S

cientists tell us that the human genome is essentially the same in all people and that genetic differences make up about one tenth of a percent of our DNA. Nevertheless, those genetic differences can have a profound impact on health problems encountered by individuals, some of which may be successfully alleviated whereas others continue to elude satisfactory treatment. One could make a similar observation about the ethical problems and dilemmas encountered in genomic research, because although studies of human genomic variation may fundamentally be similar, differences in how studies are carried out can give rise to complex or unique ethical issues, including ones that defy easy resolution. In this article, I present an overview of the basic and some of the specific ethical questions presented by genomic research with a focus on challenges to maintaining privacy and confidentiality in genome-wide association (GWA) studies and sequencing studies. This narrowing of the issues should not be taken as an indication that these are the only challenges that arise in such research. Other issues, such as when and how results of genetic analysis should be reported to individuals, are certainly worthy of attention but beyond the scope of this article. In keeping with this focus, I discuss some of the practices used by researchers for collecting, maintaining, and publishing genomic data and conclude that one radical approach (used by the highly publicized Personal Genome Project (PGP)) should be viewed as a social experiment rather than a model of best practices to be adopted by other projects.

An Initial Challenge: Engaging the Public

Although scientists have recognized genomics as a special field of study for decades, it is only recently that technological advances have made possible the rapid and cost-effective sequencing of whole human genomes. This achievement along with advances in computational capacity has paved the way for detailed explorations of human genetic variation, including GWA studies that aim to discover the relationship between genes and common, complex diseases including cancer, diabetes, and heart disease. These technological developments also prompted the formation of companies offering genome-wide scanning services directly to consumers. Consequently, the public’s engagement with genomics is simultaneously being sought by scientists attempting to uncover genetic contributions to disease and commercial businesses looking to capitalize on the public’s curiosity about their genetic selves.

Generating the level of public participation necessary for either of these enterprises to succeed can be daunting. On the commercial side, sales of genomic scans have been slow, probably because the average American cannot afford the price (which can be as high as $2500) and does not see the value in information that can be hard to understand and has little if any practical utility. To address the affordability problem, one company, 23andMe, recently dropped its rates for personal genome scans down to $399. But the utility problem is not one that the industry can solve on its own or in the short term, dependent as it is on the progress of scientific research. Recognizing this, some companies are forming alliances with researchers to address their mutual interest in accelerating genomic studies. For example, 23andMe will make the database generated through analysis of its customers’ DNA available to researchers hoping that this will spur discovery of practical applications of genomic information whereas a competitor, Navigenics, plans on charging much lower rates for individuals who enroll in research on how consumers respond to and use information generated by personal DNA scans.

Realistically, these commercial/research alliances can only go so far to help investigators in genomic research to get started. Because gene effects may be modest in complex, common diseases, it can take anywhere from 10 000 to 100 000 samples to reach the statistical power necessary to detect relevant genetic variants. Tying research enrollment to the purchase of genome scans (even at discounted rates) alone is not going to achieve those necessary numbers. Nor would such ties necessarily be a scientifically useful strategy for some studies in that the consumer base for such companies probably is not diverse enough to represent the range of genetic variation in the general US population. Furthermore, investigators who aim to understand the interaction between genetically complex diseases and environmental factors often want access not only to samples but to personal data about individuals including medical history, employment history, dietary patterns, exercise habits, and other lifestyle details that will be integrated with genetic data. Thus, genomic research inevitably depends on extremely large numbers of individuals.
from the general population being prospectively enrolled in studies. Meeting these recruitment goals presents obvious practical and logistical problems. Asking individuals to provide much (in terms of access to personal information, if not actual dollars) with little or no prospect of any direct benefit in return can sound presumptuous. The good news for researchers is that surveys indicate the public trusts researchers almost as much as they do their personal physicians insofar as being custodians of private information is concerned, and in general is willing to contribute DNA to publicly funded studies aimed at examining the interaction of genes and environment on health. Nevertheless, this trust should not be taken for granted. It assumes that researchers behave according to established standards of ethical practice as does the medical profession. The reality is, however, that despite a sense of urgency to move this area of research forward, consensus on standards for practice has not been established. Investigators who prospectively collect samples and information from individuals, as well as those who use samples originally collected in other contexts (commercial testing or prior biomedical research) are, therefore, often in the position of making up the rules as they encounter the ethical as well as practical challenges of undertaking GWA or sequencing studies.

