STUDY DESIGNS

Harnessing genomics and genome biology to understand malaria biology

Sarah K. Volkman[1,2,3], Daniel E. Neafsey[2], Stephen F. Schaffner[2], Daniel J. Park[2,4] and Dyann F. Wirth[1,2]

Abstract | Malaria is an important human disease and is the target of a global eradication campaign. New technological and informatics advancements in population genomics are being leveraged to identify genetic loci under selection in the malaria parasite and to find variants that are associated with key clinical phenotypes, such as drug resistance. This article provides a timely Review of how population-genetics-based strategies are being applied to *Plasmodium falciparum* both to identify genetic loci as key targets of interventions and to develop monitoring and surveillance tools that are crucial for the successful elimination and eradication of malaria.

[1]*Department of Immunology and Infectious Disease, Harvard School of Public Health, 665 Huntington Avenue, Boston, Massachusetts 02115, USA.*
[2]*Broad Institute, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA.*
[3]*School for Nursing and Health Sciences, Simmons College, 300 The Fenway, Boston, Massachusetts 02115, USA.*
[4]*Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, Massachusetts 02138, USA.*
Corresponding author S.K.V. e-mail: svolkman@hsph. harvard.edu

The World Health Organization (WHO) estimates that close to one million children die from malaria every year[1], and the highest mortality is found among African children. Increased control efforts have reduced the malaria burden in some areas, but evidence of rebounding malaria[2] brings these general trends into question and confirms that we have much to learn to defeat this important human pathogen. It is now more important than ever to understand key aspects of malaria biology and transmission, to identify targets that are vulnerable to intervention strategies and to create tools to interpret the changing landscape of infection. One powerful approach uses population-biology-based investigations to provide crucial insight into the causes and spread of disease. This strategy aids biological discovery by using population structure and genetic diversity to identify loci under selection or associated with clinical phenotypes and to develop tools for monitoring and evaluating interventions.

Plasmodium falciparum is a eukaryotic pathogen with a complex life cycle, spending part of its lifespan in its definitive host, the *Anopheles* mosquito, as a mostly diploid organism, and the remainder of the time in its human host as a haploid organism, where it gives rise to numerous clinical manifestations from mild to life-threatening illness. The 24 Mb genome of the parasite is distributed among 14 linear chromosomes, and the parasite contains two extra-chromosomal DNA circles that constitute the apicoplast and mitochondrial genomes. The full *P. falciparum* genome sequence was published in 2002 (REF. 3) and was followed by the publication of other *Plasmodium* genome sequences,

including that of *Plasmodium vivax*[4]. These data have allowed the elucidation of basic genome architecture and the identification of key structural elements, common metabolic and biosynthesis pathways and unique aspects that are shared among several *Plasmodium* parasites[3–6]. *P. vivax* causes extensive human malaria globally, but genomic data for this species are limited[4,7]. We therefore focus on *P. falciparum* while recognizing the need to develop and apply population-genomics approaches to *P. vivax*.

The *P. falciparum* genome is evolving in response to natural selection pressures of the human host immune system, the mosquito vector and various environmental factors, including drug treatment and changes in transmission intensity owing to specific interventions[8–10]. The data imply that parasites can 'escape' both natural and artificial selection through evolution and further provide a historical 'roadmap' of its evolutionary path. These changes to *P. falciparum* population structure can potentially be leveraged to identify and to circumvent survival strategies used by the parasite with an eye towards, for example, new drug and vaccine development for malaria control, elimination and eradication. Indeed, similar approaches have been applied to much simpler viral genomes for tracking influenza outbreaks, for example, and for developing effective influenza vaccines[11].

In this Review, we discuss opportunities that are afforded by new technologies for genomic analyses to understand the genetic diversity of *P. falciparum* further[12–14] and to identify signatures of selection. We also describe genetic studies to reveal loci that are connected to the important clinical phenotype of drug

REVIEWS

resistance. Finally, we discuss population structure and its implications for assessing the impact of intervention strategies and for monitoring malaria transmission dynamics to inform best practices for global eradication campaigns.

Technological advancements

To exploit *P. falciparum* genomics fully for studies of malaria, a new toolkit is being developed. Various genome-wide array-based methods have been developed for whole-genome genotyping[13,15–20] that can capture hundreds to thousands of markers. Depending on the technology, these methods can be used to characterize genomic variation at different levels, including SNPs, microsatellite variations (MSVs), insertions or deletions (indels) and copy number variants (CNVs). However, array-based methods need to be custom-built for each organism, requiring specific tools for *P. falciparum*[13,15–20] and *P. vivax*[21] to be constructed. Although some arrays remain valuable — for example, for identifying CNVs

— next-generation sequencing methodologies are similar in cost to array-based methods and provide information about a much larger number of SNP markers without the species-specific customization that is required of arrays. Consequently, additional *P. falciparum* genomes have been sequenced at a greater depth to identify essentially all genetic variation in those genomes, thus allowing delineation of the composition and relative proportions of parasite types within a human infection. In parallel, advances in functional genomics, including transcriptome, proteome and metabolome analysis, provide valuable insight into the basic molecular functions of the parasite. An important toolkit has been developed for such 'omics' studies (BOX 1). Interrogating the diverse populations of *P. falciparum* in humans has also required a shift in focus from animal model systems towards analysing material that has been directly isolated from infected human patients (BOX 2).

These practical advancements are enhanced by several key improvements in bioinformatics analysis, many of which were developed for the analysis of other organisms and driven by international collaborative efforts, such as the Human Genome Project[22]. New computational strategies for the identification of SNPs and other variants, including MSVs, indels and CNVs from sequencing data, are being developed and applied to *P. falciparum*, providing additional markers, such as CNVs, that are potentially associated with drug resistance. Availability of genome sequences for closely related species, including *Plasmodium reichenowi*[12] (which infects non-human primates)[23], will greatly advance our ability to make population genomic inferences by identifying derived alleles for selection analysis.

After it has been obtained, this rich genomic information can yield important information about the biology of the malaria parasite, as described in the following sections. Applications include gaining insights into parasite population structure and interventions or identifying important regions of the genome — either those that show evidence of evolutionary selection or specific loci identified by genome-wide association studies (GWASs) or mutant screens — that are responsible for particular parasite phenotypes.

P. falciparum population structure and LD

Genetic variation in parasites[12–14] reveals the exposure history of a given parasite or parasite population to selective pressures. Central to our understanding of genetic variation in *P. falciparum* is determining its current population structure, including how allele frequencies vary between different populations within the species, and the degree to which alleles at neighbouring variant sites are correlated by linkage disequilibrium (LD). Most data that are used to inform our current understanding of population structure and LD are derived from either genome-wide array-based methods[16,19,20,24,25] or from SNP information provided by sequencing data[12–14] from a few dozen published genomes. Genomic structure among isolates has been delineated using principal components analysis (PCA), from which one can infer the relatedness of samples.


Box 1 | **Transcriptome, proteome and metabolome analyses**

One of the great challenges is to understand the functional role for loci that are identified as genomically diverse or under selection in the parasite genome. Additional insight can be obtained by interrogating the parasite transcriptome, proteome and metabolome, and these technologies have been developed for *Plasmodium falciparum*. Perhaps one of the most interesting features of *P. falciparum* is the temporal and coordinated transcriptional pathway that was originally described for two strains cultured in the laboratory[93,94]. External perturbations have little impact on the transcriptional pattern, prompting work towards understanding the regulatory mechanisms that govern transcription[93–99]. Other investigations into expression patterns in parasites that are derived directly from patients[100–102] have also been undertaken, including investigation of how these patterns relate to key clinical phenotypes, such as parasite clearance associated with artemisinin resistance[103]. One major class of molecules that are involved in regulating gene transcription are the apicomplexan apetala 2 (AP2) proteins[104,105].

Additionally, epigenetic regulation involves methylation to influence chromatin remodelling and the expression of *var* virulence factor genes[106–111]. Recently, a family of long non-coding RNA (lncRNA) molecules that are coordinately expressed after DNA replication were identified in *P. falciparum*[112]. lncRNAs are important regulators of gene expression in a number of systems and may be important for modulating parasite virulence by contributing to a mechanism of antigenic variation that allows the parasite to evade host immune responses[113]. An important set of experiments would involve comparing genotypically diverse parasites with regard to their temporal transcription pattern, AP2 protein expression, epigenetic markers and lncRNA expression patterns, coupled with an expression quantitative trait linkage analysis[53,114], as has been done in human cell systems.

Both proteomic[6,115–121] and metabolomic[122–125] analyses have been published for *P. falciparum* but only for laboratory-derived strains and only under a limited set of conditions. Because proteins and, perhaps more importantly, metabolites reflect differences in parasite physiology, it will be important to relate changes in these parameters to genetic changes in parasites. Such association studies have been completed using parasites that are derived from a genetic cross under different conditions of perturbation to identify metabolites that distinguish these parasites. These approaches should be expanded to a genetically diverse set of parasites to associate changes in metabolites to specific genotypic changes through association studies. Protein modifications have important roles in cellular regulation and methodologies for system-wide mapping of modifications, including histone modifications and the phosphoproteome have been adapted for analysis of *P. falciparum*. These have yet to be applied to diverse parasite strains and, considering the indication that genes encoding proteins that are involved in protein processing or degradation and cell signalling have strong signals of selection, analysis of proteome variation (rather than just genetic variation) using GWAS approaches is warranted.

Box 2 | Technological advancements towards sample processing

— Host DNA	— 140 bp RNA oligo 'bait'	— Parasite DNA

a

b
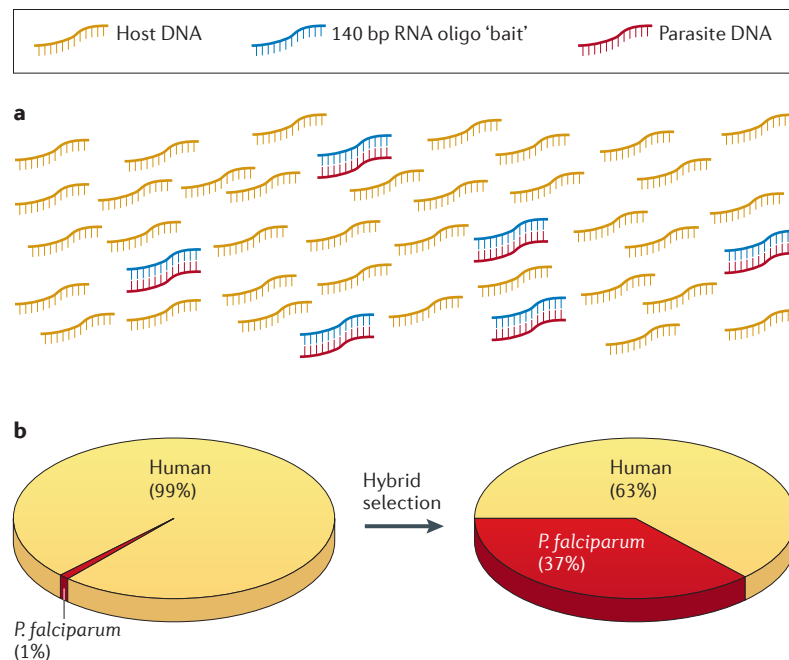
Genomic studies are hampered by a limited number of appropriate samples — a problem that is particularly acute for *Plasmodium falciparum*. Two general sample types are used for genomic analysis of *P. falciparum*: DNA derived from parasites that are adapted to tissue culture[52] and DNA from direct patient samples. Although culture adaptation is technically challenging, these parasites can be cloned by limiting dilution to enable easy acquisition of DNA from individual parasites that is void of contaminating human genetic material. These samples allow for robust and reproducible classification for various *in vitro* phenotypes, including drug response, invasion or growth parameters, gametocyte production or metabolic profiles. Questions about how representative these parasites adapted to growth in tissue culture are compared with parasites that have been directly taken from infected patients have been raised[126,127]; we are now able to apply genomic methods to answer this question.

