

Public Opinion about the Importance of Privacy in Biobank Research

David J. Kaufman,^{1,*} Juli Murphy-Bollinger,¹ Joan Scott,¹ and Kathy L. Hudson¹

Concerns about privacy may deter people from participating in genetic research. Recruitment and retention of biobank participants requires understanding the nature and magnitude of these concerns. Potential participants in a proposed biobank were asked about their willingness to participate, their privacy concerns, informed consent, and data sharing. A representative survey of 4659 U.S. adults was conducted. Ninety percent of respondents would be concerned about privacy, 56% would be concerned about researchers having their information, and 37% would worry that study data could be used against them. However, 60% would participate in the biobank if asked. Nearly half (48%) would prefer to provide consent once for all research approved by an oversight panel, whereas 42% would prefer to provide consent for each project separately. Although 92% would allow academic researchers to use study data, 80% and 75%, respectively, would grant access to government and industry researchers. Concern about privacy was related to lower willingness to participate only when respondents were told that they would receive \$50 for participation and would not receive individual research results back. Among respondents who were told that they would receive \$200 or individual research results, privacy concerns were not related to willingness. Survey respondents valued both privacy and participation in biomedical research. Despite pervasive privacy concerns, 60% would participate in a biobank. Assuring research participants that their privacy will be protected to the best of researchers' abilities may increase participants' acceptance of consent for broad research uses of biobank data by a wide range of researchers.

Introduction

Large, prospective cohort studies that use DNA samples annotated with varying amounts of medical, lifestyle, and environmental information are becoming standard research tools for examining the effects of genes, environment, and lifestyle on common complex diseases,^{1–5} but participants' concerns about the privacy of their information may interfere with recruitment of the large representative samples that are needed.^{6,7} The privacy of cohort-study and biobank participants' information is usually protected by removal of personal identifiers before data are made available to researchers. Removing fields that include names, addresses, birthdates, ages, phone and fax numbers, driver's license and identification numbers, URLs, IP and email addresses, photographs, fingerprints, and other biometric identifiers is often viewed as an effective deidentification of data.⁸ However, emerging forensic methods have shown that a third party with access to a sample of an individual's DNA could use DNA sequence data of the type collected and shared by genetic biobanks to determine that the sample belongs to a biobank participant.⁹ If the DNA data shared by biobanks are accompanied by deidentified individual-level health data, this method could lead to a more clear-cut reidentification of study participants and breach of their privacy.¹⁰ In order for researchers to recruit individuals successfully, potential participants must believe that the privacy and confidentiality of their information will be adequately protected, believe that the benefits of participating in research outweigh the risks associated with potential losses of medical and genetic privacy, or not be concerned about privacy issues.¹¹

The National Institutes of Health (NIH) and other federal agencies have contemplated the creation of a large biobank that would recruit a nationwide representative sample of at least 500,000 people. A proposed study design¹² would establish recruitment sites across the country for the collection of biospecimens and the performance of a comprehensive baseline exam on each participant. Hospital and outpatient records might be used in corroborating observed or reported phenotypes. Specimens would be sent to laboratories where DNA would be isolated and genotyped. Information from the genetic analyses, residual biospecimens, and baseline exam data would be deidentified and stored in a national repository, where they would be merged with regularly updated clinical data from participants for ten or more years, creating a national biobank. Deidentified, coded data from the biobank would be made available to institutional research board (IRB)-approved investigators for a wide range of analyses.^{12,13}

Most cohort studies and biobanks like the one outlined above are observational and do not involve experimental treatments. Participants generally undergo minimally invasive sample collection with little risk of physical harm and provide personal information through biological samples, physical and other exams, medical records, or surveys. As in most such cohort studies, participants in the proposed cohort study would initially sign a consent form that clearly outlines the data-sharing policies of the biobank. It has not been determined whether the large cohort study would seek one-time consent from participants for all IRB-approved studies, whether it would allow participants the opportunity to provide separate consents for each particular study, or whether another model of

¹Genetics and Public Policy Center, Johns Hopkins University, Washington, DC 20036 USA

*Correspondence: dkaufma2@jhu.edu

DOI 10.1016/j.ajhg.2009.10.002. ©2009 by The American Society of Human Genetics. All rights reserved.

consent, such as approval for broad disease categories or categories of research, would be used.

If the consent document accurately and clearly reflects the data-sharing policies of the proposed biobank, then prospective participants have an opportunity to consider privacy risks and make decisions about participation accordingly. For those who consent to take part in the study, most if not all instances of the sharing of participants' data by the biobank should be viewed as acceptable. It is, however, possible to envision instances of data sharing or release that could be viewed by a participant as a violation of privacy. Participants may misunderstand or underestimate the extent to which they have consented to share their data¹⁴⁻¹⁷ and subsequently view some legitimate data sharing as a loss of privacy. Large DNA-sequence data files that are made publicly available could theoretically be used by researchers outside of the biobank's purview. Finally, unintentional disclosure of biobank study data could result from carelessness or data theft. As biobanks collect increasingly larger amounts of genomic and other data and grant access to more diverse groups of researchers and others for a broader range of purposes, the risk to research participants of significant privacy losses may increase.¹⁸

The magnitude of harm occasioned by a violation of privacy may depend on the types and clinical relevance of the disclosed data and findings, the likelihood that the individual could be identified from the data, and the additional harm that could accrue as a result.¹⁹ However, even when an individual cannot be identified, or his or her study data cannot be used for harm, the perception of a loss of medical and genetic privacy may be harmful in and of itself.²⁰ Failure to maintain research subjects' personal privacy²¹ may prevent the subjects from maintaining and controlling social relationships that are affected by the information shared. For example, a person might choose to share different information with his doctor and his child in order to receive relevant care from his doctor and minimize worry of his child. The ability to control who knows what about us allows us to alter our behavior with different people.²² Losing this control can erode personal autonomy and the dignity and worth of individuals.²³ Regardless of the actual risk or magnitude of additional harms, some people will forgo participating in medical research and avoid seeking medical care²⁴ and genetic testing^{25,26} in order to prevent unwanted disclosures of their medical and genetic information.

