Genetic markers associated with dihydroartemisinin–piperaquine failure in *Plasmodium falciparum* malaria in Cambodia: a genotype–phenotype association study

Roberto Amato, Pharath Lim, Olivo Miotto, Chanaki Amarutanga, Dalin Dek, Richard D Pearson, Jacob Almagro-Garcia, Aaron T Neal, Sokunthea Sreng, Seila Suon, Eleanor Drury, Dushyanth Jyothi, Jim Stalker, Dominic P Kwiatkowski, Rick M Fairhurst

### Summary

**Background** As the prevalence of artemisinin-resistant *Plasmodium falciparum* malaria increases in the Greater Mekong subregion, emerging resistance to partner drugs in artemisinin combination therapies seriously threatens global efforts to treat and eliminate this disease. Molecular markers that predict failure of artemisinin combination therapy are urgently needed to monitor the spread of partner drug resistance, and to recommend alternative treatments in southeast Asia and beyond.

**Methods** We did a genome-wide association study of 297 *P falciparum* isolates from Cambodia to investigate the relationship of 11,630 exonic single-nucleotide polymorphisms (SNPs) and 43 copy number variations (CNVs) with in-vitro piperaquine 50% inhibitory concentrations (IC₅₀), and tested whether these genetic variants are markers of treatment failure with dihydroartemisinin–piperaquine. We then did a survival analysis of 133 patients to determine whether candidate molecular markers predict parasite recrudescence following dihydroartemisinin–piperaquine treatment.

**Findings** Piperaquine IC₅₀s increased significantly from 2011 to 2013 in three Cambodian provinces (2011 vs 2013 median IC₅₀: 20·0 nmol/L [IQR 13·7–29·0] vs 39·2 nmol/L [32·8–48·1] for Pursat; 19·6 nmol/L [11·9–33·9] vs 81·1 nmol/L [61·3–113·1] for Pursat; all p≤10⁻³; Kruskal-Wallis test). Genome-wide analysis of SNPs identified a chromosome 13 region that associates with raised piperaquine IC₅₀. A non-synonymous SNP (encoding a Glu415Gly substitution) in this region, within a gene encoding an exonuclease, associates with parasite recrudescence following dihydroartemisinin–piperaquine treatment. Genotype-wide analysis of CNVs revealed that a single copy of the *mdr*1 gene on chromosome 5 and a novel amplification of the *plasmpesin 2* and *plasmpesin 3* genes on chromosome 14 also associate with raised piperaquine IC₅₀. After adjusting for covariates, both *exo-E415G* and *plasmpesin 2–3* markers significantly associate with increased piperaquine IC₅₀ (p=3·0 × 10⁻⁸ and p=1·7 × 10⁻⁷, respectively) with decreased treatment efficacy (survival rates 0·38 [95% CI 0·25–0·51] and 0·41 [0·28–0·53], respectively).

**Interpretation** The *exo-E415G* SNP and *plasmpesin 2–3* amplification are markers of piperaquine resistance and dihydroartemisinin–piperacarque failure in Cambodia, and can help monitor the spread of these phenotypes into other countries of the Greater Mekong subregion, and elucidate the mechanism of piperaquine resistance. Since plasmpesins are involved in the parasite’s haemoglobin-to-haemozoin conversion pathway, targeted by related antimalarials, *plasmpesin 2–3* amplification probably mediates piperaquine resistance.

**Funding** Intramural Research Program of the US National Institute of Allergy and Infectious Diseases, National Institutes of Health, Wellcome Trust, Bill & Melinda Gates Foundation, Medical Research Council, and UK Department for International Development.

### Copyright
© The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY license.

### Introduction
Artemisinin combination therapy, the use of a short-acting artemisinin derivative and a long-acting partner drug, is recommended worldwide for the treatment of *Plasmodium falciparum* malaria.³ Treatment with dihydroartemisinin–piperaquine, a current front-line artemisinin combination therapy in Cambodia, Vietnam, Thailand, Myanmar, China, and Indonesia is now failing in Cambodian provinces where artemisinin resistance² has emerged and spread.¹¹,¹² This situation probably arose because the survival of parasites after artemisinin exposure increases the chance that they develop spontaneous genetic resistance to piperaquine, survive declining piperaquine plasma concentrations, and propagate via mosquitoes to other human beings. Through such processes, artemisinin resistance, which has evolved across the Greater Mekong subregion,2 threatens to compromise the efficacy of dihydroartemisinin–piperaquine and other artemisinin combination therapies in the global treatment and elimination...
of malaria." The natural selection of low-frequency, pre-existing resistance to piperaquine might also be occurring and further contributing to this problem.

