Louisiana Law Review

Volume 66 | Number 5
Special Issue
Symposium: Proceedings of "The Genomics Revolution? Science, Law and Policy"

Population Participation and Other Factors that Impact the Compilation and the Utility of Resulting Databases

Henry T. Greely

Population Participation and Other Factors that Impact the Compilation and the Utility of Resulting Databases

Henry T. Greely[*]

Let me begin by briefly addressing the social, ethical, and legal issues with respect to biobanks generally, and then the special considerations, the special problems, that arise from biobanking and particular human populations. It is important to remember that for almost every purpose, and particularly for purposes involving medical use, biobanks have to be more than just collections of physical specimens. DNA is not very useful for medical purposes unless you know something about the person's health history. You cannot find the gene for blue eyes from DNA unless you know whether people who gave you the DNA had blue eyes.

So, to be useful, biobanks really are more than just biological banks. They are combinations of physical samples of DNA—which may be tumor tissues or something else—plus medical information about a person. I have tried calling them phenotype-genotype resources, an extremely clunky term that has not caught on, but it has the advantage of being accurate. They are resources that have both physical samples of genetic material and have information about each donor's phenotype; genotype is what your genes say about you, and phenotype is your body as it exists. So a biobank needs both phenotype information and genotype information to be really useful.

Now, one important thing to know about these large genotype-phenotype resources is that their value is promising but speculative. No one really knows whether we are going to make great discoveries as a result of having 200,000 samples of DNA coupled with health information about people. It seems like a good way forward, in part because the traditional way of doing genetic research in health seems to have stalled to some extent. As we

---

[*] J.D., Professor, Stanford Law School. The author thanks Daven Williams of LSU Law School for her work with the references. This work is based on a live presentation made on February 5, 2004, and does not necessarily reflect events and changes thereafter.

have gotten past the simple genetic diseases, the ones that are caused by a single gene, the so-called simple Mendelian disorders, we have discovered, first, that many of those simple Mendelian disorders were not as simple as we thought and, second, that there are known genetic connections to much more common and significant diseases, but those connections are turning out to be really hard to tease out.

We know that there are genetic connections to schizophrenia, which affects one percent of the population,[1] to diabetes, which affects six-to-eight percent of the population,[2] to various cancers, to high blood pressure,[3] to Alzheimer's Disease,[4] to lots of diseases that affect lots of people. But finding out how the genes connect to those diseases has turned out to be really hard, and it is hoped that collecting huge amounts of data from large numbers of people will give us a way to better make those connections. At least that is the hope, and that hope may prove true.

There are two things we know for sure about such genotype-phenotype resources though. The first is that they are really expensive. The Iceland Project, if it ever actually happens in its entirety, is likely to cost several billion dollars to create a

---

1. Julian Guthrie, *A Family's Journey to Madness and Back: Son's Schizophrenia Spurs Parents to Raise Millions for Research*, San Fransisco Chronicle, Sept. 7, 2004, at A1, *available at* 2004 WLNR 7643201; *Common Genetic Influence Found in Mental Illnesses-Schizophrenia*, Genomics and Genetics Weekly, Feb. 13, 2004, at 69, *available at* 2004 WLNR 936138; Demosthenes A. Lorandos, *Secrecy and Genetics in Adoption Law and Practice*, 27 Loy. U. Chi. L.J. 277, 284 (1996); Lori Andrews and Erin Shaughnessy Zuiker, *Ethical, Legal, and Social Issues in Genetic Testing for Complex Genetic Diseases*, 37 Val. U. L. Rev. 793, 799 (2003).

2. Eric G. Stark, *Diabetes Carries a Costly Toll*, Sunday News, Lancaster, PA, Aug. 1, 2004, at 1, *available at* 2004 WLNR 11488523; Andrienne Forman, *The Second National Conference on Diabesity in America*, Nutrition Today, Nov. 1, 2004, at 245, *available at* 2004 WLNR 15747171.

3. Chris Winkelman, *Genomics: What Every Critical Care Nurse Needs to Know About the Genetic Contribution to Critical Illness*, 24 Critical Care Nurse 34, June 1, 2004, *available at* 2004 WLNR 5689392.

