



RESEARCH NOTE

Dimension reduction of Malaria Box data allows efficient compound prioritization [version 1; referees: 1 approved, 1 approved with reservations]

Gajinder Pal Singh

Molecular Medicine Group, International Centre for Genetic Engineering and Biotechnology (ICGEB), New Delhi, India

v1 First published: 18 Nov 2016, 5:2701 (doi: [10.12688/f1000research.10121.1](https://doi.org/10.12688/f1000research.10121.1))
 Latest published: 18 Nov 2016, 5:2701 (doi: [10.12688/f1000research.10121.1](https://doi.org/10.12688/f1000research.10121.1))

Abstract

Background: New anti-malarial drugs are needed to meet the challenge of artemisinin resistance and to achieve malaria elimination and eradication. The new anti-malarial compounds are expected to have many desirable properties, such as activity against multiple stages of *Plasmodium*, low host cytotoxicity, and low propensity for resistance development, but whether and how these properties might be linked to each other is not clear. A better understanding of the relationship between activities of compounds against different stages of *Plasmodium* could help in the development of strategies to prioritize compounds with maximum potential for further development. **Methods:** We utilized the large amount of data that has recently been generated on 400 anti-malarial Malaria Box compounds and performed statistical analyses, such as rank correlation, hierarchical clustering, and principal-component analyses, to test associations between activities against different stages of *Plasmodium*, other pathogens, and human cells. **Results:** We found significant positive correlations between the activities of compounds against different stages of *Plasmodium*. Our results also show toxicity associated with assays conducted at higher compound concentrations. Principal-component analyses (PCA) of the data allowed differentiation of *Plasmodium*-specific activity from general toxicity and predicted success in *in vitro* evolution of resistance. We found that a single principal-component can capture most of the desirable properties of Malaria Box compounds and can be used to rank compounds from most desirable to least desirable activity-profile. **Conclusions:** Here, we provide a systematic strategy to prioritize Malaria Box compounds for further development. This approach may be applied for prioritization of anti-malarial compounds in general.

Open Peer Review

Referee Status:

	Invited Referees	
	1	2
version 1 published 18 Nov 2016	 report	 report
1	Timothy N. C. Wells, Medicines for Malaria Venture (MMV), Switzerland	
2	Francisco-Javier Gamo, GlaxoSmithKline, Spain	

Discuss this article

Comments (0)

Corresponding author: Gajinder Pal Singh (gajinder.pal.singh@gmail.com)

Competing interests: No competing interests were disclosed.

How to cite this article: Singh GP. Dimension reduction of Malaria Box data allows efficient compound prioritization [version 1; referees: 1 approved, 1 approved with reservations] *F1000Research* 2016, 5:2701 (doi: [10.12688/f1000research.10121.1](https://doi.org/10.12688/f1000research.10121.1))

Copyright: © 2016 Singh GP. This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the [Creative Commons Zero "No rights reserved" data waiver](https://creativecommons.org/licenses/by/4.0/) (CC0 1.0 Public domain dedication).

Grant information: The work is supported by an Early Career Fellowship to G.P.S. by the Wellcome Trust/DBT India Alliance (IA/E/15/1/502297). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

First published: 18 Nov 2016, 5:2701 (doi: [10.12688/f1000research.10121.1](https://doi.org/10.12688/f1000research.10121.1))

Introduction

Malaria killed about half a million people in the year 2015, and 70% were children under the age of five¹. The emergence and spread of resistance towards frontline anti-malarial drugs in South-East Asia has created an urgent need to discover new drugs. In addition, new drugs are needed to meet the objective of malaria elimination and global eradication, for which the currently available drugs are not adequate². The desirable characteristics of new clinical candidates, also called Target Compound Profile (TCP), include high potency and fast killing of the asexual erythrocytic stage for quick relief of symptoms, high plasma half-life to reduce treatment duration, activity against the sexual stages to prevent transmission, activity against the liver-stage to avoid relapse and for prophylactic use, activity against multiple species of *Plasmodium*, and reduced propensity for the development of resistance³. New anti-malarial drugs must also be safe for mass administration, and for children and pregnant women, who are most vulnerable to malaria³.

