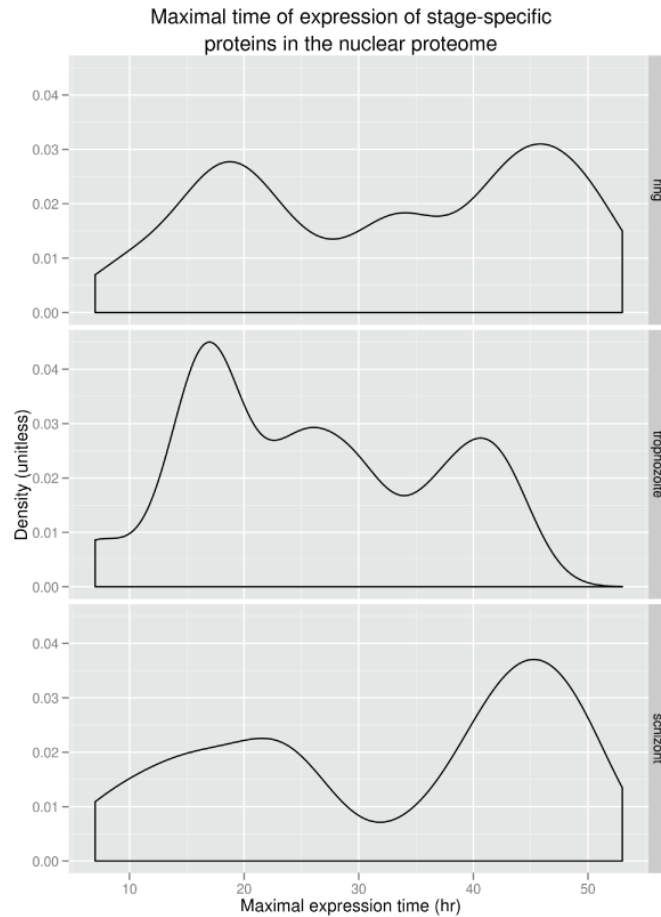
Additional file 58_Oehring et al.



Additional file 58. mRNA expression profiles of proteins found stage-specifically in the nuclear proteome. Each dot represents the expression profile of a stage-specific protein at a particular hour post-invasion during the *P. falciparum* 3D7 IDC. The x-axis refers to the hour during the IDC post-invasion. The y-axis represents relative expression of the transcript at that timepoint compared to an average over all timepoints [1]. Each protein was assigned a single dot in each timepoint, where that data was available. The dots are semi-transparent such that five overlaid dots appear as a solid black, and a single dot one fifth as dark. A minority of dots had expression ratios greater than two, and hence are not shown in the plot.

1. Llinas M, Bozdech Z, Wong ED, Adai AT, DeRisi JL: Comparative whole genome transcriptome analysis of three *Plasmodium falciparum* strains. *Nucleic Acids Res* 2006, 34: 1166-1173.

Additional file 59_Oehring et al.



Additional file 59. Maximal hour of mRNA expression of stage-specific proteins in the nuclear proteome. First, the single hour at which maximal mRNA expression occurred [1] was tallied for each protein found stage-specifically in the core nuclear proteome. Then these hours of maximal expression were put through a density estimation filter at 0.6 times the default of the density function in R, since this value appeared to be a good balance between being too noisy and being too board as to not see any patterns. Hence, a higher density represents a larger number of proteins being maximally expressed during that time period.

1. Llinas M, Bozdech Z, Wong ED, Adai AT, DeRisi JL: Comparative whole genome transcriptome analysis of three Plasmodium falciparum strains. *Nucleic Acids Res* 2006, 34: 1166-1173.

Additional file 62_Oehring et al.**Characterisation of Lineage-Specificity**

To determine how many proteins in the core nuclear proteome were phylum- or genus-specific, OrthoMCL version 4 was used. The definition of phylum-specific was simply that the gene had no orthologue outside of Apicomplexa, where all genes in the OrthoMCL group were considered the complete set of orthologous genes. Likewise, genus-specific genes had no orthologues outside *Plasmodium* spp. Specifically, a custom ruby script https://github.com/wwood/bbbin/blob/master/orthomcl_species_jumper.rb commit abf2c13 was used with the arguments “-i pfal -o pfal,pviv,bbov,pber,ncan,tann,tgon,pyoe,pcha,pkno,tpar,cmur,chom,cpar -v” and “-i pfal -o pfal,pviv,pber,pyoe,pcha,pkno -v” to search for orthologues outside the phylum and genus, respectively. Each line of the outputs were then truncated to 100 characters for computer performance reasons, and then pasted into a LibreOffice spreadsheet (<http://www.libreoffice.org/>), and saved in tab-separated values format. Taxon-specificity was then interrogated using spreadsheet formula, and summarised using GNU awk and “uniq -c” (<http://www.gnu.org/>).

Additional file 63. Primer sequences used in this study (restriction sites are in bold)

NuProc1 F	CAGT GGATCC TAAAATGAAAAATTAATAAATACAC
NuProc1 R	CAGT GCTAGC TTCTCTTTTAAATTATCCTTTTGG
NuProc2 F	GATC GGATCC TTAATATGGATGAACTAAATAAAG
NuProc2 R	GATC CCATGG TGTGTTGAATTTATTTTTTTAAC
NuProc3 F	GATC GGATCC AAAAAATGTCCAATACATTAATTTTC
NuProc3 R	GATC CCATGG TATTAACATCTGTTCAATTAATC
NuProc4 F	CAGT GGATCC TTATCAATGGATCACCAAGATCTG
NuProc4 R	CAGT CCATGG TTTTTCTTGGGTGAAGGTAATTTG
NuProc5 F	GATC GGATCC TAAAATGAACACCGAAGAAAAATTAATAC
NuProc5 R	GATC CCATGG ATCAGCTTGGCAAGTATCG
NuProc6 F	GATC GGATCC TAAATATGAATGATTATAATAATAAATAATC
NuProc6 R	GATC GC TAGC TATTTTAGCGTCATAACCTTCG
NuProc7 F	CAGT GGATCC AAAAATGGGAGATAATAACACATCC
NuProc7 R	CAGT CCATGG TTTATCATCATCTTCTTCCTC
NuProc8 F	GATC GGATCC TAAAATGTCGTCAGCAACAC
NuProc8 R	GATC CCATGG AAATTTTTGTTTGAATCAAAAAC
NuProc9 F	GATC GGATCC TAC TTA TGG GTA GAA AGA AG
NuProc9 R	GATC GCTAGC TTACAATTTTGTGTATG
NuProc10 F	CAGT GGATCC TGACCAATGATGCTTGAAG
NuProc10 R	CAGT CCATGG ATCTTTTTTTGAACTATCCTC
NuProc11 F	CAGT GGATCC AAAAATATGGCTGATAATAACCTAG
NuProc11 R	CAGT CCATGG GTAAGAATCCTTTGATGAATCG
NuProc12 F	CAGT GGATCC AAAAATGAGTAATTACCAAAATTTGAAG
NuProc12 R	CAGT CCATGG TTGTGTTATCCCAAATCTAAGAG
NuProc13 F	CAGT GGATCC TTAATATGGCTAATACATGGAGATG
NuProc13 R	CAGT CCATGG TGTTTTGTAATTAATTTTTTTTTGTAC
NuProc14 F	GATC GGATCC AAAAAATGAATGTTGAAGACATAAC
NuProc14 R	GATC CCATGG ATCCATTGTTTGCTTCC
NuProc15 F	GATC GGATCC AATAAATGACCGATATACAAATAC
NuProc15 R	GATC CCATGG CTTTTTGACAAGATATGCTGC
NuProc16 F	CAGT AGATCT TAAATAATGAAGGATCCATTAG
NuProc16 R	CAGT CCATGG ACTTTTCAGTATTGAAGTTATAT
NuProc17 F	GATC GGATCC ATAATATGTCTATAAGAAATAG
NuProc17 R	GATC CCATGG GAATCTACCATTTTTTACATTTTG
NuProc18 F	GATC GGATCC TAAATATGGGTACCTTTATAAAC
NuProc18 R	GATC CCATGG CATAAAATCCTCTAAATTATC
NuProc19 F	CAGT GGATCC AAAAATGAGTAATTACCAAAATTTGAAG
NuProc19 R	CAGT CCATGG TTGTGTTATCCCAAATCTAAGAG
NuProc20 F	CAGT GGATCC AAAAATGTCTACAATTAGATGGG
NuProc20 R	CAGT CCATGG ATTTAGAGTGTCAAATTTATGTAAC
NuProc21 F	CAGT AGATCT TAAAAATGGCAGCTAATGAAGG
NuProc21 R	CAGT CCATGG GAGAAGTAAGAGGATCCTCATTGC
NuProc22 F	GATC GGATCC AAAAATATGCAAAAGAAGCCATC
NuProc22 R	GATC CCATGG CATTTTCGGCTATTTCTTATTTG

n-NuProc1 F	CAGT GGATCC AAAGAATGAATGTTACAAATGTAG
n-NuProc1 R	CAGT CCATGG CAATTCATCATCATATTGCTC
n-NuProc2 F	GAT CGGA TCC TTA TGA TGA AGG TGT GGG G
n-NuProc2 R	GAT CCCATGG TGAAAATATTTCTTCATAATTTG
n-NuProc3 F	GAT CGGATCCT GAAAATGAGTCAACCACAAAAAC
n-NuProc3 R	GAT CCCATGG TTTTGCGTTCTGTAAACTGGC
n-NuProc4 F	GAT CGGATCC AAAAAATGAAGCTCAAATTAATTTAG
n-NuProc4 R	GAT CCCATGG TTTATTTTTTTTGGCCAACAAC
n-NuProc5 F	GAT CGGATCC AAAAATGTTGGGTTTAATATTTAAAG
n-NuProc5 R	GAT CCCATGG TAATGTCTTTAATTTCTTTACAATTTTG
n-NuProc6 F	CAGT GGATCC GTAAATTATGAAGGGTAAAGGTAG
n-NuProc6 R	CAGT CCATGG TGATGTGTTATCTGGTGGC
0442 3'repl. F	GAT CCTGCAG ATAAACGTATTTCCACATGC
0442 3'repl. R	CAT GGCGGCCG CTATTTATCATATTTTGATTC
0442 int. F	TTAATGGATGTAATAAAAAGTGATGATAGC
0442 int. R	ACCTTCAAACCTTGACTTCAGCACGTGTCTTGTAGT

Chapter 3: Complex formation of three bromo-domain proteins of *P. falciparum* (working paper)

Introduction

Within the eukaryotic nucleus, epigenetic mechanisms play a dynamic role in gene regulation and are in part defined by a complex set of reversible histone modifications that are recognised by structurally diverse proteins. Acetylation of N-terminal histone tails loosens the compact chromatin structure and makes DNA accessible to transcriptional activators. The bromo-domain is the only known 'reader' module that specifically recognises acetyl-lysine residues and that is evolutionary conserved in chromatin-associated proteins from yeast to mammals [1][2]. The 150 amino acid motif is composed of a conserved core region of 60 amino acids adopting a left-handed bundle of four helices (α_Z , α_A , α_B , α_C) with two inter-helical α_Z - α_A (ZA) and α_B - α_C (BZ) loops of variable length [3].

The 3-D structure of many bromo-domains has been solved by NMR and these studies revealed that both loops form a deep cleft, within which three conserved residues function as acetyl-lysine docking sites (Tyr-Tyr-Asn) [4]. Besides the three conserved ligand-specific binding residues, the sequence variations of the ZA and BZ loop contribute to the selective environment of the binding pocket of bromo-domains [5]. The structural cavity of the bromo-motif is suitable for being occupied by small chemical compounds that competitively block the selective interaction with acetyl-lysines. Therefore it is advantageous that the biological interaction between the bromo-domain and the acetyl-lysine residue is not very tight (binding affinities (Kd) are between 10 and 100 μ M) [6].

Bromo-domain proteins were found in single- or multicellular organisms with numbers varying from 10 in *S. cerevisiae* to 42 in humans. Human proteins carry up to six bromo-domains, whereas yeast proteins possess mostly tandem repeats [7]. Domain arrangements of a bromo-domain with other chromatin modules like PHD, B-box type zinc finger, SET or ATPase are frequently observed and particular units seem to function independently [8][9][7]. Furthermore, bromo-domain proteins often play a role as

functional subunits of large proteineous complexes. They uniformly recognise acetylated lysines and mediate gene activation in various cellular processes. The orchestrated concurrence of post-translational modifications defines the epigenetic status of chromatin [10]. Neighbouring marks could thus influence the binding characteristics of bromo-domains, either by enforcing the interaction due to the presence of multiple bromo-motifs, or by preventing interactions through contrasting marks [11]. Bromo-domain containing proteins have a wide range of functions within the transcriptional process and act as transcription factors, transcriptional co-activators, histone acetyl-transferases, methyl-transferases, chromatin structuring and chromatin remodelling proteins. Histone acetyl-transferases commonly contain bromo-domains, which is an apparent contradiction as the motif is coupled to an enzymatic activity generating its own recognition site. The acetyl-binding capability coupled with the acetyl-transferase activity suggests a spreading mechanism throughout active chromosomal regions of eukaryotic organisms [11].

The importance of the bromo-domain in the chromatin context was determined *in vivo* for several bromo-domains with the help of yeast deletion studies. Functional knockouts of some bromo-domains of the multi-protein complex RSC caused inhibition of cell growth and knockouts affecting the bromo-domain of GCN5 HAT appeared to be crucial for yeast viability [12][13]. Thus, some abundant bromo-domain proteins in yeast seem to play a global and vital role in the chromatin context.

Interestingly, chromatin-associated proteins of protozoans appear to have emerged diverse protein architectures that combine new modules in comparison to free-living eukaryotes [14]. The genome of *P. falciparum* encodes one protein (PF10_0328) carrying a bromo-domain as well as an ankyrin module and this particular combination is uniquely found in apicomplexan parasites. Ankyrin motifs are short, tandemly repeated stretches which consist of 33 residues and belong to the most abundant repeat class found in nature [15][16]. The functional involvement of ankyrin domain-containing proteins encompasses a wide range of cellular processes including cytoskeleton integrity, transcription, cell-cycle signalling and regulation. Most of these proteins contain a maximum of six repetitions, which fold into two anti-parallel α -helices and an additional β -hairpin (or long loop) leading to a curve shape as protein specific- interface [17].

P. falciparum also has a bromo-domain containing histone acetyl-transferase, PfGCN5, which participates in the formation of a large multimeric protein complex resulting in chromatin remodelling and transcriptional activation (PF08_0034) [18][19]. *In vitro* data suggested that PfGCN5 has a conserved HAT activity with substrate preference for histone H3 modifying lysine 9 and 14 [18][19]. An additional member of this complex was identified in *P. falciparum*, the PfADA2 co-activator (PF10_0143), which interacts *in vitro* and *in vivo* with PfGCN5 [20]. A second histone-modifying protein of *P. falciparum* includes a putative H3K4 histone methyl-transferase characterized by the presence of a SET-domain (PfSET1; PFF1440w) [21]. PFF1440w has been predicted to be involved in a network of chromatin related proteins using a high-throughput version of the yeast two-hybrid assay [22]. The *in silico* study of Vaquero and Bischoff classifies four additional bromo-domain containing proteins of *P. falciparum* into the group of chromatin structuring proteins, as they do not carry a catalytic domain and are therefore most probably involved in recruiting other factors (PF10_0328, PF14_0724, PFA0510w and PFL0635c) [23]. This effect is thought to happen in a long-range manner, which is in contrast to the nucleosome remodelling process that appears to be locally restricted [5]. Finally, the parasite encodes a protein predicted to be involved in chromatin remodelling that also carries a bromo-domain and shows similarity to ATPase-dependent proteins homologous to the Swi/Snf complex (PFB0730w) [23].

The bromo-domain protein PF10_0328 was found in our nuclear core proteome and localises to the nucleus in IFA experiments (see previous chapter). I anticipate it probably recognises specific acetylated lysine residues on histones and mediates gene activation. A second putative bromo-domain protein we detected in our proteomic study, PF11_0254, which shows an identical striking nuclear localisation pattern like PF10_0328 and carries a domain with weak sequence similarities to the bromo-motif (not annotated in PlasmoDB as bromo-domain protein). Finally, a third bromo-domain protein was identified as well in the nuclear core proteome annotated as PFL0635c and confirmed to be nuclear by IFA. Functional characterisation of all three proteins has the potential to reveal novel insight into *P. falciparum* gene regulation and may lead to new intervention strategies against malaria. Indeed, the recognition of acetylated histone tails by bromo-domains can be selectively prevented [24] and produces profound changes in chromatin structure of *P. falciparum*. The therapeutic potential of blocking the protein-

protein interaction of the bromo-domain and the acetyl-lysine by small-molecules is realised in the research field of human cancer since quite a while [25].

Despite their key function in epigenetic regulation, bromo-domain containing proteins have never been investigated on the functional level in *P. falciparum*. To identify proteins interacting with PF10_0328, we performed co-immunoprecipitation experiments coupled to LC-MS/MS. Our results suggest that PF10_0328 acts in concert with two additional bromo-domain proteins in regulating gene transcription. Two further co-purified proteins, a putative ApiAP2 transcription factor (PFF0670w) and the histone H2A (PFF0860c), support this hypothesis.

Results

Nuclear extraction properties and putative interaction partner candidates of the bromo-domain protein PF10_0328 in *P. falciparum*

Remarkably, the protein PF10_0328 carries an ankyrin motif in addition to the bromo-domain and consists in total of 488 aa. The single bromo-domain is predicted to be located at the C-terminal between the aa 310 to 434 and the N-terminal ankyrin motif is predicted to be located between the aa 10 to 215. In this study, I was interested in the identification of interaction partners of PF10_0328 in order to make functional predictions about its role in chromatin-related processes in *P. falciparum*. Using the 3D7/PF10_0328-3xHA transgenic parasite line expressing the full length epitope-tagged version of PF10_0328 from episomal plasmids, the putative binding partners were co-immunoprecipitated and analysed by mass spectrometry. Different nuclear extracts of 3D7/PF10_0328-3xHA were generated and tested for protein abundance and to establish ideal starting conditions for co-immunoprecipitation. The solubility of PF10_0328-3xHA was determined to be optimal under high salt conditions. Extraction in RIPA buffer containing 150mM NaCl and 0.5% of the ionic detergent deoxycholate (DOC) also revealed good protein solubility (Figure 1A).

Thus, 3D7/PF10_0328-3xHA-derived mixed nuclear extracts (high salt extract mixed with RIPA/DOC extract) of late trophozoite and schizont stage parasites were incubated with HA-affinity matrix (Roche Diagnostics) to co-purify potential interaction partners. A molecular excess of soluble HA-peptides was added as negative control to the binding

step, which specifically prevents the binding of PF10_0328-3xHA to the matrix. After several wash steps the final elution step was performed under highly specific and gentle SDS-free conditions using soluble HA-peptides. To test the quality of the complete procedure, each fraction was analysed by western blot analysis as well as by silver staining prior to the analysis of the co-IP and negative control eluates by mass spectrometry. As shown in Figure 2 the protein PF10_0328-3xHA showed efficient binding and elution properties in the positive co-IP run, whereas in the negative control incubation the excess of HA-peptides successfully blocked the binding of PF10_0328-3xHA to the affinity matrix. Silver staining of significant co-IP fractions verified that both peptide elutions derived from the positive run and the negative control show similar protein patterns except for the bait protein PF10_0328-3xHA. A band of predicted size of 56 kDA (PF10_0328) is clearly visible only in the positive co-IP and not in the negative control in presence of competing HA-peptides. The fourth and last wash steps of both runs are clean and both elutions under SDS conditions show additional release of proteins from the beads (Figure 2b).

Total protein eluted from both runs were precipitated separately, trypsinised and analysed by LC-MS/MS. Peptide spectra were searched against a combined human and *P. falciparum* annotated database using the TurboSequest software with stringent search settings. 17 and 22 *P. falciparum* proteins were identified in the eluates of the positive and the negative run, respectively. Interestingly, two proteins out of the seven that were exclusively detected in the positive sample contained a bromo-domain and were represented by more than two unique peptides (Table 1). The protein PF11_0254 carries a shortened version of the bromo-motif (predicted aa 342-399) at the C-terminus, whereas the protein PFL0635c carries a central bromo-motif of the expected size (predicted aa 476-602).

Both of the identified bromo-domain containing proteins including the bait protein PF10_0328 itself are expressed at a constant level during the asexual blood stages of *P. falciparum* [26]. Moreover, in our previous study of the nuclear core proteome we already showed that all three bromo-domain proteins localise within the nuclear compartment. The other proteins that were exclusively co-purified in the positive co-IP sample represent a putative ApiAP2 transcription factor (PFF0670w), a hypothetical protein (PF14_0379) and the merozoite capping protein 1 (MCP1; PF10_0268), each

detected by one unique peptide. Furthermore, the histone variant H2A (PFF0860c) was identified, represented by two unique peptides. Due to the high peptide coverage, and the characteristic complex formation of bromo-domain proteins found in other organisms [12][27], PF11_0254 and PFL0635c are considered to be the most probable interaction partners of PF10_0328. Additionally, the detection of the ApiAP2 transcription factor and the histone variant H2A could indicate a putative function for PF10_0328 in transcriptional regulation.

The fact that the bromo-domain protein PFL0635c was detected in a second independent co-immunoprecipitation assay, where nuclear extract generation and co-IP were performed in a slightly different way (Olivier Dietz, MSc thesis 2009), is in line with our results. In that experiment nuclei of the transgenic parasite line 3D7/PF10_0328-3xHA were extracted under high salt extraction, and the insoluble pellet was sonicated. Both samples were then separately processed as described above. The protein PFL0635c was identified twice by one unique peptide, in the eluate derived under high salt conditions as well as in the eluate of the co-IP using the extract of the sonicated pellet.

Validation of the complex formation of three *P. falciparum* bromo-domain containing proteins

To confirm the interaction between PF10_0328 and the two putative binding partners PF11_0254 and PFL0635c, the latter two were cloned in full length into a transfection vector containing a 2xTy tag and *hdhfr* resistance cassette (pHcam-2xTy) [28] and were transfected separately into the existing BSD-resistant 3D7/PF10_0328-3xHA parasite line. Again, high salt and RIPA/DOC nuclear extracts of the double transgenic parasite lines were generated from late trophozoite and schizont stages and tested for protein expression by western blot. The two putative interacting proteins were soluble under high salt conditions, but PF11_0254-Ty showed a slight reduction and PFL0635c an even stronger reduction in solubility after treatment with RIPA/DOC buffer, specifically in comparison to the last SDS-extraction (Figure 3). A small scale co-IP assay was performed using anti-HA matrix under similar conditions as described above, except for the elution conditions (a buffer containing 2% SDS was used instead of eluting in presence of HA-peptides). Figure 4 shows the Western blot results from these experiments. For each cell line the putative interacting proteins PF11_0254 and

PFL0635c co-purified with PF10_0328-3xHA specifically in the positive co-IP experiment (Western blots were probed first with anti-HA antibodies, stripped, and probed with anti-Ty antibodies). Both proteins remained undetected in the negative control co-IP (in presence of competing soluble HA-peptides) showing that these co-purifications were dependent on the presence of PF10_0328-3xHA. These results clearly confirm the mass spectrometry-based identification of PF11_0254 and PFL0635c as putative interacting partners of PF10_0328 and suggest that these three bromo-domain proteins may operate in a regulatory complex.