4. Laura McConnell et al., *Genetic Testing and Alzheimer Disease: Recommendations of the Stanford Program in Genomics, Ethics, and Society*, 3 Genetic Testing 3–12 (May 1999); Laura McConnell, Barbara A. Koenig, Henry T. Greely, & Thomas A. Raffin, *Genetic Testing and Alzheimer Disease: Has the Time Come?*, 7 Nature Medicine 757–59 (July 1998).

genotype-phenotype resource encompassing just 250,000 people.[5] The U.K. is planning to sample 500,000 British citizens, subjects of Her Majesty the Queen, and they are budgeting several billion dollars for that project.[6]

Ironically, the genetic side is the cheap part; it is not that expensive to get samples from people with a cheek swab or a blood draw. The expense is attributable to getting good health information, inputting it, and making sure that it is good information. So, first, we know biobanking is going to be expensive.[7] Second, we know that it is going to have lots of tricky ethical, social, and legal issues.[8]

About eight years ago, Bartha Knoppers, Ken Kidd, and I were at a meeting, I think in the suburban D.C. area when Bartha was Chair of the Ethics Committee of the Human Genome Organization, an international, non-profit, non-governmental group. The Committee put together a set of principles to govern research in human population genetics.[9] I am afraid that I cannot remember these principles now except that they all started with "C" and that they were Bartha's Ten Commandments.[10] I do remember at least five that are particularly important in this area:

---

5. Henry T. Greely, *Iceland's Plan for Genomic Research: Facts and Implications*, 40 Jurimetrics J. 153 (2000).

6. Roger Brownsword, *An Interest in Human Dignity as the Basis for Genomics Torts*, 42 Washburn L.J. 413 (2003).

7. Alice Hsieh, *A Nation's Genes for A Cure to Cancer: Evolving Ethical, Social and Legal Issues Regarding Population Genetic Databases*, 37 Colum. J.L. & Soc. Probs, 359, 360 (2004).

8. *Id.*; J. V. McHale, *Regulating Genetic Databases: Some Legal and Ethical Issues*, 12 Med. L. Rev. 70 (2004).

9. Meeting of the Human Genome Organization Ethical, Legal, and Social Issues Committee (since renamed the HUGO Ethics Committee) at Bethesda, Maryland, Oct. 14–15, 1995. I was not a member of that Committee but was asked, along with Professor Kidd, to give a presentation to them on the Human Genome Diversity Project. We stayed for the discussion of their proposed statement on the ethics of genetics research, which became the Committee's report to the HUGO Council entitled "Statement on Principled Conduct of Genetics Research." That report can be found at http://www.gene.ucl.ac.uk/hugo/conduct.htm. It was adopted by the HUGO Council on March 21, 1996, http://www.csu.edu.au/learning/eubios/HUGO.html.

10. Statement On The Principled Conduct Of Genetics Research, http://www.csu.edu.au/learning/eubios/HUGO.html

consent, control, confidentiality, communication, and commerce.[11] For each, biobanking adds new ethical, legal, and social complications to the existing complexities of research.

First, with respect to consent, if you are going to try to get health materials, health information, and physical genetic samples from people, from a lot of people, that takes a lot of consent, and it is very tempting to short-circuit that labor and time-intensive effort to get people to say "yes." Iceland did that short-circuiting. The Icelandic legislation does not require the consent of people to be in the health data side of the private company's (deCODE) database.[12] And it is that failure of informed consent that has, in part, kept that database tied up in public controversy and litigation in Iceland.[13] More than five years after the passage of legislation that authorized deCODE to set up that database, it still does not exist.

Control, I think in some ways, is even more important. Control over what uses biobank material can be put to is an important concern for subjects, or at least it should be. I am happy to give my samples and my health information for research on a wide variety of medically related topics. And in traditional research, someone would be interested in premature grey hair and come to me and say, "Boy, do we need your sample." And I'd say, "Sure. No problem." With biobanking, in part because it is so expensive, nobody wants to create a set of 200,000 samples and information just to look at one disease, even one affirmative, enhancing condition like premature grey hair, no matter how important it is. You want to be able to use it for everything in order to amortize that several billion dollar cost.

So if I give a sample to be biobanked, I am giving it away not just for the study of premature grey hair, but for the study of all sorts of things, some of which I may not like. Personally, I would not want my samples to be used for studies of, say, race and intelligence, because I think that would likely be misused in bad

---

11. *Id.*

12. Greely, *supra*, note 5; Michael J. Smith, *Population-Based Genetic Studies: Informed Consent and Confidentiality*, 18 Santa Clara Computer & High Tech L.J. 57, 68 (2001).

13. Greely, *supra* note 5.

ways. I might be reluctant to have my samples used for studies of Irish heritage and alcoholism, two things that run in my family, and, frankly, something that my year in New Orleans with Judge Wisdom threatened greatly to exacerbate. (I should make clear, it was a function of New Orleans, not of Judge Wisdom.) So, what makes biobanking different is that you have this large resource to be used for many purposes and, under most algorithms, when you make the decision to be a participant, you no longer hold control over what your samples are going to be used for.