Recent increases in treatment failures with dihydroartemisinin–piperaquine and piperaquine 50% inhibitory concentrations (IC50) suggest that piperaquine resistance has emerged in Cambodia. These findings, and the discovery that piperaquine-resistant parasites are sensitive to the former artemisinin combination therapy partner drug mefloquine, have recently led Cambodia’s national malaria control programme and WHO to recommend artesunate–mefloquine as the first-line artemisinin combination therapy in ten Cambodian provinces, including Pursat and Preah Vihear. Molecular markers are urgently needed for large-scale surveillance programmes to predict treatment failures with dihydroartemisinin–piperaquine in Cambodia and other countries in the Greater Mekong subregion, and investigate the molecular mechanism of piperaquine resistance.

Several genetic variations have been associated with decreased piperaquine susceptibility: a single copy of the mdr1 gene has been associated with dihydroartemisinin–piperaquine treatment failures in Cambodian patients, whereas amplification of a region downstream of mdr1 and the crt single-nucleotide polymorphism (SNP) C101F has been associated with raised piperaquine IC50 in vitro. These genetic variants are not useful as molecular markers, however, because the first is wild type, the second is very uncommon, and the third has not yet been observed in MalariaGEN P. falciparum Community Project’s global catalogue of variation in clinical samples. Another study found that a triple mutation (kelch13-C580Y, MAL10:688956 SNP, and MAL13:1718319 SNP) linked with slow parasite clearance is associated with a 5–4 times greater risk of dihydroartemisinin–piperaquine treatment failure in Cambodia’s Oddar Meanchey province; however, these SNPs were not linked to piperaquine resistance.

The purpose of this genome-wide association study (GWAS) was to discover molecular markers of dihydroartemisinin–piperaquine treatment failures in Cambodia. In designing it, we reasoned that such markers would associate with raised piperaquine IC50, increase in prevalence over time in areas where artemisinin resistance is common, and be uncommon where malaria has not been treated with artemisinins. The GWAS was performed on two large sets of P. falciparum whole-genome sequences from Cambodia and a large meta-analysis of single-nucleotide polymorphisms (SNPs) was performed.

The GWAS was performed on two large sets of P. falciparum whole-genome sequences from Cambodia and a large meta-analysis of single-nucleotide polymorphisms (SNPs) was performed.

Articles
dihydroartemisinin–piperine has not been used. Therefore, we sequenced the genomes and measured the piperine IC₅₀s of parasites collected in 2010–13 from patients with *P. falciparum* malaria in Pursat, Preah Vihear, and Ratanakiri, where the prevalences of kelch13 mutations, genetic markers for artemisinin resistance, were 77%, 34%, and 11%, respectively, and where the prevalences of dihydroartemisinin–piperquine treatment failures were 46%, 16%, and 2%, respectively, in 2012–13.

**Methods**

**Study design and participants**

To obtain samples for this GWAS study, we enrolled patients with uncomplicated *P. falciparum* malaria into parasite clearance rate and drug efficacy studies in 2010–13 in three provinces where piperquine resistance is common (Pursat), emerging (Preah Vihear), or uncommon (Ratanakiri). Written informed consent was given by adult patients, or a parent or guardian of child patients. Protocols were approved by the Cambodian National Ethics Committee for Health Research and the National Institute of Allergy and Infectious Diseases Institutional Review Board and are registered with ClinicalTrials.gov, numbers NCT00341003, NCT01240603, and NCT01736319.

Using parasitised blood samples from these studies, we measured piperine IC₅₀ ex vivo or after short-term culture in vitro, and obtained whole-genome parasite sequence data, whenever possible. The GWAS was designed to identify genetic markers of raised piperine IC₅₀s, whereas a dihydroartemisinin–piperine efficacy study was used to test for association between GWAS candidate markers and parasite recrudescence. PCR genotyping using *msp1*, *msp2*, and *glurp* microsatellites as genetic markers distinguished recrudescences from newly acquired infections. A summary of samples, according to province of origin and year of collection, is shown in the appendix.

**GWAS and survival analyses**

The preparation, sequencing, genotyping, and phenotyping of samples are described in the appendix. The GWAS and correction for population structure were done using a linear mixed model, implemented in FaST-LMM® version 2.06. We tested 11630 SNPs with minor allele frequency greater than 0.033, using genotypes encoded as the number of non-reference alleles (0 or 1), and excluding heterozygous calls to minimise confounding effects of mixed infections. Piperine IC₅₀ was used as the continuous dependent variable. When whole-genome sequence or phenotype data were available for both initial and recrudescent samples, we only analysed data from the recrudescent samples. Different approaches (using initial sample data only, using the phenotype and genotype of the same sample, averaging initial and recrudescent IC₅₀s) did not alter results. A relationship matrix was calculated using a subset of 6678 SNPs with minor allele frequency greater than 3% (options: --maf 0.03), missing data rate below 25% (options: --geno 0.25), and unlinked (in windows of 100 SNPs, shifted forward by ten SNPs each time, removing one from each pair of SNPs with linkage disequilibrium greater than 0.3; options: --indep-pairwise 100 10 0.3) and extracted using PLINK® version 1.07. In estimating the relationship matrix, we found that excluding proximal SNPs (within 10 kb or 100 kb from the tested variant) did not substantially affect results. Given the number of independent SNPs used, we applied Bonferroni correction to define a significance threshold of $p < 8.6 	imes 10^{-7}$ or less for GWAS analyses. We also defined a suggestive threshold of $p < 10^{-4}$ or less to identify relatively high-ranking loci.

