Human Molecular Genetics, 2006, Vol. 15, Review Issue 1 **R45–R49**
doi:10.1093/hmg/ddl049

Ethical issues in medical-sequencing research: implications of genotype–phenotype studies for individuals and populations

Morris W. Foster[1,*] and Richard R. Sharp[2]

[1]Department of Anthropology, University of Oklahoma, Norman, OK 73019, USA and [2]Center for Medical Ethics and Health Policy, Baylor College of Medicine, Houston, TX 77030, USA

Received January 30, 2006; Revised and Accepted March 3, 2006

Advances and declining costs in sequencing technology will result in increasing number of studies with individual sequence data linked to phenotypic information, which has been dubbed medical sequencing. At least some of this linked information will be publicly available. Medical sequencing raises ethical issues for both individuals and populations, including data release and identifiability, adequacy of consent, reporting research results, stereotyping and stigmatization, inclusion and differential benefit and culturally and community-specific concerns. Those issues are reviewed, along with possible solutions to them.

Studies of genotype–phenotype associations will play an increasingly prominent role in health research as emerging technologies continue to reduce sequencing costs and advances in bioinformatics address challenges of data interpretation (1). Mapping resources such as the dbSNP and HapMap databases will allow researchers to design studies examining relationships among detailed genomic sequence data and linked phenotypic information in a manner not previously possible due to practical and methodological constraints. In the last year, planning has begun for several such studies, including a project led by the National Human Genome Research Institute that would sequence the genomes of participants in a prospective cohort study (2), and another initiated by the National Cancer Institute that would sequence hundreds of tumor samples in common cancer types (3).

Studies in which researchers generate large amounts of genotypic and phenotypic information that can be linked to individual study participants (hereafter, 'medical-sequencing research') raise a number of ethical issues. We examine these issues and discuss possible solutions to the challenges they present.

INDIVIDUAL ISSUES

Data release and identifiability

Agencies that fund large genomic initiatives have tended to treat the data these projects produce as a community resource to be made publicly available before thorough analysis by the consortia that generate them (4). This approach insures that the benefits of large investments in genomics are fairly distributed among all potential users. Open access to genomic data also may be the best answer to lingering concerns about the potential for intellectual property claims and commercial patents to constrain future research, although public release complicates the question of whether individuals have any property rights to their own genomic sequences (5).

Although genetic data released into publicly accessible databases are not directly linked to any particular individual, these databases typically establish safeguards to further reduce the likelihood that any individual research subject might be identified. Users of the HapMap database, for example, can access individual genotypes but do not have access to phenotypic or demographic information about sample donors (with the exception of the geographical region of origin) (6).

Medical-sequencing databases, though, can include demographic information, clinical information (or portions of an individual's medical records), exposure and employment histories and family pedigrees among other linked data. As recently demonstrated in a case involving the identification of an anonymous sperm donor, in combination with other publicly available information, limited amounts of genetic data alone can uniquely identify an individual (7). Researchers also have demonstrated that relatively little phenotypic information

*To whom correspondence should be addressed at: Department of Anthropology, 455 W. Lindsey, Rm. 505C, University of Oklahoma, Norman, OK 73019, USA. Tel: +1 4053252491; Fax: +1 4053216936; Email: morris.w.foster-1@ou.edu

linked with genetic information is required to identify an individual (8,9).

Individual identifiability is of particular concern in medical-sequencing research because of the power that genotype–phenotype information may have for revealing future health status and disease susceptibility. Such information may be sought by a variety of third parties, including employers, insurers, courts of law and family members (10). As these third parties may employ that information in ways that are counter to the interests of individual research participants, it is imperative that researchers guard against such unintended uses.

Various protections could be used to limit individual identifiability in medical-sequencing research. For example, detailed sequence data could be made publicly available, but might be linked publicly with limited phenotypic information. Investigators interested in obtaining more detailed phenotypic data to examine a specific research question might then be required to submit a proposal to an oversight committee and obtain an Institutional Review Board (IRB) approval for access to particular phenotypic data fields. Such an approach would allow the oversight committee and relevant IRBs to consider possibilities of subject identification within the context of specific uses and develop appropriate protectionist strategies as new technologies emerge and social attitudes regarding the risks of identifiability change over time. Additional policies might be established to limit deliberate attempts to identify or disclose the identities of individual research participants in research reports.