To test if these proteins interact with each other through direct protein-protein interactions I performed direct yeast two-hybrid assays. The full length sequences (PF10_0328 and PF11_0254) or gene fragments of around 1000bp covering the entire coding sequence of PFL0635c were cloned into the bait vector (pGBKT7 DNA-BD) and the prey vector (pGADT7 AD) of the Matchmaker™ Gold Yeast Two-Hybrid System (Clontech). Co-transformations of the yeast strain Y2HGOLD were performed as following: pGADT7 AD-PF10_0328 in combination with pGBKT7 DNA-BD-PF11_0254, pGBKT7 DNA-BD-PFL0635c_F1 and pGBKT7 DNA-BD-PFL0635c_F3 as well as pGBKT7 DNA-BD-PF10_0328 in combination with pGADT7 AD-PF11_0254, pGADT7 AD-PFL0635c_F1 and pGADT7 AD-PFL0635c_F2. The yeast strain Y2HGOLD co-transformed with the pGADT7 AD vector containing each of the genes or fragments and the pGBKT7 DNA-BD empty vector (and *vice versa*) was used to assess autoactivation. The analysis of the potential direct interaction of the several combinations of the three bromo-domain proteins of *P. falciparum* using the Y2H system was negative. No colonies were growing, except for the yeast strain co-transformed with the plasmid pair pGADT7 AD-PF10_0328 and pGBKT7 DNA-BD-PF11_0254. Unfortunately, however, pGBKT7 DNA-BD-PF11_0254 alone was tested positive for autoactivation, showing that growth of the double transformed cells was not due to a direct interaction between these two bromo-domain proteins.

The indirect immunofluorescence assay of 3D7/PF10_0328-3xHA/PF11_0254-2xTy and 3D7/PF10_0328-3xHA/PFL0635c-2xTy parasite lines are in support of the MS results and show extensive overlap of localisation patterns within the DAPI-stained area of the nucleus (Figure 5). IFAs were performed with trophozoite stage parasites as the distinction between the cytoplasmic compartment and the nucleus is best visible here.

However, advanced microscopy techniques are absolutely needed in order to draw better conclusions about the actual extent of co-localisation within the nuclear compartment. Overall, these results lead to the conclusion that the three bromo-domain containing proteins form a multimeric protein complex which is involved in chromatin biology of *P. falciparum*.

Conclusion and Outlook

In this study, we provide first insights into possible functional roles for the conserved bromo-motif in *P. falciparum*, which is the key module to recognise the histone modification of acetylation [1][2]. In *P. falciparum*, these histone code 'readers' have not been analysed to date but knowledge about their function will be important in understanding chromatin dynamics and transcriptional regulation in this important parasite. Moreover, the protein-protein interaction between the bromo-domain and the acetylated lysine is of high relevance because its high selectivity makes it attractive for specific therapeutic intervention such as low molecular weight inhibitors [25].

NMR analysis revealed a conserved core structure of the bromo-motif which belongs to a relatively small group of conserved protein domains across eukaryotic lineages [3]. The location of this module within the protein is reported to be variable and the domain arrangements are highly diverse. Within eukaryotes, the domain architecture of chromatin-related proteins carrying a bromo-domain includes other modules like PHD, B-box type zing finger, SET or ATPase domains [8][9]. Remarkably, the *P. falciparum* protein PF10_0328 acquired by evolutionary adaptation an ankyrin motif and this combination is unique to apicomplexan parasites. Interestingly, we have found using a co-immunoprecipitation assay that the bromo-domain protein PF10_0328 somehow interacts with two other proteins containing bromo-domains as well, a full-length motif in PFL0635c and a shortened version in PF11_0254. The assembly of the three bromo-domain proteins into a multimeric protein complex makes sense, as histones are found to be acetylated at multiple positions. There is evidence in yeast that one or more acetyl-lysines are bound *in vitro* by bromo-domains of a large proteineous assembly mediating transcriptional activation [12][13]. Furthermore, the fact that PF10_0328 acquired an ankyrin repeat motif which mediates protein-protein interaction argues for a multimeric protein complex containing multiple bromo-subunits.

We were not able to show the direct interaction of any of the three bromo-domain proteins using the Y2H system. This could either mean that they are not directly interacting, or that *P. falciparum* proteins are not suitable for being tested in the yeast model. Due to lack of time I wasn't able to test if the recombinant proteins were expressed at sufficiently high levels in yeast so the exact reason for this negative result remains currently unknown. One explanation would be that we extracted a large proteinaceous complex where the three bromo-domain proteins are connected via other proteins such that they are positive for interactions by co-IP assays but not for direct interactions using the Y2H system in yeast.

The detection of the histone H2A and the ApiAP2 transcription factor is in line with a conserved role of *P. falciparum* bromo-domain proteins in transcriptional activation. The histone H2A could indicate that the protein PF10_0328 binds directly to acetylated sites on H2A even though it has to be mentioned that bromo-domains have low binding affinities. The ApiAP2 transcription factor could hint to a possible function in gene activation. Indeed, the sequence-specific DNA-binding ApiAP2 factor may bind promoters of a distinct subset of target genes and then recruit the bromo-domain proteins which recognise adjacent histone marks. Clearly, the interactions of the co-purified protein candidates have to be confirmed by using experiments such as pull downs to verify their functions involved in transcriptional regulation.

Finally, MCP1 has recently been found to be a nuclear peroxiredoxin that is able to use thioredoxin and glutaredoxin as reductants and was renamed by Richard *et al.* as PfnPrx [29]. Using chromatin-IP experiments, PfnPrx was determined to be associated with chromatin in a genome-wide manner in *P. falciparum*, which may explain the presence of this highly expressed factor in our co-IP eluate [29].

The results of the immunofluorescence assay show for the three bromo-domain proteins a nuclear localisation pattern which is rather concentrated in several regions throughout the entire nucleus. This observation could point to a more global function of the bromo-domain proteins in association with chromatin dynamics, which is consistent with the high abundance of the proteins across the complete intra-erythrocytic life cycle of *P. falciparum*.

Beside their putative regulatory role in 'reading' epigenetic marks and mediating gene activation, the three bromo-domain proteins are attractive candidates for further studies because they open an absolutely novel research area of targeting intervention strategies against malaria. The modular structure of the bromo-motif with its weak hydrophobic interaction to acetyl-lysines is also highlighted in cancer research as a target for small molecule inhibitors [6][25]. Interestingly, the bromo-motif of the protein PF10_0328 falls into the same category as the prominent BET bromo-domain of humans, which is known to be inhibited by the most potent and selective inhibitors [30][31].

Further experimental results are now needed to analyse these proteins on the functional level. Interestingly, several attempts to generate knockout cell lines failed for all three candidates (data not shown) suggesting they are essential for parasite survival. In collaboration with Zbynek Bozdech (Nanyang Technological University, Singapore) and Mike Duffy (University of Melbourne, Australia) we currently attempt to identify the specific histone residues that are recognised by the three bromo-domains. Furthermore, chromatin immunoprecipitation followed by DNA microarray analysis (ChIP-on-chip) has been performed to map the global distribution of the bromo-domain proteins in the genome of *P. falciparum*. The application of antibodies recognizing site-specific histone modifications will then allow the detection of the acetylation sites bound by the three bromo-domain proteins *in vivo*. As bromo-domains function as conserved structural units for protein-protein interactions [32] I expect that these experiments will give detailed insights into site-specific binding of these important regulators. The bromodomain-dependent recognition by the three *P. falciparum* proteins may then directly trigger the effector functions of lysine acetylation, which is believed to be exclusively transcriptional activation of defined target genes [32].

Materials and Methods

Generation of transfection constructs

To express epitope-tagged proteins in *P. falciparum*, the full-length gene sequence of PF10_0328 was amplified from cDNA and cloned into pBcam-3xHA using *Bam*HI and *Nco*I restriction sites (see previous chapter). The genes PFL0635c and PF11_0254 were amplified from cDNA/gDNA and cloned into *Bam*HI/*Nhe*I-digested pHcam-2xTy [33]. All forward primers included five wild-type nucleotides directly upstream of the natural ATG start codon. In all transfection constructs, expression of the C-terminally tagged fusion proteins is controlled by the *P. falciparum* *cam* promoter. To establish double transfectant parasite lines co-expressing PF10_0328-3xHA/PF11_0254-2xTy or PF10_0328-3xHA/PFL0635c-2xTy, the 2xTy-constructs were transfected into BSD-resistant 3D7/PF10_0328-3xHA parasites and additionally selected with WR.

Parasite culture and transfection

P. falciparum 3D7 parasites were cultured as described previously [34]. Growth synchronisation was achieved by repeated sorbitol lysis [35]. Transfections were performed as described [36] and selected on 5 µg/ml blasticidin-S-HCl and 4nM WR99210. To select for plasmid integration into the endogenous loci (3' replacement and gene deletions) transfected parasites were cycled ON/OFF WR three times (4 weeks each).

Nuclear extraction

Mixed cultures of late trophozoite and schizont parasites were used (~28-40 hpi) to release parasites from RBCs by saponin lysis. Parasites were resuspended in hypotonic cytoplasmic lysis buffer CLB to extract nuclei as previously described [37]. Proteins were either extracted in high salt buffer HSB (15% glycerol, 20mM HEPES (pH 7.9), 1M KCl, 1mM EDTA, 1mM EGTA, 1mM DTT, protease inhibitors) or in RIPA buffer containing 0.5% deoxycholate (RIPA/DOC; 50mM Tris-HCl (pH 7.5), 150mM NaCl, 1mM EDTA, 1% NP-40, 0.5% deoxycholate, protease inhibitors) for 20 min at 4°C under constant agitation. Soluble fractions were recovered by centrifugation at 13,000rpm for 3 min. Pellets were washed once in the same buffer/volume and the supernatant was combined with the previous fraction. Pellets were solubilised in SDS extraction buffer SEB (2% SDS, 10mM Tris-HCl (pH 7.5)) for 20 min at RT under constant agitation. Soluble

fractions were recovered by centrifugation at 13,000rpm for 15 min. Samples were analysed by Western blot.

Co-immunoprecipitation

High salt nuclear and RIPA/DOC nuclear extracts derived from 240ml culture (5-6% parasitaemia) of the transgenic parasite line 3D7/PF10_0328-3xHA were mixed and added at the ratio of 1:1 to binding buffer BB (50mM Tris-HCl (pH 7.5), 1mM EDTA, 1% NP-40, protease inhibitors). This input fraction (6ml) was added to HA-affinity matrix (400ul slurry) (Roche Diagnostics) and the binding step was performed for 3 hr at 4°C under constant agitation. As negative control, specific competition by an excess of soluble HA-peptides (0.1µg/µl) was applied. Four wash steps were performed in the same volume as the binding step with wash buffer WB (50mM Tris-HCl (pH 7.5), 150mM NaCl, 1mM EDTA, 1% NP-40, protease inhibitors). The last wash step was performed in the same volume (1200ul) as the subsequent elution step. Bound proteins were eluted for 1hr at 4°C under constant agitation with wash buffer containing 0.5µg/µl soluble HA-peptides. A second elution step was performed with SDS extraction buffer SEB (2% SDS, 10mM Tris-HCl (pH 7.5)). Centrifugation steps were always performed at 300-500g for 2x1min. Samples were analysed by western blot and silver staining.

Co-immunoprecipitation using the double transgenic cell lines was performed slightly different as described above. High salt nuclear and RIPA/DOC extracts from 3D7/PF10_0328-3xHA/PF11_0254-Ty and 3D7/PF10_0328-3xHA/PFL0635c-Ty were used to confirm LC-MS/MS analysis. The elution step was performed with SDS extraction buffer SEB (2% SDS, 10mM Tris-HCl (pH 7.5)). Samples were analysed by western blot using anti-HA 3F10 and anti-Ty BB2 antibodies.

Mass spectrometry analysis

Elution fractions were precipitated with 10% TCA and dissolved in 50µl 100mM Tris-HCl (pH 8.0). Proteins were digested with trypsin and analysed by capillary liquid chromatography tandem mass spectrometry (LC-MS/MS) using an Orbitrap FT hybrid instrument (Thermo Finnigan, San Jose, CA, USA). MS/MS spectra were searched against a combined *P. falciparum*/human annotated protein database. LC-MS/MS analysis was performed by Paul Jenoe and Suzette Moes (Biozentrum, University of Basel).

Western blot analysis

Protein fractions were separated on 10% SDS-PAGE gels and transferred to nitrocellulose membranes (Schleicher&Schuell). Primary antibody dilutions for rat mAb anti-HA 3F10 (Roche Diagnostics) 1:1000 and for mouse mAb anti-Ty BB2 1:2,000 (kind gift from Keith Gull).

Immunofluorescence assay

IFAs were performed with iRBCs fixed in 4% formaldehyde/0.01% glutaraldehyde as described previously [38]. Primary antibody dilutions for rat mAb anti-HA 3F10 (Roche Diagnostics) 1:100 and for mouse mAb anti-Ty BB2 1:500 (kind gift from Keith Gull). Secondary antibody dilutions for Alexa-Fluor® 568-conjugated anti-rat IgG (Molecular Probes) 1:500 and for FITC-conjugated anti-mouse IgG (Kirkegaard Perry Laboratories) 1:250. Images were taken on a Leica DM 5000B microscope with a Leica DFC 300 FX camera and acquired via the Leica IM 1000 software and processed and overlaid using Adobe Photoshop CS2.

Yeast two-hybrid system

YPDA liquid media, SD-Leu/Trp and SD-Leu/Trp/Ade2/His3+150/500/1000ng/ml aureobasidin A (10mg/ml stock solution) agar plates were prepared as described in the Appendix C of Yeast Protocol Handbook (Clontech, Mountain View, CA). Yeast co-transformation was performed according to the Small Scale LiAc Yeast Transformation Procedure in Chapter V.E.. The yeast strain Y2HGOLD of the Matchmaker™ Gold Yeast Two-Hybrid System (Clontech) was used containing the reporter genes *AbA^r*, *HIS3*, *ADE2* and *MEL1* and *trp1*, *leu2* as transformation markers. The full length gene sequences of PF10_0328 and PF11_0254 were amplified from cDNA and cloned into prey vector pGADT7 AD and the bait vector pGBKT7 DNA-BD of the Matchmaker™ Gold Y2H System using the restriction sites *EcoRI/XhoI* and *EcoRI/SalI*, respectively. The full length sequence of PFL0635c was divided into three regions of about 1000bp. Fragments 1, 2 and 3 of PFL0635c were amplified from gDNA and cloned in vectors pGADT7 AD and pGBKT7 DNA-BD. Sequences of pGADT7 AD PFL0635c fragment 3 and pGBKT7 DNA-BD PFL0635c fragment 2 had to be excluded due to sequence errors. The yeast strain Y2HGOLD was co-transformed with the pGADT7 AD vector containing each of the genes or fragments and the pGBKT7 DNA-BD empty vector (and *vice versa*) for autoactivation studies. The yeast strain Y2HGOLD was then co-

transformed with the pGADT7 AD-PF10_0328 in combination with pGBKT7 DNA-BD-PF11_0254, pGBKT7 DNA-BD-PFL0635c_F1 and pGBKT7 DNA-BD-PFL0635c_F3 and *vice versa* (pGBKT7 DNA-BD-PF10_0328 in combination with pGADT7 AD-PF11_0254, pGADT7 AD-PFL0635c_F1 and pGADT7 AD-PFL0635c_F2) for protein-protein interaction studies. Yeast strain Y2HGOLD containing two plasmids were first plated out on selective media lacking SD-Trp/Leu and then replated on SD-Trp/Leu/Ade2/His3 plates with increasing aureobasidin A concentrations.

References

1. Zeng L, Zhou M-M (2002) Bromodomain: an acetyl-lysine binding domain. *FEBS Letters* 513: 124–128. doi:10.1016/S0014-5793(01)03309-9.
2. Jacobson RH, Ladurner AG, King DS, Tjian R (2000) Structure and function of a human TAFII250 double bromodomain module. *Science* 288: 1422–1425.
3. Dhalluin C, Carlson JE, Zeng L, He C, Aggarwal AK, et al. (1999) Structure and ligand of a histone acetyltransferase bromodomain. *Nature* 399: 491–496. doi:10.1038/20974.
4. Owen DJ, Ornaghi P, Yang JC, Lowe N, Evans PR, et al. (2000) The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase gcn5p. *EMBO J* 19: 6141–6149. doi:10.1093/emboj/19.22.6141.
5. Jeanmougin F, Wurtz JM, Le Douarin B, Chambon P, Losson R (1997) The bromodomain revisited. *Trends Biochem Sci* 22: 151–153.
6. Hudson BP, Martinez-Yamout MA, Dyson HJ, Wright PE (2000) Solution structure and acetyl-lysine binding activity of the GCN5 bromodomain. *J Mol Biol* 304: 355–370. doi:10.1006/jmbi.2000.4207.
7. Schultz J, Copley RR, Doerks T, Ponting CP, Bork P (2000) SMART: A Web-Based Tool for the Study of Genetically Mobile Domains. *Nucl Acids Res* 28: 231–234. doi:10.1093/nar/28.1.231.
8. Basu MK, Poliakov E, Rogozin IB (2009) Domain mobility in proteins: functional and evolutionary implications. *Brief Bioinformatics* 10: 205–216. doi:10.1093/bib/bbn057.
9. Basu MK, Carmel L, Rogozin IB, Koonin EV (2008) Evolution of protein domain promiscuity in eukaryotes. *Genome Res* 18: 449–461. doi:10.1101/gr.6943508.
10. Strahl BD, Allis CD (2000) The language of covalent histone modifications. *Nature* 403: 41–45. doi:10.1038/47412.
11. Dyson MH, Rose S, Mahadevan LC (2001) Acetyllysine-binding and function of bromodomain-containing proteins in chromatin. *Front Biosci* 6: D853–865.
12. Cairns BR, Lorch Y, Li Y, Zhang M, Lacomis L, et al. (1996) RSC, an essential, abundant chromatin-remodeling complex. *Cell* 87: 1249–1260.
13. Cairns BR, Schlichter A, Erdjument-Bromage H, Tempst P, Kornberg RD, et al. (1999) Two functionally distinct forms of the RSC nucleosome-remodeling complex, containing essential AT hook, BAH, and bromodomains. *Mol Cell* 4: 715–723.
14. Iyer LM, Anantharaman V, Wolf MY, Aravind L (2008) Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes. *Int J Parasitol* 38: 1–31. doi:10.1016/j.ijpara.2007.07.018.
15. Marcotte EM, Pellegrini M, Yeates TO, Eisenberg D (1999) A census of protein repeats. *J Mol Biol* 293: 151–160. doi:10.1006/jmbi.1999.3136.
16. Heringa J (1998) Detection of internal repeats: how common are they? *Current Opinion in Structural Biology* 8: 338–345. doi:10.1016/S0959-440X(98)80068-7.
17. Mosavi LK, Minor DL Jr, Peng Z-Y (2002) Consensus-derived structural determinants of the ankyrin repeat motif. *Proc Natl Acad Sci USA* 99: 16029–16034. doi:10.1073/pnas.252537899.
18. Cui L, Miao J, Furuya T, Li X, Su X, et al. (2007) PfGCN5-mediated histone H3 acetylation plays a key role in gene expression in *Plasmodium falciparum*. *Eukaryotic Cell* 6: 1219–1227. doi:10.1128/EC.00062-07.
19. Fan Q, An L, Cui L (2004) *Plasmodium falciparum* histone acetyltransferase, a yeast GCN5 homologue involved in chromatin remodeling. *Eukaryotic Cell* 3: 264–276.
20. Fan Q, An L, Cui L (2004) PfADA2, a *Plasmodium falciparum* homologue of the transcriptional coactivator ADA2 and its *in vivo* association with the histone acetyltransferase PfGCN5. *Gene* 336: 251–261. doi:10.1016/j.gene.2004.04.005.

21. Volz J, Carvalho TG, Ralph SA, Gilson P, Thompson J, et al. (2010) Potential epigenetic regulatory proteins localise to distinct nuclear sub-compartments in *Plasmodium falciparum*. *Int J Parasitol* 40: 109–121. doi:10.1016/j.ijpara.2009.09.002.
22. LaCount DJ, Vignali M, Chettier R, Phansalkar A, Bell R, et al. (2005) A protein interaction network of the malaria parasite *Plasmodium falciparum*. *Nature* 438: 103–107. doi:10.1038/nature04104.
23. Bischoff E, Vaquero C (2010) In silico and biological survey of transcription-associated proteins implicated in the transcriptional machinery during the erythrocytic development of *Plasmodium falciparum*. *BMC Genomics* 11: 34. doi:10.1186/1471-2164-11-34.
24. Chung C-W (2012) Small molecule bromodomain inhibitors: extending the druggable genome. *Prog Med Chem* 51: 1–55. doi:10.1016/B978-0-12-396493-9.00001-7.
25. Chung C-W, Witherington J (2011) Progress in the discovery of small-molecule inhibitors of bromodomain–histone interactions. *J Biomol Screen* 16: 1170–1185. doi:10.1177/1087057111421372.
26. Bozdech Z, Llinás M, Pulliam BL, Wong ED, Zhu J, et al. (2003) The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS Biol* 1: E5. doi:10.1371/journal.pbio.0000005.
27. Jacobson RH, Ladurner AG, King DS, Tjian R (2000) Structure and function of a human TAFII250 double bromodomain module. *Science* 288: 1422–1425.
28. Flueck C, Bartfai R, Niederwieser I, Witmer K, Alako BTF, et al. (2010) A major role for the *Plasmodium falciparum* ApiAP2 protein PfSIP2 in chromosome end biology. *PLoS Pathog* 6: e1000784. doi:10.1371/journal.ppat.1000784.
29. Richard D, Bartfai R, Volz J, Ralph SA, Muller S, et al. (2011) A genome-wide chromatin-associated nuclear peroxiredoxin from the malaria parasite *Plasmodium falciparum*. *J Biol Chem* 286: 11746–11755. doi:10.1074/jbc.M110.198499.
30. Filippakopoulos P, Qi J, Picaud S, Shen Y, Smith WB, et al. (2010) Selective inhibition of BET bromodomains. *Nature* 468: 1067–1073. doi:10.1038/nature09504.
31. Delmore JE, Issa GC, Lemieux ME, Rahl PB, Shi J, et al. (2011) BET bromodomain inhibition as a therapeutic strategy to target c-Myc. *Cell* 146: 904–917. doi:10.1016/j.cell.2011.08.017.
32. Mujtaba S, Zeng L, Zhou M-M (2007) Structure and acetyl-lysine recognition of the bromodomain. *Oncogene* 26: 5521–5527. doi:10.1038/sj.onc.1210618.
33. Flueck C, Bartfai R, Volz J, Niederwieser I, Salcedo-Amaya AM, et al. (2009) *Plasmodium falciparum* heterochromatin protein 1 marks genomic loci linked to phenotypic variation of exported virulence factors. *PLoS Pathog* 5: e1000569. doi:10.1371/journal.ppat.1000569.
34. Trager W, Jensen JB (1978) Cultivation of malarial parasites. *Nature* 273: 621–622.
35. Lambros C, Vanderberg JP (1979) Synchronization of *Plasmodium falciparum* erythrocytic stages in culture. *J Parasitol* 65: 418–420.
36. Voss TS, Healer J, Marty AJ, Duffy MF, Thompson JK, et al. (2006) A var gene promoter controls allelic exclusion of virulence genes in *Plasmodium falciparum* malaria. *Nature* 439: 1004–1008. doi:10.1038/nature04407.
37. Voss TS, Mini T, Jenoe P, Beck H-P (2002) *Plasmodium falciparum* possesses a cell cycle-regulated short type replication protein A large subunit encoded by an unusual transcript. *J Biol Chem* 277: 17493–17501. doi:10.1074/jbc.M200100200.
38. Tonkin CJ, van Dooren GG, Spurck TP, Struck NS, Good RT, et al. (2004) Localization of organellar proteins in *Plasmodium falciparum* using a novel set of transfection vectors and a new immunofluorescence fixation method. *Molecular and Biochemical Parasitology* 137: 13–21. doi:10.1016/j.molbiopara.2004.05.009.