It is currently not well understood how TCP properties are related to each other. This makes it difficult to assess whether it would be feasible for a single compound to have all TCP properties and what strategies could be adopted to find such candidates. With the discovery of thousands of active compounds from the high-throughput assays against the erythrocytic stage of *P. falciparum*⁴⁻⁸, it has become imperative to find a prioritization strategy that can identify the most promising candidates for further development. For a subset of antimalarial compounds identified from high-throughput screens in the so-called “Malaria Box”, many of the TCP properties have been assessed. The Malaria Box is a set of 400 compounds selected based on their potent activity against the erythrocytic stage of *P. falciparum*, chemical diversity and commercial availability⁹. These compounds were made available free of cost to researchers, thus catalysing a number of studies, including the screening of these compounds against multiple *Plasmodium* stages, eukaryotic pathogens and human cells¹⁰. Some of these compounds have also been tested for their propensity for resistance generation¹¹. Here, we utilized the large amount of data generated on Malaria Box compounds and found significant associations between different TCP properties. Based on these observations, we propose a prioritization strategy for anti-malarial compounds for further development.

Methods

The screening data on Malaria Box compounds was obtained from Van Voorhis *et al.*¹⁰ who compiled the previously published data on Malaria Box compounds (55 assays), as well as their own data (236 assays). We rank transformed all assay values, such that higher values represent higher inhibition.

In case multiple assays were available for a given stage or concentration, their median values were taken: there were nine assays reporting EC₅₀ values against the asexual stage of *P. falciparum*, one assay against asexual stage at high compound concentration (10 μM), five gametocytocidal assays conducted at 0.5–1 μM compound concentrations, ten gametocytocidal assays conducted at 2.5–5 μM compound concentrations and six gametocytocidal assays conducted at 10–12.5 μM compound concentrations. There was one assay each at lower and higher compound concentrations

against liver (5 μM and 50 μM, respectively) and ookinete stages (1 μM and 10 μM, respectively). Values were also similarly combined for parasites with multiple assays, such as *Babesia* sp. and *Mycobacterium tuberculosis*.

The data on the *in vitro* resistance evolution of Malaria Box compounds was obtained from Corey *et al.*¹¹.

All statistical analyses were performed in the R software v3.3.1 (<https://www.r-project.org/>). R commands `hclust` and `prcomp` were used for hierarchical clustering and PCA analyses respectively. PCA analyses were performed on the activity data against different *Plasmodium* stages and human cells. Rank correlation values were used to create the distance matrix for the hierarchical clustering.

Results

Relationship between activity of Malaria Box compounds in different assays

Van Voorhis *et al.* have recently reported their large-scale screening data on Malaria Box compounds, which was compiled along with the previously published data¹⁰. We first reduced the dimensionality of this data by combining variables that describe activity against the same *Plasmodium* stage and pathogen. Assays conducted at higher concentration of compounds may provide different results from those performed at lower concentration, and thus assays conducted at different concentrations were combined separately.

Figure 1 provides an overview of the relationship between different properties of Malaria Box compounds in the form of a correlation matrix. Multiple observations can be made from this matrix. There is a moderate, but significant correlation, between activity against the asexual stage and the gametocyte stage of *P. falciparum* (Spearman rank correlation 0.43, between EC₅₀ values and % inhibition of gametocytes at 1 μM compound concentration). The correlation between EC₅₀ values against asexual stage and gametocyte stage gets lower when gametocytocidal activity was screened at higher compound concentrations (Spearman rank correlation 0.17, at 10μM). The gametocytocidal activity at higher concentrations shows a higher correlation with inhibitory activity against different pathogens, including *M. tuberculosis* and human fibroblast cells (Figure 1). Assays conducted at higher concentrations against asexual, liver and ookinete stages also show a higher correlation with toxicity against human cells and other cell types (Figure 1). These observations suggest that assays conducted at high compound concentration show general toxicity against a wide variety of cells, including human cells, thus hits identified from these assays should be used with caution.

