

# Balancing the Risks and Benefits of Genomic Data Sharing: Genome Research Participants' Perspectives

J.M. Oliver<sup>a</sup> M.J. Slashinski<sup>c</sup> T. Wang<sup>b</sup> P.A. Kelly<sup>d</sup> S.G. Hilsenbeck<sup>b</sup>  
A.L. McGuire<sup>a</sup>

<sup>a</sup>Center for Medical Ethics and Health Policy and <sup>b</sup>Dan L. Duncan Cancer Center, Baylor College of Medicine, and <sup>c</sup>Division of Health Promotion and Behavioral Sciences, School of Public Health, University of Texas Health Science Center, Houston, Tex.; <sup>d</sup>Department of Medicine, Tulane University School of Medicine, New Orleans, La., USA

## Key Words

Data sharing · Ethical issues · Genome research · Participant perspectives

## Abstract

**Background:** Technological advancements are rapidly propelling the field of genome research forward, while lawmakers attempt to keep pace with the risks these advances bear. Balancing normative concerns of maximizing data utility and protecting human subjects, whose privacy is at risk due to the identifiability of DNA data, are central to policy decisions. Research on genome research participants making real-time data sharing decisions is limited; yet, these perspectives could provide critical information to ongoing deliberations. **Methods:** We conducted a randomized trial of 3 consent types affording varying levels of control over data release decisions. After debriefing participants about the randomization process, we invited them to a follow-up interview to assess their attitudes toward genetic research, privacy and data sharing. **Results:** Participants were more restrictive in their reported data sharing preferences than in their actual data sharing decisions. They saw both benefits and risks associated with sharing their genomic data, but risks were seen as less concrete or happening in the future,

and were largely outweighed by purported benefits. **Conclusion:** Policymakers must respect that participants' assessment of the risks and benefits of data sharing and their privacy-utility determinations, which are associated with their final data release decisions, vary. In order to advance the ethical conduct of genome research, proposed policy changes should carefully consider these stakeholder perspectives.

Copyright © 2011 S. Karger AG, Basel

## Introduction

As technology allows genome scientists to rapidly propel the field forward, lawmakers are struggling to keep pace with advances in genome research and the risks these advances bear. Presently, ethicists and other stakeholders are collaborating to develop guidelines to provide direction for the ethical conduct of worldwide data sharing [1]. Current U.S. policies mandate federally funded researchers share generated sequence data with the scientific community by requiring deposition of DNA data in government repositories such as the National Institutes of Health database of Genotypes and Phenotypes [2].

Central to policy decisions related to genomic data sharing are two normative concerns: (1) advancing research by maximizing data efficiency and utility, and (2) protecting human subjects by minimizing risks to privacy. Research on the unique identifiability of DNA data [3, 4] has prompted a shift in policy from exclusive emphasis on sharing data in open access, or publicly accessible databases [5–7], to the creation of controlled access, or restricted databases [8]. Additional restrictions may be imposed as existing regulations [9–11], which do not consider the research use of de-identified DNA data to be research involving human subjects, are currently under review [12]. It is important that any proposed policy change takes into consideration all stakeholder perspectives, including the perspectives of genome research participants.

Several focus group and survey studies have explored participants' general perspectives on data sharing. Those studies suggest that participants want to be involved in the decision to share their genetic data and most are willing to share despite concerns regarding government oversight, privacy and confidentiality, and profiteering or misuse of data [13–18]. We conducted a randomized trial of 3 different types of consent, each affording varying levels of control over the decision about data sharing, with participants being recruited into genome research studies to assess their impact on enrollment and data sharing decisions [19]. A follow-up interview provided participants with an opportunity to share their attitudes towards genomic and scientific research, privacy and data sharing. This is the first study to our knowledge to explore genome research participants' real-time data sharing decisions and to examine the attitudes and preferences underlying those decisions. We previously reported that, despite noted concerns, the majority (53%) of participants opted for public data sharing, while a significant minority (47%) chose a more restricted data sharing option [19]. The purpose of this article is to explore the underlying factors influencing these decisions, including judgments about the risks and benefits of data sharing and issues of privacy versus data utility.

## Subjects and Methods

Participants were those recruited into one of 6 ongoing genomic research studies at Baylor College of Medicine (BCM) in Houston, Texas between January 2008 and August 2009, including pediatric autism, pediatric brain cancer, pediatric brain controls, adult/pediatric epilepsy, adult/pediatric liver cancer, and adult pancreas cancer. Participants were English proficient and

18 years of age or older and included adult patients, parents/guardians of pediatric patients and family members acting as matched case controls.

