Methods matter: Influential purification and analysis parameters for intracellular parasite metabolomics

Authors: Maureen A. Carey[1,*], Vincent Covelli[2*%], Audrey Brown[3], Jessica G. Cooper[1&], and Jennifer L. Guler[2,3,#]

*Co-first authors

[#]Corresponding author: jlg5fw@virginia.edu

Affiliations:

1: Department of Microbiology, Immunology, and Cancer Biology, University of Virginia School of Medicine, Charlottesville, VA, USA

2: Division of Infectious Disease and International Health, University of Virginia School of Medicine, Charlottesville, VA, USA

3: Department of Biology, University of Virginia, Charlottesville, VA, USA

[%]Current address: Department of Infectious Diseases, Providence-St. Joseph Health, Victorville, CA, USA

[&]Current address: Department of Plant and Soil Sciences, University of Delaware, Newark, DE, USA

ABSTRACT

Due to improved instrument sensitivity and access, the use of metabolomics is gaining traction for the study of many organisms and pathogens. For the intracellular malaria parasite, *Plasmodium falciparum*, both targeted and untargeted metabolite detection has improved our understanding of pathogenesis, host-parasite interactions, parasite response to antimalarials, and impacts of resistance. However, protocols for purification are not optimized for investigations of intracellular pathogens and noise-limiting analysis parameters are not well defined. To explore influential parameters, we purified a diverse set of *in vitro* grown intra-erythrocytic *P. falciparum* parasites for untargeted metabolomics studies. Following metabolite identification, data processing included normalization to double stranded DNA, total protein, or parasite number to correct for different sample sizes and stage differences. We found that parasite-derived variables were most appropriate for normalization as they separate sample groups and reduce noise within the data set. However, these post-analysis steps did not remove the contribution from the host erythrocyte, in the form of membrane rich ‘ghosts’, and levels of technical sample variation persisted. In fact, we found that host contamination is as influential on the metabolome as sample treatment. This analysis also identified metabolites with potential to be used as markers to quantify host contamination levels. In conclusion, purification methods and normalization choices during the collection and analysis of untargeted metabolomics heavily affect the interpretation of results. Our findings provide a basis for development of improved experimental and analytical methods for future metabolomics studies of *P. falciparum* and other intracellular organisms.

Importance: Molecular characterization of pathogens, such as the malaria parasite, can lead to effective treatment strategies and improved understanding of pathogen biology. However, the distinctive biology of the *Plasmodium* parasite, such as its repetitive genome and requirement of growth within a host cell, hinders progress towards this goal. Untargeted metabolomics is one promising approach to learn about pathogen biology and how it responds to different treatments. By measuring many small molecules in the parasite at once, we gain a better understanding of important pathways that contribute to this response. Although increasingly popular, protocols for parasite isolation from the host cell and various analysis options are not well explored. The findings presented in this study emphasize the critical need for improvements in these areas to limit misinterpretation due to host metabolites and correct for variations between samples. This will aid both basic biological investigations and clinical efforts to understand important pathogens.

Introduction

Malaria continues to be responsible for hundreds of thousands of deaths annually, most of which result from infection with the protozoan parasite, *Plasmodium falciparum* (1). Characterization of the biology of this important pathogen can lead to improved treatment strategies. The molecular mechanisms behind interesting *P. falciparum* phenotypes are challenging to understand due to a lack of traditional methods of investigation in this organism, such as forward and reverse genetics. Unbiased ‘omics approaches (transcriptomics and proteomics) are widely used but the limited annotation of the parasite genome makes these data sets challenging to interpret. One way to

alleviate this lack functional knowledge is to use network-based modeling to facilitate data interpretation (2). Additionally, the measurement of direct mediators of the phenotype, such as metabolite reactants and products of enzymatic reactions, can improve our ability to make predictions about cellular function under certain conditions. For this reason, metabolomics is becoming increasingly popular to study *P. falciparum* (3-12). These studies have allowed for a greater understanding of malaria pathogenesis (13), strain-specific phenotypes (11), and host-parasite interactions (9). Although metabolomics can successfully identify metabolic signatures that correlate well with biological function, such as time- and dose-dependent response to antimalarial treatment (3, 5) and resistance-conferring mutations (12), there are distinct challenges that need to be considered when performing metabolomic studies in *P. falciparum*.

Challenges such as host contamination, limited parasite yield, and parasite stage-specificities arise due to certain properties of this organism (see **Table 1**). For example, experimental samples typically have few parasites and abundant host material. One contributing factor is that parasitemias are limited during *in vitro* culture and clinical infections (<5% or five infected erythrocytes per 100 total (14, 15)). Additionally, *P. falciparum* is an intracellular parasite during the asexual cycle in the human blood stream; the host erythrocyte accounts for up to a 10-fold more cellular material over early state parasites (16, 17). Due to our ability to enrich for late stage parasites using magnetic purification (18), the study of the larger later stage parasite has historically allowed for efficient genomic, transcriptomic and proteomic analysis of parasite biology. These stages have typically been thought of as more metabolically active than the early stage parasites due to increased activity of well-studied cellular

pathways, including robust hemoglobin degradation (19), nuclear genome replication, and protein synthesis (20, 21). The study of the smaller early stage of the parasite is particularly hard to achieve due to difficulty isolating adequate amounts of parasite material as a result of few effective enrichment methods (22). Thus, studies must be designed in a manner to overcome these challenges, limiting sample-to-sample variation and optimizing metabolite recovery (i.e. total number of metabolites detected).

In this study, we sought to define critical parameters that would help overcome these challenges and allow the collection of high quality metabolomics data. We show that diverse sample groups can be differentiated, but the choice of analytic parameters for data processing and host cell contamination both heavily influence the parasite metabolome. In particular, we investigated normalization approaches to assess the impact of host contamination and found that the adjustment to parasite-derived variables better remove sample noise. However, even appropriate normalization fails to remove host noise completely, as host contamination is as influential on metabolome as sample treatment. Thus, we propose that the combination of improved purification and analytic parameters will generate more accurate measures of the metabolome, increasing the utility of unbiased metabolomics to investigate intracellular parasite biology.

RESULTS

Parasite sample groups are metabolically distinct

To ensure our metabolomics approach can identify obvious differences in sample groups, we compared parasite groups that differed in stage, origin, and growth

conditions (**Fig. 2A**). Distinct purification procedures were used for preparation of each sample group (see *Materials and Methods* and **Fig. 1**), resulting in different amounts of parasite material (**Fig. 2B, Table S1**). Replicates of sample group 1, which were merely lysed from host cells with a mean parasitemia of 1.14%, contained between 1.3-6.9 x 10^6 total parasites. Sample group 2 was enriched for late stage parasites using magnetic purification to a mean parasitemia of 53.6% (**Table S1**). These replicates contained between 4.7 x 10^7 to 6.7 x 10^8 total parasites (up to 100-fold more individual parasites). Despite these differences, mean protein abundance was insignificantly different across replicates of each sample group and was more variable in sample group 2 (group 1 SD: 12.7, group 2 SD: 38.2, see supplementary information for code and **Fig. 2B**). Sample group 1 had a mean of 115.3 µg/ml of protein, and sample group 2 had a mean value of 107.6 µg/ml. Cell number and DNA abundance are positively correlated, as expected (r = 0.8037, p-value = 0.00002, **Fig. 2B**); these values are not perfectly correlated because the late stage parasites in sample group 2 are actively replicating DNA, and, thus, have increased and variable genome copy number per cell. Protein does not correlate with parasite number or DNA abundance (data not shown, see supplementary information for code).

