




Research article

Gene network analysis of *Plasmodium falciparum* ubiquitin-proteasomal system pathways reveals co-expression of the E1, E2 and E3 enzymes during intraerythrocytic cycle

Lyang Higa Cano^a, Waheed Ahmed^b, Wânia Rezende Lima^c, David Corrêa Martins Jr.^d, Kleber Simônio Parreira^c, Célia R.S. Garcia^{b,*}, Ronaldo Fumio Hashimoto^{a,**} 

^a Universidade de São Paulo, Instituto de Matemática e Estatística, Departamento de Ciência da Computação, São Paulo, 05508-090, São Paulo, Brazil

^b Department of Clinical and Toxicological Analysis, School of Pharmaceutical Sciences, University of São Paulo, São Paulo, 05508-000, São Paulo, Brazil

^c Universidade Federal de Catalão, Goiás, Instituto de Biotecnologia, Departamento de Medicina, Catalão, 75704-020, Goiás, Brazil

^d Center of Mathematics, Computing and Cognition, Federal University of ABC, Santo André, 09606-045, São Paulo, Brazil

ARTICLE INFO

Keywords:

Gene co-expression networks

Malaria

Plasmodium falciparum

Ubiquitin-proteasomal system

E1-E2-E3 matching

ABSTRACT

Plasmodium falciparum is responsible for the malaria disease causing a significant number of deaths around the globe. Integrative biology (genome, transcriptome, cell biology) and host interaction of *Plasmodium* spp have been applied for decades to find solutions to fight against the malaria parasite. As an eukaryotic parasite, *P. falciparum* tightens protein activity with the ubiquitin-proteasome system (UPS) that regulates several crucial processes. In fact, the UPS pathway involves a three-step chain reaction catalyzed by enzymes categorized into three distinct groups: E1, E2, and E3. The identification of triples, composed of one enzyme from each group, that work collaboratively in the same chain reaction during the intraerythrocytic developmental cycle (IDC) in *P. falciparum* is of paramount importance, given the incomplete understanding of this phenomenon. To address this critical problem, we propose a novel approach: a Gene Co-expression Network (GCN) model for the systematic ranking of gene triples (E1, E2, E3). This model offers an innovative means of identifying triples that are most likely to operate collectively within the same biological process. Subsequently, we applied this model to analyze seven temporal RNA-Seq transcriptome datasets each representing distinct experimental conditions and temporal stages during the IDC in *P. falciparum*. Remarkably, our model has pinpointed three triples (E1, E2, E3) that potentially function together during the IDC across all seven datasets. Notably, these triplets consistently exhibit similar gene expression profiles across all seven datasets. This robust consistency underscores their resilience to variations in experimental contexts. Such findings are particularly significant given the marked divergence observed in gene expression profiles across the seven RNA-seq datasets.

* Corresponding author.

** Corresponding author.

E-mail addresses: cgarcia@usp.br (C.R.S. Garcia), ronaldo@ime.usp.br (R.F. Hashimoto).

1. Introduction

Plasmodium parasites are the causative agent of malaria disease, which is still a major public health problem. The World Health Organization estimated 627,000 deaths caused by malaria infection in 2021 [1]. Six species of *Plasmodium* are able to infect humans, among them *Plasmodium falciparum* is responsible for the most severe form of the disease and the most number of deaths caused by malaria [1]. The malaria parasite *P. falciparum* possesses a complex life cycle with sexual development in the mosquito vector *Anopheles* and asexual development in the human host hepatocytes and erythrocytes [2]. Inside the host, the parasite develops synchronously within the parasite population and in coordination with the host rhythms. This synchronization triggers the simultaneous rupture of the infected red blood cells (RBCs), releasing thousands of invasive merozoites and other metabolites produced by parasites [3]. The synchrony of the intraerythrocytic developmental cycle (IDC) is responsible for the clinical symptoms of malaria disease, characterized by fever episodes every 48 h caused by the coordinated rupture of infected RBCs. Intriguingly, this synchrony—a hallmark of in vivo parasite biology—is lost when the parasite is cultured in vitro, indicating that a host cue is essential for the synchronous development of the parasite [3].

Although RTS,S/AS01, a recently approved malaria vaccine, showed some good results in children [4], its efficacy ebbed with time, and so did not fulfill the requirements of the Malaria Vaccine Technology Roadmap [5–7]. Artemisinin in combination with other drugs is used as the first line of defense for controlling malaria. Now it is very alarming that artemisinin-resistant parasites have been reported in Southeast Asian countries [8–10]. In areas showing resistance against associated drug partners, up to about 50 % failure in treatment is now reported, and the increasing resistance of malaria parasites against artemisinins puts additional selection stress on the fewer existing partner drugs [11,12]. In reaction to this imminent crisis, the highest body, MMV (Medicines for Malaria Venture), has emphasized the development of novel antimalarial therapies to diminish the rise of resistance as well as to induce the eradication agenda. Indeed, access to the genome and transcriptome data of the *Plasmodium* ubiquitination proteasome system (UPS) can facilitate the discovery of novel drug targets, as it is considered a high-priority target due to its active role throughout the life cycle [13–18].

Ubiquitin (Ub) is a highly conserved 76-amino-acid protein across eukaryotes. Ubiquitination is the process by which ubiquitin is added to the target proteins to control their cellular levels via proteasome-mediated proteolysis or to regulate their functions through proteasome-independent reactions. Ubiquitin-like proteins (Ubl), another group of proteins structurally related to ubiquitin, share significant similarity to the ubiquitin fold [19]. The ubiquitination cascade begins with activation: Ub/Ubl is activated by E1 (ubiquitin-activating enzyme). This step involves the formation of a thioester bond between the E1 catalytic cysteine residue and Ub/Ubl, accompanied by the release of AMP. Next, in the conjugation phase, the activated Ub/Ubl is transferred via transacylation reaction to E2 (ubiquitin-conjugating enzyme). Finally, ligation is mediated by E3 (ubiquitin ligase), which facilitates the transfer of Ub/Ubl from E2 to the target protein substrate. This step assures the formation of an isopeptide bond between the glycine residue of Ub/Ubl and the lysine residue on the substrate protein. Such reactions are represented in Fig. 1.

E3 ligases are categorized into three major classes—HECT, RBR/Ubox and RING—based on their structural domains and

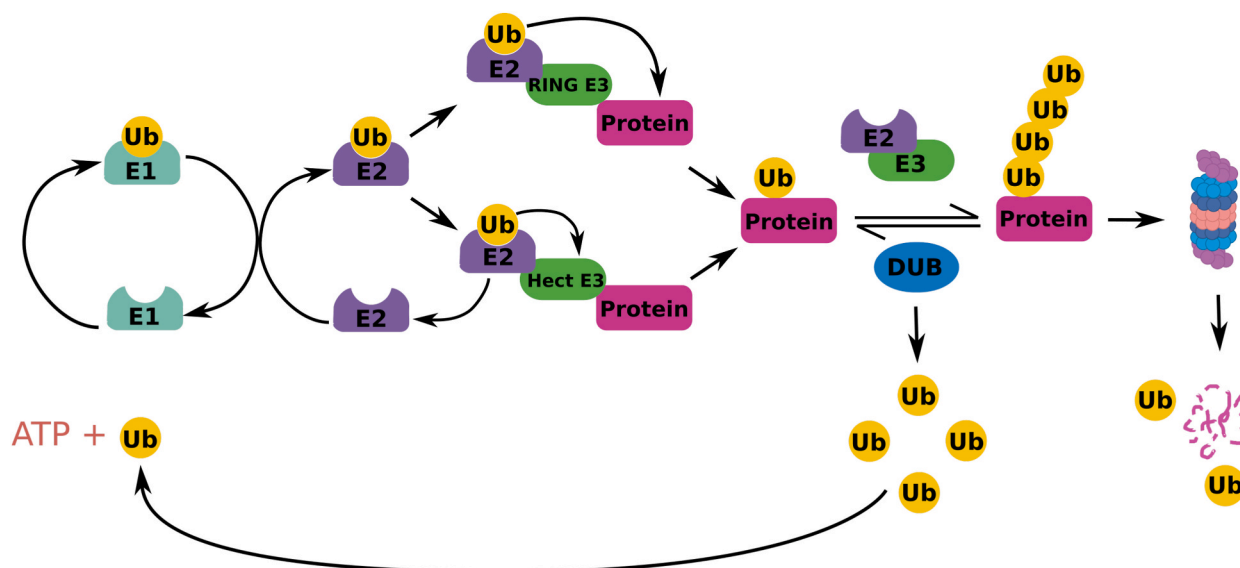


Fig. 1. Schematic diagram of the Ubiquitin Proteasome System (UPS). Players like E1, E2, E3, and DUB catalyze sequential reactions to accomplish ubiquitination. Ubiquitin-activating enzymes (E1) are employed to activate ubiquitin. Then activated ubiquitin is transferred to ubiquitin-conjugating enzymes (E2), and then carried to the target proteins with the assistance of ubiquitin ligases (E3). Two categories of E3 are HECT and RING. Ubiquitin is directly transferred to protein by RING, while in the second case, ubiquitin is first transferred to HECT E3 and then to the target substrate. Repeated cycles of this reaction convert monoubiquitinated protein into polyubiquitinated protein which may act as a signal for protein degradation by the 26S or modulate various functions depending on the position of the lysine residue bonded to the polyubiquitin chain. Deubiquitination enzymes (DUB) play their role in the recycling of ubiquitin. Figure created using BioRender (<https://www.biorender.com/>).

mechanisms of ubiquitin transfer to substrates. In RING and RBR/U-box E3 ligases, ubiquitin is directly transferred to the target substrate. In contrast, HECT E3 ligases first form a thioester intermediate with ubiquitin before transferring it to the target protein [20–22]. RING E3 ligases exist in two forms: single polypeptide chain enzymes and multi-subunit complexes (e.g., Cullin ligases), which utilizes F-box proteins to recognize substrates.