To further understand the relationship between activities against different *Plasmodium* stages, we performed hierarchical clustering of the data. Three major clusters were evident (Figure 2). Cluster 1 consists of asexual and gametocyte assays conducted at high compound concentrations. Cluster 2 consists of asexual and gametocyte assays conducted at low compound concentrations. Cluster 3 consists of assays conducted against liver and ookinete stages. Separate clustering of assays against asexual and gametocyte stage at different compound concentrations again suggests general toxicity

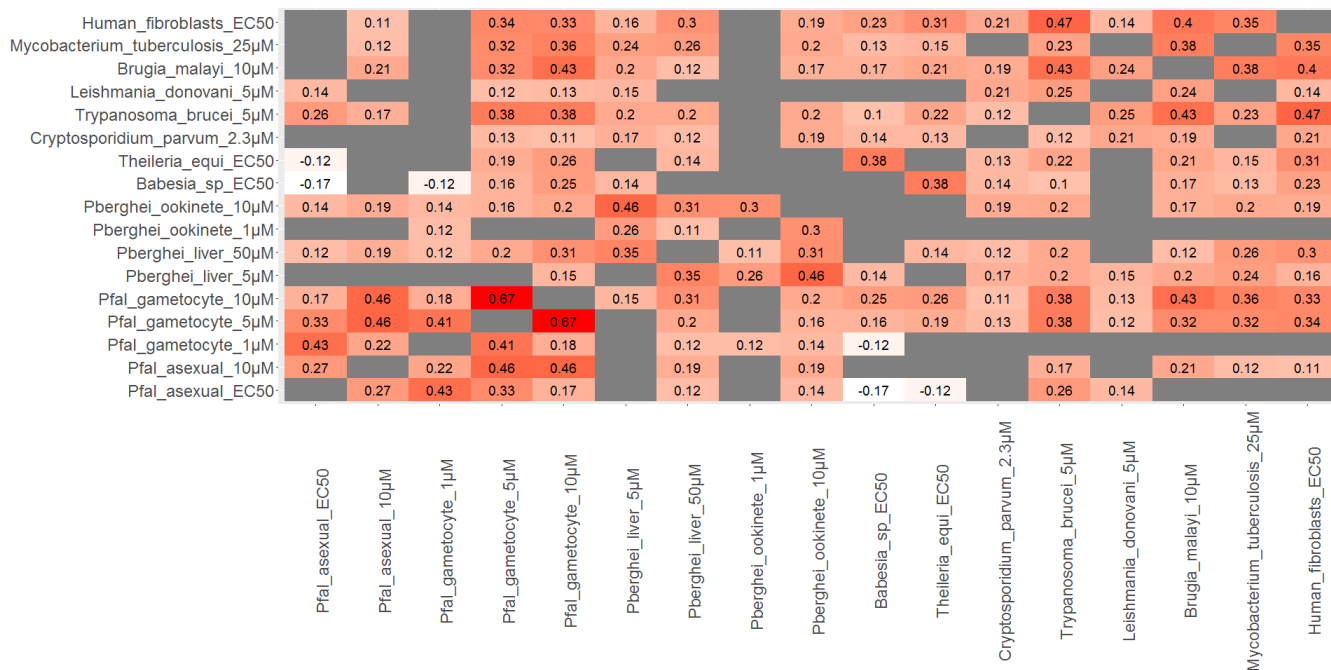


Figure 1. Correlations between inhibition values across different assays. Spearman rank correlations are shown between assays whose values were rank transformed, such that higher values indicate higher inhibition. Gray boxes indicate *p* values > 0.05. The assays performed at higher concentrations in *Plasmodium* show higher positive correlations across different assays, including activity against human cells, suggesting that assays conducted at high compound concentration show general toxicity.

