nature portfolio

Peer Review File

Population genomics and transcriptomics of Plasmodium falciparum in Cambodia and Vietnam uncover key components of the artemisinin resistance genetic background

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

Recommendation is: Accepted with major modifications

This article by Nayak et al, aimed to characterized additional genetic variations mediating the artemisinin resistant phenotype in the Greater Mekong Subregion. 290 samples were successfully analyzed by WGS and their associated transcriptomes using RNA-seq before treatment and 6 hours after artemisinin intake.

The study reports important data insights on the genetic relatedness of the resistant parasite subpopulation Kel1/Pla1 lineage in SEA. Additional markers for artemisinin resistance are described, N1131I and N821K mutations in PfRad5 and PfWD11 splicing site and transcriptomes are broadly explore to characterize the impact of artemisinin treatment on susceptible and resistant parasites.

Overall impressions
The paper is well written, clear and contains important results which I feel are important to characterized the resistant mechanism to artemisinin derivatives. However, some aspects should be addressed before the manuscript is published in Nature communication.

Major comments:
1. A large part of the result section is dedicated to the description of the transcriptomics data. This part is quite long and hard to read. The authors should reduce this part of the manuscript and try to improve them finding a way to simplify them. By this, it should be easier for the reader to measure the impact of each observations and its exact meaning regarding the initial question, the artemisinin mechanism of action and resistance.
2. In the same way, I would suggest to add in the introduction paragraph dedicated to the state of the art on RNA transcripts other than PCRTs and their function in malaria parasites.
3. Regarding additional genetic markers or other genetic markers associated for artemisinin resistance, the authors mention few similarities between the genetic architecture previously described by Amato et al. at the time of emergence and the one observed with these results.
How to distinct genetically associated markers:
- because of their close vicinity with the main driver, pfk13, and a real implication in the resistance mechanism?
- at the time of emergence because of a geographic signature of the parasite population in this area, independently of resistant mechanism?
Thinking about these and without an isogenic validation, how to interpret the presence of the different markers identified in the chromosome 13?

Minor:
4. Line 66: Africa is mentioned but it could also be interesting to mention South America at this stage and not only at line 80.

5. Line 71: a princeps publication on RSA would be useful.
6. Line 78: actualities regarding this question in Uganda and Eats Africa is important. Please update references according to the last publications from this area.
7. Lines 82-85: In fact, malaria control and elimination programs rely on ACTs, ART + partner drug. The crucial role of the partner drug in ACT efficacy should be mention even if this article is focusing on artemisinin resistance. It is the failure of the partner drug who compromises ACTs efficacies.
8. Line 91: a reference about the impact on fitness cost should be added.
9. Line 502: Please mention the ethical number of approval.
10. Line 523: It would be interesting and easier for the reader to at least mentioned the extraction method/kit used.
11. Figure 1 panel b: Does an error exist in the y axis legend. PC2??
12. Figure 5 panel a: Are the dots properly located. For South America, C580Y isolates so I assume that samples came from all around Amazonia are included. Is the dot should be more in the middle? South Asia in India? wGMS dot more below? east, more below, and right?

Reviewer #2 (Remarks to the Author):

This is an interesting study focused on identifying genes/ expression patterns that are associated with artemisinin resistance. The authors are seeking to identify factors beyond mutations in the Pfkelch 13 gene. The novelty of this study is the authors use matched samples of the same patient before and after artemisinin (ART) treatment to measure RNA species and levels in addition to the WGS of the P. falciparum parasites. Samples were collected from different geographic sites within the Greater Mekong Subregion They also measure the parasite clearance rate (PC1/2) which indicates the level of resistance with a PC1/2 > 5 hours is considered resistant). The authors justify this approach because of the widely observed differences in the impact of Pfkelch 13 mutations in different genetic backgrounds and they reason that additional genes or their expression modulate the Pfkelch13 mutation.

The authors initially analyze the WGS data and it revealed that parasites from east Cambodia are genetically different form parasites from west Cambodia and south Vietnam and generally more diverse. Three distinct clades were identified, however, clades I and II seem to be genetically related. Clade III is exclusive to western Cambodia and consists of mostly wild type parasites at K13. The GWAS studies based on PC1/2>5 identified PfK13 as the primary hit, as has been shown in previous studies. Interesting these data did not support genes discovered in a previous GWAS study (Nature Genetics volume 47, pages 226–234 (2015) on a much larger sample set derived from across the GMS region. The authors single out the polymorphisms in PfRad5 (PF3D7_1343400) on chromosome 13 and a novel WD40 domain-containing protein on chromosome 11 (PfWD11, PF3D7_1138800) that associate with artemisinin resistance by exonic and intronic SNPs, respectively. The chromosome 11 observation is novel and with the associated transcriptional regulation (see below) is potentially of interest for further investigation. It would be useful to know if this mutation occurs in other populations (Africa) where Pfkelch 13 mutations have emerged.

The characterization of the transcriptional profile including alternate splice forms by single cell RNA seq is novel in this context. Several other groups have reported RNA seq based analysis of bulk transcriptional responses to artemisinin treatment or in artemisinin resistant parasites. This was a large data set and the innovation of quantifying alternate splice variants led to some new observations. Furthermore, this opens a new opportunity of research on non-canonical RNA species and their general role in P. falciparum biology.

The terminology could lead to some confusion particularly in the malaria field – the authors define their terms as follows "…maximized the detection of all RNA transcript classes including the canonical protein-coding mRNA transcripts (protein-coding reference transcripts,PCRTs) but also their alternatively spliced variants (altPCRT) and/or overlapping noncoding antisense transcripts (asPCRT)and intergenic ncRNAtranscripts…" PfCRT is the gene encoding the chloroquine resistance transporter and this led to initial confusion on my part in reading the manuscript.

The results point to differences in the quantity of splice variants in two genes identified in the GWAS analysis, WD11 and Rad5. The follow qRTPCR studies gave additional support to changes in the splice variants for WD11 and Rad5. However, 3D7 does not have the SNP in WD11 but the splice variant can still be detected in in vitro cultures. Therefore, the SNP is not needed for the generation of the altPCRT form. It will take more experimental evidence to confirm that this mutation influences the frequency of the two RNA species. In addition, there will have to be experimental evidence for the translation of the altPCRT variant and more understanding on protein function or stability. Also, what is the epistatic relationship between Pfkelch 13 and WD11 mutations.

Expression analysis on PCRTs but also of non- PCRTs including alternative splice variants (altPCRT), anti-sense PCRT (asPCRT) and non-coding PCRTs (ncPCRT) was conducted on the samples before treatment and 6 hours after ART treatment. This is the first study to include all the different RNA species on samples taken from patients before and after ART treatment. Even before ART treatment, there were PCRTs and non PCRTs associated with resistant or susceptible parasites. Upon ART treatment, there is a down regulation many genes associated with cell cycle progression. This is much stronger in the resistant parasites with more genes being affected. These observations are in genes other than those identified in the GWAS study and it is unclear from the presentation if these are significant contributors to the phenotype. The results section is densely written and could benefit from more prioritization of the observations.

The paper would benefit from significant editing, focusing on the main observations relevant to the proposed role in modulating artemisinin resistance.

Minor:
Figure 1b: What are the squares in the PCA plot? In c. it's not easily understandable which circle corresponds to what locus. Please label the circles.

Figure 2. Move the description of the different PCRT classes into a. It might also be good to mention that these data are on the baseline RNA levels.

It might be good to keep the different PCRT classes in different colors (eg PCRT: blue, asPCRT: red, altPCRT: green

Figure 3: Please be more consistent with “resistant and susceptible” and “PC1/2 < > 5h”. The legend and the figure titles do not always correspond. B: Shouldn’t the dots be blue on a beige background to be consistent with the rest of the figure?

Figure 4b “shows up-regulation of PCRT in resistant parasites (left, blue) compared to resistant parasites” I think you mean susceptible parasites in the second instance. Please align sub figures d and e. In the legend there is an excessive “of” in line 1024. C. I don’t understand the different labels at the bottom? Is “prop” for propeller?

Supplementary figures:

7: please provide a scale for the expression
10. Are all of these performed on 3D7? Please mention what the template is.
11 What additional parasites are included in this graph and why were they excluded from main figure 4?

RE: Population genomics and transcriptomics of Plasmodium falciparum in Cambodia and Vietnam uncover key components of the artemisinin resistance genetic background, Nayak et al.

Bozdech and colleagues have been leaders in probing the complexity of genetic determinants of artemisinin resistance (Art-R) in SE Asia. There is much fanfare around mutations in K13 as being the "marker" of Art-R, and all surveillance efforts worldwide are focused solely on this gene. However, there is ample evidence that knowing K13 mutations is at best only part of the story and at worst could be completely misleading in Africa where a major drug resistance crisis looms. This research group is one of very few taking on these hard but important questions and the only group developing powerful approaches to get answers. The large point-of-care sampling that underpins this study is a very powerful resource, and this group is honing and perfecting the analytical approach needed to find the complex determinants of clinical resistance (patient clearance half-life), including a comprehensive and novel use of transcript analysis. Several aspects of this work are impressive, including their layered and integrated approach that includes GWAS, innate transcript levels, drug-induced transcription, and population genetics. Because they classify a wide range of transcripts, the authors are able to identify the previously unrecognized role of alternatively spliced gene products that contribute to the resistance profile. Nevertheless, the current version of the manuscript falls short in several ways. Because the work is pushing the boundaries for the field, the authors must do a better job outlining their rationale and their unique and comprehensive pipeline (currently cryptic and convoluted, and seemingly post hoc, which undermines the value of this work); further, they need to better highlight their adequate rigor and provide more transparency. It can be argued that their extent of validation of gene functions could be better; however, I do not think a full functional analysis is needed. However, a clearer description of the strengths and limitations of this approach, along with a more thoughtful consideration of the next steps that will further solidify this approach to put functional meaning to "genetic background" are needed. This critique is geared to making this work more accessible and convincing to a general audience, and to ensuring that ALL aspects of this work are available in the public domain for deeper follow-up and extensions of this analysis.