To adjust for potential confounder effects, we treated the geographical origin of the sample, the presence of mdr1 amplification, and the presence of kelch13 resistance alleles collectively as covariates in the linear mixed model. These covariates reduced the genomic inflation factor $\lambda_{GC}$ from 3.455 to 1.964 (appendix). We also treated genetic similarity across samples as a random effect to correct for confounding effects of population structure, which further reduced $\lambda_{GC}$ to 1.06. Although this value is still slightly above 1.0, this is unlikely due to un-accounted population stratification. When we confined the estimation of $\lambda_{GC}$ to only unlinked SNPs, its value dropped to 1.03, suggesting that residual inflation is due to extended homozygosity haplotypes in the samples and thus has little effect on the GWAS.

We did a survival analysis using the R package survival. We fitted a Cox proportional hazard regression model, in which treatment success (recrudescent vs non-recrudescent infection) represented the survival status, and we added the age of the patient (in years), parasitaemia on day 0 (log-scaled), and dose of piperine given (mg/kg, in five-unit increments starting at 35 mg/kg) as covariates. We then included the two markers (*exo-E415G* and *plasmodin 2–3*) as covariates, and estimated adjusted hazard ratios (aHRs) and adjusted survival curves.

**Copy number variations**

We subsequently tested the association of copy number variations (CNVs) with piperine IC₅₀ using the same method and parameters just described. We tested 43 CNVs present in at least five samples; genotypes represented the presence or absence (encoded as 0 or 1) of the CNV. In this analysis, mdr1 amplification was not included as a covariate in the linear mixed model. To call CNVs across the genome, we modified a procedure used previously. Briefly, we first divided the genome into 300 bp non-overlapping bins and calculated for each sample the number of reads whose alignment started within each bin. We then normalised these binned read counts by dividing by the median read count of the core regions of chromosome 9. We excluded bins where guanine–cytosine (GC) content was less than 20% due to coverage bias in most samples. Copy number state for each bin was predicted in each sample by fitting a
Between July 9, 2010, and Dec 31, 2013, in three clinical isolates collected for publication.

Results

For 43 samples with excessive variation in read coverage, we inspected each sample individually and scanned them for the presence of reads spanning the breakpoints. In total, we identified 134 regions containing CNVs across the core genome; 43 of these CNVs were present in at least ten samples. IC50s were treated as normally distributed in this set of samples, and where one of the two alleles was present in at least ten samples. IC50s were treated as normally distributed in this set of samples, and where one of the two alleles was present in at least ten samples.

Role of the funding source

The funders had no role in study design, data collection, analysis, interpretation, or report writing. The corresponding authors had full access to all data in the study and final responsibility for the decision to submit for publication.

Results

We analysed 486 P. falciparum clinical isolates collected between July 9, 2010, and Dec 31, 2013, in three Cambodian provinces where artemisinin resistance is entrenched (Pursat), emerging (Preah Vihear), and uncommon (Ratanakiri; appendix).3,4,7 To monitor the evolution of piperaquine resistance, we measured piperaquine IC50s for 297 P. falciparum clinical isolates obtained directly from patients (ex vivo, 275 [93%] of 297) or following short-term culture (in vitro, 22 [7%] of 297) in 2011, 2012, and 2013. In all three provinces, IC50s increased significantly over time, especially when comparing province-stratified data from 2011 and 2013 (Kruskal-Wallis test, all p≤10−3; appendix). Despite parasites having comparable IC50s at all sites in 2011 (medians 20·0 nmol/L [IQR 13·7–29·0], 19·3 nmol/L [15·1–26·2], and 19·6 nmol/L [11·9–33·9] for Ratanakiri, Preah Vihear, and Pursat, respectively), they were remarkably differentiated at all sites by 2013, showing about two, three, and four times increases in IC50s (medians 39·2 nmol/L [IQR 32·8–48·1], 66·2 nmol/L [49·9–83·0], and 81·1 nmol/L [61·3–113·1], respectively). These regional differences are consistent with the relative prevalences of artemisinin-resistant parasites and recrudescence parasitaemias following dihydroartemisinin-piperaquine therapy.3,4,7 To investigate the genetic basis of piperaquine resistance in vitro, we did a GWAS analysis of the 297 samples (appendix) for which we had both piperaquine IC50s and whole-genome sequences (see Methods). We tested 11 630 SNPs that were well covered in this set of samples, and where one of the two alleles was present in at least ten samples. IC50s were treated as normally distributed in this set of samples, and where one of the two alleles was present in at least ten samples.

