

## **Ethical Challenges of Genomic Epidemiology in Developing Countries**

DAVE A. CHOKSHI & DOMINIC P. KWIATKOWSKI

### **Abstract**

Ethical challenges in genomic epidemiology are the direct result of novel tools used to confront scientific challenges in the field. An orders-of-magnitude increase in scale of genetic data collection has created the need for establishing diffuse international partnerships, sometimes across developed- and developing-world countries, with ramifications for assigning research ownership, distributing intellectual property rights, and encouraging capacity-building. Meanwhile, the fact that genomic epidemiological research is so far upstream in the pipeline of therapy development has implications for the privacy rights of research participants and for a rigorous definition of valid informed consent, particularly in resource-poor settings. From these scientific underpinnings, we distill out two main categories of ethical issues: (1) How should researchers ensure that the subjects of research are appropriately protected? and (2) What is the structure of an equitable and fair system for distributing the financial and scientific rewards of research? We attempt to delineate the contours of specific problems in each category and propose steps toward solutions with reference to a particular project, known as gMap.net, that focuses on genomic epidemiological studies of malaria.

### **Introduction**

Infectious diseases constitute a significant global burden, particularly to those in poverty. Almost six million people die of HIV, malaria, and tuberculosis each year.<sup>1</sup> Simultaneously, a revolution in genomics has created high hopes for important breakthroughs in our understanding of the molecular mechanisms of disease pathogenesis. Several observers have argued that a special effort must be made to ensure that the genomics revolution does not bypass but rather is duly harnessed for the fight against global diseases of poverty.<sup>2,3</sup>

Part of this special effort is ensuring that genomic studies conform to ethical standards of medical research. However, just as the field of genomics is fraught with novel scientific challenges, it is also unique in the ethical challenges that it poses to researchers, particularly those working in developing countries. There exists an abundance of principles for the ethics of biomedical research in developing countries, but few validated methods by which to apply those principles. This article aims at that void, describing the ethical issues raised by genomic epidemiology, with an eye towards specific problems encountered in organizing a genomic database for malaria research ([www.gMap.net](http://www.gMap.net)).

## **What does genomic epidemiology have to offer?**

The sequencing of microbial pathogen genomes provides information on targets for diagnostic tests, mechanisms of virulence, and tactics used by pathogens to evade host defenses. Similarly, it is believed that genomics will yield a better understanding of differential susceptibility and response to infectious diseases in humans as well.<sup>4</sup> For instance, considerable progress has recently been made in identifying a number of gene families which are involved in modifying susceptibility to malaria. Both research strategies – investigating the genome of the pathogen and the genome of the host – promise benefits for the alleviation of global diseases of poverty through the development of drugs and vaccines.

In particular, effective vaccines may provide the best hope of a sustainable reduction in the mortality of HIV, malaria, and tuberculosis.<sup>5</sup> The facts that (1) people who are repeatedly exposed to infection acquire some level of immunity and (2) some people resist infection better than others leads to the hypothesis that it should be possible to develop effective vaccines against these three infectious diseases. The missing link is that we do not yet understand the molecular basis of acquired immunity or natural resistance. Hence, genomics, with its potential for elucidating these processes at the molecular level, holds such hope for vaccine development.

Of course, genomics is no panacea – it will not instantaneously reduce the global burden of infectious diseases. The viruses, parasites, and bacteria respectively responsible for HIV, malaria, and tuberculosis are highly evolved human pathogens skilled at evading the immune system. Nevertheless, the information harvested using genomic methods will enable us to investigate diseases in novel ways. One such method is known as genomic epidemiology. While past genetic studies have been successful at discovering major genetic defects of the immune system – mostly due to a rare mutation of a single gene – genomic epidemiology seeks to uncover genetic variants with a much more modest effect on disease susceptibility.<sup>6</sup> Why are scientists concerned with these weaker genetic effects? There are two main reasons. First, even a modest genetic effect may be of considerable public health importance if it acts on an extremely common disease, and if many different genes make a modest contribution then the overall genetic effect may be huge. Second, even modest association with specific genetic variants may be sufficient to gain insights at the molecular scale of disease pathogenesis, leading to new strategies to treat or prevent the disease.

There are a few essential elements to a genetic epidemiological study. Because the genetic effects of interest are relatively weak, a large sample size – with thousands of affected individuals matched with population controls – is required. For similar reasons, studies must be undertaken across different populations as well. At the other end of the spectrum, high-throughput genotyping technology is required to screen as much of the genome as is possible for each individual. Therefore, innovative algorithms and informatics resources must tackle the analysis of genomic diversity in different populations and fine-scale mapping of genetic association.