General observations: Nayak et al. combine genomic and transcriptomic data generated from patient samples collected at timepoints pre- and post-drug treatment as part of the TACT-CV clinical study of triple artemisinin combination therapies. They seek to clearly elucidate the genetic contributions to a complex quantitative phenotype. These genetic features are in addition to (or in place of) the well-known Kelch13 mutations with functional relevance to the mechanism of resistance, drug action or fitness to be useful as surveillance markers. Ideally, but not clear is whether these findings will be of direct relevance to Sub-Saharan Africa. The authors use population structure analysis to separate SE Asia parasites into three clades associated with varying levels of clinical resistance and drug selection histories. GWAS identified SNPs associated with clinical resistance, including PfRAD5 and PfWD11. Further transcription analysis supported and elaborated the role of these two genes, particularly highlighting alternative transcripts as a component of the parasites' mechanism of resistance/response to drug exposure. Deeper investigation of available sequence data strongly suggests a shared selection history between WD11 and K13, implying a covariant role in resistance, e.g. some kind of

evolutionary partnership with K13 that will require future study to elaborate mechanisms. The study design and data are valuable and warrant a high-level publication in Nature Communications. However, aspects of the analysis must be strengthened with an emphasis on transparency and increasing the full potential to the community of this patient clinical data for reanalysis, deeper analysis and validation of this expanded view of the genetic background of Art-R.

Major comments:

1. The clinical phenotypes are incredibly important but are getting lost. A panel in figure 1 showing the distribution of clearance times with symbols representing clades and colors for location being consistent with other panels of figure 1 would help. Include an additional indicator for K13 mutation status. A similar format to figure 5c would satisfy this concern. Showing a geographic clade distribution would set up figure 5c as building from figure 1 and would help clarify the discussion. The distribution shown in supplemental figure 3 does not do justice for the importance of the clinical phenotypes and the proposed additional panel would be a complement to supplemental figure 3.

2. The authors determined 3 clades best represent population structure of the parasites used in the study given principal component analysis and neighbor-joining tree analysis. Include further rationale to arrive at the value of K=3. Generally, elbow method is used where point of inflection denotes the value of K. Besides PCA, an MDS or K-means clustering would further confirm the variance observed and separation of population structure. A scree plot would show how this analysis impacted the determination of 3 clades and how much variation is being accounted for by each PC after the first three.

3. Figure 1c shows a smaller 4^{th} clade of roughly 6 parasites between clades 2 and 3 that would be interesting for further Tajima's D analysis. Why did the authors not note the smaller clade in the results section of this analysis? Additionally, Figure 1c has at least 4 'square' data points that do not match legend symbols for the three clades. Include a distinction for the square points in the figure legend.

4. Tajima's D shows balancing selection pressure on clade 1, active selection pressure on clade 2 and no current selection pressure on clade 3. To fully show the selections (or lack of selection) include the Tajima's D value for each SNP used in the analysis as a supplemental table and as a supplemental figure to show what across the genome is being selected for each of the three major clades, similar to figure 5 in https://doi.org/10.1177/1176934321999640 with highlighted genes of interest. What are the Tajima's d values of GWAS hits in figure 1e? Are they preferred in the observed selection sweeps or balancing selection?

5. The authors acknowledge the peak multimodality of clade 2 as part of the interpretations for ongoing selection. Could the multiple peaks also be indicative of the clades warranting separation into more subgroups rather than an overall Clade 2 being

under active selection pressure? Clade 1 also shows a multipeak profile but is described as “skewed” which does not appear to be the most accurate description of the clade 1 distribution. Be clearer about the description of clade 1.

6. To further show the commonality between clades and ancestral vs admixed natural of them it would be useful for reader interpretation to include an **identity by descent analysis (IBD)** for the parasites used in this analysis.

7. What is the rationale for using the first 10 principal components for GWAS? A scree plot for PCA should be generated and included as a supplemental figure with greater description in the methods for this choice of 10 to bolster the GWAS and population structure analyses.

8. It is stated that roughly 300 parasite samples are sufficient to be representative for the overall transmission pattern of infection in the eastern part of the GMS (cite #48) to support this claim. To fully articulate this point the authors should include a power analysis given this sample size and include a supplemental figure plotting power vs minor allele frequency and effect size. This analysis would demonstrate the ability of this study to identify genes of smaller affect size through GWAS that would constitute the genetic background of resistance.

9. The 1*10-5 threshold for significance is lower than the commonly observed threshold for GWAS of 1*10-8 Clarify this choice of threshold. There is good precedent for lower thresholds, e.g. Cerqueira et al. 2017 and Wendler JP et al., https://doi.org/10.1371/journal.pone.0096486. Power to detect is often countered by the value of available clinic samples (as the authors note). These clarifications will strengthen the paper.

10. **Provide <u>the full summary statistics for the GWAS analysis (the current supplementary material only lists significant markers and no effect size). A complete summary of statistics with necessary statistical parameters of both significant and nonsignificant hits will make this a more valuable resource.</u>** This will be help establish a new standard in a high impact journal.

11. Provide the quantile-quantile (QQ) plots for the GWAS before and after correction for population structure as a supplemental figure to demonstrate that population stratification was adequately accounted for. The plots should also include the genomic inflation factor (λ).

12. To the point of power for detection of significant SNPs, in figure 1e there is a second SNP on chromosome 11 at the opposite end compared to PfWD11 that is approaching the significance threshold, more significant than other non-significant SNPS. It is not listed in

the text or supplemental table 2. An extended value of the approach the authors take is to consider the full data as a source for further follow-up studies.

13. K13 is well known to be the major gene for artemisinin resistance in Southeast Asia, as the authors articulate well. A comprehensive analysis of the SNPs identified in linkage disequilibrium with $K13^{C580Y}$ would be useful. The authors touch on this question by their investigation of pf7 parasites and showing PfWD11 has been selected along with K13 across time. Extending this analysis to include all SNPs would further strengthen the 'pipeline' and also emphasize the outlier nature of PfWD11. A secondary GWAS analysis including K13 as a covariate as a supplemental figure could also answer emphasize this point.

14. The authors provide hypotheses for the biological role of PfWD11 heavily relying on citation #66. That publication does a fair amount of work articulating several proteins as part of the TAF1/BDP5 complex (BDP5, TAF1, TAF2, TAF10, PfWD11, PF3D7_0824800, PF3D7_1032700, PF3D7_1251800, PF3D7_0514900, PF3D7_1361200). While I agree that detailed validation of the biological function is beyond the scope of this manuscript, some deeper exploration within the data presented here of the other proteins hypothesized to be in the TAF1/BDP5 complex listed in cite #66 is warranted since PfWD11 is a key result of the study. How does the expression of these other 9 genes compare to PfWD11 and are any of them included in other analyses shown for PCRT/non-PCRT as correlated with PC1/2?

15. For the qPCR validation experiment it is unclear if it was done with multiple biological replicates or only technical triplicate of the qPCR for a single biological replicate. That distinction should be made clear in the methods and figure legend.

16. The enrichment analysis of PCRTs and non-PCRTs is presented for only down regulated genes. Theoretically upregulated genes would include proteasome components (previously highlighted by this group). State also that deceleration of the IDC by transcriptional down regulation has also already been demonstrated in early work by this group. Additionally, line 285 states enrichment analysis identified a wide array of processes related to parasite "lifecycle progression". By this interpretation, given the array of GO terms found, almost any GO term found could be seen as part of the parasite's life cycle. The interpretation is too broad to be meanginful. Also, the use of "lifecycle" throughout the manuscript is too broad. At points where the IDC is being referenced the "asexual cell cycle" or IDC should be used whereas "lifecycle" should be used in reference to transition between hosts as used in reference to lines 482-483.

17. For consistency between text and visual representation in figures, everywhere that states west vs east separation of Cambodia in the text and figures should be written, listed and described in order of west then east. This will help in interpreting figure 1.

18. The overall value of the impactful dataset will be greatly enhanced by including the scripts used for the analysis with results available through a github repository rather than available upon request. This is especially important given the analysis uses in-house scripts for detection of nonPCRTs which will be invaluable for guiding future studies.

Minor comments:

1. In the introduction be clear about “eQTL-like” acting mutations as was described in citation 25 around lines 99-100. This concept is important for the understanding the results and for the specific genes singled out for direct analysis. This will make it easier on the reader to not have to read the cited paper in depth for a full understanding.

2. For an ideal population the average Tajima’s D is expected to be around 0 but in figure 1d clades II and III appear to have an average Tajim’s D greater than 0. Can the authors provide an explanation as to why the average Tajima’s D deviates from 0?

3. Typical GWAS includes a calculation for the Hardy Weinberg equilibrium for population used within the study. Since there is a selection sweep occurring for clade 2, do the authors think there could be a change to the Hardy Weinberg equilibrium. Inclusion of the full summary statistics may also answer this question and alleviate this concern.

4. A full interpretation of the population structure of parasites would benefit from the inclusion of a kinship matrix and details of the LD pruning of the SNPs.

5. Is the expression pattern of selected genes in supplemental figure 7 generated from the transcription of the resistant parasites in the study or from a reference time series experiment? If it is the former, how were full timeseries extracted from only two timepoints of 0hr and 6hr, pre and post treatment?

6. Abbreviation of patient clearance half-life is not consistent throughout the manuscript (i.e., PC1/2 in some places and $PC^{1/2}$ in others).

7. Figure 1b has PC1 labeled as both axes.

8. Lines 74-77: Specify the increase of the R561H K13 mutation are in Africa countries. I recognize the citations for this sentence are from studies using African parasites but for flow and plainly clear transition from GMS to Africa made in this paragraph, the minor edit would be beneficial.

9. Lines 156 and 157 state transcriptome samples were generated for 310 and 300 parasites at timepoints 0hr and 6hr, respectively. However, within the methods it is stated only 284 and 251 transcriptome samples for 0hr and 6hr timepoints were suitable for analysis. With this sentence or within the results section for the transcriptome analysis this point should be restated to reiterate the samples used in GWAS vs TPAS

even though it is stated in line 232. Also line 232 states 282 samples for 0hr instead of 284 as in the methods and line 233 states 252 for 6hr instead of 251 as in methods, are these typos?

10. Line 189: Why is “(any)” in parentheses?

11. Line 215: “an SNP” should be “a SNP”.

12. Line 318: A space is needed between “PfWD11and”.

13. Lines 379-380: The percentages for each mutation listed in the text appear to be flipped from what is shown in figure 5d.

14. Line 436: Why is “architecture” in quotations? This might imply a dissatisfaction of the cited article’s (#18) analysis and claim for investing genetic architecture of resistance? If so, a clear articulated rationale of its limitations for interpreting genetic architecture of resistance would be necessary. The explanation would also be welcomed as incremental improvement to the understanding of biological phenomenon based on new or repeated studies in the basis of science.