Table 1: Genome-wide single-nucleotide polymorphisms (SNPs) most strongly associated with piperaquine 50% inhibitory concentrations (IC50s)

| Chromosome number | Position | Gene ID | Gene description | N or S | Alteration | p value |
|-------------------|----------|---------|------------------|--------|------------|--------|
| Locus L13.13      | 2504560  | PF3D7_1362500 | Exonuclease, putative | N      | p.Glu415Gly | 2·69×10−3 |
| Locus L12.12      | 418346   | PF3D7_1208500 | Conserved plasmodium protein, unknown function | S      | p.Glu132Pro | 4·80×10−4 |
| Locus L12.13      | 2728402  | PF3D7_1368700 | Mitochondrial carrier protein, putative | N      | p.Asn252Asp | 2·33×10−3 |
| Locus L13.12      | 2512415  | PF3D7_1362700 | Conserved plasmodium protein, unknown function | S      | p.687Asn | 6·25×10−3 |
| Locus L13.13      | 3519091  | PF3D7_1362800 | Conserved plasmodium protein, unknown function | N      | p.Gly202Asp | 2·12×10−4 |
| Locus L13.13      | 2447416  | PF3D7_1361000 | Arginine methyltransferase 5, putative (PRMT5) | N      | p.Asn465Ser | 3·96×10−4 |
| Locus L04.4       | 904088   | PF3D7_0420000 | Zinc finger protein, putative | S      | p.213Leu | 7·67×10−4 |
| Locus L14.14      | 2411942  | PF3D7_1458700 | Conserved plasmodium protein, unknown function | N      | p.Arg25Ile | 1·27×10−3 |
| Locus L14.13      | 2395752  | PF3D7_1458300 | Conserved plasmodium protein, unknown function | N      | p.Bel301Phe | 2·03×10−4 |
| Locus L13.13      | 2542386  | PF3D7_1363300 | Mitochondrial ribosomal protein L7 precursor, putative | S      | p.Thr218Leu | 2·48×10−4 |
| Locus L12.12      | 1787729  | PF3D7_1242200 | Conserved plasmodium protein, unknown function | N      | p.Ser518Asn | 2·90×10−4 |
| Locus L04.4       | 865807   | PF3D7_0419400 | Conserved plasmodium protein, unknown function | S      | p.242Ile | 3·05×10−4 |
| Locus L06.6       | 866478   | PF3D7_0621100 | Conserved plasmodium protein, unknown function | N      | p.Asp1144Glu | 4·04×10−4 |
| Locus L12.12      | 407838   | PF3D7_1208800 | Zinc finger protein, putative | N      | p.Ser438Asn | 4·56×10−5 |
| Locus L13.13      | 2656558  | PF3D7_1366400 | Rhoptry protein (Rhop148) | S      | p.Glu526Arg | 5·02×10−5 |
| Locus L14.14      | 2385550  | PF3D7_1457900 | Conserved plasmodium protein, unknown function | S      | p.Thr252Asp | 4·30×10−5 |
| Locus L04.4       | 865709   | PF3D7_0419400 | Conserved plasmodium protein, unknown function | N      | p.Leu239Ile | 8·96×10−5 |
| Locus L13.13      | 2746916  | PF3D7_1369100 | Conserved plasmodium protein, unknown function | N      | p.Leu430Arg | 8·98×10−5 |

Significant SNPs (Bonferroni-corrected p<8·6×10−4) and suggestive SNPs (p<10−3) associated with piperaquine IC50s according to increasing p value, are shown. For each SNP, we list the locus name, chromosome number, nucleotide position, gene ID, gene description, whether the SNP is non-synonymous (N) or synonymous (S), encoded aminoacid alteration, and association p value.
as a continuous dependent variable in a linear mixed-model algorithm, and we treated the province of sample origin, presence of *mdr1* amplification, and presence of *kelch13* mutations as covariates; to correct for the confounding effect of population structure, we treated genetic similarity across samples as a random effect (appendix).

GWAS analysis identified one major locus on chromosome 13, containing two non-synonymous SNPs, that were significantly associated with raised IC₅₀s ($p \leq 8.6 \times 10^{-7}$; table 1, figure 1A, appendix). The strongest signal (2.55 times increase in IC₅₀ after adjustment, $p=2.7 \times 10^{-9}$) is from a non-synonymous SNP (referred to as *exo-E415G*) producing a Glu415Gly substitution in a
putative exonuclease. The second strongest signal \( (p=2\cdot3\times10^{-7}) \) from a non-synonymous variant is an SNP (referred to as \( mcp-N252D \)) within the same locus, producing an Asn252Asp substitution in a putative mitochondrial carrier protein 1. A locus on chromosome 12 also reaches genome-wide significance, but only with a single synonymous SNP. To exclude the possibility of a batch effect, we repeated the GWAS analysis by including the year of sample collection as a covariate in the linear mixed model (appendix). This analysis confirmed the locus on chromosome 13 as the only one significantly associated with the phenotype. When also correcting for year, \( mcp-N252D \) is the only SNP to reach genome-wide significance \( (p=7\cdot3\times10^{-7}) \), although \( exo-E415G \) is just below the threshold \( (p=1\cdot8\times10^{-9}) \). These data indicate that the chromosome 13 locus (referred to as L13) shows multiple significant GWAS hits that are robust to multiple corrections.