Alternatively, DNA can be directly obtained from patients who are infected with malaria. The challenges here include the difficulty in obtaining robust phenotypic information, possible mixtures of several parasite genomes in the sample and contamination with human DNA. New technological approaches have dealt with the latter obstacle, as DNA from these samples can be depleted of contaminating human material at the time of collection using approaches involving white blood cell depletion (by Plasmodipur or CF-11 filtration[128]) or after DNA has been extracted using a hybrid selection protocol[129]. In the example of hybrid selection (see part **a** of the figure), biotinylated RNA probes ('baits') that are complementary to the *P. falciparum* genome are generated and incubated with DNA that has been extracted from patient blood containing both human and *P. falciparum* DNA. The RNA baits preferentially hybridize to the *P. falciparum* DNA in the mixture, and the human DNA is washed away. Captured *P. falciparum* DNA may then be eluted and amplified for sequencing. A DNA sample containing 1% *P. falciparum* DNA is subjected to hybrid selection, resulting in an enrichment of *P. falciparum* to 37% of the total DNA retained (see part **b** of the figure).

These purification processes result in substantial enrichment of *P. falciparum* DNA in the sample, making sequencing possible and cost-effective. As technical and informatics improvements are expanded, many more samples with valuable phenotypic information will be available for genomic analysis.

Genomic diversity across world regions. Advances in genomics technology have enabled analysis of many different parasites derived from distinct geographic locations. Large-scale population structure in *P. falciparum* follows continental lines, and there are major branches in Africa, South and Central America and South and East Asia (extending to Papua New Guinea)[19,26]. The picture within each group depends on the region. Within Africa, population differences between countries have ranged from undetectable (as in Uganda versus Congo[27] and Cameroon versus Congo[28]) or very small (as in Zimbabwe versus Uganda and Congo[27]) to modest over the longest distances (as in Nigeria versus Sudan versus South Africa[29]). By contrast, in South East Asia and the western hemisphere, local population structure is pronounced even within a single country[14,26]. The overall picture is consistent with a recent geographic spread from a source population in Africa, which has remained by far the largest population. Estimates of genetic diversity measurements present the same picture: the highest values consistently occur within Africa, and the lowest values occur in the Americas. From these types of analyses, we can identify mutations that are fixed in one parasite population but that are distinct from other populations and thus may be useful to identify specific parasites. Knowledge of the genetic characteristics of geographically distinct populations is important for controlling for population stratification in GWASs, tracking persistent parasite types as interventions are applied and localizing new sources of infection to maximize the effectiveness of control measures.

Explanations for observed LD structure. Levels of genomic diversity in different populations are reflected in the observed patterns of LD. Very little LD is seen in Africa, and it extends less than 1 kb[14,26]. LD is slightly higher in South East Asia (mean squared correlation coefficient (r^2) = 0.3 for markers that are less than 1 kb apart) and more so in South America (mean r^2 = 0.5 for markers that are less than 1 kb apart), where it spans ~10 kb. This difference could stem from demographic history. For example, population bottlenecks in the non-African populations could have eliminated many allele combinations, leaving strong correlations, and thus high LD, in the remaining parasites. It could also stem from a smaller effective population size outside Africa, which can sustain fewer allele combinations, or from lower outcrossing rates that small population sizes entail, both of which contribute to higher levels of LD. Detailed understanding of recombination and of the demographic history of different populations is needed to distinguish the two causes.

LD patterns also reflect the transmission history of strains within that population. The sexual phase of the *P. falciparum* life cycle occurs in the midgut of female *Anopheles* mosquitoes, following consumption of a blood meal containing male and female *P. falciparum* gametocytes. Because female *Anopheles* mosquitoes typically only bite a single human host during each egg-brooding cycle, the gametocyte pool that is available for sexual union in the midgut matches the gametocyte composition in individual infected human hosts. In geographical regions with high levels of malaria transmission, a high complexity of infection (COI) is thought to be produced through 'super-infection' (that is, multiple bites from distinct, *P. falciparum*-infected mosquitoes), although new evidence supports a model of co-infection

of distinct *P. falciparum* parasites from single bites[30]. A high COI makes recombination (outcrossing) possible between genetically distinct *P. falciparum* gametocytes during the sexual phase (providing that genetic diversity is high enough to ensure genetically distinct parasites) and results in short blocks of LD. By contrast, if only a single *P. falciparum* strain is present, the gametocytes will be identical, and the lack of recombination will result in longer blocks of LD.

Ultimately, as parasite population sizes get extremely small, it would be anticipated that LD should become extended and should theoretically approach a value of one. Thus, tools for measuring changes in parasite population structure have the potential to inform reductions in malaria transmission, as outcrossing rates are reduced to the point at which self-fertilizing (selfing) among parasites occurs in a given population, such as might be expected during a successful intervention strategy, as discussed later.

Signatures of selection

Of the two host environments of *P. falciparum*, selection by the human host rather than the mosquito places the most pressure on the parasite. Evidence abounds from recent studies of parasite diversity for two broad classes of strong natural selection in the parasite genome (FIG. 1).

Balancing selection. Genotyping and sequence analysis indicate that an unusually large fraction of the *P. falciparum* genome exhibits the polymorphism profile of immune-mediated balancing selection: a high density of high-frequency polymorphisms is seen in hundreds of antigenic genes[9,10]. Balancing selection maintains polymorphisms with the potential to encode alternative immunological identities and indefinitely keeps them at an intermediate population frequency (FIG. 1a)

When parasite populations are geographically separated, genes that are subject to balancing selection are unlikely to diverge as rapidly as other genes, because the selection prevents differences from differentially fixing in the populations. Genome-wide comparisons of diversity (within a population, denoted by π) and divergence (between populations, measured by the fixation index (F_{ST})) identify genetic loci that are more likely to be affected by this diversifying selection[25] in that they exhibit elevated diversity with low divergence. (FIGURE 1b shows an example of diversity and divergence analysis between parasites from Senegal and Thailand[25].) From these analyses, known antigens and vaccine candidates are identified, as well as novel genetic loci that encode putative antigens that trigger the human immune response. This prediction was validated when several highly polymorphic genes were expressed and recognized by human immune sera, including seven previously unknown antigens[13]. This result suggests that diverse genomic regions may encode antigenic loci that are useful for vaccine approaches. However, a number of vaccine studies suggest that the ability to target a polymorphic locus successfully, such as merozoite surface protein 1 (MSP1)[31] or apical membrane antigen 1 (AMA1)[32], may be undermined by the parasite's ability

Figure 1 | **Signatures of selection. a** | In a given population (shown by the large grey circle), there are numerous individuals (shown by the square matrix of circles) each containing alleles (shown by the red or yellow circles) across their genomes. Diversity refers to the amount of allelic variation among individuals in a population, whereas divergence refers to the amount of allelic variation between different populations. Under balancing selection, high diversity would be expected at a locus under selection but low divergence would be expected between populations. **b** | Distribution of loci based on both within-population differences (diversity, as denoted by π) and between-population differences (divergence, as measured by the fixation index (F_{ST}), between parasites from Senegal and Thailand) is shown. Loci that are classified as transporters or enzymes, including the acyl-CoA-synthetase (ACS) genes, are shown as blue diamonds, loci that are classified as antigens, including var, rifin, stevor and surfin molecules, are shown as red diamonds, and all other loci are shown as grey diamonds. Genes near the x axis (which have a high diversity and a low divergence) are under diversifying selection, and these include a number of known antigens. By contrast, genes near the y axis (which have a low diversity and a high divergence) are more likely to be under directional selection, and these include known drug resistance genes. **c** | Selective sweep as a consequence of selection for drug-resistant parasites results in the reduction of diversity (shown by the red line) compared with average diversity values for that genomic region (shown by the blue line). The sweep is caused by selection for an allele (shown by the red box) that confers survival under drug pressure. Neighbouring alleles are maintained along with the advantageous allele, resulting in a fairly large area of the genome with reduced diversity. Identification of genomic regions with reduced diversity in drug-resistant parasites as a consequence of directional selection reveals candidate drug-resistant genes. **d** | A haplotype bifurcation diagram[36] visualizes long-range associations for a given SNP. The thickness of the line represents the relative frequency of each haplotype in the population under study. Although the long-range associations between the ancestral (A) allele have been whittled away by recombination, the derived (T) allele maintains long-range associations with other SNPs, suggesting that it arose recently and that insufficient time has passed for recombination to break down these associations substantially. Part **b** of the figure is adapted from REF. 25. Part **d** of the figure is adapted, with permission, from REF. 14 © (2007) Macmillan Publishers Ltd. All rights reserved.

to survive the elicitation of a locus-specific immune response. Thus, strategies for using a combination of non-variant but immunogenic vaccine targets may be warranted.

Diversity and divergence analyses can identify loci that, conversely, diverge between parasite populations ($F_{ST} > 0.4$). Divergent loci between populations from Senegal and Thailand[25] encode proteins that are proposed to have various cellular functions, including DNA replication (for example, *PF10_0165*, *PF14_0278* and *PF14_0316*), lipid metabolism (for example, *PFB0695c*, *PFE1250w*, *PFB0685c* and *PFC0050c*), gametocytogenesis or sexual development molecules (for example,

---

Principal components analysis
(PCA). A statistical method used to simplify data sets by transforming a series of correlated variables into a smaller number of uncorrelated factors. It is commonly used to correct for stratification in genome-wide association studies.