15. Lines 457-458: The citation (67) does not support the claim the TAF complex regulates the p. falciparum life cycle. The cited article is a bioinformatic of the parasite genome searching for transcription factors and doesn’t show data describing their biological role in the parasite.

16. Line 629: “(a) (a)” should be “(a)”

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

Recommendation is: Accepted with major modifications

This article by Nayak et al, aimed to characterized additional genetic variations mediating the artemisinin resistant phenotype in the Greater Mekong Subregion. 290 samples were successfully analyzed by WGS and their associated transcriptomes using RNA-seq before treatment and 6 hours after artemisinin intake. The study reports important data insights on the genetic relatedness of the resistant parasite subpopulation Kel1/Pla1 lineage in SEA. Additional markers for artemisinin resistance are described, N1131I and N821K mutations in PfRad5 and PfWD11 splicing site and transcriptomes are broadly explore to characterize the impact of artemisinin treatment on susceptible and resistant parasites.

Overall impressions

Major comments:

1. A large part of the result section is dedicated to the description of the transcriptomics data. This partis quite long and hard to read. The authors should reduce this part of the manuscript and try to improve them finding a way to simplify them. By this, it should be easier for the reader to measure the impact of each observations and its exact meaning regarding the initial question, the artemisinin mechanism of action and resistance.

We thank the reviewer for this suggestion, and in the new version of the manuscript, we made a concerted effort to streamline the transcriptomics Results section and emphasized the main observations with more clarity. We tried to focus on the key findings, passing some of the less direct discoveries to the supplementary material.

2. In the same way, I would suggest to add in the introduction paragraph dedicated to the state of the art on RNA transcripts other than PCRTs and their function in malaria parasites.

We also thank the reviewer for this suggestion, but with all due respect, we introduce this paragraph with additional citations in the discussion section. In our view, Introduction is already quite substantial, and such a paragraph might be distracting.

The last paragraph of Discussion was reworded **line 511**: *The discovery of the non-aPCT and lncRNAs in P, falciparum dates to as early as 1991 and subsequently 2001 respectively82,83. It was, however, not until the advancement of RNA-Seq that these RNA species were captured at the genome-wide level, **…** With the advances in long-read and single-molecule sequencing, it is clear now that the complexity of the P. falciparum transcriptome goes far beyond protein-coding mRNA79,94. Here, we show that both non-aPCTs and lncRNAs exhibit marked variations in P. falciparum parasites"*

3. Regarding additional genetic markers or other genetic markers associated for artemisinin resistance, the authors mention few similarities between the genetic architecture previously described by Amato et al. at the time of emergence and the one observed with these results. How to distinct genetically associated markers: - because of their close vicinity with the main driver, pfk13, and a real implication in the resistance mechanism? - at the time of emergence because of a geographic signature of the parasite population in this area, independently of

resistant mechanism? Thinking about these and without an isogenic validation, how to interpret the presence of the different markers identified in the chromosome 13?

If we understand correctly, the main concern is the SNPs identified by GWAS that reside on chromosome 13 near PfK13. The main question is whether their associations with resistance reflect their functional involvement or, rather, a physical meiotic linkage. Although the possibility of the meiotic linkage cannot be entirely ruled out for all identified chromosome 13 SNPs, several factors suggest that at least some of these, notably the two PfRad5 SNPs, are true markers. **First,** PfRad5 is positioned three genes away from PfK13 with a large number of other SNPs in this locus (between PfK13 580Y and the PfRad5 N1131I and N821K with no association with artemisinin resistance or any other type of signature of selection. **Second**, similar to PfWD11, which is on a different chromosome, the two PfRad5 SNPs show a similar rise in frequency in the GMS, similar to PfK13. This is not the case for most SNPs on chromosome 13 (including those within the PfK13-PfRad5 locus) or any other chromosome. Third, the two PfRad5 SNPs are linked with differential mRNA levels for both the canonical and alternatively spliced form that is collectively associated with artemisinin resistance. Although we agree with the reviewer that transgenic analysis will be ultimately necessary to prove the PfRad5 involvement in artemisinin resistance, such studies are beyond the scope of this study (please see the similar argument for PfWD11 below). Nevertheless, results in this study provide a confident testable hypothesis for the reverse genetic studies of PfRad5.

In the new version of the manuscript, we show the rising frequencies of the PfRad5 mutations in the GMS throughout the last decade in **Figure S12 C**. Moreover, in **Figure S18**, we discuss the increased linkage levels at chromosome 13; however, also further triage the SNPs by PfK13 LD and covariant studies and frequency profiles.

This is also noted in the discussion part **line 484**:

*"Indeed, PfRAD5_N1131I and PfWD11 i2+30 show strong associations with PfK13 mutat phenotype when studied by GWAS of linkage disequilibrium (LD) studies in the TACT-CV cohort (**Figure S18a, b** and **Table S8**). Moreover, the temporal profile of their frequency progression in the GMS since 2009 also shows a high correlation with PfK13 C580Y, which classifies both mutations into a small group of outlier mutations with a high link with artemisinin resistance (**Figure S18c and d**). Pending validations of their biological relevance, these may now represent more suitable markers of artemisinin resistance in the future."*

Minor:

4. Line 66: Africa is mentioned but it could also be interesting to mention South America at this stage and not only at line 80.

We agree that the emergence of artemisinin resistance in other parts of the world is of significant concern, including South America. We added an additional sentence stressing the fact that PfK13 mutations were emerging in Guyana as early as 2016.

Line 68:

"In South America, several independent events also occurred that led to the emergence of resistance-associated markers, particularly in Guyana, as early as 20166,."

5. Line 71: a princeps publication on RSA would be useful.

If we are not mistaken, the princeps publication is Witkowski et al 2013. This reference is now used.

6. Line 78: actualities regarding this question in Uganda and Eats Africa is important. Please update references according to the last publications from this area.

There appear to be three more publications reporting the recent emergence of artemisinin resistance in Uganda, Eritrea, and Tanzania. These are now included along with an addition note about South America on **line 80:**

"Occurrence of artemisinin resistance was also confirmed recently in other parts of sub-Saharan Africa, such as Ethiopia14, Uganda4, Eritrea15 and Tanzania16. Together with the independent emergence in South American Guyana reported earlier6,7, these Sub-Saharan occurrences stress the fact that artemisinin resistance can emerge autonomously in essentially any other part of the malaria-endemic world17."

7. Lines 82-85: In fact, malaria control and elimination programs rely on ACTs, ART + partner drug. The crucial role of the partner drug in ACT efficacy should be mention even if this article is focusing on artemisinin resistance. It is the failure of the partner drug who compromises ACTs efficacies.

Indeed, it is correct to note that the main concern of artemisinin resistance is its ability to drive resistance to the partner drug and, as such, the failure of the entire ACT(s). We mention this on **Line 84:**

"The main concern is the fact that artemisinin resistance compromises the overall efficacies of essentially all ACT types, allowing the emergence of resistance alleles to the partner drugs, which was well documented for mefloquine19 and piperaquine20. Given that all current worldwide malaria control and elimination programs rely on the ACTs; it is imperative to understand the nature of artemisinin resistance, particularly all components of its polygenetic molecular."

8. Line 91: a reference about the impact on fitness cost should be added.

In the new version of the manuscript, we added a sentence hypothesizing the fitness cost effect on the selection of additional artemisinin resistance alleles.

Line 90:

"It is feasible to expect that other, currently unknown, components of artemisinin resistance will be in a delicate balance driven by the positive selection of resistance on one side and negative impact on parasite fitness on the other, as it was demonstrated for several of the PfK13 nonsynonymous SNPs such as C580Y and M579I21."

9. Line 502: Please mention the ethical number of approval.

We have added the following lines to the **line 524 of the Methods**:

"The trial was approved by the National Ethics Committee for Health Research (NECHR; Phnom Penh, Cambodia); the Ethical Committee, Hospital for Tropical Diseases (Ho Chi Minh City, Vietnam; and Oxford University Tropical Research Ethics Committee (OXTREC; Oxford, UK). Oxford University was the study sponsor. Monitoring was done by the Mahidol-Oxford Tropical Medicine Research Unit (MORU;

Bangkok, Thailand) and the Oxford University Clinical Research Unit (Ho Chi Minh City, Vietnam). Ethics approval was obtained from the NECHR, Cambodia (NECHR 0042); the Ethical Committee, Hospital for Tropical Diseases, Vietnam (1096); and OXTREC"

10. Line 523: It would be interesting and easier for the reader to at least mentioned the extraction method/kit used.

By all means, we wish to be completely transparent about all methodologies, hoping that this could be used by other groups. We regret this omission in the first version of the manuscript and added following lines to the Methods:

Line 551

"Briefly, all DNA samples have been extracted from white blood cell-depleted samples using the QiaSymphony Investigator kit with an extraction volume of 200 uL."

11. Figure 1 panel b: Does an error exist in the y axis legend. PC2??

This panel was redrawn with the classification of the clades. The is no error that we could spot.

12. Figure 5 panel a: Are the dots properly located. For South America, C580Y isolates so I assume that samples came from all around Amazonia are included. Is the dot should be more in the middle? South Asia in India? wGMS dot more below? east, more below, and right?

This panel was also redrawn incorporating the phenotypic information as suggested by reviewer 3. All the suggested positioning was carefully considered as well.

Reviewer #2 (Remarks to the Author):

This is an interesting study focused on identifying genes/ expression patterns that are associated with artemisinin resistance. The authors are seeking to identify factors beyond mutations in the Pfkelch 13 gene. The novelty of this study is the authors use matched samples of the same patient before and after artemisinin (ART) treatment to measure RNA species and levels in addition to the WGS of the P. falciparum parasites. Samples were collected from different geographic sites within the Greater Mekong Subregion They also measure the parasite clearance rate (PC1/2) which indicates the level of resistance with a PC1/2 > 5 hours is considered resistant). The authors justify this approach because of the widely observed differences in the impact of Pfkelch 13 mutations in different genetic backgrounds and they reason that additional genes or their expression modulate the Pfkelch13 mutation.

We are glad to see that this reviewer recognizes the importance of the genetic background in manifesting the artemisinin resistance phenotype. Indeed, the main rationale of our study is to identify the individual components of the genetic background that contribute to resistance directly or indirectly.