Figure legends

Figure 1

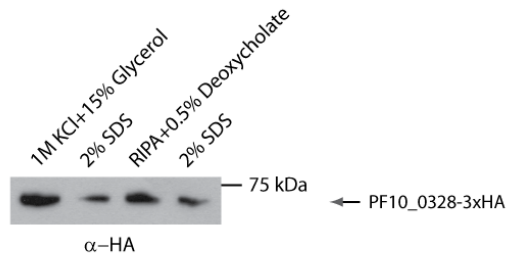
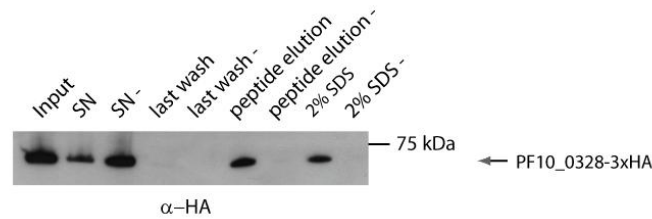


Figure 1. Optimal extraction conditions of the bromo-domain protein PF10_0328-3xHA.

Western blot analysis using rat mAb anti-HA 3F10 demonstrates good solubility of PF10_0328-3xHA under high salt (lane 1) and RIPA+0.5% deoxycholate (lane 3) conditions, followed by treatment of the insoluble pellet with 2% SDS (lane 2 and 4) (single transgenic parasite line 3D7/PF10_0328-3xHA).

Figure 2

A



B

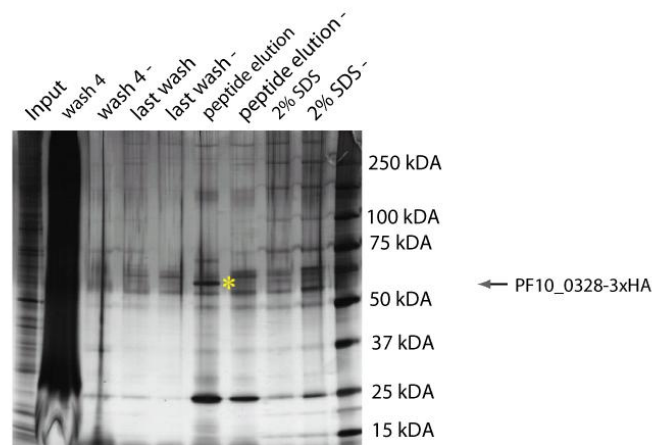


Figure 2. Large scale purification of putative PF10_0328-interacting protein candidates.

Co-immunoprecipitation analysis using mixed nuclear extracts of high salt and RIPA/DOC conditions derived from the transgenic parasite line 3D7/PF10_0328-3xHA. (A) Western blot analysis demonstrates efficient binding of PF10_0328-3xHA to the HA-affinity matrix as evident by the reduced band intensity in the supernatant fraction compared to the input fraction (lane 2 to lane 1). As negative control, an excess of HA-peptides ($0.1 \mu\text{g}/\mu\text{l}$) was added to block binding sites of the HA-affinity matrix (lane 3). Last wash steps performed in the same volume as the elution steps are clean (lane 4 and 5). PF10_0328-3xHA was eluted by HA-peptide competition ($0.5 \mu\text{g}/\mu\text{l}$) for LC-MS/MS analysis (lane 6). The peptide elution step of the negative control was performed equally and shows no band (lane 7). The second elution step was performed using SDS and detected additional PF10_0328-3xHA in the positive run (lane 8) and no protein in the negative control (lane 9). (B) Silver staining with the same samples as in (A) was performed to determine if additional proteins were pulled down together with PF10_0328-3xHA. The black smear in wash 4 (lane 2) is of unknown origin but could be explained by DNA-contamination. Eluate fractions of the positive run and the negative control of the co-immunoprecipitation were analysed by LC-MS/MS (lane 6 and 7). Note the band at roughly 65 kDa in the positive co-IP eluate (lane 6) that is absent in the negative eluate (lane 7), which most likely represents PF10_0328-3xHA (star).

Table 1

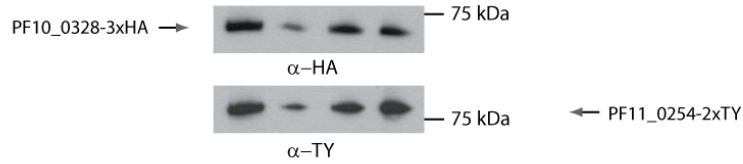
PlasmoDB Accession Nr.	Annotation	Protein score	MW kDa	Unique peptides
PFL0635c	bromodomain protein, putative	167.53	126	20
PF10_0328	bromodomain protein, putative	108.74	56	12
PF11_0254	conserved Plasmodium protein, unknown function	34.85	62	4
PFF0860c	histone H2A	11.81	14	2
PF10_0268	merozoite capping protein 1	4	44	1
PFF0670w	transcription factor with AP2 domain(s), putative	3.49	459	1
PF14_0379	conserved Plasmodium protein, unknown function	2.71	259	1

Table 1. Putative PF10_0328-interacting protein candidates identified by LC-MS/MS.

Accession numbers and corresponding annotations of the proteins exclusively detected in the PF10_0328-bound fraction are shown (excluding two annotated and putative ribosomal proteins) (www.plasmodb.org). The numbers indicate unique tryptic peptides, the protein scores and the sizes (in kDa) of the bait protein PF10_0328 (grey) and the co-purified protein candidates.

Figure 3

A



B

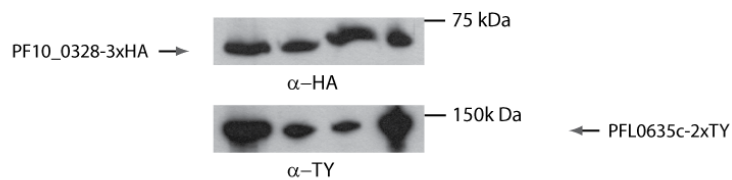


Figure 3. Optimal extraction conditions of the three bromo-domain proteins PF10_0328-3xHA, PF11_0254-2xTy and PFL0635c-2xTy using double cell lines.

Western blot analysis using rat mAb anti-HA 3F10 demonstrate good solubility of PF10_0328-3xHA (A and B, upper panel) under high salt (lane 1) and RIPA+0.5% deoxycholate (lane 3) conditions, followed by treatment of the insoluble pellet with 2% SDS (lane 2 and 4). Western blot analysis using mouse mAb anti-Ty BB2 demonstrate good solubility of PF11_0254-2xTy (A, lower panel) and PFL0635c-2xTy (B, lower panel) under high salt conditions (lane 1), whereas it detected reduced extraction efficiency of PF11_0254-2xTy (A) and PFL0635c-2xTy (B) using RIPA+0.5% deoxycholate buffer (lane 3) compared to the SDS fractions (lane 4) (double transgenic parasite lines 3D7/PF10_0328-3xHA/PF11_0254-2xTy (A) and 3D7/PF10_0328-3xHA/PFL0635c-2xTy (B)).

Figure 4

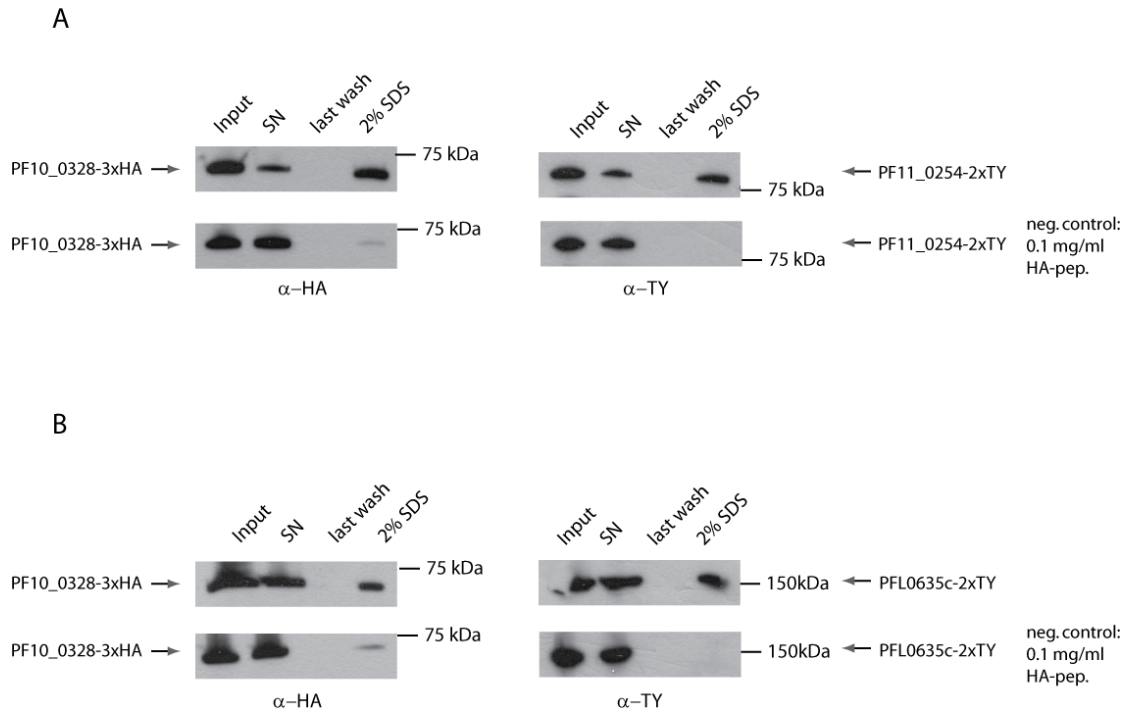


Figure 4. Identification of the two bromo-domain proteins PF11_0254 and PFL0635c as the most promising PF10_0328-interacting proteins.

Co-IPs were performed under the same conditions as for LC-MS/MS analysis using the double transgenic parasite lines 3D7/PF10_0328-3xHA/PF11_0254-2xTy and 3D7/PF10_0328-3xHA/PFL0635c-2xTy. Upper panel: Western blot analysis using rat mAb anti-HA 3F10 detect the bait protein PF10_0328-3xHA in the input, supernatant and eluates (A and B, lanes 1, 2 and 4). The co-purified binding partners PF11_0254-2xTy (A) and PFL0635c-2xTy (B) were detected using mouse mAb anti-Ty BB2 antibodies and showed the same appearance pattern. Lower panel: Negative control co-IPs using an excess of soluble HA-peptides show drastically reduced binding for the bait protein PF10_0328-3xHA (A and B, lane 4) and no binding for the interaction partners PF11_0254-2xTy (A, lane 4) and PFL0635c-2xTy (B, lane 4) in the SDS elution fractions.

Figure 5

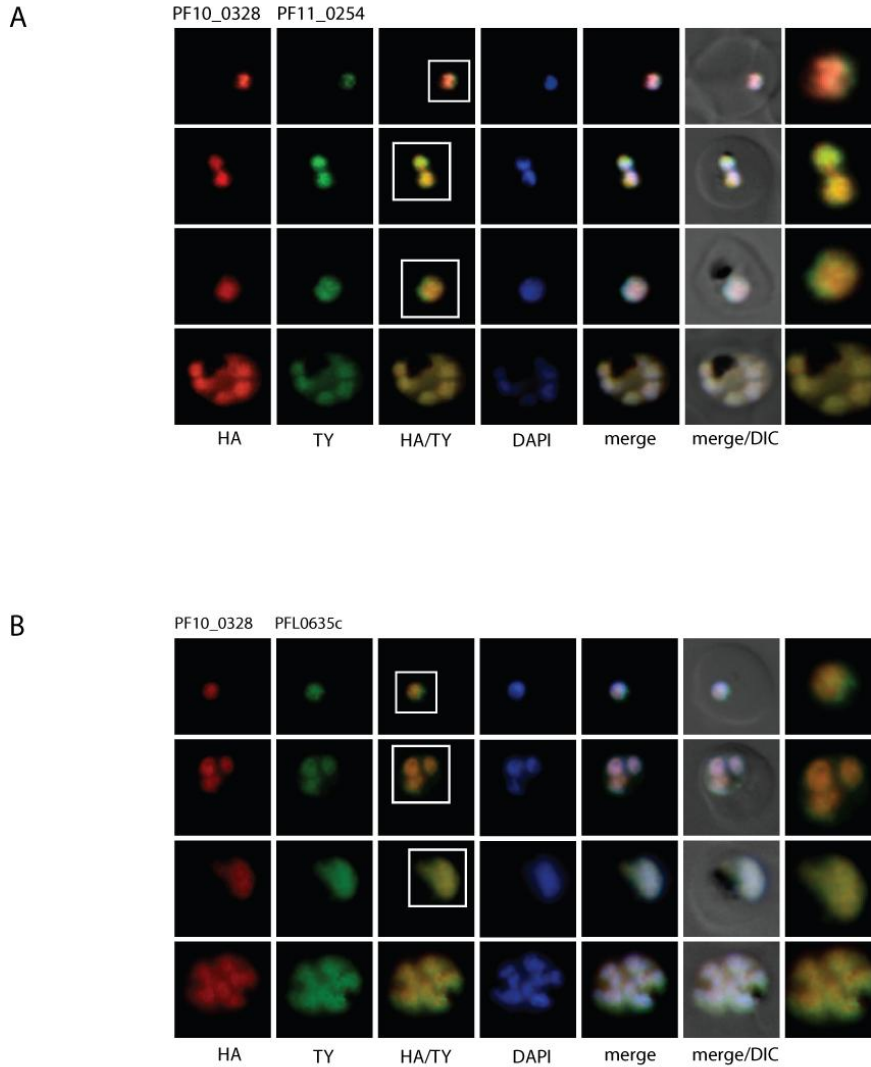


Figure 5. Combined nuclear localisation patterns of PF10_0328, PF11_0254 and PFL0635c show concentrated overlapping regions within the DAPI-stained area.

Transgenic parasites at the trophozoite stage were used and nuclei were stained with DAPI. Co-localisation of PF10_0328-3xHA and PF11_0254-2xTy as well as PF10_0328-3xHA and PFL0635c-2xTy is detectable within large overlapping areas using the double transgenic parasite lines 3D7/PF10_0328-3xHA/PF11_0254-2xTy (A) and 3D7/PF10_0328-3xHA/PFL0635c-2xTy (B).

Table 2.

Primer sequences used in this chapter (restriction sites are in bold)

pHcam-2xTY

PF11_0254_F	CAGT GGATCC AAAATATGGCTGATAATAACCTAG
PF11_0254_R	CAGT CCATGGG TAAAGAATCCTTTGATGAATCG
PFL0635c_F	GATC GGATCC TAAATATGAATGATTATAATAATAAATAATC
PFL0635c_R	GATC GCTAGC TATTTTAGCGTCATAACCTTCG

pGADT7 AD and pGBKT7 DNA-BD

PF10_0328_Y2H_F	GAT CCATATG ATGATGCTTGAAGATTTTAATAC
PF10_0328_Y2H_R	GAT CCTCGAG TCAATCTTTTTTTGAACTATCC
PF11_0254_Y2H_F	GATC GAATTC ATGGCTGATAATAACC
PF11_0254_Y2H_R	GAT CCTCGAG TTAGTAAGAATCC
PFL0635c_F1_Y2H_F	GATC CAATTG AATATGAATGATTATAATAATAAATAATC
PFL0635c_F1_Y2H_R	GAT CCTCGAG TCGTACCTTTATTATTTTCCTTTGG
PFL0635c_F2_Y2H_F	GAT CCATATG CCAAAGGAAATAATAAAGGTACGA
PFL0635c_F2_Y2H_R	GAT CCTCGAG TTATTATTTTCCTTTGG
PFL0635c_F3_Y2H_F	GATC CAATTG AATAAGAGATAATATAAAGAACAAAAAGG
PFL0635c_F3_Y2H_R	GAT CCTCGAG TTATATTTTAGCGTC

Chapter 4: Visualizing the inner life of the nuclear compartment of *P. falciparum* (working paper)

Introduction

The eukaryotic nucleus is the center of the processes that coordinate the delivery of nuclear-encoded genetic information to the protein synthesis machinery of the cytoplasm. It's a highly dynamic compartment and consists of internal structures, which are functionally connected. At the ultrastructural level, electron-sparse and electron-dense regions of chromatin can be distinguished, representing active (euchromatin) and silenced (heterochromatin) transcription areas, respectively [1]. The chromatin organisation changes during the cell cycle and many studies have demonstrated a regulatory role in gene expression, such as the transition of active genes, characterized by decondensed regions and accessible for the transcriptional machinery, to condensed, inaccessible and silent regions [2][3][4]. The interchromatin space is filled with nucleoplasm, a viscous liquid surrounded by a bilaminar nuclear envelope (NE), the outer membrane of which is continuous with the endoplasmic reticulum (ER). In higher eukaryotes, the inner surface of the double-layered membrane of the nucleus contains a network of intermediate filaments, the nuclear lamina, which binds chromatin, integral membrane proteins, and other nuclear components [5]. The NE is interrupted by nuclear pore complexes (NPC), which allow for bi-directional molecular trafficking between the cytoplasm and the nucleus. In recent publications, a broader range of cellular activities beyond the primary functions of the NPCs has been assumed [6]. Experimental evidence suggests that the inner basket structures of the NPCs are involved in the regulation of gene expression at the transcriptional and post-transcriptional level, with particular focus on perinuclear expression sites where gene positioning was linked to transcriptional activation within the commonly silenced nuclear periphery [7][8]. Using ultra-structural techniques a spatial relationship between active genes and nuclear pores as well as between silent chromatin and non-pore sites at the periphery could be shown [9][10]. Apart from that, the nuclear compartment of eukaryotes comprehends the nucleolus as its most prominent structure that is mainly composed of a fine ribonucleoprotein network [11]. The inner space of the nucleolus consists of zones rich in granules, which encircle dense microfilaments and show a close correlation between

size proportion and production level [12]. The classic nucleolar functions involve synthesis of ribosomal subunits starting with rDNA transcription and followed by multiple rounds of rRNA processing [12].

In *P. falciparum*, a recent publication of Weiner and Dahan-Pasternak provides observations of the interrelation between the nuclear pores and chromatin dynamics across the IDC [13]. Cryogenic scanning electron microscopy revealed for the first time surface views of the cytoplasmic and nucleoplasmic faces of NPCs in *P. falciparum*. The NPCs are characterized by a conserved proteineous structure of around 80 nm in size with an uneven distribution around the NE at most developmental stages of the IDC. Furthermore, a 3D model reconstructed by focused ion beam microscopy uncovered dynamic changes in the numbers of the nuclear pores as well as in its positioning coupled to chromatin organization. At the ring stage, the parasite possesses three to seven closely clustered pores per nucleus accompanied by an undifferentiated, diffuse chromatin pattern. At the trophozoite stage, the biogenesis of nuclear pores starts and their numbers increase from 12 to 58 uniformly distributed pores per nucleus. At this stage, the chromatin appears as defined heterochromatic patches filling the nucleoplasm of *P. falciparum*. With the onset of schizogony, dramatic changes take place as the number of nuclear pores decrease to two to six per divided nucleus and nuclear pores now appear clustered and are surrounded by decondensed euchromatin [13].

Immunofluorescence imaging using antibodies against the nucleolar proteins *P. falciparum* fibrillarin (PfNop1; PF14_0068) and PfNop5 (PF10_0085), two subunits of rRNA-processing ribonucleoprotein particles, characterized the nucleolus of *P. falciparum* as a 'hat-like' structure across the asexual life cycle [14][15][16]. In terms of size the *P. falciparum* nucleolus resembles that of *S. cerevisiae*; however both organisms differ in the shape of the nucleolus, which appears crescent-shaped in yeast [17]. The observation that in *P. falciparum* a single nucleolar body is formed attracts interest as its rDNA copies (18S-5.8S-28S) are found to be distributed as single units over the subtelomeres of different chromosomes (chr 1,5,7,11,13) without being linked to each other. In yeast, chromosomal deletion studies revealed that the genomic organization of tandemly repeated rDNA seems to interfere with the number of nucleoli per cell nucleus [18]. For example, budding yeast encodes around 150 rDNA genes on a single chromosome (chr 12), assembling to one large nucleolus in contrast to the

multiple nucleoli of humans where approximately 400 rDNA genes are located as tandem arrangements on five different chromosomes (chr 13,14,15,21,22) [19][20]. At the pre-replicative stage of intra-erythrocytic *P. falciparum*, two α -amanitin-resistant nucleolar transcription sites were observed using Br-UTP incorporation assays. Further, the chromosome ends harbouring rDNA genes appear to cluster at the nucleolus independent from their transcriptional status (active rDNA genes A1 and A2; silenced rDNA genes S1 and S2). With the beginning of nuclear division, nucleolar organization undergoes cell-cycle dependent changes in *P. falciparum* and clustered rDNA loci disassemble [14].