To prioritise the multiple GWAS hits, we analysed their geographical spread, on the assumption that piperaquine resistance is emerging in Cambodia and neighbouring countries where dihydroartemisinin–piperaquine therapy is used (appendix). Taken together, the results of the GWAS and global allele frequency survey identify \( exo-E415G \) as the best candidate SNP marker of piperaquine resistance.

To gain a better picture of its frequency distribution in Cambodia, we used Sanger sequencing of PCR-amplified DNA fragments to genotype this SNP (ie, \( exo-E415G \)) in 168 additional samples for which whole-genome sequence data were not available, from the same initial cohort of 241 patients. As expected, given the low \( p \) value of this SNP in the GWAS, the phenotype distribution differs significantly between the wild-type and the mutant alleles. When stratified by province and year, the association seems to be particularly strong in Pursat and Preah Vihear (figure 1B, appendix), and from 2012 onwards (appendix). Consistent with these results, the frequency of \( exo-E415G \) appears to increase in Pursat and Preah Vihear over time (figure 1C, table 2). In summary, \( exo-E415G \) is the best predictor of raised IC\(_50\)s in the current dataset, although this finding alone is insufficient to infer a causal role for it in mediating piperaquine resistance.

Since gene CNVs have also been associated with drug resistance in \( P \) falciparum, we investigated the association of piperaquine IC\(_{50}\)s with 43 genomic regions exhibiting CNVs in at least five samples. We did a GWAS using the same method we used in the SNP GWAS, except that genotypes were the presence or absence of the CNV in each sample. We found that two CNVs strongly associated with IC\(_{50}\)s (ps2·3\times10^{-4}; figure 2A, appendix). As expected, \( mdr1 \) amplification was associated \( (p=1\cdot4\times10^{-9}) \) with low IC\(_{50}\)s. However, the most significant association \( (p=7\cdot6\times10^{-7}) \) was found for a novel amplification on chromosome 14, encompassing two of the ten \( plasmsps \) genes in the \( P \) falciparum genome—\( plasmsp2 \) and \( plasmsp3 \)—that encode aspartic proteases involved in the haemoglobin-to-haemozoin conversion pathway targeted by quinolines.

A detailed analysis of sequencing reads aligned within the \( plasmsp \) 2–3 locus revealed that the amplification boundaries were identical in all samples.

Figure 1: Manhattan plot showing the statistical significance of single-nucleotide polymorphism (SNP) associations in the genome-wide association study (A), piperaquine 50% inhibitory concentrations (B), and \( exo-E415G \) frequency distribution (C).

Table 2: Joint distribution of the two piperaquine resistance markers in Cambodia.

| Ratanakiri | Preah Vihear | Pursat |
|------------|--------------|--------|
| 2010       | 2010         | 2010   |
| 2011       | 2011         | 2011   |
| 2012       | 2012         | 2012   |
| 2013       | 2013         | 2013   |

The number of samples that are wild-type (WT) or mutant (Mut) for the exonuclease \( exo-E415G \) mutation (\( exo-E415G \)) and have one (CN1) or multiple (CN2+) copies of \( plasmsp2-3 \), according to province and year of collection, are shown. Samples where data for one of the two markers were not available or the marker was present in a mixed infection were excluded.
The putative breakpoints lie near the 3’ end of both plasmepsin 1 and plasmepsin 3, so that each amplification creates an intact extra copy of plasmepsin 2 together with a new chimeric version of plasmepsin 3, with its 3’ end replaced by the 3’ end of plasmepsin 1 (appendix). Due to the degree of homology between plasmepsin 1 and plasmepsin 3, the aminoacid sequence of the chimeric plasmepsin 3 protein is identical to that of

**Figure 2:** Manhattan plot showing the statistical significance of copy number variation associations in the genome-wide association study (A) and piperacaine 50% inhibitory concentrations according to mdr1 and plasmepsin 2–3 copy number (B, C). (A) Each point represents one of the 43 copy number variations (CNVs) present in at least five samples, alternately coloured red and blue according to chromosome. Genomic location is shown on the x-axis. The p value for each CNV’s association, calculated using a linear mixed model, is shown on the y-axis; point size is proportional to significance level. The province of sample origin, status of kelch13 (mutant vs wild type), and a genetic relatedness matrix were added as fixed effects to the analysis. Two CNVs reached the Bonferroni-corrected, genome-wide significance level (p<2.3×10^-4; above the horizontal blue line), one including plasmepsin 2 and plasmepsin 3, and one including mdr1. All 43 CNVs are marked by black lines at the top and are listed in the appendix. Dashed grey vertical lines are plotted every 500 kb from the beginning of each chromosome. Solid grey vertical lines mark telomeric, sub-telomeric, and internal hypervariable regions. (B, C) Each point represents the piperacaine 50% inhibitory concentrations (IC50s) for a Plasmodium falciparum clinical isolate carrying wild-type (WT) or amplified (Amplification) mdr1 (B) or plasmepsin 2–3 (C) genes. Bold and thin horizontal lines indicate the median and IQR of each distribution, respectively. Filled circles identify samples also carrying exo-E415G.
plasmepsin 3, except that an asparagine residue is replaced by two consecutive lysines.

