MAPPINGS v1.0, a tool for network analysis of large phospho-signalling datasets: application to host erythrocyte response to *Plasmodium* infection

Jack Adderley

RMIT University https://orcid.org/0000-0001-8415-7539

Finn O'Donoghue

RMIT University

Stephen Davis

RMIT University

Christian Doerig (✉ christian.doerig@rmit.edu.au)

RMIT University https://orcid.org/0000-0002-3188-094X

Article

Abstract

Phosphorylation-based signalling implicates a complex and intertwined series of pathways and is critical to all domains of life. The interconnectivity between pathways results in the emergence of complex networks whose elucidation present a serious challenge. Large datasets of phosphorylation interactions through the activity of kinases on their numerous substrates are constantly being generated, however deciphering the complex network structure hidden in these datasets remains challenging. Many phosphorylation interactions occurring in human cells have been identified and constitute the basis for the known phosphorylation interaction network. We overlayed onto this network phosphorylation datasets obtained from an antibody microarray approach aimed at determining changes in phospho-signalling of host erythrocytes to infection with the malaria parasite Plasmodium falciparum. To analyse the datasets now mapped into the interaction network, we designed a pathway analysis tool denoted MAPPINGS that uses random walks to identify chains of phosphorylation events occurring much more or much less frequently than expected. MAPPINGS highlights pathways of phosphorylation that work synergistically, providing a rapid interpretation of the most critical pathways in each dataset. MAPPINGS confirmed several signalling interactions previously shown to be modulated by infection, and revealed additional interactions which could form the basis of numerous future studies. The MAPPINGS analysis strategy described here is widely applicable to comparative phosphorylation datasets in any context (e.g. response of cells to infection, treatment, or comparison between differentiation stages of any cell populations) and provides a rapid and reliable analysis to guide validation studies.

Introduction

Protein phosphorylation is one of the most prominent among several post-translational modifications which alter the functionality of affected proteins. Protein phosphorylation is achieved by a family of enzymes known as kinases, which transfer the gamma phosphate of adenosine triphosphate (ATP) onto hydroxyl groups of amino acids (protein kinases) [1]. In humans, protein kinases make up approximately 2% of encoded genes, yet over 50% of all proteins are phosphorylated across >200,000 known phosphorylation sites [2,3]; the inclusion of putative phosphorylation would put the number of potential phosphorylation sites in the human proteome closer to a million (see www.phosphonet.ca). Therefore, it is not surprising that kinases play essential roles in a plethora of intra- and extracellular processes [4-6]. The dynamic activation of kinases, along with the activity of protein phosphatases (enzymes which remove the phosphate groups from proteins) enable the fine control of numerous cellular functions ranging from regulating metabolism, transcription and translation, protein transport and cell growth, division and differentiation [3,7]. The capacity of protein kinases to act as substrates of other protein kinases underlies a complex interconnected web of signal transduction.

In a recent study, Olow *et al.* mapped a large number of these phosphorylation interactions in human cells, cataloguing 1733 functionally interconnected proteins into a network denoted as PhosphoAtlas [8]. This seminal work enabled large phosphorylation datasets to be mapped into sophisticated networks,

and has allowed for the exploration of how each phosphorylation on a given protein can impact its immediate neighbours in the network. Additionally, datasets that report on phosphorylation changes under various conditions (e.g. infection of a cell by a pathogen or treatment with a drug) can be analysed holistically to provide a more detailed understanding of how particular intracellular conditions impact on the phosphorylation-based signalling environment.

Large amounts of signalling information can be obtained through microarrays of phospho-specific antibodies, which report on phosphorylation changes across a large number of proteins. Antibody microarrays are highly sensitive quantitative tools capable of analysing a sample without complicated or expensive enrichment protocols, which is one of the drawbacks of mass spectrometry-based techniques (reviewed in [9]). This positions the antibody microarray as an ideal system to assess the signalling environment inside human cells in disease and infection settings. We have applied microarrays in our efforts to understand how intracellular pathogens alter their host cells during development and identified a number of key host proteins essential to the proliferation of various pathogens [10-12]. However, the datasets obtained from antibody microarray experiments are complex and difficult to interpret holistically.

Currently, typical analysis consists of identifying the key phosphorylation/protein changes manually, and the focus of any follow-up analysis is often based on the phosphorylation events that cause the largest changes observed. This often leads to less obvious changes being disregarded, despite the potential that some of these changes may play important signalling roles. To address this fundamental problem, we have developed a computational technique named MAPPINGS (Mapping and Analysis of Phosphorylation Pathways Identified through Network/Graph Signalling) that simulates and records random walks that `traverse' through an observed phospho-signalling network (available at https://github.com/FinnOD/mappings). The walks are not true random walks because of additional rules imposed that prevent repeated use of the same edge (such walks are called trails). The edges are weighted by the magnitude of the observed phosphorylation differences between the datasets (in our case, infected versus non-infected erythrocytes), so that the algorithm tends to generate paths associated with large negative (infection causes a decrease in the signal) or positive (infection causes an increased in the signal) differences (Figure 1). This strategy was formulated using PhosphoAtlas [8] as the underlying network, which was further curated to include phosphorylation connections relevant to the antibodies present on the microarray. We applied our computational technique to our published datasets on the blood-stage development of the human malaria parasite, *Plasmodium falciparum* [10]. This consisted of a control dataset (uninfected red blood cells), and three datasets obtained from erythrocytes infected with the parasite at the three major distinguishable stages of parasite development (see our original study for more detail [10]).

The benefits of the MAPPINGS approach were threefold; (i) it enabled the objective identification of pathways of signalling which went unnoticed in traditional analysis strategies, shedding light on some of the more elusive signalling dynamics; (ii) it provided a greatly accelerated starting point for future

analysis of similar signalling datasets; (iii) it enabled the identification of pivotal kinases in the network that functioned as major nodes of downstream signalling events. The significance at a practical level is that such identifications generate numerous new and testable hypotheses on the host signalling dynamics during blood-stage development of the malaria parasite, and is applicable to any phosphorylation-based dataset.

Results And Discussion

The datasets used to design this computational approach were sourced from Adderley *et al.*, 2020 [10]. This publication provided unique datasets that represent host erythrocyte signalling in the context of three stages of *P. falciparum* asexual development, namely ring-stage parasites (n=3), representing the early stages of development, trophozoite-stage parasites (n=3), representing the most metabolically active form, and schizont stage parasites (n=2), when the parasites daughter cells are assembled. This asexual development cycle is completed in 48 hours for *P. falciparum* (Figure 1). For each antibody on the array, each dataset reports the fold change from an uninfected red blood cell control which was used as the baseline in this comparison. Details about the datasets and further background information are available in the source article [10]; see [13] for a succinct review of the *P. falciparum* lifecycle. The analysis pipeline from sample generation through to the MAPPINGS analysis output is represented as a flow chart in Figure 1. The datasets were generated using the KAM-900 series antibody microarray produced by the kinomics company Kinexus, which contains 613 phosphorylation-specific antibodies. Additionally, a further 265 pan-specific antibodies recognising both the phosphorylated and unphosphorylated forms of a target protein are present, which provide information on protein abundance. As detailed in [10], a number of signals from the array were deemed unsuitable for analysis and were therefore removed from the analysis (Figure 1.3). The disregarded signals fell into at least one of the following three categories; (i) low signal intensity, (ii) relatively high error (compared to the change observed from uninfected control) and (iii) cross-reactivity to parasite proteins. This reduced the overall number of phosphorylation-specific antibody signals of each dataset to; 69 (ring), 184 (trophozoite) and 135 (schizont) (see methods section for detailed description of these categories). To account for degradation or expression of signalling proteins, each of the phosphorylation-specific signals on the array was normalised to account for changes in signalling protein abundance.