The knowledge about interrelated spatial organization of nuclear landmark structures like chromosome ends, nuclear pores and the nucleolus is surprisingly limited in *P. falciparum*. The chromosome ends are characterised by terminal telomere repeat arrays followed by six noncoding telomere-associated repeat elements (TARE 1-6). PfHP1 is the major heterochromatin component of chromosome ends as well as some internal-chromosomal regions, which are predominantly associated with *P. falciparum*-specific virulence gene clusters [21][22]. IFA revealed that PfHP1 localizes as two to seven clusters at the nuclear periphery of *P. falciparum* [22][21]. PfHP1 is a crucial silencing factor to regulate the variegated expression of variant surface protein families (such as PfEMP1, RIFIN, STEVOR, PfMC-2TM). This mediates antigenic variation, which facilitates immune evasion during blood stage infection to survive in the host environment of the erythrocyte [23][24][25][26]. In case of the 60-member *var* gene family, the parasite expresses only a single *var* gene a time while all other members maintain in a transcriptional silent state [27][28][29]. The genetic elements which regulate mutual exclusion of *var* gene transcription are controversially discussed, especially the functional role of *var* gene introns and 5' flanking regions [30][31][32]. The majority of the 60 *var* members locate at the highly polymorphic chromosome ends and some are found in central chromosomal heterochromatic regions [33][34]. *Var* gene silencing is linked to the epigenetic mark H3K9me3, which is recognised by PfHP1 leading to heterochromatin formation of subtelomeric and chromosome-internal *var* genes like already mentioned above [21][22][35]. In contrast, the histone modifications H3K9ac and H3K4me3 are associated with transcriptional active *var* genes [36][35][37]. Interestingly, switching of *var* expression from active to silent requires the 'repositioning' into an active zone at the nuclear periphery [38][39]. The mechanism of chromosomal

tethering and clustering at the nuclear periphery is not identified to date [40][41], but actin polymerisation is supposed to be involved in this process [42].

The proteinaceous composition of the *P. falciparum* nuclear pore complex is largely unknown as only two members have been identified to date and their localization was described by IFA; the nucleoporin NUP100 (PfNUP100; PFI0250c) [43] and the nucleoporin NUP116 (PfNUP116; PF14_0706) [44]. Concerning the nucleolus, several proteins are known to reside in this nuclear structure, and our nuclear proteome study detected 12 annotated nucleolar *P. falciparum* proteins. Despite that, characterisation of individual nucleolar proteins on the functional level remains open in *P. falciparum*. Thus, visualization of the nuclear architecture by its structural components like the nuclear pores and the nucleolus is an important first step towards understanding their functional interrelation.

Previously, we identified two novel *P. falciparum* nuclear porins (NUPs) based on the presence of phenylalanine-glycine (FG) pairs, and a novel nucleolar protein (see chapter 2; MSc thesis Johanna Wetzel, 2011). Remarkably, fluorescence microscopic studies revealed that the PfNUPs defined the nuclear border as a distinct ring-shape around the DAPI-stained area, and the ribosomal factory as a single intense spot within a DAPI-free cavity of the nucleus. To characterize the spatial behaviour of these nuclear structures in relation to chromosome end dynamics, I used transgenic parasite lines expressing tagged nuclear pore or nucleolus markers and analysed their localisation in relation to chromosome ends using PfHP1-specific antibodies. These preliminary results establish our cell lines as promising tools to study functional interrelations between these landmark structures using a variety of approaches including confocal and 4-D microscopy.

Results

Visualization of the chromosome ends in live and fixed *P. falciparum* parasites

The well-known localization of PfHP1 shows that chromosome ends cluster as two to seven foci at the nuclear periphery of *P. falciparum*. Flueck and colleagues used three different transgenic parasite lines to verify nuclear localization including ectopically expressed PfHP1-3xHA and PfHP1-2xTy, and 3D7/PfHP1-GFP expressing a PfHP1-GFP fusion protein under the control of its endogenous promoter [22]. However, all these results were obtained by IFA experiments using methanol- or formaldehyde/glutaraldehyde-fixed parasites and anti-tag antibodies. Here, my goal was to confirm a perinuclear localization pattern for PfHP1 for the first time by the means of the autofluorescent marker CherryFP in live cells. The established transgenic parasite line 3D7/PfHP1-CherryFP expresses the tagged version from episomal plasmids and provided intense signals suitable for live cell imaging. Thus, we were able to confirm the characteristic PfHP1 pattern as precise red spots at the nuclear periphery of merozoite-, ring-, trophozoite- and schizont-stage parasites (Fig. 1). Interestingly, chromosome ends in a subset of parasites appeared to cluster at one single location in late schizont-, merozoite and ring- stages. On average I observed between three and five PfHP1-cherryFP signals per nucleus throughout all IDC stages of *P. falciparum*. In particular, we were interested in the behaviour of PfHP1-CherryFP at the developmental stage of *P. falciparum* where it undergoes schizogony. However, the quantification of peripheral PfHP1-CherryFP spots per nucleus proved to be difficult during schizogony, as parasite nuclei are small and appeared disordered during multiple rounds of genome replication (Fig. 1). Thus, we were not able to observe whether DNA replication may cause the disruption of chromosome end clusters using a basic Leica DM 5000B epi-fluorescence microscope. To verify this possibility, time-lapse microscopy is required to monitor movements of cell structures over time, in our case the *P. falciparum* chromosome ends during the rounds of DNA synthesis/ mitosis. Moreover, the observation in merozoites and early ring stages that chromosome ends within some nuclei appear to assemble into a single chromosome end cluster with the onset of cytokinesis, has to be considered with caution as they could also be explained by microscope-based limitations.

In addition to live cell imaging, I validated affinity-purified polyclonal rabbit antibodies raised against recombinant full length PfHP1 (Igor Niederwieser, unpublished) in

methanol-fixed 3D7 wild-type parasites (Fig. 2). As seen with PfHP1-CherryFP in live cells, and as reported in the study by Flueck and colleagues using epitope-tagged PfHP1 [22], IFAs using anti-PfHP1 antibodies highlighted chromosome end clustering at the nuclear periphery throughout intra-erythrocytic development. Thus, the transgenic parasite line 3D7/PfHP1-CherryFP and the generated anti-PfHP1 antibodies are proven to be very suitable tools to visualise chromosome ends during the IDC in live and fixed cells.

Visualization of the nuclear pores across the asexual life cycle of *P. falciparum* using 3D7/PFI0170w-GFP and 3D7/PF14_0442-GFP cell lines

Only two unique member of the nuclear pore complex, a homologue to the *S. cerevisiae* protein Nup100 (PfNUP100; PFI0250c) and NUP116 (PfNUP116; PF14_0706), have been identified in *P. falciparum* to date [43][45]. We identified two novel nuclear pore candidates (PF14_0442, PFI0170w). Both proteins, PFI0170w and PF14_0442, possess repeated phenylalanine-glycine (FG) domains, which are characteristic and essential components of the proteinaceous inner pore assembly directing the permeability barrier [46][47]. Furthermore, PFI0170w and PF14_0442 share their enormous protein sizes with those of other known NUPs as found in *S. cerevisiae* [48]. We generated transgenic parasite lines 3D7/PFI0170w-GFP, 3D7/PFI0170w-3xHA and 3D7/PF14_0442-GFP, where the endogenous proteins of these novel NUP candidates were successfully tagged at the C-terminus by plasmid integration into the endogenous gene locus by single crossover recombination (Fig. 3 and MSc thesis Johanna Wetzel, 2011). These cell lines provided excellent tools to characterize *P. falciparum* nuclear pores in both fixed and live cells. As live cell microscopy of PFI0170w-GFP- and PF14_0442-GFP-expressing parasites did not show adequate fluorescent signals, parasites were released by saponin treatment from surrounding erythrocytes to increase sensitivity of detection. In this assay, the localization pattern of PF14_0442-GFP describes a uniformly perforated ring-structure surrounding the DAPI-stained area at the trophozoite stage of *P. falciparum* (Fig. 4). At anytime during schizogony, nuclear pores marked by PF14_0442-GFP were observed to cluster within several regions and aggregates appeared to persist until the ring stage of *P. falciparum* (Fig. 4). PFI0170w-GFP displayed a similar pattern but showed insufficient signals in most of the parasites despite saponin treatment, and despite the fact that PFI0170w-3xHA parasites showed very clear nuclear pore signals in IFAs using methanol-fixed cells (MSc thesis Johanna

Wetzel, 2011). Therefore, I performed IFAs of both transgenic parasite lines using antibodies against the GFP-tag. Indirect localization patterns of the NUPs show distinct ring-like structures surrounding the NE and co-staining with the ER-marker BIP confirmed their NPC origin as they occupy the 'inner face' of the ER compartment (Fig. 5).

In conclusion, our microscopy-based data of PF14_0442 and PFI0170w clearly suggest that both proteins are indeed novel structural components of *P. falciparum* NPCs. In addition, nuclear pore positioning was shown to be highly dynamic during parasite progression, which is consistent with results recently published by Weiner and Dahan-Pasternak [13].

Size increase of the nucleolar compartment at the trophozoite stage of *P. falciparum* using 3D7/PF11_0250-3xHA

To gain more insights into the morphological appearance of the nucleolus, the transgenic parasite line 3D7/PF11_0250-3xHA, which expresses the nucleolar protein PF11_0250 as C-terminal 3xHA fusion from episomal plasmids, was used in IFA analysis. This putative splicing factor was found to localize to one distinct spot within the parasitic nucleus and additional co-staining with α -human fibrillarin confirmed the nucleolar origin (see chapter 2). Here, I performed more detailed analysis of this factor during the IDC. At ring- and late schizont stages, the size of the nucleolus appeared more or less unchanged, whereas during trophozoite maturation it increased considerably in size and made up to one-third of the total volume of the nucleus (Fig. 6). This dynamic behaviour of the *P. falciparum* nucleolus is consistent with microscopic-based observations of the nucleolar compartment in budding yeast [17].

Interrelated visualization of the silenced chromosome ends, the nuclear pores and the nucleolus in *P. falciparum*

To begin to investigate interrelative aspects of heterochromatic chromosome ends, the nuclear pores and the nucleolus I used PfHP1 specific antibodies in combination with the HA-tagged transgenic parasite lines of 3D7/PFI0170w-3xHA (endogenously tagged nucleoporin; MSc thesis Johanna Wetzel, 2011) and 3D7/PF11_0250-3xHA (episomally tagged nucleolar protein). Double labeling immunofluorescence experiments demonstrated that the nucleolus was not in close association to chromosome end clusters at the nuclear periphery (Fig. 7A). Chromosome end clusters appeared to

occupy peripheral areas adjacent and internal to, but mostly not overlapping with, nuclear pores encircling the nuclear compartment (Fig. 7B).

Conclusion and Outlook

In this study, we deliver first insights into the spatial relationship of nuclear landmark structures by fluorescence-based observations across the IDC of *P. falciparum*. Using 3D7/PfHP1-CherryFP, we confirmed the perinuclear clustering of chromosome ends by live cell imaging. This finding is relevant in light of the potential importance of perinuclear chromosome end clustering in increased recombination events between heterologous chromosome ends and thus the generation of virulence gene diversity [41][33]. In particular, we turned our attention on chromosome end behaviour of *P. falciparum* during the phase of asynchronous replication where three to four rounds of DNA synthesis/mitosis are followed by cytokinesis. Unfortunately, determination of peripheral PfHP1-derived foci in relation to dividing nuclei was difficult even in live cells as the transition phases between DNA synthesis and mitosis are unclear. Only after the onset of cytokinesis single nuclei determination by DAPI-staining was feasible. Interestingly, chromosome ends of some merozoite- and ring stage parasites appeared to assemble into one cluster at a particular region of the nuclear periphery. However, this observation needs to be interpreted with caution since it could also be an artefact due to failed detection of PfHP1 foci outside the microscopic focus. Remarkably, the visualization of 'heterochromatin' through chromosomal end clusters and 'active peripheral zones' represented by the nuclear pores revealed a distinct spatial segregation. The fluorescent foci of antibodies targeting PfHP1 did not localize to the staining of the nuclear pores at any time. Our results are therefore consistent with the assumption that nuclear pores provide 'open' or active domains which are spatially separated from silent chromatin patches found in nuclei of mammalian cells [9][10]. Furthermore, the peripheral location of some PfHP1 foci, visible in relation to the nuclear pores, argues for a nuclear envelope-anchoring mechanism as observed in budding yeast [49][50].

The NPC patterns observed in live cells by GFP-tagging and after para-formaldehyde/glutaraldehyde fixation clearly show that both proteins, PFI0170w and PF14_0442, are members of the NPC in *P. falciparum*. In light of the GFP-tagged FG-NUP patterns, we observed a continuous distribution of the nuclear pores at the

trophozoite stage of *P. falciparum*, which confirmed the results of Weiner and Dahan-Pasternak [13]. This observation may indicate that morphological changes of the parasite at the trophozoite stage require an increased nucleo-cytoplasmic transport as dynamics in NPC organization was associated with developmental changes of the cell in other eukaryotes [51]. The apparent tendency of nuclear pores to cluster at ring- and late schizont stages might reflect an unknown mechanism for equal division of nuclear pores to daughter nuclei, most likely orientated at the mitotic spindles and embedded within the NE during mitosis. The fact that parasites had to be released from iRBCs in order to detect the autofluorescent GFP signal shows that the GFP signal emitted from both FG-NUP-GFP fusions is rather weak. This could be explained by the possibility that both FG-NUP proteins are positioned inside of the large proteinaceous pore complex, where they are expected to act as a selective gate for nucleo-cytoplasmic transport, or by weak expression of both proteins in general.

We characterized the nucleolus of *P. falciparum* as a distinct rounded region filling an apparent 'DAPI-free' sub-region within the nucleus, by the use of antibodies directed against the HA-tagged protein PF11_0250 (annotated as a putative splicing factor). Interestingly, the size of the compartment increased substantially, up to approximately one third of the nuclear volume at the trophozoite stage compared to the size in the other stages where it occupies only approximately one-tenth of the nuclear volume. Possibly, this reflects increased activity of ribosomal biosynthesis in *P. falciparum* as nucleolar appearance varies depending on cellular activity, proliferation or differentiation [52]. Co-staining with the nucleolar marker α -human fibrillarin confirmed the observed dynamics of the nucleolus in *P. falciparum*.

Our observations are somewhat in contrast to previous immuno-fluorescence analyses using antibodies against PfNop1/fibrillarin, which characterized the parasite nucleolus as 'hat-like' structure [14][16]. PfNop1/fibrillarin is involved in pre-rRNA processing by binding small nucleolar RNAs, which is essential for cell viability and has been observed to be conserved throughout eukaryotes. For the first time, we showed that the nucleolar body was never found in close association to the chromosome ends of *P. falciparum* using the transgenic parasite line 3D7/PF11_0250-3xHA combined with antibodies against PfHP1. Furthermore, we observed only in ring, trophozoite and late schizont stage parasites a fluorescence signal of PF11_0250-3xHA in line with the results of

Mancio-Silva *et al.* which show that the rDNA cluster disperses, the nucleolus disassembles and rRNA transcription dramatically decreases with the onset of mitosis [14].

Fluorescence imaging is a powerful tool to visualize the inner life of the *P. falciparum* nuclear compartment including several function-based structural components across the asexual life cycle stages. The nucleus of *P. falciparum* was for the first time characterized by the spatial arrangement of silenced 'heterochromatin', the ribosomal factory and the checkpoints of nuclear-cytoplasmic transport. Moreover, confocal time-lapse microscopy using our transgenic parasite lines has the power to provide a more complete range of dynamic behaviours of these and other nuclear landmarks structures across the IDC of *P. falciparum*.

Materials and Methods

Generation of transfection constructs

The full-length gene sequence of PFL1005c was amplified from gDNA and cloned into pBcam-3xHA-CherryFP using *Bam*HI and *Nco*I restriction sites (T. Voss, unpublished). Parasites expressing PF11_0250-3xHA are described in chapter 2. In both transfection constructs, expression of the C-terminally tagged fusion proteins is controlled by the *P. falciparum cam* promoter. 3' replacement constructs of pH-PFI0170w-3xHA, pH-PFI0170w-GFP and pH-PF14_0442-GFP were kindly provided by J. Wetzel (MSc thesis, 2011).

Antibodies

A polyclonal rabbit anti-PfHP1 antibody was prepared by Pacific Immunology; rabbits were immunized with recombinantly expressed PfHP1-6xHIS fusion (full length). Anti-PfHP1 antibodies were then affinity-purified on recombinant full-length PfHP1 and validated by western blot and IFA (Igor Niederwieser and N. Brancucci, unpublished).

Parasite culture and transfection

P. falciparum 3D7 parasites were cultured as described previously [53]. Growth synchronisation was achieved by repeated sorbitol treatment [54]. Transfections were

performed as described previously [55] and following selected on 5 µg/ml blasticidin-S-HCl and/or 4nM WR99210.

Selection for plasmid integration into the endogenous locus (3' replacement strategy)

In order to select for plasmid integration into the endogenous locus (3' replacement of pH-PFI0170w-GFP, pH-PF14_0442-GFP) through a single crossover event, transfected parasites were cycled ON and OFF three times for four weeks each on selective drug WR99210. Transgenic parasite line 3D7/PFI0170w-3xHA expressing the fusion protein version under its endogenous promoter was performed by J. Wetzel (MSc thesis, 2011). Successful plasmid integration was confirmed for constructs pH-PFI0170w-GFP and pH-PF14_0442-GFP by western blot as well as by PCR. For pH-PFI0170w-3xHA plasmid integration was confirmed by southern blot analysis (J. Wetzel, MSc thesis, 2011).

Live cell imaging by fluorescence microscopy

Parasites of the cell lines 3D7/PFI0170w-GFP and PF14_0442-GFP (mixed stages of rings, trophozoites and schizonts) were first released from RBCs by saponin lysis and incubated for 10 min at RT with 1 µl DAPI (200 µg/ml). Transgenic parasite line 3D7/PfHP1-CherryFP was incubated with 1 µl DAPI (200 µg/ml) for 10 min at RT. Parasites or infected red blood cells were subsequently analysed by fluorescence microscopy. Images were taken on a Leica DM 5000B microscope with a Leica DFC 300 FX camera and acquired via the Leica IM 1000 software and processed and overlaid using Adobe Photoshop CS2.

Indirect immunofluorescence assay

IFA analysis was performed either with 4% para-formaldehyde/0.01% glutaraldehyde or methanol-fixed iRBCs as described before [56]. Primary antibody dilutions: rat mAb anti-HA 3F10 (Roche Diagnostics) 1:100, rabbit pAb anti-PfHP1 (Pacific Immunology) 1:100, mouse mAb anti-GFP (Roche Diagnostics). Secondary antibody dilutions: Alexa-Fluor® 568-conjugated anti-rat IgG (Molecular Probes) 1:500; Alexa-Fluor® 568-conjugated anti-rabbit IgG (Molecular Probes), 1:500 FITC-conjugated anti-mouse IgG (Kirkegaard Perry Laboratories) 1:250. Images were taken on a Leica DM 5000B microscope with a Leica DFC 300 FX camera and acquired via the Leica IM 1000 software and processed and overlaid using Adobe Photoshop CS2.

References

1. Gahan PB (2006) Cell biology (Revised Reprint). T. D. Pollard and W. C. Earnshaw, Saunders- Elsevier, 813 pp., ISBN 0- 7216- 3360- 9 (2004). Cell Biochemistry and Function 24: 92–92. doi:10.1002/cbf.1268.
2. Chakalova L, Debrand E, Mitchell JA, Osborne CS, Fraser P (2005) Replication and transcription: shaping the landscape of the genome. Nat Rev Genet 6: 669–677. doi:10.1038/nrg1673.
3. Hakim O, Sung M-H, Hager GL (2010) 3D shortcuts to gene regulation. Curr Opin Cell Biol 22: 305–313. doi:10.1016/j.ceb.2010.04.005.
4. Madan Babu M, Janga SC, de Santiago I, Pombo A (2008) Eukaryotic gene regulation in three dimensions and its impact on genome evolution. Current Opinion in Genetics & Development 18: 571–582. doi:10.1016/j.gde.2008.10.002.
5. Strambio-De-Castillia C, Niepel M, Rout MP (2010) The nuclear pore complex: bridging nuclear transport and gene regulation. Nat Rev Mol Cell Biol 11: 490–501. doi:10.1038/nrm2928.
6. Lo SJ, Lee C-C, Lai H-J (2006) The nucleolus: reviewing oldies to have new understandings. Cell Res 16: 530–538. doi:10.1038/sj.cr.7310070.
7. Brickner JH (2009) Transcriptional memory at the nuclear periphery. Curr Opin Cell Biol 21: 127–133. doi:10.1016/j.ceb.2009.01.007.
8. Towbin BD, Meister P, Gasser SM (2009) The nuclear envelope--a scaffold for silencing? Curr Opin Genet Dev 19: 180–186. doi:10.1016/j.gde.2009.01.006.
9. Brown CR, Silver PA (2007) Transcriptional regulation at the nuclear pore complex. Curr Opin Genet Dev 17: 100–106. doi:10.1016/j.gde.2007.02.005.
10. Akhtar A, Gasser SM (2007) The nuclear envelope and transcriptional control. Nat Rev Genet 8: 507–517. doi:10.1038/nrg2122.
11. Raska I (2003) Oldies but goldies: searching for Christmas trees within the nucleolar architecture. Trends Cell Biol 13: 517–525.
12. Olson MO, Dundr M, Szebeni A (2000) The nucleolus: an old factory with unexpected capabilities. Trends Cell Biol 10: 189–196.
13. Weiner A, Dahan-Pasternak N, Shimoni E, Shinder V, von Huth P, et al. (2011) 3D nuclear architecture reveals coupled cell cycle dynamics of chromatin and nuclear pores in the malaria parasite *Plasmodium falciparum*. Cell Microbiol 13: 967–977. doi:10.1111/j.1462-5822.2011.01592.x.
14. Mancio-Silva L, Zhang Q, Scheidig-Benatar C, Scherf A (2010) Clustering of dispersed ribosomal DNA and its role in gene regulation and chromosome-end associations in malaria parasites. Proc Natl Acad Sci USA 107: 15117–15122. doi:10.1073/pnas.1001045107.
15. Mancio-Silva L, Rojas-Meza AP, Vargas M, Scherf A, Hernandez-Rivas R (2008) Differential association of Orc1 and Sir2 proteins to telomeric domains in *Plasmodium falciparum*. J Cell Sci 121: 2046–2053. doi:10.1242/jcs.026427.
16. Figueiredo LM, Rocha EPC, Mancio-Silva L, Prevost C, Hernandez-Verdun D, et al. (2005) The unusually large *Plasmodium* telomerase reverse-transcriptase localizes in a discrete compartment associated with the nucleolus. Nucleic Acids Res 33: 1111–1122. doi:10.1093/nar/gki260.
17. Warner JR (1990) The nucleolus and ribosome formation. Curr Opin Cell Biol 2: 521–527.
18. Oakes M, Aris JP, Brockenbrough JS, Wai H, Vu L, et al. (1998) Mutational analysis of the structure and localization of the nucleolus in the yeast *Saccharomyces cerevisiae*. J Cell Biol 143: 23–34.
19. Carmo-Fonseca M, Mendes-Soares L, Campos I (2000) To be or not to be in the nucleolus. Nat Cell Biol 2: E107–112. doi:10.1038/35014078.
20. McStay B, Grummt I (2008) The epigenetics of rRNA genes: from molecular to chromosome biology. Annu Rev Cell Dev Biol 24: 131–157. doi:10.1146/annurev.cellbio.24.110707.175259.
21. Pérez-Toledo K, Rojas-Meza AP, Mancio-Silva L, Hernández-Cuevas NA, Delgado DM, et al. (2009) *Plasmodium falciparum* heterochromatin protein 1 binds to tri-methylated histone 3 lysine 9 and is linked to mutually exclusive expression of var genes. Nucleic Acids Res 37: 2596–2606. doi:10.1093/nar/gkp115.
22. Flueck C, Bartfai R, Volz J, Niederwieser I, Salcedo-Amaya AM, et al. (2009) *Plasmodium falciparum* heterochromatin protein 1 marks genomic loci linked to phenotypic variation of exported virulence factors. PLoS Pathog 5: e1000569. doi:10.1371/journal.ppat.1000569.
23. Lavazec C, Sanyal S, Templeton TJ (2007) Expression switching in the *stevor* and *Pfmc-2TM* superfamilies in *Plasmodium falciparum*. Mol Microbiol 64: 1621–1634. doi:10.1111/j.1365-2958.2007.05767.x.
24. Mok BW, Ribacke U, Winter G, Yip BH, Tan C-S, et al. (2007) Comparative transcriptomal analysis of isogenic *Plasmodium falciparum* clones of distinct antigenic and adhesive phenotypes. Mol Biochem Parasitol 151: 184–192. doi:10.1016/j.molbiopara.2006.11.006.
25. Niang M, Yan Yam X, Preiser PR (2009) The *Plasmodium falciparum* STEVOR multigene family mediates antigenic variation of the infected erythrocyte. PLoS Pathog 5: e1000307. doi:10.1371/journal.ppat.1000307.
26. Witmer K, Schmid CD, Brancucci NMB, Luah Y-H, Preiser PR, et al. (2012) Analysis of subtelomeric virulence gene families in *Plasmodium falciparum* by comparative transcriptional profiling. Mol Microbiol 84: 243–259. doi:10.1111/j.1365-2958.2012.08019.x.

