



It depends whose data are being shared: considerations for genomic data sharing policies

Jill O. Robinson^{1,†}, Melody J. Slashinski^{2,‡}, Elizabeth Chiao^{3,**}
and Amy L. McGuire^{1,*,††}

1. Center for Medical Ethics and Health Policy, Baylor College of Medicine, Houston, TX 77030, USA

2. School of Public Health and Health Sciences, Division of Community Health Education, University of Massachusetts, Amherst, MA 01003, USA

3. Department of Medicine, Baylor College of Medicine, Houston, TX 77030, USA; Houston Health Services Research and Development Center of Excellence, Michael E. DeBakey Veterans Affairs Medical Center, Houston, TX 77030, USA

*Corresponding author. E-mail: amcguire@bcm.edu

ABSTRACT

There is an urgent need for consistent data sharing policies that promote the advancement of science while respecting the values and interests of those providing their genetic data for research. Responding to the article of Jalayne J. Arias, Genevieve Pham-Kanter, and Eric G. Campbell, ‘The Growth and Gaps of Genetic Data Sharing Policies in the United States’, this commentary further explores the challenges of human subjects’ protection in existing data sharing policies. We will elaborate on the need for data sharing policies to accommodate variation in individual and group preferences around data sharing and privacy concerns by comparing our previously published data on patients’ and parents’ consent to data sharing

† Ms Robinson is a Senior Project Manager in the Center for Medical Ethics and Health Policy at Baylor College of Medicine. Her research is on the ethical, legal, and social issues in genomics.

‡ Dr Slashinski is an Assistant Professor in the School of Public Health and Health Sciences at the University of Massachusetts, Amherst. Her research is on the ethical, policy, and public health implications of genomics.

** Dr Chiao is an Associate Professor in Medicine and Infectious Disease at Baylor College of Medicine. Her research is on the epidemiology and outcomes of cancer in the HIV-infected population, as well as the epidemiology, and prevention of Human Papillomavirus.

†† Dr McGuire conducts research on ethical and policy issues in genomics and human genetics. She is the Leon Jaworski Professor of Biomedical Ethics and Director of the Center for Medical Ethics and Health Policy at Baylor College of Medicine.

and attitudes about privacy to data from focus groups with HIV-positive, underserved individuals who were asked about their willingness to participate in genetic research and share their data broadly. These studies support the observation of Arias, Pham-Kanter, and Campbell that researchers, and funding agencies will need to balance the privacy interests of groups as well as individuals in future genomic data sharing policies.

KEYWORDS: data sharing, genomic, research, policy, privacy, trust

INTRODUCTION

In the article, ‘The Growth and Gaps of Genetic Data Sharing Policies in the United States’, Arias, Pham-Kanter, and Campbell provide an overview of the development of data sharing policies for genomic research from the adoption of the 1996 Bermuda Principles to the recently released 2014 National Institutes of Health (NIH) Genomic Data Sharing (GDS) Policy.¹ The authors rightly argue that the development of these policies reflects changes in scientific norms and capabilities, and that the current GDS Policy does not go far enough to address issues regarding privacy, informed consent, protections for data generators, and logistical issues with sharing data. In this commentary, we elaborate on the challenges of human subjects’ protection in existing data sharing policies. Specifically, we suggest that there is a heightened urgency to broadly share diverse sources of data, including genomic data, for research use and clinical application that will shape future policy development. At the same time, there is greater recognition of the identifiability of genomic data and the limits of genomic privacy. Examples of several highly publicized cases highlight the importance of transparency and maintaining the public’s trust, and research has shown that individuals and groups vary in their privacy risk tolerance and data sharing preferences. Data sharing policies must address all of these challenges in order to recruit large representative populations into genomic research. In addition, innovative technical solutions are needed to promote data sharing, while protecting the privacy of those whose data are made available, facilitating transparency, and respecting variability in individuals’ preferences.

DATA SHARING IMPERATIVE

As data sharing initiatives proliferate, the need for robust policies is becoming increasingly urgent. For example, President Obama announced during his 2015 State of the Union address a bold new \$215 million Precision Medicine Initiative, which would include a \$130 million investment to create a US-based research cohort of at least one million people.² In 2012, the Veteran’s Administration announced a similar program to build a ‘mega-database for genomic health’.³ On the private side, Google Genomics allows researchers to store and share genomic data through Google’s cloud platform,⁴ and Apple ResearchKit, which was launched in March 2015, offers

¹ Jalayne J. Arias et al., *The Growth and Gaps of Genetic Data Sharing Policies in the United States*, 2 J. L. & BIOSCI. 56 (2015).