The RING class is the largest among E3 ligases, characterized by a conserved core domain containing histidine and cysteine residues that enhance protein-protein interactions [23]. Two predicted RING domains (RING 1 and RING2) are present in RBR/U-box and are separated by another domain named in-between-RING domain. If only one ubiquitin is attached to the target protein, then it's called monoubiquitination, and if more than one ubiquitin moiety is added in a subsequent way then it is known as polyubiquitination. This monoubiquitinated/polyubiquitinated substrate protein is ultimately degraded by 26S proteasome. E3 ligases play a crucial role in the ubiquitination of specific proteins, facilitating selective degradation. For reversing the effect of E3 ligases, there are specific proteases called deubiquitinating enzymes (DUB). DUB targets the carbonyl group of substrate-Ub isopeptide bond via a nucleophilic attack. DUB facilitates stopping protein degradation or reversal of signaling, or ubiquitin recycling for its homeostasis, as shown in Fig. 1. There are five major classes of DUB; aspartic, serine, metallo, cysteine, and threonine proteases, based on amino acid residues involved in catalytic activity [24]. Any functional abnormality of UPS proteins may result in the deterioration of cellular homeostasis and cause numerous disorders like neurodegenerative diseases, cardiovascular diseases, systematic auto-immunity and malignancies [25–28].

Based on the ubiquitin pathway biological function, we aim to identify triples of genes (E1, E2, E3) that likely collaborate in the same chain reaction during the IDC of *P. falciparum*. More precisely, we need to solve an optimization problem: locating triples of genes (E1, E2, E3) that maximize a specific score, serving as an indicator of their collaborative functionality. By ranking these triples based on this score, we can identify the most promising candidates for further investigation. Our first task is to establish a reliable metric for detecting genes that work together. We operate under the assumption that genes exhibiting co-expression likely participate in the same biological processes [29–31]. In line with this hypothesis, gene expression correlation serves as a suitable metric to recognize genes with collaborative roles.

Our objective in this study is to develop a Gene Co-expression Network (GCN) model to identify specific E1, E2, and E3 genes within the UPS that are likely to collaborate during the IDC of *P. falciparum*. By focusing on these enzymes, which function sequentially in the ubiquitination process, we aim to uncover potential interactions that could be exploited for therapeutic purposes.

While our GCN model is general and adaptable to other pathways and organisms, we concentrate on the UPS of *P. falciparum* in this work due to its biological significance and potential as a drug target. Additionally, we present an example involving tubulin genes to demonstrate that our software predictions are effective and to emphasize the importance of the experimental context from which the RNA-seq data were obtained. Our study is strictly in silico, utilizing publicly available RNA-seq datasets to analyze gene expression patterns. Experimental validation is beyond the scope of this research but it is an important future direction.

2. Material and methods

2.1. Choosing genes of interest

Before applying the model proposed in this work (see Section 2.2), we needed to define a set of genes of interest and gather relevant information already available on them. For our particular case study of the UPS of *P. falciparum*, we focused on the E1, E2, and E3 groups. On October 18, 2021, we accessed PlasmoDB [32] to obtain RNA-seq datasets spanning the IDC of *P. falciparum*. PlasmoDB is a biological/bioinformatics database that provides access to data related to the genome, transcriptome, and proteome of *Plasmodium* species.

To achieve this goal, relevant keywords for UPS components were searched on the PlasmoDB website to retrieve transcriptome datasets containing information about UPS genes being expressed during IDC. PlasmoDB not only helped identify UPS genes but also provided access to information such as their chromosomal location, cellular localization, predicted function, and most importantly expression pattern during IDC. Later, this information was used to classify genes into different groups E1, E2 and E3. We identified 8 genes from E1, 15 from E2 and 54 from E3. This classification, along with the gene expression datasets and all other information collected about these genes, is available in Table SI of Supplementary Material.

2.2. Gene Co-expression network model

Our proposed model is a weighted graph $G = (V, E)$, where V is the set of vertices, each representing a gene of interest. V is composed of disjoint sets, which represent the groups of genes of interest—in our particular case E1, E2 and E3. Thus, we establish $V = E1 \cup E2 \cup E3$; $E1 \cap E2 = \emptyset$; $E1 \cap E3 = \emptyset$ and $E2 \cap E3 = \emptyset$.

The set of edges E is defined as all pairs of vertices from the disjoint sets, meaning $(v_i^x, v_j^y) \in E \forall x \neq y$, where x and y denote the subsets to which v belongs, in our case $E1$, $E2$ and $E3$, and i and j are indices for the elements inside the subsets, i.e., genes from $E1$, $E2$ and $E3$.

For each edge, there is an associated weight w , defined by a function $w: E \rightarrow [-1, 1] \in \mathbb{R}$. Biologically, this weight evaluates the likelihood that the genes are working together in the same chain reaction. Following the main hypothesis of this model, this weight is measured using a correlation metric. In particular, we adopt the Pearson correlation r between each pair of vertices from different groups. This approach allows us to identify combinations with the highest degree of co-expression.

It is essential to note that the correlation between genes from the same group is not calculated; edges exist only between elements from different disjoint sets since we are searching for genes from different groups that might be working in the same UPS pathway. In

fact, our model takes the cartesian product of the groups to consider all possible combinations using one gene from each group, generating a 3-partite graph, as shown in Fig. 2. Importantly, our model operates at the gene expression level, and any inference about protein or enzyme interactions is based on the hypothesis that co-expressed genes likely participate in the same biological processes [29–31].

To use the Pearson correlation r , we assume that the gene expression profiles are approximately normally distributed and that there is a linear relationship between gene expressions [33]. We used the *pearsonr* function from SciPy version 1.13.1 [34] to compute r and its associated p-value. This function calculates the p-value based on the exact distribution of the Pearson correlation coefficient between x and y under the null hypothesis of zero correlation, assuming that the data are normally distributed and that x and y are independent [34]. This method provides an exact p-value under these assumptions, as detailed in the referenced documentation. All data analyses were conducted using Python version 3.10.12. Data manipulation and statistical computations were performed with NumPy version 1.26.4 [35] and Pandas version 2.2.1 [36]. Visualizations were generated using Matplotlib version 3.7.1 [37] and Seaborn version 0.13.1 [38]. The source code is available on [github](https://github.com/LyangHiga/gcn_p_falciparum_heliyon) (https://github.com/LyangHiga/gcn_p_falciparum_heliyon).

Before building the model, we applied some preprocessing steps as detailed in Section 2.3. These operations help to normalize the data, making the gene expression values approximately normally distributed. Therefore, the assumptions underlying the Pearson correlation coefficient and its p-value calculation are reasonably satisfied in our analysis.

For each gene triple, consisting of three gene pairs (E1-E2, E1-E3, and E2-E3), we obtained three p-values. Since these p-values are dependent, we cannot use methods that assume independence, such as Fisher's or Stouffer's methods [39], to combine them into a single p-value for the entire triple. Thus we use an adaptation of Brown's method [40], which is essentially an expansion of Fisher's method that includes a correction factor to handle dependent p-values, allowing us to combine them more accurately [40].

2.2.1. Score

We have defined the correlation between pairs of genes as a measure of genes probably working collaboratively, but we still have to define how to create a ranking to classify the best candidates, the ones with the highest probability of working in the same chain reaction. To do that we just use the sum of the weight of the edges as a score. In other words, each combination, or maximal clique, i.e., a cycle path containing exactly one vertex from each group, in graph G , represents a possible reaction. The paths with the highest sum of correlations are considered to have the biggest chance of being a reaction chain that happens in practice. Such a sum can be expressed as $r(v_i^1, v_j^2) + r(v_j^2, v_k^3) + r(v_k^3, v_i^1)$, where r is the Pearson correlation function and v_i^1 , v_j^2 , v_k^3 are the genes indexed by i , j and k from E1, E2 and E3 groups, respectively. We can convert this score to a probability, a real number in $[0,1]$, by normalizing the score.

