

ARTICLE

Research participants' attitudes towards the confidentiality of genomic sequence information

Leila Jamal^{*,1,2}, Julie C Sapp³, Katie Lewis³, Tatiane Yanes⁴, Flavia M Facio³, Leslie G Biesecker³ and Barbara B Biesecker^{3,5}

Respecting the confidentiality of personal data contributed to genomic studies is an important issue for researchers using genomic sequencing in humans. Although most studies adhere to rules of confidentiality, there are different conceptions of confidentiality and why it is important. The resulting ambiguity obscures what is at stake when making tradeoffs between data protection and other goals in research, such as transparency, reciprocity, and public benefit. Few studies have examined why participants in genomic research care about how their information is used. To explore this topic, we conducted semi-structured phone interviews with 30 participants in two National Institutes of Health research protocols using genomic sequencing. Our results show that research participants value confidentiality as a form of control over information about themselves. To the individuals we interviewed, control was valued as a safeguard against discrimination in a climate of uncertainty about future uses of individual genome data. Attitudes towards data sharing were related to the goals of research and details of participants' personal lives. Expectations of confidentiality, trust in researchers, and a desire to advance science were common reasons for willingness to share identifiable data with investigators. Nearly, all participants were comfortable sharing personal data that had been de-identified. These findings suggest that views about confidentiality and data sharing are highly nuanced and are related to the perceived benefits of joining a research study.

European Journal of Human Genetics (2014) 22, 964–968; doi:10.1038/ejhg.2013.276; published online 27 November 2013

Keywords: genomics; privacy; confidentiality

INTRODUCTION

Progress in medical research involving next-generation sequencing of whole human genomes or exomes ('genomic sequencing') depends on the willingness of large numbers of individuals to contribute their genome data to research studies. In order for scientists to glean meaning from these data, they must be accompanied by phenotypic and demographic information. This increases the likelihood that data may be linked back to their original sources, even when de-identified.¹ The future of human genome research therefore relies on large-scale enrollment and public trust. Protecting the confidentiality of genome and other personal data is one way to achieve these aims.²

Familiar concerns about misuses of genetic information relate to risks of insurance discrimination and social stigma.^{3,4} However, genomic sequencing introduces new issues related to the scale of information made available to researchers.⁵ First, genomic sequencing can query nearly all protein-coding regions of the human genome at once, including the majority of genes believed to have roles in disease.⁶ Second, the significance of data generated from a human genome will almost certainly change over a lifetime. Third, it has been demonstrated that failsafe de-identification of human genomic data is not possible.⁷ A fourth point is that genomic sequencing methods have co-evolved with powerful tools for manipulating and sharing these data in large bio-repositories and databases. The relative ease of sharing data among investigators

underscores a need to modernize guidelines for using these data in research now that it can be linked to the individuals who contributed it more readily.⁸

To better understand the attitudes of research participants towards confidentiality and data sharing, we conducted 30 semi-structured interviews with participants in two National Institutes of Health (NIH) research protocols using genomic sequencing to study the basis of human disease.

MATERIALS AND METHODS

Description of studies from which participants were recruited

ClinSeq is a genomic sequencing project investigating the causal role of genetics in cardiovascular and other diseases, enrolling both symptomatic and healthy individuals.⁹ The Whole Genome Medical Sequencing (WGMS) study enrolls children and adults for genomic sequencing with the aim of discovering the genetic aetiology of rare conditions. The two studies are overseen by the same primary investigator and involve overlapping approaches to informed consent.⁶ Evidence suggests that understanding of research intentions is high in both study groups.¹⁰ During the consent process, participants are made aware that some of their study samples or data (such as blood samples or their genetic sequence) may be placed in a public database (dbGaP) and will not contain identifiers. In both protocols, sequence variants deemed clinically relevant are validated in a Clinical Laboratory Improvement Amendments-certified laboratory and returned to the corresponding participant, or their parent if the proband is a minor.⁹

¹Department of Neurogenetics, Kennedy Krieger Institute, Baltimore, MD, USA; ²Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA; ³Genetic Disease Research Branch, National Human Genome Research Institute, Bethesda, MD, USA; ⁴School of Biomolecular and Physical Sciences, Griffith University, Brisbane, Australia; ⁵Social and Behavioral Research Branch, National Human Genome Research Institute, Bethesda, MD, USA

*Correspondence: Ms L Jamal, Department of Neurogenetics, Kennedy Krieger Institute, 707 N. Broadway, Baltimore, Maryland 21205, USA. Tel: +1 44 3923 9543; Fax: +1 44 3923 2795; E-mail: jamal@kennedykrieger.org