The authors initially analyze the WGS data and it revealed that parasites from east Cambodia are genetically different form parasites from west Cambodia and south Vietnam and generally more diverse. Three distinct clades were identified, however, clades I and II seem to be genetically related. Clade III is exclusive to western Cambodia and consists of mostly wild type parasites at K13. The GWAS studies based on PC1/2>5 identified PfK13 as the primary hit, as has been shown in previous studies. Interesting these data did not support genes discovered in a previous GWAS study (Nature Genetics volume 47, pages 226–234 (2015) on a much larger

sample set derived from across the GMS region. The authors single out the polymorphisms in PfRad5 (PF3D7_1343400) on chromosome 13 and a novel WD40 domain-containing protein on chromosome 11 (PfWD11, PF3D7_1138800) that associate with artemisinin resistance by exonic and intronic SNPs, respectively. The chromosome 11 observation is novel and with the associated transcriptional regulation (see below) is potentially of interest for further investigation. It would be useful to know if this mutation occurs in other populations (Africa) where Pfkelch 13 mutations have emerged.

We thank the reviewer for recognizing these observations. To clarify this further. Our study, indeed, did not replicate the associations between artemisinin resistance and the five SNPs detected in the original study by Miotto 2025. This indicates that the artemisinin resistance phenotype underwent a broad evolutionary process likely underlined by the C580Y selective sweep. This sweep is believed to be driven by the administration of artesunate piperaquine (ACT) in that region, driving the selection of PfK13 C580Y and CNV in plasmepsin/III locus; the Kel1/Pla1 lineage. It is feasible to suggest that the five architecture SNPs were eliminated during this process. Nevertheless, the two exact mutations in the PfRAad5 detected here WERE detected in the Miotto study (presented in supplementary material classified as *DNA Helicase* gene). This suggests that the PfRad5 mutations survived the Kel1/Pla1 selective sweep, further stressing their relevance. Similarly, the intronic PfWD11 mutation was also present in the GMS parasites before the sweep and rose in its frequency along C580Y. This mutation was not detected in the original Miotto study as these focused exclusively on exonic SNPs.

Based on the currently available *P. falciparum* genome information in the Pf7K database, the frequency of both PfRad and PfWD11 mutant alleles outside of the GMS are <2%, similar to PfK13 SNPs. However, the Pf7K contains data collected prior to the emergence of artemisinin resistance on the African continent. We agree with the reviewer that it will be very interesting to study these mutations in the current time when resistance is progressively emerging and spreading in Rwanda, Uganda, Ethiopia, and other areas. We believe that this publication will provide a foundation for such studies.

The characterization of the transcriptional profile including alternate splice forms by single cell RNA seq is novel in this context. Several other groups have reported RNA seq based analysis of bulk transcriptional responses to artemisinin treatment or in artemisinin resistant parasites. This was a large data set and the innovation of quantifying alternate splice variants led to some new observations. Furthermore, this opens a new opportunity of research on non-canonical RNA species and their general role in P. falciparum biology.

To our knowledge, this is the first study showing non-canonical RNA species in natural *P. falciparum* infections. Even in that sense, we believe that our study is groundbreaking. We thank the reviewer for recognizing this.

The terminology could lead to some confusion particularly in the malaria field – the authors define their terms as follows "…maximized the detection of all RNA transcript classes including the canonical proteincoding mRNA transcripts (protein-coding reference transcripts,PCRTs) but also their alternatively spliced variants (altPCRT) and/or overlapping noncoding antisense transcripts (asPCRT)and intergenic ncRNAtranscripts…" PfCRT is the gene encoding the chloroquine resistance transporter and this led to initial confusion on my part in reading the manuscript.

We thank the reviewer for pointing this out. We change the terminology to *a, as, altPCT* instead of *PCRT*.

The results point to differences in the quantity of splice variants in two genes identified in the GWAS analysis, WD11 and Rad5. The follow qRTPCR studies gave additional support to changes in the splice variants for WD11 and Rad5. However, 3D7 does not have the SNP in WD11 but the splice variant can still be detected in in vitro cultures. Therefore, the SNP is not needed for the generation of the altPCRT form. It will take more experimental evidence to confirm that this mutation influences the frequency of the two RNA species. In addition, there will have to be experimental evidence for the translation of the altPCRT variant and more understanding on protein function or stability. Also, what is the epistatic relationship between Pfkelch 13 and WD11 mutations.

We can not agree more with the reviewer that experimental validations of these (or any other) mutant alleles associated with artemisinin resistance are absolutely crucial for drawing the final conclusion. The final goal for such a definitive study should be a complete reverse genetics where mutations are introduced into *P. falciparum* wild-type strains followed up by thorough phenotypic characterizations. Unfortunately, given the challenging nature of *P. falciparum* transgenic system, we wish to argue that such studies are beyond the scope of this work.

Nevertheless, in this new version of our manuscript, we provide additional evidence for PfWD11 as a putative contributing (modifier) gene of artemisinin resistance.

First, there is suggestive evidence for PfWD11 to play a role in artemisinin resistance that comes from two recent publications.

(i) Mass spectrometry omics study by Mok et al. (2021) showed significant downregulation of PfWD11 protein levels in CAM3.II isogenic pfk13 R539T and C580Y mutants at the trophozoites stages when compared to their isogenic wild-type control line. This provides independent evidence of lower PfD11 levels in artemisinin resistant parasites.

(ii) PiggyBac-based saturation mutagenesis indicates that PfWD11 gene is essential for parasite intraerythrocytic development. However, transposon disruption of its intergenic region between KIC5 3'UTR (downstream neighboring gene) and PfWD11 5'UTR (774bp and 734bp distance from each gene, respectively) sensitized parasites to DHA but also heat shock, oxidative stress and enhanced gametocyte conversion. Another single disruption in long non-coding RNA sequence downstream of the PfWD11 transcription start site was shown to sensitize the parasite to DHA, lumefantrine, heat shock, oxidative stress and proteasome inhibitor bortezomib. Altogether, this indicates that disruption of the regulatory network in the PfWD11 gene can affect the susceptibility of *P. falciparum* to artemisinin along with other stresses.

Second, we carried out follow-up RT-PCT measurements of the canonical and alternatively spliced form of PfWD11 transcripts to address the relationship between the PfWD11 intronic SNP and the splicing event. For that, we utilized two culture-adapted artemisinin-resistant *P. falciparum* isolates IPC4912 and IPC3445 the reference 3D7 strain as a control While IPC4912 carries a mutated form of PfK13 (I543T), IPC3445 carries PfK13 C580Y but also PfWD11i2+30C>T. *Note that both IPC strains were collected in the GMS at the time of artemisinin resistance emergence, which further supports the existence of the mutated PfWD11.*

Crucially, we observed that the alternatively spliced as well as the canonical form of the pfWD11 transcript are both significantly upregulated after DHA treatment in IPC3445, which carries the mutated form of PfWD11. In 3D7 and IPC4912 (wtPfWD11), the alternative splice form remained constitutive with only small, statistically insignificant changes. This result is consistent with the *in vivo* transcriptomes in which the alternatively spliced form was also upregulated after 6hr treatment in parasite samples with PfWD11i2+30C>T.

Third, to support the RNA-based observations, we carried out Western blot analyses of PfWD11 with a polyclonal antibody that was raised against the protein region upstream of the truncation event encompassing the C terminal end of the predicted TAF5-like domain and N-terminal end of the WD40 domain (Supplementary figure SB, Methods). Overall, we observed highly distinct patterns of the PfWD11 truncated forms between the three strains. While in the 3D7, no degradation/truncated protein products were detected, the two IPC strains each exhibited a distinct pattern. Interestingly, in the IPC3445 (PfWD11 mutant), there was one isoform (short isoform 2~52kda) that appeared inducible by DHA treatment in a dose and time-dependent fashion similar to the transcripts. In all three strains, we observed decreases of the full-length PfWD11 upon the DHA treatment, which were significantly more pronounced in IPC3445.

Taken together, pending full elucidation of PfWD11's biological role, we believe that these results provide compelling evidence of this factor playing a role in artemisinin resistance and present a strong case for initiating reverse genetics efforts that would substantiate this hypothesis. In future studies, it will be interesting to further the value of these mutations as new markers of artemisinin resistance.

It is worth noting that it took more than half a decade between the discovery of PfK13 in vivo *and partial* in vitro *characterizations, highlighting the challenging (and costly) nature of such work.*

In the new version of this manuscript, we present these three lines of evidence as follows:

First: The discussion on crossreferencing our PfWD11 observations with the previous studies is presented in Discussion **Line 461:**

”*Based on a saturation mutagenesis by transposon study, PfWD11 gene is essential for parasite intraerythrocytic development, playing a role in stress responses, gametocyte production and potentially artemisinin resistance phenotype72. Moreover multiomics study by Mok et al. et al 2021* **…** *Another single disruption identified by this study occurred in a long non-coding RNA downstream from the PfWD11 transcription start site sensitizing parasite to DHA, lumefantrine, heat shock, oxidative stress and proteasome inhibitor bortezomib.”*

Second, the RT_PCR validations of alternative splicing in three strains is presented in **Figure 4d** and the corresponding text **line 344.:**

”*Moreover, a similar transcriptional trend can be observed in two culture-adapted artemisinin resistant Cambodian strains IPC4912 carrying pfk13I543T and PfWD11wt and IPC3445 carrying pfk13C580Y and PfWD11i2+30C>T genotypes, respectively. These two strains were collected by the Pasteur Institute, Cambodia, between 2010 and 201310 and are publicly available at www.beiresources.org. Here, we have observed a dose-dependent induction of both aPCT and altPCT isoforms in IPC3445, with the most significant values in the highest dose of the drug (Figure 4d). This*

induction seems specific to the PfWD11i2+30C>T-carrying IPC3445 strain, as essentially no change in both transcript levels was observed in IPC4912 and 3D7."