Network construction and mapping to biological datasets

Phosphorylation signalling is a highly interconnected network that contains numerous feedback loops which enable finely tuned responses to external/internal stimuli. Many of the globally identified phosphorylation events have unknown functions. Additionally, the specific kinases responsible for many of these phosphorylation events remain elusive. These factors prove challenging in the understanding of how various phosphorylation events come together into a greater network. Despite this, in a study by Olow et al. [8] a network containing 1733 proteins interconnected through phosphorylation interactions was pieced together. Using this study as a framework, we made further annotations to this network map to include; the target proteins response following phosphorylation at a specific site (activation/inhibition)

and added additional phosphorylation sites reported on the microarray which were missing from the network map (See Methods for further detail). The subsequent phosphorylation network consists of kinases/substrates which are represented as nodes, while the specific phosphorylation events are represented as the network's edges (arrows). The networks edges are directed and point towards the phosphorylated substrate; an arrowhead indicates an activation effect and a square indicates that phosphorylation causes inhibition of the substrate's activity (Figure 2a).

Once the phosphorylation network had been optimised (see above), the next challenge was to overlay the reliable data from each of the malaria datasets onto the network. A number of antibody signals from the microarray reported on dual or triple phosphorylation sites for a given substrate. This is quite common among antibodies that target phosphorylation sites on proteins, and many phosphorylation sites are in relatively close proximity. These phosphorylation's also often share common roles for the substrate [14]. The network used here was structured so that each individual phosphorylation site was given its own unique edge. Therefore, the dual and triple phosphorylation signals were split, and the associated signals reassigned to the now separated sites. Additionally, in the cases where more than one antibody on the microarray recognises the same phosphorylation site, the signals from all such antibodies were averaged. The three datasets were then overlayed on the framework to generate unique networks for the ring, trophozoite and schizont developmental stages. Each of the networks underwent an edge reduction step, reducing the number of parallel edges between the nodes, thereby simplifying the overall networks and enabling a straightforward integration in Python (Networkx package). Parallel edges were defined here as multiple unique phosphorylation sites on Protein A which were all phosphorylated by Protein B. Additionally, we further separated each unique network into a positive and a negative network, to allow an independent assessment of the phosphorylation events that either increased in intensity (positive networks) or decreased in intensity (negative networks) during infection (depicted in Figure 1.4, also see methods section for more detail). The complete positive trophozoite-stage network was exported as an edge list and rendered into a network map using Cytoscape v3.8.2 to provide a visual representation of these networks. This illustrated the level of interconnectivity, while also illustrating the highly interconnected nature of the datasets (Figure 2b & c).

MAPPINGS strategy

To analyse the flow of phosphorylation through the networks (corresponding to the ring, trophozoite and schizont stages) a computational approach was designed to generate modified random walks. The walks were not completely random because the algorithm explicitly avoided (i) repetition of self-loops and (ii) cycling through the same edges in highly interconnected regions. Such paths are known as trails in network science terminology. No edge is revisited; however, nodes could be repeated to allow differential pathway development through unique edges. Each dataset was assessed for walks that were overrepresented, indicating increases in phosphorylation during infection, in the case of the positive network, or decreased activity in the case of the negative network (suppression of the pathway during infection) (Figure 1). The positive network analysis weighted edge selection towards larger positive

phosphorylation changes between healthy and infected cells and consider all negative phosphorylation changes as unusable edges. The inverse of this was performed for the negative network analysis.

Pathway Analysis - The computational technique begins with the selection of a random node in the network. The algorithm identifies all possible outbound edges (representing potential phosphorylation events) and selects one by weighting its decision on the edge weights (representing the magnitude of the associated biological signals represented as a fold change). Once an edge is selected the algorithm moves down the edge and considers the outgoing edges originating from the new node; it repeats this process until terminated and a new walk is initiated (see Equation 1). A walk is terminated under three circumstances; (i) there are no available edges which have not previously been used in the current walk, (ii) the last edge used is an inhibitory phosphorylation event (depicted in Figure 2a) or (iii) the edge fails the termination check. The termination check provides a small chance that a walk is terminated and is based on the strength of the biological signal. The probability of termination was set to 20% for the control networks (see below for more details). The strongest signals in the data have a lower likelihood to result in termination with increasing likelihood for signals which were weaker. This was implemented to distinguish edge choices where only a single edge was ultimately available at any given edge selection stage. Without this parameter the fold change value assigned to an edge in this circumstance was irrelevant due to the lack of competing choices (for more detail see methods section).

$$P(N_a \rightarrow N_b) = \frac{f_{ab}}{\sum_i f_{ai}}$$

Equation 1. Probability that the edge connecting node N_a to N_b will be chosen by the modified random walk algorithm. f_{ab} represents the fold change of phosphorylation of N_b by N_a, where entries without a connecting edge and edges that have already been visited are given a fold change value of 0.

The pathway analysis was initially repeated until 1,000,000 trails were recorded. The output of the analysis resulted in a significant number of trails with a length <3 (data not shown). This was not surprising as the node selection was random, and therefore nodes with no outbound edges (non-kinases) were often selected. The optimal output for this analysis was to distinguish pathways of phosphorylation interactions (i.e. trails), rather than simply highlight individual phosphorylation events. To this end, a minimum walk length of 3 edges was introduced as a threshold, thereby disregarding walks that terminated shorter than this. To ensure adequate trails were recorded for each analysis, we required the program to run a total of 1,000,000 trails of length ≥3. Following completion of the designated number of trails, the walks of length ≥3were decomposed into single edges and tallied to determine the number of times each unique edge was utilised. These values were denoted as the total edge usage. Analysis of the total edge usage values indicated that there was a wide variation of edge usage that occurred for every dataset analysed. Upon further analysis it was clear that certain edges were overwhelmingly more frequently traversed, regardless of the underlying fold change data (data not shown). This was a

consequence of the base network structure, which, like most biological networks, displays nodes that act as hubs in the network because they have numerous inbound and outbound edges. To address this, control networks were included as a base for comparison of the effect that the fold change data caused (Figure 1.5). The control networks were identical to the respective Positive or Negative networks; however, the edges were not assigned fold change data. Therefore, when the pathway analysis was applied to these networks, there was no inherent biological preference for any particular edge. They did however provide a baseline usage of each edge in the respective networks and therefore provided a means to account for the different interconnectivity of the nodes. The total edge usage for each of the edges in the positive network, negative network and their respective control networks were compared to determine the percentage change from control (% CFC).

The MAPPINGS tool as well as detailed instructions for its use are accessible at https://github.com/FinnOD/mappings.

MAPPINGS analysis results

The pathways yielded by MAPPINGS for each of the developmental stages are presented in Figures 3 - 5. In each case, the analysis yielded pathways whose components have previously been implicated in infection with *Plasmodium,* providing a positive control and validation with respect to the ability of the approach to detect kinases that are modulated by infection. MAPPINGS also revealed additional pathways, providing a well-grounded rationale for further experimental work and potential novel targets. Detailed below are the key findings from each of the developmental stage analysed.

Ring stage network – The ring stage, which takes approximately 24 hours to complete [15], begins as a merozoite invades a red blood cell and establishes the infection. The antibody microarray dataset used in this analysis covered a time window of 8 – 16 hours after invasion and compared a population of 33% infected erythrocytes to an uninfected control, because it is problematic to purify infected cells from uninfected cells at this early stage of infection, in contrast to the trophozoite and schizont stages, for which we can obtain preparation of >95% infected cells (see below). Despite this limitation, a number of changes in ring-infected versus uninfected erythrocyte signalling components were identified. The output of the ring stage network analysis contained 20 nodes (proteins/kinases) and 21 edges (phosphorylations) (Figure 3). The most prominent host signalling elements of the positive network were the non-receptor tyrosine kinases (TK) Src and FAK1/2 and the receptor Tyrosine kinase (RTK) RET (connected by green edges). In the negative network are the TKs Lck, Lyn, Syk, the serine/threonine kinases PKCα/δ/μ/ε, and the protein PEA15 (connected by red edges), discussed below.