Received 12 August 2013; revised 21 October 2013; accepted 24 October 2013; published online 27 November 2013

Recruitment

English-speaking adults who had completed the informed consent process for either the ClinSeq or WGMS protocols and who were willing to be re-contacted about future research opportunities were eligible to participate in this study. Prospective participants were approached by phone within 3 months of their enrollment in one of the two parent protocols. We used a purposive sampling approach to maximize the diversity of the study population with respect to personal and familial illness experiences. A member of the study team called prospective participants to ask if they would participate in a phone interview about topics covered during the informed consent process for the study they were in, including the NIH's plans for data sharing and use. Subsequently, the interviewer (LJ) followed up to obtain verbal informed consent and conduct the interviews. The Institutional Review Board of National Human Genome Research Institute approved this study and authorized a verbal consent process.

Interviews and data analysis

We conducted in-depth, semi-structured phone interviews between May 2011 and November 2012. On average, the interviews lasted for 20–30 min and were taped and transcribed by a professional transcription service. The first three interviews, which lasted ~45 min, were based on a flexible interview guide. From these interviews, we elicited themes that informed the development of a shorter, more focused interview guide. The interview transcripts were imported, coded, and analysed using NVivo 10.0 (QSR International Inc., Burlington, VT, USA). A coding framework was then developed based on these early interviews using an inductive approach.¹¹ The primary coder (LJ) coded all transcripts. A second coder (TY) independently reviewed and coded all transcripts, and consensus was achieved through discussion. Recruitment ended when the research team had agreed that themes related to confidentiality and data sharing had reached saturation. Participants used the terms 'confidentiality' and 'privacy' interchangeably when speaking with us, although strictly defined, researcher–participant confidentiality is the issue we evaluated in this study.

RESULTS

A total of 30 participants were recruited from both the ClinSeq and WGMS studies. Participant characteristics are summarized in Table 1.

We report two categories of results. The first includes prevailing attitudes towards confidentiality among research participants in our sample. The second describes their motivations for sharing personal data with researchers and the conditions that made them feel comfortable doing so.

Locating the value in confidentiality

Confidentiality as a form of control. Informational confidentiality was viewed as a way of limiting how data about oneself may be used by others. A common reason for valuing this kind of control was to prevent employers or insurance companies from using genomic information to treat individuals at risk for disease unfairly. Another reason for valuing this control was a sentiment of caution because of the uncertain duration and results of genomic research. Even among many who felt comfortable releasing their data to the public, there was a sense that relinquishing total control over it would be irresponsible:

'I guess I would think 'What would happen in 50 or 100 years, and you're all over a medical journal?' I actually don't think I would mind that necessarily, but it just made me think twice about it, you know? I guess I'd want to be asked first.' —WGMS#207

The changing nature of genomic research was a reason why many were interested in storing a copy of their genome data on a hard drive for their records or re-analysis. Control over personal data was also

Table 1 Demographic characteristics of participants
Abbreviation: WGMS, Whole Genome Medical Sequencing.

| | ClinSeq | WGMS | Overall study population |
|------------------------------------|------------------------------------|--|--------------------------|
| <i>Age at interview (in years)</i> | | | |
| Average | 50 | 54 | 52 |
| Range | 46–54 | 32–76 | 32–76 |
| <i>Gender (n)</i> | | | |
| Male | 12 | 4 | 16 |
| Female | 5 | 9 | 14 |
| <i>Ethnicity (n)</i> | | | |
| Caucasian | 12 | 13 | 25 |
| African American | 4 | 0 | 4 |
| Asian | 1 | 0 | 1 |
| <i>Relationship to proband</i> | | | |
| Self | 17 | 0 | 17 |
| Parent | 0 | 13 | 13 |
| Diagnoses of probands | Healthy or coronary artery disease | Multiple congenital anomaly and overgrowth syndromes; familial hyperparathyroidism | |

prized because of the role selective information disclosure has in relationships:

'I think what I'd be more concerned about is if for some reason it got out there in the public domain that, you know, my children now, you know, were by sperm donor that would – you know, I'd be — in some ways it wouldn't bother me, but – in fact, it wouldn't bother me at all. What it would bother – you know, the reason it would concern me is that's not the way for your children to find out that kind of information.'—WGMS #203

