A global network for investigating the genomic epidemiology of malaria

The Malaria Genomic Epidemiology Network*

Large-scale studies of genomic variation could assist efforts to eliminate malaria. But there are scientific, ethical and practical challenges to carrying out such studies in developing countries, where the burden of disease is greatest. The Malaria Genomic Epidemiology Network (MalariaGEN) is now working to overcome these obstacles, using a consortial approach that brings together researchers from 21 countries.

Each year, malaria kills about 1 million children and causes debilitating illness in more than 500 million people[1]. Underlying this massive global health problem is a remarkable biological phenomenon, the co-evolution of three eukaryotic genomes[2–9] (Table 1). The disease is caused by single-celled parasites of the genus *Plasmodium*, which invade, and reproduce in, human erythrocytes. The parasites are then transmitted from one person to another by blood-sucking mosquitoes of the genus *Anopheles*.

The evolutionary 'arms race' between the parasite, its vector and the human host is central to the problem of controlling disease. *Plasmodium* populations are continually evolving to resist antimalarial drugs and have sophisticated genetic mechanisms of evading the human immune system, presenting a major problem for the development of a vaccine against malaria[2,7,10]. *Anopheles* populations are likewise evolving to resist the insecticides that are used to control malaria, but they also have genetic defences against the parasite that might provide clues to new control strategies[11,12]. Malaria has also been a strong force for recent evolutionary selection in the human genome[9,13], and uncovering all of the human genetic factors that confer resistance to malaria would provide clues to the molecular basis of protective immunity that would be invaluable for vaccine developers.

The genetic basis of human resistance to malaria can now be investigated systematically at the level of the whole genome, by using genome-wide association (GWA) analysis. In a typical GWA study, the genotype of thousands of individuals is determined at the positions of half a million or more single nucleotide polymorphisms (SNPs)[14] (see page 728). The ultimate goal of GWA analysis is to uncover all of the DNA sequence variants that affect an individual's risk of disease, without sequencing the whole genome, by using statistical inferences based on common patterns of variation in the genome.

An important question that can be addressed by GWA analysis is why only some children develop severe malaria (that is, life-threatening forms of the disease[15,16]) in communities in which every child is repeatedly infected with *Plasmodium falciparum*, the species of parasite that is responsible for most deaths from malaria. Only a small proportion of *P. falciparum* infections progress to severe malaria, and epidemiological data indicate that about 25% of the risk is determined by human genetic factors[17] (Box 1). A typical study design is to recruit individuals with severe malaria (cases) in a hospital setting and to recruit control individuals essentially randomly from the general population. By comparing the frequency of a set of SNPs in cases and controls, it is possible to estimate the effect of different sequence variants on an individual's risk of developing severe malaria. Because the risk of developing severe malaria is probably determined by many genetic factors and environmental factors operating at different stages of infection, the effect of any one factor might be small, so a large number of individuals must be studied to obtain statistically significant results.

Similar approaches could, in principle, be used to investigate the emergence and molecular basis of drug resistance in *Plasmodium* populations or insecticide resistance in *Anopheles* populations. However, this cannot be put into practice until genomic variation in *Plasmodium* and *Anopheles* populations is better understood[3,18–21]. A complicating factor for GWA studies of *P. falciparum* is that in a single infection, the parasites that are transmitted can have different genotypes, so an individual who is infected frequently can carry a parasite population of great genetic complexity. Another is that, in Africa, where malaria is most prevalent, the *P. falciparum* genome has low levels of linkage disequilibrium[19,21]. Linkage disequilibrium is a fundamental concept to consider in GWA analysis; it refers to the correlation between genotypes that is observed at neighbouring positions in the genome. The lower the level of linkage disequilibrium, the more positions in the genome need to be genotyped for an effective GWA study. Recent technological advances in massively parallel sequencing of single DNA molecules[22] might help to overcome both of these problems, by enabling *P. falciparum* to be genotyped at a very large number of positions in the genome and by helping to distinguish the different parasite genotypes that can constitute a single infection.

In this Commentary, we describe how a global research network has been established to investigate the effects of genomic variation in humans on the biology and pathology of malaria. We focus on the human genome because the tools for genotyping and the framework for population genetics are further advanced than those for *Plasmodium* and *Anopheles* species. More specifically, we outline the practical reasons why malaria is more challenging to study by GWA analysis than many other common diseases, and we describe how we have established several projects that bring together large-scale studies carried out in multiple locations to address key scientific questions. We also describe the procedures that we use for standardizing and integrating data from different investigators, as well as the policies that we have developed to deal with issues of sample and data ownership, data release, intellectual property and ethics.

Challenges of GWA studies of malaria

The genetic analysis of human resistance to malaria is challenging at several levels, ranging from the practical and ethical issues of clinical research in the developing world to the statistical genetic issues arising from the great diversity of the populations that are affected.

*A list of participants and affiliations appears in the online version of the paper at www.nature.com/nature.