The prevalence of *plasmepsin 2–3* amplification in our cohort shows a steady increase over time in Pursat and Preah Vihear compared with Ratanakiri (appendix), and reflects our observations of increasing piperaquine IC₅₀₈ and rising prevalence of *exo-E415G* in these two provinces (table 2, appendix). Despite being on different chromosomes, *exo-E415G* and *plasmepsin 2–3* amplification are in significant linkage disequilibrium (r²=0.56, empirical p=1.6×10⁻³; appendix), and have a very similar allele frequency distribution in these Cambodian data. In particular, of the 462 samples where both markers were reliably typed, 72% (n=334) have neither marker, 19% (n=86) have both markers, and only 9% (n=42) have one of the two markers. Although the co-occurrence of the two markers in the population is interesting and surprising, the few samples where the two markers are found separately makes any conclusion regarding their relative effect on the phenotype difficult to support statistically (appendix).

To further investigate whether the *exo-E415G* and *plasmepsin 2–3* amplification markers segregate with a newly described piperaquine resistance phenotype, we subjected a subset of 12 clinical isolates in triplicate to the in-vitro piperaquine survival assay. In this assay, the survival of early ring-stage parasites following exposure to piperaquine for 48 h is assessed relative to non-exposed parasites tested in parallel. For these experiments, we selected two groups of parasites (appendix). In the first group, parasites did not recrudesce after dihydroartemisinin–piperaquine treatment, had low piperaquine IC₅₀ ex vivo and after cultivation in vitro, were wild type for *kelch13* and *exonuclease*, and carried single copies of *mdr1* and *plasmepsins 2–3*. In the second group, parasites recrudesced after dihydroartemisinin–piperaquine treatment, had high piperaquine IC₅₀ ex vivo and after cultivation in vitro, carried single copies of *mdr1*, and harbour the following mutations: *kelch13*-C580Y, *exo-E415G*, and *plasmepsin 2–3* amplification. This second group of parasites showed significantly higher piperaquine survival rates (median 61.6 nmol/L [IQR 56.8–67.0]; n=6) than the first (2.4 nmol/L [1.6–2.9]; n=6; p=0.002). The relationship between raised piperaquine survival rate and the presence of either *exo-E415G* or *plasmepsin 2–3* amplification were fully concordant.

Given that *exo-E415G* and *plasmepsin 2–3* amplification are associated with raised piperaquine IC₅₀ and survival rates, which, in turn, have been associated with parasite recrudescence, we directly tested their association with dihydroartemisinin–piperaquine failures. Of the 241 samples with clinical outcome data, we analysed a subset of 133 samples for which we had complete genetic and clinical information (appendix). Both *exo-E415G* and *plasmepsin 2–3* amplification mutants showed a highly significant enrichment in recrudescent samples (p=1.6×10⁻⁸ and 1.8×10⁻¹¹, respectively, Fisher’s exact test; appendix), with aHRs of recrudescence of 13.4 (95% CI 5.3–33.5; p=3.0×10⁻⁸) and 16.7 (5.8–48.1; p=1.7×10⁻¹⁰), respectively. Furthermore, the aHR for *plasmepsin 2–3* amplification is still significant (5.2 [95% CI 1.5–17.7]; p=8.6×10⁻³) when only kelch13 mutants were considered, suggesting that whereas artemisinin resistance is certainly a risk factor, *plasmepsin 2–3* amplification potentially has an additional, independent effect on piperaquine resistance and, ultimately, dihydroartemisinin–piperaquine treatment failure. To explicitly clarify the effect of the two markers on treatment success, we analysed the survival rate in samples with or without the two markers. We found that the adjusted survival rates at 63 days in the presence or absence of *exo-E415G* were 0.38 (95% CI 0.25–0.51) and 0.93 (0.85–0.97; figure 3A), respectively. Samples with or without *plasmepsin 2–3* amplification had survival rates of 0.41 (95% CI 0.28–0.53) and 0.95 (0.87–0.98; figure 3B), respectively. Considering that non-recrudescent infections are not necessarily indicators of drug sensitivity, these values are likely underestimated. Together, these data identify *exo-E415G* and *plasmepsin 2–3* amplification as strongly predictive markers of dihydroartemisinin–piperaquine treatment failures in Cambodia.