## **What is novel about the ethical issues that arise in genomic epidemiological studies?**

To answer this question, it is necessary to step back one level and examine the implications of the novel scientific tools utilized in genomic epidemiology that were described above. Three implications are readily drawn. First, consider the orders-of-magnitude increase in scale of data collection for genomic epidemiology. The ramifications of this increase in scale include a need for diffuse partnerships, which bring with them problems of standardizing ethical review, assigning and sharing intellectual property rights, and dividing research ownership. The second and third implications are manifestations of the fact that genomic epidemiological research is so far upstream in the pipeline of therapy development. Flowing from this is the requirement that consent be obtained for use of patient samples in future research projects, the details of which are unknown at the time of sample collection. Finally, there is a group of challenges surrounding the future social consequences of genomic research. These are particularly important because of the intuition that there is something very personal about an individual's genetic code. Because the potential uses of genomic information cannot be well defined, the potential abuses are similarly nebulous; this is what makes them difficult to prevent.

The nature of this last implication – future social consequences of genomic research – makes it worth going into some detail on the scope of this article. Societal decisions on applications of research, such as whether genetically-modified humans should be permissible, are not addressed here. Rather, we concern ourselves with two main problems: (1) protecting the subjects of genomic epidemiological research and (2) designing an equitable and fair system for allocating financial and scientific rewards of research.

### **Protecting subjects of research**

*Informed Consent.* There are several sets of guidelines on the ethics of research related to healthcare in developing countries.<sup>7,8,9</sup> For example, the principle of informed consent, codified in the Declaration of Helsinki, has been established as a cornerstone of biomedical research ethics.<sup>10</sup> However, complications arise in applying these guidelines to practical situations.<sup>11,12</sup> It is at a specific intersection of these two areas – achieving informed consent for genetics research in developing countries – where the difficulties of implementing established principles are perhaps most evident.

We can distill a number of guiding principles from a review of major reports<sup>13,14,15</sup> and notes from the field:<sup>16,17,18</sup>

- When appropriate, provide information to potential research participants in group meetings. Information should be communicated over a period of time rather than in one meeting.
- The primary source for information about a research project should be health workers rather than physicians.
- Unfamiliar concepts should be explained using analogies.

- Comprehension assessment should be a routine part of the informed consent process.
- Community consent should be respected, and, when appropriate, should be sought out, but is not a sufficient substitute for individual informed consent.

While these principles are helpful, they are only tangentially related to a number of impediments one encounters in genomic epidemiological research. For instance, is it acceptable to ask for open-ended permission to study every gene in a person's body? This question is germane because the great power of genomics is that it allows us to investigate the role of genes whose function we don't currently know. In other cases, guidelines are too vague to provide much help in solving a realistic dilemma. Unfamiliar concepts should be explained using analogies – but how does one even gauge comprehension of concepts such as a gene or DNA? Is it ethical to simply explain, “We are studying attributes that you inherited from your parents?” Finally, some questions raise the idea that perhaps informed consent is not the only ethical principle that is relevant: what are researchers' obligations for samples that were given for genetic studies five years ago, before it was considered feasible to do whole-genome analysis? Do the benefits of research ever outweigh the costs of not following up with donors to obtain re-consent?

In struggling with these issues, gMap.net collaborators are working toward a policy of sensible informed consent. Although the problem of composing appropriate consent forms is very pertinent, the organizing principle of our discussions has been that sensible consent is a process rather than a document.<sup>19</sup> The informed consent process is divided into two discrete stages: education and validation. Education requires both long-term interventions (e.g., working with particular members of a community to germinate an understanding of genetic research in that community) and short-term solutions (e.g., pictorially representing the process of infection by a mosquito). Validation corresponds to each stage of the education process – from checking local language translations to administering an exam testing comprehension. We group the categories of problems that have been encountered in the field using the classical framework of ‘valid’ consent as the union of (1) disclosure and comprehension of information, (2) voluntariness, and (3) competence.<sup>20</sup>

Disclosure and comprehension of information. Language and conceptual comprehension barriers are well-established in the informed consent literature.<sup>21</sup> In genomic epidemiology, one of the most intractable challenges is how to convey genetic concepts. We have proposed guidelines to work with linguists for both local-language translation and for word creation. Word creation involves relating a concept like ‘gene’ to attributes of heredity that are already understood in the local language. To increase understanding, creative didactic methods – such as showing the scale of how much blood is being taken from a child's body – should be employed as far as possible. At the point of communicating why consent itself is required, a clear distinction should be made between research data collection and clinical care. Comprehension assessment of consent should be used to both validate and refine these methods. Finally, the brevity of consent forms and meetings was emphasized as no less important for gMap.net than for other settings.

Voluntariness. The distinction between research data collection and clinical care is important in itself. The predicament here is that, in resource-poor settings, the provision of healthcare in research projects acts as an undue inducement, violating the criterion of voluntariness that is an essential element of informed consent.<sup>22</sup> Given the nature of our data collection – blood for DNA samples is collected during adverse malaria episodes, often from children – it is almost impossible to separate research from treatment. Our guideline therefore is to explicitly accept that the criterion of voluntariness fails in this case because other principles take priority. A corollary is that we must take steps to minimize inducement effects, such as fully ensuring that potential subjects know that healthcare provision is not contingent on participation.