As there was no practical way to obtain consent for this study without first going through the very consent process under study, we obtained a waiver of consent from the BCM institutional review board to randomize participants to one of 3 experimental consent documents with which they were enrolled into the relevant genome study. The 3 consent documents varied by genome study and in the data release options provided to participants. Informed consent was obtained in a face-to-face setting by the genome study principal investigator (PI), a research nurse or a medical resident. The consent process varied by genome study, but was the same within each study regardless of randomized consent type.

A trained research assistant from the Center for Medical Ethics and Health Policy at BCM debriefed those who consented to participate in the genome study immediately following the consent process, during a follow-up clinic visit or, for those who did not return to the clinic for a follow-up visit, by phone or by mail. The time from enrollment to debriefing varied by genome study; on average, the time lapse was 24.6 days. Debriefing entailed providing information about the randomized consent study, a detailed review of the other consent types and the data release options they provided, and an opportunity for participants to change their initial data release selection. Details of the randomization and debriefing processes have been described elsewhere [19].

Participants who enrolled in the genome study and were debriefed in person were invited to participate in a follow-up structured interview to assess their understanding, decision-making, and preferences for and attitudes toward data sharing. The face-to-face interview was administered during the debriefing by an ethics research assistant who guided participants through the questionnaire using a laptop computer and an electronic interview data warehouse program, QDS (NOVA Research Company, Bethesda, Md.). To mitigate bias, those who participated in the interview were not shown the other consent documents or data release options until partway through the interview. Interviews lasted approximately 45 min. Verbal consent was obtained, and with participants' permission, all interviews were digitally recorded. Participants were compensated with a USD 25 gift card for their participation. All materials and methods were approved by the BCM IRB.

### *Study Instruments*

*Consent Documents.* Three experimental consent documents were developed and tailored for each genomic study with expert input and a thorough review of the informed consent literature [19]. The consent documents offered participants varying combinations of choices of how broadly they could agree to release their genetic information. Data release options were described as (a) public data release (release of genetic and clinical information into both publicly accessible [open access through the internet] and restricted [accessible only to approved researchers] scientific databases), (b) restricted release (release of genetic and clinical information into restricted databases only), and (c) no release (accessible only to the genomic study PI and staff). Those who signed the *traditional* consent by default agreed to public data release (a) of their genetic and clinical information. The *binary* consent allowed participants to choose between public data release (a) and

no release (c). *Tiered* consent provided all 3 options; participants could choose public data release (a), restricted release only (b) or no release (c). Data sharing was explained similarly in each consent document, including the potential risks and benefits of sharing genetic information. While participants were informed that personally identifying information would not be released, the potential threat to privacy if DNA were traced back to the individual was described as a risk. It was noted that this risk could increase with future technological advances. Restricted access databases were described as providing an extra layer of protection because they can only be accessed by approved researchers, who have an obligation to protect privacy and maintain confidentiality. Benefits of data sharing were identified as aiding the advancement of medical research by allowing additional investigators to access the data for future research questions; it was also explained that individual direct benefits were unlikely.

*Questionnaire.* A structured interview questionnaire combining open-ended and forced-choice items was developed out of a review of the literature and with input from interdisciplinary experts. After 6 months of data collection, item response statistics were computed, and the instrument was revised to eliminate items that were not yielding significant data due to high ceiling effects and to decrease interview fatigue. The questionnaire explored 6 general areas: understanding (of participation in research generally and genetic information), comfort in decision-making (adapted from the O'Connor scale [20]), trust in medical researchers, risk-benefit assessment, preferences for and attitudes toward consent types and data sharing options, and demographic information. The types of response categories varied by item and included yes/no/don't know options; 5-point Likert-type scales rating agreement with a statement; ordinal options for ranking risks, benefits and consent type preferences; and categorical options for demographic information. The questionnaire is available by contacting the corresponding author.

#### *Data Analysis*

Participant characteristics are described with the use of frequencies for categorical variables and means or medians for continuous variables. Differences between groups were tested with chi-square tests for categorical variables and one-way analysis of variance for continuous variables. For all tests, a significance level of  $p < 0.05$  was used. All analyses were conducted using SPSS 18 (SPSS, Inc., Chicago, Ill.).