When the general public considers the importance of privacy in the context of participation in genetic and other biomedical research, they may be concerned with several aspects of the study and how it will protect their personal information. Concerns about privacy are multifaceted and may relate to the type(s) of information being collected and shared, the degree of control that participants will have over access to their information, the types of researchers (and other parties) that may have access, as

well as what, besides research, could be done with the personal information for harm or exploitation of study participants. Much of the existing literature on the importance of privacy in clinical genetics and genetic research has focused on the last of these facets, in large part because of the risks and concerns about employment and insurance discrimination against individuals, group discrimination based on genotype frequencies, and the implications of a patient's clinically relevant genetic information for family members. This body of work has consistently shown that use of clinical genetic services is tempered by public fears that a loss of genetic privacy may result in discriminatory decisions by employers or insurers.²⁷ Studies of potential genetic-research participants,²⁸ patients,²⁹ and publics³⁰ interested in genetic testing and of physicians who might refer patients for genetic tests³¹ all showed a reluctance to submit to testing or to refer patients for genetic testing because of fears about privacy violations. A small number of studies have examined privacy concerns related to participation in genetic research that would collect, analyze, and store participants' DNA samples. Findings about the relationship between privacy concerns and willingness to participate in genetic research varied considerably, but in nearly every study, privacy or confidentiality was mentioned as a primary concern of participants.^{14,15,17,32-37}

In order to assess the importance of privacy concerns in the public's support for and willingness to participate in the proposed national cohort study, how these concerns weigh against the potential benefits of participation, and whether privacy concerns relate to people's preferences about aspects of the study design, including how consent is obtained, who collects study data, and with whom the data are shared, we carried out a national survey of the general public. In addition to measuring concern about the potential for misuse of participants' personal information and study data, the survey also measured the relative importance of protecting genetic information in comparison to other types of health and personal information and whether concerns about privacy are related to the affiliation (academic, government, or industry) of researchers who might use or see cohort study data.

Subjects and Methods

On the basis of themes emerging from 15 focus groups,^{13,38} a 177-item online survey was drafted for the collection of data on public opinions about the proposed national cohort study. A large pilot study was fielded for evaluation of the online survey for length and clarity. Sample selection and survey administration were managed by the firm Knowledge Networks (KN).³⁹ A total of 7978 potential participants 18 years of age and older were randomly sampled from KN's panel of U.S. residents, including oversamples of black non-Hispanics, Hispanics, and people living outside of metropolitan statistical areas. KN selects its master panel by using list-assisted random-digit dialing to provide a probability-based sample to draw from. The main survey was fielded online

between December 14, 2007, and January 31, 2008. Information previously collected by KN regarding panel members' demographics and backgrounds was added to the data set. The survey was judged by the Johns Hopkins University IRB as imposing only minimal risks on participants and was qualified as research exempt from human subject review (application no. NA-00014533). The survey first asked several general questions about health care, privacy, and medical records, then participants were shown a video describing the proposed cohort study. The video description stated that data collected from cohort-study participants would be coded before being entered into a national database for use by researchers.⁴⁰ Respondents were then asked about their support for the proposed study, concerns about privacy, preferences about study design, and willingness to participate.

At the end of the survey, respondents were shown one of eight randomly selected study-design scenarios and asked whether they would be willing to participate in the cohort study if asked. The eight scenarios varied with respect to three factors: study burden (low burden, which consisted of a half-day exam and a yearly questionnaire, and high burden, which added a home visit by researchers and diet and exercise journals), return of individual research results (returned or not), and compensation for participation (\$50 or \$200).⁴¹ A detailed description of the scenarios and the definition provided for the term "individual research results" can be found in the [Appendix](#).

Weights corresponding to U.S. Census demographic benchmarks were calculated so that the oversamples were accounted for and bias from sampling error was minimized. Willingness to participate in the study and opinions about the study were measured with four-point Likert scales. Data were analyzed in the SUDAAN software package,⁴² which permits correction for the survey sampling scheme when judging hypothesis tests. Multiple logistic regression was used for the examination of demographic factors associated with beliefs about privacy and associations between privacy beliefs and people's willingness to participate in the study, their preferences regarding sharing samples and information, and the importance of policies for the protection of study data. Analyses of the entire data set were adjusted for age (continuous), education (categorical), household income (categorical), gender, and race or ethnic group.

Results

A total of 7978 people were contacted, and 4659 provided valid responses, for a response rate of 58.4%. The margin of error on opinion estimates based on the entire sample is $\pm 1.6\%$ after weighting of the data and correction for sampling design. Demographic characteristics of the surveyed population are found in [Table 1](#). Both weighted and unweighted demographic distributions of the sample were comparable to year 2000 U.S. Census figures.⁴¹

General Beliefs about Privacy

Before survey participants were introduced to the proposed cohort study, general questions were asked about privacy. A total of 88% were very or somewhat concerned about the privacy of their financial information, and 79% were concerned about the privacy of their medical information. Black non-Hispanics ($p = 4 \times 10^{-4}$), American Indians and

Alaska Natives ($p = 0.003$), and participants who self-identified as being of two or more races ($p = 0.04$) were all significantly more likely than white non-Hispanics to say that they were concerned about the privacy of their medical information ([Table 1](#)). Respondents over the age of 30 were also more likely to express concern about medical privacy (adjusted $p = 0.002$). Concern about medical privacy did not differ significantly by income, education, or gender.

One-third of respondents agreed with the statement that "some information in medical records is sensitive and needs extra privacy protections." The remainder agreed that "all medical information should have the same privacy protections." People who felt that some aspects of a medical record deserve extra protection ($n = 1574$) were shown a list of medical-information topics and asked which (if any) they felt needed extra privacy protection. ([Figure 1](#)) Nearly all respondents felt that social security numbers deserved extra privacy protection, whereas 44% would protect genetic test results and 28% were concerned about family histories.

Nine in ten respondents said that they would be somewhat (26%) or very (64%) concerned about "protecting my privacy" if they were to take part in the proposed cohort study. The proportion concerned about protection of privacy in the study was consistent across all demographic groups, and most groups were more concerned about protecting their privacy in the study than about protecting the privacy of medical records ([Table 1](#)). Respondents were also asked about other related concerns. Three in four were concerned about "the government having [their] samples and information," and 56% were concerned about "researchers having [their] samples and information." Black non-Hispanic respondents, women, and those without a college degree were all significantly more likely to be concerned about both researchers and the government having access to their personal information ([Table 1](#)). Asian Americans and those with incomes under \$25,000 were significantly more likely to say that they were concerned about study researchers having their samples. Finally, 37% of respondents said that they would be afraid that the information collected by the study could be used against them. Black non-Hispanics and participants under the age of 60 were significantly more likely to share this feeling. ([Table 1](#)).

