A RESEARCH REVOLUTION: GENETIC TESTING CONSUMERS BECOME RESEARCH (AND PRIVACY) GUINEA PIGS

ANGELA L. MORRISON*

* J.D., University of Colorado Law School (2011); M.S., Colorado State University (2007).

INTRODUCTION

Direct-to-consumer genetic testing companies, which offer DNA sequencing services to paying customers, and publicly-funded genetic research consortiums have both begun to conduct Internet-based genetic research studies. The offered levels of privacy and anonymity vary greatly, but even those entities that promise maximum privacy protection can no longer guarantee much, as investigators recently revealed how easy it is to identify a given individual from "anonymized" data. Prohibiting these reidentification events is unlikely to be effective, but prohibition on the opposite end (participation in research) is undesirable because individuals, the research community, and the public at large benefit from scientific and medical discoveries achieved through research. And yet, if study participants fear the consequences of participation, such as genetic discrimination or loss of control of their information and its dissemination (e.g. being identified as a Huntington's disease carrier when you had chosen not to tell anyone), participation will decrease. Because we cannot stop the advancements in computer science that enable reidentification, it is time to update existing protections for traditional human research subjects in order to meet the demands of the rapidly advancing online research community, specifically via open and interactive informed consent.

I. BACKGROUND

Advances in human science and medicine are often dependent upon research on human subjects. Shameful experiments on humans, not only in the oft-cited Nazi Germany, but also in the United States, lead to regulations on human subject research and protections for the participants. Since then, rapid developments in the fields of genetics and genomics along with the rise of an Internet society have greatly expanded

the breadth of topics on which human subject research can be conducted and the forums in which that research can take place. This section provides some background on human subject research and its regulation, and on genetic sequencing and its availability online.

A. Regulation of Human Subject Research in the United States

1. Historical Examples of Abuse

In July of 1972, an Associated Press journalist broke the story of a United States Public Health Service study, then in its fortieth year, on the effects of untreated syphilis.[1] The Tuskegee Study involved 399 syphilitic African American men in rural Macon County, Alabama who apparently did not know they had syphilis or knew, but did not know they were not receiving treatment for it.[2] Instead, they were receiving free medical examinations, hot meals, and a burial stipend for participation in a study in which researchers simply let syphilis run its course in order to investigate the ultimately fatal effects of the disease.[3] It appeared that government researchers had taken advantage of poor, illiterate men by misleading them into believing they were receiving proper medical attention.[4]

Not long before the Tuskegee story broke, researchers in New York completed a decade-long study on the effects of viral hepatitis.[5] The study subjects were residents of Willowbrook State School, a now-defunct institution for mentally challenged children.[6] Researchers intentionally infected some of the children in order to study controlled progression of the disease.[7] The investigators justified their work by pointing out that viral hepatitis was endemic in the institution, and deliberate infection with a mild strain conferred immunity against more virulent strains.[8] They also obtained consent for the artificial induction from the residents' parents.[9] Opponents denounced using children in

1. JAMES H. JONES, BAD BLOOD: THE TUSKEGEE SYPHILIS EXPERIMENT 1 (new & expanded ed. 1993).

2. *Id.* at 5-6.

3. *Id.* at 4.

4. *Id.* at 13-14.

5. *See generally* Saul Krugman, Joan P. Giles & Jack Hammond, *Infectious Hepatitis: Evidence for Two Distinctive Clinical, Epidemiological, and Immunological Types of Infection*, 200 JAMA 365 (1967) (providing background information about hepatitis at Willowbrook and describing several of the studies).

6. *Id.* at 366.

7. *Id.*

8. Saul Krugman, Letter to the Editor, *Experiments at the Willowbrook State School*, 297 LANCET 966, 966-67 (1971).

9. Krugman et al., *supra* note 5, at 366.

medical research, especially the mentally impaired.[10] They also argued that consent was in effect coerced because when the main unit of the institution refused to accept new patients (due to alleged overcrowding), the separate research unit, reserved for study participants only, continued to welcome residents; thus, the only avenue for admittance was participation in the hepatitis study.[11]

2. Current Laws

Public outrage over studies such as those at Tuskegee and Willowbrook drew some unfavorable comparisons to the experiments of Nazi Germany,[12] but at least one of the research directors failed to see any similarity to Nazi abuses or any applicability of the Nuremberg Code to his work.[13] Recommendations regarding the ethical treatment of human subjects in research had actually existed for several decades,[14] but there was almost no federal oversight until the National Research Act of 1974,[15] which was enacted after Tuskegee and Willowbrook. Years of discussions following the 1974 Act were memorialized in the influential Belmont Report,[16] which eventually lead, in 1991, to the Federal Policy for Protection of Human Subjects, more often referred to as the Common Rule.[17]

The Common Rule mandates that researchers at any public or private institution (hospital, clinic, laboratory, etc.) that receives government funding or is otherwise regulated by the government first

10. Stephen Goldby, Letter to the Editor, *Experiments at the Willowbrook State School*, 297 LANCET 749 (1971).

11. M.H. Pappworth, Letter to the Editor, *The Willowbrook Experiments*, 297 LANCET 1181 (1971).

12. JONES, *supra* note 1, at 12.

13. *Id.* at 179-80 (referring to interview with Dr. John R. Heller, director of the Division of Venereal Diseases from 1943-48); Nuremberg Medical Trial, *The Nuremberg Code (1947)*, *in* CONTEMPORARY ISSUES IN BIOETHICS 70, 70-71 (7th ed. 2008) (reprinting the Nuremberg Code, which lists ten "basic principles [that] must be observed in order to satisfy moral, ethical and legal concepts" of human medical experiments).

14. *See*, *e.g.*, Minutes of the Supplemental Session of the House of Delegates of the American Medical Association, Held in Chicago, Dec. 9-11, 1946, *reprinted in* Organization Section, 132 JAMA 1075, 1090 (1946).

15. RESEARCH COMPLIANCE SERVS., OFFICE OF RESEARCH, THE HUMAN SUBJECTS RESEARCH REVIEW SYSTEM 1 (2001). The predecessor to the U.S. Department of Health and Human Services, the Department of Health, Education, and Welfare, had issued guidelines on the protection of human subjects just three years earlier, in 1971. *Id.*

16. NAT'L COMM'N FOR THE PROTECTION OF HUMAN SUBJECTS OF BIOMEDICAL & BEHAVIORAL RESEARCH, NAT'L INSTS. OF HEALTH, THE BELMONT REPORT (1979), *available at* http://ohsr.od.nih.gov/guidelines/belmont.html.

17. Federal Policy for the Protection of Human Subjects, 56 Fed. Reg. 28,003 (June 18, 1991) (codified in scattered sections of C.F.R.).

seek approval for studies involving human subjects.[18] Approval comes from a local Institutional Review Board ("IRB"), a group of at least five people of diverse backgrounds, including one community member, at least one person whose primary concerns are scientific, and at least one person whose primary concerns are nonscientific.[19] The members must have expertise to review specific research projects, must know the applicable law, and must know the standards of professional practice.[20] No member may review his or her own research proposal.[21]

IRBs can decide to approve, demand modification of, or deny a research proposal involving human subjects.[22] Requisite criteria for an approvable proposal include the following: risks to the subjects are minimized, risks to the subjects are reasonable in relation to the anticipated benefits, selection of subjects is equitable, and informed consent is obtained.[23] Proper informed consent includes a description of risks to the subject, identification of benefits to the subject or to others, and a statement that participation is voluntary and that the subject may withdraw at any time.[24] Additionally, a statement describing the extent of confidentiality of records identifying the participant must be included.[25] The existence of adequate provisions for protecting the privacy of subjects and maintaining confidentiality of data is one of the criteria IRBs consider when they evaluate proposals.[26]

IRBs are not required at institutions that do not receive any federal funding and whose research does not otherwise fall under federal regulation.[27] This means that private pharmaceutical companies *are* subject to IRB approval because a government agency, the Food and Drug Administration ("FDA"), regulates the product of their research—pharmaceuticals.[28] IRB approval is also not required when the research project (even if it is federally funded) is limited to the collection of existing data that is publicly available, or if the subjects cannot be identified "directly or through identifiers linked to the subjects."[29]

18. 45 C.F.R. § 46.101(a) (2010).
19. *Id.* § 46.107(a), (c).
20. *Id.* § 46.107(a).
21. *Id.* § 46.107(e).
22. *Id.* § 46.109(a).
23. *Id.* § 46.111(a).
24. *Id.* § 46.116(a).
25. *Id.* § 46.116(a)(5).
26. *Id.* § 46.111(a)(7).
27. *See id.* § 46.101.
28. *See id.*; 21 C.F.R. §§ 56.101-.115 (2010).
29. 45 C.F.R. § 46.101(b)(4) (2010).