We conducted metabolomics on the samples described above (**Fig. 1**). Cultured parasites were lysed from host erythrocytes and analyzed via UPLC-MS. In comparison 1, we detected 375 total metabolites that were annotated by Metabolon, Inc.; 143 of these were detected in every sample and represented 10 energy associated metabolites, 159 lipid species, 108 peptides and amino acids, 40 nucleotides, 28 cofactors, 20 carbohydrates, and 10 others (**Fig. 2C**). Samples from group 1 contained

between 182-242 metabolites while those from group 2 contained between 267-368 metabolites (**Fig. 2C**). Fifteen metabolites are found in every group 1 sample, but not all group 2 samples, and 111 metabolites are found in every group 2 sample but not all group 1 samples. Thus, distinct samples, due to parasite origin, stage, growth conditions, and purification differences, have distinct metabolomes.

Normalization parameters influence sample variation

Normalization methods can influence results (23), but have not been explored in the use of metabolomics for *Plasmodium* nor other intracellular pathogens. To explore the importance of various normalization approaches, we performed principal component analysis with all sample metabolomes using either unnormalized data or three normalization methods: quantification of parasite number, double stranded DNA, and total protein amount. Each normalization method yields distinct principle component structures and clearly separates sample groups (**Fig. 2D**). In all cases, principle component (PC) 1 primarily represents *between* group variation, and PC2 represents *within* group variation (**Fig. 2D**). Without normalization, PC 1 and 2 summarize 78.4% of sample variation. These principal components from parasite number and DNA normalization summarize 87.7 and 80.6% of sample variation, respectively. With protein normalization, 79.1% of variation is summarized. PC2 tends to separate sample group 1 better than those samples within group 2 (**Fig. 2D**).

The metabolites that most contribute to group or sample variation are not the same with each normalization approach (**Table S2**). Thus, metabolome differences between groups are dependent on normalization approach. Yet, there are several

158 striking trends across analyses. For example, the PC structure following protein
159 normalization closely mimics that of the unnormalized data and, similar metabolites
160 contribute to PC1 and PC2 in both analyses. Sphingomyelin species contribute to within
161 group variation (PC2), and orotidine and dipeptides contribute to between group
162 variation (PC1; **Table S2**). Upon DNA or parasite number normalization, phenylalanine,
163 tryptophan, leucine, putrescine, and sedoheptulose 7-phosphate contribute to PC2, or
164 within group variation (**Table S2**). Contrary to protein amount (see *Discussion*), DNA
165 and parasite number normalization are parasite-derived and, thus, these two
166 measurements are preferable for normalization. The choice of which parasite-derived
167 variable to use for normalization should be based on the experimental question.
168 Accordingly, we normalize to parasite number during our subsequent comparison of
169 sample groups 1 (early stage) and 2 (late stage; see **Fig. 2**); normalization to DNA
170 amount would not be appropriate because these different stages have known genome
171 copy number differences (late stage parasites are actively replicating their DNA,
172 whereas ring stage parasites are haploid). Furthermore, we normalize to DNA content
173 during our subsequent comparison within a group (i.e. replicates of samples group 1,
174 see **Fig. 3**). In this case, normalization to this parasite variable is more appropriate
175 because these measurements are collected immediately prior to mass spectrometry
176 metabolite processing (**Fig. 1**) in our experimental design and are the most
177 representative of analyzed samples.
178
179 <u>Remnants of the erythrocyte host contribute to metabolite pool</u>
180 Beyond comparing the metabolomes of artificially distinct samples groups, we explored

the metabolic changes induced by antimalarial treatment. We collected metabolomics from treated and untreated early stage parasites that were identical in growth conditions and purification approach, and were matched for blood batch (**Fig. 3A**, **Table S1**, see *Materials and Methods* for group 1). Following data processing, the metabolomes of antimalarial treated and untreated parasites fail to cluster via PCA (**Fig. 3B**). Accordingly, univariate statistical analysis revealed no differentially abundant metabolites between treated and untreated samples (see supplemental information for code).

When considering possible explanations for this result, microscopy revealed that parasites lysed from host cells remain embedded in erythrocyte membranes and washes fail to isolate parasite material (**Fig. 3C**). This result emphasized that erythrocyte 'ghosts' (cell membranes with associated metabolites) remain abundant in the sample and heavily contribute to the metabolome (see *Discussion*). In fact, univariate statistical analysis only revealed one metabolite with increased abundance in one blood batch (1-arachidonoyl-GPE; see supplemental information for code). Thus, the metabolome is likely influenced by both blood batch and antimalarial treatment, with the noise induced by each variable overshadowing group differences.

To further explore the host contribution to the metabolome, we built two Random Forest classifiers to identify metabolites that are associated with either erythrocyte ghosts or antimalarial treatment. We first built a classifier to predict blood batch in early-stage parasites (**Fig. 3A**). These samples likely have large host contribution due to the inability to enrich for erythrocytes infected with early stage parasites. Ninety-five metabolites (of 298), including AMP, ADP-ribose, aspartate, and sphingosine improved

classifier accuracy in predicting blood batch (most influential depicted in **Fig. 3D**, see supplemental information for code); the remaining metabolites had no effect on classifier performance or worsened its predictive capabilities, indicating they are not associated with blood batch due to high variability or association with other features that differentiate samples. This classifier predicted blood batch with a 30% error rate. Thus, a subset of the measured metabolome was predictive of blood batch.

To determine if blood batch is as influential on metabolome as antimalarial treatment, we built a similar classifier to predict treatment within early stage samples (**Fig. 3A**). Early stage parasites were classified into two treatment conditions with a 30% class error rate. One hundred and eighteen metabolites (of 298) improved classification accuracy (see most influential in **Fig. 3E**, and supplemental information for code), including pipecolate and several dipeptides. Thus, sample metabolome can classify both blood batch and sample group, indicating sample treatment and blood batch influence the metabolome.

DISCUSSION

Here, we explore metabolomics methods used in *in vitro* study of intraerythrocytic *P. falciparum*. The parasite's intracellular lifestyle introduces challenges in implementing traditional protocols, predominately due to limited amounts of parasite material and host metabolite contamination. In our study, we sought to determine critical parameters for the collection of high quality metabolomics data despite these challenges. In particular, we investigated normalization approaches and conducted a detailed assessment of the impact of host contamination. Overall, we found that only parasite-derived variables are

best suited to use during normalization. Despite these analytic approaches, host noise permeates the analysis, as host contamination is as influential on metabolome as antimalarial treatment. Thus, improvements in both purification and analytic parameters must be combined to generate accurate metabolomes and increase our ability to learn more about the parasite's biology.

Normalization of metabolite levels aims to limit technical or non-biological variation, thus enhancing interpretation of results. Normalization can be calculated by a variety of methods and is implemented either before or after analysis (**Table 1** (24, 25)). Often, pre-analysis normalization is conducted by isolating the same number of cells for analysis (26) but this is not typically used in the study of *P. falciparum* as generating adequate biomass can be challenging. Furthermore, sample adjustments following the use of inaccurate quantification methods may introduce more variability. Post-analysis normalization methods are also routinely used; these include the use of internal standards (25, 27), corrections for protein amount (often used for supernatant or cell-free metabolomics (28)), DNA content (an approach validated in mammalian cells (29)), or cell number (typically used for bacterial populations (30)). A common approach used in the study of *P. falciparum* involves an uninfected erythrocyte control to adjust for the presence of host metabolites (7, 10, 27, 31-33). However, use of this control without other forms of normalization led to the misattribution of host metabolites to the parasite (34). Selecting the correct method of normalization in *P. falciparum* metabolomics studies is essential to ensure that parasite-derived metabolites, and not host-derived metabolites, are measured and interpreted to make conclusions.