Data Sharing, Consent, and Privacy

Despite the widespread privacy concerns described above, 73% of respondents said that they would definitely or probably "sign a consent to provide past medical records" as part of the study. After adjustment for other demographic factors, black non-Hispanics (63%, $p = 0.001$) would be less likely than white non-Hispanics (74%) to consent to provide past medical records, whereas those earning more than \$75,000 (78%, $p = 0.03$) and those with bachelor's degrees (80%, $p = 3 \times 10^{-6}$) would be more likely to consent. Among those concerned with

Table 1. Survey Respondents' Concerns about General Privacy and Issues of Privacy Related to the Large Cohort Study

	Unweighted N	Concerned about Privacy of Their Medical Information (%)	Concerned about Protecting Their Privacy in the Study (%)	Concerned about Study Researchers Having Their Samples and Information (%)	Concerned about Government Having Their Samples and Information (%)	Afraid that Data Collected by the Study Could Be Used against Them (%)
Total	4659	79	91	56	75	37
Gender						
Men	2247	78	90	53	73	39
Women	2412	81	91	59 ^a	78 ^a	36
Race or Ethnic Group						
White, non-Hispanic	2798	77	91	53	74	35
Black, non-Hispanic	774	89 ^a	94	74 ^a	84 ^a	44 ^a
Hispanic	867	79	87	56	79	41
American Indian or Alaska Native	35	94 ^a	91	52	71	41
Asian American or Pacific Islander	71	83	89	64 ^a	69	45
2+ races	114	87 ^a	93	52	73	45
Age Group						
Age 18–29	838	71 ^a	90	57	76	41
Age 30–44	1207	82	90	54	73	43
Age 45–59	1791	84	93	56	78	37
Age 60+	823	79	89	57	75	28 ^a
Education						
< High-school	502	78	90	59	74	35
High school	1380	83	90	59	78	37
Some college	1406	78	92	58	78	38
Bachelor's degree or higher	1371	79	91	50 ^a	70 ^a	36
Household Income						
< \$25,000	959	78	91	62 ^a	79	35
\$25,000–\$49,999	1499	81	89	57	77	39
\$50,000–\$74,999	1071	78	91	55	75	40
\$75,000+	1130	80	92	48	69 ^a	34
Concerned about privacy of medical information	3694	100	95 ^a	62 ^a	81 ^a	41 ^a
Not concerned about privacy of medical information	952	0	74	32	54	23

^a This demographic category differs significantly from other categories in the group ($p < 0.05$), with correction for multiple comparisons among categories and adjustment for all other variables shown in the table.

medical privacy, 70% would consent to provide past medical records, compared to 81% of those unconcerned about medical privacy ($p = 0.006$). Only 56% of respon-

dents who said that they were afraid that study data could be used against them would consent, compared to 83% who were not afraid of data misuse ($p = 1 \times 10^{-9}$). In a

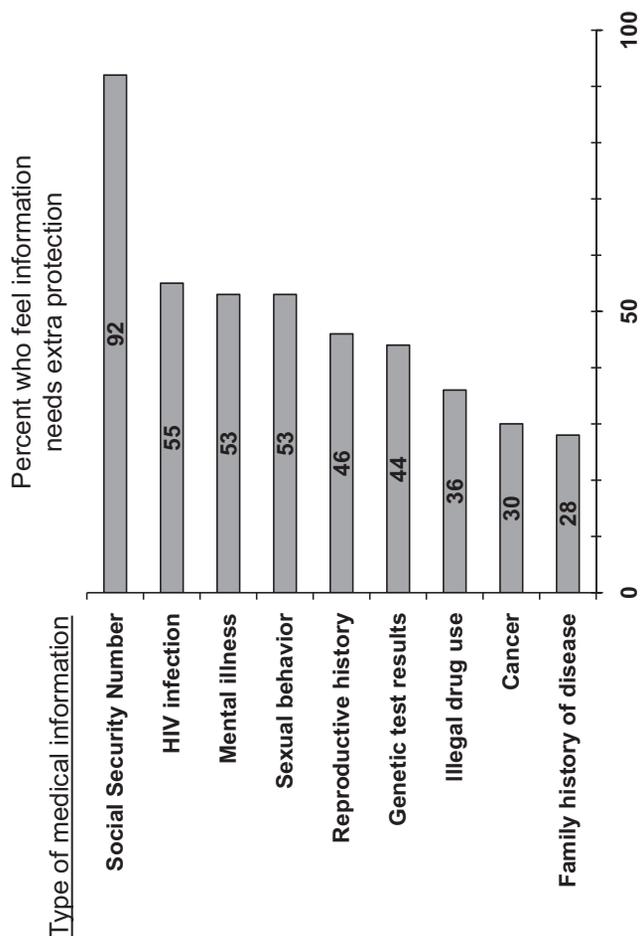


Figure 1. Percentage of Respondents Who Believe Specific Types of Medical Information Deserve Extra Privacy Protections Percentage of survey respondents who believe that various types of information in a medical record need extra privacy protections, among those who believe that some types of medical information need extra protections (n = 1574).

logistic-regression model adjusting for concern about medical privacy and fear that study data could be used against them, black non-Hispanics ($p = 0.02$) and those without a bachelor's degree ($p = 3 \times 10^{-7}$) were still significantly less likely to consent to provide past medical records.

A total of 82% said that they would definitely or probably “give blood for genetic and lab tests once during the initial physical exam.” After adjustment for other demographic factors and for the fear that study data would be used against them, black non-Hispanics (73%, $p = 0.0005$) and American Indians and Alaska Natives (65%, $p = 0.03$) would be less likely than white non-Hispanics (83%) to consent to provide a blood sample. Women (84% $p = 0.001$), participants who self-identified as being of two or more races (92%, $p = 0.006$), those earning more than \$75,000 (88%, $p = 0.04$), and those with bachelor's degrees (87%, $p = 0.009$) would be more likely to provide a sample. A total of 71% of respondents who said that they were afraid that study data could be used

against them would provide a sample, compared to 88% who were not afraid of data misuse ($p = 1 \times 10^{-9}$).

In addition to being asked about willingness to provide past medical records and a blood sample, respondents were asked what types of researchers should be allowed to submit research projects for use of the biobank's samples and information (Table 2). A total of 92% would give permission to academic and medical researchers, 80% would allow government researchers to use the samples and information, and 75% would allow pharmaceutical-company researchers to use their samples and information. Nearly half (49%) would be willing to have their deidentified information and research results “made available on the internet to anyone.”

The purpose of the large cohort study would be to provide a resource that researchers could use to study a wide variety of phenotypes. Participants in the study might be asked at the outset to provide consent that would allow their samples to be used in all types of research approved by the study, or they could be asked for consent each time that a project that would use the biobank data is approved. Respondents were asked for their preference about how the cohort study should obtain consent to share participants' samples and information with researchers. Nearly half (48%) would prefer to give permission once, at the beginning of the study, for all research approved by an oversight panel. Slightly fewer (42%) wanted to be asked permission for each research project separately, and 10% preferred to select categories of research (i.e., cancer or diabetes) for which they would or would not let their samples be used.⁴³ After adjustment for demographic factors (Table 1), people who were concerned about privacy were less likely to favor blanket consent than were respondents who were not concerned about the privacy of their medical records (45% versus 57%, $p = 0.0003$). When respondents were questioned about how being asked for consent for each study would make them feel, 81% agreed that it would make them feel “respected and involved,” and 74% agreed that they would feel that they “had control.”