Third: Western blotting is presented in Figure 4e-f and the corresponding text **line 355:**

"*Next, we carried out Westen blot analysis to test the effect of this aberrant splicing on the overall levels of PfWD11 protein (Figure 4e-f). Using a polyclonal antibody targeting the PfWD11 protein region upstream of the putative truncation, we detected the full-length protein product (~250kDa) in all three strains grown under normal conditions or after varying DHA treatments(s) (see Figure 4e). ... The abundance of all truncated protein products was insensitive to the DHA treatment, except a ~52kDa protein band that exhibited dose and time-dependent induction upon DHA treatment (Figure 4f).*"

Expression analysis on PCRTs but also of non- PCRTs including alternative splice variants (altPCRT), antisense PCRT (asPCRT) and non-coding PCRTs (ncPCRT) was conducted on the samples before treatment and 6 hours after ART treatment. This is the first study to include all the different RNA species on samples taken from patients before and after ART treatment. Even before ART treatment, there were PCRTs and non PCRTs associated with resistant or susceptible parasites. Upon ART treatment, there is a downregulation of many genes associated with cell cycle progression. This is much stronger in the resistant parasites with more genes being affected. These observations are in genes other than those identified in the GWAS study and it is unclear from the presentation if these are significant contributors to the phenotype. The results section is densely written and could benefit from more prioritization of the observations.

The paper would benefit from significant editing, focusing on the main observations relevant to the proposed role in modulating artemisinin resistance.

Similar to comment 1 of reviewer 1, we streamline the text, focusing on the main discoveries, including the mutations in PfRad5 and PfWD11 as potential new markers of artemisinin resistance. The broad downregulation of genes outside of the GWAS hits has also a potential to contribute to artemisinin resistance as these mirror the general deceleration of the IDC that was also previously associated with resistance. nIn the new version of the text we made a concerted effort to clarify these points.

Minor:

Figure 1b: What are the squares in the PCA plot? In c. it's not easily understandable which circle corresponds to what locus. Please label the circles.

The plot was redrawn, including new classifications of the clades.

Figure 2. Move the description of the different PCRT classes into a. It might also be good to mention that these data are on the baseline RNA levels.

This was added to the title of the figure legend.

It might be good to keep the different PCRT classes in different colors (eg PCRT: blue, asPCRT: red, altPCRT: green

This has been now the main and supplementary figures with *aPCT*: blue, *asPCT*: green and *altPCT*: red.

Figure 3: Please be more consistent with “resistant and susceptible” and “PC1/2 < > 5h”. The legend and the figure titles do not always correspond. B: Shouldn’t the dots be blue on a beige background to be consistent with the rest of the figure?

With the redrawing of this figure we made all the efforts to be more consistent.

Figure 4b “shows up-regulation of PCRT in resistant parasites (left, blue) compared to resistant parasites” I think you mean susceptible parasites in the second instance. Please align sub figures d and e. In the legend there is an excessive “of” in line 1024. C. I don’t understand the different labels at the bottom? Is “prop” for propeller?

This figure was reworked, including all suggestions.

Supplementary figures:

7: please provide a scale for the expression

Scale added to (now) Figure S12.

10. Are all of these performed on 3D7? Please mention what the template is.

This is now clarified in the figure legend, even though this was implied from the main text. Given that more strains were added to the new version of the manuscript, we made this absolutely clear.

11 What additional parasites are included in this graph and why were they excluded from main figure 4?

These are a subset of the TACT-CV samples randomly chosen. This is was mentioned and ow more clarified in the the main text **line 338**.:

“*qPCR of 24 paired (0hr-6hr) parasite isolates for both resistant (12) and susceptible (12) groups from the TACT-CV cohort further substantiate the RNA-Seq results detecting a statistically significant increase of PfWD11 altPCT in the resistant group (p-value 0.0008). The aPCT of PfWD11 followed a similar trend with lower statistical significance for fold-change when these 12 susceptible and 12 resistant parasites were compared (p-value 0.004) (**Figure S11**).”*

Reviewer 3.

RE: Population genomics and transcriptomics of Plasmodium falciparum in Cambodia and Vietnam uncover key components of the artemisinin resistance genetic background, Nayak et al.

Bozdech and colleagues have been leaders in probing the complexity of genetic determinants of artemisinin resistance (Art-R) in SE Asia. There is much fanfare around mutations in K13 as being the “marker” of Art-R, and all surveillance efforts worldwide are focused solely on this gene. However, there is ample evidence that knowing K13 mutations is at best only part of the story and at worst could be completely misleading in Africa where a major drug resistance crisis looms. This research group is one of very few taking on these hard but important questions and the only group developing powerful approaches to get answers. The large point-of-care sampling that underpins this study is a very powerful resource, and this group is honing and perfecting the analytical approach needed to find the complex determinants of clinical resistance (patient clearance half-life), including a comprehensive and novel use of transcript analysis. Several aspects of this work are impressive, including their layered and integrated approach that includes GWAS, innate transcript levels, drug-induced transcription, and population genetics. Because they classify a wide range of transcripts, the authors are able

to identify the previously unrecognized role of alternatively spliced gene products that contribute to the resistance profile. Nevertheless, the current version of the manuscript falls short in several ways. Because the work is pushing the boundaries for the field, the authors must do a better job outlining their rationale and their unique and comprehensive pipeline (currently cryptic and convoluted, and seemingly post hoc, which undermines the value of this work); further, they need to better highlight their adequate rigor and provide more transparency. It can be argued that their extent of validation of gene functions could be better; however, I do not think a full functional analysis is needed. However, a clearer description of the strengths and limitations of this approach, along with a more thoughtful consideration of the next steps that will further solidify this approach to put functional meaning to "genetic background" are needed. This critique is geared to making this work more accessible and convincing to a general audience, and to ensuring that ALL aspects of this work are available in the public domain for deeper follow-up and extensions of this analysis.

General observations: Nayak et al. combine genomic and transcriptomic data generated from patient samples collected at timepoints pre- and post-drug treatment as part of the TACT-CV clinical study of triple artemisinin combination therapies. They seek to clearly elucidate the genetic contributions to a complex quantitative phenotype. These genetic features are in addition to (or in place of) the well-known Kelch13 mutations with functional relevance to the mechanism of resistance, drug action or fitness to be useful as surveillance markers. Ideally, but not clear is whether these findings will be of direct relevance to Sub-Saharan Africa. The authors use population structure analysis to separate SE Asia parasites into three clades associated with varying levels of clinical resistance and drug selection histories. GWAS identified SNPs associated with clinical resistance, including PfRAD5 and PfWD11. Further transcription analysis supported and elaborated the role of these two genes, particularly highlighting alternative transcripts as a component of the parasites' mechanism of resistance/response to drug exposure. Deeper investigation of available sequence data strongly suggests a shared selection history between WD11 and K13, implying a covariant role in resistance, e.g. some kind of

evolutionary partnership with K13 that will require future study to elaborate mechanisms. The study design and data are valuable and warrant a high-level publication in Nature Communications. However, aspects of the analysis must be strengthened with an emphasis on transparency and increasing the full potential to the community of this patient clinical data for reanalysis, deeper analysis and validation of this expanded view of the genetic background of Art-R.

Major comments:

1. The clinical phenotypes are incredibly important but are getting lost. A panel in figure 1 showing the distribution of clearance times with symbols representing clades and colors for location being consistent with other panels of figure 1 would help. Include an additional indicator for K13 mutation status. A similar format to figure 5c would satisfy this concern. Showing a geographic clade distribution would set up figure 5c as building from figure 1 and would help clarify the discussion. The distribution shown in supplemental figure 3 does not do justice for the importance of the clinical phenotypes and the proposed additional panel would be a complement to supplemental figure 3.

We thank the reviewer for this comment, which will certainly improve the presentation of this manuscript. Accordingly, we added the information about the phenotypes and K13 genotype to figure 1, 5 and S3. We comment on these additions in the text:

Line 153:

"…of w & eCambodia, including Kravanh, Pursat, and Stung Treng; and southern sVietnam: Binh Phuoc and Khanh Hoa between March 18, 2018, and January 30, 2020 (Figure 1a, Table S1)52. Reflecting the status of artemisinin resistance, the majority of the samples from wCambodia and sVietnam carried a K13 mutant allele and came from infections with a slow clearance half-live (PC1/2>5 hours). In contrast, the set of samples collected in eCambodia was evenly split between artemisinin-resistant and sensitive parasites, the latter of which carried the ancestral wild-type PfK13 allele."

Line 296

"*Consistently, an admixture analysis identified a distinct ancestral subpopulation that is exclusive to (e)Cambodia and mainly comprised of artemisinin sensitive parasites (Figure S3). This population is currently admixing with two distinct subpopulations of resistant parasites, one of which appears to be spreading from (w)Cambodia. Tajima's D index value distribution is also consistent with this model (Figure 1d)."*

LINE 378: "(*Figure 5a). The marked exceptions are eGMS and wGMS, where the PfWD11 i2+30C>T frequency reached 37.7% and 11.3%, respectively. This is proportioned to the frequency of the mutant PfK13 genotype and the artemisinin resistance phenotype captured by the Pf7 samples set, with most P. falciparum infections outside the GMS still being fully sensitive to artemisinin."*

2. The authors determined 3 clades best represent population structure of the parasites used in the study given principal component analysis and neighbor-joining tree analysis. Include further rationale to arrive at the value of K=3. Generally, elbow method is used where point of inflection denotes the value of K. Besides PCA, an MDS or K-means clustering would further confirm the variance observed and separation of population structure. A scree plot would show how this analysis impacted the determination of 3 clades and how much variation is beingaccounted for by each PC after the first three.

The reviewer is correct in asking for more clarity on the population structure. We now provide the elbow analysis of the principal components, showing that the populating structure's main part falls into two clusters (K=2). Hence, in the new version of the manuscript, we changed the classification of the neighbor-joining tree analysis to two main clades (***I*** & ***II***), which reflects the actual genetic relatedness within the sample set. Nevertheless, we also note the substructure within clade ***II***, with ***IIa*** being overrepresented by parasites from wCambodia and ***IIb*** with parasites from eCambodia. In the manuscript, we wish to maintain this classification, which also reflects the stratification of the Cambodian parasite populations, consisting of drug-sensitive (clade ***I***) and (clade ***II***) resistant parasites. We also note a small subpopulation of (6) parasites (clade ***III***) that appears genetically distinct and whose origin remains unclear.

We note the elbow analysis (**line 181**):

"*Indeed, neighbor-joining tree analysis based on a pairwise distance matrix supports the PCA findings, classifying the TACT-CV cohort into two distinct clades, I and II, reflecting the K=2 stratification yielded by the elbow analysis (**Figure S2**)."*

To reflect this new population stratification we modified the tfollowing text of Results.