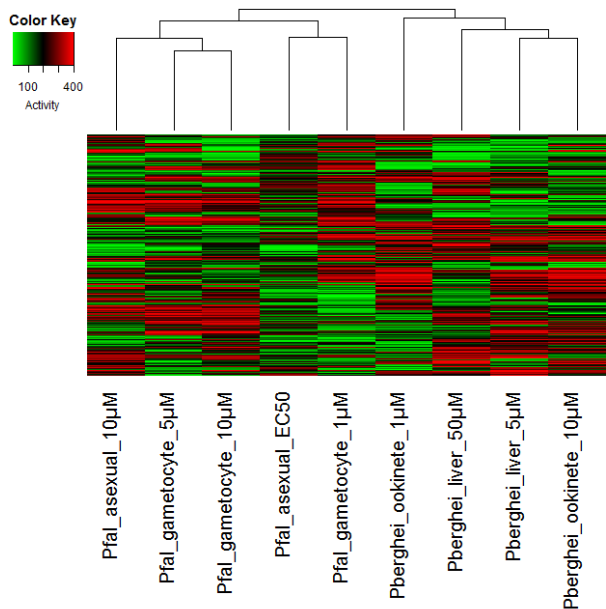


Figure 2. Hierarchical clustering of the assay data against different *Plasmodium* stages. The three major clusters are evident that correspond to activity against *P. falciparum* at high concentrations (Cluster 1, leftmost), activity against *P. falciparum* at low concentrations (Cluster 2, middle), and possibly activity against *P. berghei* (Cluster 3, rightmost). Rank correlation values were used to create the distance matrix for the clustering. The color key shows the inhibitory activity of the compounds with a higher number representing a higher activity.

of assays conducted at higher compound concentrations. There are two possibilities why liver and ookinete stages cluster together. These two stages may be physiologically more similar to each other, or it may reflect the fact that these assays were conducted against *P. berghei*, compared to other assays that were conducted against *P. falciparum*.

PCA analyses allows differentiation of general and specific toxicity

Given the possible confounding roles of compound concentration and *Plasmodium* species used in the screening, the prioritization of compounds that have pan-stage activity, but low host cytotoxicity, becomes difficult. We thus tested whether the Principal-Component Analyses (PCA) may be utilized to differentiate *Plasmodium*-specific activity from general toxicity. PCA analyses of the data from different *Plasmodium* stages and human cells lead to the identification of principal-components, which showed different properties with respect to general and specific activity. PC1 showed high correlation with assays conducted at higher compound concentrations and against a variety of cell types, including human cells (Figure 3), suggesting that PC1 is related to general toxicity. PC3, on the other hand, showed higher correlations with assays conducted at low compound concentrations, but negative or lower correlation with assays conducted at high compound concentrations in *Plasmodium*, against different pathogens and human cells (Figure 3), suggesting that PC3 is related to specific activity against *Plasmodium* across different stages. PC2 showed high positive correlation with the liver and ookinete stage assays, but showed negative correlations with asexual and gametocyte stage assays, suggesting that

Human_fibroblasts_EC50	0.44	0.13	-0.59
Mycobacterium_tuberculosis_25µM	0.38	0.1	-0.26
Brugia_malayi_10µM	0.39		-0.26
Leishmania_donovani_5µM	0.17		
Trypanosoma_brucei_5µM	0.44		-0.23
Cryptosporidium_parvum_2.3µM	0.17	0.16	-0.15
Theileria_equi_EC50	0.17		-0.3
Babesia_sp_EC50	0.12	0.1	-0.31
Pberghei_ookinete_10µM	0.52	0.56	0.16
Pberghei_ookinete_1µM	0.25	0.48	0.35
Pberghei_liver_50µM	0.54	0.36	-0.2
Pberghei_liver_5µM	0.42	0.65	0.1
Pfal_gametocyte_10µM	0.73	-0.21	-0.36
Pfal_gametocyte_5µM	0.77	-0.37	-0.13
Pfal_gametocyte_1µM	0.49	-0.31	0.53
Pfal_asexual_10µM	0.62	-0.28	
Pfal_asexual_EC50	0.46	-0.34	0.49
	PC1	PC2	PC3

Figure 3. Principal components (PC) differentiate general toxicity from *Plasmodium* specific activity. The three principal components PC1, PC2 and PC3 explained 30%, 16% and 13% of the variation in the data, respectively. PC1 showed positive correlation across different assays, suggesting that it reflects general toxicity. PC3 showed higher positive correlation only with assays conducted at lower compound concentration, but lower or negative correlation with assays conducted at higher compound concentration, other parasites and human cells, suggesting that it reflects pan-stage specific activity against *Plasmodium*.

this component reflects activity against these two stages or against *P. berghei* in which these assays were performed.