27. Gardner MJ, Hall N, Fung E, White O, Berriman M, et al. (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* 419: 498–511. doi:10.1038/nature01097.
28. Smith JD, Chitnis CE, Craig AG, Roberts DJ, Hudson-Taylor DE, et al. (1995) Switches in expression of *Plasmodium falciparum* var genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. *Cell* 82: 101–110.
29. Su XZ, Heatwole VM, Wertheimer SP, Guinet F, Herrfeldt JA, et al. (1995) The large diverse gene family var encodes proteins involved in cytoadherence and antigenic variation of *Plasmodium falciparum*-infected erythrocytes. *Cell* 82: 89–100.
30. Dzikowski R, Li F, Amulic B, Eisberg A, Frank M, et al. (2007) Mechanisms underlying mutually exclusive expression of virulence genes by malaria parasites. *EMBO Rep* 8: 959–965. doi:10.1038/sj.embor.7401063.
31. Calderwood MS, Gannoun-Zaki L, Wellems TE, Deitsch KW (2003) *Plasmodium falciparum* var genes are regulated by two regions with separate promoters, one upstream of the coding region and a second within the intron. *J Biol Chem* 278: 34125–34132. doi:10.1074/jbc.M213065200.
32. Deitsch KW, Calderwood MS, Wellems TE (2001) Malaria. Cooperative silencing elements in var genes. *Nature* 412: 875–876. doi:10.1038/35091146.
33. Hernandez-Rivas R, Mattei D, Sterkers Y, Peterson DS, Wellems TE, et al. (1997) Expressed var genes are found in *Plasmodium falciparum* subtelomeric regions. *Mol Cell Biol* 17: 604–611.
34. Rubio JP, Thompson JK, Cowman AF (1996) The var genes of *Plasmodium falciparum* are located in the subtelomeric region of most chromosomes. *EMBO J* 15: 4069–4077.
35. Chookajorn T, Dzikowski R, Frank M, Li F, Jiواني AZ, et al. (2007) Epigenetic memory at malaria virulence genes. *Proc Natl Acad Sci USA* 104: 899–902. doi:10.1073/pnas.0609084103.
36. Lopez-Rubio JJ, Gontijo AM, Nunes MC, Issar N, Hernandez Rivas R, et al. (2007) 5' flanking region of var genes nucleate histone modification patterns linked to phenotypic inheritance of virulence traits in malaria parasites. *Mol Microbiol* 66: 1296–1305. doi:10.1111/j.1365-2958.2007.06009.x.
37. Cui L, Miao J, Furuya T, Li X, Su X, et al. (2007) PfGCN5-mediated histone H3 acetylation plays a key role in gene expression in *Plasmodium falciparum*. *Eukaryotic Cell* 6: 1219–1227. doi:10.1128/EC.00062-07.
38. Volz JC, Bártfai R, Petter M, Langer C, Josling GA, et al. (2012) PfSET10, a *Plasmodium falciparum* methyltransferase, maintains the active var gene in a poised state during parasite division. *Cell Host Microbe* 11: 7–18. doi:10.1016/j.chom.2011.11.011.
39. Duraisingh MT, Voss TS, Marty AJ, Duffy MF, Good RT, et al. (2005) Heterochromatin silencing and locus repositioning linked to regulation of virulence genes in *Plasmodium falciparum*. *Cell* 121: 13–24. doi:10.1016/j.cell.2005.01.036.
40. Freitas-Junior LH, Bottius E, Pirrit LA, Deitsch KW, Scheidig C, et al. (2000) Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of *P. falciparum*. *Nature* 407: 1018–1022. doi:10.1038/35039531.
41. Marty AJ, Thompson JK, Duffy MF, Voss TS, Cowman AF, et al. (2006) Evidence that *Plasmodium falciparum* chromosome end clusters are cross-linked by protein and are the sites of both virulence gene silencing and activation. *Mol Microbiol* 62: 72–83. doi:10.1111/j.1365-2958.2006.05364.x.
42. Zhang Q, Huang Y, Zhang Y, Fang X, Claes A, et al. (2011) A Critical Role of Perinuclear Filamentous Actin in Spatial Repositioning and Mutually Exclusive Expression of Virulence Genes in Malaria Parasites. *Cell Host & Microbe* 10: 451–463. doi:10.1016/j.chom.2011.09.013.
43. Volz J, Carvalho TG, Ralph SA, Gilson P, Thompson J, et al. (2010) Potential epigenetic regulatory proteins localise to distinct nuclear sub-compartments in *Plasmodium falciparum*. *Int J Parasitol* 40: 109–121. doi:10.1016/j.ijpara.2009.09.002.
44. Lopez-Rubio J-J, Mancio-Silva L, Scherf A (2009) Genome-wide analysis of heterochromatin associates clonally variant gene regulation with perinuclear repressive centers in malaria parasites. *Cell Host Microbe* 5: 179–190. doi:10.1016/j.chom.2008.12.012.
45. Wente SR, Rout MP, Blobel G (1992) A new family of yeast nuclear pore complex proteins. *J Cell Biol* 119: 705–723.
46. Peleg O, Lim RYH (2010) Converging on the function of intrinsically disordered nucleoporins in the nuclear pore complex. *Biol Chem* 391: 719–730. doi:10.1515/BC.2010.092.
47. Strawn LA, Shen T, Shulga N, Goldfarb DS, Wente SR (2004) Minimal nuclear pore complexes define FG repeat domains essential for transport. *Nat Cell Biol* 6: 197–206. doi:10.1038/ncb1097.
48. Terry LJ, Wente SR (2009) Flexible gates: dynamic topologies and functions for FG nucleoporins in nucleocytoplasmic transport. *Eukaryotic Cell* 8: 1814–1827. doi:10.1128/EC.00225-09.
49. Hediger F, Neumann FR, Van Houwe G, Dubrana K, Gasser SM (2002) Live imaging of telomeres: yKu and Sir proteins define redundant telomere-anchoring pathways in yeast. *Curr Biol* 12: 2076–2089.
50. Gotta M, Laroche T, Formenton A, Maillet L, Scherthan H, et al. (1996) The Clustering of Telomeres and Colocalization with Rap1, Sir3, and Sir4 Proteins in Wild-Type *Saccharomyces Cerevisiae*. *J Cell Biol* 134: 1349–1363. doi:10.1083/jcb.134.6.1349.
51. Maeshima K, Yahata K, Sasaki Y, Nakatomi R, Tachibana T, et al. (2006) Cell-cycle-dependent dynamics of nuclear pores: pore-free islands and lamins. *J Cell Sci* 119: 4442–4451. doi:10.1242/jcs.03207.

52. Mélése T, Xue Z (1995) The nucleolus: an organelle formed by the act of building a ribosome. *Curr Opin Cell Biol* 7: 319–324.
53. Trager W, Jensen JB (1978) Cultivation of malarial parasites. *Nature* 273: 621–622.
54. Lambros C, Vanderberg JP (1979) Synchronization of *Plasmodium falciparum* erythrocytic stages in culture. *J Parasitol* 65: 418–420.
55. Voss TS, Healer J, Marty AJ, Duffy MF, Thompson JK, et al. (2006) A var gene promoter controls allelic exclusion of virulence genes in *Plasmodium falciparum* malaria. *Nature* 439: 1004–1008. doi:10.1038/nature04407.
56. Tonkin CJ, van Dooren GG, Spurck TP, Struck NS, Good RT, et al. (2004) Localization of organellar proteins in *Plasmodium falciparum* using a novel set of transfection vectors and a new immunofluorescence fixation method. *Molecular and Biochemical Parasitology* 137: 13–21. doi:10.1016/j.molbiopara.2004.05.009.

Figure legends

Figure 1

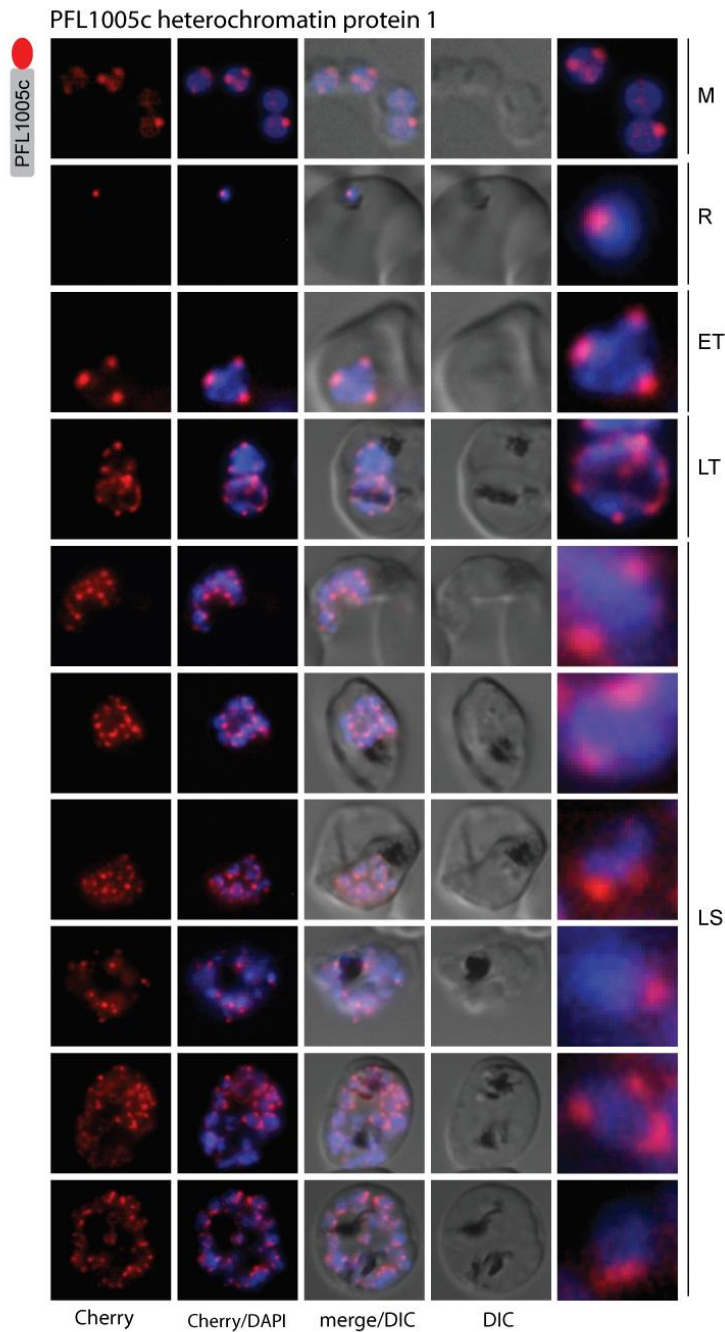


Figure 1. Live imaging shows chromosome end clusters at ring-, trophozoite-, early schizont, late schizont- and merozoite stage parasites of *P. falciparum*.

The transgenic parasite line 3D7/PfHP1-CherryFP expressing the epitope-tagged protein version from episomal plasmids was used and analyzed across the IDC (merozoite- (M), ring- (R), early (ET) and late (LT) trophozoite- and late schizont (LS) stage). Images represent PfHP1-CherryFP localisation (red), the DAPI stained nucleus (blue) and differential interference contrast (DIC).

Figure 2

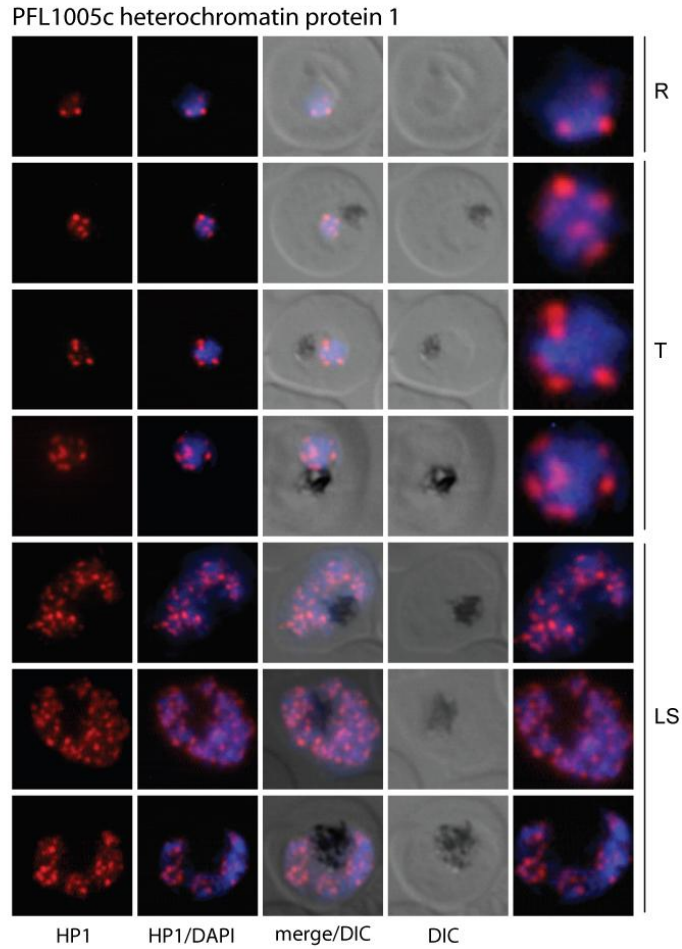


Figure 2. Immunofluorescence-based confirmation of perinuclear localization of *P. falciparum* subtelomeric clusters.

3D7-infected erythrocytes were fixed with methanol and immunostained with polyclonal antibodies against PfHP1. The images taken across the IDC of *P. falciparum* represent PfHP1 localisation (red), the DAPI stained nucleus (blue) and differential interference contrast (DIC) (ring- (R), trophozoite- (T) and late schizont stage (LS)).

Figure 3

A

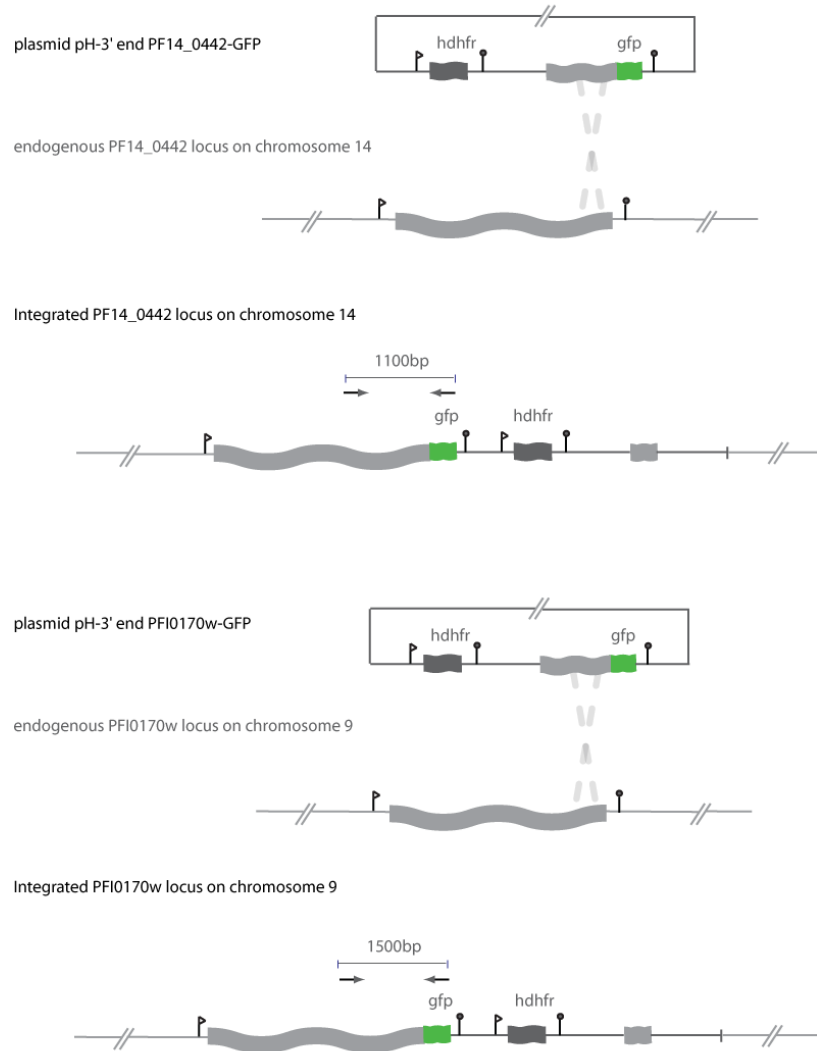
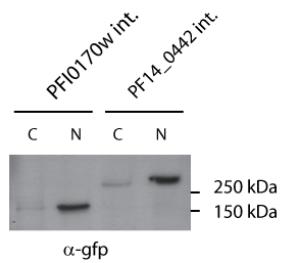


Figure 3. C-terminal tagging of two nuclear pore proteins with GFP by single crossover integration.

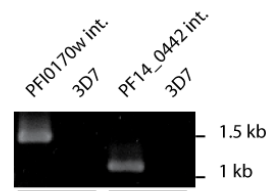
(A) Schematic representation of the integration events of pH-PF14_0442-GFP and pH-PFI0170w-GFP into the endogenous locus on chromosome 14 and 9, respectively. The sizes of the expected PCR fragments for integrated plasmids are indicated.

Figure 3

B



C

**Figure 3. C-terminal tagging of two nuclear pore proteins with GFP by single crossover integration.**

(B) Western blot analysis using antibodies against GFP detect the endogenously tagged nuclear pore proteins PFI0170w and PF14_0442 of predicted sizes in the nuclear fractions. (C) PCR on gDNA isolated from parasites transfected with pH-PFI0170w-GFP or pH-PF14_0442-GFP after three "ON-and-OFF" drug cycles. Integration occurred for PFI0170w and PF14_0442.

Figure 4

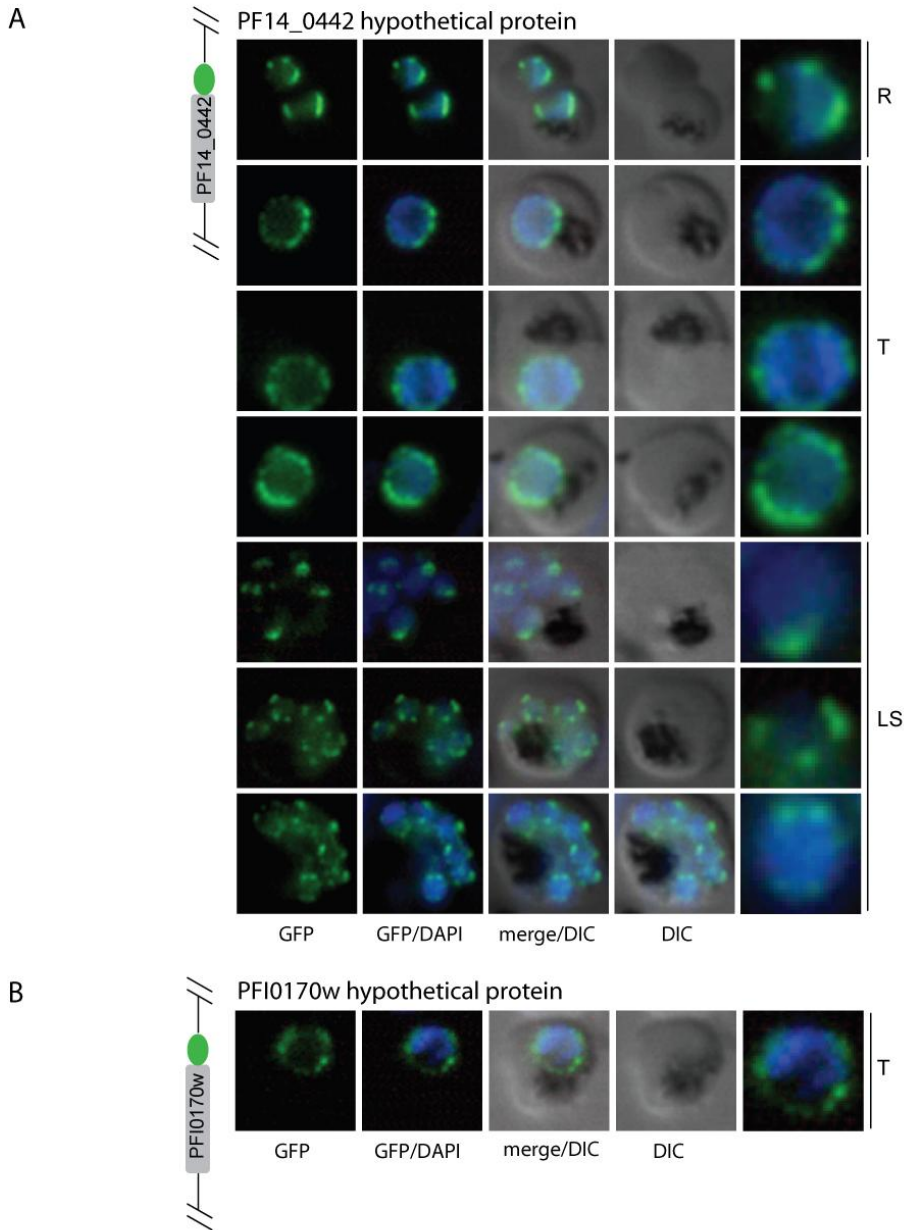


Figure 4. Dynamic positioning of the nuclear pores is stage-dependent across the IDC of *P. falciparum* by live imaging.