Participants held different opinions about the practical implications of a right to control their personal information. Some wished to be re-contacted every time their (or their child's) data were requested for a new project or felt that their children should be able to withdraw from a study when they reached adulthood. Many found it impossible to make informed decisions about how to share their personal data until they had a better understanding of what it meant:

'Well, I guess I don't know enough about how they want to use the information. I would have to—that would be one of those things, to me, that each step of the way they would have to say 'now we're going to take your information and do this with it, is that okay? Because I don't have a mindset and I'm not trained enough in genetic testing and processing to know where this might be publicized to say that now.' —WGMS#080

Others felt that by agreeing to the terms of informed consent, they had delegated decisions about future uses of their data to the NIH research team. The lack of consensus about the necessity of obtaining repeat consent for secondary research projects is significant, because some participants could envision their (or their child's) research

participation leading to ends they were uncomfortable with, like the eventual development of pre-natal diagnostics to enable targeted pregnancy termination:

'So that was my issue. And I don't want to contribute to the fact that if they know what causes [my child's condition] now when they test an embryo that has [my child's condition], they're throwing it away, because that's just – I feel like that's basically like saying [my child] is insignificant. That was one of the things I had to really come to terms with.'—WGMS#020

The idea that confidentiality is important as a form of control was strongly endorsed by several participants who believed such control to be an inherent human right, irrespective of any harm that it was intended to protect against:

'I think that, you know, some things — I just believe that people have a right to control their information. It doesn't matter whether anything bad would happen.' —ClinSeq#109

Confidentiality as a form of respect. Many agreed that a core function of confidentiality is to manifest respect between two parties exchanging information. The notion of 'confidentiality as respect' arose when some considered divulging family history information to researchers:

'Some of the questions that came up during genetic counseling I was prepared to answer... I had [family history] information from my brother. He didn't mind giving it, but he didn't give me permission to talk [to researchers] about my nephews, and so I felt uncomfortable giving their names. I'm pretty much aware, you know, that this — the nature of this particular research project is long-term and I'm going to, you know, share as much as I can about myself. With my family members I do have to be careful. I just want to give them the heads-up and be as explicit and respectful with them as I can be.' —ClinSeq#240

Confidentiality was not the only expression of respect valued by participants. Many wished to hear brief updates from investigators even if no individual research results were available. Periodic updates were viewed as an expression of courtesy and reassurance that their data had not been lost or forgotten, and that it was valued:

'Even, even if they just said that, you know, someone stepped on, on your data with a golf shoe and it's, you know, we don't have any data anymore. Because I assume if one aspect of the study might be to find out if there's increased probability of 'x' rather than 'y', then I want to find out, as time goes by, did 'x' happen or did 'y' happen? I would assume that they would contact me from time to time. It would suggest that my contribution is a little more meaningful.' — ClinSeq#050

This quote highlights a common view that within the confines of a researcher–participant relationship bound by rules of confidentiality, participants expected some form of reciprocity from researchers and were willing to remain identifiable to them for this reason.

Confidentiality as determined by personal attributes and context. Regardless of how concerned they were about confidentiality, most participants acknowledged that data sharing standards in research are not a matter of consensus. It was common for participants to have 'devil's advocate' with themselves, imagining how varied circumstances could lead people to define confidentiality differently. Respondents often qualified their opinions with statements about

how their expectations of confidentiality were partly determined by their personal attributes:

'I think it is the way that I am about myself. I think I'm more comfortable about myself and what people know about me. And some people could be embarrassed if it fell into the wrong hands and they could be discriminated against in some way if they're on some file somewhere, maybe have a disability. Those are the only things I could think of from someone else's point of view, that the information could be used against them. For me, I don't have those sorts of issues.'—WGMS #150

Related feelings of 'immunity' to confidentiality breaches were raised by many individuals with stable insurance coverage or secure employment:

'I'm tenured, so I'm not going to lose my job—so I had said to [the genetic counselor], I'm your ideal person. I can't lose my job, and I have great health insurance. So, no matter what I learn, it's not bad. And I said to her, you know, I don't—actually this is one thing that I didn't say earlier which is probably pretty important, is that, you know, if my circumstances were different, maybe I would have thought twice before saying yes I'm going to do this study. You know, could I lose my job if someone got the results? Could I, you know, lose my health insurance, even if I didn't lose my job? That kind of thing.' —WGMS#030

The notion that one's views on data sharing can vary depending on personal attributes and contextual variables was a common reason why participants desired at least some individual control over the use of their information.