Primary predictor variables were participant characteristics including genome study type (autism, epilepsy, brain cancer, brain controls, pancreas cancer, and liver cancer), consentee relationship (either adult/self consentee or parental consentee) and the time lapse (measured as continuous) between consent into the genome study and debriefing. Additional covariates included participant demographic characteristics (sex, race/ethnicity, age [measured as continuous], marital status, highest education level attained, and annual household income) and religious affiliation. Religious affiliation was categorized as Christian (response categories included Catholic, Protestant Christian and Evangelical Protestant) and other (response categories included were Jewish, Muslim, atheist or agnostic, and other). To measure participants' trust in doctors doing medical research, a validated scale [21] was used and, 6 months into data collection, reduced from a 4-item scale to 1 item due to high ceiling effects. Five-point Likert-type scale response categories, anchored by strongly disagree (1) to

strongly agree (5) were later collapsed to a dichotomous variable of no to low trust (1–3) and some to high trust (4–5).

Transcripts of verbatim transcribed interviews were analyzed using content analysis [22]. Coders initially independently coded all open-ended responses, and then the research team reviewed all interviews and resolved discrepancies using a consensus approach to ensure reliability and accuracy of coding. Qualitative analysis was managed using NVivo 8 (QSR International Inc., Cambridge, Mass.).

## **Results**

Of the 336 participants randomized to an experimental consent document and enrolled into a genome study, 13 were deemed ineligible and were removed prior to analysis of data release decisions [19]. Thirty-eight participants were deemed ineligible for the interview because they were debriefed by phone or mail ( $n = 27$ ) or they were matched-case controls ( $n = 11$ ) with other family members present who completed one interview per family. Among the 285 eligible participants invited to the structured interview, the participation rate was 80.4% ( $n = 229$ ). There were no significant demographic differences between those who completed the interview and those who declined.

Interview participants' median age was 48.8 years old (range 18–86). Participants were predominantly female (58.5%) and non-Hispanic white (58.1%). The majority reported being married (63.7%), Christian (80.8%) and completing at least some college (55.8%). Most participants also reported an annual household income over USD 40,000 (59.2%) (table 1).

#### *Attitudes Regarding the Risks and Benefits of Data Sharing*

While participants agreed that there were benefits and risks with data sharing, participants more strongly identified with the potential benefits than risks. In response to a statement that there were benefits to sharing their genetic information, 72.7% strongly agreed and 25.1% agreed. Contrastingly, 36% strongly agreed and 38.2% agreed that there were risks in sharing this information, while 16.9% strongly or somewhat disagreed (table 2).

Collectively, participants displayed more uniformity in their judgments about the benefits of data sharing. Asked to rank the most important benefit among an a priori list of 4 options, 62.7% selected advancing research to help others with a similar condition, 23% selected advancing general medical knowledge and 14.3% selected advancing research to help themselves or their family (table 2).

Qualitative data analysis of open-ended items reinforces these findings. Although some participants expressed hopes for realizing direct benefit from this research, most recognized this personal benefit was unlikely. As one participant commented (the bracketed numbers refer to the participant ID):

I'm just really serious about allowing the information from my illness to be used to help others down the line. That to me is the only benefit. Probably isn't going to help me, but it may help future patients. (231)

Some participants noted a specific desire to help others with a similar condition.

Well, the next person can avoid what I've had to put up with these past three years with the pancreas not acting nice. I just think it will benefit others with the same kind of illness that I have. (255)

Others described the more general societal benefits of participating in research.

Sharing my genetic information may be just the missing piece that the researchers need to advance good health and avoid diseases, and there may be something in my information that stands out that they didn't get in all the other people they've been studying. (369)

Survey results suggest less consistency in participants' views when ranking the most important risk in sharing their genetic information. A third (34.6%) selected the risk of having their identity revealed as most important, 30.1% selected not knowing what could happen with their genetic information in the future and 28.2% selected health insurance discrimination. The remaining 7.1% reported the most important risk was the fear of finding out unwanted information about themselves or their family member (table 2).

Open-ended items highlight participants' difficulties in identifying the concrete risks of data sharing. As one participant commented:

As a matter of fact, I'm even having a problem figuring out what the risk would be. Because I said – okay – so somebody learns about I have such-and-such. Well, what good does that do anybody? I'm just having a problem identifying the risks. (281)

Other concerns raised by participants included the lack of control over who could access their information in the public domain, fear of identity theft, anxiety about government access, apprehension over the potential commercialization of their DNA, and fear that their data would be used in morally objectionable research. Participants seemed to understand the inherent identifiability of DNA