Table 1 | Malaria involves three eukaryotic genomes

Aspect	*Plasmodium falciparum** (ref. 2)	*Anopheles gambiae*† (ref. 3)	Humans[4,5]
Genomic organization	Haploid (transiently diploid in mosquitoes) 14 chromosomes Plus the mitochondrial genome and apicoplast genome	Diploid 2 autosomal chromosomes, X chromosome, Y chromosome Plus the mitochondrial genome	Diploid 22 autosomal chromosomes, X chromosome, Y chromosome Plus the mitochondrial genome
Genome size (base pairs)	2.3×10^7	2.8×10^8	3.2×10^9
Estimated number of protein-coding genes	5,420	12,945	21,714
Databases of common genomic variants	$\sim 10^5$ unvalidated SNPs listed by dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP)	$\sim 10^6$ unvalidated SNPs listed by dbSNP	$\sim 6 \times 10^6$ validated SNPs listed by dbSNP $\sim 3 \times 10^6$ SNPs investigated by the International HapMap Project[6] (http://www.hapmap.org) $\sim 3 \times 10^4$ structural variants listed by the Database of Genomic Variants (http://projects.tcag.ca/variation)
Examples of evolutionary selection	Resistance to antimalarials[7]: to antifolates, mediated by *DHFR* and *DHPS*; to chloroquine, mediated by *CRT* and *MDR*	Resistance to insecticides[8]: to DDT and pyrethroids, mediated by the *kdr* allele, a variant of a gene encoding a voltage-gated sodium channel	Resistance to severe malaria[9]: mediated by mutations in *HBB* (resulting in different forms of haemoglobin, HbS and HbC); the presence of *HBA1* or *HBA2* (which encode different forms of the α-chain of haemoglobin, resulting in α⁺-thalassaemia); by deficiency in *G6PD*

CRT, chloroquine-resistance transporter gene; DDT, dichlorodiphenyltrichloroethane; *DHFR*, dihydrofolate reductase gene; *DHPS*, dihydropteroate synthase gene; *G6PD*, glucose-6-phosphate dehydrogenase gene; HBB, β-globin gene (which encodes the β-chain of haemoglobin); *kdr*, knockdown resistance gene; *MDR*, multidrug-resistance gene; SNP, single nucleotide polymorphism. **Plasmodium* is a genus of protozoan. There are more than 100 species, of which 5 can infect humans: *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium malariae*, *Plasmodium ovale* and some forms of *Plasmodium knowlesi*. *P. falciparum* is the most severe form and is responsible for most deaths from malaria. †*Anopheles* is a genus of mosquito. There are about 400 species, of which about 40 transmit the *Plasmodium* parasites that cause human malaria. *Anopheles gambiae* is a major vector of *P. falciparum*.

Recruiting a large number of individuals with severe malaria presents challenges because most of the burden of malaria falls on poor communities with underfunded health services and no systematic medical records. A considerable proportion of children with severe malaria die within hours of reaching a hospital; therefore, for the clinical phenotype of malaria to be classified properly, research information must be gathered at the time of hospital admission. This implies considerable responsibilities on the part of the research team for ensuring standards of medical care, particularly in a resource-poor setting. Also, it is not feasible to take large amounts of blood from children who are ill with malaria, many of whom are anaemic, so it is often necessary to use whole-genome amplification to obtain enough DNA for genotyping at numerous SNP positions. This can reduce genotyping efficiency and thus diminish statistical power, making an even larger sample size necessary[23].

In addition, designing an appropriate 'SNP genotyping' strategy for GWA studies of malaria is complicated by the large amount of genomic variation in Africa. Because of the low levels of linkage disequilibrium in populations in Africa, genotypes need to be sequenced at the positions of more SNPs than in studies of European populations. On the basis of the initial data from the International HapMap Project (http://www.hapmap.org), it was estimated that a GWA study of about 1.5 million SNPs in an African population would be approximately equivalent to a study of 0.6 million SNPs in a European population, in terms of the ability to tag a high proportion of common sequence variants[6]. But it is difficult to estimate how many SNPs will be required to tag all common variants until resequencing studies have generated a comprehensive list of common sequence variants in different African populations[24].

Furthermore, the ethnic diversity of African populations presents numerous statistical challenges for GWA studies. Many African communities consist of several ethnic groups, and minor differences in the ethnic composition of the case groups and the control groups can lead to false-positive genetic associations. To exclude such artefacts, studies need to be designed carefully, and statistical genetic methods that correct for population structure need to be applied[25].

Another problem arises from the genetic differences between populations in Africa, as opposed to within a single population. Signals of association are not expected to be constant across GWA studies carried out at different locations in Africa. For example, differences in haplotype structure can result in variable signals of association around a causal variant of a disease, particularly in 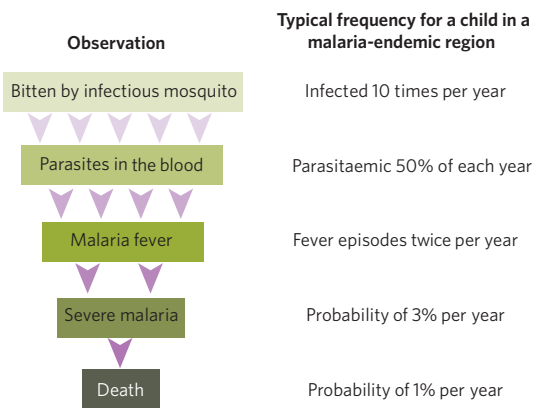genomic regions that have recently undergone evolutionary selection. Also, different populations can harbour different factors that confer resistance to malaria. One example of this is two resistance-associated forms of haemoglobin (haemoglobin S and haemoglobin C) that result from different SNPs at adjacent locations in *HBB*, the gene that encodes the β-chain of haemoglobin — these SNPs have different patterns of distribution in West Africa[26,27].

But such differences between populations can be also highly informative. For example, they can aid in uncovering genetic factors that have evolved in specific populations and in investigating interactions between genes and the environment. Importantly, differences in the patterns of linkage disequilibrium between populations can help to distinguish a causal variant from neighbouring polymorphisms. This is necessary because many SNPs that have been associated with particular diseases are not the causal variants but show an association signal simply as a result of correlation with the causal variant (because of linkage disequilibrium). Thus, GWA studies carried out at multiple sites in Africa could provide a rich resource for identifying causal variants.