2.2.2. Optional parameters

Our model also accepts two optional parameters, t and β . The first one is a threshold corresponding to the Pearson correlation's minimum value, leading to an edge in the graph. For example, using $t = 0$ leads to only edges with positive values while all negative values are discarded. The default value is $t = -1$, where all possible pairs have an edge. In its turn, the β parameter works like in the WGCNA package [41], where all edge values will be powered to this β value, which means $r(v_1, v_2)^\beta$, where r is the Pearson Correlation function, v_1 and v_2 are the expression of two given genes. This parameter is useful because for $\beta > 1$ it creates a kind of penalty for smaller values, since a small number in the range $[0,1]$ will decay more than a greater value in the same range when powered to a $\beta >$

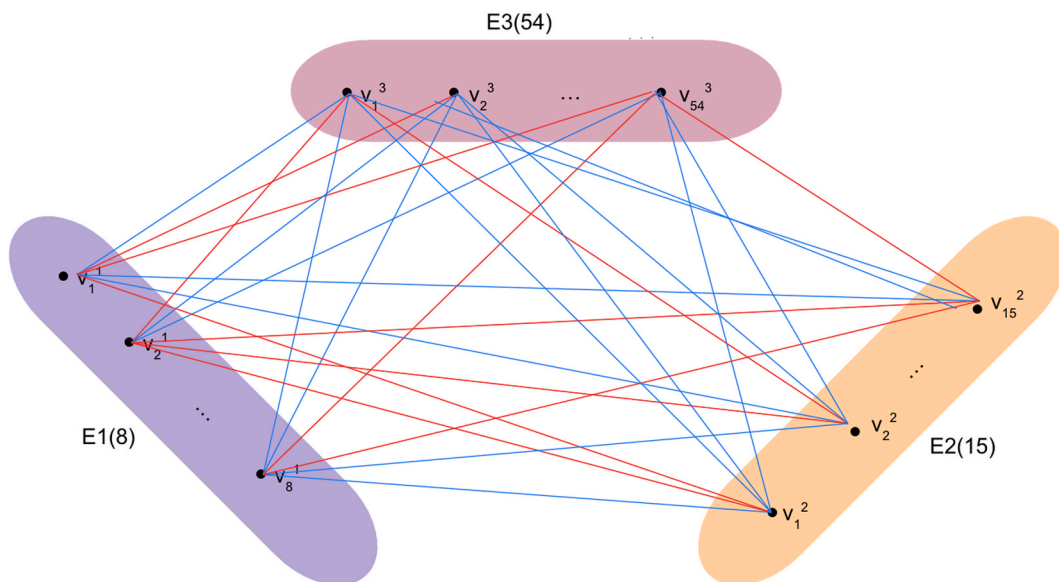


Fig. 2. Illustration of a hypothetical tripartite graph involving E1, E2, and E3 genes derived by the proposed model, with red and blue edges indicating negative and positive correlations, respectively.

1.

2.3. Applying the proposed model for ranking genes of interest

To implement our proposed model, we developed a comprehensive pipeline, as depicted in Fig. 3. To ensure the model's robustness, we evaluated its performance using several distinct RNA-seq datasets obtained from various studies as described in Table 1. All these experiments included at least two biological replicates, and the original authors reported that the high consistency between replicates demonstrates the reproducibility of the RNA-seq data [42–48]. These datasets are accessible in the supplementary materials and encompass the expression profiles of the selected 77 genes during the IDC of *P. falciparum*. The datasets were normalized for sequencing depth by the original authors using the methods specified in Table 1. In our analyses, we used the expression values from the first replicate of each dataset to maintain consistency; however, our pipeline allows users to utilize the average expression across replicates or apply other methods of combining replicate data, depending on their preferences and analytical goals.

Notably, the gene expression values in these datasets vary widely in scale. To facilitate comparative analysis and model building, we performed preprocessing steps. First, we applied a \log_2 transformation to all datasets followed by a standardization by subtracting each gene expression value from the average expression of the corresponding gene and then dividing by its standard deviation (z-score). This preprocessing strategy homogenizes the expression signals onto a common scale, a crucial step for constructing our model. This is particularly valuable as our model relies on calculating pairwise correlations between gene expression values, emphasizing the patterns of expression rather than their absolute magnitudes.

Another important observation to note is the time scale in hours in all datasets, while reactions within the UPS pathway typically occur within the order of seconds [49]. Due to this substantial disparity in temporal scales, it is reasonable to consider the reactions as happening simultaneously. As a consequence, it is plausible that genes from E1 and E3, which work in the same chain reaction, should be highly co-expressed at the same time. We have $|E1| \times |E2| \times |E3| = 8 \times 15 \times 54 = 6480$ combinations, thereby yielding 6480 positions within our ranking, or 6480 possible reactions, that can be generated using a threshold value of $t = -1$, resulting in the creation of all possible edges, as exemplified in Fig. 2. However, our focus is exclusively on capturing positive correlations. As a result, we use $t = 0$ and $\beta = 1$, a selection that exclusively permits positive edges within the network; following this approach in the Broadbent dataset, for example, resulted in 928 triples.

We also subjected these resultant triples to three filtering criteria: (1) a minimum score greater than a threshold value (which varies for each dataset); (2) a p-value smaller than 0.05, to guarantee statistical significance; and (3) a requirement that the genes involved are expressed within the same cellular location. We obtained the cellular location information from PlasmoDB, and it is available in Table SI of the Supplementary Material. The thresholds for both the score and the p-value are parameters that can be adjusted according to user preference, while the requirement for the same cellular location is fixed. We chose to use a p-value cutoff of 0.05, as it is the standard threshold for statistical significance; however, our software allows users to select different p-value thresholds based on their desired level of stringency. Users seeking more stringent results can choose a smaller p-value, whereas those who are more flexible about statistical significance can opt for a larger value. To further refine our results, we chose the minimum score that yields

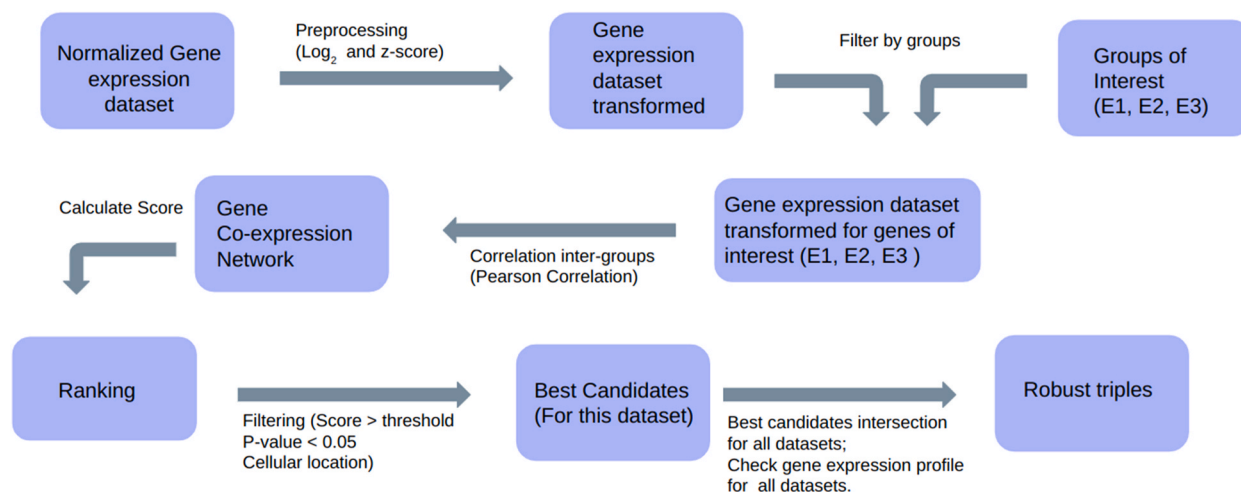


Fig. 3. Pipeline Overview. The values indicated in parentheses correspond to the parameters utilized in this specific study. The pipeline begins by selecting one RNA-seq dataset (from the seven experiments), which was normalized for sequencing depth by the original authors using the methods specified in Table 1. This entire pipeline is performed for all datasets independently. Pre-processing includes a \log_2 transformation followed by a standardization using z-scores. Next, we construct the proposed Gene Co-expression Network (GCN) model targeting the genes of interest (E1, E2, E3). Using the resultant network, we calculate Pearson correlation for each edge and assign these values as edge weights. These edge weights form the foundation for ranking gene triples. To identify the best candidate triples, we apply three filtering criteria. Subsequently, we search for similar triples across the top candidates from all seven datasets and assess whether they exhibit analogous gene expression profiles. Triples that consistently exhibit high scores and stable gene expression patterns across experiments are termed “robust”.

Table 1

This table provides an overview of the RNA-seq datasets, including the name of each dataset, the number of time points, the exact time points (in hours post-infection) used to generate our model, the normalization method, and the corresponding references for each dataset. All datasets included at least two biological replicates, with Wichers et al., 2019 [45] including three replicates.