Live cell microscopy of isolated parasites expressing PF14_0442-GFP and PFI0170w-GFP fusion proteins under their endogenous promoters. Localization patterns show equal distributions of the nuclear pores around the NE at trophozoite stage, in contrast to ring- and schizont stages, at which they appear clustered at the NE. The images taken across the IDC of *P. falciparum* represent PF14_0442-GFP and PFI0170w-GFP localisation (green), the DAPI stained nucleus (blue) and differential interference contrast (DIC) (ring- (R), trophozoite- (T) and late schizont (LS) stage).

Figure 5

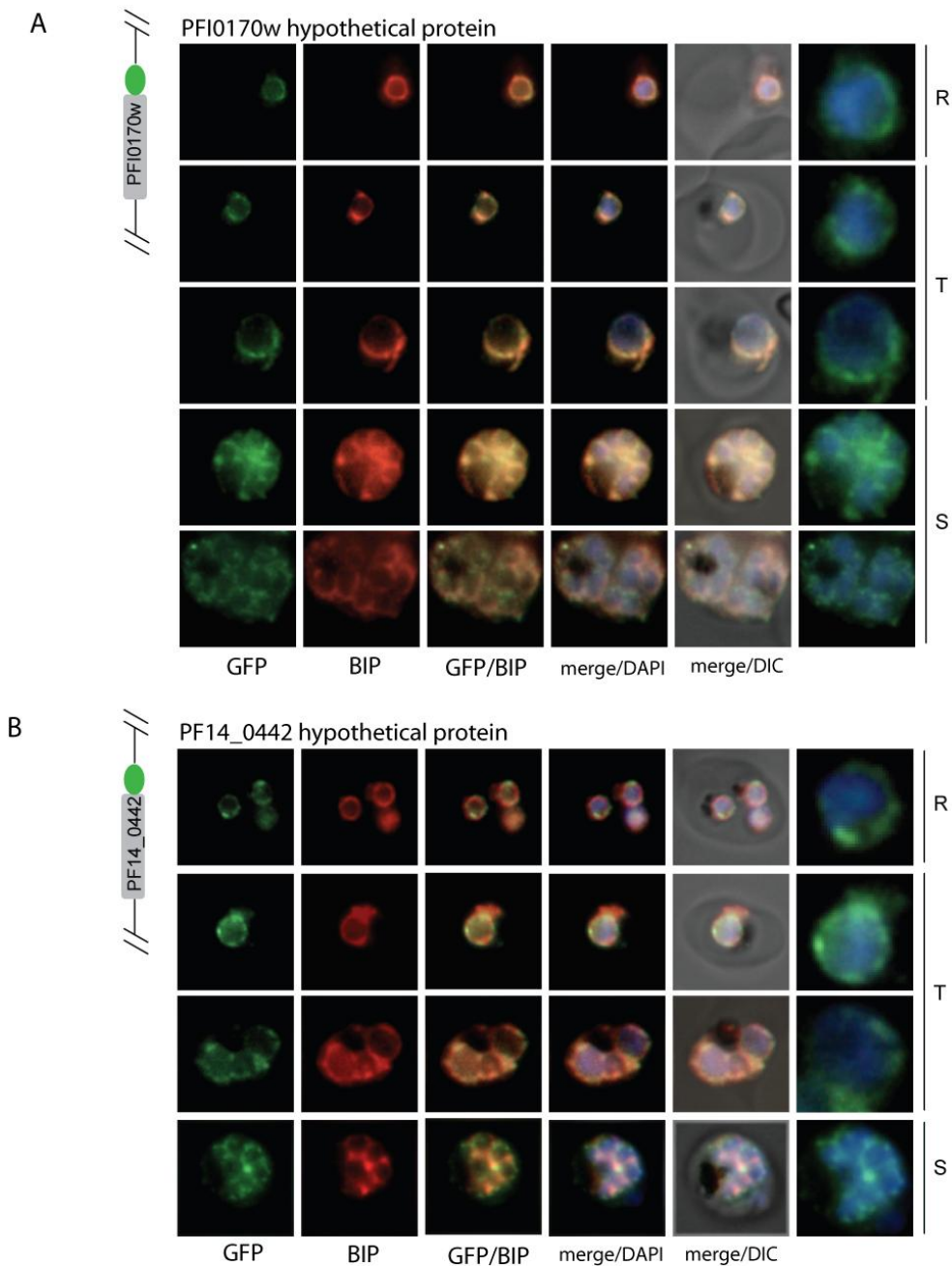


Figure 5. Immunofluorescence assay describes the nuclear pores as a ring-like structure internal to the BIP-staining in *P. falciparum*-infected erythrocytes.

Immunolocalization of FG-NUPs PF14_0442 and PFI0170w in 4% para-formaldehyde/0.01% glutaraldehyde-fixed 3D7/PF14_0442-GFP and 3D7/PFI0170w-GFP parasites which were endogenous tagged by single cross-over events. The images taken across the IDC of *P. falciparum* represent PF14_0442 and PFI0170w localisations (green), the DAPI stained nucleus (blue) and differential interference contrast (DIC) (ring- (R), trophozoite- (T) and schizont stage (S)). Co-staining was performed using antibodies against the ER-marker BIP (red).

Figure 6

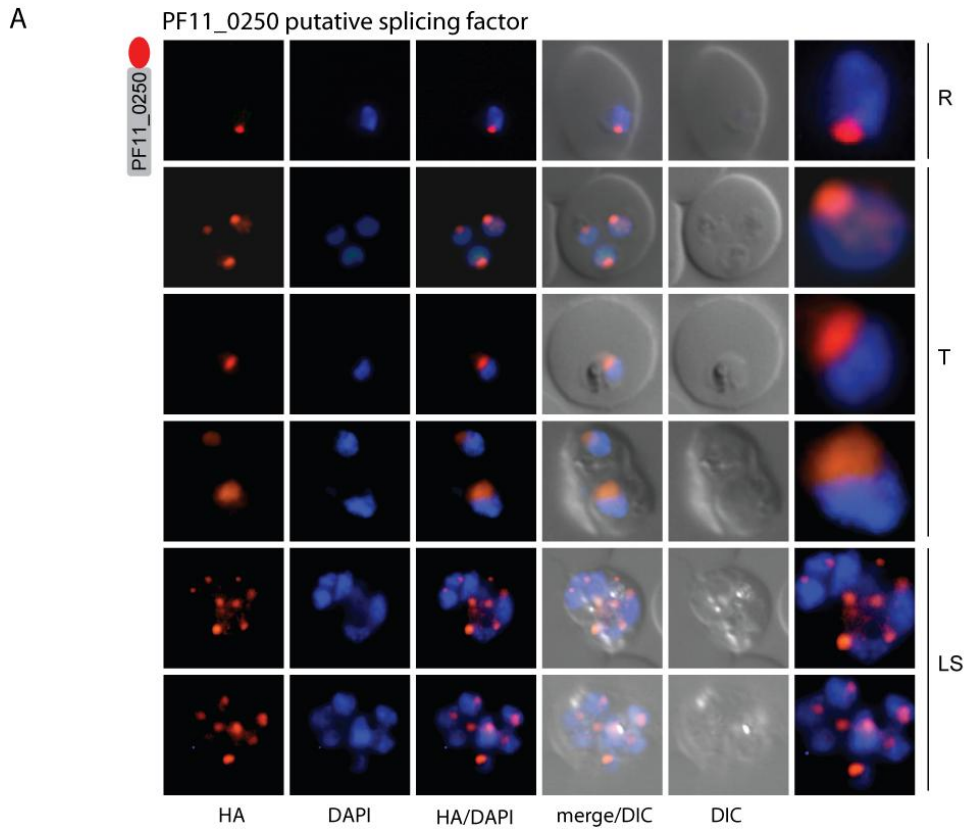


Figure 6. Dynamic size increase of the nucleolar compartment at the pre-replicative developmental stage of *P. falciparum*.

Immunofluorescence analysis of 4% para-formaldehyde/0.01% glutaraldehyde fixated parasites expressing the epitope-tagged protein version PF11_0250-3xHA from episomal plasmids. The images taken across the IDC of *P. falciparum* represent PF11_0250 localisation (red), the DAPI stained nucleus (blue) and differential interference contrast (DIC) (ring- (R), trophozoite- (T) and late schizont (LS) stage).

Figure 7

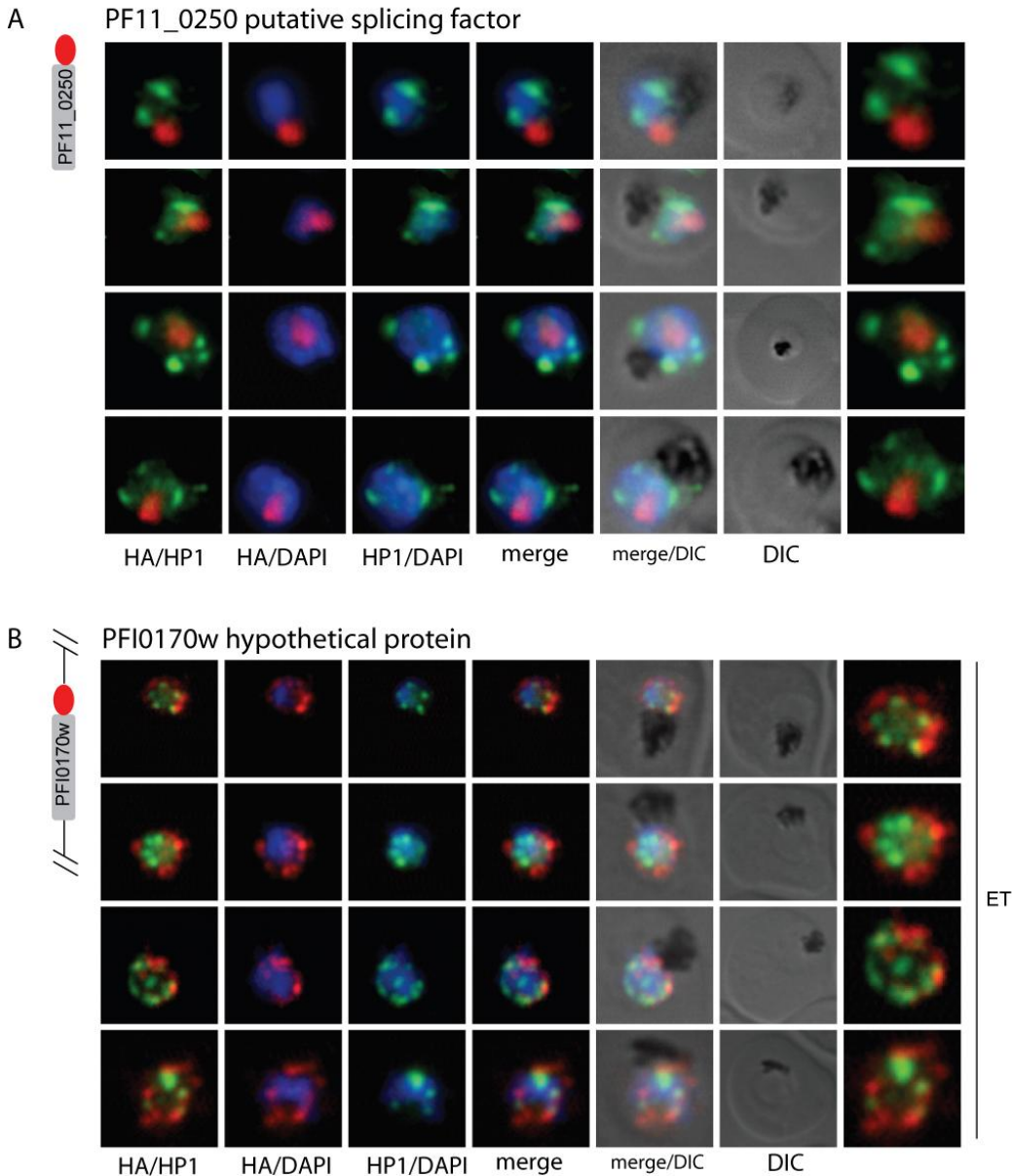


Figure 7. Spatial separation of the silenced chromosome ends, the nuclear pores and the nucleolus within the nuclear compartment at the trophozoite stage of *P. falciparum*.

Immunofluorescence analysis using the transgenic parasite lines 3D7/PFI0170w-3xHA (endogenously tagged; J. Wetzel, MSc thesis, 2011) or 3D7/PF11_0250-3xHA (episomally tagged) in combination with polyclonal antibodies against PfHP1. PFI0170w-3xHA-expressing parasites at the trophozoite stage (T) were methanol-fixed in contrast to PF11_0250-3xHA-expressing parasites, which were fixed with 4% para-formaldehyde/0.01% glutaraldehyde. The images of *P. falciparum* represent PfHP1 localisation (green), PF11_0250 or PFI0170w localisation (red), the DAPI stained nucleus (blue) and differential interference contrast (DIC).

Table 1.

Primer sequences used in this chapter (restriction sites are in bold)

pH-GFP 3' repl.

PFI0170w_F	TAG CCTGCAG TACAAAACAAAATGATGAATCCG
PFI0170w_R	TAC GCGGCCG CTACCTTTAAAATGTTTAAGTTCAG
PF14_0442_F	GAT CCTGCAG AATAACGTATTTCCACATGC
PF14_0442_R	CAT GCGGCCG CTATTTATCATATTTTGATTC

Chapter 5: General discussion and Outlook

Aspects underlying recent findings in *P. falciparum* transcriptional regulation

The major GTFs of eukaryotes including the complete set of 12 RNA polymerase II subunits were identified in *P. falciparum* [1][2][3]. However, their functional *in vivo* characterisation is sparse and only two *P. falciparum* *trans*-acting GTFs were analysed to date: the catalytic subunit of the RNA polymerase II [4] and the TATA binding protein TBP [5]. Interestingly, our organellar proteomic approach detected numerous proteins which are classified as GTFs of *P. falciparum*, namely 9 proteins out of the 12 polymerase core subunits and 15 proteins out of the 42 general coactivators involved in initiation and elongation. As a positive control set of known nuclear proteins was defined for the *in vivo* validation of the nuclear proteome, several transgenic parasite lines are available to analyse them on the functional level. In particular, the multiprotein bridging factor type 1 (PF11_0293), expected to act as a general transcriptional co-activator through connecting regulators and the TBP protein [6], would potentially give novel insights about the basic transcriptional machinery of *P. falciparum*.

The parasite *P. falciparum* shows an unusually tight program of mRNA transcription which follows a chronological course during the cell cycle and appears to be critical for survival across the complex life cycle [7][8]. Gene expression in *P. falciparum* follows the conventional model of monocistronic transcription, which was shown by using microarray as well as stable and transient transfection approaches [7][8][9][10]. In the last five years, several bioinformatic tools predicted an expanding repertoire of *cis*-regulatory sequences in the genome of *P. falciparum*, efforts that were hampered by the high AT-content of the parasite genome [11][12][13]. Nonetheless, the number of putative *cis*-acting elements identified remains little and their detailed functional determination as either transcriptional enhancers or activators is still outstanding [14]. In Apicomplexa, the identification of the novel ApiAP2 family of specific transcription factors containing 27 members in *P. falciparum* cleared up the mystery about the apparent lack of transcription-associated proteins that may regulate the genome via sequence-specific interactions [15][16]. ApiAP2 proteins are homologous to members of the Apetala2 transcription factors in plants and evolved into a lineage-specific version of the integrase

DNA-binding domain [15][16]. Several members of *P. berghei* and *P. falciparum* were analysed by using independent functional approaches, characterising them as ‘*bona fide*’ trans-acting factors with distinct DNA-binding properties [18][19][20][21]. The X-ray crystal structure of the PF14_0633 AP2 domain bound to DNA was recently determined and revealed that domain folding is highly conserved between plants and Apicomplexa [20][22]. ApiAP2 proteins of *P. falciparum* vary enormously in size, which raises the question whether ApiAP2 proteins carry additional functional domains that are undiscovered to date [16]. Interestingly, using the nuclear proteome as an input collection we identified a novel domain in ApiAP2 proteins, which we dubbed ACDC for ‘AP2-coincident domain at the C-terminus’. The co-occurrence of a second apicomplexan-specific domain in ApiAP2 proteins argues for an acquired structural and functional divergence of TAPs in apicomplexan parasites. Moreover, the group of sequence-specific transcription factors in *P. falciparum* comprehends a diverse set of zinc-coordinating DNA-binding proteins (total members 37) as well as a helix-turn helix protein family (total members 8), which includes seven members of the Myb family [1][2]. Thereof we were able to identify more the 20% of the total numbers using high accuracy mass spectrometry coupled with bioinformatics, namely 13 proteins out of the ApiAP2 family, 12 proteins out of the zinc finger family and 5 proteins out of the helix-turn helix family. Here, the transgenic parasite line of interest for further analysis expresses a putative fork head domain protein (PF13_0042) and a C2H2 zinc finger domain protein (PF10_0091), which is known to bind DNA in sequence-specific manner found in several transcription factors [17].

There is increasing evidence that mechanisms associated with gene expression evolved to an interfering control level of RNA decay in *P. falciparum* [23][24][25]. Interestingly, proteins of the PUF RNA-binding family have been identified in *P. falciparum*, which are known to target mRNAs and repress their translation [26][27]. In line with this, we detected a remarkable enrichment of putative RNA-binding proteins and factors in the nuclear proteome of *P. falciparum*, which are likely to play a role in mRNA processing. Summarized in numbers, 38 proteins were grouped according to PlasmoDB annotations as RNA-binding proteins and 13 additional proteins are likely involved in RNA processing or metabolism. Interestingly, two of them, annotated as a mRNA processing protein (PFI1600w) and a hypothetical protein (PF10_0279) share an interesting novel 56aa motif that is related to CSTF [28]. Currently, the role of this C-terminal domain is

unknown but it is interesting to speculate about an adapted function in *Plasmodium*-specific processes related to cleavage of 3' UTRs of gene transcripts. The hypothetical protein PF10_0279 exists as a tagged version in a transgenic cell line and attaches great importance in the context of RNA-decay to characterise it in more detail in *P. falciparum*. Furthermore, stage-specific clustering of the compared data set of our nuclear proteome and existing mRNA profiles revealed that ring stage parasites appear to synthesise some nuclear proteins *de novo*, whereas where carried over from schizonts of the previous life cycle. This observation, together with the phenomenon of mRNA decay and the preponderance of RNA-binding proteins, may indicate that *P. falciparum* life cycle progression relies heavily on mechanisms of post-transcriptional regulation. A recent study underlined this, as it extended previous mRNA abundance studies in *P. falciparum* showing that the average of mRNA half-life starting from the ring stage dramatically increases from 9.5 to 65 minutes across the IDC [23]. This genome-wide and stage-dependent phenomenon of post-transcriptional regulation leads to mRNA accumulation during the late schizont stage of *P. falciparum* [23]. Interestingly, two messenger ribonucleoproteins were identified in *P. berghei* named DOZI and Sm-like factor CITH, which stabilise distinct mRNAs and delay translational progression [29]. Gene deletions studies of core components of DOZI and CITH revealed that they translationally silence mRNAs essential for post-fertilization development of the parasite in the mosquito host [29].

As already mentioned, the demand for *in vivo* experimental validation of the nuclear core proteome provided several transgenic parasite lines expressing HA-tagged protein versions from episomal plasmids. Overall, 16 nuclear candidates were confirmed as such using IFA experiments and these are suitable for further analysis. To test whether the role of these proteins is essential in the regulatory context of gene expression, conventional or inducible knock-out parasite lines could be established. Micro-array analysis using the transgenic cell lines which over-express the proteins may identify up- or down-regulated target genes on a genome-wide scale, which in turn may allow to drawing conclusions about their differentiation as activators or repressors. Possible association of the proteins with specific DNA elements could be investigated using DNA-protein interaction techniques such as gel-shift assays, Southwestern analysis or genome-wide using chromatin immunoprecipitation experiments like CHIP-on-chip or CHIP-Seq. Furthermore, biochemical approaches like pull down or dot blot assays using

recombinantly expressed proteins have the potential to identify the recognition sites of histone modifications of chromatin-associated proteins.

Conserved ‘reader’ proteins of epigenetic histone marks in *P. falciparum*

Epigenetic regulation fulfils a main function in gene expression in *P. falciparum*, as eukaryotic chromatin structure, histone modifications and complementary chromatin-associated proteins are present in the parasite’s proteome [30]. The *P. falciparum* genome encodes the four core histones H2A, H2B, H3 and H4 and the variant forms H2A.Z, H2Bv, H3.3 and CenH3, which share core structures but differ in specific functions [30]. In eukaryotes, the most frequently observed histone modifications include acetylation, methylation, phosphorylation, ubiquitination, ADP-ribosylation and sumoylation, which interact either in a synergistic or an antagonistic manner [31][32]. More than 44 post-translational modifications affecting histones have been reported in *P. falciparum* by using different mass spectrometry detection systems [33]. Interestingly, more activation marks compared to silencing marks have been found compared to other unicellular eukaryotes and their combinations appear in unusual dense clusters in the epigenome of *P. falciparum* [34].