**Table 1.** Interview participant characteristics

| Participant characteristics and demographics <sup>a</sup> |             |
|---|-------------|
| Genome study  |             |
| Autism  | 3 (1.3)     |
| Epilepsy  | 55 (24.0)   |
| Brain cancer  | 12 (5.2)    |
| Brain controls  | 17 (7.4)    |
| Liver cancer  | 77 (33.6)   |
| Pancreas cancer   | 65 (28.4)   |
| Consentee relationship                                    |             |
| Adult/self-consentee                                      | 154 (67.2)  |
| Parental consentee  | 75 (32.8)   |
| Consent-to-debrief time lapse                             |             |
| Mean days (SD)  | 24.6 (37.8) |
| Median (range)  | 15 (0–380)  |
| Mean age (SD), years                                      | 47.8 (15.2) |
| Sex   |             |
| Male  | 95 (41.5)   |
| Female  | 134 (58.5)  |
| Race/ethnicity  |             |
| Hispanic or Latino  | 47 (20.7)   |
| Black, non-Hispanic                                       | 31 (13.7)   |
| White, non-Hispanic                                       | 132 (58.1)  |
| Other <sup>b</sup>  | 17 (7.5)    |
| Marital status  |             |
| Married   | 142 (63.7)  |
| Divorced, widowed, or separated                           | 60 (26.9)   |
| Never married   | 21 (9.4)    |
| Religion  |             |
| Christian   | 181 (80.8)  |
| Other <sup>c</sup>  | 43 (19.2)   |
| Education   |             |
| ≤11th grade   | 16 (7.1)    |
| High school/GED   | 55 (24.6)   |
| 1–4+ years of college                                     | 125 (55.8)  |
| Graduate degree   | 28 (12.5)   |
| Annual household income, USD                              |             |
| ≤20,000   | 31 (15.8)   |
| 20,000–40,000   | 49 (25.0)   |
| 40,000–60,000   | 26 (13.3)   |
| ≥60,000   | 90 (45.9)   |

Values denote numbers of participants with percentages in parentheses, unless noted otherwise.

<sup>a</sup> Categories do not sum to 229 because of participant non-response. <sup>b</sup> Other includes Asian, Native American/Alaskan, Pacific Islander, and other. <sup>c</sup> Christian includes Catholic, Protestant, and Evangelical; other includes Jewish, Muslim, Atheist/Agnostic, and other.

data, but many felt that it would be years before anyone could actually identify an individual on the basis of their DNA. This may explain why in response to open-ended items older patients seemed less anxious about future risks affecting them and therefore more likely to share their



**Table 2.** Participants' risk benefit assessment

|  | n (%)                             |                               |                   |                   |
|--|-----------------------------------|-------------------------------|-------------------|-------------------|
|  | strongly/<br>somewhat<br>disagree | neither agree<br>nor disagree | somewhat<br>agree | strongly<br>agree |
| There are <i>benefits</i> to sharing my genetic information<br>(n = 227) | 2 (0.9)                           | 3 (1.3)                       | 57 (25.1)         | 165 (72.7)        |
| There are <i>risks</i> to sharing my genetic information<br>(n = 225)    | 38 (16.9)                         | 20 (8.9)                      | 86 (38.2)         | 81 (36.0)         |
|  | n (%)                             |                               |                   |                   |
| Ranking most important <i>benefit</i> (n = 161)                          |                                   |                               |                   |                   |
| Advancing research that will help me/my family                           | 23 (14.3)                         |                               |                   |                   |
| Advancing research that will help others with a similar condition        | 101 (62.7)                        |                               |                   |                   |
| Gratitude for the care that my family or I have received                 | 0 (0.0)                           |                               |                   |                   |
| Advancing general medical knowledge                                      | 37 (23.0)                         |                               |                   |                   |
| Ranking most important <i>risk</i> (n = 156)                             |                                   |                               |                   |                   |
| Health insurance discrimination  | 44 (28.2)                         |                               |                   |                   |
| Fear of finding out info about myself/family that I don't want to know   | 11 (7.1)                          |                               |                   |                   |
| Having my identity revealed  | 54 (34.6)                         |                               |                   |                   |
| Not knowing what could happen with my genetic information in the future  | 47 (30.1)                         |                               |                   |                   |

data publicly than younger patients or parents of pediatric patients. As one older participant expressed it:

If I was younger, I wouldn't choose [public], but as old as I am, I'll choose that way. I don't think anybody can go on the internet now or get this information and identify me, and I don't think they will be able to in my lifetime, but I believe in the next 50 or 100 years they'll be able to do it real good, so if I was 30 years old then I'd have to watch that 50-year stretch of time. (363)

Compared to this parent of a pediatric patient:

There's a small percentage of people out there that want to get that information for misuse and misconduct. She's only 8. Her life expectancy is maybe another 70, 80 years. I don't [know] what it will be then. You can't have her running around with that risk for 80 years. (544)