We explore three post-analysis normalization approaches: protein, DNA, and

parasite number. We argue the host erythrocyte heavily contributes to protein abundance, and, thus, this metric is not solely parasite-derived. In our analysis, this was most clearly observed when comparing protein abundances between our sample groups (**Fig. 2B**). We expected a proportional increase in protein amount as parasite size increases throughout the intraerythrocytic life cycle (from sample group 1 to 2; early to late stage); however, this increase was not detected, implicating host erythrocyte contribution. Furthermore, heavy host contamination explains the observations that 1) there is an increased level of protein variability in group 2 (explained by the wider range in parasitemia level and thus host erythrocyte contribution, **Table S1**), 2) host/media metabolites such as kynurinine, phenol red, and HEPES were detected in this analysis (see below and supplemental data), and 3) protein normalization minimally changes the PCA data structure and top contributing metabolites (**Fig. 2D** and **Table S2**).

In sharp contrast, total DNA amount and parasite count are entirely parasite-derived; mature uninfected erythrocytes are anucleated, without detectable DNA (35), and are excluded when determining parasite count (see *Materials and Methods*). When metabolites were evaluated following DNA and parasite count normalization, more nuclear material and total parasites were observed in later stages (group 2, **Fig. 2B**). These data are not surprising, as late-stage parasites are known to amplify DNA content up to twenty times during their asexual life cycle (36). A greater cell count in late stage samples can be attributed to the higher parasitemia that is achieved through magnetic purification of late stage trophozoites and schizonts (37). To our knowledge, normalization to parasite-derived material has not been described in detail in previous metabolomics studies of *P. falciparum*. We propose that similar to studies in *Leishmania*

(38-40), normalization to parasite-derived measurements should become standard during metabolomics analysis of these intraerythrocytic parasites (**Table 1**).

Clearly, parasite-to-parasite sample variation can influence metabolomics data, but we also found host erythrocyte material can heavily impact a sample's metabolome. Many studies employ erythrocyte lysis prior to sample purification ((8, 32) and our current study, see *Materials and Methods*). However, this approach does not eliminate the potential for host contamination; host membrane fragments devoid of internal components, colloquially referred to as erythrocyte "ghosts," remain in purified samples (**Fig. 3C**). Despite this concerted effort to limit host metabolites through lysis, our studies support heavy erythrocyte contribution to the *P. falciparum* metabolome. Several metabolites were detected in group 1 and 2 metabolomes that have not previously been measured as produced or consumed in *Plasmodium*. For example, kynurenine is known to be present in erythrocytes and is derived from the amino acid L-tryptophan (41, 42). Although no known production or consumption has been reported in the parasite, kynurenine was detected in 13 of our 30 samples, most frequently in the group 2 (late stage parasites, see supplemental data). This finding indicates some metabolites may be from the host, not the parasite, or the parasite has greater metabolic capabilities than previously understood. Similarly, media components such as phenol red (phenolsulfonphthalein) and HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) were measured in parasite metabolomes (see supplemental data). Neither are produced or consumed by the parasite but likely remained associated with our cells following *in vitro* culture in media that contains these metabolites (i.e. RPMI, see *Materials and Methods*). The abundance of phenol red and

296 HEPES, as well as cholesterol (a metabolite excluded from parasite membranes (43,
297 44)) are correlated prior to normalization, and these correlations persist following
298 normalization. Moreover, phenol red contributed to the accuracy of our antimalarial
299 treatment classifier, further confirming that blood batch effects influenced the dataset.
300 Lastly, lipid species were the major class of metabolites detected in our analysis (**Fig.**
301 **2C**) and contributed heavily to PC2 from un- and protein-normalized data sets (**Table**
302 **S2**), perhaps due to the remaining erythrocyte membranes. These results add to the
303 overwhelming evidence of host cell and media contamination in untargeted
304 metabolomics studies of parasites.

305 Following these observations, we also explored the effect of different blood
306 batches on metabolome measurements. Because generating sufficient *Plasmodium*
307 biomass for adequate biological replicates is time-intensive, many experiments require
308 multiple batches of human blood donations. To avoid batch effects, we controlled blood
309 batches across sample groups (**Table S1**). Prior to these studies, we predicted that the
310 blood batch would have some effects on the metabolome; we did not anticipate,
311 however, that it would be as influential as known stressors, like treatment with
312 antimalarials with established metabolic effects (3, 5). Several results from our analysis
313 support this observation. First, samples from either treatment group did not cluster via
314 PCA (**Fig. 3B**). Second, we detected none-to-few metabolites whose levels were
315 significantly different between conditions (zero between with and without antimalarial
316 treatment and 1 between various blood batches). Lastly, classifiers from both treatment
317 and blood batch predicted samples with equal accuracy (30% error rate, top predictive
318 metabolites displayed in **Fig. 3D** and **E**). Overall, from these analyses, we concluded

that sample-to-sample variation exceeded variation associated with either group. We also found 1-arachidonoyl-GPE to be significantly different in abundance across blood batches, which can be explored as a potential biomarker of host contamination. To expand on this idea, we were also able to predict a set of metabolites that are most likely to be host erythrocyte-derived (or influenced by host environment) by identifying the metabolites that are most predictive of blood batch (**Fig. 3D**). Additional investigations are required since these metabolites may be parasite-derived but only produced when they are in particular environments (e.g. blood batches). Going forward, it may be possible to use these metabolites to quantify host cell contribution to metabolome and assess parasite sample purity or control for host contamination during analysis.

Overall, the methodology and findings from the current study provide a basis for the use of more streamlined *in vitro* metabolomics approaches for the future investigation of *P. falciparum* biology. We suggest a set of considerations and recommendations for enhancing the accuracy parasite metabolomics (presented in **Fig. 1 Table 1, and below**). First, samples must be better purified away from host material. Enrichment methods, whether novel or standard, should be used to increase parasitemia, reducing the number of uninfected host cells. Second, markers of host contamination must be used to evaluate the level of host contamination and resulting data. Our studies suggest that visual detection of ghost material (via microscopy) combined with assessment of host-specific metabolite markers is an effective option to assess sample purity. Finally, data must be normalized to parasite-derived measurements to limit remaining host contamination. With these considerations,