Recent studies using ChIP-on-chip allowed a big step forward in the understanding of the epigenetic processes of *P. falciparum*, as they globally correlated histone modifications to gene activity. They show a major consensus of the histone code between *P. falciparum* and other eukaryotes with only slight deviations like novel modifications on canonical or variant histones [33][35][36][37]. Interestingly, Salcedo-Amaya *et al.* uncovered significant stage-dependent specificities of the histone H3 epigenome across the IDC of *P. falciparum* [36]. ChIP-on-chip profiling of asynchronous parasites revealed an apparent dominance of euchromatic marks and heterchromatic patterns restricted to variant surface antigen gene families [36]. In detail, synchronisation of the parasite cultures showed that at the ring stage, H3K4me3 and H3K9ac were equally enriched at 5’ ends of active as well as silent genes compared to the schizont stage where they are exclusively restricted to active genes [36]. A second study using ChIP-Seq confirmed the genome-wide and dynamic localization of both histone marks H3K4me3 and H3K9ac and improved their features as they showed that H3K9ac correlates well with the transcriptional status of the marked promoters and that

H3K4me3 appears to have stage-specific regulation [37]. N-terminal histone modifications can either directly alter the chromatin structure by affecting the DNA-protein interaction or act as recognition sites for effector proteins, which then recruit other proteins [31]. The specific recognition is mediated by 'reader' proteins which carry evolutionary highly conserved structures like bromo-, chromo-, tudor- domains and PHD fingers [38]. The bromo-domain is exclusively known to be responsible for binding to acetyl-lysine residues, whereas the remaining motifs are recognising methylation marks [39][40]. Interestingly, several studies have focused to evolutionary aspects of apicomplexan proteins that create and respond to the epigenetic code [41][3]. They suggested that this protein group has proliferated throughout parasitic adaptation with regards to its domain architecture, as novel motif versions emerged or other combinations were acquired [41].

Interestingly, we were able to newly assign putative unique nuclear functions to structural regions of distinct proteins, which was only possible based on the information of the *P. falciparum* nuclear proteome. Computational methods identified a protein which has sequence similarities to the ELM2 domain and carries additionally a PHD-finger (PF11_0429). The ELM2 domain is a conserved motif found in many eukaryotic proteins and mediates transcriptional repression through the recruitment of histone deacetylases or methyl-transferases [42][43]. Furthermore, a couple of proteins were identified which show homology to the MYND zinc finger domain, a highly conserved cysteine-rich structure mediating protein-protein interactions (PF14_0310, PF10_0150, PFF0350w and PFF0105w) [44]. MYND domains are often found in combination with SET domains and these proteins have functional implications in transcriptional repression [44]. Interestingly, we found this combination of domains for one of our identified proteins annotated as histone-lysine N-methyl-transferase (PF13_0293). Thus, our findings, relying on improved and extended bioinformatic BLAST searches, confirmed the evolutionary synthesis of chromatin-related proteins in Apicomplexa and underline their structural variability.

Returning to bromo-domain proteins, the genome of *P. falciparum* encodes eight proteins which carry an annotated single bromo-motif and which may function as putative transcriptional co-activators [2]. Two proteins are histone modifying proteins (HAT PfGCN5, HKMT PfSET1), one protein is a Swi2/Snf2 ATPase and five proteins are

thought to be involved in chromatin structuring as they do not contain other discernable catalytic domains [2]. Remarkably, one protein out of the five is not annotated as a bromo-domain protein in PlasmoDB but can be handled as such because it shows sufficient sequence similarities to the motif (PF11_0254). Unfortunately, most of the chromatin structuring proteins of *P. falciparum* have not been characterised to date, despite their potential role as long term chromatin effectors. An exception is PfHP1, which mediates *var* gene silencing and recognises methylated lysines 9 of histone 3 via its chromo-domain [45][46]. To begin to understand how other chromatin-binding proteins in *P. falciparum* execute their function we concentrated our attention on three uncharacterised bromo-domain proteins that co-eluted in co-immunoprecipitation experiments. The formation of a putative multimeric bromo-domain protein complex was supported by subsequent co-immunoprecipitation experiments using nuclear extracts from two double transgenic lines co-expressing tagged versions of PF10_0328 and PF11_0254 or PF10_0328 and PFL0635c. For the first time, there is evidence that three bromo-domain proteins may have a combined or overlapping global function in regulating gene transcription across the asexual life cycle of *P. falciparum*. Moreover, we tried to verify the interaction results shown by co-IP using the Y2H yeast system to test whether the proteins bind directly or not. As no colonies were growing on selective plates, the three bromo-domain proteins may not interact directly with each other. It is conceivable that the bromo-domain proteins are connected via bridging factors. Another possibility would be that we extracted entire nucleosomes or parts thereof and thus co-purified proteins recognising other acetylated histone modifications unrelated to the bait bromo-domain protein PF10_0328. In addition, the histone H2A (PFF0860c), a putative ApiAP2 transcription factor (PFF0670w) and a hypothetical protein (PF14_0379), each detected by one or two unique peptides, were co-purified along with PF10_0328, supporting the predicted function of this bromo-domain protein in transcriptional regulation at the chromatin level. Interestingly, we identified the protein PF14_0379 as well in nuclear proteome study exclusively in the SDS fraction by mass spectrometry what argues for a specific association to the *P. falciparum* proteinous complex.

Further characterisation of the *P. falciparum* bromo-domain proteins is of course required to determine their actual role in gene expression. Our collaborators Zbynek Bozdech (Nanyang Technological University, Singapore) and Mike Duffy (University of Melbourne, Australia) have recently determined that the recombinantly expressed

bromo-domain of PF10_0328 recognises histones 3 and 4 as well as the variant H2AZ *in vitro*. Current efforts focus on the identification of the actual acetylated lysine residues that are bound by the three bromo-domains using pulldown assays employing biotinylated modified histone peptides. Furthermore, they observed that PF10_0328 binds to promoters of genes associated with active histone marks by using chromatin-IP. To obtain a genome-wide view on the function of these interesting regulators transcriptional profiling and ChIP-on-chip analysis using our established transgenic parasite lines are in progress. The anticipated results will hopefully reveal sets of genes regulated by these factors *in vivo*. For instance, as bromo-domains exclusively recognise the active histone mark of acetylation, transcriptional up-regulation of a subset of genes is expected in our overexpressing transgenic cell lines compared to the wild type line 3D7. As several attempts to establish knock-out cell lines of the three bromo-domain candidates failed indicating that they are essential for *P. falciparum* survival, a recently adapted destabilization domain system based on the ddFKBP fusion technique [47] could be applied to downregulate their expression in order to create 'loss of function' mutants.

Nuclear landmark structures in *P. falciparum*

The eukaryotic nucleus encompasses dynamic compartments and structures, which rely on an extensive network of interactions to establish the maintenance, replication, expression and translation of the genetic information. The nuclear genome is folded and packaged into higher-order structures organised as either euchromatin or heterochromatin, which participate the genome into active gene transcription or gene silencing, respectively. Spatial gene positioning has been shown to play a major role in gene regulation [48]. Indeed, silent gene loci move to regions which are permissive for transcription, permitting access to the transcriptional machinery [48]. As another example of critical nuclear sub-structures, the required exchange of macromolecules between the nucleus and its neighbouring compartment, the cytoplasm is accomplished through highly specialised apertures named nuclear pore complexes. The nucleolus is the most prominent compartment of the nucleus and has a well-established role in rDNA expression and the assembly of ribosomal subunits that are in constant flow between the nucleoplasm and other nuclear bodies.

The technology of microscopy can advantageously provide insights into the dynamic processes involved in nuclear organisation and its implication on gene regulation. Despite recent publications about the role of heterochromatin in gene expression [49][50][45], the dynamics of nuclear pores in relation to chromatin [50] and the behaviour of nucleolar assembly during parasite development across the IDC [51], little is known about the spatial arrangements of these and other sub-compartments within the *P. falciparum* nucleus. With the help of the nuclear core proteome, we were able to identify novel nuclear proteins residing in nuclear pores and the nucleolus, respectively, and to visualise them as structural representatives using fluorescence microscopy.

In *P. falciparum*, electron-microscopy of the nuclear ultrastructure revealed that the periphery consists mainly of electron-dense, heterochromatin-like material, but also contains electron-sparse gaps consistent with euchromatin [49][50]. The inside of the nucleus of *P. falciparum* is suggested to be generally more transcriptionally active, as histone modifications marking euchromatin like H3K9ac were found to localise in the central area stained by DAPI [52]. Recently, the emerging technique of serial surface view imaging by FIB revealed that loose euchromatin can also be found at the nuclear periphery, particularly in association with nuclear pores of *P. falciparum* [50]. Furthermore, this technique allowed a 3D model of the parasite nucleus to be constructed, which demonstrated dynamic changes in chromatin packaging and the local distribution of nuclear pores across the developmental stages of the IDC. These observations revealed that nuclear pores were distributed uniformly around the NE in trophozoites. Their number increased during the trophozoite maturation in contrast with the schizont, merozoite and early ring stages, where the localisation of nuclear pores was clustered and the number of nuclear pores per nucleus was decreased [50].

Var gene regulation serves as a good example for the dual role of the nuclear periphery in gene expression, because activation is assumed to involve spatial repositioning of the active locus into a distinct permissive zone at the nuclear periphery of *P. falciparum* [53][54][55][49][56][57]. FISH experiments revealed that silenced *var* genes, which occupy parts of the chromosomal ends, localise in two to seven foci at the nuclear periphery of *P. falciparum* [49][58]. Chromosome ends are the major heterochromatic or silent regions of the genome and heterochromatin was shown to spread from telomeres over 20-30 kb into the subtelomeric domains, which encode *var* and other suspected

virulence gene families [45][36][53][59][56]. Interestingly, the spatial transition of the *var* locus was proposed to require actin, as affinity purification identified it as a component of the protein complex binding the *var* intron [60] that is involved in the regulatory process [61]. Furthermore, PfSET10, a histone 3 lysine 4 methyl-transferase, was identified to co-localise with the active *var* gene expression site in *P. falciparum*, and a likely involvement of this enzyme in maintaining the permissive chromatin environment of the transcribed *var* gene was proposed [55].

To date, however, nothing is known about the perinuclear localisation of silenced *var* gene clusters or the active *var* locus in relation to the positioning of nuclear pores in *P. falciparum*. Here, we demonstrated for the first time that chromosome end clusters bearing silenced *var* genes are not associated with the 'open' domains defined by nuclear pores at the nuclear periphery. Furthermore, our observations support previous reports that chromosome end clusters of *P. falciparum* appear attached to the nuclear envelope, as some foci interfered with the circular nuclear pore structure in *P. falciparum*. The next step would be to analyse whether the perinuclear active *var* locus prefers the close proximity to the nuclear pore regions. To test this, our established transgenic parasite line 3D7/PFI0170w-3xHA can be combined with antibodies targeting the methyl-transferase PfSET10, which marks the active *var* gene transcription site, in experiments using confocal microscopy. Indeed, as shown in the model organism *S. cerevisiae*, nuclear pores create environments that stimulate gene transcription and spatial relocation of gene loci leads to transcriptional activation [62][63] supposed to facilitate mRNA export [64].

In accordance with the above mentioned results published by Weiner and Dahan-Pasternak [50], we observed that the positioning of the nuclear pores was highly dynamic during *P. falciparum* progression across the IDC. At the trophozoite stage, NPCs were equally distributed surrounding the NE, which is in contrast to the ring and schizont stages where nuclear pores appeared to cluster. The equal distribution of the NPCs at the trophozoite stage may indicate that *P. falciparum* increases nucleocytoplasmic transport during its major growth period as there is evidence that numbers and distribution of NPCs are associated with nuclear size and activity in other species [65]. The NPC clustering starting at the schizont stage may then mean decreased

activity of nucleo-cytoplasmic transport or, alternatively, may be related to unknown requirements related to the division of pores among the daughter nuclei.

Now, it will be interesting to scrutinise at which timepoint nuclear pore complexes start to assemble and disassemble and also why this phenomenon occurs in *P. falciparum*, and what the functional relevance of this dynamic behaviour is. Such studies may also shed light on different aspects of 'closed' mitosis, for example how the nuclear envelope, which is known to remain intact in *P. falciparum* [66][67], is divided into two daughter nuclei. In this context, confocal time-lapse microscopy of the established transgenic cell line 3D7/HP1-CherryFP may be used to explain the behaviour of chromosome ends during serial mitoses, including the likely disruption of the perinuclear chromosome end clusters in *P. falciparum*. With regard to nuclear pores, a co-IP approach combined with LC-MS/MS analysis using the 3D7/PFI0170w-3xHA line may be a promising approach to identify further members of the large proteinous nuclear pore complex, which is generally composed of around 30 different members [68][69], but almost entirely unknown in *P. falciparum*.

In *P. falciparum*, the nucleolus was recently described by IFA as a 'hat-like' structure at most stages across the IDC [51][70][71], which is in contrast with our observations characterising this sub-compartment as a smaller and more restricted area. This restricted area perfectly matched to a DAPI-free intranuclear cavity indicating a relatively low amount of DNA, which is consistent with the nucleolar compartment that is mainly composed of ribonucleoproteins [72]. Interestingly, the size of the parasite nucleolus increased and appeared to reach up to one third of the total nuclear volume at the pre-replicative stage of *P. falciparum*. A putative explanation for this considerable change in size may be that *P. falciparum* requires increased ribosome production ultimately leading to higher activity of the translational machinery because trophozoites are in the phase of major growth and digest large amount of hemoglobin.

P. falciparum contains single ribosomal DNA units (18S-5.8S-28S), which are positioned in the subtelomeric regions on five different chromosomes (chr 1,5,7,11,13) [51]. This arrangement is in contrast to the tandemly repeated arrays of rDNA genes on the same chromosome as found in many other eukaryotes [73][74]. As the physical separation of the subtelomeric located rDNA genes to the chromosome ends is relative short as it

varies between approximately 40-940 kb, we were interested in the spatial arrangement of the nucleolus in reference to the perinuclear clusters of the chromosome ends within the nucleus of *P. falciparum*. For the first time, we showed by using IFA that the nucleolus is not associated with chromosomal end clusters at the early trophozoite stage. It is therefore tempting to speculate that the distance of the rDNA loci to the heterochromatic chromosome ends is sufficient to separate both structural components. The next step will be to analyse whether the nucleolus actually disassembles during schizogony. This has been postulated based on the observation that the involved rDNA gene cluster disperses and RNA polymerase I activity is no longer detectable during this stage in *P. falciparum* [51]. Another step will be to analyse whether the nucleolus of *P. falciparum* reassembles at the end of the mitotic cycles from components that were preserved from the pre-mitotic phase of the cell cycle, as described for other organisms [75], or if nucleolar re-assembly occurs *de novo*. A possibility would be to add the autofluorescence marker CherryFP to the identified protein PF11_0250 and to analyse the transgenic parasite line by time-lapse microscopy combined with the technique of fluorescence recovery after photobleaching or pulse labelling with heavy isotopes [76][77][78].

Furthermore, the combination of high-throughput LC-MS/MS analysis and advanced bioinformatics would allow the detection of the nucleolar proteome at different developmental stages of the IDC, permitting us to study this sub-compartment in much greater detail and to answer important questions about nucleolar biology in *P. falciparum*. The high density and structural stability of the nucleolus in other eukaryotes [79] provides promising conditions for the establishment of a proper nucleolus isolation procedure for *P. falciparum*. A promising alternative to organellar extraction would be to use the nucleolar transgenic parasite lines and/or the antibodies directed against fibrillarin to co-purify compartment-specific protein complexes out of nuclear extracts of *P. falciparum* by using co-IP experiments coupled to LC-MS/MS. Proteomic profiling of nucleolus-enriched fractions may identify novel telomere-associated proteins, as it is hypothesised that the nucleolus might serve as a reservoir for such components [70]. Further *P. falciparum* proteins may be found that are unrelated to ribosome assembly, which would point towards novel nucleolar functions such as nuclear export, sequestration of regulatory molecules, modification of small RNAs, cell cycle control and telomerase maturation [80][81]. Nucleolar proteomics of different life cycle stages may

also provide information about changes of the protein content over time in response to parasite growth, as nucleolar components are proposed to be highly mobile [82]. Finally, this may lead to conclusions about how *P. falciparum* regulates intra-nuclear transport, which is suggested to occur by passive targeting without specific localisation signals [83].

The dataset of the nuclear proteome functions as an important platform and ‘toolbox’ to study nuclear biology in *Plasmodium* in great detail. Several findings underlined research trends in transcriptional and post-transcriptional regulation, like the functional acceptance of specific transcription factors or RNA decay in *P. falciparum*. The *in vivo* validation of the nuclear proteome offers several established transgenic parasite lines expressing promising proteins awaiting functional analyses. Indeed, high-throughput technology applied to *P. falciparum* research revealed important information on different levels and on many genes and proteins, but they usually remain descriptive as experimental functional approaches often require reverse genetics that are difficult and time consuming to perform in *P. falciparum*. While this limitation is also true for our experimentally determined nuclear proteome, our data will facilitate the targeted investigation of a large number of previously unknown nuclear proteins. The initial analysis of three chromatin structuring proteins containing bromo-domains and the visualisation of nuclear sub-compartments illustrate the power of such downstream analyses. The bromo-domain proteins are of particular interest not only to understand gene regulation in greater detail, but also because the binding of bromo-domains to acetylated lysine residues can be competitively blocked by small molecules at low nanomolar concentrations. Their subsequent functional determination may open novel targets for intervention strategies never discovered before in malaria research. My microscopy-based results visualised for the first time nuclear pores and the nucleolus related to chromosome end clustering, which revealed promising insights into the nuclear architecture of *P. falciparum* and provide a framework for future efforts to study mitosis, nuclear-cytoplasmic transport and other essential processes of nuclear biology in great detail.

References

1. Callebaut I, Prat K, Meurice E, Mornon J-P, Tomavo S (2005) Prediction of the general transcription factors associated with RNA polymerase II in *Plasmodium falciparum*: conserved features and differences relative to other eukaryotes. *BMC Genomics* 6: 100. doi:10.1186/1471-2164-6-100.
2. Bischoff E, Vaquero C (2010) In silico and biological survey of transcription-associated proteins implicated in the transcriptional machinery during the erythrocytic development of *Plasmodium falciparum*. *BMC Genomics* 11: 34. doi:10.1186/1471-2164-11-34.
3. Aravind L, Iyer LM, Wellems TE, Miller LH (2003) *Plasmodium* biology: genomic gleanings. *Cell* 115: 771–785.
4. Li WB, Bzik DJ, Gu HM, Tanaka M, Fox BA, et al. (1989) An enlarged largest subunit of *Plasmodium falciparum* RNA polymerase II defines conserved and variable RNA polymerase domains. *Nucleic Acids Res* 17: 9621–9636.
5. Ruvalcaba-Salazar OK, del Carmen Ramírez-Estudillo M, Montiel-Condado D, Recillas-Targa F, Vargas M, et al. (2005) Recombinant and native *Plasmodium falciparum* TATA-binding-protein binds to a specific TATA box element in promoter regions. *Mol Biochem Parasitol* 140: 183–196. doi:10.1016/j.molbiopara.2005.01.002.
6. Takemaru K-I, Li F-Q, Ueda H, Hirose S (1997) Multiprotein Bridging Factor 1 (MBF1) Is an Evolutionarily Conserved Transcriptional Coactivator That Connects a Regulatory Factor and TATA Element-Binding Protein. *PNAS* 94: 7251–7256.
7. Bozdech Z, Llinás M, Pulliam BL, Wong ED, Zhu J, et al. (2003) The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS Biol* 1: E5. doi:10.1371/journal.pbio.0000005.
8. Le Roch KG, Zhou Y, Blair PL, Grainger M, Moch JK, et al. (2003) Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science* 301: 1503–1508. doi:10.1126/science.1087025.
9. Wickham ME, Thompson JK, Cowman AF (2003) Characterisation of the merozoite surface protein-2 promoter using stable and transient transfection in *Plasmodium falciparum*. *Mol Biochem Parasitol* 129: 147–156.
10. de Koning-Ward TF, Sperança MA, Waters AP, Janse CJ (1999) Analysis of stage specificity of promoters in *Plasmodium berghei* using luciferase as a reporter. *Mol Biochem Parasitol* 100: 141–146.
11. Young JA, Johnson JR, Benner C, Yan SF, Chen K, et al. (2008) In silico discovery of transcription regulatory elements in *Plasmodium falciparum*. *BMC Genomics* 9: 70. doi:10.1186/1471-2164-9-70.
12. Gunasekera AM, Myrick A, Militello KT, Sims JS, Dong CK, et al. (2007) Regulatory motifs uncovered among gene expression clusters in *Plasmodium falciparum*. *Mol Biochem Parasitol* 153: 19–30. doi:10.1016/j.molbiopara.2007.01.011.
13. Wu J, Sieglaff DH, Gervin J, Xie XS (2008) Discovering regulatory motifs in the *Plasmodium* genome using comparative genomics. *Bioinformatics* 24: 1843–1849. doi:10.1093/bioinformatics/btn348.
14. Horrocks P, Wong E, Russell K, Emes RD (2009) Control of gene expression in *Plasmodium falciparum* - ten years on. *Mol Biochem Parasitol* 164: 9–25. doi:10.1016/j.molbiopara.2008.11.010.
15. Balaji S, Babu MM, Iyer LM, Aravind L (2005) Discovery of the principal specific transcription factors of Apicomplexa and their implication for the evolution of the AP2-integrase DNA binding domains. *Nucleic Acids Res* 33: 3994–4006. doi:10.1093/nar/gki709.
16. Painter HJ, Campbell TL, Llinás M (2011) The Apicomplexan AP2 family: Integral factors regulating *Plasmodium* development. *Mol Biochem Parasitol* 176: 1–7. doi:10.1016/j.molbiopara.2010.11.014.
17. Hacker U (1992) Developmentally Regulated *Drosophila* Gene Family Encoding the Fork Head Domain. *Proceedings of the National Academy of Sciences* 89: 8754–8758. doi:10.1073/pnas.89.18.8754.
18. Yuda M, Iwanaga S, Shigenobu S, Kato T, Kaneko I (2010) Transcription factor AP2-Sp and its target genes in malarial sporozoites. *Mol Microbiol* 75: 854–863. doi:10.1111/j.1365-2958.2009.07005.x.
19. Yuda M, Iwanaga S, Shigenobu S, Mair GR, Janse CJ, et al. (2009) Identification of a transcription factor in the mosquito-invasive stage of malaria parasites. *Mol Microbiol* 71: 1402–1414. doi:10.1111/j.1365-2958.2009.06609.x.
20. Lindner SE, De Silva EK, Keck JL, Llinás M (2010) Structural determinants of DNA binding by a *P. falciparum* ApiAP2 transcriptional regulator. *J Mol Biol* 395: 558–567. doi:10.1016/j.jmb.2009.11.004.
21. Flueck C, Bartfai R, Niederwieser I, Witmer K, Alako BTF, et al. (2010) A major role for the *Plasmodium falciparum* ApiAP2 protein PfSIP2 in chromosome end biology. *PLoS Pathog* 6: e1000784. doi:10.1371/journal.ppat.1000784.
22. De Silva EK, Gehrke AR, Olszewski K, León I, Chahal JS, et al. (2008) Specific DNA-binding by apicomplexan AP2 transcription factors. *Proc Natl Acad Sci USA* 105: 8393–8398. doi:10.1073/pnas.0801993105.
23. Shock JL, Fischer KF, DeRisi JL (2007) Whole-genome analysis of mRNA decay in *Plasmodium falciparum* reveals a global lengthening of mRNA half-life during the intra-erythrocytic development cycle. *Genome Biol* 8: R134. doi:10.1186/gb-2007-8-7-r134.