Motivations for data sharing

Trust in the NIH. Most participants were unconcerned about confidentiality breaches because of their trust in the NIH. They viewed their research participation as an active choice to share personal data with a specific NIH research team, motivated by a desire to advance their goals. The theme of trust arose in three distinct but mutually reinforcing forms: trust in the NIH researchers, trust in NIH as an institution, and trust in the NIH's policy of genome data de-identification.

Trust in the research team was solidified by an expectation that researchers would maintain at least some contact with participants over time. Nearly, all respondents described feeling 'engaged' with researchers and genetic counselors. By contrast, trust in NIH as an institution was rooted in perceptions of its moral integrity as a government entity. When describing the reasons for trusting the NIH, participants compared it to other types of institutions in which they had less confidence:

'Well the NIH gives me some degree of comfort. I know that the data is going to be handled right. In the commercial world, it gets a little tricky. What I particularly like about this study is that you know I had a baseline at least at one point in life where a lot of this was collected, and as we begin to know more about this you know it's a matter of being able to go back and repeat a test and see if changes occurred... On the commercial side, I get very wary about commercial products that are doing tests, particularly if we don't know how to interpret the tests...The answer would be no. Because I don't know how they would use the data, I don't know how they would interpret it but they'd sure charge a lot for it.' —ClinSeq#160

Related concerns were voiced by participants who felt that having their genomes sequenced by a physician would be more risky than

participating in an NIH study because of the disorganization and opacity of hospital recordkeeping.

Nearly every participant expressed confidence in the NIH policy of genome data de-identification. Even those who feared the consequences of confidentiality breaches were comforted by the knowledge that their genome data were being analysed in de-identified form:

'He explained to me that basically there's only one location where there's a cross-reference between the name of the participant and the identification process they're using on each individual patient's, or study participant's, file. So I don't have any issues with that. I mean, you just, you know—once you give up information you just never know where it goes. And people tell you all kinds of stuff. So, you know, I am going with the full word of the United States government' —ClinSeq#120

Many participants felt comfortable having their data de-identified because they did not believe it compromised the research aims of the study they were joining. Others referenced a tension between de-identification and the goals of genomic research. Several voiced an expectation that in the future, it will be impossible for anyone to be totally anonymous:

'I have to say I go back and forth on this all the time. I want to hope that if I were to join a study that was going to release my data I would be comfortable with it, because I think that ultimately that's what's going to happen with everybody—anononymously, obviously—and it needs to happen for genomics to progress...ultimately, you know, the way things are going...there's going to be no point in keeping it confidential' —ClinSeq#260

Benefits of data sharing. Many believed the benefits of personal data sharing outweighed the merits of anonymity. These individuals expressed a weaker desire to prohibit others from using their data in certain ways, and a stronger desire to specify ways they hoped it would be used. For example, some worried that the investigators might exclude important information from their analyses:

'I don't really have any particular issues with my name being used, you know?...I would hope to instead of being treated purely as just one of the crowd, that my ethnic background would be factored in, and I would be interested in those aspects of the study, you know? Does ethnic background make a difference? Because I'm told, for example you know, South Asians have smaller, thinner blood vessels or whatever... than Caucasians.'—ClinSeq#020

Others felt a stronger desire to advance science than to ensure their identities remained concealed. When describing their motivations for enrolling in research, some referenced their own illness histories or having benefitted from the research enrollment of others. To these individuals, the act of research participation was viewed as a contribution to society:

'The benefits outweigh. You can help people. I'm probably considered a mild case to some people out there who have all these horrible diseases. I just think if I could help someone in the future, it has to be done, really. If you do drug testing, it has to be tested, whether it is on a person or animal. Like when the doctors operated on me many times since I was born, they would not have been able to do what they do without medical research on people before me. Yes for me, it's because I've had this condition since I was born. I'm very familiar with medical stuff, how important research and genetics is.' —WGMS#105

In both protocols, those with personal and family histories of illness tended to prioritize the goal of preventing others from experiencing the same hardships as them. They recalled researching their or their loved one's symptoms on the internet and learning from others who had publicly shared information about themselves:

'My sister had [a form of cancer]...and after going through everything she said she was going to have nine kinds of radiation and each part was deadly. I said, 'are you going to go through with it' and she said 'sure, why not, if it is going to help somebody else, why not?' She didn't make it... but it was her wish that even if she had to die other people had to learn from it...so I don't have a thing if anyone wants to share my results with someone else, that would be ok with me.' —ClinSeq#170.