342 metabolomics has the potential to be a powerful tool in the study of intracellular
343 parasites, like *Plasmodium*.
344
345 **MATERIALS AND METHODS**
346 **Parasite Cultivation**
347 Laboratory-adapted *P. falciparum* lines were cultured in RPMI 1640 (Roswell Park
348 Memorial Institute medium, Thermo Fisher Scientific, Waltham, MA) containing HEPES
349 (Sigma Aldrich, St Louis, MO) supplemented with either 0.5% AlbuMAX II Lipid-rich
350 BSA (Sigma Aldrich, St Louis, MO) and 50 mg/L hypoxanthine (Thermo Fisher
351 Scientific, Waltham, MA) (referred to as AlbuMAX media) or 20% v/v pooled human
352 plasma for generation of complete RPMI (referred to as cRPMI).Parasite cultures were
353 maintained at 3% hematocrit and diluted with human red blood cells (blood batch noted
354 in **Table S1**) to maintain parasitemia between 1-3%, with change of culture medium
355 every other day (**Fig. 1; Step 1**). Cultures were incubated at 37ºC with 5% oxygen, 5%
356 carbon dioxide and 90% nitrogen (14). Some samples were treated with antimalarials
357 with metabolic effects to maximize differences between groups (see below and
358 Antimalarial treatment in **Table S1**).
359 **Parasite Isolation**
360 For isolation of sample group 1, two distinct laboratory-adapted clinical isolates of *P.*
361 *falciparum* (BEI Resources, NIAID, NIH: *Plasmodium falciparum*, Strain IPC 4884/MRA-
362 1240 and IPC 5202/MRA-1238, contributed by Didier Ménard) containing mixed stages
363 with >50% rings were synchronized using 5% sorbitol (Sigma Aldrich, St Louis, MO)
364 (45). The resultant early stage cultures were incubated at 37°C in AlbuMAX media to

allow for the development of a schizont predominant population (see *Parasite Cultivation* above). After the late stage population was confirmed using microscopy, cultures were checked every one to two hours for the development of newly invaded ring stage parasites. If the parasites were treated with antimalarials, it was performed at this stage. Fourteen flasks containing early ring-stage parasites (<3 hours post invasion) were subsequently lysed from the erythrocyte membrane using 0.15% saponin, as previously described (46) (**Fig. 1; Step 3)**. Prior to lysis, sampling of parasite material was taken for determination of erythrocyte count (hemocytometer) and parasitemia (SYBR-green based flow cytometry (47)), which contributed to parasite number determination (total erythrocytes x % parasitemia yields total parasites). Additional samples were obtained following erythrocyte lysis for protein quantification using Bradford reagent (Sigma Aldrich, St Louis, MO). A series of three wash steps were then performed using 1X PBS (Sigma Aldrich, St Louis, MO) using centrifugation at 2000 x g to remove soluble erythrocyte metabolites. Purified material was kept on ice until flash frozen using liquid nitrogen, followed by storage at -80^{o}C until sent for analysis. This procedure was performed five times for each parasite line to provide 10 replicates for group 1 metabolomic analysis. Additionally, matched parasites (same parasite lineage, media type, stage, blood batches, and purification methods) were also grown without drug treatment (**Table S1**) to generate 10 additional samples for group 1 untreated (see second comparison in **Fig. 3**).

For isolation of sample group 2, two Dd2-derived laboratory-adapted clones of *P. falciparum* (courtesy of Pradip Rathod, University of Washington, continuously cultured in the presence of antimalarial, **Table S1**) first underwent an initial sorbitol

388 synchronization step as above. The resultant early stage parasites were then incubated
389 at 37°C in cRPMI to allow for the successful transition of *P. falciparum* to the late
390 trophozoite and schizont stages, occurring 24 to 30 hours after initial synchronization.
391 Next, this predominantly late stage population was enriched through magnetic
392 purification using a MACS quad-magnet and MACS multistand (Miltenyi Biotech,
393 Bergisch Gladbach, Germany), as previously described (18) (**Fig. 1; Step 2)**. Briefly,
394 parasite cultures were passed through LS columns with attached sterile syringe needles
395 (BD Biosciences, San Jose CA) at a rate of 2-3 seconds per drop. A series of two to
396 three column washes were performed with 5 ml of warmed cRPMI. To elute the desired
397 material, the column was removed from the magnet prior to adding 5 ml of cRPMI.
398 Column flow-through from 5 flasks containing late stage parasites was allowed to
399 recover in cRPMI for 30 min at 37^{o} C prior to saponin lysis, as described above (**Fig. 1;**
400 **Step 3)**. Determination of parasite count and protein quantification, as well as
401 subsequent sample washing and freezing, were performed as described above for
402 sample group 1. This procedure was performed five times for each parasite line to
403 provide 10 samples for group 2 metabolomic analysis.

404 **Metabolite Preparation, Analysis, and Identification**

405 Metabolites were identified using Ultrahigh Performance Liquid Chromatography-Mass
406 Spectroscopy (UPLC-MS) by Metabolon, Inc. (Durham, NC). All sample preparations
407 and metabolite identifications were performed according to Metabolon, Inc, standard
408 protocols. Briefly, double stranded DNA was quantified in all samples using the Quant-it
409 Picogreen dsDNA Assay Kit (Thermo Fisher, Waltham, MA) according to the
410 manufacturer's instructions and proteins were precipitated with methanol and

411 centrifuged for extraction (**Fig. 1; Step 4**). Sample extracts were dried and reconstituted
412 in solvents containing standards (see below) at fixed concentrations to ensure injection
413 and chromatographic consistency. Waters AQUITY ultra-performance liquid
414 chromatography (UPLC) and Thermo Scientific Q-Exactive high resolution/accurate
415 mass spectrometer were used for metabolite detection (**Fig. 1**; **Step 5**). Controls that
416 were analyzed in conjunction with the experimental samples included a pooled matrix of
417 all included samples. Internal and recovery standards were used to assess variability
418 and to verify performance of extraction and instrumentation, as routinely performed by
419 Metabolon, Inc.

420 Raw data was extracted using hardware and software developed by Metabolon,
421 Inc. Metabolites were quantified using area-under-the-curve and identified by
422 comparison to a library of several thousands of pre-existing entries of purified standards
423 or recurrent unknown compounds. Each library standard was uniquely authenticated by
424 retention time/indexes, mass to charge ratios, and chromatographic data. Named
425 metabolites corresponded to library standards or were predicted with confidence
426 according to Metabolon, Inc standard protocols.

427 **Data Analysis**

428 Following the analytical protocol outlined in (48), we first preprocessed metabolite
429 abundances for each sample by imputing missing values with half of the lowest
430 detectable metabolite abundance. Next, we normalized metabolite abundances by
431 sample features, followed by normalization using metabolite features with log
432 transformation, centering, and scaling (49). To limit inter-sample variability, metabolite
433 abundances for each replicate were normalized to sample value for double stranded

DNA, protein, or parasite number. To limit inter-metabolite variability, metabolite abundances were log transformed, centered to median (50), and scaled by standard deviation (**Fig 1; Step 6**). Resultant processed metabolite abundances were used for univariate and multivariate statistics, as well as classification. All analyses were conducted using R (51-59). Welch's t-tests were used to compare group means for differential abundance determination, assuming unequal variance and normal distribution, and p-values were adjusted using a false discovery rate. The significance cutoff is 0.05. See supplementary information for code and detailed analysis.

Microscopy

Laboratory adapted *P. falciparum* clones (BEI Resources, NIAID, NIH: *Plasmodium falciparum*, Strain Patient line E/MRA-1000 or IPC 4884/MRA-1238, contributed by Didier Ménard) at >50% rings were lysed using 0.15% saponin, as previously described (46). Samples were washed twice using 1X PBS (Sigma Aldrich, St Louis, MO) and centrifugation at 2000 x g for 5 minutes. Samples were then stained on slides with either DAPI at 1:20,000 (Sigma Aldrich, St Louis, MO) and CD235a-PE antibody at 1:100 (Thermo Fisher Scientific, Waltham, MA) for fluorescence microscopy or with Giemsa stain (Sigma Aldrich, St Louis, MO) for bright field microscopy. Fluorescent images were acquired using the EVOS FL Cell Imaging System (Thermo Fisher Scientific, Waltham, MA). Bright field images were acquired using a Nikon Eclipse Ci Upright Microscope (Nikon, Melville, NY) equipped with a DMK23U274 camera (The Imaging Source, Charlotte, NC) and NIS Elements Imaging Software (Nikon, Melville, NY).