24. Sims JS, Militello KT, Sims PA, Patel VP, Kasper JM, et al. (2009) Patterns of gene-specific and total transcriptional activity during the *Plasmodium falciparum* intraerythrocytic developmental cycle. *Eukaryotic Cell* 8: 327–338. doi:10.1128/EC.00340-08.
25. Le Roch KG, Johnson JR, Florens L, Zhou Y, Santrosyan A, et al. (2004) Global analysis of transcript and protein levels across the *Plasmodium falciparum* life cycle. *Genome Res* 14: 2308–2318. doi:10.1101/gr.2523904.
26. Fan Q, Li J, Kariuki M, Cui L (2004) Characterization of PfPuf2, member of the Puf family RNA-binding proteins from the malaria parasite *Plasmodium falciparum*. *DNA Cell Biol* 23: 753–760. doi:10.1089/1044549042531413.
27. Fan Q, Miao J, Cui L, Cui L (2009) Characterization of PRMT1 from *Plasmodium falciparum*. *Biochem J* 421: 107–118. doi:10.1042/BJ20090185.
28. Takagaki Y, Manley JL, MacDonald CC, Wilusz J, Shenk T (1990) A multisubunit factor, CstF, is required for polyadenylation of mammalian pre-mRNAs. *Genes Dev* 4: 2112–2120.
29. Mair GR, Lasonder E, Garver LS, Franke-Fayard BMD, Carret CK, et al. (2010) Universal features of post-transcriptional gene regulation are critical for *Plasmodium* zygote development. *PLoS Pathog* 6: e1000767. doi:10.1371/journal.ppat.1000767.
30. Miao J, Fan Q, Cui L, Li J, Li J, et al. (2006) The malaria parasite *Plasmodium falciparum* histones: organization, expression, and acetylation. *Gene* 369: 53–65. doi:10.1016/j.gene.2005.10.022.
31. Kouzarides T (2007) Chromatin modifications and their function. *Cell* 128: 693–705. doi:10.1016/j.cell.2007.02.005.
32. Jenuwein T, Allis CD (2001) Translating the Histone Code. *Science* 293: 1074–1080. doi:10.1126/science.1063127.
33. Trelle MB, Salcedo-Amaya AM, Cohen AM, Stunnenberg HG, Jensen ON (2009) Global histone analysis by mass spectrometry reveals a high content of acetylated lysine residues in the malaria parasite *Plasmodium falciparum*. *J Proteome Res* 8: 3439–3450. doi:10.1021/pr9000898.
34. Cui L, Miao J (2010) Chromatin-mediated epigenetic regulation in the malaria parasite *Plasmodium falciparum*. *Eukaryotic Cell* 9: 1138–1149. doi:10.1128/EC.00036-10.
35. Cui L, Miao J, Furuya T, Li X, Su X, et al. (2007) PfGCN5-mediated histone H3 acetylation plays a key role in gene expression in *Plasmodium falciparum*. *Eukaryotic Cell* 6: 1219–1227. doi:10.1128/EC.00062-07.
36. Salcedo-Amaya AM, van Driel MA, Alako BT, Trelle MB, van den Elzen AMG, et al. (2009) Dynamic histone H3 epigenome marking during the intraerythrocytic cycle of *Plasmodium falciparum*. *Proc Natl Acad Sci USA* 106: 9655–9660. doi:10.1073/pnas.0902515106.
37. Bártfai R, Hoeijmakers WAM, Salcedo-Amaya AM, Smits AH, Janssen-Megens E, et al. (2010) H2A.Z Demarcates Intergenic Regions of the *Plasmodium falciparum* Epigenome That Are Dynamically Marked by H3K9ac and H3K4me3. *PLoS Pathog* 6: e1001223. doi:10.1371/journal.ppat.1001223.
38. Basu MK, Carmel L, Rogozin IB, Koonin EV (2008) Evolution of protein domain promiscuity in eukaryotes. *Genome Res* 18: 449–461. doi:10.1101/gr.6943508.
39. Zeng L, Zhou M-M (2002) Bromodomain: an acetyl-lysine binding domain. *FEBS Letters* 513: 124–128. doi:10.1016/S0014-5793(01)03309-9.
40. Jacobson RH, Ladurner AG, King DS, Tjian R (2000) Structure and function of a human TAFII250 double bromodomain module. *Science* 288: 1422–1425.
41. Iyer LM, Anantharaman V, Wolf MY, Aravind L (2008) Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes. *Int J Parasitol* 38: 1–31. doi:10.1016/j.ijpara.2007.07.018.
42. Wang L, Rajan H, Pitman JL, McKeown M, Tsai C-C (2006) Histone deacetylase-associating Atrophin proteins are nuclear receptor corepressors. *Genes Dev* 20: 525–530. doi:10.1101/gad.1393506.
43. Ding Z, Gillespie LL, Paterno GD (2003) Human MI-ER1 alpha and beta function as transcriptional repressors by recruitment of histone deacetylase 1 to their conserved ELM2 domain. *Mol Cell Biol* 23: 250–258.
44. Lutterbach B, Sun D, Schuetz J, Hiebert SW (1998) The MYND motif is required for repression of basal transcription from the multidrug resistance 1 promoter by the t(8;21) fusion protein. *Mol Cell Biol* 18: 3604–3611.
45. Flueck C, Bartfai R, Volz J, Niederwieser I, Salcedo-Amaya AM, et al. (2009) *Plasmodium falciparum* heterochromatin protein 1 marks genomic loci linked to phenotypic variation of exported virulence factors. *PLoS Pathog* 5: e1000569. doi:10.1371/journal.ppat.1000569.
46. Pérez-Toledo K, Rojas-Meza AP, Mancio-Silva L, Hernández-Cuevas NA, Delgado DM, et al. (2009) *Plasmodium falciparum* heterochromatin protein 1 binds to tri-methylated histone 3 lysine 9 and is linked to mutually exclusive expression of var genes. *Nucleic Acids Res* 37: 2596–2606. doi:10.1093/nar/gkp115.
47. Armstrong CM, Goldberg DE (2007) An FKBP destabilization domain modulates protein levels in *Plasmodium falciparum*. *Nat Methods* 4: 1007–1009. doi:10.1038/nmeth1132.
48. Sexton T, Schober H, Fraser P, Gasser SM (2007) Gene regulation through nuclear organization. *Nat Struct Mol Biol* 14: 1049–1055. doi:10.1038/nsmb1324.

49. Ralph SA, Scheidig-Benatar C, Scherf A (2005) Antigenic variation in *Plasmodium falciparum* is associated with movement of var loci between subnuclear locations. *Proc Natl Acad Sci USA* 102: 5414–5419. doi:10.1073/pnas.0408883102.
50. Weiner A, Dahan-Pasternak N, Shimoni E, Shinder V, von Huth P, et al. (2011) 3D nuclear architecture reveals coupled cell cycle dynamics of chromatin and nuclear pores in the malaria parasite *Plasmodium falciparum*. *Cell Microbiol* 13: 967–977. doi:10.1111/j.1462-5822.2011.01592.x.
51. Mancio-Silva L, Zhang Q, Scheidig-Benatar C, Scherf A (2010) Clustering of dispersed ribosomal DNA and its role in gene regulation and chromosome-end associations in malaria parasites. *Proc Natl Acad Sci USA* 107: 15117–15122. doi:10.1073/pnas.1001045107.
52. Volz J, Carvalho TG, Ralph SA, Gilson P, Thompson J, et al. (2010) Potential epigenetic regulatory proteins localise to distinct nuclear sub-compartments in *Plasmodium falciparum*. *Int J Parasitol* 40: 109–121. doi:10.1016/j.ijpara.2009.09.002.
53. Duraisingh MT, Voss TS, Marty AJ, Duffy MF, Good RT, et al. (2005) Heterochromatin silencing and locus repositioning linked to regulation of virulence genes in *Plasmodium falciparum*. *Cell* 121: 13–24. doi:10.1016/j.cell.2005.01.036.
54. Voss TS, Healer J, Marty AJ, Duffy MF, Thompson JK, et al. (2006) A var gene promoter controls allelic exclusion of virulence genes in *Plasmodium falciparum* malaria. *Nature* 439: 1004–1008. doi:10.1038/nature04407.
55. Volz JC, Bártfai R, Petter M, Langer C, Josling GA, et al. (2012) PfSET10, a *Plasmodium falciparum* methyltransferase, maintains the active var gene in a poised state during parasite division. *Cell Host Microbe* 11: 7–18. doi:10.1016/j.chom.2011.11.011.
56. Lopez-Rubio J-J, Mancio-Silva L, Scherf A (2009) Genome-wide analysis of heterochromatin associates clonally variant gene regulation with perinuclear repressive centers in malaria parasites. *Cell Host Microbe* 5: 179–190. doi:10.1016/j.chom.2008.12.012.
57. Dzikowski R, Li F, Amulic B, Eisberg A, Frank M, et al. (2007) Mechanisms underlying mutually exclusive expression of virulence genes by malaria parasites. *EMBO Rep* 8: 959–965. doi:10.1038/sj.embor.7401063.
58. Freitas-Junior LH, Hernandez-Rivas R, Ralph SA, Montiel-Condado D, Ruvalcaba-Salazar OK, et al. (2005) Telomeric heterochromatin propagation and histone acetylation control mutually exclusive expression of antigenic variation genes in malaria parasites. *Cell* 121: 25–36. doi:10.1016/j.cell.2005.01.037.
59. Tonkin CJ, Carret CK, Duraisingh MT, Voss TS, Ralph SA, et al. (2009) Sir2 paralogues cooperate to regulate virulence genes and antigenic variation in *Plasmodium falciparum*. *PLoS Biol* 7: e84. doi:10.1371/journal.pbio.1000084.
60. Zhang Q, Huang Y, Zhang Y, Fang X, Claes A, et al. (2011) A Critical Role of Perinuclear Filamentous Actin in Spatial Repositioning and Mutually Exclusive Expression of Virulence Genes in Malaria Parasites. *Cell Host & Microbe* 10: 451–463. doi:10.1016/j.chom.2011.09.013.
61. Deitsch KW (2001) Gene silencing and antigenic variation in malaria parasites. *ScientificWorldJournal* 1: 650–652. doi:10.1100/tsw.2001.369.
62. Taddei A, Houwe GV, Hediger F, Kalck V, Cubizolles F, et al. (2006) Nuclear pore association confers optimal expression levels for an inducible yeast gene. *Nature* 441: 774–778. doi:10.1038/nature04845.
63. Casolari JM, Brown CR, Komili S, West J, Hieronymus H, et al. (2004) Genome-wide localization of the nuclear transport machinery couples transcriptional status and nuclear organization. *Cell* 117: 427–439.
64. Blobel G (1985) Gene gating: a hypothesis. *Proc Natl Acad Sci USA* 82: 8527–8529.
65. Lim RYH, Ullman KS, Fahrenkrog B (2008) Biology and biophysics of the nuclear pore complex and its components. *Int Rev Cell Mol Biol* 267: 299–342. doi:10.1016/S1937-6448(08)00632-1.
66. Striepen B, Jordan CN, Reiff S, van Dooren GG (2007) Building the perfect parasite: cell division in apicomplexa. *PLoS Pathog* 3: e78. doi:10.1371/journal.ppat.0030078.
67. Gerald N, Mahajan B, Kumar S (2011) Mitosis in the Human Malaria Parasite *Plasmodium Falciparum*. *Eukaryotic Cell* 10: 474–482. doi:10.1128/EC.00314-10.
68. Cronshaw JM, Krutchinsky AN, Zhang W, Chait BT, Matunis MJ (2002) Proteomic analysis of the mammalian nuclear pore complex. *J Cell Biol* 158: 915–927. doi:10.1083/jcb.200206106.
69. Rout MP, Aitchison JD, Suprpto A, Hjertaas K, Zhao Y, et al. (2000) The yeast nuclear pore complex: composition, architecture, and transport mechanism. *J Cell Biol* 148: 635–651.
70. Mancio-Silva L, Rojas-Meza AP, Vargas M, Scherf A, Hernandez-Rivas R (2008) Differential association of Orc1 and Sir2 proteins to telomeric domains in *Plasmodium falciparum*. *J Cell Sci* 121: 2046–2053. doi:10.1242/jcs.026427.
71. Figueiredo LM, Rocha EPC, Mancio-Silva L, Prevost C, Hernandez-Verdun D, et al. (2005) The unusually large *Plasmodium* telomerase reverse-transcriptase localizes in a discrete compartment associated with the nucleolus. *Nucleic Acids Res* 33: 1111–1122. doi:10.1093/nar/gki260.
72. Raska I (2003) Oldies but goldies: searching for Christmas trees within the nucleolar architecture. *Trends Cell Biol* 13: 517–525.
73. McStay B, Grummt I (2008) The epigenetics of rRNA genes: from molecular to chromosome biology. *Annu Rev Cell Dev Biol* 24: 131–157. doi:10.1146/annurev.cellbio.24.110707.175259.

74. Carmo-Fonseca M, Mendes-Soares L, Campos I (2000) To be or not to be in the nucleolus. *Nat Cell Biol* 2: E107–112. doi:10.1038/35014078.
75. Dundr M, Misteli T, Olson MO (2000) The dynamics of postmitotic reassembly of the nucleolus. *J Cell Biol* 150: 433–446.
76. Sekar RB, Periasamy A (2003) Fluorescence resonance energy transfer (FRET) microscopy imaging of live cell protein localizations. *J Cell Biol* 160: 629–633. doi:10.1083/jcb.200210140.
77. Lippincott-Schwartz J, Snapp E, Kenworthy A (2001) Studying protein dynamics in living cells. *Nat Rev Mol Cell Biol* 2: 444–456. doi:10.1038/35073068.
78. Lam YW, Lamond AI, Mann M, Andersen JS (2007) Analysis of nucleolar protein dynamics reveals the nuclear degradation of ribosomal proteins. *Curr Biol* 17: 749–760. doi:10.1016/j.cub.2007.03.064.
79. Andersen JS, Lyon CE, Fox AH, Leung AKL, Lam YW, et al. (2002) Directed proteomic analysis of the human nucleolus. *Curr Biol* 12: 1–11.
80. Olson MO, Dundr M, Szebeni A (2000) The nucleolus: an old factory with unexpected capabilities. *Trends Cell Biol* 10: 189–196.
81. Olson MOJ, Hingorani K, Szebeni A (2002) Conventional and nonconventional roles of the nucleolus. *Int Rev Cytol* 219: 199–266.
82. Olson MOJ, Dundr M (2005) The moving parts of the nucleolus. *Histochem Cell Biol* 123: 203–216. doi:10.1007/s00418-005-0754-9.
83. Misteli T (2001) The concept of self-organization in cellular architecture. *J Cell Biol* 155: 181–185. doi:10.1083/jcb.200108110.

Curriculum vitae

Sophie Clara Oehring
Dornacherstrasse 329
4053 Basel
Switzerland

Phone: 0041 61 331 65 14
Mobile: 0041 79 746 56 26
Email: Sophie.Oehring@unibas.ch

Experience

April 2008 - June 2012 **PhD thesis**, Swiss Tropical and Public Health Institute, University of Basel, Switzerland, supervised by Prof. T. Voss

Topic: A proteomic approach to identify and characterise novel nuclear proteins in *Plasmodium falciparum*

- > global analysis of the nuclear proteome by high accuracy mass spectrometry coupled with bioinformatics and *in vivo* localisation experiments
- > detailed analysis of selected nuclear proteins of *P. falciparum* on the functional level using various approaches including reverse genetics and co-immunoprecipitation
- > description of the *P. falciparum* nuclear compartment by visualisation of landmark structures based on fluorescence microscopy

Feb 2007 - Feb 2008 **Post-graduate position**, Syngenta Crop Protection AG, Stein, Switzerland

Topic: Applied research to investigate the efficacy of fungicides and to characterize the resistance of field isolates

- > identification of relevant fungal genes and specific point mutations playing a significant role in the resistance

June 2005 - June 2006 **Master thesis**, Institute of Plant Biology, University of Zurich, Switzerland, supervised by Dr. Markus Geisler

Topic: Mutational analysis of auxin transporter AtPGP1 in yeast

- > identification of motifs that might alter the auxin transport activity of AtPGP1, especially putative phosphorylation sites in the linker region that might be a possible regulation mechanism

Publications

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

Henrichs, S., Wang, B., Fukao, Y., Zhu, J., Charrier, L., Bailly, A., **Oehring, SC.**, Linnert, M., Weiwad, M., Endler, A., Nanni, P., Pollmann, S., Mancuso, S., Schulz, A. and Geisler M. (2012) Regulation of ABCB1/PGP1-catalysed auxin transport by linker phosphorylation The EMBO Journal

Curriculum vitae

Bouchard, R., Bailly, A., Blakeslee, J.J., **Oehring, SC.**, Vincenzetti, V., Paponov, I., Palme, K., Mancuso, S., Murphy, A., Schulz, B. and Geisler, M. (2006) Immunophilin-like TWISTED DWARF1 modulates auxin efflux activities of *Arabidopsis* p-glycoproteins. J. Biol. Chem.

Education

April 2008 - June 2012	PhD position, Swiss Tropical and Public Health Institute, University of Basel, Switzerland
June 2000 - June 2006	Master of Science in Biology, University of Zurich, Switzerland 1 st subject: Molecular biology, University of Zurich, Switzerland 2 nd subject: Immunology, ETH Zurich, Switzerland
Oct 2004	Practical course in molecular plant physiology, Faculty of Life Science, Dep. of Plant Biology, University of Copenhagen, Denmark

Technical Skills

molecular biology	PCR, genome walking, cloning, transformation, transfection (of bacteria, plants, fungi, plasmodium) site-directed mutagenesis and yeast "in vivo" mutagenesis DNA/RNA isolation sequencing and analysis (Geneious, SeqMan, MegAlign, BLAST) southern blot western blot, co-immunoprecipitation
biochemistry	yeast-two hybrid system (plant and plasmodium)
physiology	auxin transport assays (yeast) semi-quantitative growth assays (yeast)
microscopy	immunofluorescence assay and live cell imaging (plasmodium)
general	culture and maintenance of plants, yeast, fungi and plasmodium microbial bioassays

Employment

Feb 2007 - Feb 2008	Post-graduate position, Syngenta Crop Protection AG, Stein-Säckingen, Switzerland
---------------------	---

Teaching

Jan 2006	Course assistant: molecular plant physiology for undergraduate students, University of Zurich, Switzerland
Oct 2005 - Jan 2006	Course assistant: "Principles of life" (Microscopy of plant/cell structures) for High School teachers, University of Zurich, Switzerland

Curriculum vitae

Languages

German	mother tongue
English	fluent spoken & written (Language exchange in England)
French	intermediate spoken & written (Language exchange in France)
Spanish	intermediate spoken & written (7 years Residence in Costa Rica and Language exchange in Cuba)

Poster Presentations

Sept 2010	Marine biological laboratory MBL 21 th Annual Molecular Parasitology Meeting MPM Woods Hole, MA, USA <u>Poster title:</u> A proteomic approach identifies a large number of novel nuclear and ER-associated proteins in <i>P. falciparum</i>
Dez 2009	PEG/DFG 7th Malaria Meeting, Hamburg, Germany <u>Poster title:</u> Nuclear proteomics to identify novel factors involved in nuclear function and gene expression in <i>P. falciparum</i>
Sep 2009	Marine biological laboratory MBL (20 th) Annual Molecular Parasitology Meeting MPM Woods Hole, MA, USA <u>Poster title:</u> Nuclear proteomics to identify novel factors involved in nuclear function and gene expression in <i>P. falciparum</i>
Mai 2009	EMBL European Molecular Biology Laboratory 5th Annual BioMalPar Conference on the Biology and Pathology of the Malaria Parasite Heidelberg Heidelberg, Germany <u>Poster title:</u> A proteomic approach to identify and characterise nuclear proteins in <i>P. falciparum</i>
Nov 2008	PEG/DFG 6th Malaria Meeting, Tübingen, Germany <u>Poster title:</u> A proteomic approach to identify and characterise nuclear proteins in <i>P. falciparum</i>
Aug 2005	2 nd Tri-national Arabidopsis Meeting, Advancing the genomics frontier, Neuchâtel, Switzerland <u>Poster title:</u> Mutational analysis of auxin transporter AtPGP1 in yeast

Oral Presentations

Nov 2010	SSTMP (Swiss Society for Tropical Medicine and Parasitology) PhD Student Meeting, Basel, Switzerland <u>Presentation title:</u> A proteomic approach identifies a large number of novel nuclear and ER-associated proteins in <i>P. falciparum</i>
Nov 2010	PEG/DFG 8th Malaria Meeting, Basel, Switzerland <u>Presentation title:</u> A proteomic approach identifies a large number of novel nuclear proteins in <i>P. falciparum</i>
Oct 2009	SSTMP (Swiss Society for Tropical Medicine and Parasitology) PhD Student Meeting, Spiez, Switzerland <u>Presentation title:</u> Nuclear proteomics in <i>Plasmodium falciparum</i>
Oct 2008	SSTMP (Swiss Society for Tropical Medicine and Parasitology) PhD Student Meeting, Jongny, Switzerland <u>Presentation title:</u> Nuclear proteomics in <i>Plasmodium falciparum</i>

Curriculum vitae

Scholarships

Dez 2008 - Dez 2009 PhD scholarship Novartis Foundation for Medical-Biological Research

References

Prof. Dr. Till Voss Swiss Tropical and Public Health Institute, University of Basel,
Switzerland
Phone: 0041 61 284 81 61
Email: till.voss@unibas.ch

Dr. Helge Sierotzki Syngenta Crop Protection AG, Research Biology, Stein, Switzerland
Phone: 0041 62 866 01 80
Email: helge.sierotzki@syngenta.com

Dr. Markus Geisler University of Fribourg, Department of Biology - Plant Biology,
Switzerland
Phone: 0041 26 300 8827
Email: markus.geisler@unifr.ch

During my studies, I attended lectures of following lecturers:

C. Schönenberger, B. Fahrenkrog, M. Tanner, R. Brun, G. Pluschke, P. Mäser, H.P. Beck,
I. Felger, C. Daubenberger,

Basel, 12th of June 2012