Another set of problems related to voluntariness of genetic research revolves around the question of how far consent applies to future research. At the time that consent is given, the participant has no way of knowing exactly what research projects he or she is consenting to.<sup>23</sup> One idea that has surfaced is to narrow the scope of possible research by a disease-specific constraint – that is, participants who give their consent to a research collaborator are assured that their sample will only be used for malaria-related projects, for instance. This could be combined with a continuous opt-out process by which participants can withdraw from any study at any point in time in the future. One limitation of this approach is the logistical difficulty of continuously communicating what research projects are being conducted using a given person's genetic information.

Competence. There are two particular problems relating to vulnerable populations and competent consent which we must address: (1) At what age is it appropriate for an adolescent to veto a parent's consent (i.e., to give assent)? and (2) How do we ensure that a participant's competence is not undermined by severe illness? Our thinking on obtaining assent from adolescents has evolved from specifying an age threshold to charging the investigator (again, with review from a local ethics board) with examining individual cases based on the criteria of maturity and understanding. To ensure that a patient's competence is not compromised by severe illness, it has been proposed that a two-stage consent process be implemented. The first stage is simply obtaining consent for taking blood when a patient comes in for treatment during a malaria episode. The second stage is following up with the patient (or the patient's parents or guardians, in the case of children) to ensure that permission was indeed granted for the blood to be used in research. That way, the decision to participate in research is not made when the potential donor's faculties may be affected by severe illness.

The final category of problems revolves around the issue of conflicts between social units – for instance, an individual, a family, an ethnic group, or a local government – for granting consent. The only definite principle to have emerged in the literature is that community consent is no substitute for individual informed consent. There is no precise guidance on *when* community consent should be sought out. In our discussions, we proposed that the investigator – with review from the local ethics board – should decide which social units beyond the individual should be consulted. It was also proposed that each social unit consulted has veto power; that is, only if all social units grant consent can it be considered sensible.

*Confidentiality.* Anonymity and confidentiality of research participants are safeguards against the potential harm arising from misuse of genetic information, such as discrimination, stigmatization, and procedures as practical as unwanted paternity testing.<sup>24</sup> It is important to be clear about levels of confidentiality in genetic databases. We follow the structure outlined by the American National Bioethics Advisory Commission (NBAC) in distinguishing the extent to which a research sample can be linked with the identity of its source.<sup>25</sup> “Unidentified” samples are originally collected without identifiers and are impossible to link with their sources. “Unlinked” samples are those that were originally identified, but have been irreversibly stripped of all identifiers and thus impossible to link to their sources. “Coded” samples are unidentified for research purposes, but can be linked to their sources through the use of a code. Decoding is the responsibility of the principal investigator or another designated researcher. “Identified” samples are those that allow the researcher to link the biological information derived from research directly to the individual from whom the material was obtained.

Aside from ethical directives for protecting participants, the source of genetic material and the purpose of research drive the decision regarding what level of identification is appropriate. Again, novel characteristics of genomic epidemiological research shape the ethical discussion. Collaboration in gMap.net occurs between developing-world laboratories largely responsible for collecting DNA samples and developed-world laboratories responsible for high-throughput genotyping of those samples. A pragmatic approach to confidentiality might be to code samples both during sample collection and again during processing of samples received in the developed-world genotyping center. The lab code and field code would not be able to be linked up except by the principal investigator in the genotyping center. Making the link all the way back to the name of the sample contributor requires the willing participation of the principal investigator of the developing-world laboratory from which the sample originated. From this description there is no difficulty in classifying this structure as “coded” confidentiality. However, there are issues that escape the NBAC framework given above. For instance, how much phenotypic information—which is necessary for disease-specific genomic epidemiological studies—can be stored with a “coded” sample before it effectively becomes an “identified” sample?

In some circumstances, there are reasons to have “coded” rather than “unidentified” or “unlinked” samples. The long-term nature of genomic epidemiological research might require that further information—or re-consent—be collected from participants for future studies. Although there should be a high threshold for re-tracing the steps back from a genotyped sample to the contributing individual—and this threshold should be guarded by researchers as well as local ethics committees—the possibility of doing so should not be completely precluded. It is worth mentioning also that despite best intentions and efforts in ensuring confidentiality, modern DNA identification techniques can link a sample with an individual if one wishes to spend the effort and the individual provides a sample for matching.<sup>26</sup> Taken together, these parameters argue for a confidentiality policy that accomplishes the twin goals of protecting participants and facilitating research. It has been suggested that the best way to do this is to adopt a ‘charitable trust’ model where encryption of identifiers is one step removed from researchers themselves.<sup>27</sup>

Finally, there are two issues regarding protection of participants which we have omitted discussion of here. First, we have not addressed the informed consent process for archived samples which may have been taken from subjects who were not made specifically aware that the samples were going to be used for genetic studies. Practical guidelines to regulate these archived specimens—in almost all circumstances they can be used only after they have been unlinked—have been developed and explained elsewhere.<sup>28</sup> Second, we have not given details on database access conditions here because problems of database management will be addressed in the next section.