RNA-seq	Number of time points	Time Points (hpi)	Normalization Method	Ref.
Otto et al., 2010	7	Every 8h from 0h to 48h.	TPM	[42]
Broadbent et al., 2015	9	[6h, 14h, 20h, 24h, 32h, 36h, 40h, 44h, 48h]	FPKM	[43]
Toenhake et al., 2018	8	Every 5h from 5h to 40h	TPM	[44]
Wichers et al., 2019	8	[0h, 8h, 16h, 24h, 32h, 40h, 44h, 48h]	TPM	[45]
Subudhi et al., 2020	25	Every 2 h from 0h to 48h.	TPM	[46]
Chappell et al., 2020	7	Every 8h from 0h to 48h	TPM	[47]
Kucharski et al., 2020	24	[4h, 6h, 8h, 10h, 12h, 14h, 16h, 18h, 22h, 24h, 26h, 28h, 30h, 32h, 34h, 36h, 38h, 40h, 42h, 44h, 46h, 48h, 50h, 52h]	TPM	[48]

the top 2 %–5 % of triples. In Fig. 3, we refer to these as “Best Candidates.” In this way, we obtained the triples with the highest likelihood of working in the same chain reaction, while simultaneously reducing the number of triples from 6480 to a manageable quantity for manual inspection. This reduction is crucial because it allows biologists to focus on the most promising candidates.

3. Results

Following the pipeline from Fig. 3 for all datasets yielded 25 triples for the Broadbent et al., 2015 [43] dataset, 33 triples for the Otto et al., 2010 [42] dataset, 35 triples for the Toenhake et al., 2018 [44] dataset, 32 triples for the Wichers et al., 2019 [45] dataset, 42 triples for the Subudhi et al., 2020 [46] dataset, 34 triples for the Chappell et al., 2020 [47] dataset, and 29 triples for the Kucharski et al., 2020 [48] dataset. The detailed list of these triples and their corresponding minimum scores are available in the Supplementary Material.

From these finalized results for each dataset, we sought to identify similar triples, essentially finding intersections among the results. This comparison revealed only three triples that are consistently present across all datasets, which we refer to in Fig. 3 as “robust triples”. These shared triples are presented in Table 2. Table 3 shows the results of the first robust triple (PF3D7_1333200, PF3D7_1345500, PF3D7_0319100), and Panels A–G of Fig. 4 display their gene expression profiles for each dataset. The results of the other two robust triples are available in Supplementary Material.

Upon examining the final outcomes for each dataset, a noticeable trend emerges: a significant proportion of genes from E1, E2, and E3 do not exhibit similar gene expression profiles across the various RNA-seq experiments. Wichers et al., 2019 [45], in their investigation of the *stevor* gene family, also observed a substantial variability in gene expression patterns across different experiments. The underlying cause of this phenomenon is attributed to various confounding experimental factors, including disparities in experimental designs (such as replicates, read numbers, sequencing length, and time points of harvest), distinct culture conditions (such as variations in serum versus AlbuMAX supplementation, gas mixtures, passage numbers, and synchronization methods), and accumulated genetic variations across different cell lines [45].

The considerable divergence in gene expression patterns among different experiments significantly impacts our study, as these diverse profiles serve as the input for our model. As a result, gene triples identified as potential components of the same reaction in one RNA-seq experiment may receive a score indicative of non-cooperativity in another. This inconsistency limits our ability to make reliable predictions that are universally applicable across different datasets.

To illustrate this variability, we selected the best triple—PF3D7_1225800 (UBA1), PF3D7_1033900 (2ONU), PF3D7_0826500 (UFD2)—from the Broadbent et al., 2015 [43] dataset and compared the results of this same triple with the other six datasets, as shown in Table 4. The observed disparities in results among these datasets are substantial, highlighting the challenge of drawing consistent conclusions from heterogeneous data sources.

Analysis of these results reveals that the top candidate from the Broadbent et al., 2015 dataset performed poorly in all other datasets. This discrepancy suggests that these genes may not collaborate within the same chain reaction in other experimental contexts.

Table 2

Triples (E1, E2, E3) present in the final results of all datasets. Scores, p-values and gene expression profiles for each dataset are available in Supplementary Material.

E1	E2	E3
PF3D7_1333200	PF3D7_1345500	PF3D7_0319100
PF3D7_1333200	PF3D7_1345500	PF3D7_1210900
PF3D7_1333200	PF3D7_1345500	PF3D7_0303800

Table 3
Results of the triple (PF3D7_1333200, PF3D7_1345500, PF3D7_0319100), the first robust triple, for all datasets.

Dataset	Score	Normalized Score	P-value
Broadbent et al., 2015 [43]	2.302	0.767	0.001375
Otto et al., 2010 [42]	2.743	0.914	0.000329
Toenhake et al., 2018 [44]	2.613	0.871	0.000270
Wichers et al., 2019 [45]	2.956	0.985	8.18e-08
Subudhi et al., 2020 [46]	2.263	0.754	5.75e-08
Chappell et al., 2020 [47]	2.371	0.790	0.008552
Kucharski et al., 2020 [48]	2.772	0.924	2.13e-14

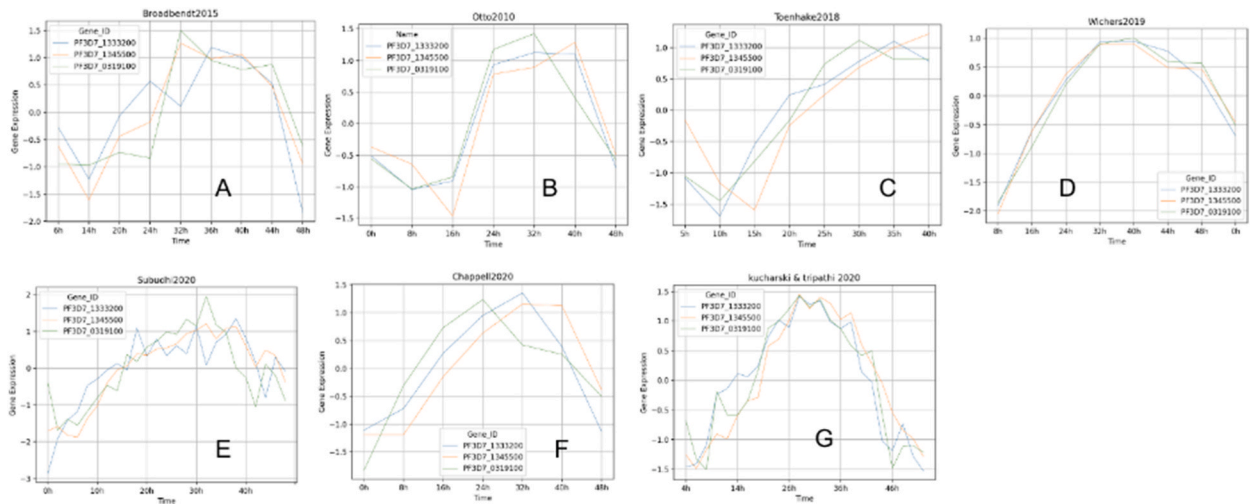


Fig. 4. Gene expression profiles of the triple (PF3D7_1333200, PF3D7_1345500, PF3D7_0319100) for all datasets. A: Broadbent et al., 2015; B: Otto et al., 2010; C: Toenhake et al., 2018; D: Wichers et al., 2019; E: Subudhi et al., 2020; F: Chappell et al., 2020; G: Kucharski et al., 2020.

Examining their gene expression profiles across each dataset, as illustrated in Figs. S3 and S4 in the Supplementary Material, reveals substantial variations.

These findings underscore a crucial point: while our model effectively predicts potential gene interactions, these predictions are intricately tied to the context of individual RNA-seq experiments. The inherent variability in gene expression profiles results in distinct outcomes across experiments, necessitating caution when generalizing predictions to different experimental conditions.

To determine if this phenomenon is unique to UPS-related genes, we investigated PF3D7_0903700 (Alpha tubulin 1) and PF3D7_1008700 (Tubulin beta chain)—two genes widely acknowledged to operate collaboratively within the Tubulin complex. Surprisingly, our investigation revealed that despite achieving high scores in six out of the seven datasets, as shown in Table 5, these well-known gene pairs still exhibit distinct gene expression profiles across each dataset. For instance, around the 20-h post-infection (20 hpi) timeframe, while certain experiments demonstrate elevated expression levels for both genes, others show a low expression pattern. This variability is illustrated in Panels A and B of Fig. 5: Panel A displays low expression of both genes at 20 hpi, and Panel B shows a case where both are highly expressed. The gene expression profiles for all datasets are provided in Fig. S5 (Supplementary Material).

These discrepancies imply that our method can identify correlated genes within a dataset, the variability across datasets hinders the generalization of predictions for gene expression profiles at specific times to different experimental conditions. This observation

Table 4
Scores, normalized scores, and p-values for the triple (PF3D7_1225800, PF3D7_1033900, PF3D7_0826500). This triple was identified as the top candidate in Broadbent et al., 2015 but showed varying results in other datasets.

Dataset	Score	Normalized Score	P-value
Broadbent et al., 2015	2.514	0.838	0.0001
Otto et al., 2010	0.574	0.191	0.22
Toenhake et al., 2018	1.086	0.362	0.01
Wichers et al., 2019	1.461	0.487	0.018
Subudhi et al., 2020	0.766	0.255	0.000017
Chappell et al., 2020	1.279	0.426	0.052
Kucharski et al., 2020	1.535	0.512	1.584e-07

Table 5

Scores and p-values for the pair PF3D7_0903700 (Alpha tubulin 1) and PF3D7_1008700 (Tubulin beta chain). The results show a high score for six out of the seven datasets.