² President Obama’s State of the Union Address: Remarks as prepared for delivery (2015).

³ US Department of Veterans Affairs: Office of Research & Development, *Million Veteran Program*, <http://www.research.va.gov/MVP/> (accessed Apr. 20, 2015).

⁴ Google Cloud Platform, *Google Genomics*, <https://cloud.google.com/genomics/> (accessed Apr. 20, 2015).

an open source software framework that allows researchers to create apps to collect and store health data, including genomic data, directly from consumers through their iPhone.⁵

There are numerous other international public and private efforts to capitalize on the reduced cost of sequencing, the knowledge gained from the Human Genome Project, and developments in digital technology to advance research and improve clinical care. Central to all of these initiatives is their mission to promote open, responsible data sharing. Their success will depend on coherent data sharing policies, as well as technical solutions that allow for interoperability, and the willingness of individuals to broadly share their data.

IDENTIFIABILITY OF GENOMIC DATA

As Arias, Pham-Kanter, and Campbell acknowledge, it is not possible to completely de-identify genomic data. With a reference sample, individuals can be uniquely identified on the basis of a very small amount of sequence data,⁶ and can even be recognized in aggregated data sets.⁷ Even without a reference sample, and with relative ease from a technical perspective, it is possible to link publicly available data in order to identify individuals in research databases with only minimal associated data available.⁸

The NIH GDS policy recognizes the potential identifiability of genomic data and requires informed consent for prospectively collected samples,⁹ even though the federal Common Rule and Office of Human Research Protections still consider research on coded data not to be research involving human subjects.¹⁰ In 2011, the Department of Health and Human Services (HHS) issued an Advanced Notice of Proposed Rule Making to potentially change the Common Rule in recognition of the identifiability of genomic data and require broad consent for unspecified future research on all prospectively collected biospecimens.¹¹ More recently, the Newborn Screening Saves Lives Reauthorization Act of 2014 was passed, which considers research on newborn blood spots to be research involving human subjects, and requires the HHS to promulgate proposed regulations to update the Common Rule and address consent requirements for research involving biospecimens and genomic data.¹²

IMPORTANCE OF PUBLIC ACCEPTANCE AND TRUST

Given this backdrop, the importance of transparent and responsible data sharing policies that address risks to privacy and respect the rights and interests of participants cannot be overstated. The long-term success of the many new research initiatives

⁵ Apple, *Apple ResearchKit*, <https://www.apple.com/researchkit/> (accessed Apr. 20, 2015).

⁶ Zhen Lin et al., *Genomic Research and Human Subject Privacy*, 305 *SCIENCE* 183 (2004).

⁷ Nils Homer et al., *Resolving Individuals Contributing Trace Amounts of DNA to Highly Complex Mixtures Using High-Density SNP Genotyping Microarrays*, 4 *PLOS GENET.* e1000167 (2008).

⁸ Melissa Gymrek et al., *Identifying Personal Genomes by Surname Inference*, 339 *SCIENCE* 321 (2013).

⁹ Final NIH Genomic Data Sharing Policy, 79 *Fed. Reg.* 51,345 (Aug. 28, 2014).

¹⁰ US Department of Health and Human Services, *Office for Human Research Protection: Guidance on Research Involving Coded Private Information or Biological Specimens*, Oct. 16, 2008.

¹¹ Human Subjects Research Protections: Enhancing Protections for Research Subjects and Reducing Burden, Delay, and Ambiguity for Investigators, 76 *Fed. Reg.* 44,512 (July 26, 2011).