Population stratification
The presence of multiple population subgroups that show limited interbreeding. When such subgroups differ in both allele frequency and disease prevalence, this can lead to erroneous results in association studies.
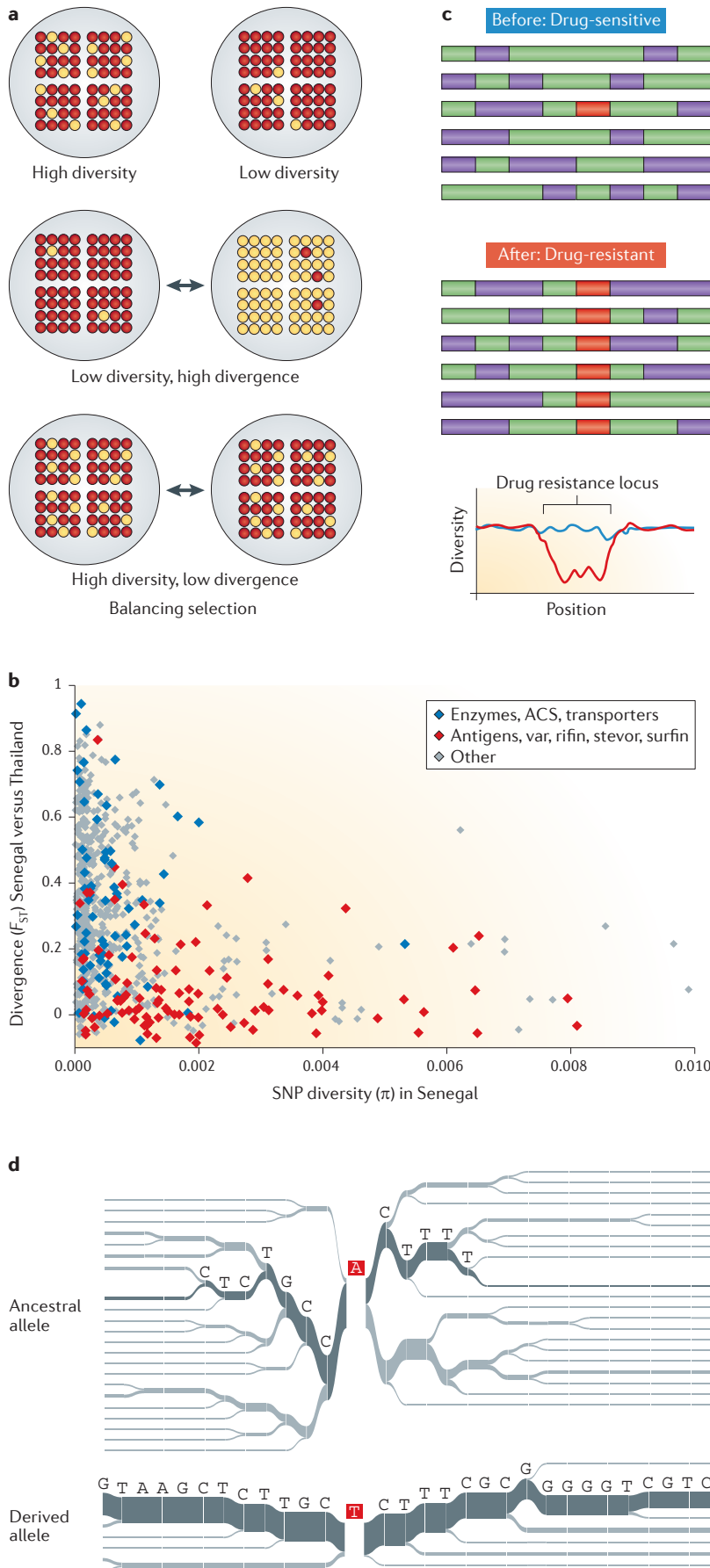
Effective population size
The number of individuals in a population genetic model that most closely matches the allele frequency distributions in the data. This can differ from actual population size owing to historical demographic events, such as population bottlenecks, migration and other factors.

Complexity of infection
(COI). The number of different parasite types, as determined by counting distinct parasite genomes, within an infected individual.

Balancing selection
Selection that favours the maintenance of more than one polymorphic allele in a population by mechanisms such as frequency-dependent selection or heterozygote advantage.

a

High diversity   Low diversity

Low diversity, high divergence

High diversity, low divergence

Balancing selection

b

- ◆ Enzymes, ACS, transporters
- ◆ Antigens, var, rifin, stevor, surfin
- ◆ Other

Divergence (F_{ST}) Senegal versus Thailand

SNP diversity (π) in Senegal

c

Before: Drug-sensitive

After: Drug-resistant

Drug resistance locus

Diversity

Position

d

Ancestral allele

Derived allele

PF13_0248 and *PFC0640w*) and transporters (for example, *PFL1125w*, *PF14_0342* and *PF14_0455*) (FIG. 1b). The reasons for the divergence are currently unclear and require further investigation but may be a consequence of differences in vector populations or other distinct selective pressures between Senegal and Thailand.

Directional selection. Directional selection in the context of a traditional selective sweep[33] leaves a distinctive genomic imprint that consists of depleted polymorphism and enhanced LD (FIG. 1c,d). This genomic signature is detectable via 'haplotype-based' tests of natural selection, such as the long-range haplotype test (LRH test). In *P. falciparum*, there is generally short LD as a consequence of considerable recombination[35]. In response to strong selective pressures, long haplotype signals that result from the rapid rise of variants that are linked to flanking mutations are easily detected as they stand out from the normally short LD of the genomic background[36] (FIG. 1d). It is important to note that these signals may be absent if directional selection began on common or standing variation[37]. Equally important is the difficulty in identifying a clear demarcation between selective sweeps and neutral processes without having a detailed understanding of demographic history and recombination rate variation — knowledge that is lacking for *P. falciparum*. Nevertheless, in genome-wide scans for selective sweeps, a number of loci show strong evidence for recent directional selection, and they all point to a single, recent evolutionary pressure: drugs. Loci that are known to confer resistance to formerly effective anti-malarial drugs — including the chloroquine resistance transporter (*pfcrt*) for chloroquine[38] and bifunctional dihydrofolate reductase thymidylate synthase (*dhfr-ts*) for pyrimethamine[39] — show all of the signs: a local desert of diversity and a strong LD between those SNPs found in the swept region (FIG. 1c). Other key modifier genes, including *P. falciparum* multiple drug resistance gene (*pfmdr1*)[40,41] and the GTP cyclohydrolase gene (*gch1*)[18,42], have been implicated in some drug responses, generally through adaptive changes in their copy number.

A plethora of additional genes show weaker evidence for recent positive selection[24,25,43], raising the possibility that they are also associated with drug responses. Their products have various putative functional roles[25], including cell surface adhesion, membrane transport, genome maintenance, transcriptional regulation, metabolism and post-translational modification, such as ubiquitylation. Evidence for sweeps at multiple genes in a single pathway suggests that selection has been involved. For example, several genes in the ubiquitylation pathway[44] are under positive selection, as seen in worldwide populations and in a thorough population analysis of parasites recently isolated in Senegal. Similarly, proteins in the fatty acid and lipid metabolism pathway have among the highest signals of selection, implicating the human or mosquito physiological state as strong selective forces on parasite survival and propagation. Key to the success of these approaches is the functional characterization of candidate loci and demonstrating their involvement in conferring important clinical phenotypes, such as

drug resistance[25]. Preliminary functional data suggest that members of the cytoadherence-linked asexual gene (*clag*) family can modulate parasite drug responses[45]: a finding that is consistent with the observation that some surface molecules are under positive selection[24,25].

Demographic confounders. As mentioned above, genomic patterns that are indicative of selection can be difficult to interpret if they can also be caused by demographic history. For example, population bottlenecks and expansion, which have occurred in the *P. falciparum* lineage[46], can also reduce diversity, increase haplotype lengths and alter allele frequency ranges; in these instances, assessing the significance of any selection test result usually requires computational simulations across a range of reasonable demographic scenarios[47]. Past studies that have estimated the demographic parameters for *P. falciparum* (that is, the ages and scales of bottlenecks and expansions) vary somewhat in their methodologies and outcomes[46,48,49], but truly genome-wide studies[24,25] benefit from being able to use most of the genome as the null distribution, using the assumption that most SNPs reflect demography more than selection. Moreover, recent computational methods relax that assumption by jointly modelling demography and selection together[50,51].

Monitoring transmission using signatures of selection. Understanding the relationship between the distribution of mutations within and between parasite populations has important implications for the discovery of alleles under either balancing or directional selection, and knowledge of this variation can potentially provide powerful tools for monitoring malaria transmission. For example, identifying divergent alleles that have become fixed in distinct populations could be useful as biomarkers to identify sources of new infection in an endemic population when transmission becomes very low or to distinguish whether new cases are coming from changing epidemiology within that endemic population or from distinct areas of transmission. This knowledge has great use in terms of informing the types and timing of intervention strategies ultimately to eliminate malaria in those places where it is endemic.

Identification of causal loci for specific phenotypes

Parasite phenotypes. A key challenge for identifying causal loci from genomic or functional studies is the classification and quantification of robust and reliable phenotypes. Clinical phenotypes that are related to pathogenesis (such as anaemia or severe disease), immunity or parasite clearance rates (that are associated with drug resistance) are the most informative and most reliable when they are assessed in parasites that are directly taken from the patient during a natural infection. However, the human host unavoidably complicates interpretation of these traits, and their assessment can often only be obtained once. A thorough phenotypic assessment would therefore necessitate large sample numbers and thoughtful study design to account for variation both within the human and

within the parasite populations. Thus far, phenotyping has mainly been carried out on culture-adapted[52] parasites that have been isolated from patients. These can then be tested for various *in vitro* phenotypes, including drug response, invasion types, cytoadherence properties, the ability to produce gametocytes and their metabolic profiles.

Below, we elaborate on the use of genomics to identify mutations that are associated with altered drug responses, but these approaches (for example, GWASs) could be extended to other biologically important phenotypes.

Linkage analysis in P. falciparum. Linkage mapping in *P. falciparum* has been accomplished using laboratory genetic crosses to correlate segregation patterns in the progeny that is associated with specific phenotypes, including drug response[53–58], pathogenesis[59] or mosquito infectivity[60]. Originally, MSVs were used as genomic markers[61], but these have now been augmented with variants determined using whole-genome methods (whether array- or sequencing-based). Linkage studies leverage the reasonably high level of recombination in *P. falciparum* to map the genetic determinants for specific traits: one round of recombination between parents and progeny results in large haplotype blocks that require fewer markers to identify than for population-based association studies[8]. In some geographic regions, where low diversity means that recombination rarely results in the reassortment of haplotype blocks, it may even be possible to carry out similar analyses using field isolates[62,63]. Combining linkage analysis with other independent tests, such as association mapping, provides a potentially powerful means of prioritizing candidate genes that are responsible for a given phenotype.

Challenges of GWASs in P. falciparum. The molecular and genetic mechanisms of many phenotypic traits that are most relevant to elimination and eradication, such as variability in parasite responsiveness to drugs, are poorly understood. Because GWASs do not require prior knowledge of gene functions or trait mechanisms, they are useful for identifying important genetic variants in organisms such as *P. falciparum* that have many genetic loci with no known functional homologues. Although these candidate variants require functional validation, the use of GWASs as hypothesis-generating experiments provides a powerful starting point for identifying traits and is one of the most effective approaches available in our modern genomic toolkit.

Undertaking GWASs in *P. falciparum* requires overcoming various challenges, including identifying heritable traits, coping with low LD, using appropriate sample collection or other methods to deal with population stratification and functionally validating associations. These challenges are described below, followed by examples of successful GWASs.

When surveying the *P. falciparum* genome for genotype–phenotype associations, only phenotypes with a strong genetic basis (those with high heritability) will be detected by a GWAS. Heritability of *P. falciparum* traits, such as

---

Directional selection
Selection that favours one allele over all other alleles of a gene. The frequency of this beneficial allele can rise or can be held in check by recurrent mutation.

Selective sweep
Describes the reduced levels of genetic diversity observed around a selected locus. A selective sweep (also referred to as genetic hitch-hiking) arises because positive selection for an advantageous allele increases the frequency of not only that allele but also other alleles contained within the same haplotypes.

Long-range haplotype test
(LRH test). A test for haplotypes with significantly longer than expected ranges of linkage disequilibrium; this indicates a lack of recombination between genetic markers caused by natural selection.

Heritability
The proportion of the total phenotypic variation in a trait that can be attributed to genetic effects.

drug resistance, can be variable: recent studies of parasite clearance rates in South East Asia found that the heritability of this phenotype depends on when and where samples were collected[64,65]. Confounding this complication, anti-malarial responses can be quantified in various ways, including *in vitro* based metrics, such as IC_{50}, or clinically derived metrics, such as *in vivo* parasite clearance rates.