Willingness to Participate in a National Cohort Study

Most survey respondents (84%) supported the general idea of the large cohort study. At the conclusion of the survey, each participant was randomly selected to view one of eight different study scenarios, as described in the Methods section (see Appendix, as well). With the responses to all eight scenarios combined, 60% said that they definitely or probably would be willing to participate in the study if asked, given the scenario that they read. Support for the study and willingness to participate did not vary substantially across the demographic factors found in Table 1.⁴¹

Willingness to participate in the cohort study was related to the types of researchers that respondents would allow access to their samples and information. For example, 66% of people who said that they would allow U.S. academic researchers use of the collected data would be

Table 2. Results of Multiple Logistic Regressions Examining Demographic Differences in Willingness to Share Data

	"I Would Allow These Researchers to Use My Samples and Information for Research."				"If I Could Not Be Identified, I Would Be Willing to Have My Information and Research Results Available on the Internet to Anyone."			
	Academic or Medical Researchers in the United States		Government-Funded Researchers		Pharmaceutical-Company Researchers		Agree ^b	p Value
	Agree ^a	p Value	Agree ^a	p Value	Agree ^a	p Value		
Gender								
Men	92%	0.23	81%	0.39	75%	0.95	53%	0.01
Women	91%		78%		73%		45%	
Household Income								
\$0–24,999	89%	0.004	77%	0.02	72%	0.47	49%	0.91
\$25,000–49,999	90%		76%		75%		47%	
\$50,000–74,999	94%		80%		75%		48%	
\$75,000+	95%		88%		77%		54%	
Education								
Bachelor's degree or higher	95%	0.01	87%	0.0004	74%	0.40	53%	0.39
No bachelor's degree	90%		77%		75%		48%	
Race or Ethnic Group								
Black, non-Hispanic	85%	0.004	71%	0.06	71%	0.07	49%	0.13
Hispanic	89%	0.47	78%	0.48	69%	0.04	46%	0.33
White, non-Hispanic	93%	reference	81%	reference	76%	reference	50%	reference

All findings are adjusted for general concern about medical privacy and for concern about protecting privacy in the study. All p values are based on results of multiple logistic regressions containing all covariates in the table, as well as age, which was entered into the models as a continuous variable.

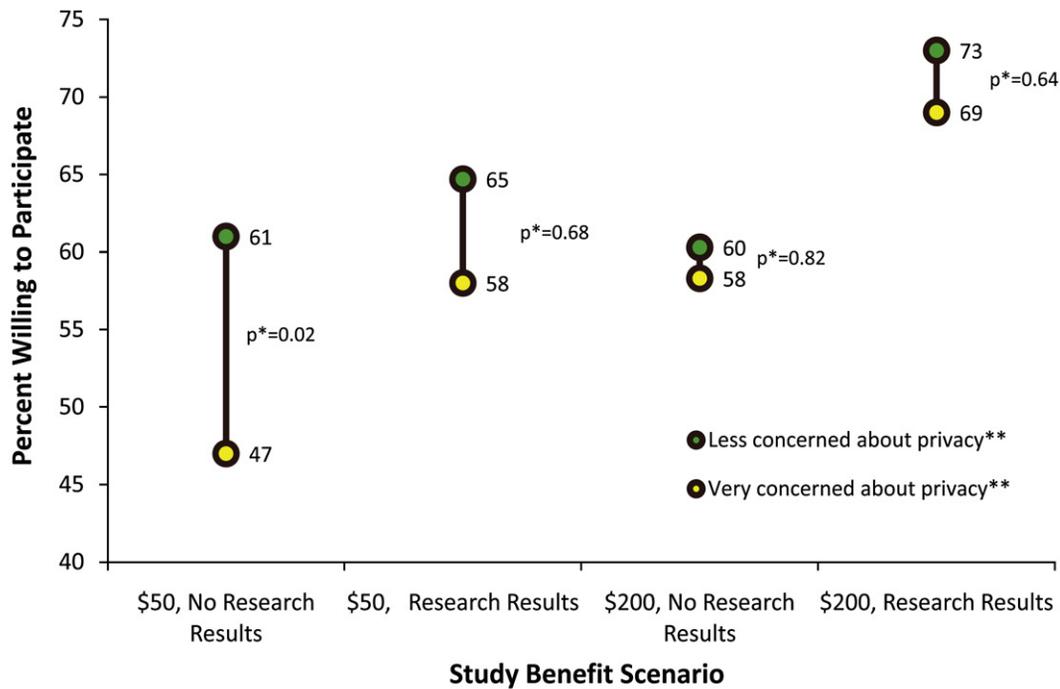
^a Percentage of respondents who agree with the statement "I would allow these researchers to use my samples and information for research."

^b Percentage of respondents who agree with the statement "If I could not be identified, I would be willing to have my information and research results available on the internet to anyone."

willing to participate, compared to only 19% of those who would not permit academics to use their data (adjusted $p < 1 \times 10^{-9}$). Additionally, respondents who were worried that study results could be used against them were less likely to say that they would participate than were those who were not concerned (48% versus 68%, adjusted $p = 2 \times 10^{-7}$). Respondents who would be willing to have their "information and research results made available on the internet to anyone" if they could not be identified were more likely to say that they would participate (75% versus 46%, $p = 2 \times 10^{-6}$). After adjustment for survey respondents' willingness to share information with academic researchers and their concern about study data being used against them, responses to questions about people's general concern about the privacy of medical information ($p = 0.29$) and about protection of privacy in the study ($p = 0.87$) were not associated with willingness to participate.

As Figure 2 shows, the relationship between concerns about protecting one's privacy in the study and willingness

to participate varied depending on the study benefits in the scenario that respondents were shown. Among people shown the scenario in which cohort-study participants would not receive individual research results and would be given only \$50 for their time, concerns about protection of privacy were significantly related to willingness to join the study; those who were very concerned about protection of their privacy were significantly less likely to say that they would participate (47% versus 61%, respectively; adjusted odds ratio [OR] = 0.68, $p = 0.02$). However, among respondents who were told that they would receive \$200 (OR comparing "very concerned" to "less concerned" = 0.96, $p = 0.82$), those told that they would receive individual research results (OR = 0.92, $p = 0.68$), and those told they would receive both \$200 and research results (OR = 0.90, $p = 0.64$), there was no significant difference in willingness to participate between people who were very concerned about protecting their privacy and those with less concern.



*p values are for the comparison of the odds that those less concerned with privacy would participate to the odds that those very concerned would participate, adjusting for age, gender, race and ethnicity, household income, and education.