### **Financial and scientific rewards**

*Data access and intellectual property rights.* Policies governing data access to DNA sequences span a broad spectrum. At one extreme, private firms generate data that are used internally or licensed to pharmaceutical companies for hefty fees. For example, in 1996, the company Human Genome Sciences sold exclusive rights of access to its database of cDNAs for three years to SmithKline Beecham for \$125 million.<sup>29</sup> Private genetic databases, therefore, are effectively unavailable to most potential users because of licensing terms and nondisclosure agreements. While the sequence data in private databases may eventually contribute to therapeutic development, their utility as an academic research tool is necessarily limited. Proponents of these private databases counter that their advantage lies in attracting private capital to help develop practical applications of genetic research. Indeed, selling exclusive rights to database access leads to a neat interface with the normal framework for development in the pharmaceutical sector – intellectual property rights. Companies like SmithKline Beecham are able to mine private databases for ‘patentable’ sequences; the patenting of DNA sequences then provides the incentives necessary for subsequent development.

At the other end of the spectrum are the ‘Bermuda Principles,’ named after the agreement on data release reached by the International Human Genome Sequencing Consortium in Bermuda during 1996.<sup>30</sup> The major government- and nonprofit-funded high-throughput sequencing centers originally agreed to the Bermuda Principles in order to ensure rapid sequence data release. Updates to the Principles called for release of DNA sequence assemblies of 2 kb or greater within 24 hours of generation and of raw shotgun sequence data within one week of generation. It was believed that without such an agreement, the wait for information sufficient to meet patent criteria would lead to long delays and thus be a significant impediment scientifically.<sup>31</sup>

The Human Genome Project also had to address two problems created by its novel data release policy. The first problem was breaches in scientific etiquette involving pre-publication data. Several papers submitted to peer-reviewed journals using data made available by the Human Genome Project did not acknowledge the contribution of the producers of that sequence data. To address this, Project leaders made clear that publications are expected to acknowledge the source of sequence data through the use of appropriate citations; they also urged the broader scientific community to recognize that producers of sequence data have a legitimate interest in publishing their own data.<sup>32</sup> The second problem was operationalizing the belief that the genome should not be patented. Project leaders added an explicit directive in 1997 against

patenting newly discovered DNA with the threat of penalizing researchers in future grant reviews if the directive were not obeyed.<sup>33</sup> It should be noted that the Project chose this route rather than filing noncommercial patent applications to block other claimants or putting restrictions on who could access the public database.

Policies on access to genetic databases that lie intermediate to these two ends of the spectrum have also emerged. The SNP Consortium, funded by a group of 11 private companies collaborating with research centers and the Wellcome Trust, aimed to create a collection of sequence differences in the human genome. Just as with the Human Genome Project, collaborators had a commitment to produce a publicly-available end product—in this case, a human genome map of single nucleotide polymorphisms (SNPs) of a certain density. The members of the Consortium agreed that while the SNP map itself should not be patentable, any innovations made using data from the Consortium was fair game for patenting.<sup>34</sup> The SNP Consortium also employed a unique approach to keeping data in the public domain. The Consortium applied for patents, but only to establish a priority date for the discovery and secure standing as inventors, not to secure commercial patent rights.<sup>35</sup> Such patents are sometimes called ‘blocking patents’ because they serve the purpose of obstructing other claimants’ potential patent filings. Patent applications were abandoned before they were actually issued, making it clear that their sole purpose was to ensure open access to the SNP database.

Another genetic variation resource, the International HapMap Project, also utilized innovative methods to protect open access for its database. Just as with the Human Genome Project and the SNP Consortium, the coordinators of the HapMap Project did not believe that the resource which they sought to catalogue should be patented.<sup>36</sup> However, the HapMap Project employed a distinct strategy to fend off restrictive patents that would hinder their efforts toward open access. This strategy developed as a result of a specific problem related to the nature of the data compiled by the HapMap Project. One can broadly divide this data into two sets: (1) SNPs, SNP assays, allele and genotype frequencies by population, and haplotype information and (2) genotype information in a chromosome region that is of insufficient density to derive haplotype information. HapMap’s data access policy mandates that the former set of data be released to public databases as quickly as possible with no restrictions.<sup>37</sup>

However, the latter set of data required a more complicated access tool. The reason for this is that it would be possible for external parties to combine the public HapMap Project’s genotype with their own, to then construct haplotypes, to file for patents on those haplotypes, and thereby restrict others from using those haplotypes and underlying data.<sup>38</sup> In order to defend against this possibility, genotype data is made available under a ‘click-wrap’ license agreement, which is a type of ‘copy-left’ license.<sup>39</sup> This provision mandates that users are granted access to all data only if they agree not to restrict use of the data by others and to share the data only with others who have agreed with the same condition. When genotype information is of a sufficient density to construct haplotypes, then the individual genotypes and haplotypes are publicly released (both on the HapMap website and to dbSNP, a public SNP database with no licensing restrictions). The HapMap data access policy also specifically mentions that the ‘click-wrap’ licensing approach is not meant to block



downstream patents on haplotypes for which associated phenotypes, such as disease susceptibility, have been discovered.<sup>40</sup>