**Ethical Groundwork**

As is the case with most applied ethics, it is easier to articulate general principles or norms that should be adhered to, than it is to develop specific or unambiguous rules of conduct consistent with those principles and to adhere to those rules in practice. Consequently, when new areas of research develop it is not unusual for those involved to look to rules and practices applied in similar fields to guide their activities. In the case of GWA studies, researchers naturally turn to the established body of ethics and law governing clinical trials for guidance on fundamental issues, which in this context are informed consent, respect for privacy, and dissemination of results. Unfortunately, differences between these areas of biomedical research work against this being a very effective or satisfactory process. To begin with, genomic studies do not involve administering and monitoring the effects of any medical intervention on human subjects. Instead, genomic research is structured around collection, analysis, and dissemination of information about individuals and populations with whom the researchers may have very limited, if any, direct interaction. This affects the nature of the relationships involved, so that it is a mischaracterization to use the term “subject” both in reference to individuals who provide samples and information for GWA studies and individuals who volunteer for clinical trials.

However, finding a term that more aptly fits the role in GWA studies is not easy. The term “participant” has popular usage in the field, but it can be taken to infer an on going engagement with other participants, including the investigators, or imply that the individual is in a position to influence how the research is carried out. If a study was designed to include this level of active participation, the label would fit the role. However, not all genomic studies will be structured in this way, and therefore the label would not be appropriate for all such studies. Furthermore, secondary uses of data by researchers other than the initial investigators are often intended in genomic research, so that even if the label accurately described the role in the initial study, it would not represent the role in such subsequent studies. Inconsistency and confusion over the labels and characterizations of relationships for this type of research is not limited to the research community. Institutional review boards, oversight agencies, and courts also struggle under, and sometimes compound, this confusion as they sort out questions of rights and responsibilities in this arena. For example, by acknowledging those who provide samples for research as “donors,” and the act of providing those samples as making a gift, courts have established the rights of such individuals in so far as questions of ownership and control over samples are concerned. Unfortunately, this does not resolve questions about their privacy rights or the duties of those who might extract information from the samples and in the absence of clear legal standards on these matters, those questions must be resolved by the application of ethical standards and best practices within the field. To avoid adding to this same confusion, for purposes of this article I refer to individuals who provide samples and personal information for genomic research as “contributors” instead of subjects or participants. Unlike donors, contributors have no special status or significance within law so using that label does not infer any established rights or corresponding duties. This usage also signifies the value of the resources they provide without equating them with sponsors and funders who underwrite research.

Other aspects of genomic studies also distinguish them from clinical trials and make application of standards designed for clinical research (including standards in the federal regulations governing research) problematic. The overarching goal of GWA and sequencing studies is the advancement of general knowledge on the significance of human genetic variation. To serve that end, such studies are designed around the creation and governance of massive databases capable of initially supporting basic genomic research and eventually serving as a resource for more targeted genetic studies if and when associations between gene variants and common diseases are observed. Accordingly, GWA studies typically have no defined end point and are anticipated to last for decades. In contrast, clinical research studies are designed to test hypotheses, are most often focused on a specific disease or health condition, and can be completed within a predictable and limited time frame. The open-endedness of GWA studies affects how the fundamental issues of informed consent, protection of privacy, and dissemination of results are handled and makes adaptation of the rules from one context to the other especially problematic.