Adequacy of consent

The possibility of medical sequencing of individual genomes <3 years after the completion of the human genome sequence demonstrates the speed with which new technologies and approaches for analyzing the data they generate are being developed. The pace of these developments raises concerns about whether it is possible for a potential study participant who is not a geneticist (or perhaps even one who is) to appreciate the scope of future applications of the information and biological materials he or she provides, particularly when some of those future uses may employ as yet undeveloped technologies and analytic approaches (11). Because strong international consensus supports the need to obtain informed consent from research subjects prior to conducting research (12), the inability to anticipate the types of benefits and risks associated with future research using donated biological materials raises major ethical worries.

There are several possible strategies for addressing this challenge. The first is to discuss with prospective subjects how medical-sequencing research could pose risks to individuals that may be unforeseeable at present. This approach is taken by the Personal Genome Project, which will make select genotype–phenotype data publicly available (13). The second is to incorporate more extensive educational counseling into the consenting process. For example, attendance at extended educational sessions might be required of participants, as well as some demonstrated level of understanding before an invitation to participate in medical-sequencing research is extended. The third is to involve participants in the ongoing management of project data (14). This involvement may be through a proxy, such as a community advisory board or through periodic renewals of participant consent (15,16). These approaches are not mutually exclusive, and it is likely that none is sufficient alone.

Reporting research results

By its very nature, a fully or partially sequenced individual genome can reveal information about genetically based or -contributed characteristics that is unknown to the participant. An individual sequence may contain information that shows variants for disease susceptibilities or drug responses already identified from other studies, such as the APOE variant that is associated with Alzheimer's, and also may have information about variants that are not yet identified but one day will be. What are the responsibilities of researchers, their sponsoring institutions and funders to inform donors of these findings? Corollary questions include obligations to provide clinically certified analyses, to provide counseling or care, to notify other biological kin who also may be affected and to have open-ended obligations for new genetic discoveries. As we learn more about environmental contributors to disease, similar questions may arise about responsibilities with respect to predictive phenotypic information.

Although 'honest brokers' can be used to limit researcher access to participant contact information and participants can be told up front that information about findings (anticipated or unanticipated) will not be provided to them, these arrangements do not resolve the underlying moral questions about researchers' responsibilities toward participants. Moreover, the potential for identifying one's individual sequence or that of someone else on a public database and finding known or suspected variants within it will be increasingly within the reach of the lay public including employers, insurers, family members and other potentially interested parties.

POPULATION ISSUES

Stereotyping and stigmatization

We know that members of some populations are more likely to be affected by some diseases than members of other populations. We also know that many polymorphisms vary in frequency between populations, although most are present at some frequency in all populations. Medical-sequencing projects will add many more examples of ways in which health-related variants differ in frequencies between populations, in addition to differences in variants related to other complex traits and in haplotype frequencies and length, extents of linkage disequilibrium and instances of natural selection.

There is the potential for these and other genomic differences, which are based primarily on variations in frequency, nonetheless to be treated as reductionary, mutually exclusive biological definitions or stereotypes of populations (17). Where higher or lower frequencies may contribute to health disparities, particularly those that are viewed in adverse ways by some, these genetic stereotypes also may result in

stigmatization for all population members whether they posses the genetic feature at issue or not (18).

Stereotyping can be particularly harmful when population identities are used as exclusionary criteria rather than as proxies for likelihood in clinical diagnosis and treatment. Those clinical applications often are reinforced by researchers' common practice of uncritically using racial, ethnic and other identities as though genetic findings apply to all members of a named population (19). More nuanced, critical use of population identities in genetic research, especially in communications with the media and the public, can reduce the potential for stereotyping, although any popularization of research findings will entail some simplification.