**These categories represent those who said they would be “very concerned” about “protecting my privacy” if they were participating in the study and those who said they would be somewhat concerned, a little concerned, or not at all concerned

Figure 2. The Relationship between Privacy Concerns and Willingness to Participate in Research When the Benefits of Participation Vary

Relationship between concerns about protecting privacy in the large cohort study and willingness to participate in the study, under differing scenarios of study benefits.

Nearly all respondents (93%) said that it would be somewhat or very important that it be illegal for insurers and employers to get their study information, and 84% felt that it would be important to have a law protecting research information from law-enforcement officials. Concerns about one’s medical privacy ($p = 0.00002$), concerns about the government having access to samples and information ($p < 1 \times 10^{-9}$), and fear that data collected by the cohort study could be used against a study participant ($p = 0.01$) were all independently and significantly associated with support for a law barring access by employers and insurers. Similarly, concern about the government having access to samples and information ($p \leq 1 \times 10^{-9}$) and general concern about medical privacy ($p = 0.009$) were associated with support for a law prohibiting access by law enforcement.

Conclusions

As has been observed in many other studies of privacy in research, survey respondents in this study strongly valued both privacy and participation in biomedical research. Other studies of public attitudes have consistently shown

that broad public support for biomedical research is often balanced against concerns about maintaining confidentiality.^{17,44–46} Despite ubiquitous concerns about protection of privacy among our survey respondents, six in ten would participate in the large cohort study if asked, and most would share their research data with academic, government, and industry researchers. This finding agrees with other surveys that observed that more than half of respondents would be willing to share clinical data and samples for research, provided that either the patients’ permission would be sought beforehand or their data would be deidentified to protect their privacy.^{44–53}

Genetic information has been viewed by some scholars as an exceptionally sensitive class of information,⁵⁴ and it has been targeted for specific heightened privacy protections.⁵⁴ In this study, only 15% of those surveyed felt that genetic test results were a class of medical information that was especially sensitive and needed extra privacy protection. Although 37% did worry that data collected by the study could be used against them, fear of genetic discrimination and the exceptional nature of genetic information were not strongly supported here. It is striking that although 90% of respondents were concerned about protecting their privacy, less than half that many said that

they feared that the data would be used against them. In addition to being related to worries about discrimination or loss of insurance, concerns about maintaining privacy may be strongly related to issues of control over information about oneself. The fact that large majorities agreed that study-by-study consent would make them feel that they were in control and respected by researchers supports the notion that maintaining a sense of personal autonomy may be as important as minimizing the harms that might accrue from sharing personal information.

The issue of whom study data would be shared with was a salient one. Survey respondents consistently were more worried about government researchers and agencies accessing data from the cohort study than about academic and medical researchers doing so. Sharing information with pharmaceutical companies was even less palatable, though it is unclear whether this was due to privacy concerns or disapproval of the industry's profit motive. These findings are consistent with several other studies that observed that people are more willing to share medical information with academics than with governments^{17,33,35,50–52,55,56} or industries.^{17,48,51,53,55} As others have suggested,^{19,57} this implies that informed-consent documents should clarify, to the extent possible, what types of researchers will have access to study data.

General concerns about the protection of privacy in the study were not significantly related to people's stated willingness to participate in the large cohort study, but some specific beliefs related to privacy were correlated with willingness to join the study. The 37% who feared that study data could be used against them were significantly less likely to join the cohort study, but nearly half would participate despite this fear. Although the respondents who would not let academic researchers apply for use of the cohort data were significantly less likely to say that they would join the study, this group included only 8% of respondents, so this unwillingness seems unlikely to have a particularly large overall impact. Several studies of the importance of privacy and consent in genetic research participation have observed high rates of willingness to participate in genetic research despite broad general concerns about privacy. A Singapore-based survey found that people who were less concerned about privacy were more willing to donate a blood sample for genetic research, but privacy concerns were less important to people than the perceived societal benefits of participation.³⁵ Members of a genetic study of epilepsy felt that it was important to have general control over access to their DNA, but understood that not agreeing to a full release of the data could compromise the utility of the data set.¹⁷ A Swedish study found that 86% of participants would donate a blood sample and information to a secure database linked to personal identifiers and that only 3% more would participate if the data were delinked.³⁶ Another study found that similar percentages of people would be willing to participate in genetic studies whether the studies used identifiable or deidentified data.³⁷

The relationship between general privacy concerns and willingness to participate in genetic research may be of most significance in recruiting a representative sample. People who have not completed a college education and members of the black non-Hispanic community were observed to have greater privacy concerns across several privacy-related questions. Efforts to enhance both overall privacy protections and the information that people in these demographic groups receive about such efforts may be of particular use in minimizing selection biases due to differential levels of concern.

Although this survey does not explain with certainty why pervasive privacy concerns do not translate into more inhibition about participation in research, the data suggest at least two possible explanations. One possibility, echoed in other studies, is that although people may recognize the risk of significant loss of privacy if they participate, they may also recognize and accept that their privacy cannot be guaranteed despite researchers' best efforts.⁴⁸ Many people may be willing to accept the risks that they feel remain. For example, respondents to our survey may have gained confidence in the ability of the large cohort study to protect their privacy because they were told that their data would be coded before being entered into the database, that a committee would oversee who is allowed the use of the data, and that they might be able to consent for each use of their data. It may also be that people underestimate the extent to which their data will be shared by biobanks.^{14–17}

A second possible explanation for the observation that ubiquitous privacy concerns are not strongly related to willingness to join a research study is that for many people, the perceived potential risk of a loss of privacy is overcome by the benefits of participating in such research. Generally speaking, there are very few individual benefits to participating in a biobank or observational cohort study. Benefits might include monetary compensation for a participant's time or the return of health information or research findings from the study. Although such benefits do nothing to mitigate the actual risk of suffering a loss of privacy, our survey offers some evidence that modest payment or the return of individual research results may outweigh some people's privacy concerns as they decide whether or not to participate in such research.

The notion that payment or return of health information can outweigh a person's privacy concerns may have implications for the ethical enrollment of subjects. When considering issues of privacy, the first goal of a researcher should be to protect research participants to the greatest extent that is practically possible. However, once these protections have been put in place, it is not clear that providing health information or a reasonable, fair incentive is ethically wrong simply because some people view it as a sufficient counterbalance to their remaining privacy concerns. For many people, their simple desire to contribute to research will be sufficient to overcome their privacy concerns—for example, in [Figure 2](#), 47% of those

who said that they would be very concerned about their privacy would participate even when offered the lower level of compensation and no individual research results. However, it would seem odd to forbid recruiters from mentioning the potential societal benefits of research participation because of concerns that doing so might be coercive. For many people, concerns about privacy may not be an absolute deal-breaker that precludes participation in research but, rather, may be one of a list of several pros and cons that are weighted in deciding whether or not to participate. If the information or incentives to be given to participants are deemed appropriate to the risks and burden of participation, it may be acceptable that they tip the balance toward participation for some people who are concerned about privacy.