The short blocks of LD in *P. falciparum*, particularly in African populations[19,25], are an important consideration for study design. Traditional GWASs rely on the genotyped markers being correlated to causal mutations through high LD[66]. In a population with low LD, an array-based GWAS may not have sufficient detection power unless the causal mutation is present on the array. However, when a signal is found, short LD makes localizing the signal to a single gene much easier[67]. Loss of detection power owing to limited LD can therefore be circumvented by using whole-genome sequencing to identify all variants in the genome. Use of sequence data for GWASs reveals stronger association signals, provides more supporting markers in areas of high LD and can detect candidate loci in areas of low LD that were previously missed by array-based GWASs. Sequencing-based approaches are thus a promising avenue for future GWASs in malaria.

Population demography — particularly in the form of population stratification — can hinder GWAS analyses if it is not appropriately controlled for. The presence of closely related individuals in the data set or, conversely, broad genetic differences between groups of samples owing to differing population histories can erroneously inflate associations and can produce false positives[68]. The ideal GWAS would eliminate such confounders by studying a phenotype that is heritable, measurable and strongly apparent in a subset of parasites (such as drug resistance) while entirely sampling from a single, non-stratified population. However, this is not always possible, and many approaches have been developed to eliminate false positives[69–71] while analysing stratified data sets. In particular, mixed-model approaches[72–75] have successfully been used to control for population stratification in malaria studies[25].

For studies with small sample sizes, which include all malaria-based GWASs to date, gains in study power can be achieved by using multi-marker or haplotype-based association approaches[25,66] instead of standard single-marker tests. Positively selected variants typically lie on long haplotypes[36] that are more easily detectable by multi-marker tests, such as the LRH test.

GWAS approaches towards identifying drug resistance loci.
Although the GWAS approach is promising, we are still in the early days of applying this methodology for loci discovery in the malaria parasite, and the few GWASs that have been carried out have primarily investigated novel variants that are associated with drug resistance. These studies[24,25,43] generate long lists of loci that are hypothesized to be associated with specific phenotypes, and a current challenge is winnowing the list to the most likely candidates. One strategy involves combining results of independent tests to identify the most likely candidate genes for a functional follow-up (FIG. 2).

Providing validation for GWAS-based approaches, several genes that are already known to be associated with drug resistance have been identified, such as *pfcrt*, *pfmdr1* and *dhfr-ts*; additionally, one novel candidate has been identified and functionally validated. The GWAS by Van Tyne *et al.*[25] identified a highly polymorphic locus, *PF10_0355*, as being associated with halofantrine resistance based on a small set of globally diverse parasites. *PF10_0355* was classified as a member of the *msp3* gene family[76]. When a variant of *PF10_0355* from a drug-resistant parasite was introduced into a drug-sensitive parasite through transfection, the parasite was rendered resistant not only to halofantrine but also to the chemically similar drugs mefloquine and lumefantrine, but not to the chemically distinct compounds chloroquine, artemisinin and atovaquone. This is the first functional demonstration that a potential drug resistance locus identified by a GWAS confers a drug resistance phenotype. This study used a modest number of parasites that had been sampled from many populations using a limited marker set; current studies now assay larger parasite numbers from single populations using sequencing data to capture essentially all genetic variation.

In addition to profiling sequence variants, array-based methods identified a novel *P. falciparum* CNV corresponding to the first gene in the folate pathway, GTP cyclohydrolase (*gch1*)[77], which was later validated[42] as being associated with antifolate use in *P. falciparum* in Thailand.

We anticipate increased power to detect mutations that are associated with, or responsible for, key clinical phenotypes using sequencing-based GWASs, which are now being applied to several phenotypes.

Laboratory-based selection of drug resistance.
A second application of genomics, which is useful for understanding drug resistance and for promoting drug development, involves the study of artificially selected drug-resistant parasites and the use of sequencing to find the causal mutations conferring resistance. This approach allows the identification of genetic variants that are capable of conferring drug resistance, which in many cases have been found in the drug target. Typically, it involves placing a clonal parasite line under drug selection *in vitro* to create mutant forms of the parasite that are resistant to the compound; then, whole-genome analysis (either array- or sequencing-based) is performed to identify the causal mutation that is associated with drug resistance. This approach has successfully been implemented for a number of anti-malarial agents[16,78–82], demonstrating its use for identifying drug targets.

Selection studies using different drugs have revealed intriguing distinctions in the drug sensitivity profiles between different mutant strains, and there is a potential relevance to therapeutic strategies for malaria control. For example, mutations in *pfcrt* conferring chloroquine resistance were also found to confer sensitivity to some

---

IC_{50}
Also known as the half maximal inhibitory concentration, this is the concentration of a small molecule that results in 50% inhibition of a molecular target or cellular process.

Mixed-model approaches
A class of linear model in which quantitative phenotypes are modelled using both fixed, genetic effects (including SNP genotypes and the principal components of population structure) and random effects (which explicitly models both the heritable and non-heritable components of phenotypic variation).
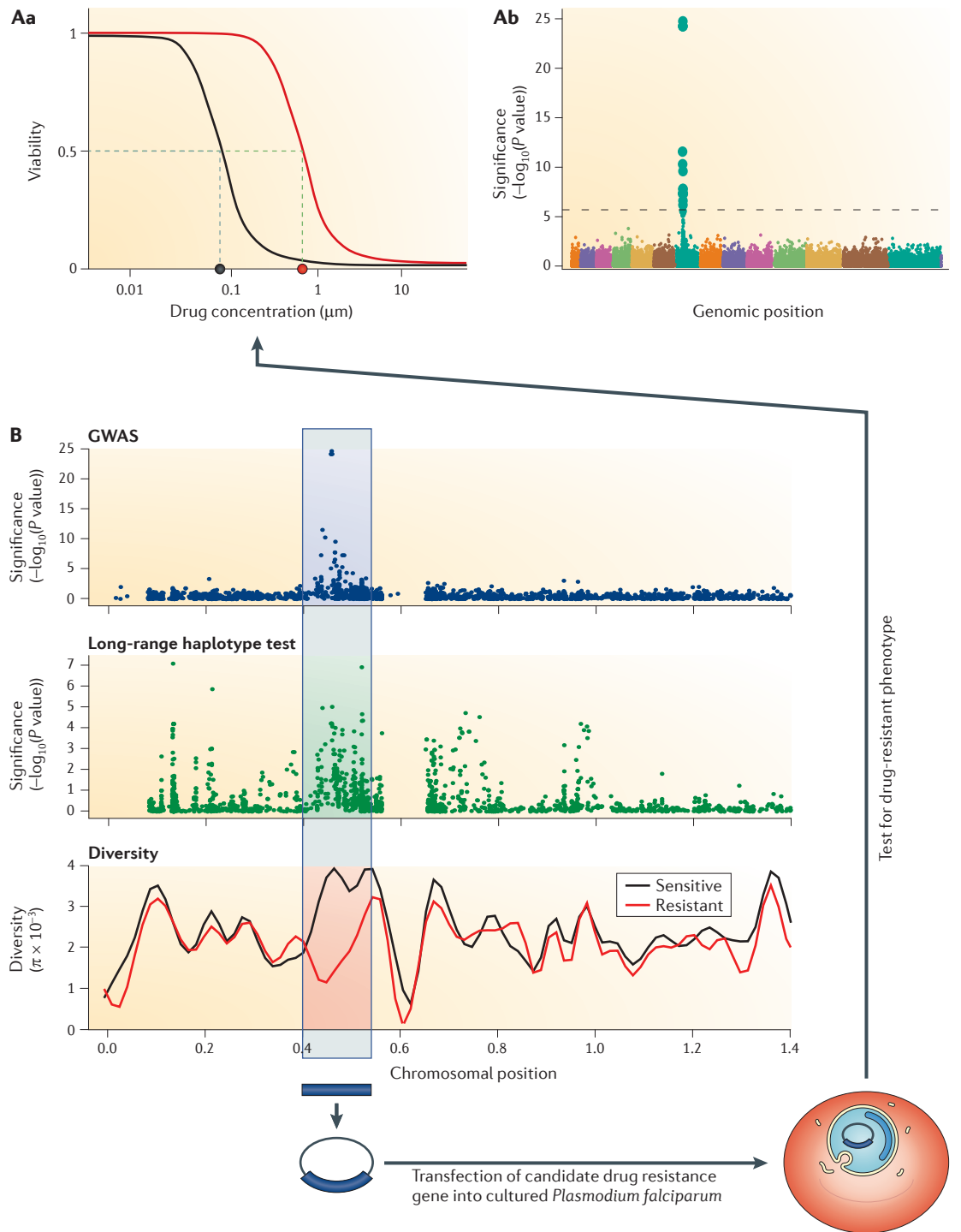
Figure 2 | **Identifying drug resistance loci using GWASs and functional studies. Aa** | Schematic representing data from a drug-treatment parasite viability assay, which allows the determination of the IC_{50}; this represents the drug concentration at which parasite viability is decreased by half. Such assays can be used to classify parasite responses to anti-malarial compounds, thus distinguishing a drug-resistant parasite (shown by the red line) from a drug-sensitive parasite (shown by the black line). **Ab** | Genome-wide association study (GWAS) analysis results shown as a Manhattan plot in which P values for variants across the 14 chromosomes (represented by different colours across the x axis) are shown. The Bonferroni level for genome-wide significance is shown as a dotted line, and genetic variants that rise above this level are associated with the drug phenotype that was observed in panel **Aa**. **B** | Multiple independent analysis approaches can be combined to improve the power of locus identification for functional follow-up studies. For example, alleles identified by GWASs, long-range haplotype (LRH) tests or diversity and divergence analyses are identified, and a gene expression construct is created comprising the genetic variant that is associated with drug resistance. To test whether each gene variant is necessary and sufficient to confer the observed drug-resistant phenotype, each putative drug resistance variant is introduced into a drug-sensitive parasite by transfection followed by testing for drug resistance.

novel compounds[43]. As a consequence, these novel compounds were effective against chloroquine-resistant but not chloroquine-sensitive parasites. Furthermore, continued exposure to these novel compounds selected for reversion of the *pfcrt* mutation, thus resensitizing parasites to chloroquine[43]. Such observations give hope to the idea of repurposing chloroquine, which is an excellent anti-malarial drug that is only compromised by drug resistance. It could be speculated that giving compounds with these complementary effects in combination might prevent the emergence of drug-resistant parasites.

Many drug selection studies focus on compounds that are not yet in clinical use, so we do not currently expect to see these mutations in wild populations. Nevertheless, these studies suggest that the application of anti-malarial pressure in nature will select for drug-resistant forms of the parasite and that the use of drugs in combination is a key strategy for protecting against the selection of drug-resistant mutants. Although there may be problems with creating evolved strains, and laboratory culture may not mimic what happens in humans, *in vitro* studies have the advantage of selecting only for a few mutations; this leads to low ambiguity and, providing that the results are replicated in independent clones, there can be high statistical confidence in the involvement of a specific allele. This level of statistical power would be more difficult to achieve with GWASs.

Emerging strategies for assessing interventions

Genetic and genomic information can identify novel mechanisms of drug resistance and can assist with new drug development and target identification. We propose that understanding parasite population structure and LD would allow a novel (and as of yet untested) means of monitoring parasite dynamics related to transmission, assessing interventions such as drugs or vaccines and identifying sources of new infection. Such assessments, which are carried out through easily deployed genomic tools, are key requirements of the current malaria eradication campaign[83].