These trends toward rapid pre-publication release of data from large-scale biological research projects crystallized in a meeting on data sharing sponsored by the Wellcome Trust in January 2003.<sup>41</sup> Meeting participants embraced the spirit of the Bermuda Principles. The report issued as a result of the meeting argues that the benefits of pre-publication release of sequence data have been significant and that therefore rapid, open access should be the standard pursued in other large-scale projects. ‘Community resource projects’ were defined as those devised to create a set of data, reagents, or other material “whose primary utility will be as a resource for the broad scientific community.”<sup>42</sup> The report also specifically addressed the issue of conflicts between pre-publication data release and the norms of publishing the first analysis of one’s own data. One solution brought forth was to create a new type of scientific publication known as a Project Description to inform the scientific community about the project and to provide citations to reference data sources.

It is worth going a bit deeper in examining whether the recommendations of the Wellcome Trust meeting apply to genomic databases involving developing-country partnerships. One can distill out three main issues in such an analysis:

- The importance of the positive-feedback nature of databases, that is, to what extent overall progress of the enterprise depends on rapid release because results build on one another
- Research credit, that is, publications or patents taken out to recognize researchers responsible for particular discoveries
- 3<sup>rd</sup> party development, that is, implications for downstream exploitation of basic science research

Figure 1 (see end of paper) shows how existing database access policies address these three issues. Here, we seek to go beyond that in order to point up difficulties that arise particularly in a developing-world context for genomic epidemiology.

The first issue, the positive-feedback nature of databases, is an important filter – if results do not build on one another in some way, a distributed collaboration does not have to agree to standards of rapid pre-publication data release. It is likely that many genomic epidemiological databases, such as gMap.net, do satisfy the positive-feedback criterion because of the large sample sizes required for studies.

A review of the data access policies of the community resource projects mentioned to this point (the Human Genome Project, the SNP Consortium, and the HapMap Project) reveals that the positive-feedback problem was most likely the dominant factor in the construction of those policies. However, for the developing-world context, the second and third issues, research credit and 3<sup>rd</sup> party development, increase in significance. Research credit (described more fully below in the *Research Ownership* section) is important not just for individual scientists, but for capacity-building aims as well—publications establish the reputation of scientists, which is important in attracting funds to a growing lab. Web publication of data yields neither resume building through publications nor intellectual property rights. Thus,

considerations for research credit have ramifications for exactly how rapid data release should be – and indeed whether it should be ‘pre-publication’ at all.

3<sup>rd</sup> party development, or downstream development of basic science research, is a thornier issue still. There are a number of factors at play here. Some are logistical: how does one assign patents in large collaborations where there is no single firm assignee? Others are more fundamental. The case of genomic epidemiology again breaks from the precedents we have examined in that disease-specific projects like gMap.net will include disease association studies, going beyond map- or sequence-building to the discovery of specific genetic variants that cause resistance or susceptibility to disease. This would seem to move towards the domain of patentable innovations. Of course, database projects must take into account that at some point compiled data must interface with intellectual property regimes, no matter how open-access initial data sharing is. The added challenges for diffuse genomic epidemiological projects are manifold: Which party holds the intellectual property? Where are patent rights taken out? What steps should the network take to ensure that research is translated into therapeutic innovation as quickly as possible? Do researchers have a responsibility in ensuring that sample populations have access to these fruits of research? To whom should the financial rewards, such as royalty revenue, flow?

The most elegant solution to the data access problem seems to be encapsulated in the Bermuda Principles – placing data immediately without restriction into the public domain. This line of reasoning gains force because fundamentally the data yielded by association studies are descriptions of natural phenomena very similar to human genome sequences. However, a cogent argument can also be made in favor of upstream proprietary rights – even for projects that are fully publicly-funded – because the network could thereby have more leverage over what occurs downstream in the process, such as ensuring access, sharing benefits, and ensuring expeditious development. A range of intermediate solutions to these problems may also be emerging as the open-access research movement begins to interface with public or nonprofit drug- and vaccine-development initiatives.<sup>43</sup>