Developing a global research network

In the past, research into the human genetic factors that affect resistance to malaria has been characterized by multiple research groups each pursuing relatively small studies on their own samples. But the chance of making a discovery, and replicating the finding, is greatly increased if there are effective mechanisms for different research groups to share data and thereby enlarge the number of samples that are studied. The concept of forming a network for sharing data on the genomic epidemiology of malaria — which was to become the Malaria Genomic Epidemiology Network (MalariaGEN) — originated from work that was funded in 2003 by the Bill & Melinda Gates Foundation and by the UK Medical Research Council. The purpose of this funding was to develop web-based software that would allow the integration of clinical and genetic data collected by different research groups. This funding also supported a workshop on the ethical and ownership issues involved in sharing data, which was held in Accra, Ghana, in January 2004 and attended by scientists and clinical researchers from ten research groups in Africa.

MalariaGEN was established in 2005, with joint funding from the Bill & Melinda Gates Foundation (through the Foundation for the National Institutes of Health) and the Wellcome Trust, as part of the Grand Challenges in Global Health initiative[28] (http://www.gcgh.org). The purpose of this joint funding was to discover mechanisms of protective immunity

Box 1 | Progression of malarial disease in a malaria-endemic region

Observation	Typical frequency for a child in a malaria-endemic region
Bitten by infectious mosquito	Infected 10 times per year
Parasites in the blood	Parasitaemic 50% of each year
Malaria fever	Fever episodes twice per year
Severe malaria	Probability of 3% per year
Death	Probability of 1% per year

In a malaria-endemic region, there is a large variation in the clinical severity of infections with *Plasmodium falciparum*. When a young child becomes infected, he or she usually becomes ill with fever but eventually recovers. A small proportion of infections progress to severe malaria; that is, forms of the disease that have life-threatening complications, such as profound anaemia or cerebral malaria[15,16]. After repeated infections, older children and adults acquire clinical immunity to malaria, meaning that they can tolerate infection without developing symptoms.

The figure illustrates the likelihood of progression from infection with *P. falciparum* to death for a young child living in a malaria-endemic region, showing frequencies that are representative for such a child. In any given situation, the frequencies will depend greatly on environmental factors, such as mosquito biting rates. Despite the importance of environment, human genetic factors are estimated to account for approximately 25% of the variation among African children in the risk of developing severe malaria[17]. Known genetic factors that confer resistance to malaria, such as the allele that encodes the HbS form of haemoglobin (ref. 9) (Table 1), account for only a small proportion of this variation, implying that many genetic factors involved in resistance remain to be uncovered[17].

Such genetic factors might operate at any stage of the progression from receiving a bite from a malaria-carrying mosquito to dying as a result of severe malaria. For example, individuals who carry the HbS-encoding allele have a tenfold lower risk of severe malaria than those who do not carry this allele, and the mechanism of protection seems to be that HbS acts to suppress the number of parasites in the blood.

to malaria by combining analysis of human genome variation with large-scale epidemiological studies in malaria-endemic regions. Five objectives necessary for achieving this goal were identified: building a global network for sharing data on the genomic epidemiology of malaria; collecting DNA and clinical data from individuals with different phenotypes of malaria; characterizing genetic variation in populations in malaria-endemic regions; identifying genetic variants that provide protection against severe malaria; and defining the immunological mechanisms by which such genetic variants exert their protective effect.

The group of researchers who came together to tackle these objectives, the MalariaGEN investigators, are mainly leaders of clinical, epidemiological or immunological research projects in malaria-endemic areas, and they contribute samples and data to the MalariaGEN programme. Other MalariaGEN investigators contribute expertise and technical resources related to high-throughput analysis of genomic variation, statistical genetics or biomedical ethics. The host institutions of MalariaGEN investigators, the MalariaGEN partner institutions, are located in 15 malaria-endemic countries and 6 other countries (for additional information, see http://www.malariagen.net/resource/1), and the institutions in malaria-endemic countries have well-established study sites, where individuals are recruited to participate in research. Most of these study sites are in sub-Saharan Africa: in Burkina Faso, Cameroon, Gambia, Ghana, Kenya, Malawi, Mali, Nigeria, Senegal, Sudan and Tanzania. There are also MalariaGEN study sites in Papua New Guinea, Sri Lanka, Thailand and Vietnam.

To address the complexities involved in setting up such a global research network, MalariaGEN investigators agreed, at an inaugural meeting in Oxford, United Kingdom, in July 2005, to establish the network in four stages. The first stage was to establish a set of principles and processes, agreed by all investigators, to regulate a central resource of DNA samples and phenotypic data (Box 2 and see http://www.malariagen.net/resource/1). More specifically, this involved standardizing scientific definitions and procedures, enabling partners to gain secure access to the data resource via the Internet, and developing rules about data sharing, intellectual property and appropriate consent.

The second stage was to define a core scientific programme of large-scale experiments and statistical analysis, which would use data and expertise from multiple investigators, and the results of which would belong jointly to all of the investigators involved. Projects that are part of this core programme are called Consortial Projects (Box 2). There are four such projects so far, and each has a specific objective and a plan of action (Table 2). After a Consortial Project has been defined, each investigator decides whether he or she wishes to contribute to the project.

The third stage was to find ways of assisting investigators in malaria-endemic countries to develop clinical and epidemiological studies that would advance the core scientific programme. Investigators were invited to submit funding proposals for projects at their study sites that would contribute to Consortial Projects, using the research infrastructure of the local partner institution and founded on the scientific interests and expertise of the local investigators. Funding was allocated after proposals had been reviewed by a group of investigators that represented the network as a whole (with members from Cameroon, Gambia, Ghana, Italy, Kenya, Malawi, Mali, Sri Lanka, Sudan, Tanzania and the United Kingdom). This group evaluated both the scientific design and the feasibility of the clinical and epidemiological studies proposed, taking into account the infrastructure and expertise of the local partner institution and study site.