As discussed above, decreases in transmission levels reduce outcrossing and ultimately shift allele frequencies and increase LD. As intervention strategies (such as insecticide spraying, bed net use or vaccination) are deployed, transmission intensity is expected to decline, thus reducing COI. Furthermore, as we approach parasite elimination, we anticipate observation of clonal populations of parasites in a given transmission area. Ultimately, as individual parasite types have their own 'fingerprint' or genomic signature, we should be able to track these parasite types using molecular barcodes as interventions are applied.

Determining and applying COI. Classic means of evaluating transmission intensity are based on indicators such as the entomologic inoculation rate (EIR), seropositivity for malaria antigens or gametocyte carriage rates; however, these measures can be difficult to obtain, are indirect measures of the parasite types within individuals and provide mainly population-level information

rather than specific information about the parasites that survive or escape an applied intervention. By contrast, evaluation of patient COI (FIG. 3a) can assess transmission directly in humans. Empirical observations of LD in parasites from regions that exhibit differing levels of COI support this prediction: there is much lower LD in high-transmission African sites than in moderate-to-low-transmission sites in South East Asia or low-transmission sites in the Americas[13,14,26].

FIGURE 3b shows the expected relationship between parasite LD and mean COI in a population. This indicates that LD might be useful for determining the approximate historical COI and therefore the relative rate of transmission in populations that undergo low-to-moderate levels of infection. However, population-level LD is not sensitive to change in locations where infection rates are high; for example, a further increase in transmission in a high-transmission area where LD is already very low will have a negligible effect, as nearly all infections already permit outcrossing. Population-level LD also does not respond rapidly to change: a dramatic increase in transmission in a high-LD region might be detectable within a year or two, but more modest increases and decreases would take substantially longer. Thus, rather than measuring LD, determining the actual distribution of COI among individuals in the population might be a better means of inferring changes in transmission rates over short periods of time.

Historically, COI has been determined by assaying highly polymorphic loci[84–86] (FIG. 3c). Several groups are now using PCR-directed next-generation sequencing of highly polymorphic loci to estimate COI. If a locus with suitably high levels of polymorphism is chosen, such that the probability that two strains will share the same haplotype within an individual infection is small, tallying the number of distinct haplotypes can sensitively — and, to some extent, quantitatively — be used to measure COI. Broadly deploying this kind of deep sequencing of highly polymorphic loci could provide a practical approach towards monitoring COI as interventions are applied (FIG. 3d).

Molecular barcoding to identify parasite types. Measuring COI is anticipated to be useful for distinguishing populations that undergo moderate and high levels of transmission (FIG. 3); there is less use for detecting populations that undergo very low levels of transmission where infected individuals would typically share a COI of 1 (hence measuring COI would provide no additional information). In these cases, alternative population genomic tools that can identify particular parasite types rather than just the number of distinct parasites have potential use. An example is the molecular barcode[87], which is a tool that evaluates 24 independent, highly variable SNPs to identify a parasite type uniquely within a population of parasites (FIG. 4a). Under moderate- or high-transmission conditions, almost every parasite in a population will exhibit a distinct molecular barcode signature. Under extremely low transmission conditions, such as those that might result from a disease elimination campaign,

Entomologic inoculation rate (EIR). The number of infectious mosquito bites a person is exposed to in a certain time period, typically a year.
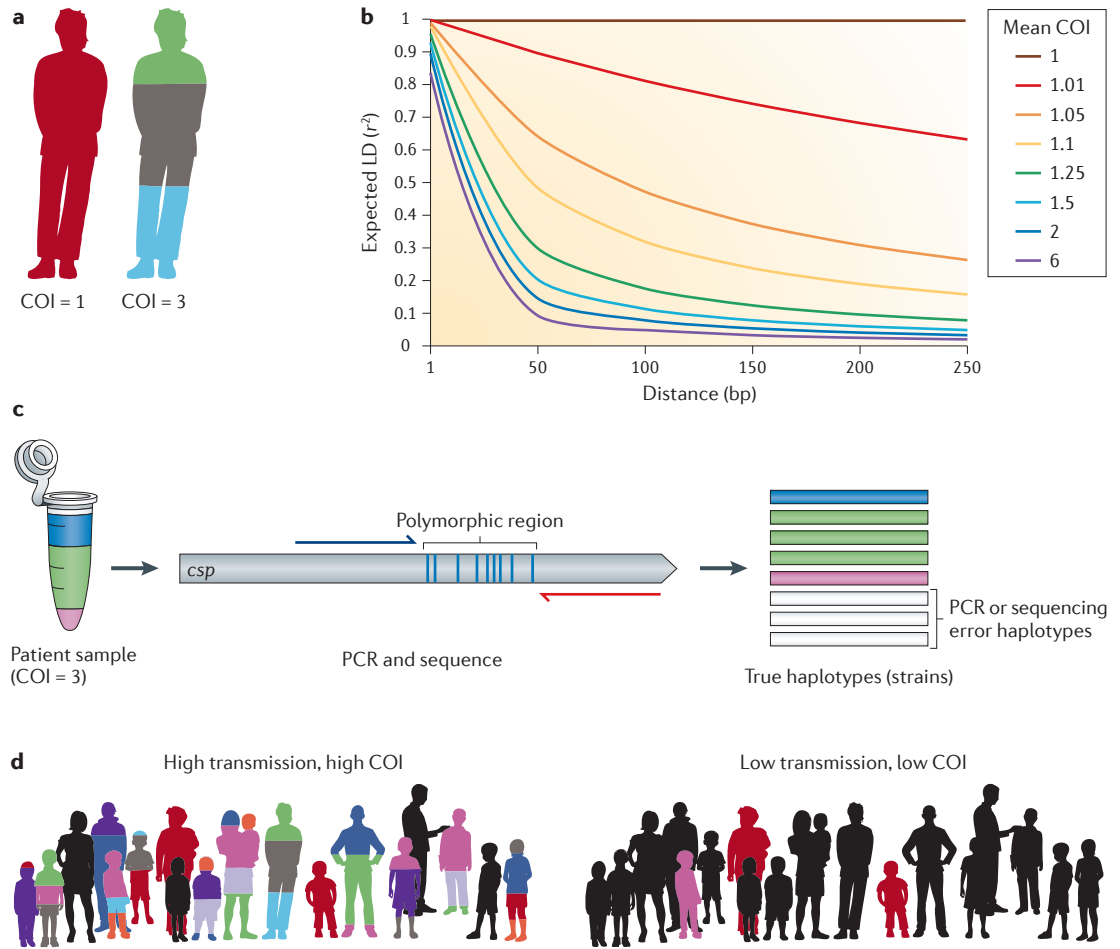
Figure 3 | ***Plasmodium falciparum* complexity of infection. a** | Complexity of infection (COI) represents the number of different parasite types in an infected individual. A person infected with one parasite type is represented by a homogeneous colour, whereas a person infected with multiple parasite types is represented with multiple colours. **b** | When the mean COI is exactly 1, meaning that a single parasite type exclusively determines infection, LD is perfect in the parasite genome, irrespective of chromosomal distance (shown by the brown line). However, when the mean COI is barely above 1 (1.01 would be equivalent to 99% single infections and 1% double infections among infected individuals in a population), expected LD begins to drop precipitously as a function of distance to the extent that mean COI levels above 2 are incapable of producing significant further reduction in the statistic. **c** | Next-generation sequencing approaches can be used to determine COI. In the example shown, DNA is derived from a patient sample with a COI of 3 and subjected to PCR amplification across a highly polymorphic locus, such as the circumsporozoite (*csp*) gene. PCR primers can be engineered with unique sequence tags to identify each individual patient sample. The PCR products from multiple samples are pooled and then sequenced to high coverage. Bioinformatic tools are used to filter out systematic errors (to distinguish true haplotypes from PCR or sequencing error haplotypes) and to deconvolute the pooled sequence reads based on the sample-specific sequence tags. The number and type of different haplotypes for each sample are tallied to quantify COI values for each individual across the population. **d** | Expectations under high transmission intensity would be that individuals within that population would be infected with multiple parasite types (with a high COI). By contrast, under low transmission intensity, COI would be expected to be low, representative of infected individuals harbouring only a single parasite type (shown as a homogenous colour; black represents uninfected individuals).

Reinfection
A subsequent infection by a different strain of parasite compared with the original infection.

Recrudescence
A persistent erythrocytic infection that results in the original strain of parasite re-emerging following anti-malarial treatment.

transmission of individual strains will be stochastic, resulting in the random extinction of some lineages and the amplification of others, which can then be observed repeatedly. Parasite clonality, where LD has reached a value of 1, is a direct theoretical consequence of small population size. The molecular barcode tool can follow individual parasite types and can identify micro-epidemiological factors. Under the right conditions, this type of tool could be deployed to distinguish reinfection or recrudescence to see whether specific parasite types are escaping vaccine efficacy and to follow parasite transmission patterns both across and within transmission seasons (FIG. 4b,c). Leveraging parasite genetic diversity to fingerprint parasites and to follow them in a population can give insight into parasite population dynamics that will be a key advancement towards elimination.
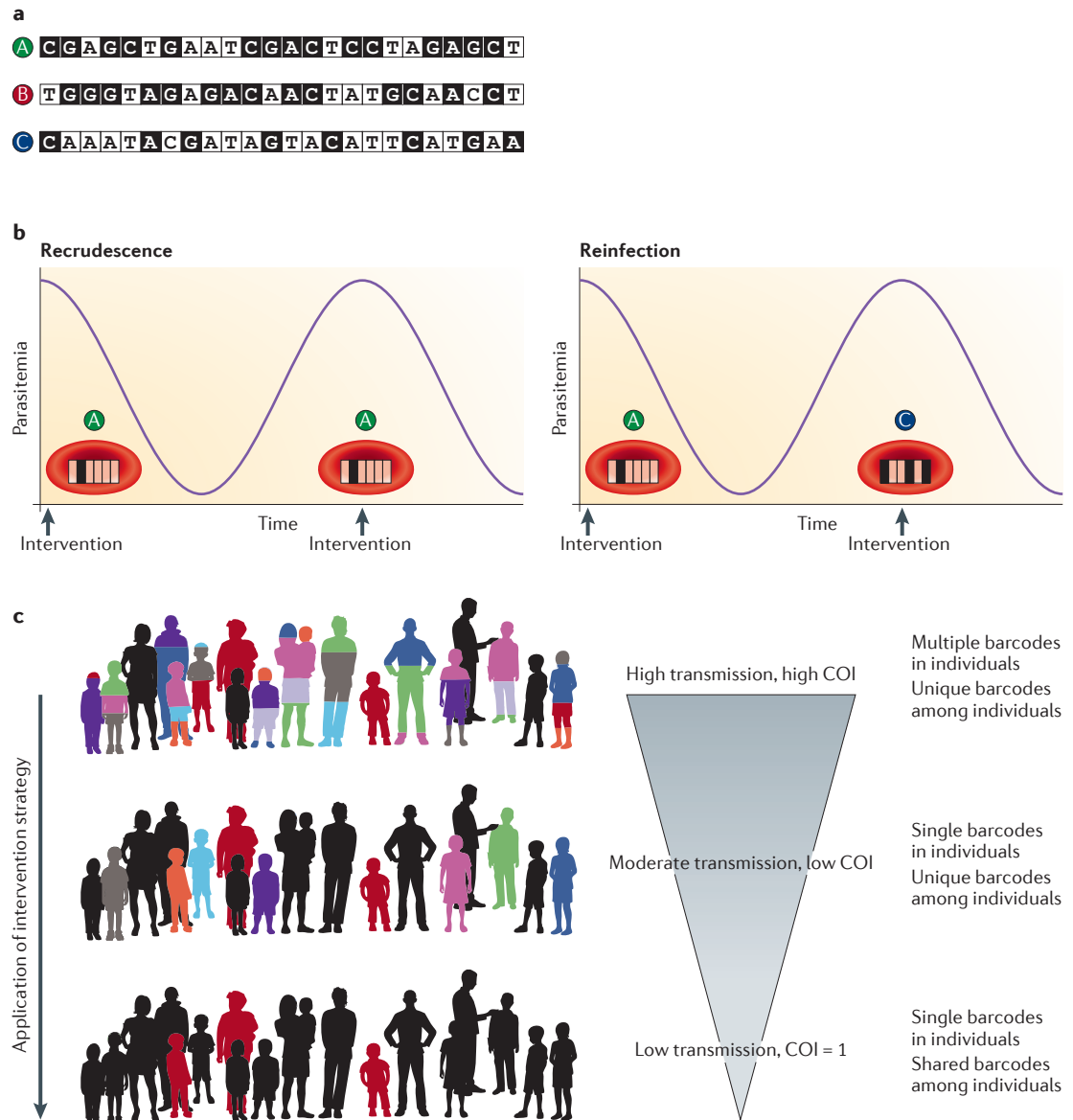
a