**The Illusive Goal of Valid, Informed Consent**

To protect individual autonomy, ethics (and law) require that enrollment in research be preceded by and based on the voluntary, informed and explicitly given consent of subjects. The scope of information to be disclosed in order for consent to be informed is governed by the general principle that it should encompass whatever would be material to the decision of whether or not to volunteer. Based on this principle,
specific disclosure requirements intended to maximize the chances that individuals who enroll in clinical research have a genuine understanding of what is involved have been encoded in the federal regulations referred to as the Common Rule. Not all the disclosures required under those rules are relevant to the circumstances of GWA studies. Those that are relevant include an explanation of the purposes of the research, a description of any reasonably foreseeable risks, and any reasonably expected benefits to the subject, a statement that participation is voluntary, whom to contact for answers to questions about the research, and perhaps most pertinent, “the extent, if any, to which confidentiality of records identifying the subject will be maintained.”

Describing the purposes of the research is not especially burdensome or complicated, given the general nature of the goals of GWA studies. Similarly, the preliminary status of this field of research makes it unlikely that any direct benefits would accrue to contributors, so that disclosures of a lack of anticipated benefits would be easy to convey. The tasks of first determining and then communicating the foreseeable risks are much more difficult. Unlike clinical research where attention to risk is appropriately fixed on risks to the physical and mental well-being of subjects who volunteer to be subjected to some medical intervention, the risks in GWA studies stem from the nature and quantity of information on individuals that will be collected at the outset and then added to as the research progresses.

Although the nature of the risk presented (loss of privacy) is readily apparent, the likelihood of the risk occurring and the severity of the consequences of any such occurrence are not easily determined. The risks also extend into the future in a way that other risks might not. To begin with, a loss of privacy could occur in 2 different ways: through the disclosure of information to a third party in violation of a confidentiality agreement or from an intruder illicitly gaining access to a database containing identifiable information. To guard against the former, researchers can avoid creating unrealistic or unintended expectations regarding confidentiality in the first place. For example, broad statements made during the consent process to the effect that confidentiality will be adhered to at all times, if not accompanied by any further explanation of what that really means, are not effective disclosures in terms of informed consent. For research of this magnitude to take place, a great number of people over a very long period of time will necessarily need access to the personal information that is being supplied. In other words, the notion of maintaining absolute confidentiality does not fit the reality of such information driven research. Nevertheless, when disclosures are being made for consent purposes, it may very well not be possible to identify those who will be granted access to data as the study progresses. A reasonable alternative used by some studies is to describe the qualifications or conditions that would have to be met before access to data would be granted to anyone beyond the initial investigators, and to identify the person or title of the position within the study with responsibility for applying those procedures. For example, under the National Institutes of Health’s (NIH) policy, which was arrived at after notice and an opportunity for public comment, investigators and institutions seeking access to GWA study datasets containing individual genotype and phenotype information must submit requests for access to a Data Access Committee and as a condition of access must agree (among other things), to not attempt to identify individuals within an obtained dataset, to not share elements of the dataset beyond individuals listed in the request, to use the data only for approved research, and to follow appropriate data security protections. When procedures like these are established, disclosed to potential contributors and consistently adhered to, risks to privacy via unconsented to disclosures may be minimized.

However, a further complication for reducing and disclosing risks of privacy arises when results of analyses conducted on personal information, including genomic information are distributed or made available to the broader research community. Under the access policies of some funders (including NIH) to accelerate gene discoveries and maximize the use of resources, investigators are expected to make some data freely and publicly available. The likelihood that such postings threaten privacy, of course depends on the chances that an individual’s identity can be gleaned from scrutinizing publicly posted data. Some data experts contend that removal of apparent identifiers (eg, names, addresses, unique account numbers) although necessary, may not be a sufficient method for deidentifying individual profiles when datasets are replete with details such as phenotypic profiles. Additional strategies that have been used by data managers to further deidentify datasets include use of computational systems designed specifically for that purpose. How successful these strategies are for rendering data sets completely anonymous is a matter of debate. It is important to keep in mind that research would not be possible unless some individuals were willing to assume some risks to privacy. After all, in research, information is collected and created with the intent that it will be distributed, not kept hidden or secret.