The problem of stigmatization is somewhat different, involving little, if any, contribution from the scientific community and arising almost entirely from pre-existing public attitudes toward specific health statuses and populations (20). As such, there is less that researchers can do about stigmatization aside from being active in appropriately framing findings about health statuses and populations that historically have experienced stigma and discrimination, and so are likely to be targets in future.

Inclusion and differential benefit

For the most part, the funders of large-scale genomics projects such as HapMap and the SNP Consortium have made the assumption that the discovery of common variants and patterns will benefit all populations (21). The medical-sequencing projects proposed by NIH maintain the assumption that discoveries in some populations (however defined) will benefit members of other populations not only with respect to common inherited variants, but also in the form of common pathways the modification of which by a variety of environmental contributors can result in differential disease susceptibility and drug response. Thus, planning so far has focussed on recruiting racially and ethnically diverse samples to discover common features. Although the seemingly broad utility of the HapMap data (which were generated from four specific populations) provide good support for a focus on genomic patterns that are common to all populations (22), the much greater levels of investment required by medical-sequencing projects render the assumption of common applicability a much bigger gamble.

The alternative to the assumption that common variants, patterns and pathways will capture most of the effects of genetic and environmental contributors to common diseases (although perhaps not most of the specific environmental contributors themselves, which are likely to vary by locality, occupation, social status, cultural practices and other criteria) is to target sampling from particular populations to identify contributors and pathways that may be specific or occur in higher frequencies among their members, which would emphasize certain genetic and environmental contributors to the exclusion of others (23).

If there are a small number of variants and pathways common to most lung cancers, for example, then both the targeted and the diverse designs will benefit most members of most populations. However, if some variants and pathways are more likely to be discovered in certain populations than others (for example, among smokers or people with particular ancestries), then a design that targets those populations may be necessary to ensure that their members share equally in the benefits of medical sequencing. These are not necessarily mutually exclusive approaches, and it is likely in the latter case that some variants and pathways can be identified through a diverse design, whereas others will require targeted populations. Obviously, that combination would be more expensive and involve more ethical considerations about the distribution of potential benefits than a broadly diverse design with common benefits alone.

In addition, the investment of significant resources in large-scale medical-sequencing projects will have consequences both for downstream research opportunities and for investments in other research projects. With respect to the former, those populations and phenotypes that are investigated more comprehensively in medical-sequencing projects are more likely to benefit from subsequent discoveries and studies based on those rich data. With respect to the latter, the still-significant costs of sequencing will preclude the funding of some other studies that would provide somewhat different benefits (24). Finally, sequencing can be used to identify alterations that result from both inherited as well as environmental contributors, whereas the data generated are more likely to be used for genetically targeted diagnostics and therapeutics rather than in developing behavioral and environmental interventions to promote health through prevention (25).

Community and culturally specific issues

Community consultation in planning biomedical studies increasingly is common in research involving indigenous and minority groups, largely due to their historical experiences of research abuses and continuing economic and political vulnerability for stereotyping and stigmatization (26). In addition, where targeted populations comprise organized cultural and/or political entities (such as American Indian tribes), consultation demonstrates respect for the moral authority of those communities (27). Most populations, however, are both too large and too diverse to comprise a single moral community. Who, for example, would be consulted for a study targeting African Americans or Hispanic Americans? Nonetheless, it may be useful to learn how some African Americans and Hispanic Americans view biomedical research in general and medical sequencing in particular. The approach used by the HapMap Project, for example, was to consult in those localities from which donors would be recruited later, which allowed researchers to identify some culturally specific concerns and perspectives about genetic variation that may not have been anticipated otherwise and to use that information to better educate potential participants about the project as well as to better frame population-specific genetic findings to reduce stereotyping and stigmatization (28).

Community consultation or engagement, however, should not be mistaken for community consent nor does every medical-sequencing project or every population warrant advance consultation. A diverse cohort that includes participants from multiple racial and ethnic populations and has the goal of identifying common features across populations, for example, presents fewer recruiting and reporting challenges than a study that proposes to recruit and sequence a

population-specific cohort. Moreover, community consultation is primarily a localized process that is not a substitute for broader regional, national and international policy forums for evaluating investments in large-scale medical-sequencing projects, although findings from specific community engagements can inform broader policy discussions.