General toxicity predicts *in vitro* resistance evolution

In vitro resistance evolution has recently been attempted against 30 Malaria Box compounds with three independent lines for each compound¹¹. We next tested whether the *Plasmodium*-specific activity or general toxicity estimated from the principal components might predict *in vitro* resistance evolution. Compounds for which resistance could not be developed showed significantly higher PC1 values (Figure 4A). These compounds also showed higher human toxicity (Figure 4B) and enrichment of probe-like compounds, which have chemical properties associated with higher non-specific activity⁹ (Figure 4C). These results suggest that general toxicity of compounds may lead to lower success in *in vitro* resistance evolution. On the other hand, high PC3 values were associated with higher likelihood of *in vitro* resistance generation (Wilcox test $p = 0.02$, not shown).

Prioritization of Malaria Box compounds

In general, our results suggest that compounds that show high PC3 values should be prioritized for further development, including target identification by *in vitro* resistance evolution. Table 1 lists the top 20 Malaria Box compounds with the highest PC3 values. These compounds show high activity against multiple stages at a low concentration, but low activity against human cells. In total, 11 of these compounds show favourable oral bioavailability values. Some of these are active against other pathogens (Table 1). The values of three principal components for all Malaria Box compounds are available in Dataset 1¹⁴.

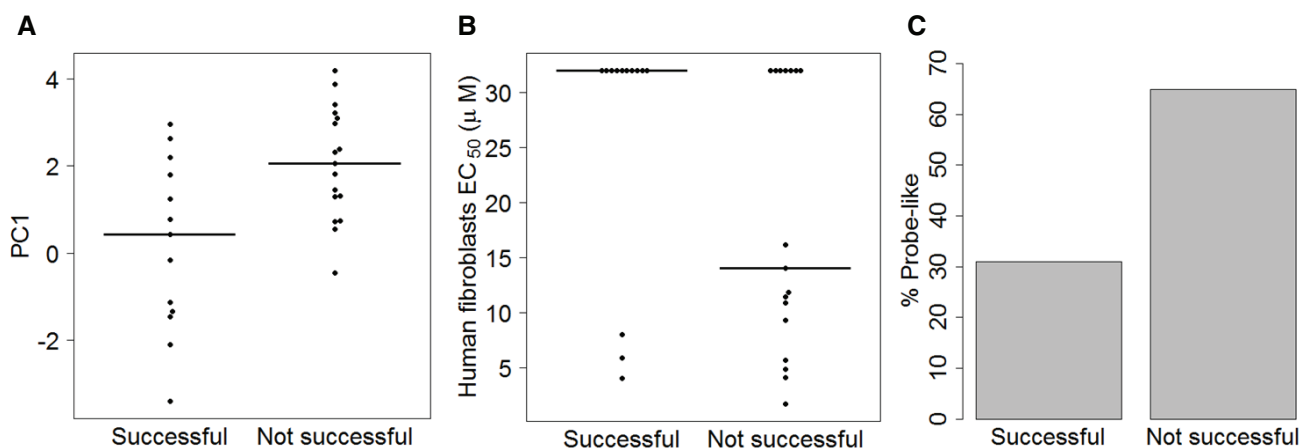


Figure 4. Higher general toxicity of compounds against which *in vitro* resistance development was not successful. The *in vitro* resistance evolution was attempted for 30 Malaria Box compounds and was successful for 13 compounds¹¹. Compounds for which resistance evolution was not successful showed (A) higher PC1 values (Wilcox p 0.004), (B) lower EC₅₀ against human fibroblasts cells (Wilcox p 0.130), and (C) a higher proportion of probe-like compounds, as classified by Medicines for Malaria Venture⁹ (Fisher p 0.100).