*Benefit sharing.* Here, benefit is an umbrella term for therapies developed by genomic epidemiological research and financial rewards that might be derived from those therapies. International guidelines that speak to how these benefits should be shared with research participants have often incorporated some form of an ‘assured availability agreement’—a guarantee that those exposing themselves to the risks of research be assured access to the products of that research.<sup>44,45</sup> As Bhutta has pointed out, there is considerable complexity in trying to put these principles into practice.<sup>46</sup> In the case of genomic epidemiological research, there is substantial lag time between data collection and the development of therapies. There are connections with the issue of protecting subjects: (1) the promise of benefits might act as an undue inducement to an impoverished population and (2) any obligation to assure availability of an end product to original participants impinges upon the confidentiality of the database. It is also unclear who deserves to gain financially from, for instance, the discovery of a novel anti-malarial molecule from studies of natural genetic diversity. Any of at least five groups can make a claim: the subjects themselves, the health professionals who diagnosed and treated them, the

epidemiologists who managed the study, the geneticists who produced the result, and the company that makes the end product. As Chadwick and Berg have pointed out, while our moral intuitions may sympathize most with the subjects' claim, it is the scientists who have actually made the subjects' samples 'valuable.'<sup>47</sup>

But this leads us to deeper issues which are also at work here. Genomic epidemiology describes the relationship of patterns of disease with natural human diversity. If we assert first that the reference human genome sequence belongs to mankind and second that, given the positive-externality effects of vaccines and therapies for infectious diseases, research is of potential benefit to all, it follows that the aims of benefit-sharing should shift from purely local interests to broader interests.

Nevertheless, companies may have special moral obligations to local participants, and perhaps more broadly to local healthcare systems. Keeping in mind the more expansive view of benefit-sharing, a social return to the community might take the form of technology transfer, local training, provision of healthcare or information infrastructures, or the possible use of a percentage of royalties for humanitarian purposes. For example, the Human Genome Organization Ethics Committee proposed that pharmaceutical industries should set aside a certain proportion of their income for healthcare development or as broad humanitarian assistance for developing countries.<sup>48</sup> Such a set-aside has the added advantage of skirting the difficulties, both logistical and ethical, of tracing back to research participants after a number of years have elapsed.

Another idea that draws on this broader conception of benefit-sharing is the idea of a 'developing country license.'<sup>49</sup> The idea originated in efforts to convince universities to adopt publicly-minded licensing policy. Policies dictated that universities' intellectual property rights would be leveraged to ensure access to essential medicines, such as anti-retrovirals, in resource-poor countries. Similarly, the organizers of a genetic database could mandate the adoption of such 'reach-through' licensing provisions for any research conducted using the group's data. Provisions could ensure access to end products for the broad geographic areas in which original research participants reside. Ethical guidelines developed for the Human Genome Diversity (HGD) Project serve as precedent here.<sup>50</sup> The HGD Project requires that financial benefits from commercial use of samples and of the information derived from them should be in some way returned to the community. Only those researchers who agree to this clause are granted access to the HGD Project's samples and data.

*Research ownership.* Consider the structure of a collaboration for an international malaria genomic database. Laboratories in the developed world must establish partnerships with research groups doing large-scale clinical and epidemiological studies of malaria in developing-world locales. The developed-world labs must develop pipelines for high-throughput genotyping of thousands of DNA samples as well as database technologies to share that data directly with the developing-world research partners. At some point in the future, the database would be sufficiently populated for very large-scale studies across multiple sites (e.g., requiring 10,000 samples from patients with severe malaria) to dissect complex genetic effects. This structure immediately leads to fundamental questions regarding ownership: how is credit distributed among research partners? How are implications for local capacity-

building and career development incorporated into that calculus? What sort of mechanism should be in place to deal with conflicting results?

The Swiss Commission for Research Partnership with Developing Countries contends that the organizing principle here should be capacity-building.<sup>51</sup> It is not that research ownership is important as an end in itself, but rather as a means to garner increased funding or human capital. For that reason, there should be discussion amongst collaborators as to what types of research ownership are most important in developing local capacity—publication in journals, inclusion on grant proposals, or technical training. In the gMap.net group, there are links being formed between the African labs and the Oxford labs via student exchange and periodic meetings. The distributed nature of analysis in gMap.net—genotyping data are made directly available to researchers via the web—facilitates the mission of capacity building. It is hoped that these processes will grow into a governance structure that can handle the more significant challenges of diffused research ownership, such as how to deal with conflicting opinions on scientific findings.

## **Conclusions**

We have attempted to review some of the ethical challenges confronted in doing genomic epidemiological research in developing countries, to describe the guidelines currently in place to help resolve those challenges, and to outline novel elements of genomic epidemiological research that require further ethical analysis. To conclude, we point out some limitations of the perspective which we have adopted in this article. First, because we have divided the issues into discrete categories – for example, issues of informed consent and issues of data access – there has not been an emphasis on the interconnections between categories. For instance, does the level of consent granted set constraints on who is able access a particular database? That is, if the criterion of consent is that it be disease-specific, does that necessitate that data access be given only to those researchers confirming that their interest is in that disease? The logistical difficulties of such an inference are self-evident. Another limitation of our focus is that we have ignored broad areas of ethical complexity in an attempt to make progress on particular problems. Improving local capacity in bioethics in developing countries is essential to ensure that the philosophical principles of genomic ethics are informed by a practical understanding of what will work at the local level.