DISCUSSION

In this qualitative study, we explored the meanings attached to confidentiality and data sharing among participants in two genomic research studies. Our results suggest that beliefs about information use are informed by factors related to situational security and uncertainty, altruism, personality traits, illness histories, and other attributes of context. Because these factors are dynamic and inter-related, we hypothesize that they may change over time. Among the individuals we interviewed, a willingness to share data was often accompanied by feelings of trust in and engagement with the NIH research team.

These findings have implications for efforts to update and harmonize rules governing the use of personal data in genomic research, including those recently announced by a large group of research institutions that plan to form an international framework for genomic data sharing.¹² In the United States, it has been over 2 years since the Office for Human Research Protections (OHRPs) invited the public to comment on its suggestions for modernizing a regulatory infrastructure that predates many of today's research modalities.¹³ One of OHRP's proposals is to establish universal data security protections that are commensurate with the level of 'identifiability' of personal information collected from study participants. OHRP has also suggested that more open-ended informed consent should be obtained at the time of collection for all research specimens, even if they have been stripped of identifiers. However, the proposed reforms say little about how data generated from these specimens should be managed over the long term.

In the European Union, legislation proposed in 2012 was intended to adapt existing information management guidelines (called the Data Protection Directive) to reflect the increased use of digital information. The proposed rule circumscribes the acceptable uses of personal electronic data, including health data, but allows for cases in which it may be used broadly to advance research. Advocates of civil liberties worry that not all research justifies a waiver of individual data protections, whereas others feel the proposed regulations should be relaxed to foster health research.^{14,15} It remains unclear when and how personal data may be appropriated for research without the explicit consent of the people from whom it originated.

Both the OHRP and European Union reform proposals emphasize a view that informational risk is a straightforward function of individual identifiability from research data. Our results echo other studies that have found attitudes towards personal data sharing in research to be more complex than this, influenced by contextual factors and trust in research institutions.^{16,17} Our findings reinforce

that when researchers are trusted, many participants do not mind contributing identifiable personal data to multiple research projects provided that they are kept informed, to some extent, about the nature of the research they are contributing to^{18–20} and that personal data sharing is undertaken more willingly by those who believe that research will yield concrete benefits, either for themselves, society, or both.^{19–21}

We have shown that concerns about personal data sharing are not fully addressed by focusing on the magnitude of the risk of an anonymity breach, and that views about confidentiality have plural origins. This implies that it will remain challenging to include diverse populations in genomic research using a ‘one-size fits all’ approach to informed consent. Future studies should explore an expanded range of strategies for managing participant interests in informational confidentiality. For instance, instead of using a unitary approach to risk reduction, researchers may ameliorate the terms of study participation by increasing the benefits of enrolling in a study, both direct and indirect.

Some proposed reforms would expand the scope of data use research participants may authorize in a single informed consent session, with no requirement to follow-up with them about secondary data uses over the long term. This is concerning given our finding that many participants were motivated to share data with a specific research team and valued occasional updates from them. Given the complexity of attitudes towards data sharing and the emergence of more adaptive, flexible approaches to data governance,^{22–24} we believe new communication strategies should be investigated as means of balancing the benefits and risks of research more favourably. Examples of strategies to explore include: ‘dynamic’ informed consent platforms that provide opt-in choices to research participants as new questions and issues arise over time^{25,26} and user-friendly websites to help participants refresh their understanding of genomics as the field evolves, such as those used in the direct-to-consumer genomics industry.²⁷

Although qualitative data provide valuable insight into conceptually nuanced topics such as confidentiality, our findings are not externally generalizable. Furthermore, our interviews did not systematically elicit separate views on informational privacy and as opposed to researcher–participant confidentiality, a distinction that warrants further exploration. Although studies are needed in larger, more demographically representative groups, our findings may guide the scope and content of these future research projects.

CONFLICT OF INTEREST

LJ, JCS, KL, TY, FF, and BBB declare no conflicts of interest. LGB is an uncompensated adviser to Illumina Corporation and receives royalties from Genentech.

ACKNOWLEDGEMENTS

We thank the participants in this study, Lauren Ivey and Toby Emanuel for coordinating the interviews, and Nancy Kass for reviewing an earlier version of

this manuscript. The Intramural Research Program of the National Human Genome Research Institute funded this research.