Another relevant law is the Health Information Portability and Accountability Act ("HIPAA").[30] HIPAA prohibits covered entities—health care providers, health plans, health care clearinghouses—from "us[ing] or disclos[ing] protected health information without an authorization" (meaning, informed consent) from the patient whose information is subject to use or disclosure.[31] However, this prohibition does not apply to "de-identified" information; covered entities may use and disclose personal health information, without restriction, if they first remove eighteen listed identifiers such as name, social security number, and home address.[32] Additionally, no patient consent is required for the use or disclosure of otherwise protected health information if that use or disclosure is for research purposes, although a waiver of consent must first be approved by an IRB or a similar privacy board.[33]

B. Human Genome Sequencing and Testing

1. Sequencing the Human Genome

During the 1990s, the publicly-funded Human Genome Project and the private company Celera raced to sequence the human genome, that is, to report all of the nucleotide "letters" of all of the DNA comprising all 24 chromosomes.[34] The groups jointly announced their first working drafts of the human genome in 2000 and they released approximately 90 percent complete annotated drafts the following year.[35] Additional information has been added ever since. In 2007 the first complete sequence of a single individual's genome was published.[36]

---

30. Health Insurance Portability and Accountability Act, Pub. L. No. 104-191, 110 Stat. 1936 (1996).

31. 45 C.F.R. § 164.508(a)(1) (2010). Health information does include genetic information. 42 U.S.C. § 1320d-9(a) (2006).

32. 45 C.F.R. §§ 164.502(d)(2), 164.514(b)(2) (2010). Alternatively, if a statistician determines that "the risk [of reidentification] is very small," then the information can be used and disclosed without restriction. *Id.* §§ 164.502(d)(2), 164.514(b)(1).

33. *Id.* § 164.512(i)(1)(i).

34. *See generally* James D. Watson, *The Human Genome Project: Past, Present, and Future*, 248 SCIENCE 44, 45 (1990) (explaining the project, including that all 22 autosomes plus the X and Y sex chromosomes would be sequenced). Celera joined the race full time in 1999. J. CRAIG VENTER, A LIFE DECODED: MY GENOME: MY LIFE 286 (2007).

35. Int'l Human Genome Sequencing Consortium, *Initial Sequencing and Analysis of the Human Genome*, 409 NATURE 860, 875 (2001); J. Craig Venter et al., *The Sequence of the Human Genome*, 291 SCIENCE 1304, 1315 (2001).

36. Samuel Levy et al., *The Diploid Genome Sequence of an Individual Human*, 5 PLOS BIOLOGY 2113, 2114 (2007) (identifying J. Craig Venter as the single DNA donor). The first published DNA sequences had been derived from several anonymous donors. Int'l Human Genome Sequencing Consortium, *supra* note 35, at 865-66; VENTER, *supra* note 34, at 285-86.

2. Direct-to-Consumer Genetic Testing

Advancements in sequencing equipment and technology that were developed in conjunction with the sequencing of the human genome soon paved the way for several Internet-based companies to provide genome sequencing services directly to members of the general public. These companies are referred to as direct-to-consumer genetic testing companies ("DTC-GTCs").[37] For the average individual interested in genealogy or predisposition to a given disease, full sequencing (returning the individual nucleotide letters of one's entire genome), although rapidly decreasing in cost, is still prohibitively expensive and time-consuming.[38] A less expensive option is exome sequencing, which examines just the exons within genes (which are already only a portion of the entire genome).[39] An even cheaper and faster approach is to examine single nucleotide polymorphisms ("SNPs," pronounced "snips"). Each SNP involves a single change in our DNA code; one nucleotide may have been substituted for a more common one, an extra nucleotide may have been inserted, or one may have been deleted.[40] More than one million SNPs have already been identified and reported by researchers, and they correspond to both non-clinical (e.g. eye color) and clinical (e.g. sickle cell anemia) traits. By using DNA chip technology, DTC-GTCs can study a million SNPs almost simultaneously.[41]

37. *See*, *e.g.*, 23ANDME, https://www.23andMe.com (last visited Feb. 24, 2011); DECODEME, http://www.decodeme.com (last visited Feb. 24, 2011); LUMIGENIX, http://www.lumigenix.com (last visited Feb. 24, 2011); NAVIGENICS, http://www.navigenics.com (last visited Feb. 24, 2011); PATHWAY GENOMICS, https://www.pathway.com (last visited Feb. 24, 2011).

38. Illumina's full genome sequencing service, available only with a physician's referral, dropped from $48,000 in early 2010 to $19,500 by July of that year. And the company charges just $9,500 if a physician certifies that sequencing could lead to treatment of that patient's condition. Randall Parker, *Illumina Full Genome Sequencing Costs Below $20k*, FUTUREPUNDIT (July 18, 2010, 11:37 AM), http://www.futurepundit.com/archives/007347.html. Soon, full genome sequencing may cost one tenth of that. John Markoff, *I.B.M. Joins Pursuit of $1,000 Personal Genome*, N.Y. TIMES, Oct. 6, 2009, at D2.

39. Pauline C. Ng et al., *Genetic Variation in an Individual Human Exome*, 4 PLOS GENETICS 1, 1 (2008), http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2493042/pdf/pgen.1000160.pdf.

40. *See*, *e.g.*, Alain Vignal et al., *A Review on SNP and Other Types of Molecular Markers and Their Use in Animal Genetics*, 34 GENETICS SELECTION EVOLUTION 275, 277-78 (2002).

41. *How Does 23andMe Genotype My DNA?*, 23ANDME, https://www.23andme.com/you/faqwin/chip (last visited Apr. 3, 2011).

II. ONLINE COMMUNITIES AS A SOURCE OF HUMAN SUBJECTS FOR GENETIC RESEARCH

Most research studies on humans have traditionally been conducted by, and physically at, institutions such as universities, hospitals, and pharmaceutical companies. For example, a pulmonologist might have invited her cystic fibrosis patients to participate in a study on a new treatment, or a clinical professor might have recruited sets of identical and fraternal twins for a nature-versus-nurture study.

While studies on new surgical methods might still have to be conducted in person on an operating room table, the Internet has greatly expanded the types of research studies that can be conducted without personal interaction between researcher and subject. The pulmonologist could have her cystic fibrosis patients and their families submit DNA samples to an online-based testing company in order to investigate the genetic basis of the disease, and she could also create online surveys to gather information about how her patients respond to a new treatment. Similarly, the twins in the nature-versus-nurture study could submit their DNA samples to an online-based twins' community and answer survey questions there, without ever meeting the professor. The Internet has created a whole new forum for scientific research, and both for-profit and not-for-profit groups are utilizing it.

A. Research Arms of Direct-to-Consumer Genetic Testing Companies

The initial premise of DTC-GTCs was that consumers paid to have a DNA sample sequenced and analyzed for information related to their ancestry, risk of disease, and non-disease traits (e.g. earwax type). More recently, companies have added research opportunities whereby customers can share their purchased genetic information for use in research. For the first time, subjects are paying to be enrolled in research studies.[42]

One of the largest DTC companies, 23andMe, has a research arm called 23andWe for the purpose of investigating the basic causes of disease, developing drugs and other treatments, and predicting an individual's risk of disease.[43] It aims to accomplish these goals by creating a larger pool of samples than can be achieved through typical

42. In the typical scenario, research subjects are paid for their participation in a study. Council for International Organization of Medical Sciences (CIOMS), in Collaboration with the World Health Organization (WHO), *International Ethical Guidelines for Biomedical Research Involving Human Subjects (2002)*, *in* CONTEMPORARY ISSUES IN BIOETHICS, *supra* note 13, at 79.

43. *Consent Document*, 23ANDME, https://www.23andme.com/about/consent (last visited Apr. 3, 2011).

location-based clinical trials.[44] Participants contribute their genetic information and answer online surveys, which cover a range of topics from non-clinical traits (e.g. freckles, right- or left-handedness) to serious diseases (e.g. Parkinson's disease, diabetes), as well as less-serious conditions (e.g. migraines, lactose intolerance) and responses to drugs.[45] Two clinical research communities, one for Parkinson's disease and one for sarcoma, have also been established. Those who have been diagnosed with Parkinson's, sarcoma, or related disorders may join the respective community if they pledge to contribute their sequenced DNA and to take online surveys about their experiences with the diseases.[46] 23andMe will then "correlate [customers'] responses to online surveys with their genetic data" in order to conduct research studies.[47]

Although not clearly defined or promoted like 23andMe's 23andWe, Navigenics does have a research arm through which it conducts its own genetic and medical research.[48] Navigenics's customers' genetic information may also be used for external research studies,[49] as is true for the customers of most, if not all, DTC-GTCs.[50]

B. Non-Profit Research Consortiums

Publicly funded research projects aim to advance the understanding

---

44. *Featured Research*, 23ANDME, https://www.23andme.com/slideshow/research (slide 1) (last visited Apr. 3, 2011). Until 2010, 23andMe was also running a program called "Research Revolution" which, like 23andWe, was aimed at creating, via the 23andMe customer base, a large enough data pool such that statistically meaningful research on a particular health condition could be conducted. Customers could donate their genetic information to studies on one or more of ten health problems, which included migraines, epilepsy, multiple sclerosis, and rheumatoid arthritis. Those who were suffering from the conditions were called "patients" and when 1,000 patients had enrolled, 23andMe promised to start a research study using both in-house resources and outside experts. Customers who were not suffering from a particular condition could still sign up as "supporters" of that condition and their genetic information would serve as experimental controls. *Research Revolution*, 23ANDME, https://www.23andme.com/researchrevolution (last visited Apr. 1, 2010) (archived copy on file with Journal on Telecommunications and High Technology Law). It appears that 23andWe now encompasses the general themes of Research Revolution, but without customer-directed resource allocation to, or topic selection of, research studies.