The fourth stage was to strengthen the capacity to manage data, and to carry out statistical and genetic analyses, at partner institutions in malaria-endemic countries. A fellowship programme in data analysis was established. After an open application process, a data fellow was appointed at each partner institution. Most of the MalariaGEN data fellows work on the team of a MalariaGEN investigator and have responsibilities for managing the team's data. All data fellows receive training and support in data management, statistical genetics and computing skills. This training is provided by a team of expert statisticians, geneticists and computer programmers who work at the MalariaGEN Resource Centre, which is based at two locations in the United Kingdom, at the University of Oxford and at the Wellcome Trust Sanger Institute near Cambridge. Members of the MalariaGEN resource centre organize regular data-analysis workshops, both in the United Kingdom and at partner institutions in malaria-endemic countries. These workshops provide structured teaching, together with an opportunity for data fellows to share their experiences and to analyse their own data with hands-on assistance from an expert.

Dealing with data

Sharing data is a simple concept but, when many investigators and partner institutions are involved, it can be complex to put into practice. There is the technical issue of how to amalgamate data from different research groups. There needs to be transparency about the ownership and permitted uses of the data and samples contributed by investigators. Procedures need to be established for releasing data and, where appropriate, for protecting intellectual property. This section outlines how MalariaGEN has dealt with each of these areas.

Standardizing and integrating data

Standardizing and integrating data from multiple study sites is central to MalariaGEN's mission. As an example, Consortial Project 1 (Table 2), which is the core project of MalariaGEN's programme, depends on there being a standardized clinical definition of severe malaria. Severe malaria consists of several overlapping clinical syndromes, often referred to as

subphenotypes: these include cerebral malaria (which is characterized by coma), profound anaemia and respiratory distress. Some genetic factors confer resistance generally to severe malaria, whereas others might be specific for a subphenotype. The clinical definition of severe malaria therefore depends on a combination of observations, some of which (for example, respiratory distress) can be quantified less precisely than others (for example, anaemia, through measuring haemoglobin concentration), and there is ongoing research into how to minimize the diagnostic error rate. After consulting MalariaGEN investigators — and after a joint meeting with the Severe Malaria in African Children network[16], in Yaoundé, Cameroon, in November 2005 — a standardized case report form was agreed (see http://www.malariagen.net/resource/1). This form is not intended to replace the case report forms used by individual investigators but rather to provide a template for extracting core information from different clinical data sets in a standardized manner, while giving investigators the freedom to collect data in the way that is most appropriate to their own research.

In a large research network, there will be site-to-site variation in the way in which clinical and epidemiological information is recorded and stored at the local level, so investigators and data fellows are encouraged to have an active role in data standardization and integration. This is facilitated by web-based software developed specifically for this purpose by the MalariaGEN resource centre. Investigators collect data using the database format that is best supported at their institution, and they periodically upload their data via a secure, password-protected interface to a personalized section of the MalariaGEN website, which cannot be accessed by others. Tools are provided for checking data integrity and for transferring data into the database for the relevant Consortial Project. The process of data transfer generally requires the investigator to recode or transform certain variables in their own data set to match the format of the project database, and the web-based software assists and documents this process.

MalariaGEN investigators are also working on the standardization of immunological assays as part of Consortial Project 2, which involves investigating the genetic determinants of the immune response in different populations and environmental settings (Table 2). In the first phase of this project, antibody measurements are being carried out at a central reference laboratory to ensure that data from different study sites can be directly compared. In the long term, the project seeks to develop robust methods and standardized reagents that will enable reference laboratories to be established at partner institutions.

Sharing data and establishing rules of ownership

MalariaGEN is a data-sharing community in which independent investigators with different projects and research objectives contribute to a central repository of DNA samples and a central database of core phenotypic data for each Consortial Project. General principles of data sharing and ownership were agreed at the inaugural meeting of MalariaGEN (Box 2 and see http://www.malariagen.net/resource/1). The major findings of each Consortial Project will be published in scientific journals, with all investigators who contributed to the project listed as authors. In addition, investigators are encouraged to analyse the data that have been generated from their own samples, and to incorporate any additional clinical or experimental data that they have for these samples; these analyses are then permitted to be published independently of the findings of the Consortial Project.

One of the most important considerations when building the database for each Consortial Project was protecting the anonymity of research participants. The MalariaGEN database contains no personal identifiers and is not linked to databases at local study sites. However, one of MalariaGEN's key principles is that investigators should be able to analyse data generated from samples that they contributed and to amalgamate these data with locally held phenotypic data. A standard operating procedure was therefore developed to ensure that the local databases held by partner institutions that contain data generated by MalariaGEN are designed and used according to appropriate ethical guidelines (see http://www.malariagen.net/resource/1).