3. Figure 1c shows a smaller 4th clade of roughly 6 parasites between clades 2 and 3 that would be interesting for further Tajima's D analysis. Why did the authors not note the smaller clade in the results section of this analysis?

There are indeed 6 samples that appear distinct from the rest of the cohort and, as such, create a small clade in the neighbor-joining tree analysis. We did not mention this class in our original version as this small group does not support any meaningful statistical power for any downstream analyses such as the Tajima's D. At the moment we remain puzzled about the origin of these parasites. Although it is possible to search for the ancestors of these parasites by comparisons to the genetic database (Pf7k), we believe that this could create an unnecessary distraction in this already complex analysis as it will not contribute to the main objective, which is the genetic bases of artemisinin resistance. Based on the new clade classification, this is not termed clade III.

Nevertheless, the existence of this clade is now mentioned in the text (**Line 193**):

"It is also important to note the existence of a small clade of 6 parasite isolates (clade III) that showed a high genetic distance from the TACT-CV cohort, whose origin is currently unknown."

Additionally, Figure 1c has at least 4 'square' data points that do not match legend symbols for the three clades. Include a distinction for the square points in the figure legend.

This figure has been redrawn with new clade classifications, and the square points are now assigned to clade III

4. Tajima's D shows balancing selection pressure on clade 1, active selection pressure on clade 2 and no current selection pressure on clade 3. To fully show the selections (or lack of selection) include the Tajima's D value for each SNP used in the analysis as a supplemental table and as a supplemental figure to show what across the genome is being selected for each of the three major clades, similar to figure 5 in https://doi.org/10.1177/1176934321999640 with highlighted genes of interest. What are the Tajima's d values of GWAS hits in figure 1e? Are they preferred in the observed selection sweeps or balancing selection?

To address this reviewer's comment, we carried out Tajima's D value for a non-overlapping window of 5 Kb for each clade separately to understand the selection signature across different genome regions and commented on the selection parameters based on the distribution of Tajima's D. This analysis supported the overall observations showing the clade IIa and IIb (formally clade 1 and 2); albeit unevenly distributed. Indeed, there are several genomic regions, particularly those on chromosomes 2, 6, and 12, that appear to be under positive selection in both clades. This is now depicted in **Figure S4. Table S2.**

The GWAS hits fall into genomic regions with positive Tajima'sD values, which indicates balancing selection. The exception is the region on chromosome 13 (PfK13-Prad5), which appears to be under positive selection in clade IIa and represents the recent spread of artemisinin resistance in (e)Cambodia.

We discuss this in the text **Line 243**

"Interestingly, the PfWD11 falls in the regions with positive Tajima's D values in both clades II a and IIb, suggesting a balancing selection in both parts of Cambodia. On the other hand, the region of chromosome 13 with PfK13, PRad5, and other GWAS hits appears to be under positive selection in clade IIa, representing the recent spread of artemisinin resistance in (e)Cambodia."

5. The authors acknowledge the peak multimodality of clade 2 as part of the interpretations for ongoing selection. Could the multiple peaks also be indicative of the clades warranting separation into more subgroups rather than an overall Clade 2 being under active selection pressure? Clade 1 also shows a multipeak profile but is described as "skewed" which does not appear to be the most accurate description of the clade 1 distribution. Be clearer about the description of clade 1.

We thank the reviewer for this comment, allowing us to describe the Tajima's D index distribution with greater clarity. Indeed, multiple peaks in the distribution (now) Clade IIa and IIb reflect a subgroup segregation within both clades, most being shifted towards the positive values. This indicates that at least a part of the parasite populations in these clades are under selection. This contrasts the Clade I Tajima' D distribution that appears neutral, indicating no selection. Indeed, the main purpose of this study was to show the difference between Clade I on one side and Clades IIa and IIb on the other in terms of putative drug selection. We believe that this analysis still depicts this but, at the same time, corrects the interpretation of the multiple peak distribution in the text.

Line 200: *"Clade I peak at 0, indicating an absence of (any) current selection pressure. On the other hand, Clade IIa and IIb exhibit complex distributions with multiple peaks, most of which shift towards the positive Tajima's D index values. This suggests a finer subgrouping of the parasite populations in Clade IIa and IIb with ongoing under recent positive selection."*

6. To further show the commonality between clades and ancestral vs admixed natural of them it would be useful for reader interpretation to include an identity by descent analysis (IBD) for the parasites used in this analysis.

Indeed, in the new version of the manuscript, we carried out a comprehensive IBD analysis with the goal of finding the IBD fractions shared within each cluster of parasites as well as across the parasite clusters. Here, we found that parasites in Clades IIa and IIb show some degree of genetic relatedness even though they still retain their own identity. Clade I, on the other hand, appear genetically distinct but at the same time exhibit broader diversity within itself. Clade I is more related to IIb compared to IIa, which appears distant. This reflects the geographic distribution with Clade I and IIb being from eCambodia and Clade IIa from wCambodia. Overall, the IBD results support the overall picture of the population structure of the TACV -CV cohort compromised of three subpopulations, two of which coexist in eCambodia. Here, artemisinin-resistant parasites expanding from Cambodia are gradually admixing into the drug-sensitive population preexisting in this region and/or spreading there from the neighboring Laos.

Findings from the analysis are presented in Figure S5 and described in text (**line 220**):

*"This is also consistent with the identity by descent (IBD) analysis that revealed close genetic relatedness of Clade IIa and IIb, both of which are distant from Clade I (**Figure S5**). Interestingly, parasites in Clade I exhibit broader genetic variability, possibly reflecting their diversion unaffected by (any) recent drug selection. Clade I is, however, more related to Clade IIb, reflecting their common geographic distribution. Collectively, these results suggest that artemisinin-resistant parasites expanding from Cambodia*

are gradually admixing into the drug-sensitive population preexisting in this region and/or spreading there from neighboring Laos."

7. What is the rationale for using the first 10 principal components for GWAS? A scree plot for PCA should be generated and included as a supplemental figure with greater description in the methods for this choice of 10 to bolster the GWAS and population structure analyses.

We regret our error in the incorrect description of the GWAS analysis.

Initially, we used the 10 principle components (PC) correction, which is one of the standard methods used by the FaSTLMM package. Generally, 10 PC correction is a benchmark default setting used by common GWAS applications. Alternative to the PC-based correction, we also tested the Genetic Relatedness Matrix (GRM) calculated from SNP data that uses a matrix to control the effect of population structure present on the data set. In our internal evaluation, FaSTLMM with GRM showed less inflated statistical significance (λ = 1.05) compared to the benchmarking 10PC approach (λ = 2.25) (see **Figure S6**). Hence, in this study, we used the GRM-based approach as the final method. Details of this method and the utilized setting is provided in the materials and methods. Regretfully, in the last version of the manuscript, we misrepresented the GWAS analysis, failing to indicate the 10PC was used only as an initial benchmarking, not as the final method. This error was caused by miscommunication between the contributing authors. Being aware of this, we now inspected all details of the method descriptions to ensure their correctness.

We corrected this error in the text of the new version of this manuscript (**line 223**): *"Here, we carried out GWAS with the Genetic Relatedness Matrix (GRM) that mitigates the population structure effect as implemented by the FastLMM program. The GRM method was chosen instead of the standard 10 principal component correction due to a lower inflation factor Λ^{GRM} = 1.05 compared to Λ^{10PC} = 2.25 (**Figure S6**).*

8. It is stated that roughly 300 parasite samples are sufficient to be representative for the overall transmission pattern of infection in the eastern part of the GMS (cite #48) to support this claim. To fully articulate this point the authors should include a power analysis given this sample size and include a supplemental figure plotting power vs minor allele frequency and effect size. This analysis would demonstrate the ability of this study to identify genes of smaller affect size through GWAS that would constitute the genetic background of resistance.

Here, we regret a slight misunderstanding of our original statement, which gave the impression that 300 parasite samples were chosen as a sample size for sufficient statistical power. Instead, the statement that these samples represent the population refers to the extremely low transmission rate. In other words, these 300 samples collected in the TACT-CV cohort were more or less ALL *falciparum* malaria infections detected in these four regions between 2018 and 2020. Hence these 300 *P. falciparum* isolates represent the core/most (if not all) falciparum infections with no other available in these sites. Because of this sample size, we calculate the statistical power in the new version to show its sufficiency.

9. The 1*10-5 threshold for significance is lower than the commonly observed threshold for GWAS of 1*10-8 Clarify this choice of threshold. There is good precedent for lower thresholds, e.g. Cerqueira et al. 2017 and Wendler JP et al., https://doi.org/10.1371/journal.pone.0096486. Power to detect is often countered by the value of available clinic samples (as the authors note). These clarifications will strengthen the paper.

We thank the reviewers for recognizing this issue and acknowledging that the lower P-value (10-5) was used instead of the commonly applied (10-8) due to sample scarcity. We managed to obtain only 300 samples because no other infections existed during the survey period. We thank the reviewer for pointing us to the precedence publication, and we include this in our manuscript for more clarity.

Amendments addressing both comments 8 and 9 is now mentioned in the text **line 226**:

"Indeed, even this limited size cohort of ~ 300 samples of the TACT-CV can provide sufficient statistical power for genetic analysis of artemisinin resistance factors (Figure S7). For that, however, we applied a lower statistical threshold of pValue>10-5, instead of commonly used 10-8, to accommodate for the sample scarcity as previously described56*."*

10. Provide the full summary statistics for the GWAS analysis (the current supplementary material only lists significant markers and no effect size). A complete summary of statistics with necessary statistical parameters of both significant and nonsignificant hits will make this a more valuable resource. This will be help establish a new standard in a high impact journal.

As suggested, the new version of the manuscript now provides the modified table of GWAS results in the supplementary **Table S3**, which contains all the SNPs along with summary statistics.

11. Provide the quantile-quantile (QQ) plots for the GWAS before and after correction for population structure as a supplemental figure to demonstrate that population stratification was adequately accounted for. The plots should also include the genomic inflation factor (λ).

In **Figure S6** we now provide all QQ plots depicting h PC10, GRM based corrections. Here, we show that inflation point dropped from uncorrected λ=9.95, to the λ^{PC10} =2.25, λ^{GRM} =1.05 and $\lambda^{PC10\&GRM}$ =0.962. For the final GWAS analysis we selected the GRM method. (also see comment 7 above)

12. To the point of power for detection of significant SNPs, in figure 1e there is a second SNP on chromosome 11 at the opposite end compared to PfWD11 that is approaching the significance threshold, more significant than other non-significant SNPS. It is not listed in the text or supplemental table 2. An extended value of the approach the authors take is to consider the full data as a source for further follow-up studies.