Among the public policy topics that genotype–phenotype studies might raise are the appropriateness of using racial, ethnic and other social identities to analyze the data (especially in studies of complex traits such as intelligence), the emphasis on developing genetic therapies rather than largely non-genetic (and lower cost) prevention strategies, the differential benefits of diverse and targeted designs and which phenotypes to study (given limited research resources). Medical sequencing also may raise questions from members of particular cultural or religious groups about the implications of more detailed genetic and phenotypic information for pre-natal screening and decision-making, reproductive assistance and selection and genetic therapy or enhancement, among other potentially controversial areas.

CONCLUSION

The challenges of using linked genotype–phenotype data for medical-sequencing projects prefigure issues that will arise in future uses of many existing biological samples linked to phenotypic information, including disease registries, hospital-based tissue collections and prospective cohort studies. Thus, developing effective strategies for addressing these challenges in medical-sequencing research can inform a much broader set of issues in the ethical conduct of research.

ACKNOWLEDGEMENTS

This publication was made possible by grant numbers CA114626-01 from the National Cancer Institute, ES11174 from the National Institute on Environmental Health Sciences and HG02691, HG03042 and HG03083 from the National Human Genome Research Institute. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NCI, NIEHS, NHGRI or the National Institutes of Health.

Conflict of Interest statement. None declared.