> **Box 2 | Key elements of MalariaGEN's policy**
>
> MalariaGEN investigators have agreed on a set of principles and processes for sharing samples and data. An important step was to define several Consortial Projects, each of which has a specific objective and project plan (Table 2). Investigators can control how their samples and data are used by MalariaGEN by specifying which Consortial Projects they wish to contribute to.
>
> **Definitions**
> • Consortial Project: a project that uses data and expertise from multiple investigators, is carried out with core MalariaGEN funds, and is agreed by the Project Management Committee, the funding bodies and all of the investigators taking part in the project.
> • Contributing investigator: an investigator who contributes data, samples or expertise to a Consortial Project.
> • Investigator's own analysis: analysis carried out by a contributing investigator on his or her own samples, either using data generated by MalariaGEN or facilitated in another way by MalariaGEN.
>
> **Key principles**
> • The ownership of physical samples and clinical data contributed to Consortial Projects remains with the contributing investigator.
> • The contributing investigator can request that DNA samples be returned at any stage after the agreed experiments have taken place and can use the samples for purposes other than those of MalariaGEN.
> • Genotyping data generated by a Consortial Project is fully accessible to the investigator who contributed the samples and may be used for the investigator's own analysis.
> • The contributing investigator is responsible for ensuring that samples are taken with the participants' informed consent and for gaining local ethical approval, with support from the MalariaGEN ethics team.
> • The laboratories that process samples and analyse data are responsible for the safety and integrity of the samples, and for maintaining security and confidentiality of information.
> • Authorship of publications by MalariaGEN will reflect the contributions of all who have provided data and expertise, in accordance with normal academic practice.
> • Data at the level of the individual from human GWA studies will be made available to the scientific community through an independent data-access committee.
> • Intellectual-property protection will be sought only if it will expedite the translation of a scientific discovery into affordable health benefits for the populations that are most in need.
> • For further details of MalariaGEN consortial policies, see http://www.malariagen.net/resource/1

Releasing data and protecting data as intellectual property

Because the scientific benefits of GWA studies are cumulative, the value of a single study can be increased substantially if the data for individual subjects are available to the wider scientific community, provided that the identity of these individuals is securely protected[14,29]. MalariaGEN's policy on this topic was developed in consultation with all MalariaGEN investigators and with ethics-review boards at several MalariaGEN partner institutions (see http://www.malariagen.net/resource/1). In broad terms, the data-release policy seeks to permit research that is consistent with the nature of informed consent and the uses of the samples agreed by the relevant ethics-review boards. A key concern that arose from the consultation was to guard against the data being used in a way that might lead to any form of ethnic stigmatization. Another concern was to ensure that the timeline for data release is fair for investigators in malaria-endemic countries who have contributed resources and data to a project, because these investigators generally have less capacity for analysing genetic data than researchers in rich countries. Balancing the benefits of prompt data release with the need to protect the interests of partner institutions, MalariaGEN's current policy is to release GWA data 9 months after contributing investigators have had access to the complete data set. Data are placed in the European Genotype Archive (http://www.ebi.ac.uk/ega) and are then made available on application

Table 2 | MalariaGEN Consortial Projects

Project	Study sites	Project outline
1. Analysis of human genome variants associated with resistance or susceptibility to severe malaria	Burkina Faso, Cameroon, Gambia, Ghana (Navrongo and Kumasi), Kenya, Malawi, Mali, Nigeria, Papua New Guinea, Tanzania and Vietnam	• Study sites have recruited >10,000 cases and >10,000 ethnically matched controls for genetic association studies of severe malaria, with family trios being recruited at four study sites. • GWA studies of >10,000 individuals to be completed in 2009 for three study sites.
2. Analysis of human genetic factors that determine antibody responses to malaria	Burkina Faso, Kenya, Mali, Senegal, Sri Lanka, Sudan and Tanzania	• Standardized assays for antibodies specific for malaria antigens are being developed. • Assays will be used in a range of community-based studies with various epidemiological designs, to test associations between candidate genes and antibody responses.
3. Analysis of human genome variation in populations in which malaria is endemic	Burkina Faso, Cameroon, Gambia, Ghana, Kenya, Mali, Nigeria, Papua New Guinea, Sudan, Tanzania and Vietnam	• A DNA repository is being established to allow detailed studies of human genome variation in malaria-endemic populations. • Each site has recruited ≥90 individuals and has ethical approval for sequence data to be publicly released.
4. Genetic linkage studies of quantitative traits associated with malaria	Ghana, Senegal and Thailand	• A data-sharing collaboration has been set up between investigators working on genetic linkage studies of the intensity of malaria infection and other quantitative traits.

to an independent data-access committee (as described on the MalariaGEN Data Access web page, http://www.malariagen.net/access). As an additional check and balance, a working group is being established to represent partner institutions and ethics-review boards in malaria-endemic countries, and this group will be kept informed about applications for access to data and consulted about any proposed changes to the data-release policy.

It was important for MalariaGEN to develop guidelines on the circumstances in which data should be protected as intellectual property before publication, with careful consideration of arguments for and against patenting discoveries[30]. On the one hand, if a scientific discovery could lead to health benefits, then every effort should be made to make these benefits available to those who need them most, a process that could involve patenting the discovery. On the other hand, there is an argument for releasing data as openly as possible when there are no immediate applications for improving health and when open access to the data could drive innovations that might lead to health benefits. Arguably, for genomic epidemiology data, the prompt release of scientific findings is, in general, the appropriate course of action, but occasionally there might be discoveries that are exceptions to this. MalariaGEN's current policy is that intellectual-property protection should be sought if all three of the following conditions are satisfied: the discovery must be directly relevant to a medical application; it must be probable that the intellectual property will be licensed for development immediately; and the discovery must have been shown to require intellectual-property protection as a stimulus for further development (see http://www.malariagen.net/resource/1). In such cases, intellectual property will be licensed to non-profit organizations if possible. And, if financial benefits arise, then MalariaGEN will seek to ensure that these benefits flow to the communities who participated in the research.