(A) `CGAGCTGAATCGACTCCTAGAGCT`

(B) `TGGGTAGAGACAACTATGCAACCT`

(C) `CAAATACGATAGTACATTCATGAA`

b

Recrudescence

Parasitemia

(A) (A)

Time

Intervention   Intervention

Reinfection

Parasitemia

(A) (C)

Time

Intervention   Intervention

c

Application of intervention strategy

High transmission, high COI

Multiple barcodes in individuals
Unique barcodes among individuals

Moderate transmission, low COI

Single barcodes in individuals
Unique barcodes among individuals

Low transmission, COI = 1

Single barcodes in individuals
Shared barcodes among individuals

Figure 4 | **Molecular barcoding to identify parasite types. a** | The molecular barcode comprises 24 unlinked SNPs that are present in a global parasite population at a high minor allele frequency[87]. The molecular barcode can distinguish one parasite type (A, B or C) from another. **b** | This schematic indicates the molecular barcode for the parasite causing the peak in parasitemia (A or C). The molecular barcode tool can be applied in drug or vaccine trials to determine whether the parasite that is present before the intervention (parasite A) is the cause of subsequent infection (recrudescence) or whether a distinct parasite type (parasite C) is the cause of that subsequent infection (reinfection). **c** | The molecular barcode can be applied to assess changing transmission levels during therapeutic interventions and is particularly useful when the mean complexity of infection (COI) among infected individuals approaches a value of 1. In populations with a high transmission and high COI levels, both multiple parasite genomes in each sample and highly variable combinations of barcodes among individuals in the population would be expected to be seen. Under moderate transmission and low COI levels, although single barcodes would be found in infected individuals (COI = 1), these are likely to differ among individuals in the population. Finally, as the disease nears eradication, not only will individuals harbour parasites with unique barcodes, but these barcodes will be shared among the few infected individuals, representing clonal parasite populations.

Application of genomic tools to evaluate vaccines.
Genomic tools for estimating COI and for identifying particular parasite infections by molecular barcode can now be applied to field-based investigations, including vaccine studies. Monovalent vaccines that are designed to enhance immunity to highly polymorphic targets, such as the *csp* locus, may not completely confer cross-protective immunity to the range of segregating *csp* variants in the parasite population. Deployment of genomic tools in clinical trials will enable investigators to detect this phenomenon and

anticipate the evolution of vaccine resistance in the parasite population[85]. Genomic approaches that are applied to the ongoing RTS,S (also known as Mosquirix) vaccine Phase III trial[88], in which approximately half of all vaccinated individuals are protected from disease, might help us to understand why only that subset was protected. For example, the vaccine may work to reduce the burden of disease by decreasing the number of distinct parasite types in the host. Alternatively, the vaccine might be successfully eliminating all variants that are represented in the vaccine, and individuals who are not protected may harbour parasites with novel variants that are not represented in the original vaccine. Although molecular analysis of RTS,S Phase I and Phase IIb samples did not suggest selection of non-vaccine genotypes[89–92], these studies only evaluated individual polymorphisms rather than haplotypic differences. The deep coverage that is afforded by next-generation sequencing, when applied with appropriate analysis methods, allows us to identify the specific variants that are present in individuals within a large patient population with confidence, enabling us to distinguish between these two cases. This kind of approach will enable researchers not only to monitor the success of interventions but also to evaluate the causes of success and failure.

Population history suggests that natural immunity induces balancing selection but that monovalent vaccines may induce directional selection. Similar to the situation for drug resistance, where parasites escape drug pressure by becoming resistant through changes at a defined locus, it might be anticipated that parasites escaping monovalent vaccine pressure have forms of the target allele that do not elicit specific immune responses through vaccination, resulting in parasite populations that are refractory to the given vaccine. Thus, vaccine pressure may enrich for parasite forms containing intractable versions of the vaccine target, as they are selected to high prevalence in the population. It might be imagined that careful monitoring using genetic approaches — as is currently done for influenza vaccine strategies — could identify parasites that escape this immune response and inform the development of a modified vaccine to include these specific parasite subtypes.

Conclusions and future directions

Genomic analysis is providing rich and unique insight into the biology and life history of *P. falciparum*. There are several natural selective forces that leave their imprint on the parasite genome: the human immune response is implicated in the diversifying selection that is evident in over 500 genes encoding membrane and membrane-associated proteins, and there is strong support for the directional selection of anti-malarial drugs. Perhaps what is most surprising is the evidence for directional selection of many genes encoding metabolic, cell-signalling and protein-processing and protein-degradation pathways. The underlying selective mechanisms are not known, but this work provides the basis for developing hypotheses that can be tested with functional analysis. For example, the positive selection that has recently been seen for genes in the ubiquitylation pathway leads us to propose that regulation of protein degradation might be an important fitness component of parasite survival. The goal of contemporary disease elimination efforts should be to prevent selective sweeps from occurring in response to novel compounds, even if resistance mutations arise, through careful genomic surveillance and flexible, combinatorial application of drug treatments.

In addition to uncovering new biology, we hypothesize that genetic measurements of *P. falciparum* population structure in infected human populations can also be used to infer transmission levels. This idea will need to be tested using field-based samples, but it opens the possibility of directly measuring changes in transmission, particularly during elimination and eradication efforts, from patient-derived samples, thereby providing outcome-based data in real time. In the decade since the publication of the *P. falciparum* genome sequence[3], malaria genomics and genome biology have advanced our understanding of malaria biology. However, we are just beginning to leverage population genetic information from this organism to discover novel variants that are responsible for crucial clinical phenotypes and to develop tools for assessment of interventions as we work towards elimination of this important human pathogen.

1. WHO. in World Malaria Report 2011 (World Health Organization, Switzerland, 2011).
2. Trape, J. F. *et al.* Malaria morbidity and pyrethroid resistance after the introduction of insecticide-treated bednets and artemisinin-based combination therapies: a longitudinal study. *Lancet Infect. Dis.* **11**, 925–932 (2011).
3. Gardner, M. J. *et al.* Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* **419**, 498–511 (2002).
4. Carlton, J. M. *et al.* Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*. *Nature* **455**, 757–763 (2008).
5. Carlton, J. M. *et al.* Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii*. *Nature* **419**, 512–519 (2002).
6. Hall, N. *et al.* A comprehensive survey of the *Plasmodium* life cycle by genomic, transcriptomic, and proteomic analyses. *Science* **307**, 82–86 (2005).
7. Dharia, N. V. *et al.* Whole-genome sequencing and microarray analysis of *ex vivo Plasmodium vivax* reveal selective pressure on putative drug resistance genes. *Proc. Natl Acad. Sci. USA* **107**, 20045–20050 (2010).
8. Anderson, T., Nkhoma, S., Ecker, A. & Fidock, D. How can we identify parasite genes that underlie antimalarial drug resistance? *Pharmacogenomics* **12**, 59–85 (2011).
   This is a comprehensive review article that outlines genome-scale approaches for identifying genetic variants that contribute to drug resistance.
9. Escalante, A. A., Cornejo, O. E., Rojas, A., Udhayakumar, V. & Lal, A. A. Assessing the effect of natural selection in malaria parasites. *Trends Parasitol.* **20**, 388–395 (2004).
10. Weedall, G. D. & Conway, D. J. Detecting signatures of balancing selection to identify targets of anti-parasite immunity. *Trends Parasitol.* **26**, 363–369 (2010).
    This review describes population genetic approaches for identifying signatures of balancing selection and the use of new genomic information to discover new potential vaccine candidates.
11. Ghedin, E. *et al.* Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution. *Nature* **437**, 1162–1166 (2005).
12. Jeffares, D. C. *et al.* Genome variation and evolution of the malaria parasite *Plasmodium falciparum*. *Nature Genet.* **39**, 120–125 (2007).
13. Mu, J. *et al.* Genome-wide variation and identification of vaccine targets in the *Plasmodium falciparum* genome. *Nature Genet.* **39**, 126–130 (2007).
14. Volkman, S. K. *et al.* A genome-wide map of diversity in *Plasmodium falciparum*. *Nature Genet.* **39**, 113–119 (2007).
15. Cheeseman, I. H. *et al.* Gene copy number variation throughout the *Plasmodium falciparum* genome. *BMC Genomics* **10**, 353 (2009).
16. Dharia, N. V. *et al.* Use of high-density tiling microarrays to identify mutations globally and elucidate mechanisms of drug resistance in *Plasmodium falciparum*. *Genome Biol.* **10**, R21 (2009).
17. Jiang, H. *et al.* Detection of genome-wide polymorphisms in the AT-rich *Plasmodium falciparum* genome using a high-density microarray. *BMC Genomics* **9**, 398 (2008).
18. Kidgell, C. *et al.* A systematic map of genetic variation in *Plasmodium falciparum*. *PLoS Pathog.* **2**, e57 (2006).
19. Neafsey, D. E. *et al.* Genome-wide SNP genotyping highlights the role of natural selection in *Plasmodium falciparum* population divergence. *Genome Res.* **9**, R171 (2008).

20. Tan, J. C. *et al.* Optimizing comparative genomic hybridization probes for genotyping and SNP detection in *Plasmodium falciparum*. *Genomics* **93**, 543–550 (2009).

21. Westenberger, S. J. *et al.* A systems-based analysis of *Plasmodium vivax* lifecycle transcription from human to mosquito. *PLoS Negl. Trop. Dis.* **4**, e653 (2010).

22. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).

23. Liu, W. *et al.* Origin of the human malaria parasite *Plasmodium falciparum* in gorillas. *Nature* **467**, 420–425 (2010).

24. Mu, J. *et al.* Plasmodium falciparum genome-wide scans for positive selection, recombination hot spots and resistance to antimalarial drugs. *Nature Genet.* **42**, 268–271 (2010).

25. Van Tyne, D. *et al.* Identification and functional validation of the novel antimalarial resistance locus *PF10_0355* in *Plasmodium falciparum*. *PLoS Genet.* **7**, e1001383 (2011).
 This was the first demonstration of a GWAS in *P. falciparum* that included functional follow-up of a novel locus identified using this approach.