Syk/Lyn – Phosphorylation of the membrane protein Band 3 by the Syk TK is essential to destabilise the erythrocyte membrane during parasite egress, and Syk inhibitors have been shown to block this process [16]. Our network analysis detected a decrease in Syk phosphorylation of the activation-associated residue Y323 during ring stage infection. The decrease in Syk phosphorylation in the early stages of infection is consistent with the parasite preventing the premature lysis of the RBC. This block may simply be released

at the end of schizont stage development to allow parasite egress; this would explain why there is no detected increase (relative to uninfected red blood cells) in schizonts. The network analysis indicates that the TKs Lck and/or Lyn may be implicated in the reduction of Syk phosphorylation in rings. Both Lck and Lyn are non-receptor TKs that belong to the Src family, with wide functionality from proliferation through to apoptosis and metabolic signalling [17]. Lck is primarily expressed in T-cells and the brain [18], while Lyn has been identified to have roles across several cells types of hematopoietic origin [19], and has been detected in the erythrocyte proteome [20]. Consequently, it is more likely that Lyn is responsible for Syk phosphorylation observed; however, this may reveal a novel function for Lck in the erythroid lineage. Lyn was previously suggested to be involved in Band 3 phosphorylation in uninfected erythrocytes, implicating a possible role in *P. falciparum* infection that warrants further exploration [21,22].

Protein kinase C (PKC) – The human PKC family comprises 10 isoforms of serine/threonine kinases that function in the phosphoinositide pathway to regulate a range of cellular processes. A decrease in overall host PKC activity during *P. falciparum* infection of red blood cells was first reported more than 20 years ago [23], which is consistent with the decrease in PKCα, PKCδ, PKCμ and PKCε phosphorylation detected by our analysis. PKCδ phosphorylation was also decreased in the trophozoite analysis as well (see Figure 4). The biological function of this decrease is not understood.

Focal adhesion kinases (FAK1/2) – FAKs are non-receptor TKs which serve to promote signalling through recruitment to activated cell surface receptors, notably of the integrin family [24]. Our analysis indicated that FAK1/2 are phosphorylated on the activating residues Y397/Y402 during infection. Further, the activation of FAK1 was notable across all time points examined during parasite development, suggesting it may have a continuous role in the infected host cell. These activating phosphorylation event can be mediated by a number of kinases, including the aforementioned Src TK [25]. No implication of FAKs or RET in infection has been reported, but our data suggest it may be of interest to explore this further (Figure 3 and 4). Indeed, FAKs can be activated by membrane deformation, therefore, it is tempting to speculate that the ontogeny of knobs (made of proteins exported by *P. falciparum* to the red blood cell membrane to provide cytoadherence, see [26]) in the plasma membrane of infected red blood cells may trigger FAK signalling.

Trophozoite stage network – Trophozoite are the most metabolically active stage of development and are notable for extensive host cell modification and increased haemoglobin digestion [27]. The antibody microarray dataset used in this analysis covered a window of 24 – 28 hours post invasion and compared a population of >95% infected cells to an uninfected counterpart [10]. This was possible because the parasite's digestion of host red blood cell haemoglobin leads to the accumulation of a paramagnetic pigment known as hemozoin, enabling magnet-mediated enrichment of infected cells [28]. The higher level of activity of the parasite at its trophozoite stage, combined with the greater level of enrichment, resulted in a greater number of signals of sufficient quality to be included in this analysis. The output of the trophozoite stage analysis contained a subnetwork of 53 nodes (proteins/kinases) and 66 edges (phosphorylation's) (Figure 4). The most prominent host signalling elements of the positive network were

the aforementioned TKs FAK and Ret, and several MAPKs (connected by green edges; see below). In the negative network were the Syk TK and the PAK1 serine/threonine kinase, and the substrate protein RPS6 (connected by red edges), discussed below.

Syk – The reduction in Syk phosphorylation observed at ring stage is not maintained at the trophozoite stage. This suggests the possible inhibition occurring during the ring stage begins to be alleviated at the trophozoite stage, which is fully consistent with the appearance of Band 3 phosphorylation in mature trophozoites [29].

FAK-Ret-MAPK – Phosphorylation of the FAK and RET TKs observed in the ring stage analysis (see above) appears to still be present at the trophozoite stage, though to a lesser extent (Figure 4). In addition to Src (see above), the hepatocyte growth factor receptor (MET) is activated in trophozoites and represents a possible additional activator of FAK1, which suggest multiple activation pathways may converge on FAK1 (Figure 4). MET phosphorylation in trophozoite-infected erythrocytes has been previously confirmed by Western blot [10]. Interestingly, a strong candidate for an effector of RET in trophozoites, the mitogen activated protein kinase 3 (MAPK3, or ERK1) was not observed in the ring stage analysis. ERK1 phosphorylation has not been investigated in the context of *P. falciparum* development in red blood cells. Two of the alternative activators of ERK1 from this analysis were MAP2K1 and MAP2K2 (otherwise known as MEK1/2). Though phosphorylation of MEK1/2 was not detected at the trophozoite stage, active ERK1 strongly suggest this is indeed the case. MEK1/2 phosphorylation during the later stages of *P. falciparum* blood stage development has been reported previously [30]. Unfortunately, the signals for these phosphorylation sites were included in those that were disregarded because of low reliability (see above and reference [10]). Interestingly, abnormal MEK1 phosphorylation of ERK signalling in erythrocytes of patients with sickle cell disease (SCD) is critical for the adhesive interactions of these cells with the endothelium [31]. Together with our observations on MAPK pathway activation, this may have profound implications with respect to the mechanisms of cytoadherence of *P. falciparum* infected red blood cells.

P21 activated kinase 1 (PAK1) – PAK1 is a serine/threonine kinase with multiple roles in regulation the cytoskeleton and apoptosis [32]. The activation of PAK1 by *Plasmodium* infection has been reported [30], but the mechanism of its activation remains unclear. Our analysis suggest it may be of interest to determine whether PDK1 plays a role upstream of this pathway.

Ribosomal protein S6 (RPS6) – While there are no published data on RPS6 during *P. falciparum* red blood cell infection, it has been shown that infected hepatocytes, during the liver stage of *P. falciparum* life cycle, show elevated levels of RPS6 phosphorylation [33]. In our analysis we noted a decrease in phosphorylation of the S235 site during trophozoite development. The primary activator of RPS6 is the Ribosomal Protein S6 Kinase (S6K), but PKC (which is activated by infection, see above) has also been reported to be involved in RPS6 phosphorylation [34]. Interestingly, the sites T421/S424 on RPS6, which acts to enhance activation, can be phosphorylated by ERK1/2 [35]. The T421/S424 site was not part of the

microarray dataset; it would be of great interest to investigate this site further in relation with MAPK pathways and RPS6 S235 phosphorylation.

Schizont stage network – The final notable stage of *P. falciparum* asexual development within human erythrocytes is the schizont stage, which accounts for the final hours of the asexual lifecycle. Schizont stage parasites are less metabolically active than the trophozoites, with a large proportion of the host cells cytosol now digested [36]. The antibody microarray dataset used in this analysis was performed on infected red blood cells 44 – 48 hours post invasion, corresponding to the final few hours before daughter cell release. This dataset, like that of the trophozoite dataset compared a population of >95% infected cells to an uninfected counterpart [10]. The output of the schizont stage network analysis contained a subnetwork of 42 nodes (proteins/kinases) and 45 edges (phosphorylation's) (Figure 5). The trophozoite and schizont analysis shared a number of similar signalling connections, suggesting an overall signalling trend at the later stages of blood stage development. These commonalities include Src, FAK, RET MAPK3, MAPK14 (also known as p38α) and many of the PKC isoforms. The most striking pathway from the schizont stage analysis is the pathway from MAPK14 (p38α) to Mdm2, discussed below.