Figure 1

Data access policies for genetic databases	Private databases	SNP Consortium	HapMap Consortium	Updated Bermuda Principles (Human Genome Project)
Positive-feedback nature of databases: overall progress of enterprise depends on rapid release because results build on one another	- Feedback process hindered by access limitations such as user fees	- Public-private consortium involved collaboration across research entities that involved commitment to publicly available SNP map	- Conditional rapid data release policy to the public via dbSNP. Only genotype information that could be used to file patents restricting the access of others is subject to 'click-wrap' agreement.	- Sequence assemblies larger than 2 kb must be released into public databases within 24 hours; raw shotgun sequences must be released within a week.
Research credit: publications or patents taken out to recognize researchers responsible for particular discoveries	- Not applicable; individual researchers' rewards are financial rather than publication- or patent-related	- Since data was immediately released publicly via a website, intermediate publications were not applicable. Patent applications were taken out solely to establish the dates of scientific discoveries but are abandoned after a period sufficient to prevent others from filing patent applications using Consortium data.	- Since data is immediately released publicly via a website, intermediate publications are not applicable. No patent applications are filed on discoveries.	- Other researchers are permitted to use publicly available data for all purposes aside from "publication of the results of a complete genome sequence assembly or other large-scale analyses." No patent applications are filed on discoveries.
3 <sup>rd</sup> party development: implications for downstream exploitation of basic science research	- Interfaces with intellectual property milieu almost immediately; companies pay for access to private databases in order to mine them for 'patentable' sequences	- SNP Consortium founders believed that the SNP map that would be developed is itself not something that should be patentable. However, there are no restrictions on patents taken out using SNP data as a result of, for instance, association studies.	- The HapMap Consortium believes that SNP, genotype, and haplotype data in the absence of specific utility do not constitute patentable inventions. However, data-release policy does not block users from filing for appropriate intellectual property on, for instance, association studies as long as any ensuing patent is not used to prevent others' access to HapMap data.	- One part of the Bermuda principles (revised in 1997) is an explicit directive to participants against patenting newly discovered DNA. However, there are no conditions placed on users of the GenBank database; nor are blocking patents filed by the Project. As with SNP and HapMap data, there are no restrictions on patents taken out using Project data.

### Acknowledgements

The authors are indebted to Mike Parker for his thoughtful review of multiple drafts of this article.