Table 1. Top Malaria Box compound with the highest PC3 values. These compounds show high activity against multiple stages at a low concentration, but low activity against human cells. The mouse oral bioavailability was obtained by measuring the plasma concentration of the compounds with a single high oral dose (140 $\mu\text{M}/\text{kg}$)¹⁰. Compounds with favourable plasma concentration (plasma $C_{\text{max}} > 1\mu\text{g}/\text{ml}$) are indicated. Activity of compounds against other parasites is also indicated. The oral bioavailability data and compound activity data against other parasites was obtained from Van Voorhis *et al.*¹⁰. The PC3 values for all Malaria Box compounds are available in [Dataset 1](#)¹⁴.

MMV ID	Mouse oral single dose $C_{\text{max}} > 1\mu\text{M}$	<i>Trypanosoma brucei brucei</i>	<i>Leishmania donovani</i>	<i>Trypanosoma cruzi</i>	<i>Brugia malayi</i>	<i>Schistosoma mansoni</i>
MMV007116	Y					
MMV666021						
MMV006389	Y		Y			Y
MMV007127	Y	Y	Y			
MMV020439	Y					
MMV665827						
MMV665977				Y		
MMV000563	Y	Y	Y	Y		
MMV666596						
MMV666693					Y	
MMV396678						
MMV007430	Y					
MMV665878	Y			Y		
MMV006820	Y					
MMV007228						
MMV665782						
MMV396595	Y					
MMV019266					Y	
MMV020942	Y					
MMV007199	Y					

Dataset 1. PC1, PC2 and PC3 values for 400 Malaria Box compounds

<http://dx.doi.org/10.5256/f1000research.10121.d142570>

Discussion

The wide availability of Malaria Box has catalysed a number of studies on these compounds¹⁰. Prioritization of compounds based on a large number of variables is not straightforward. Here, we analysed this data and found that a single variable (PC3) can capture most of the desirable compound properties: activity against multiple *Plasmodium* stages and low host cytotoxicity, thus greatly simplifying the task of compound prioritization. Our analyses

suggest that screening at high compound concentrations can lead to general toxicity and thus should be avoided. Thus the idea that hits identified from multiple assays should be more confident¹⁰ needs to be considered carefully when hits are identified from high concentration assays. The consensus approach might lead to the selection of compounds with general toxicity.

We found significant correlation between activity against the asexual stage and the gametocyte stage of *P. falciparum* (Spearman rank correlation 0.43), which suggests that it might be easier to find compounds that have activity against both these stages, even though the two stages have different growth properties. The correlations between asexual stage with the liver and

ookinete stages were low (Figure 1 and Figure 2). This could reflect different physiological states of liver and ookinete stages from asexual and gametocyte stages, but it might also reflect that liver and ookinete stage assays were performed in *P. berghei*, rather than *P. falciparum*. Thus, the development of higher throughput liver and ookinete stage assays in *P. falciparum* could be valuable. It is important to note that the correlation values that we report should be considered an underestimate, as inhibition values for assays against the same stage show large variability, e.g. median rank correlation among nine EC₅₀ values against asexual stage of *P. falciparum* was 0.51. The possible reasons for this variability have previously been discussed¹⁰.

The difficulty in the evolution of *in vitro* resistance is considered a very desirable property of a compound¹² given that a number of anti-malarial drugs are becoming less effective because of resistance generation¹³. Our observations suggest caution in interpreting the results of *in vitro* resistance evolution experiments. The failure to obtain resistance *in vitro* could be because of general toxicity of the compound on the erythrocyte hosts. Thus we suggest that the host toxicity of compounds should be thoroughly evaluated before conducting the labour-intensive *in vitro* resistance evolution experiments.

While we have prioritized compounds according to their pan-stage activity and low human toxicity, we would like to stress that compounds that show activity across pathogens and human cells may also be potential leads, if their toxicity could be managed. One possibility to reduce the toxicity of a compound is to identify its target in the parasite and its human ortholog, and utilize the three-dimensional structures of the compound with the target to modify the compound to increase selectivity. However, target

identification of these compounds might be more difficult using *in vitro* resistance development.