**Discussion**

In GWAS analyses of piperaquine IC₅₀ phenotypes, we identified two genetic markers of piperaquine resistance in vitro and of dihydroartemisinin–piperaquine treatment failures in patients: a non-synonymous SNP encoding a Glu415Gly mutation in a putative exonuclease (*exo-E415G*), and amplification of the *plasmepsin 2* and *plasmepsin 3* genes (*plasmepsin 2–3* amplification). The prevalence of these two markers has increased substantially in recent years in Pursat and Preah Vihear, where artemisinin resistance is prevalent and where dihydroartemisinin–piperaquine has been...
the front-line artemisinin combination therapy since 2008 and 2010, respectively. In a global dataset of SNP allele frequencies, the exo-E415G allele is observed only in eastern mainland southeast Asia, where dihydroartemisinin–piperazine is used, and completely absent where this artemisinin combination therapy has not been used.

Of the associated variations found in this study, plasmepsin 2–3 amplification is the strongest candidate for mediating piperazine resistance because of the role of plasmepsins in haemozoin synthesis pathways in the parasite food vacuole. Since piperazine, like chloroquine, is believed to inhibit the conversion of toxic haem moieties to non-toxic haemozoin crystals during haemoglobin digestion, it is possible that piperazine targets plasmepsin 2, plasmepsin 3, or both, and that overproduction of these plasmepsins counteracts the drug’s action. The SNP markers, exo-E415G and mcpN252D, do not lend themselves to an equally simple explanation and, until functional studies are done, it will not be possible to determine whether either is directly involved in mediating piperazine resistance, whether their role is compensatory for lost fitness in piperazine-resistant mutants, or whether they are associated with piperazine resistance simply because of their strong linkage to plasmepsin 2–3 amplification or some other functional mutation in the Cambodian parasite populations.

Assigning a role to the single copy variant of mdr1, which is associated with lower sensitivity to piperazine, is equally problematic. Although it is possible that mdr1 plays a functional part in piperazine resistance (eg, by restricting the amount of drug that enters the parasite’s food vacuole), recent changes in Cambodia’s anti-malarial drug policy have caused a decline in mefloquine pressure, which might have promoted the loss of mdr1 amplifications in the parasite populations. In view of the pronounced population structure following the emergence of artemisinin resistance in Cambodia, it is possible that the association with a single copy of mdr1 is the result of piperazine resistance having emerged in specific artemisinin-resistant, mefloquine-sensitive populations.

Taken together, these findings identify plasmepsin 2–3 amplification as the most likely candidate to be a causal variant of piperazine resistance. In the samples studied here, it is strongly linked to exo-E415G, which is on a different chromosome and might represent some other functional component of the resistance phenotype. In practice, this means that the exo-E415G SNP can be presently used as a predictive marker of piperazine resistance in Cambodia; however, this might not be the case elsewhere. Therefore, plasmepsin 2–3 amplification should be monitored in areas where dihydroartemisinin–piperazine is used, although it is somewhat more laborious to genotype. Despite unresolved questions about causal mechanism, we now have tools to identify areas where treatment failures with dihydroartemisinin–piperazine are likely to occur, and thereby to empower national malaria control programmes to make informed decisions about whether to switch to alternative artemisinin combination therapies for first-line treatment of P falciparum malaria.

Contributors
RA, PL, OM, DPK, and RMF designed the study. RA, PL, CA, DD, ATN, SSr, SSu, and ED collected data. RA, PL, OM, RDP, JA-G, DJ, JS, DPK, and RMF analysed and interpreted data. RA, PL, DPK, and RMF wrote the manuscript.

Declaration of interests
We declare no competing interests.

Acknowledgments
This work was funded by the Intramural Research Program of the National Institute of Allergy and Infectious Diseases, US National Institutes of Health; Bill & Melinda Gates Foundation (OPP1040463, OPP1188166); Medical Research Council (G0600718); and UK Department for International Development (MO06212). This publication uses data from the MalariaGEN Plasmodium falciparum Community Project. Genome sequencing was done by the Wellcome Trust Sanger Institute and the Community Project was coordinated by the MalariaGEN Resource Centre with funding from the Wellcome Trust (098051, 090770). We thank all patients for participating in this study; the staff of the WTSI Sample Logistics, Sequencing, and Informatics facilities for their contribution; Marcus Lee for helpful discussions; and Robert Gwadz and Thomas Wellemes for their efforts in support of this work.