Engaging with ethical issues

A range of ethical and social issues arise in establishing a network to share data between investigators in many countries. Ensuring ethical standards for the conduct of clinical research in developing countries raises many complex issues[31]. And the accumulation of detailed genomic information about individuals is raising new questions for society in general[32]. This combination of ethical and social issues needs to be addressed appropriately[33]. MalariaGEN has therefore established a team with expertise in medical ethics, which works with investigators and partner institutions to assess the ethical and social issues at different study sites, with the aim of establishing best practices for the ethical conduct of research carried out by MalariaGEN. This ethics team also develops training materials for investigators and ethics-review boards and has held workshops in Kenya, Mali, Thailand and Vietnam. To support investigators in tackling specific ethical issues and to gain an understanding of local practices, members of the team have also visited study sites in Cameroon, Gambia, Ghana, Kenya, Malawi, Mali, Papua New Guinea, Senegal and Sudan.

One of the most important aspects of this work is to find effective ways of communicating with research participants[34]. For example, when a very sick child is brought from her village to a busy government hospital, and her parents are asked whether part of the diagnostic blood sample can be used for a research project, it is often difficult to convey the distinction between medical diagnosis and medical research. Terms such as 'research', 'genetics', 'laboratory' and 'database' might be meaningless unless a concise and effective way is found of translating these concepts into the local language (by using examples and metaphors drawn from local experience), without creating anxiety by information overload. After consulting investigators and ethics-review board members, MalariaGEN has developed a template and guidelines for obtaining informed consent from participants in genetic studies of resistance to malaria (see http://www.malariagen.net/resource/1). To understand how guidelines can be put into practice most effectively, the ethics team is also undertaking empirical research on the process of gaining informed consent at different study sites, with the objective of establishing best practice across MalariaGEN study sites, while being sensitive to local culture and practices.

The ethics team is also working to develop models of consultation at the community level that are appropriate for diverse cultural settings. A sensitive issue for many communities is the potential abuse of genetic data relating to ethnicity, which could result in stigmatization. Qualitative research is being carried out to understand the perspectives of communities and other stakeholders on the collection and use of information about ethnicity in genomic epidemiology projects. The aim is to develop guidelines for the publication and release of data about ethnicity that will provide the maximum scientific benefit while safeguarding the interests of participants and their communities.

Many of the ethical and social challenges confronting MalariaGEN stem from the diversity inherent in a large scientific enterprise with partners in rich and poor countries that span multiple disciplines, from clinical research and community-based research to state-of-the-art genomics and bioinformatics. Often, partners need to agree on an appropriate balance between standardization and shared practices on the one hand, and diversity and sensitivity to local circumstances on the other hand. MalariaGEN's procedures for data integration and guidelines for informed consent are examples of this process.

Looking forward

In September 2008, an ambitious plan for the elimination of malaria was announced — the Global Malaria Action Plan (http://www.rbm.who.int/gmap). This plan, which is supported by major international development agencies and governments around the world, seeks to halve the number of malaria cases worldwide by 2010 and to eliminate deaths from malaria almost completely by 2015. But it cannot succeed without effective insecticides and antimalarial drugs. And even if the plan's goals for the next decade are achieved, the chance of controlling and eliminating malaria over the long term will be greatly increased if an effective vaccine becomes available.

The new science of genomic epidemiology could assist these efforts to eliminate malaria, by providing more effective ways of monitoring the emergence of parasite resistance to antimalarial drugs and of mosquito resistance to insecticides, and by providing new leads for malaria vaccine development based on a better understanding of the natural mechanisms of protective immunity.

If genomic epidemiology is to make a contribution in this way, there need to be mechanisms in place to help researchers both in malaria-endemic countries and worldwide to pool their resources. Research groups in malaria-endemic countries need access to the technical expertise and infrastructure for the large-scale analysis of genomic variation. And research groups worldwide need to combine forces to analyse the massive amounts of data being generated by these studies, leading the way for important discoveries to be made. The MalariaGEN community is endeavouring to learn how to build and maintain the relationships, shared values and best practices that underpin this new type of scientific collaboration. ∎