Dataset	Score(Range [0; 1])	P-value
Broadbent et al., 2015	0.996	1.6e-5
Otto et al., 2010	0.960	0.012
Toenhake et al., 2018	0.996	6e-5
Wichers et al., 2019	0.997	5.4e-5
Subudhi et al., 2020	0.992	1.3e-13
Chappell et al., 2020	0.729	0.246
Kucharski et al., 2020	0.987	1.3e-11

underscores the need for caution when comparing results across different experimental contexts, emphasizing that while predictions within a specific experiment context may hold, establishing universal rules is considerably more intricate.

While acknowledging this observed dependence on the experiment inputs, our investigation led us to identify triples that consistently ranked well across all seven datasets. Remarkably, only three triples emerged as common denominators in the final results of all seven datasets, as detailed in Table 2. Upon analyzing their gene expression profiles across the diverse experiments, we observed a new trend. Despite significant inconsistencies and variability in gene expression patterns among the different RNA-seq experiments, these triples not only achieved high scores across all datasets, but also exhibited similar gene expression patterns. This remarkable consistency implies that these genes are robust, consistently displaying a congruent result and gene expression pattern regardless of the specific RNA-seq experiment analyzed, as demonstrated by the uniformly high scores for the first robust triple in Table 3 and the compatible expression profiles across all datasets shown in Panels A–G of Fig. 4. This finding underscores the impact of dataset inconsistency on our results, as only a few gene triples remain consistently correlated and exhibit similar expression patterns across different experimental conditions, making them prime candidates for further study. In contrast, as demonstrated by the tubulin example, some gene pairs may be highly correlated within datasets but do not show consistent expression patterns across different experiments, limiting their predictive utility.

4. Discussion

In the course of this research, the proposed model identified 25 triples of genes on the Broadbent et al., 2015 dataset, available in the Supplementary Material. Our observations revealed that there is considerable diversity among the E1, E2, and E3 genes within these triples. Specifically, we identified 3 different E1 genes, 9 different E2, and 9 different E3, among these 25 triples. It is widely recognized in eukaryotes that, despite the abundance of E2, they are less diverse in comparison to E3, which exhibit a greater diversity of members [50]. Within the three robust triples, the E1 (PF3D7_1333200/PF13_0182) and the E2 (PF3D7_1345500) remained the same. However, the E3 varied: in the first triple, we identified the E3 ubiquitin-protein ligase RBX1, (PF3D7_0319100); in the second triple, the GPI mannosyltransferase 1, (PF3D7_1210900); and in the third triple, the IBR domain protein, putative, (PF3D7_0303800), as shown in Table 2.

As the molecular and functional characterization of E1, E2, and E3 ubiquitin enzymes in *P. falciparum* remains limited, there is a scarcity of information regarding their activities and interactions with partner molecules. In light of this, we focused on exploring the potential functional activities within the highest-scoring robust triple identified in this study. The *Pfsuba1* protein, encoded by the

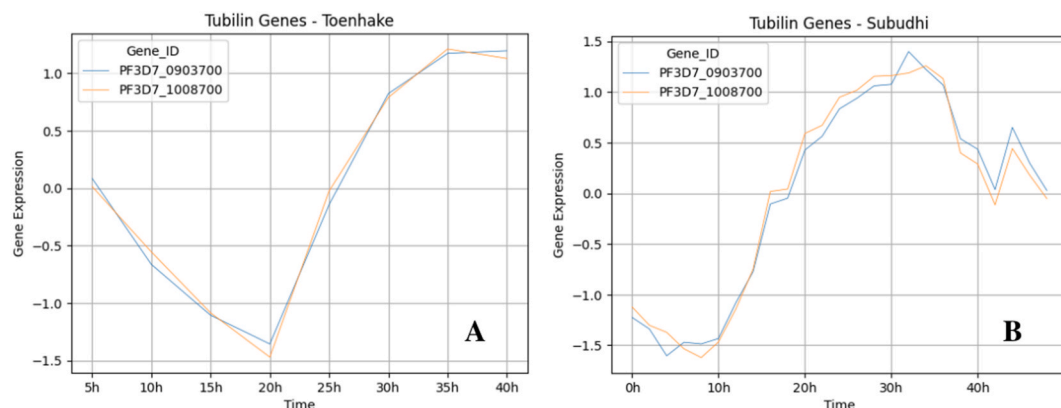


Fig. 5. Gene expression profiles of PF3D7_0903700 (alpha-tubulin 1) and PF3D7_1008700 (tubulin beta chain). A: Expression profiles in Toenhake et al., 2018 dataset. Both genes show low expression at 20 hpi; B: Expression profile in Subudhi et al., 2020 dataset. Both genes exhibit high expression at 20 hpi. The comparison highlights that while the two genes share closely related expression profiles within the same experiment, their profiles diverge significantly between experiments.

PF3D7_1225800 gene, is a ubiquitin-activating enzyme involved in the ubiquitin pathway in *P. falciparum* and is one of the best-characterized members of the E1 superfamily in this species [51]. On the other hand, the uba1 protein encoded by the PF3D7_1333200 gene, which we identified in our research, exhibits the highest score across the analysis of seven different transcript datasets but remains relatively unexplored. Sequence alignment analysis revealed low sequence similarity between the amino acid sequences of the two uba1 proteins, XP_001350655.1 and XP_001350063.1, encoded respectively by the PF3D7_1225800 and PF3D7_1333200 genes (Fig. S6). These enzymes share common features and mechanisms of action, such as the adenylation of ubiquitin and the formation of a thioester bond with the activated ubiquitin molecule [52]. In our research, we identified two conserved ThiF domains in uba1/PF3D7_1333200, which were also detected in *Pfsuba1* according to the Pfam analysis available in the Swiss Browser (Fig. S7). The ThiF domain is a NAD/FAD-binding fold found in ubiquitin-activating E1 family and members of the bacterial ThiF/MoeB/HesA family [53–55]. Indeed, *Pfsuba1* features a ubiquitin-activating enzyme active site, two ubiquitin-activating enzyme catalytic domains, two ThiF repeats, and a catalytic cysteine at the N-terminus [51,56].

The ubiquitin-conjugating enzyme found in our analyses is encoded by the gene PF3D7_1345500, also known as UBC or 2H2Y. This *ubc* protein has been related to the family of specific ERAD-like proteins in the parasite, just like the E1 enzyme [51,56]. In addition to containing the necessary ubiquitination domains, ERAD-like proteins contain a signal peptide that is responsible for directing the proteins to the apicoplast. The apicoplast is a non-photosynthetic plastid that resulted from a secondary endosymbiosis event involving a red alga that created a plastid attached to four membranes. This organelle is responsible for fatty acid metabolism and isoprenoid biosynthesis in *Apicomplexa* such as *Plasmodium* and *Toxoplasma* [56–58]. In addition to the E2 enzyme targeting the apicoplast, once the ubiquitin-conjugated E2 binds to the E3 ubiquitin ligase it plays a role in the UPS pathway as well. Likewise, in the course of High-Throughput Mass Spectrometry research on *Apicomplexan* parasites, proteins associated with a parasite-specific ERAD-like system is comprised by all the essential ubiquitination enzymes necessary for activation (PF3D7_1333200 and PF3D7_1365400), conjugation (PF3D7_1345500), binding (PF3D7_0316900 and PF3D7_0312100), and deconjugation (PF3D7_1031400) [56,58].

E3s play a crucial role in the UPS as they ensure a high level of specificity and selectivity when targeting substrates within cells [59]. In the *P. falciparum* 3D7 strain, the PF3D7_0319100 transcript encodes a putative E3 ubiquitin-protein ligase RBX1, characterized by its catalytic RING domain. In higher eukaryotes, the E3 RBX1 is part of a tetrameric E3 ubiquitin ligase complex known as SCF (Skp1-Cullin1-F-box protein) which comprises four key components: RBX1 or RBX2, SKP1, an F-box protein, and a cullin [49]. Recent evidence has shed light on the functional characterization of two distinct types of E3 ubiquitin ligases in *P. falciparum*: the typical SCF-like E3 ubiquitin ligase, known as PfSCF, and a human CRL4-like E3 ubiquitin ligase, referred to as PfCRL4 [60]. PfSCF is believed to consist of the following components based on in vitro immunoprecipitation and ubiquitination assays: PfCullin-1, PfSkp1, PfRbx1, PfFBXO1, and PfCacyBP. On the other hand, PfCRL4 is composed of PfCullin-2, PfCPSF_A, two WD40 repeat proteins, and PfRbx1. PfCRL4 has been found to play a significant role in cell division and membrane integrity, and it has been suggested to be essential for the proper functioning of mitochondria, the endoplasmic reticulum (ER), and genomic maintenance. However, the role of the E3 ubiquitin-protein ligase RBX1 remains relatively obscure. In a recent study conducted by Rizvi and colleagues [60], evidence emerges of the interaction between the cullin protein and PfRbx1 (PF3D7_0319100) in *P. falciparum*. The findings from this research suggest that PfRbx1 is expressed in multiple cellular compartments, including the cytoplasm, nucleus, and chromatin, in the D10 strain of *P. falciparum* during the trophozoite stage. The result predicted, as illustrated in Fig. 4A–G, showed the expression of E1, E2, and E3 transcripts increases around the 16-h mark of the intraerythrocytic cycle and remains elevated until approximately 46 h. However, there is some variability depending on the specific transcriptome dataset considered. Despite evidence of PfRbx1 expression in trophozoites, we can suggest that transcript expression is particularly high during the early and late trophozoite stages, as well as during the schizont stage.