P38α - MAPKAPK2 - Mdm2 pathway - p38α is a mitogen-activated protein kinase (MAPK) with key responsibilities in erythroblast enucleation during stress erythropoiesis [37]. Interestingly p38α activity has also been linked to stress responses in red blood cells, with a suspected role in eryptosis (the red blood cell equivalent of apoptosis) [38]. The reduction in phosphorylation observed in our analysis could indicate that this cell death pathway is being circumvented by the parasite to facilitate prolonged survival of its host cell. The downstream effector of p38α, Mdm2 is a E3 ubiquitin-protein ligase and is responsible for the ubiquitination of TP53 (or p53) which flags p53 for proteasomal degradation [39]. Its role in mature red blood cells is unknown, but it is essential for regulating erythropoiesis [40]. Our analysis points to a reduction in Mdm2 phosphorylation at S166, which is known to enhance the proteins capacity to inactivate p53 signalling [41]. Interestingly, p53 is shown to be activated from our analysis, which could be in part due to a reduction in Mdm2 activity. Together this illustrates a possible 'pro survival' pathway being activated, which could be the result of direct signalling manipulation by the parasite. In the absence of transcriptional activity in erythrocyte, this concept is intriguing and warrants further exploration.

Concluding Remarks

The combination of (i) increasingly (very) large datasets originating from phospho-proteomic approaches (including those obtained through arrays comprising phospho-specific antibodies as exemplified here), and (ii) the tremendous underlying complexity of phospho-signalling networks in mammalian cells, raises serious issues with respect to deciphering the modulation of signalling pathways between biological samples. This calls for system-wide computational approaches to deconvolute raw data. MAPPINGS, the network-based analysis tool developed here aims at contributing to this important task, and consists of a pathway-based analysis tool that uses random walks to identify chains of consistent phosphorylation events. MAPPINGS highlights pathways of phosphorylation that work synergistically to provide a rapid

interpretation of the consistent pathways in a dataset. A limitation of this approach is in the treatment of the inhibitory associated edges: MAPPINGS currently terminates trails following the selection of an inhibitor edge; however if a subsequent edge out of this node indicated the inverse effect (i.e increase in inhibition then a decrease on outbound edges) this would be congruent with the expected flow of phosphorylation. Consequently, this leads to a small number of trails not being explored and therefore not represented in the output. This limitation is currently being addressed and may be amended into the program in a later version.

The MAPPINGS tool developed here is applicable to any antibody microarray or phospho-proteomic dataset comprising signalling proteins. MAPPINGS can be utilised to provide detailed pathway analysis on already published datasets, and will be an effective tool to screen new datasets in any context where there is a need to compare phospho-signalling between two biological samples. We illustrate this concept through the identification of host erythrocyte pathway modulation during the development of malaria parasites. A notable proportion of the pathways identified with MAPPINGS are consistent with published host signalling studies, validating the strategy. Additionally, new and exciting host signalling interactions were observed, for example signalling pathways that implicate FAK, p38α and RET. MAPPINGS analysis is essentially a hypothesis-generating exercise, and these findings now need functional validation. The findings in the specific example used here are far-reaching. Currently all deployed antimalarials and those in development target parasite-encoded proteins, with many derivatives of current or previous deployed compounds [42]. Parasite resistance and ensuing treatment failure is becoming apparent for every deployed antimalarial [43]. This calls for the development of next-generation drugs with (i) untapped modes of action to prevent cross-resistance and (ii) have low propensity for the emergence of *de novo* resistance. In recent years *P. falciparum* has been shown to require the activity of several of its host kinases [30,44,45], which when inhibited chemically, result in parasite death. This suggests that host targeted drug discovery (HDT) may be feasible avenue for malaria treatments as it has for other infectious diseases (reviewed in [46]).

Methods

Microarray datasets – Datasets used to design this random walks-based network analysis were published in Adderley *et al*. 2020 [10], containing a total of 8 datasets, including replicates. The replicate datasets were unified with the replicate values averaged for the analysis. These datasets were ring-stage parasites (n=3), trophozoite-stage parasites (n=3), and schizont stage parasites (n=2). Each dataset reports the fold change from an uninfected red blood cell control which was used as the base line of signalling. See source article for further details on these samples and the manual data analysis performed [10].

Signal filtering – A number of signals reported by the antibody microarray were deemed unreliable. These signals fell into three categories, and were removed from the dataset and the subsequent analysis in the present work. These categories were; (i) low signal intensity, (ii) relatively high error (compared to change observed from control) and (iii) cross-reactivity to parasite proteins. <u>Low intensity signals:</u> were defined as signals were both the control (uninfected sample) and infected sample were below 1000 relative units. This threshold is recommended by the manufacturer, as signals below this intensity are often difficult to validate. <u>High error relative to signal change:</u> in some instances, signals appeared to vary notably between the biological replicates of that datasets. To account for this, we combined the uninfected and infected signals error for each unique antibody on the microarray (which are reported as percentage error) and disregarded any antibody whose total signal error was greater than the percentage change reported from the uninfected control. <u>Cross-reactive signals:</u> we removed the antibodies that were identified as cross-reactive (see [10] for more detail on cross-reactive signal determination).

The substrate effect for each phosphorylation in the network – As mentioned in the introduction, phosphorylation fundamentally results in either the activation or inhibition of the target substrate. The network we based our study on [8] did not record the effect that each phosphorylation event had on the target substrate. As this information is crucial biological interpretation of the output data, where possible we annotated this information into the network. This information was provided by Kinexus, otherwise literature searching was undertaken to classify as many phosphorylation effects as possible as activation or inhibition. Despite these efforts, a number of phosphorylation sites have unknown effects, consequently these sites could not be annotated in our networks. To enable continuity of the MAPPINGS analysis these sites are treated as though they were activation sites; whenever future studies uncover the function of these sites the base network used in this analysis can be updated accordingly.

Assignment of biological data to duplicated edges – One limitation of this approach is that the upstream kinase responsible for each of the substrate phosphorylation's is unknown. Consequently, as multiple kinases can often phosphorylate the same target substrate at the same phosphorylation site, the fold change data was assigned to each of these possible interactions in the network. This means that in some instances a single antibody signal is mapped to multiple edges. As it is unlikely that all possible kinases contribute to the phosphorylation of a single substrate at once, the MAPPINGS analysis (see analysis strategy section) was designed to determine which of the possible kinases was most likely resulting in the phosphorylation event observed.

Developing a positive and negative network for independent dataset analysis – Once the biological data and network were combined, there were a hand full of nodes which were connected by multiple parallel directed edges. To enable a straightforward analysis strategy, we applied an edge reduction step that left

a single directed edge between each of the two nodes. To account for the multiple directed edges that had varying phosphorylation-specific antibody signals, we developed two independent networks using the antibody microarray data. One of these networks was designated the positive network, which retained the phosphorylation-specific antibody signals which increased during infection, while the other network was designated the negative network, which retained the decreasing signals during infection. In both instances the greater magnitude values were preferentially selected. This enabled the independent assessment of the phosphorylation interactions that increased during infection and those that decreased.

Analysis strategy walk termination settings – There were three termination checks placed into the function used in this analysis: <u>(i) No edges to choose</u> - If there were no remaining usable edges available the walk would terminate, this was to avoid self-loop and to stop cyclic connections being over reporting. <u>(ii) Inhibitory signalling</u> - If the last edge used during a walk was an inhibition phosphorylation (depicted in Figure 2a) the walk would terminate. Inhibitory phosphorylation results in the de-activation of the target protein or kinase. Therefore, walks were terminated following usage to remain consistent with what would happen in a biological setting. <u>(iii) Weighted termination chance</u> – The weighted termination chance enabled the function to discriminate edge usage due to fold change data when a single edge was available during edge selection. In the circumstance where a single edge is available, the function will select it, as there are no other options. This was problematic, as situations where the edge option was 1 the output would report no change in edge usage between the control and microarray data networks, regardless of the strength of the microarray data. By including a weighted chance for edge termination based on the magnitude of the edges fold change we were able to account for this in our analysis. A weighted chance which scaled from 0 – 20% was applied after each step in a walk. This scaling was linearly assigned to the fold change data in the positive and negative networks with the largest magnitude fold changes being assigned no termination chance (0%) and smallest fold changes being assigned a chance of 20%. For the control networks, where no fold change values were assigned, the weighted termination chance was set at 20% globally.