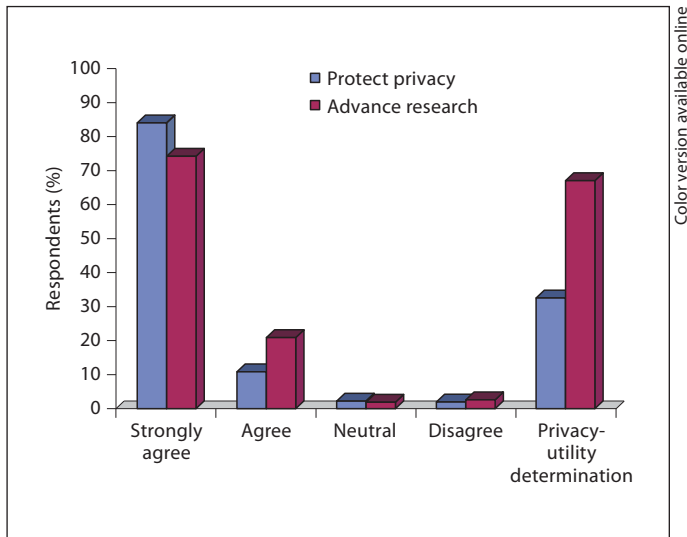
This is also consistent with our previous finding that parents of pediatric patients were significantly more restrictive in their actual data sharing decisions than adult participants [19].

#### *Comparing Hypothetical and Actual Preferences*

As we previously reported, 83.9% of participants initially consented to public data release when they were enrolled into the genome study, and the majority (53%) chose public data release after debriefing [19]. During the

interview, participants were given an opportunity to review the consent documents they did not receive through randomization, express their opinions about these consents and determine what data release selection they think they would have chosen had they been enrolled with a different consent. Results indicate participants' hypothetical choices were inconsistent with their actual data sharing decisions, as participants were generally more restrictive in their hypothetical data sharing preferences than in their actual data release selections.

After reviewing the traditional consent, 30.8% reported they would have declined participation if enrolled with this type of consent. However, as previously reported, all participants randomized to the traditional consent initially agreed to participate and have their data released publicly [19]. Participants reviewing the binary consent also reported that they would have been more restrictive in sharing their genetic data than those who were randomized to this type of consent; hypothetically, 32.9% indicated they would have chosen not to release their data beyond the genome study PI, and 1.3% reported they would have declined participation completely. Of the participants randomized to the binary consent, one participant declined enrollment due to data sharing concerns (the only one in the entire study), but only 15.1% of those



**Fig. 1.** Participants' attitudes toward having their privacy protected and advancing research. In separate questions we asked participants how important it is to them to have their privacy protected ( $n = 228$ ) and to advance research ( $n = 227$ ). Response categories were collapsed from a 5-point Likert scale to a 4-point scale, anchored by strongly agree (5) to disagree (combining strongly disagree (1) and disagree (2)). The final column, privacy-utility determination ( $n = 196$ ), represents respondents' answer when asked to choose which of these, protecting privacy or advancing research, they considered most important.

who enrolled opted out of data sharing. Likewise, the majority (55.8%) of those reviewing the tiered consent reported that they would have chosen restricted data release, but only 19.5% of those randomized to the tiered consent initially chose this option.

Qualitatively, many participants expressed a dislike of public data sharing, despite having already agreed to full public data release upon enrollment into the genome study. For example, one participant who was randomized to the binary consent and originally chose public data release said:

The public thing; it's a little scary. Because I would think that it would only be people that were researching such a diagnosis or prognosis would need to read such a thing. (263)

Another participant also randomized to binary consent and chose public data release said:

I don't want public access. I want it to be scientific databases in this hospital only. I don't want it everywhere. Now if this hospital calls and says, this database we have ... St. Jude's hospital now feels like there's just more research than they can do and we could help them – then yeah. (710)

Interestingly, not all participants who expressed concerns about public data sharing changed to a more restrictive option when provided the opportunity after debriefing.

### Privacy-Utility Determination

One possible explanation for the discordance between hypothetical and actual preferences is the notion of a privacy-utility trade-off [13, 23]. Many of the participants in this study who opted not to change their original consent, despite noted concerns, may have done so because they considered the risks to be minimal, occurring in the future and outweighed by the scientific benefits of making their data broadly accessible. In other words, they may have ultimately decided that the utility of public data release outweighed their real, but less concrete, privacy concerns.