For data managers in GWA studies, the challenge is choosing a format for data publication that is sufficiently opaque to protect individual interests and sufficiently transparent to suit the purposes of the intended distribution. Under NIH’s approach to this dilemma, distribution of individual-level data (genotypes and phenotypes) has been restricted to approved researchers, whereas posting of summary-level information and aggregate genotype data on public sites has been allowed. Use of the past tense “allowed” is intentional here. NIH’s policy suddenly changed in August 2008, when it blocked further access to the public databases following the publication of a study on a technique for sorting through mingled samples of DNA to determine whether a particular individual’s DNA were included in the pool. The technique, if it works, has obvious applications in criminal investigations. Whether this new technique could be used or would likely be used by anyone to determine whether a known individual were a contributor to a GWA study is a critical question. Based on initial comments by officials at NIH (and other entities that maintain similar databases), the likelihood of anyone using this technique to violate the privacy of a contributor to research seems “quite low.” This is, in part, because the person applying the technique would have to already have a genetic profile from a known individual.
Moreover, as a representative from NIH has explained, “the technology to obtain the required genomic profile is not commonly used outside of the research community. And, even if an individual’s SNP profile was found within a pooled dataset, all that would be learned is that this profile was contained in the dataset and, thus, it could then be associated with the characteristics of that dataset (eg, disease or control population).” Remote as this possibility may be, where public confidence in researchers commitments to protecting the privacy of those who contribute their DNA and personal data are concerned, some institutions prefer to take a cautious approach rather than lose the public’s trust by appearing to be cavalier about the privacy of those who contribute to the research. Others take a much bolder course in resolving the issues of who has responsibility for assessing or assuming risks to privacy in the context of GWA studies. And at the same time that NIH is reining access to aggregate data in, a project at Harvard is opening up access to individually identified data to demonstrate that our presumptions and worries over the need to protect the privacy of health information are unfounded and in the process it has transformed the role of contributors in problematic ways.

The PGP: Taking Assumption of the Risk to an Extreme

On October 20, 2008, a press conference was held at Harvard Medical School to announce the progress of the PGP, a GWA study that had been underway for about a year. The newsworthiness of the project is based on the incorporation of a rather unusual aim for a GWA study. To recruit individuals interested in obtaining and openly sharing their genome sequences, related health and physical information, and reporting their experiences as a participant of the project on an ongoing basis. What makes this goal and the role of volunteers in this study unique is that “openly sharing” means sharing personal information with the general public, not just within an expanded research community. But this study actually goes beyond this aspect of participation to other areas that make this study difficult to categorize in relation to volunteers. The additional problems can be seen by a careful examination of its enrollment process and the expectations that are placed on volunteers.

Individuals interested in taking part in the PGP are asked to first familiarize themselves with different aspects of the project, including the risks and potential benefits as described in project materials. If they are still interested, they then are asked to demonstrate their comprehension of concepts relevant to the project, including basic genetics, by completing an entrance examination, which serves as a screening tool for eligibility to participate. Those who are determined to be eligible and who are willing then complete an online consent and the process of collecting information from the individual begins. However, individuals are not fully enrolled in the project until the issue of fees has been addressed. According to information posted on the PGP website, a standard fee of $1000 is charged to subsidize the costs of research activities, including genome sequencing. Once fees have been paid volunteers are interviewed for purposes of verifying their understanding of the project and their consent to participation. After that verification, DNA is collected from the individual and genome analysis begins. As data are generated from this analysis it is added to files accessible by that individual.