45. *Featured Research,* 23ANDME, https://www.23andme.com/slideshow/research (slide 2) (last visited Apr. 3, 2011); *Consent Document*, *supra* note 43.

46. *23andMe Parkinson's Community: Strength in Numbers*, 23ANDME, https://www.23andme.com/pd (last visited Apr. 3, 2011); *23andMe Sarcoma Community: A Patient-Driven Revolution in Sarcoma Research*, 23ANDME, https://www.23andme.com/sarcoma (last visited Apr. 3, 2011).

47. *Featured Research*, *supra* note 44.

48. *See Privacy Policy*, NAVIGENICS, http://www.navigenics.com/visitor/what_we_offer/our_policies/privacy (last updated June 19, 2009).

49. *Id.*

50. *See infra* Part III.B.1.

of genetic and environmental contributions to human traits, as well as to improve medical professionals' ability to diagnose, treat, and prevent illness.[51] The 1000 Genomes Project is an international research consortium formed "to create the most detailed and medically useful picture to date of human genetic variation."[52] The project's goal is to sequence the genomes of approximately 1200 people worldwide and to make that data "swiftly available to the worldwide scientific community through freely accessible public databases."[53] Similarly, the mission of Harvard School of Medicine's Personal Genome Project ("PGP") is to develop personal genomics technology and practices that "yield identifiable and improvable benefits at manageable levels of risk."[54] It is currently recruiting almost 100,000 volunteers to share their personal and genetic information.[55]

III. GENETIC INFORMATION PRIVACY IN ONLINE RESEARCH COMMUNITIES

Sharing genetic information—and sharing a lot of it—has become very easy very quickly. Technological advances continue to improve the speed and accuracy of DNA sequencing, and scientific research continues to increase our understanding of what those sequences mean as predictors of health, disease, or response to pharmaceuticals. Further, the Internet allows for convenient, rapid dissemination of all of that information: the reports about improved sequencing techniques, the announcements of the latest medical breakthroughs, and uploads of the sequence data itself. For-profit companies and non-profit consortiums are already taking advantage of the Internet as a venue for genetic research studies and as a source of research subjects. But all of this is happening at least one step ahead of legislators. This section explores the extent to which genetic information is protected by law, along with how online-based researchers protect the genetic information of their subjects.

51. *See*, *e.g.*, PERSONAL GENOME PROJECT, http://www.personalgenomes.org (last visited Apr. 3, 2011).

52. Press Release, 1000 Genomes Project, International Consortium Announces the 1000 Genomes Project 1 (Jan. 22, 2008), http://www.1000genomes.org/sites/1000genomes.org/files/docs/1000genomes-newsrelease.pdf.

53. *Id.*

54. *Mission*, PERSONAL GENOME PROJECT, http://www.personalgenomes.org/mission.html (last visited Apr. 5, 2011).

55. PERSONAL GENOME PROJECT, *supra* note 51.

A. Relevant Legislation

Genetic information is unique compared to other types of health information because it "can reveal information about an individual's current family members and future offspring" including "predispositions and personal characteristics" even when those predispositions and characteristics are not readily apparent from a person's appearance or current health status.[56] Additionally, genetic information is "remarkably identifiable."[57] In the United States, federal and state legislation provides some protection against misuse of this information.

1. The Genetic Information Nondiscrimination Act of 2008

Reports of discrimination based on genetic information began as early as 1991.[58] Federal legislation, including the Americans with Disabilities Act, Title VII of the Civil Rights Act of 1964, and the Health Insurance Portability and Accountability Act, did not directly address genetic information and protection under these acts remained uncertain, and at best, limited.[59] States attempted to supplement federal legislation,[60] but the resulting patchwork of laws provided "inadequate" protection from discrimination.[61]

The Genetic Information Nondiscrimination Act ("GINA") was passed in 2008 to provide "national and uniform" protection against genetic discrimination.[62] It prohibits employers from discriminating against job applicants or current employees based on their genetic information.[63] It also amended the Employee Retirement Income Security Act of 1974 ("ERISA"),[64] the Public Health Service Act ("PHS"),[65] the Internal Revenue Code of 1986 (relating to group health

56. Jeffrey P. Braff et al., Am. Health Lawyers Ass'n's Advisory Council on Racial & Ethnic Diversity, *Patient-Tailored Medicine, Part Two: Personalized Medicine and the Legal Landscape*, J. HEALTH & LIFE SCI. L., Jan. 2009, at 1, 37 (citing James G. Hodge, Jr., *Ethical Issues Concerning Genetic Testing and Screening in Public Health*, 125C AM J. MED. GENETICS 66, 69 (2004)).

57. *Id.*

58. Morse Hyun-Myung Tan, *Advancing Civil Rights, The Next Generation: The Genetic Information Nondiscrimination Act of 2008 and Beyond*, 19 HEALTH MATRIX 63, 73-77 (2009) (reviewing instances of genetic discrimination).

59. *See* Daniel Schlein, *New Frontiers for Genetic Privacy Law: The Genetic Information Nondiscrimination Act of 2008*, 19 GEO. MASON U. C.R. L.J. 311, 318-40 (2009).

60. *See id.* at 347-50; Tan, *supra* note 58, at 89-93.

61. *See* Genetic Information Nondiscrimination Act of 2008, Pub. L. No. 110-233, 122 Stat. 881, § 2 (2008).

62. *Id.*

63. 42 U.S.C.A. §§ 2000ff to ff-11 (West 2010).

64. Genetic Information Nondiscrimination Act of 2008, Pub. L. No. 110-233, § 101, 122 Stat. 881 (2008) (to be codified in scattered sections of 29 U.S.C.).

65. *Id.* § 102 (to be codified at 42 U.S.C. §§ 300gg to gg-53, *amended by* Patient

insurance),[66] and Title XVIII of the Social Security Act (relating to Medicare supplemental insurance)[67] to uniformly prohibit issuers of both individual and group health insurance plans from using genetic information for underwriting purposes, both when issuing a policy and when setting its price.[68]

Although GINA fills in many gaps left by the previous patchwork of legislation and common law, it is not comprehensive. It does not protect information about the actual manifestation of a disease or disorder.[69] GINA also does not cover "spheres of life" outside of employment and insurance.[70] And within the context of insurance, it covers only health insurance; disability, long-term care, and life insurance are not included. Although insurers cannot "request, require, or purchase" genetic information, collecting that information "incidental" to acquiring other allowable information is permitted.[71] Furthermore, GINA says nothing about genetic discrimination by financial service providers or in social contexts. Importantly, GINA does not protect information in the public domain, such as information on the Internet.

2. State Laws

GINA preempts state laws only to the extent of mandating minimum standards;[72] states are free to enact legislation with stricter genetic information protections. Although GINA's insurance provision is limited to health insurance, 17 states currently have laws that cover other types of insurance.[73] Fourteen states restrict genetic information discrimination in life insurance, 15 states do so for disability insurance, and nine do so for long-term care insurance.[74]

Protection and Affordable Care Act, Pub. L. No. 111-148, 124 Stat. 119 (2010)).

66. *Id.* § 103 (to be codified in scattered sections of 26 U.S.C., *amended by* Patient Protection and Affordable Care Act, Pub. L. No. 111-148, 124 Stat. 119 (2010)).

67. *Id.* § 104 (to be codified at 42 U.S.C. § 1395ss, *amended by* Patient Protection and Affordable Care Act, Pub. L. No. 111-148, 124 Stat. 119 (2010)).

68. *Id.* §§ 101-05 (to be codified in scattered sections of 26 U.S.C., 29 U.S.C., 42 U.S.C.).

69. Amy L. McGuire & Mary Anderlik Majumder, *Two Cheers for GINA?*, 1 GENOME MED. 6.1, 6.2 (2009).

70. Patrick Taylor, *When Consent Gets in the Way*, 456 NATURE 32, 33 (2008).

71. *E.g.*, 29 U.S.C.A. § 1182 (West 2010); 42 U.S.C.A. § 300gg-1 (West 2010).

72. Tan, *supra* note 58, at 103.

73. *Genetics and Life, Disability and Long-Term Care Insurance*, NAT'L CONFERENCE OF STATE LEGISLATURES, http://www.ncsl.org/IssuesResearch/Health/GeneticNondiscriminationLawsinLifeDisability/tabid/14283/Default.aspx (last updated Jan. 28, 2008).

74. *Id.*

B. Privacy Policies, Terms of Service, and Informed Consent

DTC-GTCs and non-profit research consortiums acknowledge genetic information privacy risks to varying degrees. Given the online nature of these groups, traditional signed consent forms have been replaced by click-through agreements, and traditional face-to-face conversations with medical professionals have been replaced by privacy policies and terms of service documents posted on websites. This section examines how online-based genetic research groups are—or are not—using these surrogate documents to adequately inform research participants of potential risks inherent in their population.

1. Direct-to-Consumer Genetic Testing Companies

23andMe uses a lengthy Privacy Statement and Terms of Service agreement to address the types of information the company collects, how the company uses that information, and how they protect it.[75] Personal Information ("information that can be used to identify [a user] either alone or in combination with other information") is sub-categorized in part as Registration Information (name, email address, etc. used to create an account or purchase services), Genetic Information (such as the data generated by processing a user's DNA sample), and Self-Reported Information (such as a user's survey responses).[76] Aggregated Genetic and Self-Reported Information, which has been "stripped of Registration Information and combined with data from . . . other users . . . to minimize the possibility of exposing individual-level information,"[77] from *all* users may be shared with both non-profit and commercial third parties.[78] If a user chooses to participate in 23andWe Research—which the company encourages its users to do[79]—then that user's information, in aggregated form, may also be disclosed to third parties for the purpose of publication in scientific journals.[80] In other words, the only difference in information disclosure between 23andWe participation and non-participation is that personal information of the former group, but not the latter, may be published in scientific journals.