ACKNOWLEDGEMENTS

457 This project was supported with funding through the National Institute of Allergy and
458 Infectious Disease R21AI119881 and the University of Virginia CHARGE Enhancement
459 Fellowship (to JLG). MAC and VC are supported by institutional training grants
460 (T32GM008136 and T32AI007046, respectively).

461 We would like to thank Dr. Jason Papin and Gregory Medlock, as well as the
462 members of the Guler, Papin, and Petri labs at the University of Virginia for helpful
463 discussion and feedback on experimental design and analysis. Additionally, we would
464 like to thank Michelle Warthan (University of Virginia) for laboratory support, and
465 Webster Santos (Virginia Tech) and Kevin Lynch (University of Virginia) for antimalarial
466 synthesis and purification.

467

468

REFERENCES

1. **WHO.** 2015. Global Technical Strategy for Malaria 2016–2030. Geneva: World Health Organization (WHO).
2. **Carey MA, Papin JA, Guler JL.** 2017. Novel Plasmodium falciparum metabolic network reconstruction identifies shifts associated with clinical antimalarial resistance. BMC Genomics **18:**543.
3. **Allman EL, Painter HJ, Samra J, Carrasquilla M, Llinás M.** 2016. Metabolomic profiling of the malaria box reveals antimalarial target pathways. Antimicrobial agents and chemotherapy **60:**6635-6649.
4. **Babbitt SE, Altenhofen L, Cobbold SA, Istvan ES, Fennell C, Doerig C, Llinas M, Goldberg DE.** 2012. Plasmodium falciparum responds to amino acid starvation by entering into a hibernatory state. Proc Natl Acad Sci U S A **109:**E3278-3287.
5. **Creek DJ, Chua HH, Cobbold SA, Nijagal B, MacRae JI, Dickerman BK, Gilson PR, Ralph SA, McConville MJ.** 2016. Metabolomics-Based Screening of the Malaria Box Reveals both Novel and Established Mechanisms of Action. Antimicrob Agents Chemother **60:**6650-6663.
6. **Olszewski KL MJ, Wilinski D, Burns JM, Vaidya AB, Rabinowitz JD, et al. .** 2009. Host-parasite interactions revealed by Plasmodium falciparum metabolomics. Cell Host Microbe **5:**191-199.
7. **Park YH, Shi YP, Liang B, Medriano CA, Jeon YH, Torres E, Uppal K, Slutsker L, Jones DP.** 2015. High-resolution metabolomics to discover potential parasite-specific biomarkers in a Plasmodium falciparum erythrocytic stage culture system. Malar J **14:**122.
8. **Parvazi S, Sadeghi S, Azadi M, Mohammadi M, Arjmand M, Vahabi F, Sadeghzadeh S, Zamani Z.** 2016. The Effect of Aqueous Extract of Cinnamon on the Metabolome of Plasmodium falciparum Using 1HNMR Spectroscopy. Journal of Tropical Medicine **2016:**5.
9. **Sana TR, Gordon DB, Fischer SM, Tichy SE, Kitagawa N, Lai C, Gosnell WL, Chang SP.** 2013. Global Mass Spectrometry Based Metabolomics Profiling of Erythrocytes Infected with Plasmodium falciparum. PLOS ONE **8:**e60840.
10. **Sengupta A, Ghosh S, Das BK, Panda A, Tripathy R, Pied S, Ravindran B, Pathak S, Sharma S, Sonawat HM.** 2016. Host metabolic responses to Plasmodium falciparum infections evaluated by 1H NMR metabolomics. Mol Biosyst **12:**3324-3332.
11. **Teng R, Lehane AM, Winterberg M, Shafik SH, Summers RL, Martin RE, van Schalkwyk DA, Junankar PR, Kirk K.** 2014. 1H-NMR metabolite profiles of different strains of Plasmodium falciparum. Bioscience reports **34:**e00150.
12. **Siddiqui G, Srivastava A, Russell AS, Creek DJ.** 2017. Multi-omics Based Identification of Specific Biochemical Changes Associated With PfKelch13-Mutant Artemisinin-Resistant Plasmodium falciparum. The Journal of Infectious Diseases **215:**1435-1444.
13. **Park YH, Shi YP, Liang B, Medriano CAD, Jeon YH, Torres E, Uppal K, Slutsker L, Jones DP.** 2015. High-resolution metabolomics to discover potential parasite-specific biomarkers in a Plasmodium falciparum erythrocytic stage

culture system. Malaria Journal **14:**122.

14. **Trager W, Jensen JB.** 1976. Human malaria parasites in continuous culture. Science **193**.
15. **WHO.** 2015. Guidelines for the treatment of malaria, 3rd ed. Geneva: World Health Organization (WHO).
16. **Langreth SG, Jensen JB, Reese RT, Trager W.** 1978. Fine structure of human malaria in vitro. J Protozool **25:**443-452.
17. **Canham PB, Burton AC.** 1968. Distribution of size and shape in populations of normal human red cells. Circ Res **22:**405-422.
18. **Paul F, Roath S, Melville D, Warhurst DC, Osisanya JOS.** 1981. SEPARATION OF MALARIA-INFECTED ERYTHROCYTES FROM WHOLE BLOOD: USE OF A SELECTIVE HIGH-GRADIENT MAGNETIC SEPARATION TECHNIQUE. The Lancet **318:**70-71.
19. **Francis SE, Sullivan Jr DJ, Goldberg, E D.** 1997. Hemoglobin metabolism in the malaria parasite Plasmodium falciparum. Annual Reviews in Microbiology **51:**97-123.
20. **de Rojas MO, Wasserman M.** 1985. Temporal relationships on macromolecular synthesis during the asexual cell cycle of Plasmodium falciparum. Trans R Soc Trop Med Hyg **79:**792-796.
21. **Foth BJ, Zhang N, Chaal BK, Sze SK, Preiser PR, Bozdech Z.** 2011. Quantitative time-course profiling of parasite and host cell proteins in the human malaria parasite Plasmodium falciparum. Mol Cell Proteomics **10:**M110.006411.
22. **Jackson KE, Spielmann T, Hanssen E, Adisa A, Separovic F, Dixon MW, Trenholme KR, Hawthorne PL, Gardiner DL, Gilberger T, Tilley L.** 2007. Selective permeabilization of the host cell membrane of Plasmodium falciparum-infected red blood cells with streptolysin O and equinatoxin II. Biochem J **403:**167-175.
23. **van den Berg RA, Hoefsloot HC, Westerhuis JA, Smilde AK, van der Werf MJ.** 2006. Centering, scaling, and transformations: improving the biological information content of metabolomics data. BMC Genomics **7:**142.
24. **Kohl SM, Klein MS, Hochrein J, Oefner PJ, Spang R, Gronwald W.** 2012. State-of-the art data normalization methods improve NMR-based metabolomic analysis. Metabolomics **8:**146-160.
25. **Ejigu BA, Valkenborg D, Baggerman G, Vanaerschot M, Witters E, Dujardin JC, Burzykowski T, Berg M.** 2013. Evaluation of normalization methods to pave the way towards large-scale LC-MS-based metabolomics profiling experiments. OMICS **17:**473-485.
26. **Hoffmann R, Seidl T, Dugas M.** 2002. Profound effect of normalization on detection of differentially expressed genes in oligonucleotide microarray data analysis. Genome Biology **3:**research0033.0031.
27. **Babbitt SE, Altenhofen L, Cobbold SA, Istvan ES, Fennell C, Doerig C.** 2012. Plasmodium falciparum responds to amino acid starvation by entering into a hibernatory state. Proc Natl Acad Sci U S A **109**.
28. **Silva AM, Cordeiro-da-Silva A, Coombs GH.** 2011. Metabolic variation during development in culture of Leishmania donovani promastigotes. PLoS Negl Trop Dis **5:**e1451.