¹² Newborn Screening Saves Lives Reauthorization Act of 2014, H.R. 1281 (Dec. 18, 2014).

mentioned above will depend on widespread public acceptance and an unprecedented number of individuals agreeing to participate in genomic research.¹³

With the burden of a history that includes widespread discrimination and social eugenics policies in the United States and abroad,¹⁴ it is conceivable that the field will face serious challenges in garnering the necessary public support required for global data sharing and uptake in research participation. As Arias, Pham-Kanter, and Campbell point out, several widely publicized and unfavorable stories, such as the public release of the sequenced HeLa cell line and the secondary use of the Havasupai tribe's genetic data,¹⁵ have resulted in lawsuits, impeded research, and raised serious concerns about the erosion of public trust.¹⁶

Despite these scandals, empirical evidence suggests that a considerable portion of the general public is willing to participate in genomic research and share their genomic information.¹⁷ However, this willingness to participate should not be interpreted as general license to use individuals' genomic data, as seeking explicit permission for data sharing through informed consent is perceived as a sign of respect¹⁸ and non-negotiable¹⁹ by some research participants.

Furthermore, research suggests that a substantial minority of individuals is willing to participate in genomic research, but only if their data remain private and protected. In a randomized study of consent for data sharing, we found that when given a choice, 47 per cent of research participants ultimately chose not to share their genetic information in an open-access database.²⁰ One third of those participants who chose not to have their data shared publicly opted to restrict sharing of their genetic information to databases accessible only to approved researchers, and 15 per cent opted out of all data sharing. Similarly, a recent study of focus groups with African Americans found that while 80 per cent would hypothetically participate in genomic research if their data were going into a restricted access database, only 50 per cent would do so if their data were going to be shared in an open-access database.²¹

An important factor in data sharing decisions, and a noted limitation of our data sharing study, underscores a crucial challenge potentially facing genomic researchers—participant trust. All participants in our data sharing study were recruited in a clinical

¹³ Jane Kaye & Naomi Hawkins, *Data sharing policy design for consortia: challenges for sustainability*, 6 *GENOME MED.* 4 (2014).

¹⁴ Garland E. Allen, *Eugenics and Modern Biology: Critiques of Eugenics, 1910–1945*, 75 *ANN. HUM. GENET.* 314 (2011).

¹⁵ Arias, *supra* note 1, at 2.

¹⁶ Michelle M. Mello & Leslie E. Wolf, *The Havasupai Indian Tribe Case—Lessons for Research Involving Stored Biologic Samples*, 363 *NEW ENG. J. MED.* 204 (2010).

¹⁷ Amy L. McGuire et al., *To Share or Not to Share: A Randomized Trial of Consent for Data Sharing in Genome Research*, 13 *GENET. MED.* 948 (2011); Amy A. Lemke et al., *Public and Biobank Participant Attitudes Toward Genetic Research Participation and Data Sharing*, 13 *PUB. HEALTH GENOM.* 368 (2010); David J. Kaufman et al., *Public Opinion about the Importance of Privacy in Biobank Research*, 85 *AM. J. HUM. GENET.* 643 (2009).

¹⁸ Jill M. Oliver et al., *Balancing the Risks and Benefits of Genomic Data Sharing: Genome Research Participants' Perspectives*, 15 *PUB. HEALTH GENOM.* 106 (2012); Leila Jamal et al., *Research Participants' Attitudes Towards the Confidentiality of Genomic Sequence Information*, 22 *EUR. J. HUM. GENET.* 964 (2014).

¹⁹ Evette J. Ludman et al., *Glad You Asked: Participants' Opinions of Re-consent for dbGap Data Submission*, 5 *J. EMPIR. RES. HUM. RES. ETHICS* 9 (2010).

²⁰ McGuire, *supra* note 17, at 4.

²¹ Susanne B. Haga & Jennifer O'Daniel, *Public Perspectives Regarding Data-Sharing Practices in Genomics Research*, 14 *PUB. HEALTH GENOM.* 319 (2011).

setting, and the majority exhibited a high degree of trust in medical researchers. Although not an independent predictor of one's data sharing preferences, trust was significantly associated with one's privacy-utility determination; participants expressing high trust in researchers were more likely to agree that advancing research was more important to them than their privacy concerns.²² These findings suggest participant populations that have less trust in medical researchers, such as those who might be vulnerable to discrimination and/or stigmatization may have additional privacy concerns, which could affect their willingness to share their genomic data.