75. *Privacy Statement*, 23ANDME, https://www.23andme.com/about/privacy (last updated June 24, 2010); *Terms of Service*, 23ANDME, https://www.23andme.com/about/tos (last visited Apr. 5, 2011).

76. *Terms of Service*, *supra* note 75, § 1.

77. *Id.*

78. *Privacy Statement*, *supra* note 75.

79. *Consent Document*, *supra* note 43.

80. *Privacy Statement*, *supra* note 75.

23andMe requires informed consent in order to participate in 23andWe,[81] and the company suggests that its Consent Document is "based upon an IRB-approved consent document."[82] 23andMe also requires informed consent for the company to share Registration Information with third-party research partners.[83] The company does not mention whether third parties with whom it shares information will be required to protect the confidentiality of that information or to refrain from attempts to reidentify individual contributors.[84]

23andMe has posted information about GINA on its website and acknowledges some, but not all, of the gaps in GINA's protection.[85] In general, however, the company does warn consumers that sharing genetic data can lead to unintended consequences, such as a third party discovering additional information about a user, or future scientific advances causing a revelation that could not have been predicted.[86]

Despite not having research arms as apparent as 23andMe's, other DTC-GTCs are no less likely to use their customers' genetic information for research purposes.[87] For example, in its Privacy Policy, Navigenics states that it "believes in . . . helping further scientific and medical research."[88] In promotion of that belief, it may use its customers' Genetic Data (i.e. the DNA genotyping results) linked to their Phenotype Information (defined to include, for example, gender, height, weight, ethnicity, and ancestry, as well as health conditions and diseases of the user and the user's family members) to "[d]iscover or validate associations between certain genetic variations and certain health conditions or traits."[89] The company may also publish its findings "without disclosing [a user's] Genetic Data in a quantity sufficient to uniquely identify [a user]."[90] Navigenics's Informed Consent document does state that Genetic Data and Phenotype Information will be used by the company

81. *Consent Document*, *supra* note 43.

82. *Terms of Service*, *supra* note 75, § 5. It is unclear whether the Terms of Service is referring to the actual Consent Document, and it is also unclear from where the "IRB-approved" document that 23andMe based its own document upon came.

83. *Privacy Statement*, *supra* note 75.

84. *See infra* Part IV.A.1.

85. "GINA does not cover life or disability insurance providers." *What is Gina?*, 23ANDME, https://www.23andme.com/you/faqwin/gina (last visited Apr. 12, 2011).

86. *Considerations*, 23ANDME, https://www.23andme.com/more/considerations (last visited Feb. 24, 2011). *See also Consent Document*, *supra* note 43.

87. *Cf.* Heidi C. Howard et al., *Blurring Lines*, 11 EMBO REPORTS 579, 579 (2010) (reporting that DTC genetic testing companies vary in their candidness about and clarity on use of customers' genetic information for research purposes).

88. *Privacy Policy*, *supra* note 48.

89. *Id.*

90. *Id.* Presumably Navigenics is aware of how little genetic data is required to identify an individual. *See* Lin et al., *infra* note 129; *see also infra* Part IV.A.1.

for internal research,[91] but it does not give any specifics about the potential subject areas or scope of that research, nor does it state whether that research has been approved by an independent authority, such as an IRB.[92]

Navigenics also gives its customers opportunities to share their Genetic Data and Phenotype Information with not-for-profit third-party organizations who conduct genetic or medical research.[93] These organizations may also publish their study results and "deposit such [Genetic Data and Phenotype Information] . . . into public data repositories or otherwise make them publicly available."[94] It is unclear whether this is an opt-in or opt-out system.

deCODEme, another DTC-GTC, may invite its customers "to participate in studies or other research activities,"[95] which suggests that participation is voluntary. However, both the company's Privacy Policy and Service Agreement also state that deCODEme may use its customers' information to gather statistical aggregate data,[96] which suggests at least minimal automatic use for internal research purposes. Additionally, that data may include "associating genetic variants with any of the self-reported user attributes,"[97] which further suggests that the company is linking Genetic Information with Self-Reported Information (as 23andMe would describe it). Because the Service Agreement (which users must agree to before their DNA samples are analyzed) doubles as deCODEme's "informed" consent document,[98] it is unclear whether the company would seek additional consent for participation in the research studies that it invites its users to join. The company's website also does not mention whether deCODEme has sought independent approval for use of its data in research studies.

Currently, fellow DTC-GTCs Lumigenix and Pathway Genomics do not appear to be conducting any in-house research, but both

91. *Informed Consent, Health Compass*, NAVIGENICS, http://www.navigenics.com/visitor/what_we_offer/our_policies/informed_consent/health_compass (last visited Feb. 24, 2011).

92. Seeking independent approval of a study plan would be voluntary because private DTC-GTCs such as Navigenics are currently not obligated to obtain permission for research studies. *See infra* Part V.B.2.

93. *Privacy Policy*, *supra* note 48.

94. *Id.*

95. *Terms of Use*, DECODEME, http://www.decodeme.com/terms-of-use (last updated June 2, 2007).

96. *Privacy Policy*, DECODEME, http://www.decodeme.com/privacy-policy (last updated Nov. 12, 2007); *Service Agreement*, DECODEME, http://www.decodeme.com/service-agreement (last updated Nov. 30, 2009).

97. *Id.*

98. The full name of the document is "deCODEme Genetic Scan Service Agreement and Informed Consent." *Service Agreement*, *supra* note 96.

companies may collaborate with other organizations who conduct research.[99] Not only do both companies pledge to obtain their customers' express consent before sharing any data with collaborators, they will also require that collaborators obtain study permission from an IRB.[100]

2. Non-Profit Research Consortiums

The privacy risks associated with participation in the 1000 Genomes Project are seemingly low because the researchers are not collecting any personally identifying or medical information from subjects.[101] Most of the genome sequences will come from DNA previously submitted to another project for which no personally identifying or medical information was collected either.[102]

In complete contrast, privacy and confidentiality for subjects in the Personal Genome Project will be almost non-existent.[103] Because one of the PGP's goals is to explore research and commercial uses of human genetic data linked to trait information,[104] almost all information submitted by a participant, including physical trait and medical information, and even photographs, will be posted on the PGP's public website and database along with genetic information.[105]

To help ensure that study participants understand the implications of their involvement, interested individuals must first submit to a rigorous IRB-approved pre-enrollment process that begins with an eligibility questionnaire and an entrance exam.[106] The latter assesses "comprehension of concepts . . . includ[ing] . . . potential risks of participating, project protocols, and basic genetics";[107] a passing score of no less than 100 percent

99. *Privacy Policy*, LUMIGENIX, http://www.lumigenix.com/privacy (last updated Nov. 10, 2010); *Privacy Statement*, PATHWAY GENOMICS, https://www.pathway.com/about-us/privacy-policy (last visited Apr. 12, 2011).

100. *Supra* note 99.

101. Press Release, *supra* note 52, at 2.

102. *Id.* at 4 (noting that samples collected, without any medical or identifying information, for the HapMap project will be used for much of the 1000 Genomes Project). *But see infra* Part IV.A.1 (discussing reidentifiability of an individual anonymous contributor to a pool of thousands of DNA samples, such as that compiled by the HapMap project).

103. PERSONAL GENOME PROJECT, CONSENT FORM: PERSONAL GENOME PROJECT § X:10.1 (2010), *available at* http://www.personalgenomes.org/consent/PGP_Consent_Approved03312010.pdf [hereinafter FULL CONSENT] (explaining that PGP will not keep data in a confidential or anonymous fashion, nor will it require third parties who access the data to do so).

104. PERSONAL GENOME PROJECT, CONSENT FORM: ELIGIBILITY SCREENING FOR THE PERSONAL GENOME PROJECT § I (2010), *available at* http://www.personalgenomes.org/consent/PGP_MiniConsent_Approved03312010.pdf [hereinafter MINI CONSENT].