The second SNP at chromosome 11 is a nucleotide change A>T in an intergenic region between genes PF3D7_1119600 (metalloprotease) and PF3D7_1119700 (unknown). We agree that even lower statistical significance in our GWAS approach might be useful information. We provide all GWAS related information in the newly updated **Table S3**. Nevertheless, for consistency, we do not discuss this second mutation or any other SNP with pValue>10^{-5} in the text.

13. K13 is well known to be the major gene for artemisinin resistance in Southeast Asia, as the authors articulate well. A comprehensive analysis of the SNPs identified in linkage disequilibrium with K13C580Y would be useful.

LD analysis with PfK13C580Y identified 91 SNPs with R-squared>=0.3 and d-prime>0.5. Out of these PfRAD5_N1131 (13:1718288) showed the highest association with high LD (R^2=0.96) while PfWD11 i2+30 was R^2=0.31.

A secondary GWAS analysis including K13 as a covariate as a supplemental figure could also answer emphasize this point.

GWAS analysis with PfKC380Y as a binary phenotype also identified several regions on chromosome 11 (with PfWD11), 13 (with PfK13 and PfRa5), and 14 associated with the PfK13 genotype.

The authors touch on this question by their investigation of pf7 parasites and showing PfWD11 has been selected along with K13 across time. Extending this analysis to include all SNPs would further strengthen the 'pipeline' and also emphasize the outlier nature of PfWD11.

Subsequently, we investigated the frequency profiles of the SNPs identified by the LD and PfK13 GWAS study, showing that both PfWD11 and PfRad5 and several other polymorphisms followed the evolution of the PfK13 SNPs in the GMS since the emergence of artemisinin resistance in 2009. Here, we show that in addition to their GWAS associations with artemisinin resistance, the SNPs in both Pfad5 and PfWD11 fall into the group of ~20 outlier SNPs associating with PfK13 throughout the evolution of artemisinin resistance in the GMS over the last 15 years.

These results are presented in **Figure S18** and discussed in Discussion, **Line 481:**

"Taken together, unlike PfK13, for which different alleles emerged, spread, and subsequently disappeared, being replaced by others, the SNPs in both PfRad5 and PfWD11 remained constant throughout the last 15 years during which artemisinin resistance spread through the GMS. Indeed, PfRAD5_N1131I and PfWD11 i2+30 show strong associations with PfK13 mutat phenotype when studied by GWAS of linkage disequilibrium (LD) studies in the TACT-CV cohort (Figure S18a, b and Table S8). Moreover, the temporal profile of their frequency progression in the GMS since 2009 also shows a high correlation with PfK13 C580Y, which classifies both mutations into a small group of outlier mutations with a high link with artemisinin resistance (Figure S18c and d). Pending validations of their biological relevance, these may now represent more suitable markers of artemisinin resistance in the future. Currently, little is known about the putative genetic background of the PfK13 mutant parasites emerging in Sub-Saharan Africa11,77. It will be particularly interesting to investigate if these alleles also contribute to artemisinin resistance in African parasite lines; owing to otherwise low frequencies of…"

14. The authors provide hypotheses for the biological role of PfWD11 heavily relying on citation #66. That publication does a fair amount of work articulating several proteins as part of the TAF1/BDP5 complex (BDP5, TAF1, TAF2, TAF10, PfWD11, PF3D7_0824800, PF3D7_1032700, PF3D7_1251800, PF3D7_0514900, PF3D7_1361200). While I agree that detailed validation of the biological function is beyond the scope of this manuscript, some deeper exploration within the data presented here of the other proteins hypothesized to be in the TAF1/BDP5 complex listed in cite #66 is warranted since PfWD11 is a key result of the study. How does the expression of these other 9 genes compare to PfWD11 and are any of them included in other analyses shown for PCRT/non-PCRT as correlated with PC1/2

We investigated the expression value at the baseline level in resistant and susceptible parasites and the response upon ACT exposure. None of the interacting genes were found to be differentially expressed with statistical significance. As such, adding this to the manuscript would not be beneficial as it can distract from this already highly substantial manuscript. Naturally, we are willing to do so if the reviewer finds it necessary.

A table is provided below.

	TPAS			RESPONSE(PC$^{1/2}$ < 5Hr)			RESPONSE(PC1/2 > 5Hr		
	Resistant	Susceptible		Hour0	Hour6		Hour0	Hour6	
ID	mean_exp	mean_exp	P	mean	mean	P	mean	mean	P
TAF1	0.01	-0.02	0.66	-0.27	0.32	0.00	-0.18	0.20	0.2(
TAF2	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/
TAF10	0.25	-0.26	0.04	0.09	-0.13	0.19	0.28	-0.38	-0.3
PF3D7_0824800	0.26	-0.28	0.01	0.10	-0.13	0.09	0.36	-0.39	-0.3
PF3D7_1032700	-0.16	0.18	0.11	-0.15	0.17	0.07	0.06	-0.07	-0.0
PF3D7_1251800	0.25	-0.27	0.05	0.02	-0.02	0.83	0.17	-0.18	-0.1
PF3D7_0514900	-0.03	0.03	0.72	0.13	-0.15	0.21	0.13	-0.14	-0.1
PF3D7_1361200	0.19	-0.21	0.13	-0.10	0.12	0.08	0.09	-0.10	-0.1

15. For the qPCR validation experiment, it is unclear if it was done with multiple biological replicates or only a technical triplicate of the qPCR for a single biological replicate. That distinction should be made clear in the methods and figure legend.

All qPCR measurements with the in vitro cultured parasites were done in three biological replicas. This is now clearly stated in the legends of the figure and the description of materials and methods.

16. The enrichment analysis of PCRTs and non-PCRTs is presented for only down regulated genes. Theoretically upregulated genes would include proteasome components (previously highlighted by this group). State also that deceleration of the IDC by transcriptional down regulation has also already been demonstrated in early work by this group. Additionally, line 285 states enrichment analysis identified a wide array of processes related to parasite "lifecycle progression". By this interpretation, given the array of GO terms found, almost any GO term found could be seen as part of the parasite's life cycle. The interpretation is too broad to be meaningful. Also, the use of "lifecycle" throughout the manuscript is too broad. At points where the IDC is being referenced the "asexual cell cycle" or IDC should be used whereas "lifecycle" should be used in reference to transition between hosts as used in reference to lines 482-483.

We agree that the term "lifecycle" is imprecise and thank the reviewer for highlighting this discrepancy. Throughout the text we refer to that strictly as the IDC as this represents the main focus of our observations.

Moreover, it is true that in our previous work, we demonstrated that compared to sensitive parasites, artemisinin-resistant parasites upregulate several cellular pathways involved in protein metabolism and unfolded protein response, etc. This was previously shown in steady-state parasites prior to treatment (e.g., H0). Due to the limited statistical power of this study due to the limited number of samples, these upregulations in the H0 transcriptomes did not reach sufficient statistical significance and, as such, were not presented. Instead, this manuscript focuses on factors that could be observed with sufficient statistical power, such as noncanonical aPCTs.

The only statistically significant functional enrichments that the reviewer refers to were those in the H6 transcriptomes that showed the deceleration of the IDC that was also previously shown in our earlier work.

We acknowledge the lower statistical power of this study and its impact on the TPAS in the text **line 284**:

"*The H0 transcriptomes analyzed in this study did not recapitulate the previously reported ARTP29, identifying only a small portion of differentially expressed genes between the resistant and susceptible parasites. This is due to the limited statistical power of the TACT-CV samples set (n<300), in which most differential gene expressions were below the statistical significance achieved with previous larger datasets of TRACI27 and TRACII29."*

17. For consistency between text and visual representation in figures, everywhere that states west vs east separation of Cambodia in the text and figures should be written, listed and described in order of west then east. This will help in interpreting figure 1.

As advised, we corrected this aspect, mentioning west and east in this order.

18. The overall value of the impactful dataset will be greatly enhanced by including the scripts used for the analysis with results available through a github repository rather than available upon request. This is especially important given the analysis uses in-house scripts for detection of nonPCRTs which will be invaluable for guiding future studies.

The scripts are submitted to github repository. And can be accessed through the link (www.github.com/sourav/tact_cv_analysis_protocol)

Minor comments:

1. In the introduction be clear about "eQTL-like" acting mutations as was described in citation 25 around lines 99-100. This concept is important for the understanding the results and for the specific genes singled out for direct analysis. This will make it easier on the reader to not haveto read the cited paper in depth for a full understanding.

WE rephrased this sentence for clarity **line 106**:
"*The differential expression of most, if not all, individual genes of the ARTP is also associated with specific genetic polymorphism[30], and in one case, it was shown to be mediated by sequence length polymorphisms at the AT-rich tandem repeats in the gene promoter[31]*."

2. For an ideal population the average Tajima's D is expected to be around 0 but in figure 1d clades II and III appear to have an average Tajim's D greater than 0. Can the authors provide an explanation as to why the average Tajima's D deviates from 0?

The deviations from 0 indicate that the selection processes are ongoing in these clades. Please see the response to the **major comment 4**.

3. Typical GWAS includes a calculation for the Hardy Weinberg equilibrium for population used within the study. Since there is a selection sweep occurring for clade 2, do the authors think there could be a change to the Hardy Weinberg equilibrium. Inclusion of the full summary statistics may also answer this question and alleviate this concern.

We believe that the populations in clade IIa and IIb are not in Hardy Weinberg equilibrium as they are under selection. In our GWAS study, we used the GRM population stratification, which is equivalent to Hardy Weinberg equilibrium.

4. A full interpretation of the population structure of parasites would benefit from the inclusion of a kinship matrix and details of the LD pruning of the SNPs.

In the new manuscript, we provide LD calculations for the C580Y SNPs, showing a limited association with only 91 SNPs mainly located at different chromosomes. This highlights the low levels of LD in the *P. falciparum* genomes in natural infections, which was previously reported. In addition, the IDB analysis provided further support of the overall populations structure.

5. Is the expression pattern of selected genes in supplemental figure 7 generated from the transcription of the resistant parasites in the study or from a reference time series experiment? If it is the former, how were full timeseries extracted from only two timepoints of 0hr and 6hr, pre and post treatment?