The MAPPINGS tool as well as detailed instructions for its use are accessible at https://github.com/FinnOD/mappings.

Declarations

Author contributions

J.A. and F.O. performed the analysis and designed the program.

J.A., C.D. and S.D. designed the project and guided the analysis.

All authors contributed the writing of the manuscript.

Competing interests

The authors declare no competing interests.

References

1   Fabbro, D., Cowan-Jacob, S. W. & Moebitz, H. Ten things you should know about protein kinases: IUPHAR Review 14. *Br J Pharmacol* **172**, 2675-2700, doi:10.1111/bph.13096 (2015).

2   Taylor, S. S. & Kornev, A. P. Protein kinases: evolution of dynamic regulatory proteins. *Trends Biochem Sci* **36**, 65-77, doi:10.1016/j.tibs.2010.09.006 (2011).

3   Endicott, J. A., Noble, M. E. & Johnson, L. N. The structural basis for control of eukaryotic protein kinases. *Annual review of biochemistry* **81**, 587-613, doi:10.1146/annurev-biochem-052410-090317 (2012).

4   Cohen, P. The origins of protein phosphorylation. *Nat Cell Bio* **4**, E127+ (2002).

5   Kannan, N., Taylor, S. S., Zhai, Y., Venter, J. C. & Manning, G. Structural and functional diversity of the microbial kinome. *PLoS Biol* **5**, e17, doi:10.1371/journal.pbio.0050017 (2007).

6   Walsh, C. T., Garneau-Tsodikova, S. & Gatto, G. J., Jr. Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew Chem Int Ed Engl* **44**, 7342-7372, doi:10.1002/anie.200501023 (2005).

7   Cargnello, M. & Roux, P. P. Activation and Function of the MAPKs and Their Substrates, the MAPK-Activated Protein Kinases. *Microbiology and Molecular Biology Reviews : MMBR* **75**, 50-83, doi:10.1128/mmbr.00031-10 (2011).

8   Olow, A. *et al.* An Atlas of the Human Kinome Reveals the Mutational Landscape Underlying Dysregulated Phosphorylation Cascades in Cancer. *Cancer Res* **76**, 1733-1745, doi:10.1158/0008-5472.CAN-15-2325-T (2016).

9   Yue, L. & Pelech, S.

10   Adderley, J. D. *et al.* Analysis of erythrocyte signalling pathways during Plasmodium falciparum infection identifies targets for host-directed antimalarial intervention. *Nat Commun* **11**, 4015, doi:10.1038/s41467-020-17829-7 (2020).

11   Haqshenas, G. *et al.* A Role for the Insulin Receptor in the Suppression of Dengue Virus and Zika Virus in Wolbachia-Infected Mosquito Cells. *Cell Rep* **26**, 529-535 e523, doi:10.1016/j.celrep.2018.12.068 (2019).

12   Haqshenas, G. *et al.* Signalome-wide assessment of host cell response to hepatitis C virus. *Nat Commun* **8**, 15158, doi:10.1038/ncomms15158 (2017).

13   Venugopal, K., Hentzschel, F., Valkiunas, G. & Marti, M. Plasmodium asexual growth and sexual development in the haematopoietic niche of the host. *Nat Rev Microbiol* **18**, 177-189, doi:10.1038/s41579-019-0306-2 (2020).

14   Li, Y., Zhou, X., Zhai, Z. & Li, T. Co-occurring protein phosphorylation are functionally associated. *PLoS Comput Biol* **13**, e1005502, doi:10.1371/journal.pcbi.1005502 (2017).

15   Gruring, C. *et al.* Development and host cell modifications of Plasmodium falciparum blood stages in four dimensions. *Nat Commun* **2**, 165, doi:10.1038/ncomms1169 (2011).

16   Pantaleo, A. *et al.* Syk inhibitors interfere with erythrocyte membrane modification during P falciparum growth and suppress parasite egress. *Blood* **130**, 1031-1040, doi:10.1182/blood-2016-11-748053 (2017).

17   Parsons, S. J. & Parsons, J. T. Src family kinases, key regulators of signal transduction. *Oncogene* **23**, 7906-7909, doi:10.1038/sj.onc.1208160 (2004).

18   Bommhardt, U., Schraven, B. & Simeoni, L. Beyond TCR Signaling: Emerging Functions of Lck in Cancer and Immunotherapy. *Int J Mol Sci* **20**, doi:10.3390/ijms20143500 (2019).

19   Ingley, E. Functions of the Lyn tyrosine kinase in health and disease. *Cell Commun Signal* **10**, 21, doi:10.1186/1478-811X-10-21 (2012).

20   Bryk, A. H. & Wisniewski, J. R. Quantitative Analysis of Human Red Blood Cell Proteome. *J Proteome Res* **16**, 2752-2761, doi:10.1021/acs.jproteome.7b00025 (2017).

21   Pantaleo, A. *et al.* Band 3 Erythrocyte Membrane Protein Acts as Redox Stress Sensor Leading to Its Phosphorylation by p (72) Syk. *Oxid Med Cell Longev* **2016**, 6051093, doi:10.1155/2016/6051093 (2016).

22   Brunati, A. M. *et al.* Sequential phosphorylation of protein band 3 by Syk and Lyn tyrosine kinases in intact human erythrocytes: identification of primary and secondary phosphorylation sites. *Blood* **96**, 1550-1557 (2000).

23   Hall, B. S., Daramola, O. O., Barden, G. & Targett, G. A. Modulation of protein kinase C activity in Plasmodium falciparum-infected erythrocytes. *Blood* **89**, 1770-1778 (1997).

24   Zhao, X. & Guan, J. L. Focal adhesion kinase and its signaling pathways in cell migration and angiogenesis. *Adv Drug Deliv Rev* **63**, 610-615, doi:10.1016/j.addr.2010.11.001 (2011).

25   Lu, Y. & Sun, H. Progress in the Development of Small Molecular Inhibitors of Focal Adhesion Kinase (FAK). *J Med Chem* **63**, 14382-14403, doi:10.1021/acs.jmedchem.0c01248 (2020).

26   Subramani, R. *et al.* Plasmodium falciparum-infected erythrocyte knob density is linked to the PfEMP1 variant expressed. *mBio* **6**, e01456-01415, doi:10.1128/mBio.01456-15 (2015).

27   Elliott, D. A. *et al.* Four distinct pathways of hemoglobin uptake in the malaria parasite Plasmodium falciparum. *Proc Natl Acad Sci U S A* **105**, 2463-2468, doi:10.1073/pnas.0711067105 (2008).

28   Inyushin, M. *et al.* Superparamagnetic Properties of Hemozoin. *Sci Rep* **6**, 26212, doi:10.1038/srep26212 (2016).

29   Billett, H. H. Syk TKIs "strengthen" RBCs against malaria. *Blood* **130**, 960-961, doi:10.1182/blood-2017-07-796474 (2017).

30   Sicard, A. *et al.* Activation of a PAK-MEK signalling pathway in malaria parasite-infected erythrocytes. *Cell Microbiol* **13**, 836-845, doi:10.1111/j.1462-5822.2011.01582.x (2011).

31   Zennadi, R. MEK1/2 as a Therapeutic Target in Sickle Cell Disease. *Int J Blood Res Disord* **6**, doi:10.23937/2469-5696/1410038 (2019).

32   Dummler, B., Ohshiro, K., Kumar, R. & Field, J. Pak protein kinases and their role in cancer. *Cancer Metastasis Rev* **28**, 51-63, doi:10.1007/s10555-008-9168-1 (2009).