105. FULL CONSENT, *supra* note 103, §§ IV:4.1, V:5.1, V:5.5.

106. MINI CONSENT, *supra* note 104, § IV:4.1-.2.

107. *How It Works*, PERSONAL GENOME PROJECT,

correct is required.[108] If an applicant is deemed eligible to continue, the next pre-enrollment steps include submitting baseline trait data[109] and verifying identity.[110] After the PGP enrolls an individual, the baseline trait data is published on the PGP's public website and database, and the enrollee submits tissue samples.[111] Those tissue samples are used for DNA analysis, which in turn is used to generate a Preliminary Research Report that should help participants decide whether or not to release their genetic data to the public website and database[112] where it will be associated with the previously-submitted baseline trait data.[113] Once released, neither the participant nor the PGP can control who has access to, makes copies of, or otherwise uses the information.[114]

This ready accessibility suggests a myriad of worrisome scenarios limited only by imagination. According to the PGP, "anyone with sufficient knowledge and resources" could use the online data to truthfully claim that a participant is, for example, predisposed to a disease or related to criminals.[115] If someone altered and republished the data the same claims of disease predisposition or criminal relationships could be falsely made.[116] Someone could even "make synthetic DNA and plant it at a crime scene,"[117] thus implicating an innocent person in a crime. Additionally, although the information will not intentionally become part of a participant's medical record, the information could be identified with a particular individual and added to that person's record,[118] possibly affecting access to health insurance despite GINA. Further, GINA permits employers to acquire genetic information from the public domain, which would include the PGP's public website and database.[119]

---

http://www.personalgenomes.org/howitworks.html (last visited Feb. 24, 2011). This type of informed consent, in which researchers openly admit that subjects' privacy and anonymity cannot be guaranteed, and subjects then consent to the possibility of complete public disclosure of their information, is called "open consent." Jeantine E. Lunshof et al., *From Genetic Privacy to Open Consent*, 9 NATURE REV. GENETICS 406, 409 (2008).

108. MINI CONSENT, *supra* note 104, § IV:4.2.

109. Baseline trait data may include "date of birth, medications, allergies, vaccines, personal medical history, race/ethnicity/ancestry, and vital signs" as well as family medical history and a facial photograph. FULL CONSENT, *supra* note 103, § IV:4.1.

110. *Id.* § IV.

111. FULL CONSENT, *supra* note 103, § V:5.1-.3. An enrollee may choose not to publish her baseline trait data, but that may make her ineligible to participate in other aspects of the study. *Id.* § V:5.1(b).

112. *Id.* § V:5.4-.5.

113. *Id.* § V:5.5(c).

114. *Id.* § VII:7.1(c).

115. *Id.* § VII:7.1(a)(iii).

116. *Id.*

117. *Id.*

118. *Id.* § X:10.4.

119. *See*, *e.g.*, 42 U.S.C.A .§ 2000ff-1(b)(4) (West 2010).

IV. PRIVACY VERSUS PUBLIC BENEFIT

We have a lot to gain, and a lot to lose, from genetic research. Medical and scientific advancements depend on information about how genes influence traits and diseases. Yet mere collection of DNA for the research that provides that information can compromise an individual's privacy and anonymity. This section highlights some of the competing values and concerns that shape the debate over how best to balance the benefits and risks of online-based genetic research.

A. Privacy Concerns

An individual who chooses to participate in an online-based research study may not be aware of the risks to his privacy, especially if the researcher does not fully inform him of these risks. These risks include loss of anonymity (even if the researcher "promises" confidentiality or anonymity) and genetic discrimination (despite GINA). This section highlights some of the privacy risks associated with participation in an online-based genetic research study.

1. Reidentification: The Loss of Anonymity

One's DNA sequence or carrier status for a particular disease-related gene is undoubtedly information that most people would want to keep private or at least limit the dissemination of. When other types of private information, such as name, social security number, or address, are collected, they are usually anonymized before being shared or released. Typically this means the identifying information is deleted from the rest of the data. For decades this has led to a "robust anonymization assumption,"[120] the belief that by removing certain pieces of information from data the individual to whom the data corresponds would remain anonymous. Unfortunately, recent studies have shown the failure of that assumption. For example, in the mid-1990s a Massachusetts government agency that purchased health insurance for state employees, the Group Insurance Commission ("GIC"), made its patient records available to researchers.[121] First, of course, GIC removed "explicit identifiers" such as name, address, and social security number.[122] Still, one industrious researcher was able to use birth date, sex, and ZIP code to identify the governor of Massachusetts (who had assured the public that patients' information would remain private) from within the

120. Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701, 1706 (2010).

121. *Id.* at 1719.

122. *Id.*

"deidentified" database of patient records.[123] This matching of seemingly anonymous information to the specific individual from whom the information was derived is known as *reidentification*.

Similarly, a 2008 study shocked the genetic research community by proving it was possible to identify a single individual's DNA contribution from a pool of thousands of DNA samples.[124] One commentator explained that "[b]ecause the pool consists of DNA from so many people, the assumption ha[d] been that it would be impossible to identify any one individual's DNA."[125] The National Institutes of Health ("NIH") had been so "confident in the anonymity of pooled genetic data that it recommended it be made public for all researchers to use."[126] After the 2008 study was published, the NIH and other similar research institutions removed some genetic data from their publicly accessible websites,[127] stating that "[t]he greatest concern is that identifying an individual this way could reveal sensitive health information."[128]

Before reidentification concerns surfaced, a 2004 computational study had already determined that as few as 30 SNPs will uniquely identify a single person.[129] This is troubling because up to a million SNPs are usually examined in DTC genetic testing.[130] Thus, someone with access to both individual and public genetic data could identify the individual using just a small set of SNPs.[131]

What this means for participants in DTC genetic testing research studies, as well as for subjects in non-profit research consortium studies, is that there is no safety in numbers. Even though hundreds or thousands of individuals' genetic information may be pooled for so-called genome-wide association studies, a single person could be pinpointed within that pool. To the extent that these research projects *offer* privacy or anonymity, they can no longer *guarantee* either.

Any sharing of genetic data, even deidentified data, with fellow researchers or third parties (as DTC-GTCs acknowledge they may do, and non-profit research consortiums readily do) opens up the possibility

123. *Id.* at 1719-20.

124. Nils Homer et al., *Resolving Individuals Contributing Trace Amounts of DNA to Highly Complex Mixtures Using High-Density SNP Genotyping Microarrays*, 4 PLOS GENETICS 1, 7 (2008), http://www.ncbi.nlm.nih.gov/pmc/articles/ PMC2516199/pdf/pgen.1000167.pdf.

125. Jennifer Couzin, *Whole-Genome Data Not Anonymous, Challenging Assumptions*, 321 SCIENCE 1278, 1278 (2008).

126. *Id.*

127. *Id.*

128. *Id.*

129. Zhen Lin et al., *Genomic Research and Human Subject Privacy*, 305 SCIENCE 183, 183 (2004).

130. *How Does 23andMe Genotype My DNA?*, *supra* note 41.

131. Lin et al., *supra* note 129.

that someone—for good or for ill—will "reverse engineer" the data and defeat intended privacy protections.

2. Informed Consent and Autonomy

Participants in traditional clinical trials usually meet with a health care provider who can explain informed consent forms and answer related questions. In contrast, research participants who enroll online do not interact personally with anyone, let alone a medical professional. If a potential subject has a question, it is more likely to be answered by email than by telephone, certainly not in person, and not necessarily by a medical professional. This physical and emotional separation between parties can create a false sense of security, and the casualness of the online environment in general can create the impression that joining a genetic study is trivial. After all, enrollment takes just a few clicks of a mouse. Computer users are already accustomed to click-through agreements, and there is no reason to believe that they treat informed consent forms any differently than software use agreements.[132] For enrollees who do actually read the consent forms, concerns remain that they do not comprehend the content.[133]

An individual's choice to participate in a research study or otherwise share his or her genetic information should be respected. However, one of the unique aspects of genetic information—that parts of it are common to one's blood relatives—implicates choice and autonomy for family members who do not want to share their genetic information, and perhaps do not want to know any secrets that their or their relatives' DNA sequences might reveal.[134] For example, identical twins have identical DNA sequences, so if one twin contributes his genetic

132. Gabrielle Kohlmeier, *The Risky Business of Lifestyle Genetic Testing: Protecting Against Harmful Disclosure of Genetic Information*, UCLA J.L. & TECH., Fall 2007, at 42 n.149.

133. Amy L. McGuire & Richard A. Gibbs, *No Longer De-Identified*, 312 SCIENCE 370, 371 (2006). "As is widely recognized in a medical context, a signature on a consent document does not necessarily indicate . . . understanding." Katherine Wasson, *Direct-to-Consumer Genomics and Research Ethics: Should a More Robust Informed Consent Process Be Included?*, 9 AM. J. BIOETHICS 56, 57 (2009). Yet lack of comprehension appears to be legally sufficient. Kohlmeier, *supra* note 132, at 40 (citing Canterbury v. Spence, 464 F.2d 772 (D.C. Cir. 1972), for the proposition that legal informed consent does not require comprehension).

134. *See* Amy L. McGuire et al., *Research Ethics and the Challenge of Whole-Genome Sequencing*, 9 NATURE REV. GENETICS 152, 154 (2008) (recommending that "participants . . . include close genetic relatives in decisions about research participation" because "[c]linically relevant . . . information about family members' health risks can be revealed during the course of data analysis"). *Cf.* Erin Murphy, *Relative Doubt: Familial Searches of DNA Databases*, 109 MICH. L. REV. 291 (2010) (arguing against the practice of familial searching—the practice of searching for partial matches to crime scene DNA evidence in order to identify potential relatives of the source, who is often the suspected perpetrator—on numerous grounds, including privacy).

information to a research study, his twin essentially does as well, perhaps without consent or even knowledge.

3. Genetic Discrimination

Despite the protections that GINA does afford, the Act does not prohibit genetic discrimination based on publicly or commercially available information, such as information on the Internet. Nor do its provisions protect against discrimination by financial service providers, by insurers other than health insurers, or by society as a whole. Although GINA's provisions prohibit an employer from using private genetic data to effect discrimination, concerns remain about the employer who "receive[s] genetic information in legal ways, us[es] that information illegally, and then rationaliz[es] such use on legal grounds."[135] Participation in genetic research studies, especially in those that make their data or findings publicly accessible,[136] raises the likelihood that an employer or service provider will intentionally or unintentionally come to possess that information, which in turn increases the likelihood that discrimination will result.