We regret this confusion. The transcriptomic profiles in Figure S7 (now S12) are extracted from the reference data of the P. falciparum IDC transcriptome published previously. We correct the legend of Figure S12 to make this clear.

6. The abbreviation of patient clearance half-life is not consistent throughout the manuscript (i.e., PC1/2 in some places and PC1/2 in others).

We scanned through the text and made this consistent.

7. Figure 1b has PC1 labeled as both axes.

Figure 1b has been corrected.

8. Lines 74-77: Specify the increase of the R561H K13 mutation are in Africa countries. I recognize the citations for this sentence are from studies using African parasites but for flow and plainly clear transition from GMS to Africa made in this paragraph, the minor edit would be beneficial.

Also, in line with reviewer 1 comment 6, this paragraph was rewritten with new most recent citations were added.

9. Lines 156 and 157 state transcriptome samples were generated for 310 and 300 parasites at timepoints 0hr and 6hr, respectively. However, within the methods it is stated only 284 and 251 transcriptome samples for 0hr and 6hr timepoints were suitable for analysis. With this sentence or within the results section for the transcriptome analysis this point should be restated to reiterate the samples used in GWAS vs TPAS even though it is stated in line 232. Also line 232 states 282 samples for 0hr instead of 284 as in the methods and line 233 states 252 for 6hr instead of 251 as in methods, are these typos?

We regret this confusion. The differing numbers were caused by a triage process in which we first generated transcriptomes from the initial 310 and 300 RNA preparations from the initial parasite samples. Nevertheless, subsequently, we applied quality control criteria for transcriptome integrity, which led to the exclusion of a certain number of samples for the final dataset. In the new version of the manuscript, we only include the final high-quality samples. In the new version, we only mentioned the high-quality transcriptomes, which included 282 for H0 baseline transcriptomes and 252 for h6 induced transcriptomes. In the material methods section describing the TPAS, the transcriptomes were classified not by H0 and H6 but instead by the resistant status of the parasite, which is 284 for resistant and 251 for susceptible. Although those appear similar numbers these are not the same. ..

10. Line 189: Why is “(any)” in parentheses?

Parentheses were removed.

11. Line 215: “an SNP” should be “a SNP”.

With all due respect, an SNP is a correct spelling.

12. Line 318: A space is needed between “PfWD11and”.

Indeed, we thank the reviewer for the correction.

13. Lines 379-380: The percentages for each mutation listed in the text appear to be flipped from what is shown in figure 5d.

We regret this typo and correct it.

14. Line 436: Why is “architecture” in quotations? This might imply a dissatisfaction of the cited article’s (#18) analysis and claim for investing genetic architecture of resistance? If so, a clear articulated rationale of its limitations for interpreting genetic architecture of resistance would be necessary. The explanation would also be welcomed as incremental improvement to the understanding of biological phenomenon based on new or repeated studies in the basis of science.

By no means did we imply any disrespect for the study of the artemisinin resistance genetic architecture. The initial notion of the parentheses was simply to indicate that architecture is not exactly a genetic term but rather a metaphor. In the new version we removed the parentheses.

15. Lines 457-458: The citation (67) does not support the claim the TAF complex regulates the p. falciparum life cycle. The cited article is a bioinformatic of the parasite genome searching for transcription factors and doesn’t show data describing their biological role in the parasite.

We regret this confusion. There is proteomics evidence that PfWD11 resides within the TAF complex, which functions as an epigenetic reader that regulates IDC-dependent transcription. This is shown in reference (now) 74. This is line with generic function of TAF complexes in other eukaryotes (reference 75).

WE rephrase these statements on **line 490; “***Although the precise biological function of PfWD11 remains unknown, there is evidence of it being a part of the TAF epigenetic reader complex along with the TAF1/BDP5 chromatin-binding protein[74]. The key role of the TAF complex in many eukaryotes is in the regulation of cell cycle progression by transcriptional initiation[75].*”

16. Line 629: “(a) (a)” should be “(a)”

Corrected.

REVIEWERS' COMMENTS

Reviewer #2 (Remarks to the Author):

The authors have responded to the reviewer comments in the revised manuscript. They have clarified several issues by including additional analyses as requested by reviewers. The manuscript represents a major experimental effort deriving data from patient samples before and after treatment with ACTs. As stated in the previous review, this is a major technical breakthrough in the field and the data generated will be of value to others in the field beyond this initial analysis by the authors.

There remain some weaknesses. The authors have made some progress in streamlining the descriptions, the manuscript remains densely written. The authors now acknowledge that these results do not recapitulate genes they have identified in previous of parasite samples from the same region. They attribute this statistical power, however it does call into question whether they are uncovering key new insights into the biology of artemisinin response or are detecting secondary or tertiary downstream effects. The authors should address the issue of primary verses secondary effects more critically both in the presentation and in the discussion.

The authors focus on SNP variants in two genes, PfRAD5 and a WD40 domain protein on chromosome 11, PfWD11. These variants appear in the population simultaneously with the emergence of partial artemisinin resistance and mutations in Pfkelch13. This temporal data is compelling, however, without mechanistic explanations, their role in partial ART resistance remains unclear. The authors should acknowledge that these variants might represent factors in parasite fitness/transmission that could indirectly impact the response to ART. While outside the scope of this manuscript the authors should acknowledge the limitations of their results and suggest a broader range of next steps.

Reviewer #3 (Remarks to the Author):

The authors have comprehensively and thoughtfully addressed my wide range of detailed critiques. I am impressed with the care and detail of their improvements to the manuscript that will make it more accessible to the reader and more impactful in the field. We accept their point that adding the expression data of interacting genes would be a distraction given that no significant differences were observed. Please ensure that the "currently private" scripts are comprehensively and coherently accessible to the reader via github.

Responses to the REVIEWERS' COMMENTS

Reviewer #2 (Remarks to the Author)

The authors have responded to the reviewer comments in the revised manuscript. They have clarified several issues by including additional analyses as requested by reviewers. The manuscript represents a major experimental effort deriving data from patient samples before and after treatment with ACTs. As stated in the previous review, this is a major technical breakthrough in the field and the data generated will be of value to others in the field beyond this initial analysis by the authors.

We thank the reviewer for recognizing the advances made in this study. This is the first study analyzing alternative splicing and noncoding RNA expression in clinical malaria samples, showing their associations with a key phenotype, drug resistance. Moreover, here we also demonstrate the mechanism of the spread of drug resistance that involves admixing the resistance-carrying allele into the otherwise susceptible population.

There remain some weaknesses. The authors have made some progress in streamlining the descriptions, the manuscript remains densely written. The authors now acknowledge that these results do not recapitulate genes they have identified in previous of parasite samples from the same region. They attribute this statistical power, however it does call into question whether they are uncovering key new insights into the biology of artemisinin response or are detecting secondary or tertiary downstream effects. The authors should address the issue of primary verses secondary effects more critically both in the presentation and in the discussion.

The authors focus on SNP variants in two genes, PfRAD5 and a WD40 domain protein on chromosome 11, PfWD11. These variants appear in the population simultaneously with the emergence of partial artemisinin resistance and mutations in Pfkelch13. This temporal data is compelling, however, without mechanistic explanations, their role in partial ART resistance remains unclear. The authors should acknowledge that these variants might represent factors in parasite fitness/transmission that could indirectly impact the response to ART. While outside the scope of this manuscript the authors should acknowledge the limitations of their results and suggest a broader range of next steps.

If we understand correctly, the secondary effects that this referee refers to are factors of parasite fitness and transmission as opposed to the (primary) factors involved directly in the artemisinin resistance. Indeed, the possibility that the identified genetic polymorphisms mediate the secondary effects cannot be ruled out. This is, for example, possibly true for PfWD11, whose involvement in gametocytogenesis has been hinted at in previous studies using saturating mutagenesis. On the other hand, PfRad5 likely represents a factor of the resistance mechanism, given its role in the DNA damage repair pathway linked with the artemisinin mode of action in numerous previous studies. Hence, the transcriptional differences associated with the increased clearance half-life in this, as well as the earlier studies, will likely represent both the primary and/or secondary factors.

On the one hand, this could be seen as a limitation (the fact that not all identified factors contribute directly to the mechanism *per se*) but at the same time, this is the main strength of this study, observing the clinically relevant genetic makeup of resistant parasites; the genetic background. Naturally, *in vitro,* validations are necessary to establish the causality of the identified factors, which is beyond the scope

of this study and the knowledge of all three reviewers. In the new version of the manuscript, we did our best to be more critical about this.

We note this in more detail in the text at several points:
Introduction line 95 *"The latter may represent mutation that does not necessarily mediate the artemisinin resistance per se but instead help to alleviate the fitness cost due to the resistance mutation and/or boost gametocyte production to elevate transmission of the resistant parasite lines."*
Results line 288 "*Nevertheless, it is feasible to suggest that the identified transcriptional events contribute to the artemisinin resistance genetic background either as direct factors of the resistance mechanism or by alleviating fitness costs and/or boosting gametocyte production (see discussion)."*
Discussion line 523 *"In future in vitro studies, exploring the casualty of the identified genetic factors for artemisinin resistance will be crucial. In particular, it will be crucial to assess whether these factors contribute directly to the multifaceted mechanism of artemisinin resistance96 or contribute to its spread in the clinical/epidemiological setting. The latter may involve factors that can alleviate the fitness cost caused by different PfK13 mutations in the asexual parasite growth97,98 or boost of sexual stage induction observed even in the initial study of artemisinin resistance in the GMS1. The natural limitation of this study, similar to any association studies, is the incapability to distinguish these two categories of the artemisinins resistance from each other. In turn, however, here we identify new genetic factors that contribute to artemisinin resistance of P. falciparum malaria infection in real-life clinical settings, which, in our view, represent the main strength of this work."*

Reviewer #3 (Remarks to the Author):

The authors have comprehensively and thoughtfully addressed my wide range of detailed critiques. I am impressed with the care and detail of their improvements to the manuscript that will make it more accessible to the reader and more impactful in the field. We accept their point that adding the expression data of interacting genes would be a distraction given that no significant differences were observed. Please ensure that the "currently private" scripts are comprehensively and coherently accessible to the reader via github.
We thank the reviewer for the praise about our analysis and presentation. We now provide a direct online link to the R scripts developed for this analysis. This is now directly obtainable from https://github.com/sourav-bioinfo/tact_cv_analysis_protocol.