- Lin Z, Owen AB, Altman RB: Genetics. Genomic research and human subject privacy. *Science* 2004; **305**: 183.
- Anderlik MR, Rothstein MA: Privacy and confidentiality of genetic information: what rules for the new science? *Annu Rev Genomics Hum Genet* 2001; **2**: 401–433.
- Anderlik MR, Rothstein MA: Privacy and confidentiality of genetic information: what rules for the new science? *Annu Rev Genomics Hum Genet* 2001; **2**: 401–433.
- Johnson S, Kass NE, Natowicz M: Disclosure of personal medical information: differences among parents and affected adults for genetic and nongenetic conditions. *Genetic Testing* 2005; **9**: 269–280.
- Heehey C, Hawkins N, de Vries J, Boddington P, Kaye J: Assessing the privacy risks of data sharing in genomics. *Public Health Genomics* 2011; **14**: 17–25.
- Facio FM, Sapp JC, Linn A, Biesecker LG: Approaches to informed consent for hypothesis-testing and hypothesis-generating clinical genomics research. *BMC Med Genomics* 2012; **5**: 45.
- Gymrek M, McGuire AL, Golan D, Halperin E, Erlich Y: Identifying personal genomes by surname inference. *Science* 2013; **339**: 321–324.
- Tabor HK, Berkman BE, Hull SC, Bamshad MJ: Genomics really gets personal: how exome and whole genome sequencing challenge the ethical framework of human genetics research. *Am J Med Genetics A* 2011; **155A**: 2916–2924.
- Biesecker LG, Mullikin JC, Facio FM *et al*: The ClinSeq Project: piloting large-scale genome sequencing for research in genomic medicine. *Genome Res* 2009; **19**: 1665–1674.
- Sapp J, Dong D, Stark C *et al*: Parental attitudes, values, and beliefs toward the return of results from exome sequencing in children. *Clin Genet* 2013; e-pub ahead of print 20 August 2013; doi:10.1111/cge.12254.
- Bradley EH, Curry LA, Devers KJ: Qualitative data analysis for health services research: developing taxonomy, themes, and theory. *Health Serv Res* 2007; **42**: 1758–1772.
- Hayden EC: Geneticists push for global data-sharing. *Nature* 2013; **498**: 16–17.
- Katsnelson A: Delays in updates to ethics guidelines for research spark concern. *Nature Med* 2013; **19**: 511.
- Carinci F, Di Iorio CT, Ricciardi W, Klazinga N, Verschuuren M: Revision of the European Data Protection Directive: opportunity or threat for public health monitoring? *Eur J Public Health* 2011; **21**: 684–685.
- Coppieters Y, Leveque A: Ethics, privacy and the legal framework governing medical data: opportunities or threats for biomedical and public health research? *Arch Public Health* 2013; **71**: 15.
- Tabor HK, Stock J, Brazg T *et al*: Informed consent for whole genome sequencing: a qualitative analysis of participant expectations and perceptions of risks, benefits, and harms. *Am J Med Genetics A* 2012; **158A**: 1310–1319.
- 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.
- Ludman EJ, Fullerton SM, Spangler L *et al*: Glad you asked: participants’ opinions of re-consent for dbGap data submission. *J Empir Res Hum Res Ethics* 2010; **5**: 9–16.
- Hobbs A, Starkbaum J, Gottweis U, Wichmann HE, Gottweis H: The privacy-reciprocity connection in biobanking: comparing German with UK strategies. *Public Health Genomics* 2012; **15**: 272–284.
- 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.
- Trinidad SB, Fullerton SM, Bares JM *et al*: Genomic research and wide data sharing: views of prospective participants. *Genet Med* 2010; **12**: 486–495.
- Terry SF, Shelton R, Biggers G, Baker D, Edwards K: The haystack is made of needles. *Genet Testing Mol Biomarkers* 2013; **17**: 175–177.
- Trakadis YJ: Patient-controlled encrypted genomic data: an approach to advance clinical genomics. *BMC Med Genomics* 2012; **5**: 31.
- O’Doherty KC, Burgess MM, Edwards K *et al*: From consent to institutions: designing adaptive governance for genomic biobanks. *Social Sci Med* 2011; **73**: 367–374.
- Kaye J: The tension between data sharing and the protection of privacy in genomics research. *Annu Rev Genomics Hum Genet* 2012; **13**: 415–431.
- Kuehn BM: Groups experiment with digital tools for patient consent. *JAMA* 2013; **310**: 678–680.
- Salari K, Karczewski KJ, Hudgins L, Ormond KE: Evidence that personal genome testing enhances student learning in a course on genomics and personalized medicine. *PLoS One* 2013; **8**: e68853.