B. Public Benefit: Promoting, Not Impeding, Genetic Research

On the other side of the scale from protecting privacy is many researchers' demand for greater access to and easier sharing of genetic as well as other types of information. This demand arises because genetic information is valuable for research purposes when genotype is linked to phenotype, when what your genes say about you is combined with what you (and your doctor) say about you. For example, researchers cannot determine which genes influence Alzheimer's disease if they do not know which subjects are suffering from Alzheimer's disease, and therefore whose DNA to study. And complex human traits and conditions, those influenced by genetic as well as environmental and other factors, can be accurately studied only by combining individuals' genetic information with their personal information, such as that about medical history, diet, and exercise.[137] Furthermore, longitudinal studies (those that track the same

135. Tan, *supra* note 58, at 117. Employers can acquire genetic information through new-hire medical examinations, such as drug testing and fitness-for-duty exams, as well as through health claims submitted to employer-sponsored health plans, workers' compensation claims, and requests for sick, family, and medical leave. Joanne L. Hustead & Janlori Goldman, *The Genetics Revolution: Conflicts, Challenges and Conundra*, 28 AM. J.L. & MED. 285, 293-94 (2002).

136. This includes at least DTC-GTCs 23andMe and Navigenics, who may publish their results, as well as most non-profit research consortiums. *See supra* Part III.B.

137. John M. Conley et al., *Enabling Responsible Public Genomics*, 20 HEALTH MATRIX 325, 328 (2010).

participants over an extended period of time) require the ability to connect a particular research participant with that person's data in order to see what traits or conditions have changed over time.[138] The need for linked data to advance health care, then, conflicts with concerns about reidentifiability and discrimination that result from ineffective privacy protections.[139]

The Human Genome Project inspired the view that collections of genetic information were "global public goods"[140] such that "all human genomic sequence information . . . should be freely available and in the public domain."[141] This uninhibited sharing of data was considered an "ethical imperative" in order to maximize the data's value and best promote research and development for the benefit of society.[142] Studies such as the PGP follow this open-access model.

Further support for easier access to data comes from scholars who have noted that informed consent requirements obstruct research by decreasing the data pool because not all individuals will consent.[143] The decreased pool leads to biased results because the elderly, illiterate, and those of lower socioeconomic status are less likely to consent,[144] and consent is less likely to be sought from the very ill and very impaired.[145] Those who do participate "represent a self-selected group that could skew research results."[146]

According to some medical professionals, HIPAA's consent requirements have also had a "profoundly negative impact" on research.[147] These requirements include that "patients have to give consent for each use of their data."[148] In practice, this has prevented separate institutions

138. *See id.* at 341.

139. It is worth noting that linking genotypic and phenotypic data does not automatically mean that the individual who supplied that data will be identified, but rather that "the content of the data renders it inherently identifiable." *Id.*

140. Bartha Maria Knoppers, *Consent to "Personal" Genomics and Privacy*, 11 EMBO REP. 416, 417 (2010) (quoting HUGO Ethics Committee, *Statement on Human Genomic Databases of 2002*, 13 EUBIOS J. ASIAN & INT'L BIOETHICS 99 (2003)).

141. *Id.* (quoting HUMAN GENOME ORGANIZATION, PRINCIPLES AGREED AT THE FIRST INTERNATIONAL STRATEGY MEETING ON HUMAN GENOME SEQUENCING (1996)).

142. *Id.* (quoting European Soc'y of Human Genetics, *Data Storage and DNA Banking for Biomedical Research*, 11 EUR. J. HUM. GENETICS 906 (2003)).

143. Julie R. Ingelfinger & Jeffrey M. Drazen, *Registry Research and Medical Privacy*, 350 NEW ENG. J. MED. 1452, 1453 (2004).

144. Taylor, *supra* note 70, at 32.

145. Ingelfinger & Drazen, *supra* note 143.

146. Francis S. Collins & James D. Watson, *Genetic Discrimination: Time to Act*, 302 SCIENCE 745, 745 (2003).

147. Jocelyn Kaiser, *Privacy Rule Creates Bottleneck for U.S. Biomedical Researchers*, 305 SCIENCE 168, 169 (2004) (referring to a survey by the American Association of Medical Colleges).

148. *Id.*

that have already conducted research studies—with patient consent—from pooling their genetic data to look for medical and genetic trends.[149]

Additionally, some commentators have pointed out that SNP data has little predictive value on its own,[150] so we should not rush to prioritize privacy over research progress when dealing with data disclosures. Indeed, genetic determinism—the belief that genes alone determine all of the physical and behavioral characteristics of an individual—is a discredited notion.[151] The vast majority of traits (e.g. height, hair color) and diseases (e.g. cardiovascular disease, colon cancer) arise from a complex mixture of many gene products, environmental factors, and individual choice.[152] If revelation of one's genetic information does not actually reveal that much, then suppressing its release and use harms medical research more than it helps an individual's privacy. At the very least, research and privacy must be balanced instead of requiring that "research has to demonstrate that the public interest substantially outweighs privacy protection."[153]

C. Balancing Concerns

The more information to which researchers have access, and the more complete that information is, the more accurate and reliable their study results are. Ease of communication on the Internet could lead to more study participation by a broader group of people, thereby improving data breadth and the robustness of research conclusions. In turn, this information would contribute to medical and scientific advancements that help the general public. But with that ease of communication comes ease of reidentification, increased opportunities for discrimination, and concerns that participants do not fully appreciate the consequences of a click of a mouse. These potential benefits and harms must be balanced so as to not stifle scientific progress in the preservation of individual privacy, nor sacrifice personal rights for the sake of medical advances.

149. *See id.* (reporting that 14 institutions involved in a prostate cancer genetics study could not pool their data to look for cancer susceptibility genes and instead were limited to sharing summaries of their analyses).

150. Barbara Prainsack et al., *Misdirected Precaution*, 456 NATURE 34, 35 (2008).

151. *See*, *e.g.*, David B. Resnik & Daniel B. Vorhaus, *Genetic Modification and Genetic Determinism*, 1 PHIL. ETHICS & HUMAN. MED. 1, 3-4 (2006), http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1524970/pdf/1747-5341-1-9.pdf.

152. *See*, *e.g.*, *id.* at 4.

153. Knoppers, *supra* note 140, at 418.

V. RECOMMENDATIONS

Unlike the Tuskegee men, direct-to-consumer genetic testing subscribers are not in danger of becoming unwitting participants in a study on untreated syphilis. Likewise, subjects in a non-profit research consortium's genetic study need not fear being intentionally infected with hepatitis, as were the children at Willowbrook. Yet the members of this novel cohort of genetic research participants still deserve respect for their privacy, deference to their autonomy, and protection from discrimination. However, ensuring that these criteria are met should not come at the expense of scientific and technological progress. This final section proposes several approaches to balancing these competing private and public concerns.

A. Addressing Reidentifiability

Ultimately, problems with participation in Internet-based human genetics research studies lie not in the participation itself, but in the unintended consequences of participation. The primary unintended consequence is reidentification—that someone (inside or outside of the research study) will determine not only that a given individual participated, but what the participant's genetic contribution says about that person. If a study organizer could guarantee that his subjects' identities would remain confidential and their genetic data would remain private, the remaining concerns and arguments would be relatively benign. For example, advocates for patients' autonomy might still worry about truly informed consent in the absence of a doctor-patient relationship, or statisticians might bemoan that DTC-GTCs' solicitation of customers as research subjects is skewing the data pool.

Unfortunately, the solution is not as simple as an outright ban on reidentification; because the act of connecting the dots from anonymized data to the corresponding individual cannot itself be detected, a ban would be ineffective.[154] Even if it could be detected and prevented, technology will always leave enforcers one step behind reidentifiers: "[I]n the arms race between . . . anonymization and reidentification, the reidentifiers hold the permanent upper hand."[155]

One proposal for addressing reidentification suggests restricting the flow of (in the instant case, genetic) information such that disclosures

154. Ohm, *supra* note 120, at 1758.

155. *Id.* at 1752. *See also* Bradley Malin & Latanya Sweeney, *How (Not) to Protect Genomic Data Privacy in a Distributed Network: Using Trail Re-identification to Evaluate and Design Anonymity Protection Systems*, 37 J. BIOMEDICAL INFORMATICS 179, 191 (2003) ("[W]e are developing more robust . . . re-identification algorithms.").

occur only to the extent that the benefits outweigh the costs to privacy.[156] In the medical research community, regulators could build upon intrinsic "human networks of trust" that already exist among professionals.[157] This is a great suggestion for data generated from traditional human subject research projects set at a hospital, university, or other collegial institution, but it is unlikely to be effective in an arms-length commercial model such as 23andMe's or Navigenics's collaborations with outside researchers. Also, it is inapplicable to an endeavor such as the Personal Genome Project, because one of the PGP's goals is to determine the results, both positive and negative, of free and unrestricted public sharing of genetic information. If reidentification is one of those results, the investigators would certainly want to know.