We tested this by asking participants how important it was for them to protect their privacy versus advance research. When asked in separate questionnaire items, the majority (84.2%) of participants strongly agreed that it is important to them to protect their privacy. A smaller, but still significant majority (74%) also strongly agreed that it is important to them to advance research. However, when forced to choose between the two in a single questionnaire item, participants predominantly (67.3%) chose advancing research as more important (fig. 1). This was reflected in open-ended responses to questions about participants' data sharing decisions. For example, one participant explained:

Like I said, the public – the word public scares me. But to me at the time, since I only had that choice, I thought it was worth the risk to help the overall good. (566)

Privacy-utility determinations were significantly associated with participants' actual final data release selections (chi-square test,  $p < 0.001$ ). After debriefing, 42.2% of participants who felt privacy protection was more important than advancing research chose restricted data release, and 26.6% chose no release beyond the study PI, while 56.8% of those who felt advancing research was more important chose full public data release (table 3).

Although participants' trust in doctors doing medical research was not significantly associated with their final data release decisions, trust was significantly associated with their privacy-utility determination (chi-square test,  $p = 0.014$ ). In general, more study participants expressed some to high trust ( $n = 165$ ) than no to low trust ( $n = 31$ ). While participants expressing no to low trust were evenly divided in their privacy-utility determinations, those

**Table 3.** Participants' privacy-utility determination by final data release selection and trust in medical researchers

|   | Which of the following is more important to you? n (%) |                             | p value <sup>a</sup> |
|---|--|-----------------------------|----------------------|
|   | advancing research (n = 132)                           | protecting privacy (n = 64) |                      |
| Final data release selection <sup>b</sup> |  |                             |                      |
| Public                                    | 75 (56.8)  | 20 (31.3)                   | <0.001               |
| Restricted                                | 49 (37.1)  | 27 (42.2)                   |                      |
| No release                                | 8 (6.1)  | 17 (26.6)                   |                      |
| Trust in researchers <sup>c</sup>         |  |                             |                      |
| No to low trust                           | 15 (11.4)  | 16 (25.0)                   | 0.014                |
| Some to high trust                        | 117 (88.6)   | 48 (75.0)                   |                      |

<sup>a</sup> p values reflect the bivariate association (chi-square test) between final data sharing selections and privacy-utility determination.

<sup>b</sup> Final data release selections were made after debriefing when participants were able to choose from all data release options.

<sup>c</sup> Trust was measured with a 5-point Likert scale anchored by strongly disagree (1) to strongly agree (5); for analysis and reporting categories were collapsed to no to low trust (1–3) and some to high trust (4–5).

reporting some to high trust more often selected advancing research as more important than protecting their privacy (table 3).

## Discussion

In the face of evolving regulations and guidelines for genome research, it is critical to understand the perspectives of those who may be impacted by proposed changes; in particular, genome research participants' perspectives should be included. While public opinion data provides valuable insight into general attitudes toward genetic research, this may not necessarily reflect the views of actual genome research participants. A review of 22 studies comparing actual to hypothetical willingness found studies reported more actual participation than suggested in hypothetical survey findings [24]. The authors provide a psychological explanation that these survey respondents may not be as emotionally invested as real research participants. Our findings suggest that even real research participants, presumed to be emotionally invested, make different judgments when responding to hypothetical versus actual choices. Participants were generally much more restrictive in their hypothetical data sharing preferences. Responding to open-ended interview questions on these preferences, participants often expressed concerns about releasing their genetic information into publicly accessible databases. Throughout the study, however, actual data re-

lease decisions were not as restrictive; the majority of participants initially agreed to public data release and a smaller majority still chose public data release after debriefing.

One explanation for this is that participants are making a deliberate privacy-utility trade-off. A significant challenge with data sharing is balancing the risks of the inherent identifiability of DNA data, and the implications this has on privacy protection, with the utility associated with amassing genetic data for analysis. Although participants were concerned with the protection of their privacy, when forced to choose, participants more often chose to help advance research. This privacy-utility determination could underscore the reason why, when faced with no option, all participants randomized to the traditional consent chose to participate in the genome study, despite facing the prospect of open access (i.e. public) data release and why most participants in the study chose public access as their final data release selection. However, because individuals make different privacy-utility determinations, and because these are correlated with actual data sharing decisions, we must recognize and respect the substantial minority of participants who are concerned about privacy protection and prefer not to broadcast their data publicly.

There could be other explanations for why participants were less restrictive in their actual decisions compared to hypothetical preferences. For example, it is possible that some participants did not appreciate that their data sharing decisions were inconsistent with their stated

preferences. Many participants had difficulty understanding complex concepts (like data sharing) and other key elements of their participation (data reported elsewhere, forthcoming), a finding that is consistent with other studies [25, 26] and could have impacted participants' final choices. Alternatively, it may simply have been easier for participants not to change their data sharing decision after debriefing. The majority (67.8%) did not change from their original data release option [19]. The field of behavioral economics provides explanations of cognitive biases that may help explain this trend. The status quo bias posits that when provided alternative choices, participants are more likely to maintain their current choice (i.e. the status quo), although this decision may be inconsistent with their true preferences [27–29]. The study design may have influenced these biases, resulting in fewer people opting to change their decision. For example, after debriefing, participants were given additional data sharing options and then asked to make their final data sharing decision; these added options may have increased the difficulty of their decision-making, resulting in inaction, which is the least complicated decision to make. Additionally, only participants who opted to change their data sharing decision had to sign a new consent form, possibly making the original consent the established, and therefore, the easier choice.