29. **Silva LP, Lorenzi PL, Purwaha P, Yong V, Hawke DH, Weinstein JN.** 2013. Measurement of DNA concentration as a normalization strategy for metabolomic data from adherent cell lines. Analytical chemistry **85:**10.1021/ac401559v.
30. **Rijpma SR, van der Velden M, Bilos A, Jansen RS, Mahakena S, Russel FG, Sauerwein RW, van de Wetering K, Koenderink JB.** 2016. MRP1 mediates folate transport and antifolate sensitivity in Plasmodium falciparum. FEBS Lett **590:**482-492.
31. **Olszewski KL.** 2009. Host-parasite interactions revealed by plasmodium falciparum metabolomics. Cell Host Microbe **5**.
32. **Sana TR, Gordon DB, Fischer SM, Tichy SE, Kitagawa N, Lai C.** 2013. Global mass spectrometry based metabolomics profiling of erythrocytes infected with Plasmodium falciparum. PLoS One **8**.
33. **Teng R, Lehane AM, Winterberg M, Shafik SH, Summers RL, Martin RE.** 2014. 1H-NMR metabolite profiles of different strains of Plasmodium falciparum. Biosci Rep **34**.
34. **Olszewski KL, Mather MW, Morrisey JM, Garcia BA, Vaidya AB, Rabinowitz JD, Llinas M.** 2013. Retraction: Branched tricarboxylic acid metabolism in Plasmodium falciparum. Nature **497:**652.
35. **Whaun JM, Rittershaus C, Ip SH.** 1983. Rapid identification and detection of parasitized human red cells by automated flow cytometry. Cytometry **4:**117-122.
36. **Margos G, Bannister LH, Dluzewski AR, Hopkins J, Williams IT, Mitchell GH.** 2004. Correlation of structural development and differential expression of invasion-related molecules in schizonts of Plasmodium falciparum. Parasitology **129:**273-287.
37. **Ribaut C, Berry A, Chevalley S, Reybier K, Morlais I, Parzy D, Nepveu F, Benoit-Vical F, Valentin A.** 2008. Concentration and purification by magnetic separation of the erythrocytic stages of all human Plasmodium species. Malaria Journal **7:**45.
38. **Rojo D, Canuto GA, Castilho-Martins EA, Tavares MF, Barbas C, Lopez-Gonzalvez A, Rivas L.** 2015. A Multiplatform Metabolomic Approach to the Basis of Antimonial Action and Resistance in Leishmania infantum. PLoS One **10:**e0130675.
39. **Westrop GD, Williams RA, Wang L, Zhang T, Watson DG, Silva AM, Coombs GH.** 2015. Metabolomic Analyses of Leishmania Reveal Multiple Species Differences and Large Differences in Amino Acid Metabolism. PLoS One **10:**e0136891.
40. **Akpunarlieva S, Weidt S, Lamasudin D, Naula C, Henderson D, Barrett M, Burgess K, Burchmore R.** 2017. Integration of proteomics and metabolomics to elucidate metabolic adaptation in Leishmania. J Proteomics **155:**85-98.
41. **Wang Y, Liu H, McKenzie G, Witting PK, Stasch J-P, Hahn M, Changsirivathanathamrong D, Wu BJ, Ball HJ, Thomas SR, Kapoor V, Celermajer DS, Mellor AL, Keaney JF, Hunt NH, Stocker R.** 2010. Kynurenine is an endothelium-derived relaxing factor produced during inflammation. Nat Med **16:**279-285.
42. **Hartai Z, Klivenyi P, Janaky T, Penke B, Dux L, Vecsei L.** 2005. Kynurenine metabolism in plasma and in red blood cells in Parkinson's disease. J Neurol Sci

239:31-35.

43. **Wunderlich F, Fiebig S, Vial H, Kleinig H.** 1991. Distinct lipid compositions of parasite and host cell plasma membranes from Plasmodium chabaudi-infected erythrocytes. Mol Biochem Parasitol **44:**271-277.
44. **Gulati S, Ekland EH, Ruggles KV, Chan RB, Jayabalasingham B, Zhou B, Mantel P-Y, Lee MCS, Spottiswoode N, Coburn-Flynn O, Hjelmqvist D, Worgall TS, Marti M, Di Paolo G, Fidock DA.** 2015. Profiling the Essential Nature of Lipid Metabolism in Asexual Blood and Gametocyte Stages of Plasmodium falciparum. Cell host & microbe **18:**371-381.
45. **Lambros C, Vanderberg JP.** 1979. Synchronization of Plasmodium falciparum erythrocytic stages in culture. The Journal of parasitology**:**418-420.
46. **Doolan DL.** 2002. Malaria Methods and Protocols, vol 72. Humana Press.
47. **Bei AK, Desimone TM, Badiane AS, Ahouidi AD, Dieye T, Ndiaye D, Sarr O, Ndir O, Mboup S, Duraisingh MT.** 2010. A flow cytometry-based assay for measuring invasion of red blood cells by Plasmodium falciparum. Am J Hematol **85:**234-237.
48. **Xia J, Wishart DS.** 2011. Web-based inference of biological patterns, functions and pathways from metabolomic data using MetaboAnalyst. Nat Protocols **6:**743-760.
49. **Sugimoto M, Kawakami M, Robert M, Soga T, Tomita M.** 2012. Bioinformatics Tools for Mass Spectroscopy-Based Metabolomic Data Processing and Analysis. Curr Bioinform **7:**96-108.
50. **Evans AM, DeHaven CD, Barrett T, Mitchell M, Milgram E.** 2009. Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems. Anal Chem **81:**6656-6667.
51. **Borchers HW.** 2015. Pracma: practical numerical math functions. R package version 1.8. 3.
52. **Champely S.** 2012. pwr: Basic functions for power analysis. R package version **1**.
53. **Kuhn M.** 2015. Caret: classification and regression training. Astrophysics Source Code Library.
54. **Liaw A, Wiener M.** 2002. Classification and regression by randomForest. R news **2:**18-22.
55. **Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F.** 2015. e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien, 2015. R package version**:**1-6.
56. **Scherer R, Scherer MR.** 2016. Package 'samplesize'.
57. **Team RC.** 2014. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2014.
58. **Wickham H.** 2007. Reshaping data with the reshape package. Journal of Statistical Software **21:**1-20.
59. **Wickham H.** 2016. tidyverse: Easily install and load 'tidyverse'packages [Software].

TABLES AND FIGURE

Table 1. **Parameters in metabolomics analysis of intracellular parasites, including *Plasmodium*.** Note: most parameters do not have strict recommendations, as they are dependent on experimental design. Grey highlights indicate methods that were employed and evaluated during this study.