In mammals, it is well-established that this enzyme is expressed in various tissues and plays a crucial role in cell survival and division. Notably, RBX1 and RBX2 have been extensively studied in the context of anti-cancer therapy. Inhibiting these proteins has been shown to induce apoptosis and cell senescence, while their overexpression is directly associated with the proliferation of tumor cells [61,62].

While the prediction in this analysis identified only the RBX1 transcript, it is reasonable to speculate that enzymes such as E1, E2, and particularly E3 ligase also play pivotal roles in the parasite's survival by regulating protein activity through the UPS pathway. Given the potential of the E3 ligase enzyme RBX1 as a therapeutic target, it becomes relevant to undertake the characterization of this protein in *P. falciparum*. Additionally, the development of inhibitors holds promise for studying its function during the parasite's intraerythrocytic cycle.

5. Conclusion

The model we have developed serves as a predictive tool for identifying potential gene interactions; however, these predictions are constrained to the context of individual RNA-seq experiments. Nonetheless, it is worth noting that certain genes appear to exhibit robustness in this regard. These genes attain high scores and consistently display similar gene expression patterns across various experiments, regardless of the experimental conditions. In our specific study involving genes from E1, E2, and E3 of *Plasmodium falciparum*, we identified three triples, as shown in Table 2, that demonstrate robustness across all seven RNA-seq experiments used in this study. In addition, this research shows that the E1, E2 and E3 enzymes must participate in ubiquitination to promote parasite development and replication during the trophozoite and schizont stages in the parasite's IDC.

CRediT authorship contribution statement

Lyang Higa Cano: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Conceptualization. **Waheed Ahmed:** Writing – review & editing, Writing – original draft, Methodology, Data curation, Conceptualization. **Wânia Rezende Lima:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Conceptualization. **David Corrêa Martins:** Writing – review & editing, Writing – original draft, Supervision, Conceptualization. **Kleber Simônio Parreira:** Writing – review & editing, Writing – original draft, Data curation, Conceptualization, Methodology. **Célia R.S. Garcia:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Funding acquisition, Conceptualization. **Ronaldo Fumio Hashimoto:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Methodology, Funding acquisition, Formal analysis, Conceptualization.

Data and code availability

The source code used in this study has been deposited in the GitHub repository at: https://github.com/LyangHiga/gcn_p_falciparum_heliyon.

All datasets supporting the findings of this study are included in the Supplementary Material.

Funding

We thank Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil (CAPES) - Finance Code 001. We also thank São Paulo Research Foundation (FAPESP) for funding CRSG (Proc. 2017/08684-7, 2018/07177-7 and 2023/07656-0), RFH (Proc. 2015/22308-2 and 2023/07656-0), and DCM-jr (Proc. 2018/18560-6 and 2018/21934-5). This work was also supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) fellowship no. 88887.639065/2021-00.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors acknowledge the National Laboratory for Scientific Computing (LNCC/IME, Brazil) for providing HPC resources of the SDumont supercomputer, which have contributed to the research results reported within this paper. URL: <http://sdumont.lncc.br>.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.heliyon.2025.e44019>.

References

- [1] World Health Organization, World Malaria Report 2022, 2022.
- [2] A.F. Cowman, C.J. Tonkin, W.H. Tham, M.T. Duraisingh, The molecular basis of erythrocyte invasion by malaria parasites, *Cell Host Microbe* 22 (2017) 232–245, <https://doi.org/10.1016/j.chom.2017.07.003>.
- [3] M.K. Singh, B.K.d.M. Dias, C.R.S. Garcia, Role of melatonin in the synchronization of asexual forms in the parasite *Plasmodium falciparum*, *Biomolecules* 10 (2020) 1243, <https://doi.org/10.3390/biom10091243>. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7563138/>.
- [4] RTS,S Clinical Trials Partnership, Efficacy and Safety of RTS,S/AS01 Malaria Vaccine with or Without a Booster Dose in Infants and Children in Africa: Final Results of a Phase 3, Individually Randomised, Controlled Trial, 2015.
- [5] R. Gosling, L. von Seidlein, The future of the RTS,S/AS01 malaria vaccine: an alternative development plan, *PLoS Med.* 13 (2016) e1001994, <https://doi.org/10.1371/journal.pmed.1001994>.
- [6] K. Matuschewski, Vaccines against malaria-still a long way to go, *FEBS J.* 284 (2017) 2560–2568, <https://doi.org/10.1111/febs.14107>.
- [7] L. von Seidlein, The advanced development pathway of the RTS,S/AS01 vaccine, *Methods Mol. Biol.* 2013 (2019) 177–187, https://doi.org/10.1007/978-1-4939-9550-9_13.
- [8] H. Noedl, Y. Se, K. Schaefer, B.L. Smith, D. Socheat, M.M. Fukuda, Evidence of artemisinin-resistant malaria in Western Cambodia, *N. Engl. J. Med.* 359 (2008) 2619–2620, <https://doi.org/10.1056/NEJMc0805011>.
- [9] W.L. Hamilton, R. Amato, R.W. van der Pluijm, C.G. Jacob, H.H. Quang, N.T. Thuy-Nhien, T.T. Hien, B. Hongvanthong, K. Chindavongsa, M. Mayxay, R. Huy, R. Leang, C. Huch, L. Dysoley, C. Amaratunga, S. Suon, R.M. Fairhurst, R. Tripura, T.J. Peto, Y. Sovann, P. Jittamala, B. Hanboonkunupakarn, S. Pukrittayakamee, N.H. Chau, M. Imwong, M. Dhorda, R. Vongprommek, X.H.S. Chan, R.J. Maude, R.D. Pearson, T. Nguyen, K. Rockett, E. Drury, S. Gonçalves, N.J. White, N.P. Day, D.P. Kwiatkowski, A.M. Dondorp, O. Miotto, Evolution and expansion of multidrug-resistant malaria in southeast Asia: a genomic epidemiology study, *Lancet Infect. Dis.* 19 (2019) 943–951, [https://doi.org/10.1016/S1473-3099\(19\)30392-5](https://doi.org/10.1016/S1473-3099(19)30392-5).
- [10] R.W. van der Pluijm, M. Imwong, N.H. Chau, N.T. Hoa, N.T. Thuy-Nhien, N.V. Thanh, P. Jittamala, B. Hanboonkunupakarn, K. Chutasmit, C. Saelow, R. Runjarern, W. Kaewmok, R. Tripura, T.J. Peto, S. Yok, S. Suon, S. Sreng, S. Mao, S. Oun, S. Yen, C. Amaratunga, D. Lek, R. Huy, M. Dhorda, K. Chotivanich, E. A. Ashley, M. Mukaka, N. Waithira, P.Y. Cheah, R.J. Maude, R. Amato, R.D. Pearson, S. Gonçalves, C.G. Jacob, W.L. Hamilton, R.M. Fairhurst, J. Tarning, M. Winterberg, D.P. Kwiatkowski, S. Pukrittayakamee, T.T. Hien, N.P. Day, O. Miotto, N.J. White, A.M. Dondorp, Determinants of dihydroartemisinin-