It should be emphasized that people's responses on a cross-sectional survey about participation in a hypothetical study will not necessarily correlate with actual behaviors in real situations at a different point in time. This study is able to provide valid estimates of the relationships between concerns about privacy, study-design factors, and people's willingness to participate, but is likely to be less accurate in estimating actual participation rates. Additionally, this survey was fielded before the passage of the Genetic Information Nondiscrimination Act of 2008 (GINA), which prevents health insurers and employers from denying coverage, adjusting premiums, or otherwise discriminating on the basis of genetic information.^{58,59} With the passage and implementation of GINA, the risks and potential harms of misuse of genetic research data will decrease. However, changes in public perceptions of these risks may lag behind these legislative advances, and researchers should be prepared to address ongoing privacy concerns about the potential misuse of both genetic and nongenetic medical information.

The actions that the public views as serious misuses of study data, the people whom they deem most likely to commit these actions, and what they believe could or should be done to improve the public trust and prevent such actions are all closely related topics that warrant additional attention and further research. Changes in the landscape of privacy risks and protections have occurred since the survey was fielded, including the passage of GINA, the emergence of new methods enabling reidentification of deidentified study data through DNA, the increased publication of genotype data in manuscripts of genome-wide association studies, and the increased use of electronic medical records. The effects of these changes on public views of the privacy risks associated with participation in biobanks or biomedical research warrant further investigation.

Regardless of what such research might find, the results of this study support the argument that during the consent process, potential research participants should be told about the different levels of deidentification of data that are possible,¹⁰ the fact that studies including DNA may not be completely deidentifiable, explicit details of the protections offered by the study protocol, and the privacy

risks that remain.^{57,60–62} In addition to providing research participants with transparent, forthright explanations of the privacy risks that they may face, consent documents should detail what data could be gathered through study protocols, whom the data could be shared with, how the data might be analyzed, and what formats the data are likely to be published in. The desire of research participants to know what risks they face up front will be satisfied, and trust, based on an honest assessment of risks and protections, may be established between researchers and those who choose to participate.

In addition to clear communications of outlying privacy risks that accompany participation in research, the research enterprise must work to fortify the protections that it offers participants. Policies about the publication of and public access to deidentified data that include genetic sequences should be reviewed by parties that share or publish such data. Researchers should be encouraged to use certificates of confidentiality to protect participants from forced disclosure of their identities for use in civil, criminal, administrative, or legislative proceedings. The NIH should consider adoption of a different model of certificate of confidentiality, such as the one used by researchers at the U.S. Department of Justice that does not permit the researcher discretion about whether to release study information to law-enforcement officials and instead forbids studies with certificates from all such disclosures. It may also be worthwhile to examine what practices researchers and data-access committees of biobanks and large cohort studies are using to maintain privacy, where they view vulnerabilities, and what problems they have experienced or observed in protecting subjects' privacy, because research practitioners may identify problems and potential solutions long before policy makers become aware that the problems exist.

Appendix: Definition of Individual Research Results and Scenario Variables Used in the Pilot Study

Definition of Individual Research Results

The following text was used in the survey as an explanation of the concept of individual research results for survey participants:

Blood samples would be sent to a lab, where a genetic analysis would be done. Genetic and medical information would be stored in a databank. Researchers could apply to use the samples and information to study genes, environment, and lifestyle.

Researchers might find that a certain genetic, environmental or lifestyle factor is related to a specific disease. This kind of general study finding would be released to the public.

An individual participant's research result would be the information researchers find about whether the person had a specific genetic, environmental, or lifestyle risk factor. These individual research results would not be released to the public.

Scenarios Used in Pilot Study Instrument

Each survey respondent viewed one of eight randomly selected scenarios that varied with respect to three factors: the burden of the cohort study that they would be asked to participate in, the amount of compensation that they would receive, and whether or not they would receive individual research results back. There were two versions of each factor, resulting in eight possible scenarios:

- (1) Higher Burden, Higher Compensation, No Individual Research Results
- (2) Higher Burden, Higher Compensation, Individual Research Results Returned
- (3) Higher Burden, Lower Compensation, No Individual Research Results
- (4) Higher Burden, Lower Compensation, Individual Research Results Returned
- (5) Lower Burden, Higher Compensation, No Individual Research Results
- (6) Lower Burden, Higher Compensation, Individual Research Results Returned
- (7) Lower Burden, Lower Compensation, No Individual Research Results
- (8) Lower Burden, Lower Compensation, Individual Research Results Returned

The alternate versions of the three factors were defined as follows:

Lower Burden: *Let's say the study is going forward and you were invited to participate. At the beginning of the study, you would be asked to travel to a local health clinic for one half day of exams. You would provide samples (blood, urine, etc.) for laboratory tests and fill out questionnaires on your health, diet and lifestyle. In addition, you would be asked to complete a health assessment questionnaire once a year for the next ten years.*

Higher Burden: *Let's say the study is going forward and you were invited to participate. At the beginning of the study, you would be asked to travel to a local health clinic for one half day of exams. You would provide samples (blood, urine, etc.) for laboratory tests and fill out questionnaires on your health, diet and lifestyle. Researchers would come to your home to collect environmental samples and to place a device to monitor air quality. You would be asked keep a diet and exercise journal for one week and to complete a health assessment questionnaire once a year for the next ten years.*

Receive Research Results: *You would receive results from your initial physical examination and laboratory tests. You would also receive general research findings from the study. You could also find out your individual research results if you wanted to.*

Do Not Receive Research Results: *You would receive results from your initial physical examination and laboratory tests. You would also receive general research findings from the study. However, you would not be given any individual research results.*

Lower Compensation: *The study would reimburse you for the cost of any travel to and from the initial exam. You would receive \$50 to compensate you for time spent at the initial exam.*

Higher Compensation: *The study would reimburse you for the cost of any travel to and from the initial exam. You would receive \$200 to compensate you for time spent at the initial exam and an additional \$20 for each completed health-assessment questionnaire.*

Acknowledgments

This work was funded by the National Human Genome Research Institute under grant number 1 U01 HG004206-0, Public Consultation to Inform the Design of Possible Large Scale Studies of Genes and Environment in Common Disease. The authors wish to thank our anonymous reviewers for providing advice to improve this manuscript.