RESPECTING VARYING PERSPECTIVES

Arias, Pham-Kanter, and Campbell argue that data sharing policies need to account for the different issues of addressing risks to individuals' privacy versus considering possible group harms and stigmatization, especially in vulnerable populations. To date, few studies exploring perspectives on genomic data sharing have been conducted with individuals from groups vulnerable to discrimination and/or stigmatization, such as those from marginalized or historically discriminated against groups or those with a stigmatizing condition, like HIV, and thus, may have less trust in medical researchers. Findings from a systematic literature review on the ethical implications of conducting genomic research in populations with a stigmatizing mental health condition, bipolar disorder highlight the importance of protecting these individuals' data and the difficulty in tailoring the informed consent process to do so, given the complexities of genomic research.²³ Additional lessons learned from the lawsuit involving downstream use of the Havasupai tribe's genetic information signify the importance of transparency in obtaining informed consent with groups who might be less trusting of the medical research enterprise.²⁴

As a supplemental project with the Center for AIDS Research in Houston, Texas, we conducted focus groups with HIV-positive individuals (six focus groups, total $n = 40$) who also completed a brief survey that explored attitudes about genetics, research, and their HIV-status, prior research experience, and demographic information. The participants were recruited from individuals receiving social and health services at a public health care clinic that provides comprehensive care for indigent HIV-positive patients. The clinic also conducts social, behavioral, and clinical research, and the population is comprised of individuals who are predominantly from minority backgrounds.

In contrast to findings from our previous study and demonstrating the importance of Arias, Pham-Kanter, and Campbell's argument, when surveyed, the majority (63 per cent) of focus group participants selected privacy protection as more important than advancing research when asked to choose between the two.²⁵ Focus group participants' privacy concerns seemed to be associated with a fear of discrimination and were inextricably linked to their HIV status and the accompanying stigma of this condition. Focus group participants also had less trust in medical researchers (54 per cent reported some to high trust) compared to our HIV-negative data sharing study

²² Oliver, *supra* note 18, at 4.

²³ Iris J. Groisman et al., *Use of Next Generation Sequencing Technologies in Research and Beyond: Are Participants with Mental Health Disorders Fully Protected?*, 13 BMC MED. ETHICS 36 (2012).

²⁴ Mello, *supra* note 16, at 4.

²⁵ Oliver, *supra* note 18, at 4.

population (84 per cent reported some to high trust). This finding is particularly important because, as previously noted, trust has been shown to be associated with privacy-utility determinations.²⁶

Although focus group participants were decidedly more concerned about their health and HIV status, and for some the desire to protect that status, than about advancing research, most were willing to participate in genomic research. Most participants were also willing to allow their genomic information to be stored indefinitely and used for future, unspecified research use, although they questioned why their data would need to be stored rather than analysed now. Also, because this type of research does not provide a direct clinical benefit, participants' willingness to participate was conditional on certain criteria: permission is sought; a thorough explanation of the study is provided, and there is either a direct clinical benefit or some form of reciprocity. We are unsure to what extent these participants clearly understood the meaning of genomic research, data sharing, and its risks and benefits, as we did not objectively measure their understanding and many seemed to have difficulty fully grasping the basic concepts. Interestingly, when surveyed, 47 per cent of focus group participants viewed genetic information as most similar to a medical diagnosis, from a list that included name, social security number, credit card number, and fingerprint. We previously reported findings from our data sharing study that the majority (72 per cent) selected 'fingerprint' as most similar to genetic information.²⁷ These results could indicate that focus group participants understood the personally identifiable nature of DNA data differently and potentially less well than our previous research participants. Although this does not necessarily mean that 'meaningful informed consent is not attainable in genomics research',²⁸ further research is needed to assess what types of information participants must have to provide valid informed consent and how well they need to understand that information to be comfortable and feel like they are making an informed decision.

IMPLICATIONS FOR FUTURE POLICY

Some have suggested that because it is not possible to guarantee privacy, we should 'abandon the traditional concept of medical confidentiality' and only enroll research participants who agree to unrestricted redisclosure of their health information and unrestricted disclosure of any information that is generated in the research, without any promises of confidentiality or privacy.²⁹ However, this would significantly limit participation and may result in biased representation in research. We have therefore argued that instead of limiting research participation to 'information altruists',³⁰ individuals be given control over how broadly they want their data shared.³¹ Our research supports the argument that participants want control, which is viewed as a fulfillment

²⁶ Kaufman, *supra* note 17; Oliver, *supra* note 18, at 4.

²⁷ Jill O. Robinson et al., *Participants' Recall and Understanding of Genomic Research and Large-Scale Data Sharing*, 8 J. EMPIR. RES. HUM. RES. 42 (2013).