- piperazine treatment failure in *Plasmodium falciparum* malaria in Cambodia, Thailand, and Vietnam: a prospective clinical, pharmacological, and genetic study, *Lancet Infect. Dis.* 19 (2019) 952–961, [https://doi.org/10.1016/S1473-3099\(19\)30391-3](https://doi.org/10.1016/S1473-3099(19)30391-3).
- [11] M.D. Spring, J.T. Lin, J.E. Manning, P. Vanachayangkul, S. Somethy, R. Bun, Y. Se, S. Chann, M. Ittiverakul, P. Sia-ngam, W. Kuntawunginn, M. Arsanok, N. Buathong, S. Chaorattanakawee, P. Gosi, W. Taaksorn, N. Chanarat, S. Sundrakes, N. Kong, T.K. Heng, S. Nou, P. Teja-isavadharm, S. Pichyangkul, S. T. Phann, S. Balasubramanian, J.J. Juliano, S.R. Meshnick, C.M. Chour, S. Prom, C.A. Lanteri, C. Lon, D.L. Saunders, Dihydroartemisinin-piperazine failure associated with a triple mutant including kelch13 C580Y in Cambodia: an observational cohort study, *Lancet Infect. Dis.* 15 (2015) 683–691, [https://doi.org/10.1016/S1473-3099\(15\)70049-6](https://doi.org/10.1016/S1473-3099(15)70049-6). [https://www.thelancet.com/journals/laninf/article/PIIS1473-3099\(15\)70049-6/fulltext](https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(15)70049-6/fulltext). publisher: Elsevier. Lancet (London, England) 386, 31–45. doi:10.1016/S0140-6736(15) 60721-8.
 - [12] C. Amaratunga, P. Lim, S. Suon, S. Sreng, S. Mao, C. Sopha, B. Sam, D. Dek, V. Try, R. Amato, D. Blessborn, L. Song, G.S. Tullo, M.P. Fay, J.M. Anderson, J. Tarning, R.M. Fairhurst, Dihydroartemisinin-piperazine resistance in *Plasmodium falciparum* malaria in Cambodia: a multisite prospective cohort study, *Lancet Infect. Dis.* 16 (2016) 357–365, [https://doi.org/10.1016/S1473-3099\(15\)00487-9](https://doi.org/10.1016/S1473-3099(15)00487-9).
 - [13] F.C. Koyama, D. Chakrabarti, C.R.S. Garcia, Molecular machinery of signal transduction and cell cycle regulation in *Plasmodium*, *Mol. Biochem. Parasitol.* 165 (2009) 1–7, <https://doi.org/10.1016/j.molbiopara.2009.01.003>.
 - [14] M.N. Aminake, H.D. Arndt, G. Pradel, The proteasome of malaria parasites: a multi-stage drug target for chemotherapeutic intervention? *International Journal for Parasitology. Drugs and Drug Resistance* 2 (2012) 1–10, <https://doi.org/10.1016/j.ijpddr.2011.12.001>.
 - [15] F.C. Koyama, R.Y. Ribeiro, J.L. Garcia, M.F. Azevedo, D. Chakrabarti, C.R.S. Garcia, Ubiquitin proteasome system and the atypical kinase PfPK7 are involved in melatonin signaling in *Plasmodium falciparum*, *J. Pineal Res.* 53 (2012) 147–153, <https://doi.org/10.1111/j.1600-079X.2012.00981.x>.
 - [16] F.C. Koyama, T.L.G. Carvalho, E. Alves, H.B. da Silva, M.F. de Azevedo, A.S. Hemery, C.R.S. Garcia, The structurally related auxin and melatonin tryptophan-derivatives and their roles in *Arabidopsis thaliana* and in the human malaria parasite *Plasmodium falciparum*, *J. Eukaryot. Microbiol.* 60 (2013) 646–651, <https://doi.org/10.1111/jeu.12080>.
 - [17] F.C. Koyama, M.F. Azevedo, A. Budu, D. Chakrabarti, C.R.S. Garcia, Melatonin-induced temporal up-regulation of gene expression related to ubiquitin/proteasome system (UPS) in the human malaria parasite *Plasmodium falciparum*, *Int. J. Mol. Sci.* 15 (2014) 22320–22330, <https://doi.org/10.3390/ijms151222320>.
 - [18] P.H.S. Pereira, C. Curra, C.R.S. Garcia, Ubiquitin proteasome system as a potential drug target for malaria, *Curr. Top. Med. Chem.* 18 (2018) 315–320, <https://doi.org/10.2174/1568026618666180427145308>.
 - [19] M. Hochstrasser, Evolution and function of ubiquitin-like protein-conjugation systems, *Nat. Cell Biol.* 2 (2000) E153–E157, <https://doi.org/10.1038/35019643>. URL: <https://www.nature.com/articles/ncb0800.E153>. number: 8 Publisher: Nature Publishing Group.
 - [20] J.M. Hübregtse, M. Scheffner, S. Beaudenon, P.M. Howley, A family of proteins structurally and functionally related to the E6-AP ubiquitinprotein ligase, *Proc. Natl. Acad. Sci. U. S. A.* 92 (1995) 2563–2567. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC42258/>.
 - [21] S. Hatakeyama, M. Yada, M. Matsumoto, N. Ishida, K.I. Nakayama, U box proteins as a new family of ubiquitin-protein ligases, *J. Biol. Chem.* 276 (2001) 33111–33120, <https://doi.org/10.1074/jbc.M102755200>.
 - [22] R. Budhidarmo, Y. Nakatani, C.L. Day, RINGs hold the key to ubiquitin transfer, *Trends Biochem. Sci.* 37 (2012) 58–65, <https://doi.org/10.1016/j.tibs.2011.11.001>.
 - [23] A.J. Saurin, K.L. Borden, M.N. Boddy, P.S. Freemont, Does this have a familiar RING? *Trends Biochem. Sci.* 21 (1996) 208–214.
 - [24] S.M.B. Nijman, M.P.A. Luna-Vargas, A. Velds, T.R. Brummelkamp, A.M.G. Dirac, T.K. Sixma, R. Bernards, A genomic and functional inventory of deubiquitinating enzymes, *Cell* 123 (2005) 773–786, <https://doi.org/10.1016/j.cell.2005.11.007>.
 - [25] J. Wang, M.A. Maldonado, The ubiquitin-proteasome system and its role in inflammatory and autoimmune diseases, *Cell. Mol. Immunol.* 3 (2006) 255–261.
 - [26] M. Frezza, S. Schmitt, Q.P. Dou, Targeting the ubiquitin-proteasome pathway: an emerging concept in cancer therapy, *Curr. Top. Med. Chem.* 11 (2011) 2888–2905, <https://doi.org/10.2174/156802611798281311>.
 - [27] S.R. Powell, J. Herrmann, A. Lerman, C. Patterson, X. Wang, The ubiquitin–proteasome system and cardiovascular disease, *Progress in molecular biology and translational science* 109 (2012) 295–346, <https://doi.org/10.1016/B978-0-12-397863-9.00009-2>. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3743449/>.
 - [28] N.P. Dantuma, L.C. Bott, The ubiquitin-proteasome system in neurodegenerative diseases: precipitating factor, yet part of the solution, *Front. Mol. Neurosci.* 7 (2014). URL: <https://www.frontiersin.org/articles/10.3389/fnmol.2014.00070>.
 - [29] E.J. Williams, D.J. Bowles, Coexpression of neighboring genes in the genome of *Arabidopsis thaliana*, *Genome Res.* 14 (2004) 1060–1067, <https://doi.org/10.1101/gr.2131104>. URL: <http://genome.cshlp.org/lookup/doi/10.1101/gr.2131104>.
 - [30] E.A.R. Serin, H. Nijveen, H.W.M. Hilhorst, W. Ligterink, Learning from Co-expression networks: possibilities and challenges, *Front. Plant Sci.* 7 (2016), <https://doi.org/10.3389/fpls.2016.00444>. URL: <http://journals.frontiersin.org/Article/10.3389/fpls.2016.00444>.
 - [31] W. Yin, L. Mendoza, J. Monzon-Sandoval, A.O. Urrutia, H. Gutierrez, Emergence of co-expression in gene regulatory networks, *PLoS One* 16 (2021) e0247671, <https://doi.org/10.1371/journal.pone.0247671>. URL: <https://dx.plos.org/10.1371/journal.pone.0247671>.
 - [32] C. Aurecochea, J. Brestelli, B.P. Brunk, J. Dommer, S. Fischer, B. Gajria, X. Gao, A. Gingle, G. Grant, O.S. Harb, M. Heiges, F. Innamorato, J. Iodice, J. C. Kissinger, E. Kraemer, W. Li, J.A. Miller, V. Nayak, C. Pennington, D.F. Pinney, D.S. Roos, C. Ross, Jr. C.J. Stoeckert, C. Treatman, H. Wang, PlasmoDB: a functional genomic database for malaria parasites, *Nucleic Acids Res.* 37 (2009) D539–D543, <https://doi.org/10.1093/nar/gkn814>, doi:10.1093/nar/gkn814.
 - [33] L.L. Havlicek, N.L. Peterson, Robustness of the Pearson correlation against violations of assumptions, *Percept. Mot. Skills* 43 (1976) 1319–1334, <https://doi.org/10.2466/pms.1976.43.3f.1319>. URL: <http://journals.sagepub.com/doi/10.2466/pms.1976.43.3f.1319>.
 - [34] P. Virtanen, R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S.J. van der Walt, M. Brett, J. Wilson, K.J. Millman, N. Mayorov, A.R.J. Nelson, E. Jones, R. Kern, E. Larson, C.J. Carey, I. Polat, Y. Feng, E.W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E.A. Quintero, C.R. Harris, A.M. Archibald, A.H. Ribeiro, F. Pedregosa, P. van Mulbregt, SciPy 1.0 Contributors, SciPy 1.0: fundamental algorithms for scientific computing in python, *Nat. Methods* 17 (2020) 261–272, <https://doi.org/10.1038/s41592-019-0686-2>.
 - [35] C.R. Harris, et al., Array programming with NumPy, *Nature* 585 (7825) (Sep. 2020) 357–362, <https://doi.org/10.1038/s41586-020-2649-2>.
 - [36] The pandas development team, Pandas-Dev/Pandas: Pandas, Zenodo, 2023, <https://doi.org/10.5281/zenodo.8301632> [Online]. Available: Version 2.2.1.
 - [37] J.D. Hunter, Matplotlib: a 2D graphics environment, *Comput. Sci. Eng.* 9 (3) (May 2007) 90–95, <https://doi.org/10.1109/MCSE.2007.55>.
 - [38] M.L. Waskom, Seaborn: statistical data visualization, *J. Open Source Softw.* 6 (60) (Apr. 2021) 3021, <https://doi.org/10.21105/joss.03021>.
 - [39] N.A. Heard, P. Rubin-Delanchy, Choosing between methods of combining p-values, *Biometrika* 105 (1) (Mar. 2018) 239–246, <https://doi.org/10.1093/biomet/asx076>.
 - [40] W. Poole, D.L. Gibbs, I. Shmulevich, B. Bernard, T.A. Knijnenburg, Combining dependent P-values with an empirical adaptation of Brown's method, *Bioinformatics* 32 (2016) i430–i436, <https://doi.org/10.1093/bioinformatics/btw438>. URL: <https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btw438>.
 - [41] P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis, *BMC Bioinf.* 9 (2008) 559, <https://doi.org/10.1186/1471-2105-9-559>. URL: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-9-559>.
 - [42] T.D. Otto, D. Wilinski, S. Assefa, T.M. Keane, L.R. Sarry, U. Böhme, J. Lemieux, B. Barrell, A. Pain, M. Berriman, C. Newbold, M. Llinás, New insights into the blood-stage transcriptome of *Plasmodium falciparum* using RNA-seq, *Mol. Microbiol.* 76 (2010) 12–24, <https://doi.org/10.1111/j.1365-2958.2009.07026.x>.
 - [43] K.M. Broadbent, J.C. Broadbent, U. Ribacke, D. Wirth, J.L. Rinn, P.C. Sabeti, Strand-specific RNA sequencing in *Plasmodium falciparum* malaria identifies developmentally regulated long non-coding RNA and circular RNA, *BMC Genom.* 16 (2015) 454, <https://doi.org/10.1186/s12864-015-1603-4>. URL: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12864-015-1603-4>.
 - [44] C.G. Toenhake, S.A.-K. Frischka, M.S. Vijayabaskar, D.R. Westhead, S.J. van Heeringen, R. Bártfai, Chromatin accessibility-based characterization of the gene regulatory network underlying *Plasmodium falciparum* blood-stage development, *Cell Host Microbe* 23 (2018) 557–569.e9, <https://doi.org/10.1016/j.chom.2018.03.007>.