33   Glennon, E. K. K. *et al.* Alterations in Phosphorylation of Hepatocyte Ribosomal Protein S6 Control Plasmodium Liver Stage Infection. *Cell Rep* **26**, 3391-3399 e3394, doi:10.1016/j.celrep.2019.02.085 (2019).

34   Valovka, T. *et al.* Protein kinase C phosphorylates ribosomal protein S6 kinase betaII and regulates its subcellular localization. *Mol Cell Biol* **23**, 852-863, doi:10.1128/mcb.23.3.852-863.2003 (2003).

35   Biever, A., Valjent, E. & Puighermanal, E. Ribosomal Protein S6 Phosphorylation in the Nervous System: From Regulation to Function. *Front Mol Neurosci* **8**, 75, doi:10.3389/fnmol.2015.00075 (2015).

36   Krugliak, M., Zhang, J. & Ginsburg, H. Intraerythrocytic Plasmodium falciparum utilizes only a fraction of the amino acids derived from the digestion of host cell cytosol for the biosynthesis of its proteins. *Mol Biochem Parasitol* **119**, 249-256, doi:10.1016/s0166-6851(01)00427-3 (2002).

37   Schultze, S. M. *et al.* p38alpha controls erythroblast enucleation and Rb signaling in stress erythropoiesis. *Cell Res* **22**, 539-550, doi:10.1038/cr.2011.159 (2012).

38   Gatidis, S. *et al.* p38 MAPK activation and function following osmotic shock of erythrocytes. *Cell Physiol Biochem* **28**, 1279-1286, doi:10.1159/000335859 (2011).

39   Karni-Schmidt, O., Lokshin, M. & Prives, C. The Roles of MDM2 and MDMX in Cancer. *Annu Rev Pathol* **11**, 617-644, doi:10.1146/annurev-pathol-012414-040349 (2016).

40   Maetens, M. *et al.* Distinct roles of Mdm2 and Mdm4 in red cell production. *Blood* **109**, 2630-2633, doi:10.1182/blood-2006-03-013656 (2007).

41   Chen, J. The Roles of MDM2 and MDMX Phosphorylation in Stress Signaling to p53. *Genes Cancer* **3**, 274-282, doi:10.1177/1947601912454733 (2012).

42   Tse, E. G., Korsik, M. & Todd, M. H. The past, present and future of anti-malarial medicines. *Malar J* **18**, 93, doi:10.1186/s12936-019-2724-z (2019).

43   Thu, A. M., Phyo, A. P., Landier, J., Parker, D. M. & Nosten, F. H. Combating multidrug-resistant Plasmodium falciparum malaria. *FEBS J* **284**, 2569-2578, doi:10.1111/febs.14127 (2017).

44   Adderley, J. *et al.* Signalome-wide assessment of erythrocyte response to Plasmodium reveals novel targets for host-directed antimalarial intervention. *bioRxiv*, 2020.2006.2009.143206, doi:10.1101/2020.06.09.143206 (2020).

45   Kesely, K. *et al.* Identification of tyrosine kinase inhibitors that halt Plasmodium falciparum parasitemia. *PLoS One* **15**, e0242372, doi:10.1371/journal.pone.0242372 (2020).

46   Zumla, A. *et al.* Host-directed therapies for infectious diseases: current status, recent progress, and future prospects. *Lancet Infect Dis* **16**, e47-63, doi:10.1016/S1473-3099(16)00078-5 (2016).

47   Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504, doi:10.1101/gr.1239303 (2003).

48   Adderley, J., Williamson, T. & Doerig, C. Parasite and host erythrocyte kinomics of Plasmodium infection. *Trends Parasitol*, doi:10.1016/j.pt.2021.01.002 (2021).
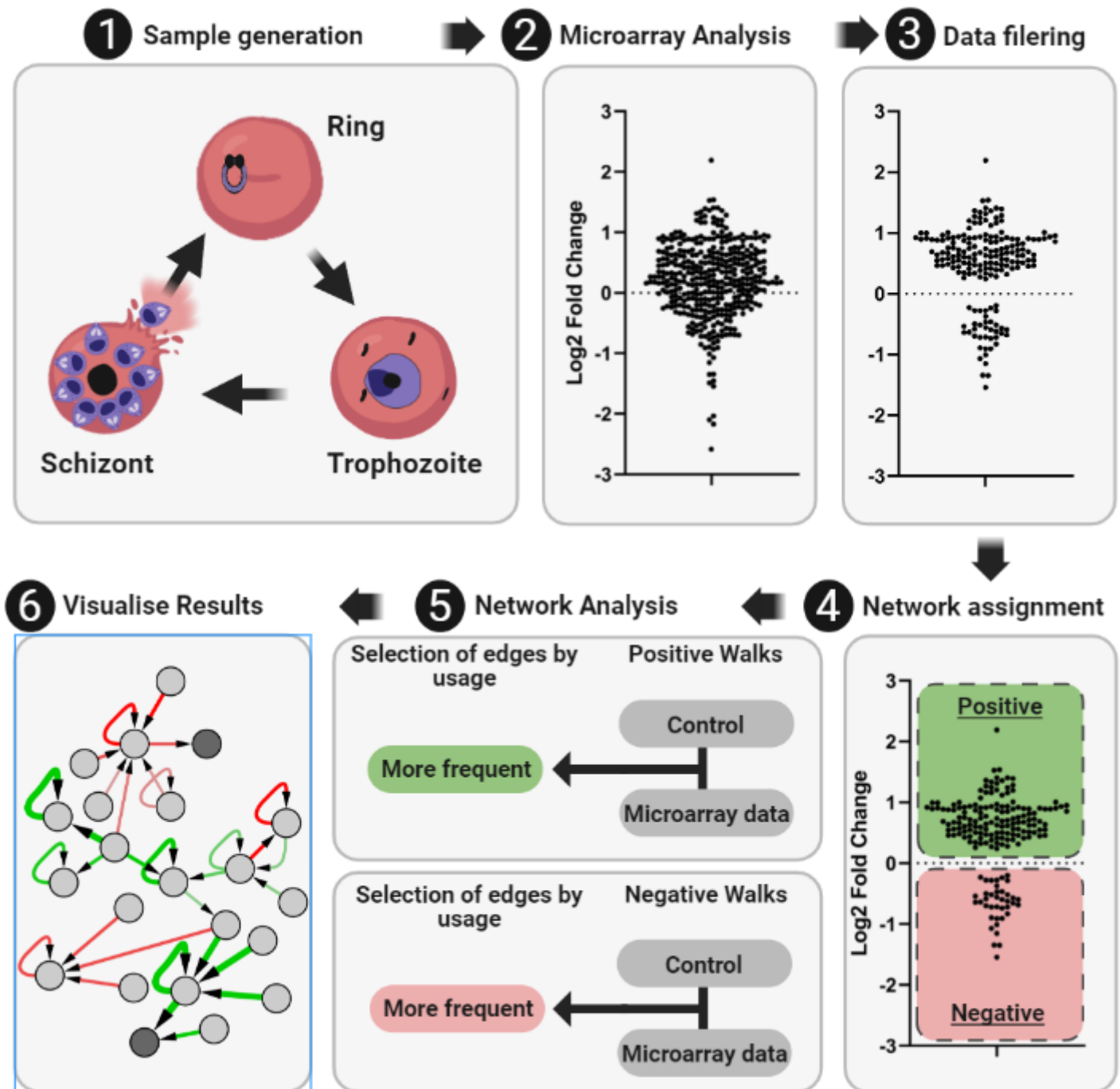
Figures