Participants in this study were all recruited into the genome study within a clinical setting and most expressed some to high trust in medical researchers. In most cases, the investigator recruiting them to the genome study was either their own physician or a physician at the hospital where they or their family member was being treated. Additionally, all the genome studies were conducted at BCM in Houston, Texas within the Texas Medical Center, a highly respected institution in the area. These factors could have influenced their decision to participate and to release their data as other studies have noted the influence of trust on willingness to participate in genetic research [24, 30, 31]. Trust was significantly associated with participants' privacy-utility determination as those who expressed some to high trust more often selected advancing research as more important than privacy protection. However, trust was not a significant independent predictor of final data release decisions. Qualitatively, many participants described their trust in their doctor as a factor in their willingness to participate. More research is needed on perspectives of genome research participants from rural areas and groups who historically have displayed no to low trust in medical researchers to further explore this effect.

As genome science continues to advance and lawmakers race to develop and implement sound policies governing this research, the ethical, legal and social implications of these advancements on genome research participants must be carefully considered. In our study, participants expressed a strong desire to be included in data sharing decisions. However, participants varied in terms of their risk-benefit assessments and judgments about the privacy-utility trade-off inherent in decisions about data sharing. To foster public trust and encourage research participation, genome researchers should consider participants' preferences, as well as the overall study design, when deciding upon consent procedures. Small, investigator-initiated studies where data sharing is the secondary, not primary, goal may want to adopt the tiered consent as a way to respect participants' desire for control over who can access their genetic data. However, in studies where the primary goal is to create a community resource (e.g. a biobank), data sharing may be a condition of participation and so tiered consent would not be practical or easy to implement. Future research should focus on alternative, feasible consent procedures for this type of research.

### Acknowledgements

This work was supported by grant NIH R01 HG004333 (A.L. McGuire, S.G. Hilsenbeck, P.A. Kelly, R.A. Gibbs); The Greenwall Foundation Faculty Scholars Program in Bioethics (A.L. McGuire); DLDC P30CA125123 (S.G. Hilsenbeck). We are grateful for the commitment and unwavering support of William Fisher, Alica Goldman, John Goss, Ching Lau, Jeffrey Noebels, Mehmet Okcu, and Diane Treadwell-Deering, and sincerely appreciate all of the thoughtful and generous patients and research participants who participated in this study. We thank Laura Beskow, Wylie Burke, Mildred Cho, Rebecca Fisher, Gail Geller, Laura Lyman Rodriguez, Louise Strong, and Richard Gibbs for their expert advice throughout this project, and Claudette Campbell, Sally E. Hodges, Liz Hinojosa, Melissa Lambeth, Morgan Lasala, Melissa Pagaoa, Suzanne Wheeler, Tiffany Zgabay-Hunsucker, and Jennifer L. Graves for their valuable assistance and research coordination.

### References

- 1 Knoppers BM, Harris JR, Tasse AM, Budin-Ljosne I, Kaye J, Deschenes M, Zawati MH: Towards a data sharing code of conduct for international genomic research. *Genome Med* 2011;3:46.
- 2 National Human Genome Research Institute: Reaffirmation and extension of NHGRI rapid data release policies: large-scale sequencing and other community resource projects, February, 2003.