Data availability

This publication uses data on Malaria Box compounds, as reported and compiled by Van Voorhis *et al.*¹⁰ (DOI: [10.1371/journal.ppat.1005763.s002](https://doi.org/10.1371/journal.ppat.1005763.s002)). The data on the *in vitro* resistance evolution is reported by Corey *et al.*¹¹.

Dataset 1: PC1, PC2 and PC3 values for 400 Malaria Box compounds, DOI: [10.5256/f1000research.10121.d142570](https://doi.org/10.5256/f1000research.10121.d142570)¹⁴

Author contributions

G.P.S. conceived and designed the study, performed the research and wrote the manuscript.

Competing interests

No competing interests were disclosed.

Grant information

The work is supported by an Early Career Fellowship to G.P.S. by the Wellcome Trust/DBT India Alliance (IA/E/15/1/502297).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements

Suggestions and criticism on the manuscript from Dr. Amit Sharma and Ms. Preeti Goel from the author's lab are gratefully acknowledged.

References

- World Health Organization: **The World Malaria Report 2015**. World Health Organization, Geneva; 2014.
[Reference Source](#)
- Alonso PL, Brown G, Arevalo-Herrera M, *et al.*: **A research agenda to underpin malaria eradication**. *PLoS Med.* 2011; **8**(1): e1000406.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Burrows JN, van Huijsdijnen RH, Möhrle JJ, *et al.*: **Designing the next generation of medicines for malaria control and eradication**. *Malar J.* 2013; **12**: 187.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hovlid ML, Winzeler EA: **Phenotypic Screens in Antimalarial Drug Discovery**. *Trends Parasitol.* 2016; **32**(9): 697–707.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Avery VM, Bashyam S, Burrows JN, *et al.*: **Screening and hit evaluation of a chemical library against blood-stage *Plasmodium falciparum***. *Malar J.* 2014; **13**: 190.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Meister S, Plouffe DM, Kuhen KL, *et al.*: **Imaging of *Plasmodium* liver stages to drive next-generation antimalarial drug discovery**. *Science.* 2011; **334**(6061): 1372–1377.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Guiguemde WA, Shelat AA, Bouck D, *et al.*: **Chemical genetics of *Plasmodium falciparum***. *Nature.* 2010; **465**(7296): 311–315.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Gamo FJ, Sanz LM, Vidal J, *et al.*: **Thousands of chemical starting points for antimalarial lead identification**. *Nature.* 2010; **465**(7296): 305–310.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Spangenberg T, Burrows JN, Kowalczyk P, *et al.*: **The open access malaria box: a drug discovery catalyst for neglected diseases**. *PLoS One.* 2013; **8**(6): e62906.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Van Voorhis WC, Adams JH, Adelfio R, *et al.*: **Open Source Drug Discovery with the Malaria Box Compound Collection for Neglected Diseases and Beyond**. *PLoS Pathog.* 2016; **12**(7): e1005763.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Corey VC, Lukens AK, Istvan ES, *et al.*: **A broad analysis of resistance development in the malaria parasite**. *Nat Commun.* 2016; **7**: 11901.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ding XC, Ubben D, Wells TN: **A framework for assessing the risk of resistance for anti-malarials in development**. *Malar J.* 2012; **11**: 292.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Woodrow CJ, White NJ: **The clinical impact of artemisinin resistance in Southeast Asia and the potential for future spread**. *FEMS Microbiol Rev.* 2016; pii: fuw037.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Singh GP: **Dataset 1 in: Dimension reduction of Malaria Box data allows efficient compound prioritization**. *F1000Research.* 2016.
[Data Source](#)

Open Peer Review

Current Referee Status:



Version 1

Referee Report 16 January 2017

doi:10.5256/f1000research.10902.r18780



Francisco-Javier Gamo

Tres Cantos Medicines Development Campus Unit, Malaria Unit, GlaxoSmithKline, Tres Cantos, Spain

The article shows an analysis of the available biological data on the 400 compounds of the Malaria Box set trying to understand if there is any correlation between favorable parasitological properties and undesirable unspecific or toxicity aspects. Most of the data have been retrieved from Van Voorhis et al., 2016. Some of these data have been produced by GSK and this reviewer is one of the co-authors. Statistical analyses reveal some correlations that would allow prioritization of phenotypic hits based on the most desirable antimalarial compound profiles.