References
1 WHO. Guidelines for the treatment of malaria, 3rd edition. Geneva: World Health Organization Press, 2015.
2 Fairhurst RM. Understanding artemisinin-resistant malaria: what a difference a year makes. Curr Opin Infect Dis 2015; 28: 417–25.
3 Amarantunga C, Cheng S, Suon S, et al. Artemisinin-resistant Plasmodium falciparum in Pursat province, western Cambodia: a parasite clearance rate study. Lancet Infect Dis 2012; 12: 851–58.
4 Amarantunga C, Lim P, Suon S, et al. Dihydroartemisinin–piperazine resistance in Plasmodium falciparum malaria in Cambodia: a multisite prospective cohort study. Lancet Infect Dis 2016; 16: 357–65.
5 Dondorp AM, Nosten F, Yi P, et al. Artemisinin resistance in Plasmodium falciparum malaria. N Engl J Med 2009; 361: 455–67.
6 Arisz F, Witkowski B, Amarantunga C, et al. A molecular marker of artemisinin-resistant Plasmodium falciparum malaria. Nature 2014; 505: 50–55.
7 Ashley EA, Dhoura M, Fairhurst RM, et al. Spread of artemisinin resistance in Plasmodium falciparum malaria. N Engl J Med 2014; 371: 484–85.
8 Takala-Harrison S, Jacob CG, Arce Z, et al. Independent emergence of artemisinin resistance mutations among Plasmodium falciparum in southeast Asia. J Infect Dis 2015; 211: 670–79.
9 Leong R, Barrette A, Bouth DM, et al. Efficacy of dihydroartemisinin–piperazine for treatment of uncomplicated Plasmodium falciparum and Plasmodium vivax in Cambodia, 2008 to 2010. Antimicrob Agents Chemother 2013; 57: 818–26.
10 Saunders DL, Vanachayankul P, Lon C, et al. Dihydroartemisinin–piperazine failure in Cambodia. N Engl J Med 2014; 371: 16: 357–65.
11 Spring MD, Lin JT, Manning JE, et al. Dihydroartemisinin–piperazine failure associated with a triple mutant including kelch13 C5807Y in Cambodia: an observational cohort study. Lancet Infect Dis 2015; 15: 683–91.
12 Leong R, Taylor WR, Bouth DM, et al. Evidence of Plasmodium falciparum malaria multidrug resistance to artemisinin and piperazine in western Cambodia: dihydroartemisinin–piperazine open-label multicenter clinical assessment. Antimicrob Agents Chemother 2015; 59: 4719–26.
13 Tun KM, Imwong M, Lwin KM, et al. Spread of artemisinin-resistant *Plasmodium falciparum* in Myanmar: a cross-sectional survey of the K13 molecular marker. *Lancet Infect Dis* 2015; 15: 415–21.

14 Dondorp AM, Fairhurst RM, Slutsker L, et al. The threat of artemisinin-resistant malaria. *N Engl J Med* 2011; 365: 1073–75.

15 Chaorattanakawee S, Saunders DL, Sea D, et al. Ex vivo drug susceptibility testing and molecular profiling of clinical *Plasmodium falciparum* isolates from Cambodia from 2008 to 2013 suggest emerging piperazine resistance. *Antimicrob Agents Chemother* 2015; 59: 4631–43.

16 Eastman RT, Dharia NY, Winzeler EA, Fidock DA. Piperazine resistance is associated with a copy number variation on chromosome 5 in drug-pressed *Plasmodium falciparum* parasites. *Antimicrob Agents Chemother* 2011; 55: 3908–16.

17 Motoo O, Amato R, Ashley EA, et al. Genetic architecture of artemisinin-resistant *Plasmodium falciparum*. *Nat Genet* 2015; 47: 226–34.

18 Snounou G, Beck HP. The use of PCR genotyping in the assessment of recrudescence or reinfection after antimalarial drug treatment. *Parasitol Today* 1998; 14: 462–67.

19 Lippert C, Listgarten J, Liu Y, Kadie CM, Davidson RI, Heckerman D. FaST linear mixed models for genome-wide association studies. *Nat Methods* 2011; 8: 813–15.

20 Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; 81: 559–75.

21 WorldWide Antimalarial Resistance Network (WWARN) DP Study Group. The effect of dosing regimens on the antimalarial efficacy of dihydroartemisinin-piperaquine: a pooled analysis of individual patient data. *PLoS Med* 2013; 10: e1001564; discussion e1001564.

22 Miles A, Iqbal Z, Vauterin P, et al. Indels, structural variation, and recombination drive genomic diversity in *Plasmodium falciparum*. *Genome Res* 2016; 26: 1288–99.

23 Pearson RD, Amato R, Auburn S, et al. Genomic analysis of local variation and recent evolution in *Plasmodium vivax*. *Nat Genet* 2016; 48: 959–64.

24 Price RN, Uhlemann AC, Brockman A, et al. MQ resistance in *Plasmodium falciparum* and increased *pfmdr1* gene copy number. *Lancet* 2004; 364: 438–47.

25 Chugh M, Sundararaman V, Kumar S, et al. Protein complex directs hemoglobin-to-hemozoin formation in *Plasmodium falciparum*. *Proc Natl Acad Sci USA* 2013; 110: 5392–97.

26 Duru V, Khim N, Leang R, et al. *Plasmodium falciparum* dihydroartemisinin-piperaquine failures in Cambodia are associated with mutant K13 parasites presenting high survival rates in novel piperaquine in vitro assays: retrospective and prospective investigations. *BMC Med* 2015; 13: 305.

27 Motoo O, Almagro-Garcia J, Manske M, et al. Multiple populations of artemisinin-resistant *Plasmodium falciparum* in Cambodia. *Nat Genet* 2013; 45: 648–53.