Parameter	Options	Factors to consider
Growth conditions	Ring stage	-Limited biomass (1-2µm, **Figs. 2A and 3A**), haploid genome -Few enrichment options
	Late stage	-Larger in size (3-10µm, **Fig. 2A**), polyploid genome -Can use magnetic enrichment (**Fig. 1**)
	Mixed stages	-Effects of stage variation on data
	Media batches	-Relevant if using serum-based media formulations
	Blood batches	-Must be recorded and ideally matched within comparisons (**Table S1**) -Useful to assess host contamination levels (**Fig. 3D**)
Additional controls	Uninfected erythrocytes	-Use to identify host metabolites -Does not replace normalization
Enrichment methods	Saponin, other lytic reagents	-Compatible with all stages (**Fig. 1**) -Parasites remain in ghosts (**Fig. 3C)** *-Need improved methods that isolate parasite from host cell*
	Magnetic purification	-Increases parasite to host ratio (**Fig. 1**)
Metabolite Detection	NMR	-Limited metabolite detection but higher confidence
	Mass Spectrometry	-Industry standard for broad detection
Pre-analysis normalization	Cell number normalization	-Can be combined with any post-analysis normalization but requires sample manipulation
Post-analysis normalization	Parasite-derived parameters	-Selection requires knowledge of experimental design (i.e. parasite number or DNA amount)
	Parameters with mixed derivation (host, parasite)	-Can fail to remove undesired noise (**Fig. 2**)
	Internal standards	-Dependent on metabolomics facilities
Centering	Mean	-Standard centering
	Median	-Less sensitive to outliers
	Other	-See (23) for summary of alternative approaches
Scaling	Within group SD	-Requires no additional samples
	Z-scoring	-Requires control samples (can use uninfected erythrocytes)
Statistical analysis	Univariate	-Requires multiple comparison corrections
	Multivariate	-Reveals group differences based on multiple variables
	Machine learning (Random Forest)	-Classification is more stringent than univariate tests, but can identify nonlinear effects

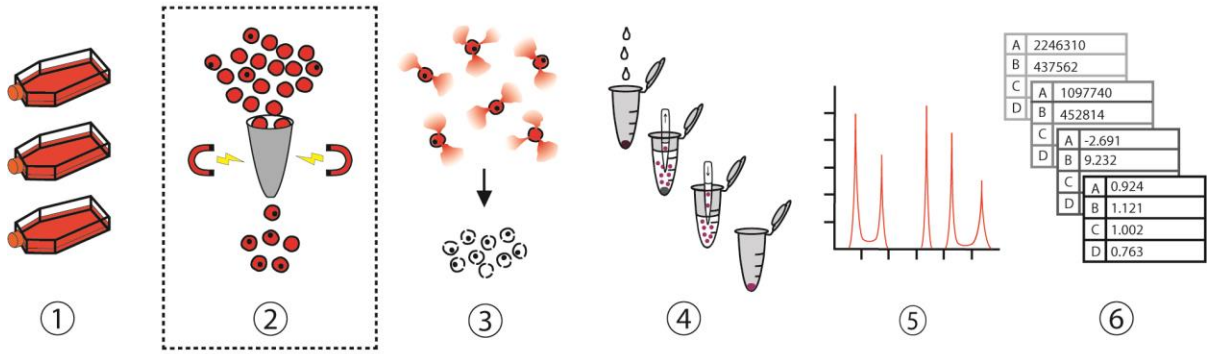
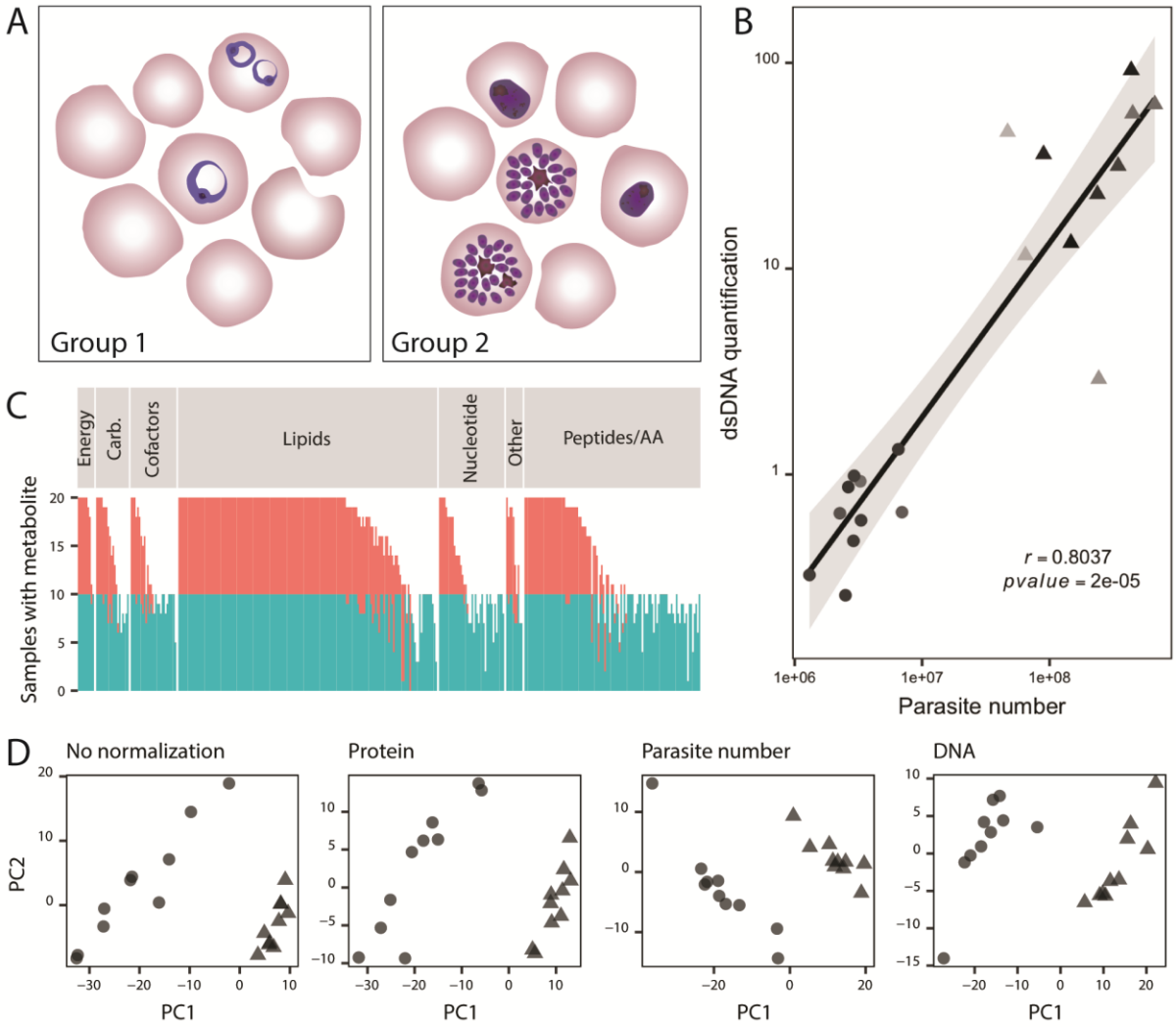
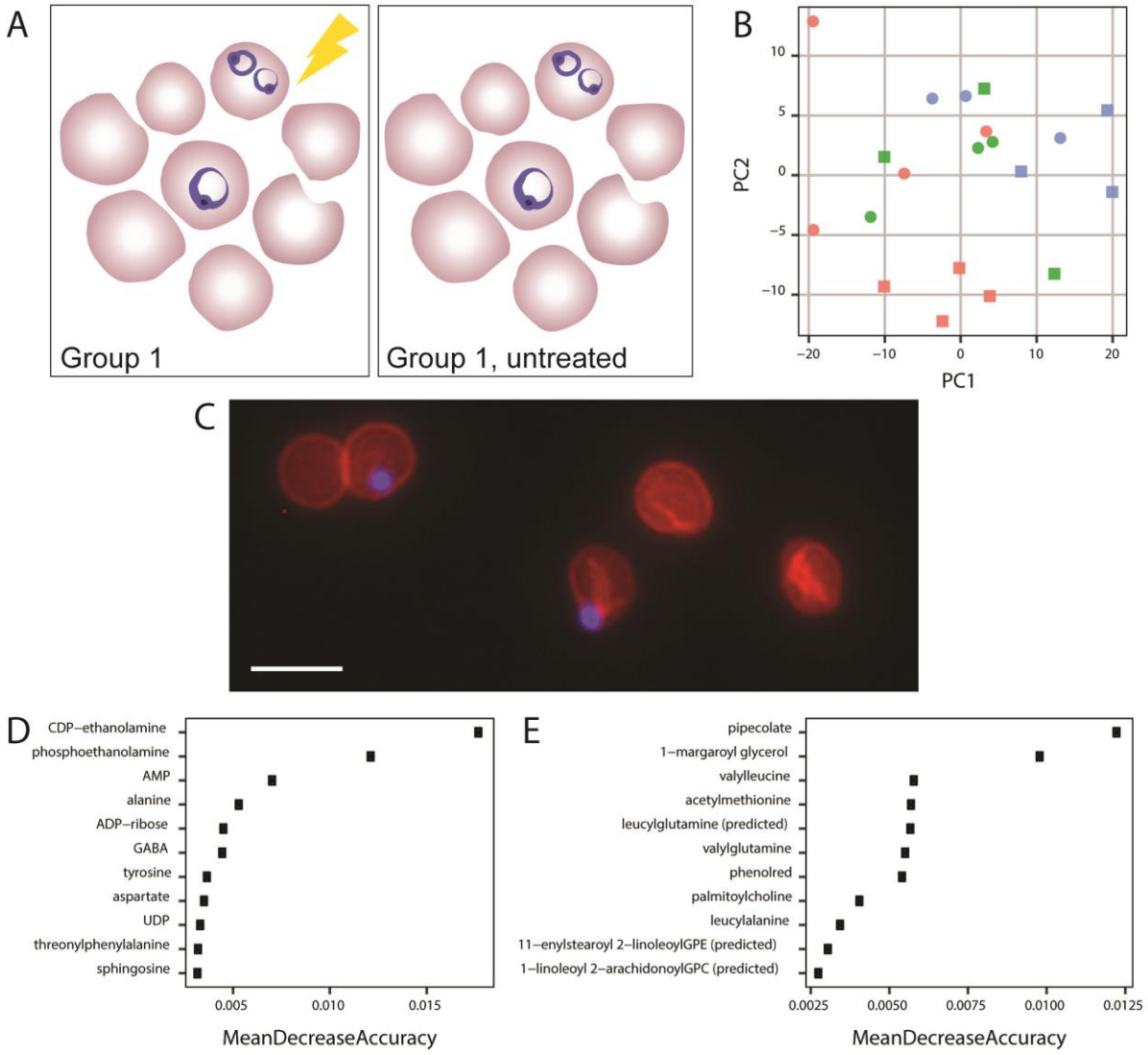