Figure 1

Overview of data generation, processing, modelling and output using antibody microarrays and MAPPINGS. 1) Sample generation, samples used for the development of the network analysis approach used here were of human red blood cells following infection with Plasmodium falciparum. 2) Application

of the samples onto the antibody microarrays. 3) Selection of reliable antibodies based on the signal intensity observed, the error observed between the replicates and removal of signals identified as cross reactive to parasite material (see Methods for more detail). 4) The reliable signals are split into positive (green) and negative (red) networks and are mapped to the known global phosphorylation interactions. 5) The path analysis simulates and records large numbers of modified random walks (see main text) where paths are weighted by the observed intensity differences between infected and healthy red blood cells. Separate analyses were performed using the positive and negative values of the log2 fold change for the edge weights to tease apart signalling pathways enhanced (positive weights) or suppressed (negative weights) by the parasite. 6) Output is captured as an edge list consisting of comma-separated values (.csv) which can be visualised in programs such as Cystoscope 47. Here we have represented the more frequent edges for the positive network (green) and the negative network (red) Figure adapted from 48 and modified using BioRender.com.
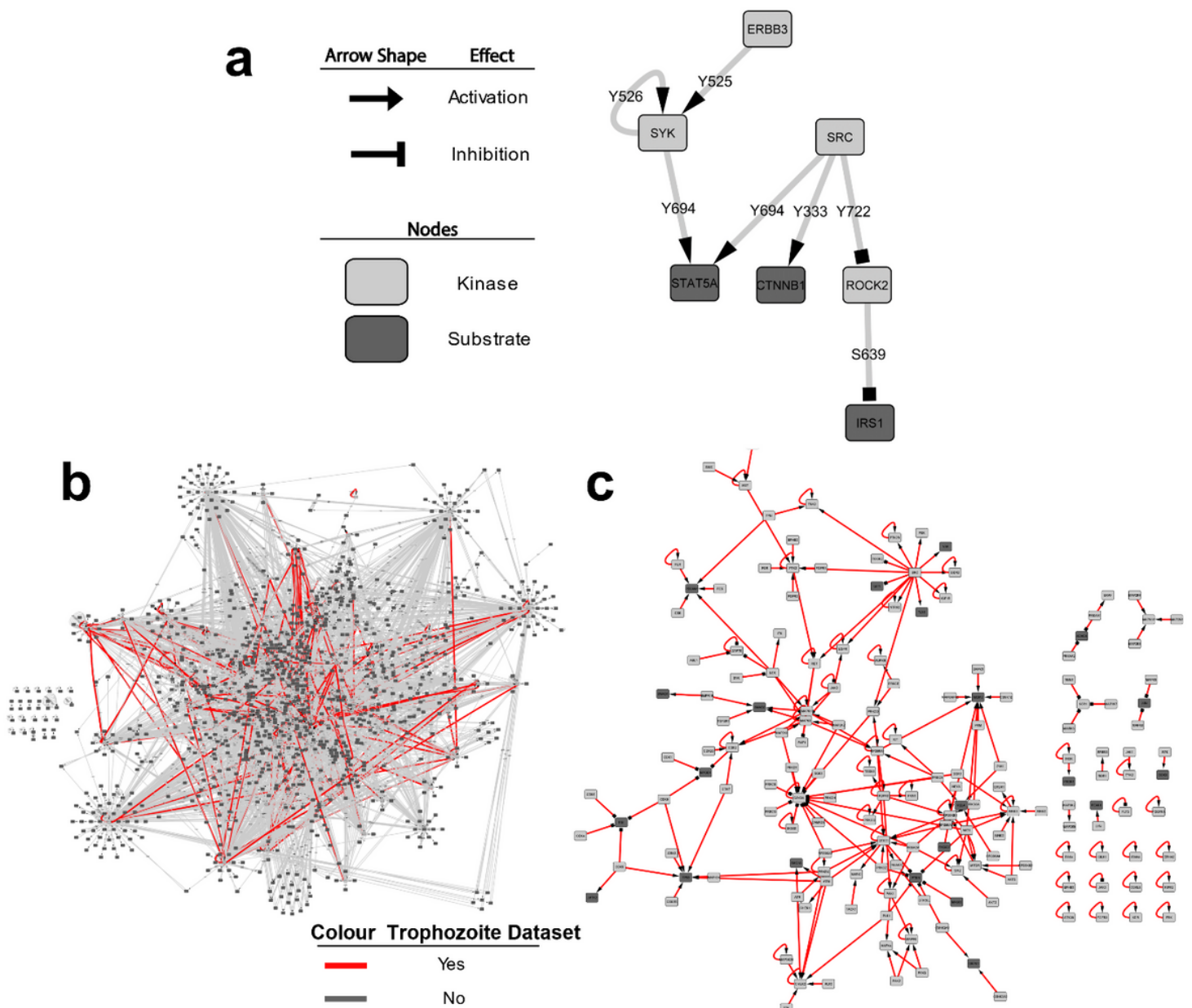

Figure 2

Example of the phosphorylation network structure and an overlay of trophozoite specific antibody microarray data analysed. a) Definition of edge type and node type in the networks utilised throughout this study. Kinases and substrates are represented as nodes (dark nodes = substrates, light nodes = kinases). Phosphorylation's events are represented as directed edges, with edges pointing towards the substrate of the interaction. The effect of the edge is represented in the arrowhead (arrow = activation, square = inhibition). b) The basal human phospho-signalling network used here contains 1156 proteins (nodes, dark = substrate, light = kinases) and 6224 phosphorylation connections (edges, grey = not in trophozoite dataset, red = in trophozoite dataset). c) Subnetwork of the connections in the human phosphorylation network which were assigned antibody microarray data from the trophozoite dataset. This subnetwork contains 167 proteins (nodes) and 237 phosphorylation connections (edges). The nodes and edges were identified as having reliable phosphorylation changes during the trophozoite stage of P. falciparum blood stage development following signal filtering.
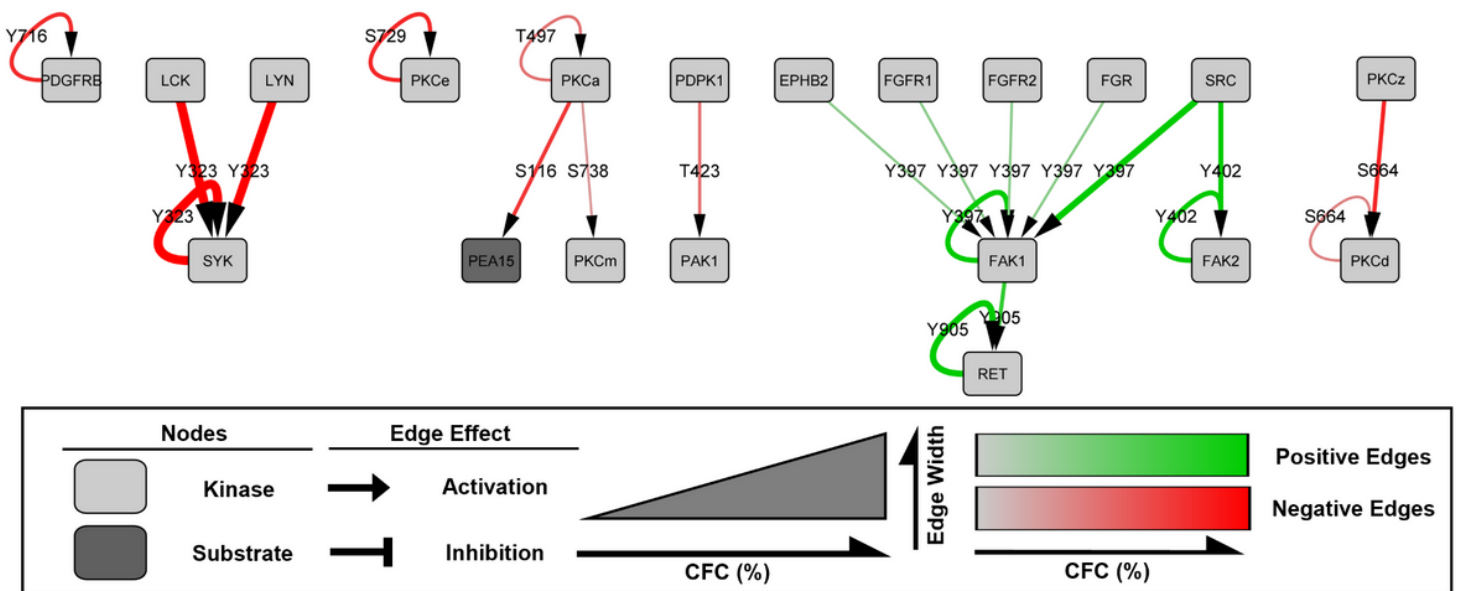