Third is confidentiality, preventing unauthorized uses of information that a patient or a subject handed over in confidence. ("Confidentiality" is a narrower term than "privacy," which can include attempts by others to learn information that the individual has not voluntarily handed over to someone, as well as more fundamental issues of bodily autonomy, such as those involved in reproductive decisions.) It is important for researchers to try to keep identifying information as confidential as possible and, yet, biobanking can make that difficult—particularly biobanks that want to keep adding new information as it comes up, as is the case in Iceland. If you are going to keep adding new medical information, somebody or something has to know that Computer File No. 1473892 is Hank Greely. Otherwise, when you get new information on Hank Greely, you cannot add it to his file.

Well, once you have got that, once you no longer can completely anonymize and break the link between the person's file and the person, you have the potential for a variety of different breaks of confidentiality—there is a chance of that information getting out in one way or the other. In addition, because this is not very well appreciated but very important, it must be emphasized even anonymous data these days really is not very anonymous.

If you want to have good health information on somebody, for good research purposes, you can take away their name, you can take away their Social Security number, you can take away their address, but you probably do not want to take away, for example, where they were born, or when they were born, or their sex, their race, and so forth. That can make a difference in some research.

So, I was born in Columbus, Ohio in 1952. Based on the size of Columbus, Ohio at that time, my guess is there were seven children born the day I was born. They would, most likely, have

been four boys and three girls. One of the boys was probably black; three of the boys were probably white. You look at my birthplace and my date of birth, and you check the records of Franklin County, Ohio, and you are down to only three people in the world that I might be.

So that confidentiality, even with anonymous data, cannot be guaranteed any more because a rich data set—and you want a rich data set of health information for your scientific purposes—will have information that somebody could use to track down and identify a person relatively easily. The greater computerization of data, the greater accessibility of data through the Internet, and the genealogy folks, bless their hearts, are trying to get all sorts of public records information onto the Internet as fast as possible, but this only increases this issue. So, confidentiality is a real problem.

Fourth, the problem of communication. Let us say my DNA is in a biobank and some researchers analyze it. They discover that I have a mutation in the MLH1 gene that predisposes me to hereditary colon cancer. This would be a really interesting piece of information for me because people with this mutation are highly likely to get colon cancer, but dying from colon cancer is fairly preventable if you get regular checkups, regular colonoscopies, or even have your colon taken out entirely. A colectomy is not a great thing, but it is a lot better than dying of metastatic colon cancer.

I would really like to know if I have a mutation in my MLH1 gene. But do the people running the biobank have an obligation to tell me that? If they are watching 200,000 people, every time one comes up with a health risk, do they have an obligation to go back and report? Every time they discover a new health risk, do they have to go back and look through the bank and say, "You know, we never realized that this apoE-4 variant is not only associated with high blood pressure and heart disease, but also with Alzheimer's Disease, and now we know it. Should we go back and tell all the people we know have apoE-4 alleles that they're at risk for Alzheimer's Disease?" Good questions I think. I would love to be the plaintiff's lawyer for the family of the person who died with colon cancer when a biobank knew that he had the gene mutation but did not tell him about it.

The last issue is commerce. Because these biobanks are so expensive, they are largely, except in the U.K., being funded or supported with private money. People want to get a return on their investments of several billion dollars, and that certainly is understandable. Now, in traditional medicine and traditional genetics disease research, I, a research subject, might be from a family with a history of breast cancer. I would go to the researcher because I would hope it is going to help me, or my family members, my daughters, my granddaughters, and so forth.

In biobanking, we could be looking at anything. So, I give you my DNA, you may use it to find things about diseases that I care about a lot, or for diseases that I do not care about very much. You are going to make millions of dollars out of it, or at least so you hope. You are the researcher, with your stock options and your Scientific Advisory Board memberships, and some people are going to make a lot of money out of this. I am not getting anything for it. I feel like a sucker. Commerce, commercial concerns, I think, should be greater for subjects in a world of widespread genotype-phenotype resources. So I think those are really big, important issues with respect to these biobanks, and they're issues that we don't have good answers for yet.

I want to really stress what Michael Malinowksi said, that we mainly have questions. We do not have well-accepted answers, we do not have answers that are written into law, or that even are clearly understood and agreed upon by researchers, let alone by research subjects. The potential here that I worry about is the potential for subjects to feel they have been cheated, betrayed, or lied to. The potential is, on a lesser level, a genetic Tuskegee.