- [45] J.S. Wichers, J.A.M. Scholz, J. Strauss, S. Witt, A. Lill, L.-I. Ehnold, N. Neupert, B. Liffner, R. Lühken, M. Petter, S. Lorenzen, D.W. Wilson, C. Löw, C. Lavazec, I. Bruchhaus, E. Tannich, T.W. Gilberger, A. Bachmann, Dissecting the gene expression, localization, membrane topology, and function of the *Plasmodium falciparum* STEVOR protein family, *mBio* 10 (2019) e01500, <https://doi.org/10.1128/mBio.01500-19>.
- [46] A.K. Subudhi, A.J. O'Donnell, A. Ramaprasad, H.M. Abkhallo, A. Kaushik, H.R. Ansari, A.M. Abdel-Haleem, F. Ben Rached, O. Kaneko, R. Culleton, S.E. Reece, A. Pain, Malaria parasites regulate intra-erythrocytic development duration via serpentine receptor 10 to coordinate with host rhythms, *Nat. Commun.* 11 (2020) 2763, <https://doi.org/10.1038/s41467-020-16593-y>.
- [47] L. Chappell, P. Ross, L. Orchard, T.J. Russell, T.D. Otto, M. Berriman, J.C. Rayner, M. Llinás, Refining the transcriptome of the human malaria parasite *Plasmodium falciparum* using amplification-free RNA-seq, *BMC Genom.* 21 (2020) 395, <https://doi.org/10.1186/s12864-020-06787-5>.
- [48] M. Kucharski, J. Tripathi, S. Nayak, L. Zhu, G. Wirjanata, R.W. van der Pluijm, M. Dhorda, A. Dondorp, Z. Bozdech, A comprehensive RNA handling and transcriptomics guide for high-throughput processing of *Plasmodium* blood-stage samples, *Malar. J.* 19 (2020) 363, <https://doi.org/10.1186/s12936-020-03436-w>.
- [49] R.J. Deshaies, C.A. Joazeiro, RING domain E3 ubiquitin ligases, *Annu. Rev. Biochem.* 78 (2009) 399–434, <https://doi.org/10.1146/annurev.biochem.78.101807.093809>. URL: <https://www.annualreviews.org/doi/10.1146/annurev.biochem.78.101807.093809>.
- [50] E.P. Risseuw, T.E. Daskalchuk, T.W. Banks, E. Liu, J. Cotelesage, H. Hellmann, M. Estelle, D.E. Somers, W.L. Crosby, Protein interaction analysis of SCF ubiquitin E3 ligase subunits from Arabidopsis, *Plant J.* 34 (6) (2003 Jun) 753–767. PMID 12795696.
- [51] S. Spork, J.A. Hiss, K. Mandel, M. Sommer, T.W. Kooij, T. Chu, G. Schneider, U.G. Maier, J.M. Przyborski, *Eukaryot. Cell* 8 (2009) 1134–1145 [PMC free article] [PubMed] [Google Scholar].
- [52] A. Zuin, M. Isasa, B. Crosas, Ubiquitin signaling: extreme conservation as a source of diversity, *Cells* 3 (3) (2014 Jul 10) 690–701, <https://doi.org/10.3390/cells3030690>. PMID: 25014160; PMCID: PMC4197634.
- [53] M.W. Lake, M.M. Wuebbens, K.V. Rajagopalan, H. Schindelin, Mechanism of ubiquitin activation revealed by the structure of a bacterial MoeB-MoaD complex, *Nature* 414 (2001) 325–329. View article PMID: 11713534.
- [54] L.M. Lois, C.D. Lima, Structures of the SUMO E1 provide mechanistic insights into SUMO activation and E2 recruitment to E1, *EMBO J.* 24 (2005) 439–451. View article PMID: 15660128.
- [55] I. Lee, H. Schindelin, Structural insights into E1-catalyzed ubiquitin activation and transfer to conjugating enzymes, *Cell* 134 (2008) 268–278. View article PMID: 18662542.
- [56] S. Agrawal, D.W. Chung, N. Ponts, G.G. van Dooren, J. Prudhomme, C.F. Brooks, E.M. Rodrigues, J.C. Tan, M.T. Ferdig, B. Striepen, K.G. Le Roch, An apicoplast localized ubiquitylation system is required for the import of nuclear-encoded plastid proteins, *PLoS Pathog.* 9 (6) (2013) e1003426, <https://doi.org/10.1371/journal.ppat.1003426>. Epub 2013 Jun 13. PMID: 23785288; PMCID: PMC3681736.
- [57] B.J. Foth, L.M. Stimmer, E. Handman, B.S. Crabb, A.N. Hodder, G.I. McFadden, The malaria parasite *Plasmodium falciparum* has only one pyruvate dehydrogenase complex, which is located in the apicoplast, *Mol. Microbiol.* 55 (1) (2005 Jan) 39–53, <https://doi.org/10.1111/j.1365-2958.2004.04407.x>. PMID: 15612915.
- [58] J.D. Fellows, M.J. Cipriano, S. Agrawal, B. Striepen, A plastid protein that evolved from ubiquitin and is required for apicoplast protein import in *Toxoplasma gondii*, *mBio* 8 (3) (2017 Jun 27) e00950, <https://doi.org/10.1128/mBio.00950-17>. PMID: 28655825; PMCID: PMC5487736.
- [59] V. Nagy, I. Dikic, Ubiquitin ligase complexes: from substrate selectivity to conjugational specificity, *Biol. Chem.* 391 (2010) 163–169.
- [60] Z. Rizvi, G.S. Reddy, S.M. Gorde, P. Pundir, D. Das, P.S. Sijwali, *Plasmodium falciparum* contains functional SCF and CRL4 ubiquitin E3 ligases, and CRL4 is critical for cell division and membrane integrity, *PLoS Pathog.* 20 (2) (Feb. 2024) e1012045, <https://doi.org/10.1371/journal.ppat.1012045>.
- [61] L. Jia, M.S. Soengas, Y. Sun, ROC1/RBX1 E3 ubiquitin ligase silencing suppresses tumor cell growth via sequential induction of G2-M arrest, apoptosis, and senescence, *Cancer Res.* 69 (2009) 4974–4982 [PMC free article] [PubMed] [Google Scholar] [Ref list].
- [62] J. Shao, Q. Feng, W. Jiang, Y. Yang, Z. Liu, L. Li, W. Yang, Y. Zou, E3 ubiquitin ligase RBX1 drives the metastasis of triple negative breast cancer through a FBXO45-TWIST1-dependent degradation mechanism, *Aging (Albany NY)* 14 (13) (2022 Jul 8) 5493–5510, <https://doi.org/10.18632/aging.204163>. Epub 2022 Jul 8. PMID: 35802537; PMCID: PMC9320552.