Figure 3

Subnetwork of the MAPPINGS analysis performed on P. falciparum ring stage microarray dataset. The network contains 20 nodes with 21 connecting edges. Kinases and substrates are represented as nodes (dark nodes = substrates, light nodes = kinases) and phosphorylation events are represented as edges, which are designated with the specific phosphorylation site. Edges are represented in a colour gradient from grey to green (positive edges) and grey to red (negative edges) and a size gradient which corresponds to the percentage change from the control network trails (%CFC). The effect of the edge is represented in the arrowhead (arrow = activation, square = inhibition).
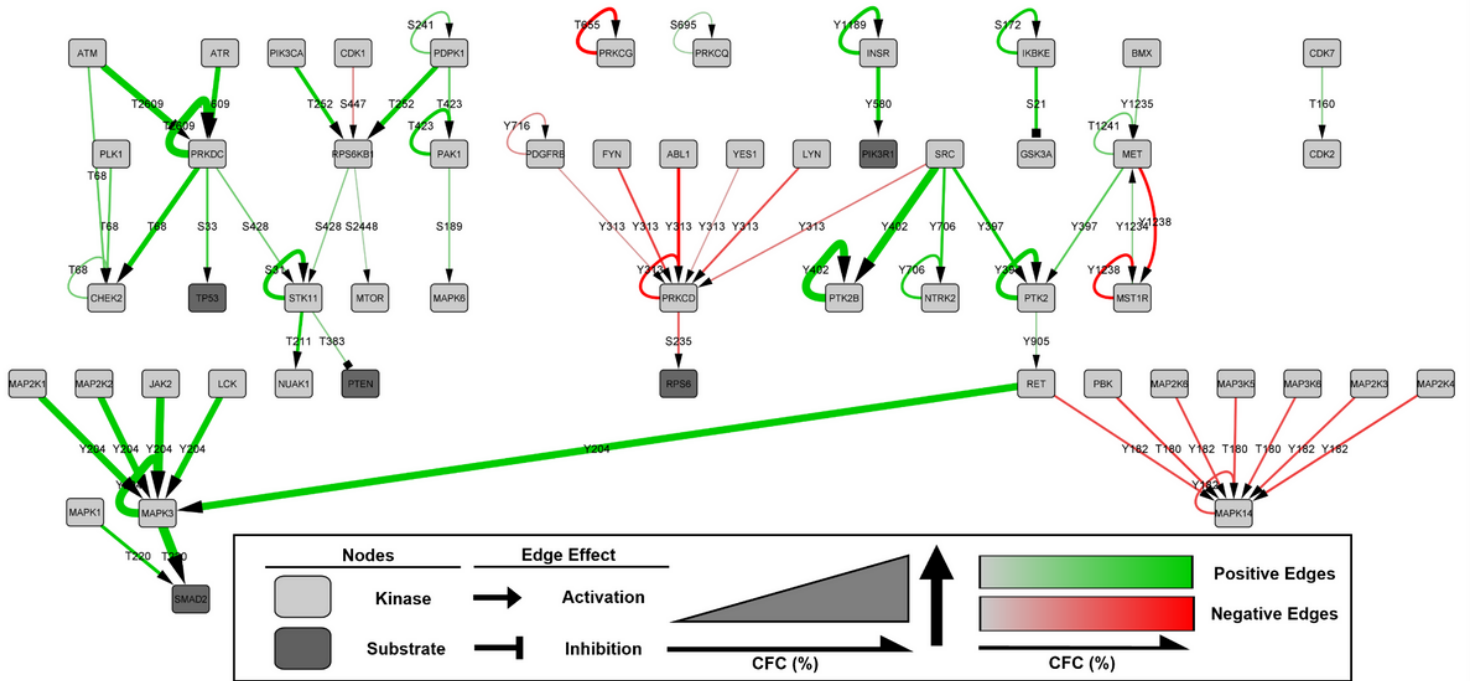
Figure 4

Subnetwork of the MAPPINGS analysis performed on P. falciparum trophozoite stage microarray dataset. The network contains 53 nodes with 66 connecting edges. Kinases and substrates are represented as nodes (dark nodes = substrates, light nodes = kinases) and phosphorylation events are represented as edges, which are designated with the specific phosphorylation site. Edges are represented in a colour gradient from grey to green (positive edges) and grey to red (negative edges) and a size gradient which corresponds to the percentage change from the control network trails (%CFC). The effect of the edge is represented in the arrowhead (arrow = activation, square = inhibition).
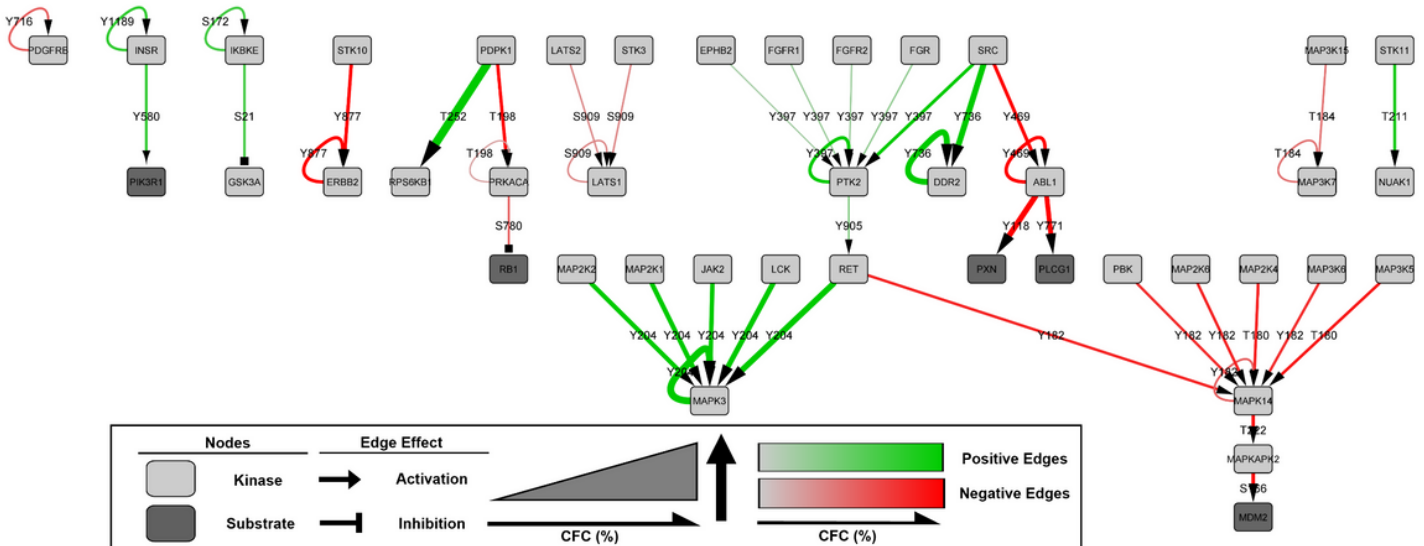
Figure 5

Subnetwork of the MAPPINGS analysis performed on P. falciparum schizont stage infection. The network contains 42 nodes with 45 connecting edges. Kinases and substrates are represented as nodes (dark nodes = substrates, light nodes = kinases) and phosphorylation events are represented as edges, which are designated with the specific phosphorylation site. Edges are represented in a colour gradient from grey to green (positive edges) and grey to red (negative edges) and a size gradient which corresponds to the percentage change from the control network trails (%CFC). The effect of the edge is represented in the arrowhead (arrow = activation, square = inhibition).