The Tuskegee study, where African-American farm laborers were watched but not treated for syphilis for over forty years by the U.S. Government, was a horrific event.[14] Moreover, it has augmented the distrust by African-Americans of medical research and, in fact, even serves as a focus for concerns by African-Americans about medical research.[15] Tuskegee has, at the very

14. *See generally* James H. Jones, Bad Blood; The Tuskagee Syphilis Experiment (2d ed. Free Press 1993).

15. Dorothy E. Roberts, *The Nature of Blacks' Skepticism About Genetic Testing: Purpose, Voice, and Values Essay*, 27 Seton Hall L. Rev. 971 (1997).

least, contributed to the generally lower rate of participation in medical research by African-Americans compared to European-Americans.[16] We do not want that in genetics, for African-Americans, cystic fibrosis patients, or any other group. We do not want people to feel cheated and betrayed. It is bad for the research if people do not feel that they have been treated well by researchers. And all of these issues with respect to biobanks have the potential for leading to that kind of disillusionment and unhappiness.

Let me talk now a little bit about populations. When I talk about population research, I could mean two different things. One possibility is to encompass entire populations. Iceland is attempting to get all of its population to participate. Now, it has been five years, and I think it is unlikely that DeCode will ever achieve what the film clip that Professor Malinowski showed implies already exists. Alternatively, population research can focus on a particular disease among a particular population. Tay-Sachs disease among Cajuns, alcoholism among the Irish, schizophrenia among the Hopi, whatever it is, this approach looks at genetic variation within a particular culturally defined population. Researchers do that a lot because we know, from epidemiology, that some populations have higher rates of certain diseases than others. We do not know in advance whether the cause is genetic, environmental, or a combination of the two, but we know that disease rates differ some from population to population.

It is often recounted that F. Scott Fitzgerald once said to Ernest Hemingway, "The rich are different from you and me," and that

16. That African-Americans are less willing to participate in biomedical research is widely believed and the reason has often been traced to distrust. *See, e.g.*, Giselle Corbie-Smith, Stephen B. Thomas, Mark V. Williams, Sandra Moody-Ayers, *Attitudes and Beliefs of African-Americans Toward Participation in Medical Research*, 14 J. Gen. Int. Med. 537 (1999). *But see* David Wendler, et. al., *Are Ethnic Minorities Less Willing to Participate in Health Research?*, 3-PLOS-Medicine, No. 2, e19 (Feb. 2006), *available at* http://medicine.plosjournals.org/perlserv/?request=get-document&doi=10.1371/journal.pmed.0030019 (arguing that the reason for lower African-American participation in medical research is insufficient outreach to African-American populations).

Hemmingway replied, "Yes, they have more money."[17] If I say "Population research is different from individual research," you could say, "Yeah, there are more individuals involved." I do not think that is right. I think population research is qualitatively different because, when you do research concerning a population, those results affect the entire population, including the people who chose not to participate, or who were never asked whether they wanted to participate. If you put out a research finding that says the Irish have a particularly high rate of carrying an alcoholism gene, that affects every person of Irish ancestry. If you put out something that says the Hopi have a particularly high rate of schizophrenia—please note, I am making up these examples, but if that is what you say—that affects all the Hopi, those who participated in the study and those who did not. Population research has broader effects because it says something, not just about the individual who participates or that individual's immediate family, but about everybody who falls into that culturally defined group.

In general, we like to think that people who might be negatively affected by research have the right to decide whether they want to participate in it or not. And that brings me to some of the same five issues I discussed above, now played through in a population context. First, consider consent. In addition to requiring individual consent, when research is about a group, should we require group consent? Now I have been out on one end of this argument, arguing that when it is feasible, we should. That is not a popular end, in part because it is often going to be very hard to figure out what the group is, how you can get their consent,

17. This excellent repartee is apparently apocryphal. See the discussion by Professor James West, a Fitzgerald scholar, in an on-line discussion of Fitzgerald, http://www.washingtonpost.com/wp-dyn/content/discussion/2005/08/26/DI2005082601396.html. Hemingway did include a version of the exchange, between Fitzgerald and an unnamed third party, in his short story, The Snows of Kilamanjaro, where the autobiographical narrator says:

> The rich were dull and they drank too much or they played too much backgammon. They were dull and they were repetitious. He remembered poor Scott Fitzgerald and his romantic awe of them and how he had started a story once that began, "The rich are different from you and me." And how someone had said to Scott, "Yes, they have more money."