Received: July 28, 2009

Revised: September 21, 2009

Accepted: October 5, 2009

Published online: October 29, 2009

References

1. Watts, G. (2007). UK Biobank gets 10% response rate as it starts recruiting volunteers. *BMJ* 7595, 659.
2. National Children's Study. (2007). National Children's Study Homepage. <http://www.nationalchildrensstudy.gov/>.
3. Roden, D.M., Pulley, J., Basford, M., Bernard, G., Clayton, E., Balsler, J., and Masys, D. (2008). Development of a large-scale de-identified DNA biobank to enable personalized medicine. *Clin. Pharmacol. Ther.* 84, 362–369.
4. Kaiser Permanente. (2008). Division of Research: Research program on genes, environment, and health. <http://www.dor.kaiser.org/studies/rpgeh/index.html>.
5. Biobank, U.K. (2008). UK Biobank Homepage. <http://www.ukbiobank.ac.uk/>.
6. Palmer, L.J. (2007). UK Biobank: bank on it. *Lancet* 9578, 1980–1982.
7. Manolio, T.A., Bailey-Wilson, J.E., and Collins, F.S. (2006). Genes, environment and the value of prospective cohort studies. *Nat. Rev. Genet.* 10, 812–820.
8. U.S. Department of Health & Human Services. Summary of the HIPAA Privacy Rule. <http://www.hhs.gov/ocr/privacy/hipaa/understanding/summary/index.html>.
9. Homer, N., Szeling, S., Redman, M., Duggan, D., Tembe, W., Muehling, J., Pearson, J.V., Stephan, D.A., Nelson, S.F., and Craig, D.W. (2008). Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. *PLoS Genet.* 8, e1000167.
10. Lowrance, W.W., and Collins, F.S. (2007). Ethics. Identifiability in genomic research. *Science* 5838, 600–602.
11. Kohane, I.S., and Altman, R.B. (2005). Health-information altruists—a potentially critical resource. *N. Engl. J. Med.* 19, 2074–2077.
12. National Human Genome Research Institute. (2004). Design considerations for a potential United States population-based cohort to determine the relationships among genes, environment, and health: recommendations of an expert panel. <http://www.genome.gov/Pages/About/OD/ReportsPublications/PotentialUSCohort.pdf>.

13. Kaufman, D., Geller, G., Leroy, L., Murphy, J., Scott, J., and Hudson, K. (2008). Ethical implications of including children in a large biobank for genetic-epidemiologic research: a qualitative study of public opinion. *Am. J. Med. Genet. C. Semin. Med. Genet. 1*, 31–39.
14. McCarty, C.A., Nair, A., Austin, D.M., and Giampietro, P.F. (2007). Informed consent and subject motivation to participate in a large, population-based genomics study: the Marshfield Clinic Personalized Medicine Research Project. *Community Genet. 1*, 2–9.
15. Ormond, K.E., Cirino, A.L., Helenowski, I.B., Chisholm, R.L., and Wolf, W.A. (2009). Assessing the understanding of biobank participants. *Am. J. Med. Genet. A. 2*, 188–198.
16. Schmidt, S., Allen, K.D., Loiacono, V.T., Norman, B., Stanwyck, C.L., Nord, K.M., Williams, C.D., Kasarskis, E.J., Kamel, F., McGuire, V., et al. (2008). Genes and environmental exposures in veterans with amyotrophic lateral sclerosis: the GENEVA study. Rationale, study design and demographic characteristics. *Neuroepidemiology 3*, 191–204.
17. McGuire, A.L., Hamilton, J.A., Lunstroth, R., McCullough, L.B., and Goldman, A. (2008). DNA data sharing: research participants' perspectives. *Genet. Med. 1*, 46–53.
18. Lowrance, W.W. (2006). Summary of the NHGRI workshop on privacy, confidentiality and identifiability in genomic research. <http://www.genome.gov/19519198>.
19. Beskow, L.M., Burke, W., Merz, J.F., Barr, P.A., Terry, S., Penchaszadeh, V.B., Gostin, L.O., Gwinn, M., and Khoury, M.J. (2001). Informed consent for population-based research involving genetics. *JAMA 18*, 2315–2321.
20. Nass, S.J., Levit, L.A., and Gostin, L.O. (2009). Beyond the HIPAA privacy rule: enhancing privacy, improving health through research. <http://www.nap.edu/catalog/12458.html>.
21. Pritts, J. (2008). The importance and value of protecting the privacy of health information: roles of HIPAA privacy rule and the Common Rule in health research. <http://www.iom.edu/CMS/3740/43729/53160.aspx>.
22. Rachels, J. (1975). Why privacy is important. *Philos. Public Aff. 4*, 323–333.
23. Institute of Medicine. (2009). *Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research*, 320.
24. Forrester Research. (2005). National consumer health privacy survey 2005. <http://www.chcf.org/topics/view.cfm?itemID=115694>.
25. Satia, J.A., McRitchie, S., Kupper, L.L., and Halbert, C.H. (2006). Genetic testing for colon cancer among African-Americans in North Carolina. *Prev. Med. 1*, 51–59.
26. Dickson, M.R., Carter, C.L., Carpenter, M.J., McClure, R.L., McGee, D.A., Zapka, J.G., and Strange, C. (2008). Barriers to genetic testing among persons at risk for alpha-1 antitrypsin deficiency. *Genet. Test. 4*, 501–505.
27. Baruch, S., Kaufman, D., and Hudson, K. (2007). U.S. public opinion on uses of genetic information and genetic discrimination. http://www.dnnapolicy.org/resources/GINAPublic_Opinion_Genetic_Information_Discrimination.pdf.
28. Hadley, D.W., Jenkins, J., Dimond, E., Nakahara, K., Grogan, L., Liewehr, D.J., Steinberg, S.M., and Kirsch, I. (2003). Genetic counseling and testing in families with hereditary nonpolyposis colorectal cancer. *Arch. Intern. Med. 5*, 573–582.
29. Issa, A.M., Tufail, W., Hutchinson, J., Tenorio, J., and Baliga, M.P. (2009). Assessing patient readiness for the clinical adoption of personalized medicine. *Public Health Genomics 3*, 163–169.
30. Hoop, J.G., Roberts, L.W., and Hammond, K.A. (2009). Genetic testing of stored biological samples: views of 570 U.S. workers. *Genet. Test. Mol. Biomarkers. 3*, 331–337.
31. Lowstuter, K.J., Sand, S., Blazer, K.R., MacDonald, D.J., Banks, K.C., Lee, C.A., Schwerin, B.U., Juarez, M., Uman, G.C., and Weitzel, J.N. (2008). Influence of genetic discrimination perceptions and knowledge on cancer genetics referral practice among clinicians. *Genet. Med. 9*, 691–698.
32. Godard, B., Marshall, J., and Laberge, C. (2007). Community engagement in genetic research: results of the first public consultation for the Quebec CARTaGENE project. *Community Genet. 3*, 147–158.
33. Beskow, L.M., and Dean, E. (2008). Informed consent for biorepositories: assessing prospective participants' understanding and opinions. *Cancer Epidemiol. Biomarkers Prev. 6*, 1440–1451.
34. Bogner, H.R., Wittink, M.N., Merz, J.F., Straton, J.B., Cronholm, P.F., Rabins, P.V., and Gallo, J.J. (2004). Personal characteristics of older primary care patients who provide a buccal swab for apolipoprotein E testing and banking of genetic material: the spectrum study. *Community Genet. 4*, 202–210.
35. Wong, M.L., Chia, K.S., Yam, W.M., Teodoro, G.R., and Lau, K.W. (2004). Willingness to donate blood samples for genetic research: a survey from a community in Singapore. *Clin. Genet. 1*, 45–51.
36. Kettis-Lindblad, A., Ring, L., Viberth, E., and Hansson, M.G. (2006). Genetic research and donation of tissue samples to biobanks. What do potential sample donors in the Swedish general public think? *Eur. J. Public Health 4*, 433–440.
37. Hull, S.C., Sharp, R.R., Botkin, J.R., Brown, M., Hughes, M., Sugarman, J., Schwinn, D., Sankar, P., Bolcic-Jankovic, D., Clarridge, B.R., et al. (2008). Patients' views on identifiability of samples and informed consent for genetic research. *Am. J. Bioeth. 10*, 62–70.
38. Murphy, J., Scott, J., Kaufman, D., Geller, G., LeRoy, L., and Hudson, K. (2008). Public expectations for return of results from large-cohort genetic research. *Am. J. Bioeth. 11*, 36–43.
39. Baker, L.C., Bundorf, M.K., Singer, S., and Wagner, T.H. (2003). Validity of the survey of health and internet and knowledge network's panel and sampling. <http://www.knowledgenetworks.com/ganp/docs/Appendix%20Survey%20of%20Health%20and%20the%20Internet.pdf>.
40. Genetics and Public Policy Center. (2008). Proposed study video. <http://www.youtube.com/watch?v=m-x81nkCP8A>.
41. Kaufman, D., Murphy, J., Scott, J., and Hudson, K. (2008). Subjects matter: a survey of public opinions about a large genetic cohort study. *Genet. Med. 11*, 831–839.
42. International, R.T.I. (2008). SUDAAN version 9.0 online reference manual (Durham, NC: Research Triangle Institute).
43. Murphy, J., Scott, J., Kaufman, D., Geller, G., LeRoy, L., and Hudson, K. (In press, 2009.). Perspectives on informed consent for biobanking. *Am. J. Public Health*.
44. Kass, N.E., Hull, S.C., Natowicz, M.R., Faden, R.R., Plantinga, L., Gostin, L.O., and Slutsmann, J. (2004). Medical privacy and the disclosure of personal medical information: the beliefs and experiences of those with genetic and other clinical conditions. *Am. J. Med. Genet. A. 3*, 261–270.
45. Robling, M.R., Hood, K., Houston, H., Pill, R., Fay, J., and Evans, H.M. (2004). Public attitudes towards the use of