- <sup>1</sup> World Health Organization. 2002. World Health Report: Reducing Risks, Promoting Healthy Life. Geneva.
- <sup>2</sup> T. Pang and D. Weatherall. Genomics and global health. *BMJ* 2002; 324: 1051-52.
- <sup>3</sup> World Health Organization. 2002. Genomics and World Health. Geneva.
- <sup>4</sup> D. Kwiatkowski. 2002. Genetics. In Manson's Tropical Diseases 21<sup>st</sup> edition. G. Cook and A. Zumla, eds. Philadelphia, PA. W.B. Saunders
- <sup>5</sup> N. Letvin et al. Prospects for Vaccines to Protect Against AIDS, Tuberculosis, and Malaria. *JAMA* 2001; 285(5): 606-11.
- <sup>6</sup> N. Risch. Searching for genetic determinants in the new millennium. *Nature* 2000; 405(6788): 847-56.
- <sup>7</sup> Nuffield Council on Bioethics. 2002. The ethics of research related to healthcare in developing countries. London.
- <sup>8</sup> Council for International Organizations of Medical Sciences. 2002. International Ethical Guidelines for Biomedical Research Involving Human Subjects. Geneva.
- <sup>9</sup> Swiss Commission for Research Partnership with Developing Countries. 1998. Guidelines for research in partnership with developing countries. 11 principles. Geneva.
- <sup>10</sup> World Medical Organization. Declaration of Helsinki. *BMJ* 1996; 313(7070): 1448-1449.
- <sup>11</sup> Z. Bhutta. Ethics in international health research: a perspective from the developing world. *Bulletin of the World Health Organization* 2002; 80(2): 114-120.
- <sup>12</sup> Africans discuss ethics of biomedical research. *Lancet* 2002; 359: 16.
- <sup>13</sup> Nuffield Council on Bioethics, op. cit. 7.
- <sup>14</sup> World Health Organization, op. cit. 3.
- <sup>15</sup> Council for International Organizations of Medical Sciences, op. cit. 8.
- <sup>16</sup> D. Fitzgerald et al. Comprehension during informed consent in a less-developed country. *Lancet* 2002; 360: 1266.
- <sup>17</sup> N. Lynoe et al. Obtaining Informed Consent in Bangladesh. *NEJM* 2001; 344(6): 460-1.
- <sup>18</sup> M. Preziosi et al. Practical Experiences in Obtaining Informed Consent for a Vaccine Trial in Rural Africa. *NEJM* 1997; 336(5): 370-3.
- <sup>19</sup> D. Kwiatkowski. 2004. Global diseases of poverty—what has genomic epidemiology got to offer and what are the ethical challenges? Presented, Workshop on Human Genome Technologies for Clinical and Epidemiological Research on Malaria. Accra.
- <sup>20</sup> T. Beauchamp and J. Childress. 1979. Principles of Biomedical Ethics. New York, NY. Oxford University Press.
- <sup>21</sup> National Bioethics Advisory Commission. 2001. Ethical and Policy Issues in International Research; Clinical Trials in Developing Countries. Vol. 1. Washington, D.C.
- <sup>22</sup> Nuffield Council on Bioethics, op. cit. 7.
- <sup>23</sup> J. Kaye and P. Martin. Safeguards for research using large scale DNA collections. *BMJ* 2000; 321: 1146-49.
- <sup>24</sup> P. Sankar. Genetic privacy. *Ann Rev Med.* 2003; 54: 393.
- <sup>25</sup> National Bioethics Advisory Commission. 1999. Research involving human biological materials: ethical issues and policy guidance. Washington, D.C.
- <sup>26</sup> Z. Lin et al. Genomic Research and Human Subject Privacy. *Science* 2004; 305: 183.
- <sup>27</sup> L. Burnett et al. The 'GeneTrustee': a universal identification system that ensures privacy and confidentiality for human genetic databases. *J Law Med* 2003; 10(4): 506-13.
- <sup>28</sup> G. Sirugo et al. A national DNA bank in The Gambia, West Africa, and genomic research in developing countries. *Nat Genet* 2004; 36(8): 785-6.
- <sup>29</sup> A. Marks and K. Steinberg. The ethics of access to online genetic databases: private or public? *Am J Pharmacogenomics* 2002; 2(3): 207-12..
- <sup>30</sup> National Human Genome Research Institute. 2003. Reaffirmation and Extension of NHGRI Rapid Data Release Policies. <http://www.genome.gov/10506537>. Accessed 14 Feb 2005.
- <sup>31</sup> R. Cook-Deegan and S. McCormack. Patents, Secrecy, and DNA. *Science* 2001; 293(5528): 217.
- <sup>32</sup> T. Powledge. Proposed revision of data-sharing principles recognizes changing times and technologies. *The Scientist*. 11 March 2003.
- <sup>33</sup> E. Marshall. Bermuda Rules: Community Spirit, With Teeth. *Science* 2001; 291(5507): 1192.
- <sup>34</sup> The SNP Consortium: Frequently Asked Questions. <http://snp.cshl.org/about/faq.shtml>. Accessed 14 Feb 2005.
- <sup>35</sup> Cook-Deegan and McCormack, op. cit. 31.
- <sup>36</sup> The International HapMap Consortium. Integrating ethics and science in the International HapMap Project. *Nat Rev Genet* 2004; 5: 467-475.

- <sup>37</sup> International HapMap Project. Data Access Policy for the International HapMap Project. <http://www.hapmap.org/cgi-perl/registration>. Accessed 14 Feb 2005.
- <sup>38</sup> The International HapMap Consortium. The International HapMap Project. *Nature* 2003; 426: 789-796.
- <sup>39</sup> A. Rai. 2004. 'Open-Source' Biomedical Research: Promises and Perils. AEI-Brookings Joint Center Briefing. Washington, D.C.
- <sup>40</sup> International HapMap Project, op. cit. 37.
- <sup>41</sup> The Wellcome Trust. Sharing Data from Large-scale Biological Research Projects: A System of Tripartite Responsibility. Report of Wellcome Trust meeting held on 14-15 January 2003 in Fort Lauderdale, Florida, USA.
- <sup>42</sup> Ibid.
- <sup>43</sup> S. Maurer. Finding Cures for Tropical Diseases: Is Open Source an Answer? *PLoS Med* 2004; 1: 183-186.
- <sup>44</sup> World Medical Association. 2000. Declaration of Helsinki, ethical principles for medical research involving human subjects. Amended by the World Medical Association 52<sup>nd</sup> General Assembly. Edinburgh, Scotland.
- <sup>45</sup> Council for International Organizations of Medical Sciences, op. cit. 8.
- <sup>46</sup> Bhutta, op. cit. 11.
- <sup>47</sup> R. Chadwick and K. Berg. Solidarity and equity: new ethical frameworks for genetic databases. *Nat Rev Genet* 2001; 2: 318-321.
- <sup>48</sup> Human Genome Organization Ethics Committee. Statement on the principled conduct of genetics research. *Law Hum Genome Rev* 1996; Jan-Jun(12): 253-5.
- <sup>49</sup> Y. Benkler. Commons-Based Strategies and the Problems of Patents. *Science* 2004; 305(5687): 1110-1
- <sup>50</sup> North American Regional Committee, Human Genome Diversity Project. Model Ethical Protocol for Collecting DNA Samples. *Houst Law Rev* 1997; 33(5): 1431-74.
- <sup>51</sup> Swiss Commission for Research Partnership with Developing Countries, op. cit. 9.