See the discussion of this incident in John Updike, Poor Little Rich Boy, The Guardian, June 21, 2003.

or who speaks for them. But I think, when you can do it, you should do it. Even when there is no organized group from which to seek consent, in this kind of research you should always pay attention to the group. Maybe you cannot get all Cajuns to agree on anything. There is no Cajun government or organization that speaks for all Cajuns, but you can talk to Cajun groups, you can talk about the effects of the research on all Cajuns. You can give talks in parishes up and down south Louisiana and get the Cajun population at least to think about these implications.

Control in this context is, in some ways, even more important. Some groups will be happy to participate in some research, particularly research aimed at diseases that affect them. Giving subjects some control over how their samples can be used can be important. The late Ryk Ward, a genetic anthropologist, ran into trouble a few years ago. He had collected samples from a tribe in British Columbia to be used in disease-focused research—arthritis.[18] He ended up using the samples later in some anthropological research, looking at connections between that tribe and other tribes.[19] The tribe found out and was really annoyed. They said, "We didn't give it to you for that. We gave it to you to find out about our diseases. We don't care whether you think we came over the Bering Strait or not. That's not the kind of research we want." And they felt unhappy with him—betrayed and cheated.

Confidentiality is the third factor. Confidentiality in a group context may sound kind of silly, but perhaps it is not. Do you have to say, "We found a gene that correlates with alcoholism in the Irish?" Could you say in Celtic populations, or in northern European populations? Do you have to say that the schizophrenia gene is from the Hopi? Could you say Puebloan peoples? Could you say a tribe from southwest United States? Of course, every time you fuzz the identity, you do lose something. Literature, the scientific literature, is not as clear and concise. Somebody might

18. *They Need Your DNA*, New Scientist, Sept. 30, 2000, *available at* 2000 WLNR 17410; Kurt Kleiner, *Blood Fued*, New Scientist, Sept. 30, 2000, *available at* 2000 WLNR 17453.

19. Ann Gibbons, *The Peopling of the Americas*, Science, Oct. 4, 1996, 1996 WLNR 3527432.

try to repeat the schizophrenia research on the Hopi, not realizing that you have already done it. But the Hopi might insist on it, or the Irish might insist on it, or others might insist on it as a way of protecting their reputations from your research.

Fourth, let us consider communication in the group research context. Too often, scientists think that, when they have done research with groups, communicating the results back to them means sending them reprints of their articles—which the scientists think are written in English, although real people do not think that they are written in English. And some populations do not read English. Communicating results back should often mean going back in person, talking to people, telling them your results, giving them a chance to ask questions, and explaining it in language they understand. They should understand that the samples they gave you and the interview they gave you three years ago did not just disappear into a black hole; rather, it generated some results. Maybe they are great, exciting, wonderful results. Ninety-nine percent of the time they are negative or equivocal results, but you have found out something and you have told them what it is. This process of communication respects their participation.

Commerce, the last of these issues, also takes on a different aspect with populations, particularly, though not exclusively, when you are talking about populations that are not a majority and may have been oppressed by the majority population. They say, "You came, you took our land, you took our people, you took our resources. Now, you're trying to take our genes. And you want to make a lot of money out of it, and none of it is coming back to us." This concept, I think, has been greatly exaggerated.

I do not think this happened at all in human genetics, in part because I do not think any particular human gene—more accurately, "allele"—found in any particular group has been all that important. Groups do not have genes different from those of other people, and they rarely even have variations that are not found anyplace else. But a lot of people who have been exploited in the past, who feel that they have been exploited, have real concern that this is just another way of exploiting them. And so, having some sort of sharing of financial gains is, I think, an important way to try to assuage those fears to both be more fair and to appear more fair.

Now, like all of these concerns and possible solutions, the solution is tricky as well. You cannot get research subjects by promising them too much of a return. That is unfair inducement. If groups think that their participation in research is like a lottery ticket because of a three percent royalty, then you run the risk that you are misrepresenting the real benefits to them by inducing them to take part in research under false pretenses. They may believe they are all going to be rich when, at least 99.9 times out of 100, they will not be. You know that you will get only about sixty-seven cents back on the dollar if you play the lottery long enough, but people still buy lots of lottery tickets. Yet, to completely ignore the one in a hundred possibility and to take from a group without any possibility of a return also seems wrong to me.

I have given you far more questions, I think, than answers. That is because we are at a point where this research has begun, but it has not become common enough to create shared expectations. We are the people, we are the generation that has both the opportunity and the duty to create some settled expectations, some rules, some guidelines, some standards, about how people in biobanks and populations in biobanks should be treated. We need to think that through and do a good job, because how well the research subjects feel they have been treated is at stake. In the long run, the willingness of them to participate in research and to support research is at stake. This is an issue, not just of ethics and for ethicists, but of science and for scientists.