Reviewer finds conclusions very interesting specially those related to the potential misleading information provided by malaria assays carried out at high concentration of compound. Principal Component Analyses shows strong correlation between toxic effects in human cells and antimalarial activities produced at the highest compound concentration, suggesting these antimalarial effects not being specific for parasites. Identification of a single principal component capturing most desirable properties of malaria box compounds is an important result and could be the basis to rank compounds in larger datasets.

There are some additional points that could help to improve current version or be considered for potential follow ups.

1. Author describes the most desirable properties for novel antimalarial compounds but comment that how these properties are linked to each other is not clear. However most of them (speed of action, activity against different stages of Plasmodium, propensity to select for resistance, etc) are strongly influenced by antimalarial target. So, mode of action (MoA) should be a clear link among the different properties and compounds displaying similar MoA should show similar properties. Reviewer agrees that pharmacokinetic aspects of antimalarial compounds (e.g. long half life), are structure based and should be much less related to antimalarial MoA.
2. Despite there is still a lack of understanding on the antimalarial targets of phenotypic hits there is already some information at this regard. I suggest author to include in the analysis current knowledge on mode of action to identify those targets providing the most desirable antimalarial properties. This would be specially interesting if the analysis is expanded to larger datasets.
3. Conclusions based on resistance evolution should be put in the context of the small number of compounds with available results. Larger sets of data would be needed to confirm this trend. Nevertheless reviewer agrees with poly-pharmacology and general toxicity as two of the main properties that negatively influence selection of resistance.

Competing Interests: Reviewer is one of the co-authors of Van Voorhis et al., 2016. That paper contains most of the biological data used in current analysis. However, reviewer does not feel this to have an influence in the output of my review.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Referee Report 29 November 2016

doi:[10.5256/f1000research.10902.r18052](https://doi.org/10.5256/f1000research.10902.r18052)



Timothy N. C. Wells

Medicines for Malaria Venture (MMV), Geneva, Switzerland

This is an analysis of the Malaria Box set, and an attempt to try to understand whether there are any correlations between activity and safety signals or resistance development using Principal Component Analyses. The author defines a single principal component which can be used to capture most of the desirable compounds of the Malaria Box. The data used are from the Malaria Box and the recent summary in Van Voorhis et al., 2016 (of which this reviewer is one of the 150 co-authors).

The initial correlations show that there are some links between the different life cycle stages of inhibition for malaria, and certainly these are strong correlations compared to say the correlation between killing Plasmodium and killing distantly related worm pathogens.

The most useful insight from the paper is in Figure 4 where it's clear that

1. Compounds where no resistance could be generated have higher PC1 values.
2. Compounds where resistance could not be generated are more significantly represented by fibroblast inhibition at <10 uM.

Given that compounds where no resistance can be made are deemed by some to be part of higher value scaffolds (we all want drugs which don't generate resistance) – then this observation warrants more exploration.

Some comments for the author – which would add value to this study (or could form the basis of the next study):

1. The PCA based on cytotoxicity is built from the human fibroblast data. This is a good start, but the real value would be to use all the NCI59 cell data from the US National Cancer Institute and to see whether there is anything else that emerges.
2. As a set of compounds, the Malaria Box was the fruit of a very small investment – and so the compounds were often selected based on availability and cost. It would be good to run the same PCA analysis on the Pathogen Box structures. For pathogen box there was flexibility to make any compound that was required for the collection, and so the quality of the structures chosen was arguably higher.
3. Two other types of compound data could be analysed. First it would be interesting to see the same analysis used on the TCAMS set – does that help in some ways to reprioritize these structures.

Second, what happens with PC1 if you look at say the 8000 compounds taken in development by the Pharma industry (and not yet launched) – do they appear to be higher quality.

Competing Interests: The paper uses the data from van Voorhis et al., 2016 as the basis of its data set. I am an author on that paper, and also on the original concept of the malaria box. I have never heard of, nor met the author however, and do not feel that this influences me in any significant way.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.