Figure 1. **Metabolite purification and analysis pipeline**. 1) Laboratory-adapted *P. falciparum* clones are cultured in host erythrocytes. 2) If enrichment of late stage parasites is desired (dotted line), cultures can be passed through a magnetic column to retain paramagnetic late stage-infected erythrocytes (black dots inside red circles). Samples for parasite count determination were collected at steps 1 and 2, depending on the sample group (see *Materials and Methods*). 3) Erythrocytes (infected and uninfected) are lysed using saponin, but parasites remain intact (black dots). Wash steps are used to remove hemoglobin and other intracellular erythrocyte contents (red material). Samples for total protein determination were collected at this step. 4) Soluble metabolites (purple dots) are extracted from precipitated protein (grey pellet) using methanol (droplets). Samples for DNA content determination were taken at this step, prior to methanol extraction. 5) Metabolites are detected by liquid chromatography followed by mass spectroscopy. Metabolites are identified by comparison to a library of authenticated standards. 6) Abundance data for each metabolite is normalized to an appropriate parameter (i.e. DNA content or parasite number), log transformed, centered to median, and scaled to variance, prior to employing statistical comparisons.



675 Figure 2. **Comparison 1: Metabolomes are distinct and influenced by**
676 **normalization approach. A. Comparison made throughout Figure 2**. Group 1
677 contains ten samples of early-stage parasites grown in AlbuMAX-based media in three
678 blood batches, treated with antimalarial (see *Materials and Methods* and **Table S1**).
679 These parasites were isolated at a low parasitemia and purified from host material using
680 saponin lysis. Group 2 contains ten samples of late-stage parasites grown in a human
681 serum-based media in four blood batches, treated with antimalarial (see *Materials and*
682 *Methods* and **Table S1**). Group 2 parasites were purified magnetically to achieve high
683 parasitemia and lysed from host cells with saponin. **B. Sample characteristics**.

684 Samples (group 1 in circles and group 2 in triangles) were evaluated for DNA, parasite
685 count, and protein amount prior to analysis. **C. Summary of detected metabolites**. Not
686 all metabolites were detected in each sample. The majority of metabolites detected
687 were lipid species. Sample groups are color coded with group 1 in red and group 2 in
688 blue. A full list of identified metabolites are listed in supplemental data (see Github). **D.**
689 **Normalization affects measured metabolome.** Principle component (PC) analysis
690 was performed prior to normalization (left), as well as using three different normalization
691 methods (left to right, total protein, parasite number, and DNA). Circles indicate group 1
692 samples and triangles indicate group 2 samples. For PC decompositions, see **Table S2**.
693



694 Figure 3. **Comparison 2: Blood batch and antimalarial treatment influence**
695 **metabolomes. A. Comparison made throughout Figure 3**. All samples were grown in
696 AlbuMAX-based media in three blood batches and purified from host material using
697 saponin lysis during the early stage. Group 1 was treated with antimalarial for 6 hours
698 and group 1, untreated, did not undergo treatment (see **Table S1**); samples were
699 matched for blood batch. **B. Metabolome principle component analysis**. PCA of
700 DNA normalized, median-centered metabolomes of early stage parasites from

Comparison 2. Squares indicate antimalarial treated samples and triangles indicate untreated samples. Blood batches are indicated by color. **C. Visualization of erythrocyte ghosts containing parasites**. Fluorescent imaging (40X) reveals parasites (blue, DAPI) retained within erythrocyte ghosts (red, phycoerythrin conjugated CD235a antibody) following saponin treatment. Scale bar represents 10µm. **D. Metabolites predictive of blood batch.** Top ten most predictive variables in the blood batch Random Forest classifier. **E. Metabolites predictive of antimalarial treatment.** Top ten most predictive variables in the antimalarial treatment Random Forest classifier.