What all this detail reveals is that the role of being a volunteer in the PGP goes beyond being a mere contributor or even participant. In designing the study, the investigators have significantly transformed the role of volunteer from the traditional notion of a research subject. What it ends up being instead is difficult to characterize. The role is not wholly that of a funder, yet it involves underwriting research. It is not simply that of a consumer, yet it has aspects of purchasing a service (genome sequencing). The investigators view the role more as that of “research collaborators.” This is particularly interesting because it signals that the status of the volunteer has been transformed into that of a researcher. If that is the case, it leaves us to wonder what has happened to the rights and duties associated with the subject/researcher relationship. Have they been dissolved in the role transformation process, so that what is left is simply two equal parties in a joint venture that is governed by the terms of a contractual agreement? That could very well be what the study designers intended. It seems that the investigators do not see themselves as having any obligations to protect the volunteers other than to be honest with them in disclosing the terms of the venture. Nor do the investigators seem to accept any responsibilities for what comes of publishing personal data in an obviously identifiable format. As the lead investigator, George Church put it, “It’s not the personal genome project that’s releasing the data, it’s the individuals releasing their own data, which is perfectly within their rights. We just want to make sure they’re well informed before making this decision.”

As of October 20, 2008, 10 volunteers had enrolled in the PGP and were about to release their data via public postings on the PGP website. If and how this has an impact on their lives remains to be seen. In interviews several of them indicated that they were unworried about any negative consequences it might have for them; nevertheless, a Data Safety Monitoring Board will send an email to each of them (and subsequent enrollees) every 3 months asking if they have had any adverse events as a result of participation. However, what might constitute an “adverse event” related to publicly posting one’s medical history, and genomic information is not spelled out on the PGP website. Things like discriminatory actions taken by an employer or insurer based on the published information would undoubtedly qualify as adverse events, but making a causal link between public postings of information and any such economic losses could be difficult. Nevertheless, if reports of adverse events should be received, there is the additional unanswered question of what actions the PGP would take in response. Given the statement above that it is not the PGP but rather the individual who releases identified data, the investigators seem to think that the individuals assume any and all risks when they enroll in the project. This is a radical departure from the traditional rules governing the protection of subjects in research. Consequently, the PGP should be seen as a loosely structured social experiment within a GWA study and should not be
adopted as a proven model of best practices for GWA studies. In reaching this conclusion, I am inclined to agree with Francis Collins’s assessment that although it is an “interesting and thought provoking”26 model, asking people to give up any shred of privacy is neither the best way to get the public more comfortable with genomics nor is it the best way to nurture the public’s trust in researchers.

### Conclusion

The tension between maintaining confidentiality and permitting wide access to data is much greater in GWA studies than clinical research, and different studies have taken different approaches to reduce this tension. On one end of the spectrum is a highly conservative approach exemplified by the NIH polices governing access to aggregate data, where the tension is resolved in favor of protecting contributors’ interests and preserving the public’s trust. On the other end is a riskier approach exemplified by the PGP that resolves the tension in favor of wide access and moving genomic science forward into genomics. Neither individual nor the public interests are well served when the risks to privacy encountered in genomic research are either overly exaggerated or completely shrugged off. The best approach lies somewhere in between these extremes. It acknowledges that research is not ever entirely risk free and researchers bear responsibilities toward volunteers, including in relation to the risks involved. In GWA studies in particular, researchers are in a better position than contributors to assess the scope of the risks involved and are better able to take steps to minimize the risks. Fulfilling such responsibilities requires expertise in the areas of data management and security as well as genomic science and a commitment to good ethical practice.

### Disclosures

None.

### References

13. 45 CFR §46.116(a) (5).

**Key Words:** genetics ■ ethics ■ research
Ethical Challenges Encountered in Genomic Research
Patricia A. Roche

doi: 10.1161/CIRCGENETICS.108.846758
Circulation: Cardiovascular Genetics is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2009 American Heart Association, Inc. All rights reserved.
Print ISSN: 1942-325X. Online ISSN: 1942-3268

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circgenetics.ahajournals.org/content/2/3/293

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation: Cardiovascular Genetics can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation: Cardiovascular Genetics is online at:
http://circgenetics.ahajournals.org//subscriptions/