- primary care patient record data in medical research without consent: a qualitative study. *J. Med. Ethics* 1, 104–109.
46. Robert Wood Johnson Foundation. (2005). Americans support online personal health records; patient privacy and control over their own information are crucial to acceptance. <http://www.rwjf.org/pr/product.jsp?id=21844>.
 47. Baker, R., Shiels, C., Stevenson, K., Fraser, R., and Stone, M. (2000). What proportion of patients refuse consent to data collection from their records for research purposes? *Br. J. Gen. Pract.* 457, 655–656.
 48. Damschroder, L.J., Pritts, J.L., Neblo, M.A., Kalarickal, R.J., Creswell, J.W., and Hayward, R.A. (2007). Patients, privacy and trust: patients' willingness to allow researchers to access their medical records. *Soc. Sci. Med.* 1, 223–235.
 49. Kass, N.E., Natowicz, M.R., Hull, S.C., Faden, R.R., Plantinga, L., Gostin, L.O., and Slutsman, J. (2003). The use of medical records in research: what do patients want? *J. Law Med. Ethics* 3, 429–433.
 50. Whiddett, R., Hunter, I., Engelbrecht, J., and Handy, J. (2006). Patients' attitudes towards sharing their health information. *Int. J. Med. Inform.* 7, 530–541.
 51. Willison, D.J., Schwartz, L., Abelson, J., Charles, C., Swinton, M., Northrup, D., and Thabane, L. (2007). Alternatives to project-specific consent for access to personal information for health research: what is the opinion of the Canadian public? *J. Am. Med. Inform. Assoc.* 6, 706–712.
 52. Page, S.A., and Mitchell, I. (2006). Patients' opinions on privacy, consent and the disclosure of health information for medical research. *Chronic Dis. Can.* 2, 60–67.
 53. Stone, M.A., Redsell, S.A., Ling, J.T., and Hay, A.D. (2005). Sharing patient data: competing demands of privacy, trust and research in primary care. *Br. J. Gen. Pract.* 519, 783–789.
 54. McGuire, A.L., Fisher, R., Cusenza, P., Hudson, K., Rothstein, M.A., McGraw, D., Matteson, S., Glaser, J., and Henley, D.E. (2008). Confidentiality, privacy, and security of genetic and genomic test information in electronic health records: points to consider. *Genet. Med.* 7, 495–499.
 55. Kettis-Lindblad, A., Ring, L., Viberth, E., and Hansson, M.G. (2007). Perceptions of potential donors in the Swedish public towards information and consent procedures in relation to use of human tissue samples in biobanks: a population-based study. *Scand. J. Public Health* 2, 148–156.
 56. Buchwald, D., Mendoza-Jenkins, V., Croy, C., McGough, H., Bezdek, M., and Spicer, P. (2006). Attitudes of urban American Indians and Alaska Natives regarding participation in research. *J. Gen. Intern. Med.* 6, 648–651.
 57. Lunshof, J.E., Chadwick, R., Vorhaus, D.B., and Church, G.M. (2008). From genetic privacy to open consent. *Nat. Rev. Genet.* 5, 406–411.
 58. Hudson, K.L., Holohan, M.K., and Collins, F.S. (2008). Keeping pace with the times—the Genetic Information Nondiscrimination Act of 2008. *N. Engl. J. Med.* 25, 2661–2663.
 59. Baruch, S., and Hudson, K. (2008). Civilian and military genetics: nondiscrimination policy in a post-GINA world. *Am. J. Hum. Genet.* 4, 435–444.
 60. Greely, H.T. (2007). The uneasy ethical and legal underpinnings of large-scale genomic biobanks. *Annu. Rev. Genomics Hum. Genet.* 343–364.
 61. McGuire, A.L., and Gibbs, R.A. (2006). Genetics. No longer de-identified. *Science* 5772, 370–371.
 62. Hull, S.C., and Wilfond, B.S. (2008). What does it mean to be identifiable? *Am. J. Bioeth.* 10, W7–W8.