REFERENCES

1. Chan, E.Y. (2005) Advances in sequencing technology. *Mut. Res. Fundam. Mol. Mech. Mutagen.*, **573**, 13–40.
2. Collins, F.S. (2004) The case for a US prospective cohort study of genes and environment. *Nature*, **429**, 475–477.
3. Bonetta, L. (2005) Going on a cancer gene hunt. *Cell*, **123**, 735–737.
4. Wellcome Trust (2003) *Sharing Data from Large-scale Biological Research Projects: A System of Tripartite Responsibility*. Wellcome Trust, http://www.wellcome.ac.uk/assets/wtd003207.pdf.
5. Robertson, J.A. (2003). The $1000 genome: ethical and legal issues in whole genome sequencing of individuals. *Am. J. Bioethics*, **3**, W35–W42.
6. International HapMap Consortium (2003) The International HapMap Project. *Nature*, **426**, 789–796.
7. Motluk, A. (2005) Anonymous sperm donor traced on internet. *New Scient.*, **2524**, 6.
8. Malin, B., Sweeney L. (2004) How (not) to protect genomic data privacy in a distributed network: using trail re-identification to evaluate and design anonymity protection systems. *J Biomed. Inform.*, **37**, 179–192.
9. Malin, B.A. (2005) An evaluation of the current state of genomic data privacy protection technology and a roadmap for the future. *J. Am. Med. Inform. Assoc.*, **12**, 28–34.
10. Clayton, E.W. (2003) Ethical, legal, and social implications of genomic medicine. *New Engl. J. Med.*, **349**, 562–569.
11. Rothstein, M.A. (2005) Expanding the ethical analysis of biobanks. *J. Law Med. Ethics*, **33**, 89–101.
12. Brody, B.A., McCullough, L.B. and Sharp, R.R. (2005) Consensus and controversy in clinical research ethics. *J. Am. Med. Assoc.*, **294**, 1411–1414.
13. Church, G.M. (2005) The Personal Genome Project. *Mol. Syst. Biol.*, doi:10.1038/msb4100040, http://www.nature.com/msb/journal/vl/nl/full/msb4100040.html.
14. Eskenazi, B., Gladstone, E.A., Berkowitz, G., Drew, C.H., Faustman, E.M., Holland, N.T., Lanphear, B., Meisel, S.J., Perera, F.P., Rauh, V.A. *et al.* (2005) Methodologic and logistic issues in conducting longitudinal birth cohort studies: Lessons learned from the Centers for Children's Environmental Health and Disease Prevention Research. *Environ. Health Perspect.*, **113**, 1419–1429.
15. Strauss, R.P., Sengupta, S., Quinn, S.C., Goeppinger, J., Spaulding, C., Kegeles, S.M. and Millett, G. (2001) The role of community advisory boards: involving communities in the informed consent process. *Am. J. Pub. Health*, **91**, 1938–1943.
16. Wendler, D., Prasad, K. and Wilfond, B. (2002) Does the current consent process minimize the risks of genetics research? *Am. J. Med. Genet.*, **113**, 258–262.
17. Foster, M.W. and Sharp, R.R. (2004) Beyond race: towards a whole genome perspective on human populations and genetic variation. *Nat. Rev. Genet.*, **5**, 790–796.
18. Foster, M.W. and Sharp, R.R. (2000) Genetic research and culturally specific risks: one size does not fit all. *Trends Genet.*, **16**, 93–95.
19. Sankar, P. and Cho, M.K. (2002) Toward a new vocabulary of human genetic variation. *Science*, **298**, 1337–1338.
20. Condit, C. (2004) Science reporting to the public: does the message get twisted? *Can. Med. Assoc. J.*, **170**, 1415–1416.
21. Foster, M.W. and Sharp, R.R. (2005) Will investments in biobanks, prospective cohorts, and markers of common patterns of variation benefit other populations for drug response and disease susceptibility gene discovery? *Pharmacogenomics J.*, **5**, 75–80.
22. International HapMap Consortium. (2005) A haplotype map of the human genome. *Nature*, **437**, 1299–1320.
23. Foster, M.W. and Sharp, R.R. (2005) Will investments in large-scale prospective cohorts and biobanks limit our ability to discover weaker, less common genetic and environmental contributors to complex diseases? *Environ. Health Perspect.*, **113**, 119–122.
24. Garber, K. (2005) Human cancer genome project moving forward despite some doubts in community. *J. Natl Cancer Inst.*, **97**, 1322–1324.
25. Sze, J. and Prakash, S. (2004) Human genetics, environment, and communities of color: ethical and social implications. *Environ. Health Perspect.*, **112**, 740–745.
26. Dickert, N. and Sugarman, J. (2005) Ethical goals of community consultation in research. *Am. J. Pub. Health*, **95**, 1123–1127.
27. Sharp, R.R. and Foster, M.W. (2002) Community involvement in the ethical review of genetic research: lessons from American Indian and Alaska Native populations. *Environ. Health Perspect.*, **110**, 145–148.
28. International HapMap Consortium (2004) Integrating ethics and science in the International HapMap Project. *Nat. Rev. Genet.*, **5**, 467–475.

APPENDIX

Recommendations for best practices

Identifiability

- Weigh the scientific value of rapid public release of sequence information against the increasing potential for using those data to identify individual participants.

- Limit access to linked phenotypic data that may make some participants more identifiable.

Adequacy of consent

- Explicitly highlight and discuss the likelihood of unforeseeable risks with participants.
- Incorporate more extensive educational counseling into the consenting process.
- Involve participants in the ongoing management of project data.

Reporting research results

- Using honest brokers does not completely insulate researchers from ongoing moral obligations toward sample donors.
- This is a difficult issue, complicated by decreasing sequencing costs, that requires careful attention from researchers and bioethicists to develop consensus best practices.

Stereotyping and stigmatization

- Researchers should be more sophisticated in using population identities, particularly in media contacts.
- Researchers should carefully frame findings about health statuses and populations with histories of stigma and discrimination.

Inclusion and differential benefit

- Choices about which phenotypes and populations to sequence can have consequences for those that are not targeted, depending on the extent to which contributors and pathways are common across different phenotypes and populations.

Community and culturally specific issues

- Community consultation should be done when organized cultural and/or political entities are specifically targeted, can inform ethical evaluations about how some members of other populations view medical sequencing, but is less critical when studies include participants from diverse populations and focus on features common across populations.
- Community consultation should not be substituted for broader policy discussions about implications for larger racial, ethnic, cultural and other populations.