26. Mu, J. *et al.* Recombination hotspots and population structure in *Plasmodium falciparum*. *PLoS Biol.* **3**, e335 (2005).

27. Anderson, T. J. *et al.* Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. *Mol. Biol. Evol.* **17**, 1467–1482 (2000).

28. Pumpaibool, T. *et al.* Genetic diversity and population structure of *Plasmodium falciparum* in Thailand, a low transmission country. *Malaria J.* **8**, 155 (2009).

29. Conway, D. J. *et al.* Extreme geographical fixation of variation in the *Plasmodium falciparum* gamete surface protein gene *Pfs48/45* compared with microsatellite loci. *Mol. Biochem. Parasitol.* **115**, 145–156 (2001).

30. Nkhoma, S. C. *et al.* Close kinship within multiple-genotype malaria parasite infections. *Proc. R. Soc. B* 7 Mar 2012 (doi:10.1098/rspb.2012.0113).

31. Takala, S. L. *et al.* Dynamics of polymorphism in a malaria vaccine antigen at a vaccine-testing site in Mali. *PLoS Med.* **4**, e93 (2007).

32. Thera, M. A. *et al.* A field trial to assess a blood-stage malaria vaccine. *N. Engl. J. Med.* **365**, 1004–1013 (2011).

33. Smith, J. M. & Haigh, J. The hitch-hiking effect of a favourable gene. *Genet. Res.* **23**, 23–35 (1974).

34. Conway, D. J. *et al.* High recombination rate in natural populations of *Plasmodium falciparum*. *Proc. Natl Acad. Sci. USA* **96**, 4506–4511 (1999).

35. Jiang, H. *et al.* High recombination rates and hotspots in a *Plasmodium falciparum* genetic cross. *Genome Biol.* **12**, R33 (2011).

36. Sabeti, P. C. *et al.* Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419**, 832–837 (2002).

37. Laufer, M. K. *et al.* Return of chloroquine-susceptible falciparum malaria in Malawi was a reexpansion of diverse susceptible parasites. *J. Infect. Dis.* **202**, 801–808 (2010).

38. Wootton, J. C. *et al.* Genetic diversity and chloroquine selective sweeps in *Plasmodium falciparum*. *Nature* **418**, 320–323 (2002).

39. Nair, S. *et al.* A selective sweep driven by pyrimethamine treatment in southeast asian malaria parasites. *Mol. Biol. Evol.* **20**, 1526–1536 (2003).

40. Foote, S. J., Thompson, J. K., Cowman, A. F. & Kemp, D. J. Amplification of the multidrug resistance gene in some chloroquine-resistant isolates of *P. falciparum*. *Cell* **57**, 921–930 (1989).

41. Wilson, C. M. *et al.* Amplification of a gene related to mammalian *mdr* genes in drug-resistant *Plasmodium falciparum*. *Science* **244**, 1184–1186 (1989).

42. Nair, S. *et al.* Adaptive copy number evolution in malaria parasites. *PLoS Genet.* **4**, e1000243 (2008).

43. Yuan, J. *et al.* Chemical genomic profiling for antimalarial therapies, response signatures, and molecular targets. *Science* **333**, 724–729 (2011).
 This study used a chemical screen and GWASs to identify a number of antimalarial compounds and confirms the involvement of the identified SNPs in response to drugs using both linkage analysis and genetically modified parasites. Demonstrates that most of these activities are associated with only three previously identified genes and that combinations of drugs targeting the *pfcrt* locus may be clinically useful.

44. Ponts, N. *et al.* Deciphering the ubiquitin-mediated pathway in apicomplexan parasites: a potential strategy to interfere with parasite virulence. *PLoS ONE* **3**, e2386 (2008).

45. Nguitragool, W. *et al.* Malaria parasite *clag3* genes determine channel-mediated nutrient uptake by infected red blood cells. *Cell* **145**, 665–677 (2011).
 This paper shows that the plasmodial surface anion channel (PSAC) is an anti-malarial target of compounds inhibiting nutrient uptake. The authors localize this response to the *clag3* locus, which is previously assumed to function in cytoadherence. This work demonstrates that surface molecules can alter cell permeability to diverse solutes, including drugs.

46. Su, X. Z., Mu, J. & Joy, D. A. The 'Malaria's Eve' hypothesis and the debate concerning the origin of the human malaria parasite *Plasmodium falciparum*. *Microbes Infect.* **5**, 891–896 (2003).

47. Sabeti, P. C. *et al.* Positive natural selection in the human lineage. *Science* **312**, 1614–1620 (2006).
 This is a review of the different genomic patterns left by positive natural selection and the many methods and tests used to detect them.

48. Hartl, D. L. The origin of malaria: mixed messages from genetic diversity. *Nature Rev. Microbiol.* **2**, 15–22 (2004).

49. Hartl, D. L. *et al.* The paradoxical population genetics of *Plasmodium falciparum*. *Trends Parasitol.* **18**, 266–272 (2002).

50. Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H. & Bustamante, C. D. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* **5**, e1000695 (2009).

51. Li, J. *et al.* Joint analysis of demography and selection in population genetics: where do we stand and where could we go? *Mol. Ecol.* **21**, 28–44 (2012).

52. Trager, W. & Jensen, J. B. Human malaria parasites in continuous culture. *Science* **193**, 673–675 (1976).

53. Ferdig, M. T. *et al.* Dissecting the loci of low-level quinine resistance in malaria parasites. *Mol. Microbiol.* **52**, 985–997 (2004).
 This paper describes the use of quantitative trait analysis to identify the *pfnhe-1* locus, which is related to quinine resistance. It also demonstrates the power of linkage studies to locate genes contributing to drug resistance.

54. Hayton, K. & Su, X. Z. Drug resistance and genetic mapping in *Plasmodium falciparum*. *Curr. Genet.* **54**, 223–239 (2008).

55. Nkrumah, L. J. *et al.* Probing the multifactorial basis of *Plasmodium falciparum* quinine resistance: evidence for a strain-specific contribution of the sodium–proton exchanger PfNHE. *Mol. Biochem. Parasitol.* **165**, 122–131 (2009).

56. Sa, J. M. *et al.* Geographic patterns of *Plasmodium falciparum* drug resistance distinguished by differential responses to amodiaquine and chloroquine. *Proc. Natl Acad. Sci. USA* **106**, 18883–18889 (2009).

57. Sanchez, C. P., Mayer, S., Nurhasanah, A., Stein, W. D. & Lanzer, M. Genetic linkage analyses redefine the roles of PfCRT and PfMDR1 in drug accumulation and susceptibility in *Plasmodium falciparum*. *Mol. Microbiol.* **82**, 865–878 (2011).

58. Wellems, T. E., Walker-Jonah, A. & Panton, L. J. Genetic mapping of the chloroquine-resistance locus on *Plasmodium falciparum* chromosome 7. *Proc. Natl Acad. Sci. USA* **88**, 3382–3386 (1991).

59. Hayton, K. *et al.* Erythrocyte binding protein PfRH5 polymorphisms determine species-specific pathways of *Plasmodium falciparum* invasion. *Cell Host Microbe* **4**, 40–51 (2008).

60. Vaidya, A. B. *et al.* A genetic locus on *Plasmodium falciparum* chromosome 12 linked to a defect in mosquito-infectivity and male gametogenesis. *Mol. Biochem. Parasitol.* **69**, 65–71 (1995).

61. Su, X. *et al.* A genetic map and recombination parameters of the human malaria parasite *Plasmodium falciparum*. *Science* **286**, 1351–1353 (1999).

62. Dharia, N. V. *et al.* Genome scanning of Amazonian *Plasmodium falciparum* shows subtelomeric instability and clindamycin-resistant parasites. *Genome Res.* **20**, 1534–1544 (2010).
 This study used genome scanning to reveal an almost clonal population structure for *P. falciparum* parasites from Peru, which suggests that these parasites could be useful in linkage or quantitative trait analyses. It also identifies mutations in the apicoplast that confer resistance to clindamycin, which is used in Peru in combination with quinine to treat pregnant women and infants.

63. Griffing, S. M. *et al.* South American *Plasmodium falciparum* after the malaria eradication era: clonal population expansion and survival of the fittest hybrids. *PLoS ONE* **6**, e23486 (2011).

64. Anderson, T. J. *et al.* High heritability of malaria parasite clearance rate indicates a genetic basis for artemisinin resistance in western Cambodia. *J. Infect. Dis.* **201**, 1326–1330 (2010).

65. Anderson, T. J. *et al.* Inferred relatedness and heritability in malaria parasites. *Proc. Biol. Sci.* **277**, 2531–2540 (2010).

66. de Bakker, P. I. *et al.* Efficiency and power in genetic association studies. *Nature Genet.* **37**, 1217–1223 (2005).

67. Pulit, S. L., Voight, B. F. & de Bakker, P. I. Multiethnic genetic association studies improve power for locus discovery. *PLoS ONE* **5**, e12600 (2010).

68. Altshuler, D., Daly, M. J. & Lander, E. S. Genetic mapping in human disease. *Science* **322**, 881–888 (2008).

69. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).

70. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genet.* **38**, 904–909 (2006).

71. Pritchard, J. K., Stephens, M., Rosenberg, N. A. & Donnelly, P. Association mapping in structured populations. *Am. J. Hum. Genet.* **67**, 170–181 (2000).

72. Kang, H. M. *et al.* Variance component model to account for sample structure in genome-wide association studies. *Nature Genet.* **42**, 348–354 (2010).

73. Kang, H. M. *et al.* Efficient control of population structure in model organism association mapping. *Genetics* **178**, 1709–1723 (2008).

74. Price, A. L., Zaitlen, N. A., Reich, D. & Patterson, N. New approaches to population stratification in genome-wide association studies. *Nature Rev. Genet.* **11**, 459–463 (2010).
 This is a Review of GWAS methods that deal with population structure, focusing mostly on mixed-model approaches.

75. Thornton, T. & McPeek, M. S. ROADTRIPS: case-control association testing with partially or completely unknown population and pedigree structure. *Am. J. Hum. Genet.* **86**, 172–184 (2010).

76. Ochola, L. I. *et al.* Allele frequency-based and polymorphism-versus-divergence indices of balancing selection in a new filtered set of polymorphic genes in *Plasmodium falciparum*. *Mol. Biol. Evol.* **27**, 2344–2351 (2010).

77. Dvorin, J. D. *et al.* A plant-like kinase in *Plasmodium falciparum* regulates parasite egress from erythrocytes. *Science* **328**, 910–912 (2010).

78. Istvan, E. S. *et al.* Validation of isoleucine utilization targets in *Plasmodium falciparum*. *Proc. Natl Acad. Sci. USA* **108**, 1627–1632 (2011).
 This study takes a pharmacogenomic approach to select for *P. falciparum* parasites by using compounds that interfere with isoleucine use, which must be obtained by the parasite exogenously. Resistant parasites had mutations in isoleucyl-tRNA synthetase, showing that this approach can be used to elucidate new targets for drug development.

79. McNamara, C. & Winzeler, E. A. Target identification and validation of novel antimalarials. *Future Microbiol.* **6**, 693–704 (2011).

80. Meister, S. *et al.* Imaging of *Plasmodium* liver stages to drive next-generation antimalarial drug discovery. *Science* **334**, 1372–1377 (2011).

81. Nam, T. G. *et al.* A chemical genomic analysis of decoquinate, a *Plasmodium falciparum* cytochrome *b* inhibitor. *ACS Chem. Biol.* **6**, 1214–1222 (2011).