Given that the real concerns from study participation stem from privacy and discrimination issues attendant to reidentification, all of the suggested reforms below must be considered in light of, or as attempts to mitigate, the reidentifiability risk.

B. Legislative Reform

Current federal legislation does not go far enough to protect the privacy of participants in online-based genetic research studies. Some reforms should be made, but others may not protect participants enough to justify the negative impact they could have on research. This section proposes and analyzes some possible legislative amendments.

1. Expansion of the Genetic Information Nondiscrimination Act

The protections afforded by GINA should be expanded to cover all genetic information, including that found in the public domain, such as on the Internet, and information derived via reidentification. Specifically, and at a minimum, GINA should provide protection against the use of genetic information that has been shared in support of a valid research project.

Expanding GINA in this way will help ameliorate some of the effects of reidentification, as well as promote personal autonomy and consumer choice (i.e. individuals still get to make decisions about accessing, or not accessing, their genetic information) while simultaneously boosting participation in research due to the (perceived) protection from discrimination based on one's genetic information.

156. *See* Ohm, *supra* note 120, at 1768-69.
157. *Id.* at 1770.

The weakness in this suggestion, however, is that discrimination is hard to prove, and appealing to insurance companies' and employers to "not peek" when the information is readily accessible is unlikely to be successful. Additionally, expanding GINA does nothing to address informed consent and may even mislead potential research participants into thinking that they are completely protected by the law.

However, despite the fact that reforming GINA will not prevent reidentification or completely eliminate genetic-based discrimination, the proposed changes should still be enacted because they fill in legislative gaps from which abuse by an employer or insurer could arise.

2. Revising the Common Rule

Currently, the Common Rule requires Institutional Review Board-approval only for institutions that receive federal funds or who are otherwise federally regulated. Research carried out by a private company that does not produce a pharmaceutical or other regulated product is exempt from the IRB requirements. Therefore, a private Internet-based company such as 23andMe, which offers only genetic sequencing and analysis services, is not required to seek IRB approval for the research projects its research arm, 23andWe, undertakes.

The Common Rule could be expanded to include research projects implemented by companies currently outside of the Rule's reach. Then, like investigators at large academic universities or scientists at pharmaceutical companies, researchers at companies like 23andMe would have to draft research proposals and seek IRB approval before commencing genetic information studies. Their studies would be evaluated to ensure that the risks to the subjects are both minimized and reasonable in relation to the anticipated benefits. Selection of subjects would need to be equitable and informed consent would have to be obtained. The consent form would need to include a description of risks to the subject, an identification of benefits to the subject or to others, and a statement that participation is voluntary and the subject may withdraw at any time. A statement describing the extent of confidentiality of records identifying the subject will also have to be included, as the existence of adequate provisions for protecting the privacy of subjects and maintaining confidentiality of data is one of the criteria considered by IRBs when they evaluate proposals.

This is likely to be a politically unpopular approach, as it would greatly expand the scope of oversight to currently unregulated private companies—those that receive no federal funding and do not produce a good regulated by the federal government. The tests offered by DTC-GTCs (and hence the vehicle by which the public participates in one of

their research studies) are currently not regulated by the FDA. Several scholars have argued for the tests to fall under the auspices of the FDA because a similar (if not as comprehensive) test ordered by one's personal physician *does* require FDA regulation and approval.[158] Additionally, or alternatively, DTC genetic tests could be regulated by the Clinical Laboratory Improvement Amendment of 1988 ("CLIA").[159]

Changes to the regulatory scheme do appear to be on the horizon,[160] but even if DTC-GTCs' tests become subject to FDA regulation, and if the companies are required to seek IRB approval for their research studies,[161] what change will be effected? Historically, the "harm" that IRB committees look for, and try to ensure that investigators minimize, is physical harm.[162] DTC research participants face no physical harm or intrusion; they spit into a tube. The psychological and emotional harm that could come from learning something devastating about one's genetic profile—such as carrying the gene for the aggressive and fatal Huntington's disease—is no more extensive than if the subject opted only for the testing service and chose not to participate in the study.

158. Scholars have been concerned that DTC genetic testing services are not safe or efficacious, have not been clinically or analytically validated, and even have suspect clinical utility. *See*, *e.g.*, Lauren B. Solberg, Note, *Over the Counter But Under the Radar: Direct-to-Consumer Genetics Tests and FDA Regulation of Medical Devices*, 11 VAND. J. ENT. & TECH. L. 711, 720, 722 (2009). More problematic is that no federal agency is responsible for ensuring that the tests meet any particular quality standards. *See*, *e.g.*, Jennifer A. Gniady, Note, *Regulating Direct-to-Consumer Genetic Testing: Protecting the Consumer Without Quashing a Medical Revolution*, 76 FORDHAM L. REV. 2429, 2436-37 (2008). The Federal Trade Commission can regulate advertisement of DTC tests, but critics contend the Commission has not taken action against advertisements that may be false or misleading. Solberg, *supra*, at 722. If regulated, DTC genetic tests would likely fall under the guise of medical devices, which the Food and Drug Administration regulates. *See* Gail H. Javitt, *In Search of a Coherent Framework: Options for FDA Oversight of Genetic Tests*, 62 FOOD & DRUG L.J. 617, 618-19 (2007).

159. CLIA governs all laboratories that perform tests designed to provide information about a person's health. Specifically, it aims to regulate protocols, reagents, quality control procedures and even the qualifications of laboratory personnel. Although CLIA governs many of the tangible aspects of genetic testing, it does not ensure any specific standards for accuracy, reliability, or clinical validity. Gniady, *supra* note 158, at 2440; Douglas A. Grimm, *FDA, CLIA, or a "Reasonable Combination of Both": Toward Increased Regulatory Oversight of Genetic Testing*, 41 U.S.F. L. REV. 107, 121 (2006).

160. Dan Vorhaus, *Update: FDA Taking Another (Public) Look at DTC Genetic Tests*, GENOMICS L. REP. (Feb. 8, 2011), http://www.genomicslawreport.com/index.php/tag/dtc (reporting that the FDA announced a public meeting, set for March 8, 2011, to discuss DTC genetic tests).

161. This could occur as a direct consequence of FDA regulation or via expansion of the Common Rule.

162. *See*, *e.g.*, 45 C.F.R. § 46.102(i) (2010) (defining minimal risk—which IRBs seek to achieve—as "mean[ing] that the probability and magnitude of harm . . . anticipated in the research [is] not greater . . . than [that] ordinarily encountered in daily life or during . . . routine physical or psychological examinations").

(Because DTC-GTCs recruit study participants via their existing consumer pool, all potential consequences of study participation must be compared to a baseline of receiving only the genetic testing service, rather than by using "no genetic testing" as a baseline and comparing it to participation in a study that includes discovering genetic information about oneself.[163]) Only under a much broader conception of *harm* might research proposals by DTC-GTCs be denied by IRBs based on "unjustifiable harm" grounds. That is, to the list of possible adverse consequences from participating in a research study, IRBs would have to add genetic discrimination and other social harms,[164] especially those attendant to reidentification. However, those harms are "difficult to forecast" and the "most significant risks . . . may presently be unknown."[165] Ultimately, then, the Common Rule's "limited conception of risk leaves it ill-equipped to protect human subjects"[166] in DTC-based research.

Proper informed consent is also an important criterion for gaining IRB approval. 23andMe's online consent form already meets the requirements for identifying risk to the subject (although it could be beefed up[167]) and the benefits to the participant or others.[168] The consent form also includes the requisite statement that participation is voluntary and the subject may withdraw at any time.[169] However, this is qualified by acknowledging that it takes time to withdraw, and participants' genetic information that has already been used for research purposes cannot be withdrawn.[170]

The Common Rule's IRB provision already applies to academic or otherwise publicly-funded research consortiums, except when data is collected from the public domain or when the data is from subjects that cannot be identified. We have seen the fallacy of the claim that subjects "cannot" be identified, so the Rule could be amended to eliminate these (the public domain and de-identified information) exceptions. However, if follow-on investigators want to analyze existing data that was previously collected for a different purpose, these proposed amendments would prohibit that use because the original data contributor (the human

163. *But see 23andMe Parkinson's Community: Strength in Numbers* and *23andMe Sarcoma Community: A Patient-Driven Revolution in Sarcoma Research*, *supra* note 46 (offering genetic testing and research enrollment for free—compared to at least $199 for the standard testing kit—for Parkinson's disease and sarcoma patients, which could be considered coercive).

164. *See* Conley et al., *supra* note 137, at 363.

165. *Id.* at 363-64.

166. *Id.* at 364.

167. *See infra* Part V.D.

168. *Consent Document*, *supra* note 43.

169. *Id.*

170. *Id.*

subject) most likely did not consent to the use of his information in a second study.[171] The follow-on investigator would then have to plan and execute her own study, which could be cost-prohibitive for academic researchers already struggling for research funding. Additionally, even if cost were not a factor, increasing the number of studies also increases the amount of data collected, which thereby increases the chances for breaches of privacy and confidentiality of that data, and that is certainly not a beneficial solution.

Overall, these changes to the Common Rule could restrict research without enough counterbalancing privacy protections to justify the stifling effects. Having to seek research approval from an IRB would force DTC-GTCs to create comprehensive research proposals and informed consent documents, and if the IRBs, for their part, started considering reidentification and its consequences as potential harms, the resulting documents could be quite robust. Yet informed consent procedures are already known to inhibit research.[172] Furthermore, the amendments would do nothing to prevent reidentification, and could address it only if and to the extent that IRBs weighed it as a harm. Finally, the proposed changes also would not address or attempt to prevent genetic discrimination.