- 3 Homer N, Szelinger S, Redman M, Duggan D, Tembe W, Muehling J, Pearson JV, Stephan DA, Nelson SF, Craig DW: Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. *PLoS Genet* 2008;4:e1000167.
- 4 Lin Z, Owen AB, Altman RB: Genetics. Genomic research and human subject privacy. *Science* 2004;305:183.
- 5 The Wellcome Trust: Summary of principles agreed at the international strategy meeting on human genome sequencing. Bermuda, February 25th–28th, 1996.
- 6 The Wellcome Trust: Sharing data from large-scale biological research projects: a system of tripartite responsibility. Fort Lauderdale, January 14th–15th, 2003.
- 7 National Human Genome Research Institute: Notice on development of data sharing policy for sequence and related genomic data, 2009.
- 8 National Institutes of Health: Policy for sharing of data obtained in NIH supported or conducted genome-wide association studies (GWAS), 2007.
- 9 Code of Federal Regulations: Title 45, part 46, protection of human subjects.
- 10 Code of Federal Regulations: Title 45, parts 160 and 164, standards for privacy of individually identifiable health information.
- 11 Office for Human Research Protection: Guidance on research involving coded private information or biological specimens. Washington, US Department of Health and Human Services, 2008.
- 12 Human Subjects Research Protections: Enhancing protections for research subjects and reducing burden, delay, and ambiguity for investigators; advance notice of proposed rulemaking. 76 Federal Register 143, July 2011, pp 44512–44531.
- 13 McGuire AL, Hamilton JA, Lunstroth R, McCullough LB, Goldman A: DNA data sharing: research participants' perspectives. *Genet Med* 2008;10:46–53.
- 14 Ludman EJ, Fullerton SM, Spangler L, Trinidad SB, Fujii MM, Jarvik GP, Larson EB, Burke W: Glad you asked: participants' opinions of re-consent for dbGap data submission. *J Empir Res Hum Res Ethics* 2010;5:9–16.
- 15 Lemke AA, Wolf WA, Hebert-Beirne J, Smith ME: Public and biobank participant attitudes toward genetic research participation and data sharing. *Public Health Genomics* 2010;13:368–377.
- 16 Kaufman DJ, Murphy-Bollinger J, Scott J, Hudson KL: Public opinion about the importance of privacy in biobank research. *Am J Hum Genet* 2009;85:643–654.
- 17 McGuire AL, Basford M, Dressler LG, Fullerton SM, Koenig BA, Li R, McCarty CA, Ramos E, Smith ME, Somkin CP, Waudby C, Wolf WA, Clayton EW: Ethical and practical challenges of sharing data from genome-wide association studies: the eMERGE Consortium experience. *Genome Res* 2011;21:1001–1007.
- 18 Pullman D, Etchegary H, Gallagher K, Hodgkinson K, Keough M, Morgan D, Street C: Personal privacy, public benefits, and biobanks: a conjoint analysis of policy priorities and public perceptions. *Genet Med* 2011, E-pub ahead of print.
- 19 McGuire AL, Oliver JM, Slashinski MJ, Graves JL, Wang T, Kelly PA, Fisher W, Lau CC, Goss J, Okcu M, Treadwell-Deering D, Goldman AM, Noebels JL, Hilsenbeck SG: To share or not to share: a randomized trial of consent for data sharing in genome research. *Genet Med* 2011, E-pub ahead of print.
- 20 O'Connor AM: Validation of a decisional conflict scale. *Med Decis Making* 1995;15:25–30.
- 21 Hall MA, Camacho F, Lawlor JS, Depuy V, Sugarman J, Weinfurt K: Measuring trust in medical researchers. *Med Care* 2006;44:1048–1053.
- 22 Hsieh HF, Shannon SE: Three approaches to qualitative content analysis. *Qual Health Res* 2005;15:1277–1288.
- 23 McGuire AL, Gibbs RA: Genetics. No longer de-identified. *Science* 2006;312:370–371.
- 24 Johnsson L, Helgesson G, Rafnar T, Halldorsdottir I, Chia KS, Eriksson S, Hansson MG: Hypothetical and factual willingness to participate in biobank research. *Eur J Hum Genet* 2010;18:1261–1264.
- 25 Beskow LM, Dean E: Informed consent for biorepositories: assessing prospective participants' understanding and opinions. *Cancer Epidemiol Biomarkers Prev* 2008;17:1440–1451.
- 26 Ormond KE, Cirino AL, Helenowski IB, Chisholm RL, Wolf WA: Assessing the understanding of biobank participants. *Am J Med Genet A* 2009;149A:188–198.
- 27 Boonen LH, Donkers B, Schut FT: Channeling consumers to preferred providers and the impact of status quo bias: does type of provider matter? *Health Serv Res* 2011;46:510–530.
- 28 Iyengar SS, Lepper MR: When choice is demotivating: can one desire too much of a good thing? *J Pers Soc Psychol* 2000;79:995–1006.
- 29 Redelmeier DA, Shafir E: Medical decision making in situations that offer multiple alternatives. *JAMA* 1995;273:302–305.
- 30 Bussey-Jones J, Garrett J, Henderson G, Moloney M, Blumenthal C, Corbie-Smith G: The role of race and trust in tissue/blood donation for genetic research. *Genet Med* 2010;12:116–121.
- 31 Critchley CR, Nicol D, Otlowski MF, Stranger MJ: Predicting intention to biobank: a national survey. *Eur J Public Health* 2010, E-pub ahead of print.