82. Rottmann, M. *et al.* Spiroindolones, a potent compound class for the treatment of malaria. *Science* **329**, 1175–1180 (2010).

83. Alonso, P. L. *et al.* A research agenda to underpin malaria eradication. *PLoS Med.* **8**, e1000406 (2011).

84. Farnert, A. *et al.* Genotyping of *Plasmodium falciparum* infections by PCR: a comparative multicentre study. *Trans. R. Soc. Trop. Med. Hyg.* **95**, 225–232 (2001).

85. Juliano, J. J. *et al.* Exposing malaria in-host diversity and estimating population diversity by capture-recapture using massively parallel pyrosequencing. *Proc. Natl Acad. Sci. USA* **107**, 20138–20143 (2010).
 This study used high-throughput sequencing to characterize malaria parasite diversity among patients and to evaluate parasite populations before and after drug treatment. These data are combined with computational strategies to study the evolution of drug resistance.

86. Snounou, G. *et al.* Biased distribution of *msp1* and *msp2* allelic variants in *Plasmodium falciparum* populations in Thailand. *Trans. R. Soc. Trop. Med. Hyg.* **93**, 369–374 (1999).

87. Daniels, R. *et al.* A general SNP-based molecular barcode for *Plasmodium falciparum* identification and tracking. *Malar. J.* **7**, 223 (2008).
 This paper describes the generation of TaqMan assays for 24 SNPs that have a high minor-allele frequency to distinguish individual parasite genomes.

88. Agnandji, S. T. *et al.* First results of phase 3 trial of RTS,S/AS01 malaria vaccine in African children. *N. Engl. J. Med.* **365**, 1863–1875 (2011).

89. Alloueche, A. *et al.* Protective efficacy of the RTS,S/AS02 *Plasmodium falciparum* malaria vaccine is not strain specific. *Am. J. Trop. Med. Hyg.* **68**, 97–101 (2003).

90. Bojang, K. A. *et al.* Efficacy of RTS,S/AS02 malaria vaccine against *Plasmodium falciparum* infection in semi-immune adult men in The Gambia: a randomised trial. *Lancet* **358**, 1927–1934 (2001).

91. Enosse, S. *et al.* RTS,S/AS02A malaria vaccine does not induce parasite CSP T cell epitope selection and reduces multiplicity of infection. *PLoS Clin. Trials* **1**, e5 (2006).

92. Waitumbi, J. N. *et al.* Impact of RTS,S/AS02$_A$ and RTS,S/AS01$_B$ on genotypes of *P. falciparum* in adults participating in a malaria vaccine clinical trial. *PLoS ONE* **4**, e7849 (2009).

93. Bozdech, Z. *et al.* The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS Biol.* **1**, E5 (2003).

94. Le Roch, K. G. *et al.* Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science* **301**, 1503–1508 (2003).

95. Duraisingh, M. T. *et al.* Heterochromatin silencing and locus repositioning linked to regulation of virulence genes in *Plasmodium falciparum*. *Cell* **121**, 13–24 (2005).

96. Foth, B. J., Zhang, N., Mok, S., Preiser, P. R. & Bozdech, Z. Quantitative protein expression profiling reveals extensive post-transcriptional regulation and post-translational modifications in schizont-stage malaria parasites. *Genome Biol.* **9**, R177 (2008).

97. Freitas-Junior, L. H. *et al.* Telomeric heterochromatin propagation and histone acetylation control mutually exclusive expression of antigenic variation genes in malaria parasites. *Cell* **121**, 25–36 (2005).
 This paper describes the use of high-throughput RNA sequencing approaches to detect novel *P. falciparum* transcripts, to correct gene models, to propose alternative splicing events and to predict untranslated regions. Differential expression patterns across the developmental life cycle were connected with DNA microarray analysis to advance our understanding of gene expression substantially.

98. Otto, T. D. *et al.* New insights into the blood-stage transcriptome of *Plasmodium falciparum* using RNA-seq. *Mol. Microbiol.* **76**, 12–24 (2010).

99. Raabe, C. A. *et al.* A global view of the nonprotein-coding transcriptome in *Plasmodium falciparum*. *Nucleic Acids Res.* **38**, 608–617 (2010).

100. Daily, J. P. *et al.* Distinct physiological states of *Plasmodium falciparum* in malaria-infected patients. *Nature* **450**, 1091–1095 (2007).

101. Lemieux, J. E. *et al.* Statistical estimation of cell-cycle progression and lineage commitment in *Plasmodium falciparum* reveals a homogeneous pattern of transcription in *ex vivo* culture. *Proc. Natl Acad. Sci. USA* **106**, 7559–7564 (2009).

102. Mackinnon, M. J. *et al.* Comparative transcriptional and genomic analysis of *Plasmodium falciparum* field isolates. *PLoS Pathog.* **5**, e1000644 (2009).

103. Mok, S. *et al.* Artemisinin resistance in *Plasmodium falciparum* is associated with an altered temporal pattern of transcription. *BMC Genomics* **12**, 391 (2011).
 DNA microarrays were used in this study to show that parasites that are obtained from patients exhibiting delayed clearance rates under artemisinin drug treatment show decreased metabolic activities in the ring stages. The data also suggest that the increased protein metabolism in the schizont stage may protect against oxidative stress and other effects of artemisinin.

104. De Silva, E. K. *et al.* Specific DNA-binding by apicomplexan AP2 transcription factors. *Proc. Natl Acad. Sci. USA* **105**, 8393–8398 (2008).

105. Balaji, S., Babu, M. M., Iyer, L. M. & Aravind, L. Discovery of the principal specific transcription factors of *Apicomplexa* and their implication for the evolution of the AP2-integrase DNA binding domains. *Nucleic Acids Res.* **33**, 3994–4006 (2005).

106. Aravind, L., Iyer, L. M., Wellems, T. E. & Miller, L. H. *Plasmodium* biology: genomic gleanings. *Cell* **115**, 771–785 (2003).

107. Coleman, B. I. & Duraisingh, M. T. Transcriptional control and gene silencing in *Plasmodium falciparum*. *Cell. Microbiol.* **10**, 1935–1946 (2008).

108. Hakimi, M. A. & Deitsch, K. W. Epigenetics in *Apicomplexa*: control of gene expression during cell cycle progression, differentiation and antigenic variation. *Curr. Opin. Microbiol* **10**, 357–362 (2007).

109. Horrocks, P., Wong, E., Russell, K. & Emes, R. D. Control of gene expression in *Plasmodium falciparum* — ten years on. *Mol. Biochem. Parasitol.* **164**, 9–25 (2009).

110. Iyer, L. M., Anantharaman, V., Wolf, M. Y. & Aravind, L. Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes. *Int. J. Parasitol.* **38**, 1–31 (2008).

111. Cui, L. & Miao, J. Chromatin-mediated epigenetic regulation in the malaria parasite *Plasmodium falciparum*. *Eukaryot. Cell* **9**, 1138–1149 (2010).

112. Broadbent, K. *et al.* A global transcriptional analysis of *Plasmodium falciparum* malaria reveals a novel family of telomere-associated lncRNAs. *Genome Biol.* **12**, R56 (2011).
 Uses a high-resolution DNA-tiling microarray to identify long non-coding RNAs (lncRNAs) localized to the telomeres, where they are proposed to have a role in telomere maintenance, virulence gene regulation and potentially in other processes that are related to chromosome end biology in the parasite.

113. Bright, A. T. & Winzeler, E. A. Noncoding RNA, antigenic variation, and the virulence genes of *Plasmodium falciparum*. *BMC Biol.* **9**, 50 (2011).

114. Su, X., Hayton, K. & Wellems, T. E. Genetic linkage and association analyses for trait mapping in *Plasmodium falciparum*. *Nature Rev. Genet.* **8**, 497–506 (2007).

115. Florens, L. *et al.* A proteomic view of the *Plasmodium falciparum* life cycle. *Nature* **419**, 520–526 (2002).

116. Khan, S. M. *et al.* Proteome analysis of separated male and female gametocytes reveals novel sex-specific *Plasmodium* biology. *Cell* **121**, 675–687 (2005).

117. Lasonder, E. *et al.* Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry. *Nature* **419**, 537–542 (2002).

118. Lasonder, E. *et al.* Proteomic profiling of *Plasmodium* sporozoite maturation identifies new proteins essential for parasite development and infectivity. *PLoS Pathog.* **4**, e1000195 (2008).

119. Patra, K. P., Johnson, J. R., Cantin, G. T., Yates, J. R. & Vinetz, J. M. Proteomic analysis of zygote and ookinete stages of the avian malaria parasite *Plasmodium gallinaceum* delineates the homologous proteomes of the lethal human malaria parasite *Plasmodium falciparum*. *Proteomics* **8**, 2492–2499 (2008).

120. Sinden, R. E. Reality check for malaria proteomics. *Genome Biol.* **10**, 211 (2009).

121. Wuchty, S., Adams, J. H. & Ferdig, M. T. A comprehensive *Plasmodium falciparum* protein interaction map reveals a distinct architecture of a core interactome. *Proteomics* **9**, 1841–1849 (2009).

122. Besteiro, S., Vo Duy, S., Perigaud, C., Lefebvre-Tournier, I. & Vial, H. J. Exploring metabolomic approaches to analyse phospholipid biosynthetic pathways in *Plasmodium*. *Parasitology* **137**, 1343–1356 (2010).

123. Olszewski, K. L. *et al.* Branched tricarboxylic acid metabolism in *Plasmodium falciparum*. *Nature* **466**, 774–778 (2010).

124. Olszewski, K. L. *et al.* Host-parasite interactions revealed by *Plasmodium falciparum* metabolomics. *Cell Host Microbe* **5**, 191–199 (2009).

125. Teng, R. *et al.* Metabolite profiling of the intraerythrocytic malaria parasite *Plasmodium falciparum* by ^{1}H NMR spectroscopy. *NMR Biomed.* **22**, 292–302 (2009).

126. Nsobya, S. L., Kiggundu, M., Joloba, M., Dorsey, G. & Rosenthal, P. J. Complexity of *Plasmodium falciparum* clinical samples from Uganda during short-term culture. *J. Infect. Dis.* **198**, 1554–1557 (2008).

127. Pologe, L. G. & Ravetch, J. V. Large deletions result from breakage and healing of *P. falciparum* chromosomes. *Cell* **55**, 869–874 (1988).

128. Venkatesan, M. *et al.* Using CF11 cellulose columns to inexpensively and effectively remove human DNA from *Plasmodium falciparum*-infected whole blood samples. *Malar J.* **11**, 41 (2012).

129. Melnikov, A. *et al.* Hybrid selection for sequencing pathogen genomes from clinical samples. *Genome Biol.* **12**, R73 (2011).
 This paper describes the use of the hybrid selection process to capture *P. falciparum* DNA from a DNA sample derived from patient blood samples containing a mixture of human and *P. falciparum* DNA.

FURTHER INFORMATION
Sarah K. Volkman's homepage: http://www.hsph.harvard.edu/research/sarah-cooke
Dyann F. Wirth's homepage: http://www.hsph.harvard.edu/wirthlab
Broad Institute: http://www.broadinstitute.org
Harvard Malaria Initiative: http://www.hsph.harvard.edu/research/hmi
PlasmoDB: http://plasmodb.org/plasmo
Nature Reviews Genetics Series on Study designs: http://www.nature.com/nrg/series/studydesigns/index.html

ALL LINKS ARE ACTIVE IN THE ONLINE PDF