²⁸ Arias, *supra* note 1, at 2.

²⁹ Jeantine E. Lunshof et al., *From Genetic Privacy to Open Consent*, 9 NAT. REV. GENET. 406 (2008).

³⁰ Isaac S. Kohane & Russ B. Altman, *Health-information Altruists—A Potentially Critical Resource*, 353 NEW ENG. J. MED. 530 (2005).

³¹ Amy L. McGuire & Richard A. Gibbs, *Genetics. No Longer De-Identified*, 312 SCIENCE 370 (2006).

of the ethical obligation of respect for persons, and when given the option, individuals and groups vary in their data sharing preferences. Further, these findings support Arias, Pham-Kanter, and Campbell's observation that there is an imperative to balance the privacy interests of groups as well as individuals in future genomic data sharing policies.³²

The Global Alliance for Genomics and Health was founded in 2013 as an international collaboration to facilitate data sharing.³³ In the process of catalysing collaborative projects aimed at sharing data, the Global Alliance developed an Application Programming Interface to allow for interoperable exchange of data, and published a Framework for Data Sharing that provides principles for responsible sharing of genomic and health data.³⁴ This Framework calls for, among other things, transparency, participant engagement, and appropriate protection of privacy. The Global Alliance has also developed a consent policy that recognizes that consent is an 'open, communicative, and continuing process', and the right of individuals not to participate in international data sharing, or to withdraw their consent for participation.³⁵

Some have gone further and suggested that, because preferences may change over time, consent should be dynamic, leveraging technological advances to create an opportunity for participants to change their data sharing options over the course of the research.³⁶ Allowing participants to choose how broadly they want their data be shared not only shows respect for variability in individual and group preferences, but can also go a long way toward building and maintaining the public's trust. Future research is needed to assess the ethical and practical impact of affording participants the opportunity to change their consent and control how broadly their data are shared over time.

Finally, substantial resources should go toward creating more comprehensive and cohesive regulatory and governance structures that provide enhanced data privacy protections.³⁷ These new regulations and governance structures must be accompanied by stronger repercussions that carry meaningful penalties for violating codes of use conduct.³⁸ Data sharing policies that emphasize the importance of transparency and respect for individual participants' privacy preferences, rely on technical solutions to protect privacy and data use agreements to ensure confidentiality, and include substantial penalties for intentional breaches in privacy and/or misuse of data, will go a long way toward engendering trust and gaining the public acceptance required for large-scale genomic research.

³² Arias, *supra* note at 1, at 2.

³³ Global Alliance for Genomics & Health, *Regulatory and Ethics Working Group*, <http://genomicsandhealth.org/our-work/working-groups/regulatory-and-ethics-working-group/work-products> (accessed Apr. 20, 2015).

³⁴ Global Alliance for Genomics & Health, *Framework for Responsible Sharing of Genomic and Health-Related Data*, <http://genomicsandhealth.org/about-the-global-alliance/key-documents/framework-responsible-sharing-genomic-and-health-related-data> (accessed Apr. 20, 2015)

³⁵ Global Alliance for Genomics & Health, *Global Alliance for Genomics & Health: Consent Policy*, <http://genomicsandhealth.org/our-work/work-products/consent-policy> (accessed Apr. 20, 2015).

³⁶ Jane Kaye et al., *Dynamic Consent: A Patient Interface for Twenty-First Century Research Networks*, 23 EUR. J. HUM. GENET. 141 (2015).

³⁷ Kaye, *supra* note 13, at 4.

³⁸ COMPILED PUBLIC COMMENTS TO THE DRAFT NIH GENOMIC DATA SHARING POLICY (Sept. 20–Nov. 20, 2013).

ACKNOWLEDGEMENTS

Research supporting this manuscript was funded by the National Human Genome Research Institute (R01HG004333) and the Center for AIDS Research (5P30AI036211-18) 'Ethical Issues in HIV/AIDS Research Involving Specimens' supplement supported by the National Institute of Allergy and Infectious Disease to study ethical issues in HIV/AIDS research involving specimens. Additionally, the authors would like to acknowledge the focus group participants and study coordinators Lena Williams and Donica Beckett. The authors have no financial, personal, academic, or other conflicts of interest in the subject matter discussed in this manuscript.