1. Snow, R. W., Guerra, C. A., Noor, A. M., Myint, H. Y. & Hay, S. I. The global distribution of clinical episodes of *Plasmodium falciparum* malaria. *Nature* **434,** 214–217 (2005).
2. Gardner, M. J. *et al.* Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* **419,** 498–511 (2002).
3. Holt, R. A. *et al.* The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* **298,** 129–149 (2002).
4. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409,** 860–921 (2001).
5. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291,** 1304–1351 (2001).
6. The International HapMap Consortium. A haplotype map of the human genome. *Nature* **437,** 1299–1320 (2005).
7. 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).
8. Hemingway, J., Field, L. & Vontas, J. An overview of insecticide resistance. *Science* **298,** 96–97 (2002).
9. Kwiatkowski, D. P. How malaria has affected the human genome and what human genetics can teach us about malaria. *Am. J. Hum. Genet.* **77,** 171–190 (2005).
10. Wootton, J. C. *et al.* Genetic diversity and chloroquine selective sweeps in *Plasmodium falciparum*. *Nature* **418,** 320–323 (2002).
11. Ranson, H. *et al.* Evolution of supergene families associated with insecticide resistance. *Science* **298,** 179–181 (2002).
12. Riehle, M. M. *et al.* Natural malaria infection in *Anopheles gambiae* is regulated by a single genomic control region. *Science* **312,** 577–579 (2006).
13. Sabeti, P. C. *et al.* Positive natural selection in the human lineage. *Science* **312,** 1614–1620 (2006).
14. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447,** 661–678 (2007).
15. Marsh, K. *et al.* Indicators of life-threatening malaria in African children. *N. Engl. J. Med.* **332,** 1399–1404 (1995).
16. Taylor, T. *et al.* Standardized data collection for multi-center clinical studies of severe malaria in African children: establishing the SMAC network. *Trans. R. Soc. Trop. Med. Hyg.* **100,** 615–622 (2006).
17. Mackinnon, M. J., Mwangi, T. W., Snow, R. W., Marsh, K. & Williams, T. N. Heritability of malaria in Africa. *PLoS Med.* **2,** e340 (2005).
18. Mu, J. *et al.* Genome-wide variation and identification of vaccine targets in the *Plasmodium falciparum* genome. *Nature Genet.* **39,** 126–130 (2007).
19. Mu, J. *et al.* Recombination hotspots and population structure in *Plasmodium falciparum*. *PLoS Biol.* **3,** e335 (2005).
20. Jeffares, D. C. *et al.* Genome variation and evolution of the malaria parasite *Plasmodium falciparum*. *Nature Genet.* **39,** 120–125 (2007).
21. Volkman, S. K. *et al.* A genome-wide map of diversity in *Plasmodium falciparum*. *Nature Genet.* **39,** 113–119 (2007).
22. Hillier, L. W. *et al.* Whole-genome sequencing and variant discovery in *C. elegans*. *Nature Methods* **5,** 183–188 (2008).
23. Teo, Y. Y. *et al.* Whole genome-amplified DNA: insights and imputation. *Nature Methods* **5,** 279–280 (2008).
24. Bhangale, T. R., Rieder, M. J. & Nickerson, D. A. Estimating coverage and power for genetic association studies using near-complete variation data. *Nature Genet.* **40,** 841–843 (2008).
25. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genet.* **38,** 904–909 (2006).
26. Agarwal, A. *et al.* Hemoglobin C associated with protection from severe malaria in the Dogon of Mali, a West African population with a low prevalence of hemoglobin S. *Blood* **96,** 2358–2363 (2000).
27. Modiano, D. *et al.* Haemoglobin C protects against clinical *Plasmodium falciparum* malaria. *Nature* **414,** 305–308 (2001).
28. Varmus, H. *et al.* Grand challenges in global health. *Science* **302,** 398–399 (2003).
29. Manolio, T. A. *et al.* New models of collaboration in genome-wide association studies: the Genetic Association Information Network. *Nature Genet.* **39,** 1045–1051 (2007).
30. Chokshi, D. A., Parker, M. & Kwiatkowski, D. P. Data sharing and intellectual property in a genomic epidemiology network: policies for large-scale research collaboration. *Bull. World Health Organ.* **84,** 382–387 (2006).
31. Nuffield Council on Bioethics. *The Ethics of Research Related to Healthcare in Developing Countries.* <http://www.nuffieldbioethics.org/go/ourwork/developingcountries/publication_309.html> (Nuffield Council on Bioethics, 2002).
32. Lunshof, J. E., Chadwick, R., Vorhaus, D. B. & Church, G. M. From genetic privacy to open consent. *Nature Rev. Genet.* **9,** 406–411 (2008).
33. Chokshi, D. & Kwiatkowski, D. Ethical challenges of genomic epidemiology in developing countries. *Genomics Soc. Policy* **1,** 1–15 (2005).
34. Chokshi, D. A. *et al.* Valid consent for genomic epidemiology in developing countries. *PLoS Med.* **4,** e95 (2007).

The Malaria Genomic Epidemiology Network

Lead Investigators Eric Akum Achidi[1], Tsiri Agbenyega[2], Stephen Allen[3,4], Olukemi Amodu[5], Kalifa Bojang[6], David Conway[6], Patrick Corran[7], Panos Deloukas[8], Abdoulaye Djimde[9], Amagana Dolo[9], Ogobara Doumbo[9], Chris Drakeley[10,11], Patrick Duffy[12,13], Sarah Dunstan[14], Jennifer Evans[2,15], Jeremy Farrar[14], Deepika Fernando[16], Tran Tinh Hien[14], Rolf Horstmann[15], Muntaser Ibrahim[17], Nadira Karunaweera[16], Gilbert Kokwaro[18], Kojo Koram[19], Dominic Kwiatkowski[8,20], Martha Lemnge[21], Julie Makani[22], Kevin Marsh[18], Pascal Michon[3], David Modiano[23], Malcolm E. Molyneux[24], Ivo Mueller[3], Theonest Mutabingwa[12], Michael Parker[25], Norbert Peshu[18], Chris Plowe[26,27], Odile Puijalon[28], Jiannis Ragoussis[20], John Reeder[3], Hugh Reyburn[10,11], Eleanor Riley[10], Jane Rogers[8], Anavaj Sakuntabhai[28], Pratap Singhasivanon[29], Sodiomon Sirima[30], Giorgio Sirugo[6], Adama Tall[31], Terrie Taylor[26,32], Mahamadou Thera[9], Marita Troye-Blomberg[33], Tom Williams[18] & Michael Wilson[19]

Data Fellows Lucas Amenga-Etego[19,34], Tobias O. Apinjoh[1], Edith Bougouma[30], Rajika Dewasurendra[16], Mahamadou Diakite[9], Anthony Enimil[2], Ayman Hussein[17], Deus Ishengoma[21], Muminatou Jallow[6], Enmoore Lin[3], Alioune Ly[31], Valentina D. Mangano[20,23], Alphaxard Manjurano[10,11], Laurens Manning[3], Carolyne M. Ndila[18], Vysaul Nyirongo[24], Tom Oluoch[18], Nguyen T. N. Quyen[14], Prapat Suriyaphol[35] & Ousman Toure[9]