C. Protective Approaches

Other approaches to safeguarding genetic privacy, while still promoting research, attempt to add an additional layer of protection either between a subject and those who work with the subject's genetic information (stewardships) or between a subject and those who might try to force revelation of a subject's identity (Certificates of Confidentiality). These approaches are explored in turn below.

1. Stewardships

As mentioned earlier, "the utility and privacy of data are intrinsically connected"[173] such that genetic data must be linked to some personal information in order to benefit research. Researchers often use codes to keep a subject linked (keyed) to his or her data (i.e. genetic information) without employing a ready identifier such as the subject's name. The goal of these systems is to maintain privacy without completely losing the importance and value of linked information.

171. HIPPA creates similar problems. *See supra* text accompanying notes 147-149.
172. *See supra* Part IV.B.
173. Ohm, *supra* note 120, at 1705.

Some writers have suggested that independent third parties should hold the keys.[174] These stewards of the identifying linkages would not only hold the keys, but "turn" them, using them to "run queries that have passed independent ethical scrutiny."[175] Although this approach keeps any individual researcher or group of researchers (e.g. sample collectors, DNA sequencers, data analyzers) from knowing "too much," it is vulnerable to security breaches. These breaches may include "misconduct by the person who retains the key . . . , theft of the key by hackers, and the loss of laptops or other storage devices that contain the keys."[176]

More importantly, stewardships do nothing to address reidentification. Possessing a key would actually make reidentification much easier, as the point of a key is to connect research subjects to the samples they contribute, and lacking the key would not prevent reidentification. Stewardships also do not address other privacy concerns such as improving informed consent and autonomy or reducing the potential for genetic discrimination. Although stewardships are more of an inconvenience to researchers rather than an actual impediment, they do so little, in practice, to protect privacy that they do not create a viable balance between the two.

2. Certificates of Confidentiality

Certificates of Confidentiality for research studies using genetic information are another means by which participants' confidentiality could be (partially) protected. Certificates allow researchers to protect the privacy of their subjects by "withholding from all persons not connected with the . . . research the names or other identifying characteristics" of the study participants.[177] Specifically, researchers may not be compelled by law enforcement officials or in the context of any legal proceeding to identify subjects.[178] In a 1973 murder investigation, a Certificate successfully prevented disclosure of drug treatment program participants despite a grand jury subpoena for their photographs.[179]

However, the strength of Certificates was tested again in 2006 with less success.[180] A defense attorney was able to obtain access to

174. Taylor, *supra* note 70, at 33; Ingelfinger & Drazen, *supra* note 143.

175. Taylor, *supra* note 70, at 33.

176. Conley et al., *supra* note 137, at 347 (internal citations omitted).

177. 42 U.S.C. § 241(d) (2006).

178. *Id.*

179. People v. Newman, 298 N.E.2d 651, 657 (N.Y. 1973) (vacating adjudication of contempt against Dr. Robert Newman, then director of the New York City Methadone Maintenance Treatment Program, for refusing to comply with subpoena).

180. Laura M. Beskow et al., *Certificates of Confidentiality and the Compelled Disclosure of Research Data*, 322 SCIENCE 1054, 1054-55 (2008).

information about his client's participation in a research study, which had included collection of genetic information, conducted under a Certificate.[181] The court clearly was unfamiliar with Certificates and their intended imperviousness.[182]

Perhaps before courts can be expected to uphold the unimpeachable status of Certificates, additional legislation, and education, is necessary. However, the NIH still puts faith in their effectiveness and "explicitly encourages investigators" to obtain a Certificate as protection against compelled disclosure of participants in genome-wide association studies.[183]

In the context of research studies by DTC-GTCs, however, Certificates will usually be unavailable because they are limited to IRB-approved research.[184] In fact, most DTC-GTCs warn their clients that they may disclose personal information, including genetic information, if required to do so by law.[185] More importantly, the utility of Certificates for online-based genetic research studies will be limited because the greatest risks to individual privacy and confidentiality come from reidentification by technologically savvy computer users rather than from zealous attorneys. The request of a bioinformatician who demands that study participants be reidentified may be denied due to the presence of a Certificate, but one who does not bother to seek permission will not be thwarted. Additionally, Certificates do not address the privacy concerns of promoting informed consent and reducing genetic discrimination. Overall, Certificates are "insufficient to underwrite absolute privacy promises."[186]

D. Interactive Informed Consent

If we acknowledge the real threat of reidentification, but also that it is almost impossible to stop, we must concomitantly recognize that genetic research participants are vulnerable to breaches of privacy and anonymity. To ensure that the participants themselves are fully aware of this vulnerability, enhanced informed consent procedures should be implemented by Internet-based research studies. First, to approach full

181. State v. Bradley, 634 S.E.2d 258, 260 (N.C. 2006).

182. Beskow et al., *supra* note 180, at 1054.

183. NAT'L INSTS. OF HEALTH, POLICY FOR SHARING OF DATA OBTAINED IN NIH SUPPORTED OR CONDUCTED GENOME-WIDE ASSOCIATION STUDIES (GWAS) (2007), *available at* http://grants.nih.gov/grants/guide/notice-files/NOT-OD-07-088.html.

184. *But see supra* Part V.B.2 (suggesting an expansion of the Common Rule to cover DTC-GTCs, thus requiring IRB approval for research, and therefore opening up the availability of Certificates of Confidentiality).

185. *See*, *e.g.*, *Terms of Service*, *supra* note 75, § 8.

186. Conley et al., *supra* note 137, at 349.

comprehension of risk, ignorance of risk must be acknowledged. Informed consent documents must fully disclose not only all known risks, including reidentification and genetic discrimination, but also state that there are unknown risks,[187] which may even be the most significant ones.[188] Second, these risks must be prominently displayed—perhaps in bolded text and all capital letters like waivers of liability are required to have—not buried in the fine print of an unread privacy policy.

The first ten enrollees in the PGP were required to have advanced degrees in genetics or similar fields to ensure that they fully appreciated the risks of their genetic information being made publicly available. The next wave of PGP subjects must submit to a lengthy and stringent screening process, including an entrance exam, to help ensure that they are similarly aware of the risks associated with public access to their genetic information. Other research consortiums and DTC-GTCs could also be required to include a screening process or entrance exam for their study participants.[189]

Another way to ensure that potential subjects fully appreciate the risks of participating is to prohibit companies from enrolling participants solely via the Internet. Interested individuals would be required to contact, preferably by phone rather than email, a genetic counselor on staff at the DTC-GTC[190] or research consortium who talks one-on-one with the subject, evaluates comprehension of possible risks, and answers questions.

Less radical, and more in sync with the online environment, is to leave consent procedures solely online, but implement something more robust than click-through forms. Currently, for example, a subject views a scroll-through screen full of caveats which he or she likely does not read before clicking the "I Consent" button at the bottom. Instead, subjects could be prompted to type a provided sentence that expresses comprehension of risk: "I, [enter your name], understand that by participating in this study, privacy and anonymity of my genetic information cannot be guaranteed and that someone, not authorized by me, may figure out that I participated in this study and may learn the content of my genetic information which could lead to discrimination or other negative effects."

187. *Id.* at 354.

188. *See supra* text accompanying note 165.

189. *But see supra* Part IV.B (noting that consent procedures can decrease and therefore bias data sets).

190. Pathway Genomics currently appears to be the only DTC genetic testing company that provides its customers with access to genetic counselors pre- or post-testing, but the company does not have a research arm. PATHWAY GENOMICS, *supra* note 37.

The PGP is already using a similarly interactive tool, the multiple choice quiz, which interested participants must complete perfectly before being considered for enrollment. DTC-GTCs and research consortiums aside from the PGP could employ comparable online questionnaires that are designed to test comprehension of risk of study participation. Successful completion of these quizzes would help ensure that interested subjects read the consent forms and privacy statements well enough to appreciate the risks of their participation.

Overall, enhanced informed consent procedures strike a balance between protecting privacy and promoting research. Although consent procedures can inhibit research,[191] these enhanced requirements are not nearly as burdensome on researchers as the expansion of the Common Rule could be.[192] Additionally, although these procedures do not stop reidentification or genetic discrimination, they do acknowledge those potential adverse consequences. Armed with as much information as possible, research participants are less likely to suffer "post-enrollment regret."[193]

CONCLUSION

Both public and private Internet-based research studies provide the opportunity for individuals to contribute their genetic information to scientific and medical research projects, but the Internet also provides an environment in which individual privacy and anonymity are almost impossible to guarantee. Because so much can be gained from genetic research in the way of medical, scientific, and even bioinformatics advancements, participation should not be prohibited, but rather should be protected where possible and fully informed where protection is not possible. At a minimum, Congress should expand GINA to cover all genetic information, no matter how it is obtained. More importantly, online research studies should implement more robust enrollment procedures based on full disclosure of known potential risks (plus acknowledgment that others are unknown) and active, even interactive, acceptance of those risks by participants. Only then can we realize a true research revolution.

191. *See supra* Part IV.B.
192. *See supra* Part V.B.2.
193. Conley et al., *supra* note 137, at 354.