Resource Centre Kirk A. Rockett (Lab Projects Lead)[20], Aaron Vanderwal (Informatics Lead)[20], Taane Clark (Statistics Lead)[8,20], Michael Parker(Ethics Lead)[20,25], Rebecca Wrigley (Network Development Lead)[20], Dominic Kwiatkowski[8,20] (Director), Daniel Alcock[8], Sarah Auburn[8], David Barnwell[20], Susan Bull[20,25], Susana Campino[8], Jantina deVries[20,25], Abier Elzein[17,20], Julie Evans[20], Kathryn Fitzpatrick[20], Anita Ghansah[19,20], Angie Green[20], Lee Hart[20], Eliza Hilton[20], Christina Hubbart[20], Catherine Hughes[20], Anna E. Jeffreys[20], Katja Kivinen[8], Bronwyn MacInnis[8], Magnus Manske[8], Gareth Maslen[8], Marilyn McCreight[20], Alieu Mendy[20], Catherine Moyes[20], Aceme Nyika[8], Claire Potter[20], Paul Risley[7], Kate Rowlands[20], Miguel SanJoaquin[20,24], Kerrin Small[20], Elilan Somaskantharajah[8], Marryat Stevens[20], YikYing Teo[20] & Renee Watson[20]

Project Management Committee Tsiri Agbenyega[2], Dan Carucci[36], Katharine Cook[37], Alan Doyle[37], Ogobara Duombo[9], Jeremy Farrar[14], Michael Gottlieb[36], Kevin Marsh[18], Odile Puijalon[28], Terrie Taylor[26,32] & Dominic Kwiatkowski (Chair)[8,20]

[1]The University of Buea, PO Box 63, Buea, South West Province, Cameroon. [2]Kwame Nkrumah University of Science and Technology, Private Mail Bag, Kumasi, Ghana. [3]Papua New Guinea Institute of Medical Research, PO Box 378, Madang, Papua New Guinea. [4]Swansea Medical School, Swansea University, Singleton Park, Swansea, West Glamorgan SA2 8PP, UK. [5]Institute of Child Health, College of Medicine, University of Ibadan, Ibadan, Nigeria. [6]MRC Laboratories, Atlantic Road, Fajara, PO Box 273, Banjul, Gambia. [7]National Institute for Biological Standards and Control, Blanche Lane, South Mimms, Potters Bar, Hertfordshire EN6 3QG, UK. [8]The Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK. [9]The Malaria Research & Training Centre, University of Bamako, PO Box 1805, Bamako, Mali. [10]London School of Hygiene & Tropical Medicine, Keppel Street, London WC1E 7HT, UK. [11]Joint Malaria Programme, Kilimanjaro Christian Medical Centre, PO Box 3010, Moshi, Tanzania. [12]Genome Science Center, Sokoine University of Agriculture, PO Box 3000, Chuo Kikuu, Morogoro, Tanzania. [13]Seattle Biomedical Research Institute, 307 Westlake Avenue North, Seattle, Washington 98109, USA. [14]Oxford University Clinical Research Unit, The Hospital for Tropical Diseases, 190 Ben Ham Tu, Quan 5, Ho Chi Minh City, Vietnam. [15]Department of Molecular Medicine, Bernhard Nocht Institute for Tropical Medicine, Postfach 30 41 2, D-20324 Hamburg, Germany. [16]Faculty of Medicine, University of Colombo, PO Box 271, Kynsey Road, Colombo 8, Sri Lanka. [17]Institute of Endemic Disease, University of Khartoum, Medical Service Science Campus, PO Box 102, Khartoum, Sudan. [18]Kenya Medical Research Institute (KEMRI)–Wellcome Trust Programme, PO Box 230, Kilifi, Kenya. [19]Noguchi Memorial Institute for Medical Research, University of Ghana, PO Box LG 581, Accra, Ghana. [20]Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK. [21]National Institute for Medical Research, PO Box 9653, Dar es Salaam, Tanzania. [22]Muhimbili University of Health and Allied Sciences, PO Box 65001, Dar es Salaam, Tanzania. [23]University of Rome 'La Sapienza', Piazzale Aldo Moro 5, 00185 Rome, Italy. [24]Malawi–Liverpool–Wellcome Trust Clinical Research Programme, College of Medicine, University of Malawi, PO Box 30096, Chichiri, Blantyre 3, Malawi. [25]The Ethox Centre, Department of Public Health and Primary Health Care, University of Oxford, Badenoch Building, Old Road Campus, Headington, Oxford OX3 7LF, UK. [26]Blantyre Malaria Project, PO Box 32256, Chichiri, Blantyre 3, Malawi. [27]University of Maryland School of Medicine, 655 West Baltimore Street, Baltimore, Maryland 21201, USA. [28]Institut Pasteur, Unité d'Immunologie Moléculaire des Parasites, 28 Rue du Dr Roux, 75724 Paris Cedex 15, France. [29]Faculty of Tropical Medicine, Mahidol University, 420/6 Ratchawithi Road, Ratchathewi, Bangkok 10400, Thailand. [30]Centre National de Recherche et Formation sur le Paludisme, Avenue de l'Oubritenga, BP 2208, Ouagadougou 01, Burkina Faso. [31]Institut Pasteur de Dakar, BP 220 Dakar, Senegal. [32]Michigan State University, Department of Internal Medicine, College of Osteopathic Medicine, East Lansing, Michigan 48825, USA. [33]The Wenner-Gren Institute, Stockholm University, SE-106 91 Stockholm, Sweden. [34]Navrongo Health Research Centre, PO Box 114, Navrongo, Ghana. [35]Faculty of Medicine, Siriraj Hospital, Mahidol University, 2 Prannok road, Siriraj, Bangkoknoi, Bangkok 10700, Thailand. [36]Foundation for the National Institutes of Health, 9650 Rockville Pike, Bethesda, Maryland 20814, USA. [37]The Wellcome Trust, Gibbs Building, 215 Euston Road, London NW1 2BE, UK.