

INVITED REVIEW

Society and personal genome data

Anna Middleton^{1,2,*}¹Society and Ethics Research Group, Connecting Science, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SA, UK and ²Faculty of Education, University of Cambridge, UK

*To whom correspondence should be addressed. Tel: +44 1223492391; Email: am33@sanger.ac.uk

Abstract

Genomic data offer a goldmine of information for understanding the contribution of genetic variation makes to health and disease. The potential of genomic medicine, to predict, diagnose, manage and treat genetic disease, is underpinned by accurate variant interpretation. This in itself hinges on the ability to access large and varied genomic databases. There is now recognition that international collaboration between research and healthcare systems are paramount to delivering the scale of genomic data required. No single research group, institute or country will liberate our understanding, it is only through global cooperation, together with super computing power, will we truly make sense of how genotype and phenotype correlate. Whilst it is logistically possible to create computing systems that talk to each other and aggregate datasets ready to reveal novel correlations, the bottom line is that this will only happen if *people* (whether they be scientists, clinicians, patients, research participants, policy makers, politicians, law makers) support the principle that we should be donating, accessing and sharing our DNA data in this way. And in order to make the most sense of genomics, given the geographical and ancestral variation between us, such *people* are likely to be the majority of society. Within this review, a perspective is proffered on the human story that underpins genomic 'big data' access and how we are at a tipping point as a society—we need to decide collectively, are we in? and if so, what needs to be in place to protect us? or are we out?

Introduction

According to the Chief Medical Officer to the United Kingdom (UK) government, 'we need to welcome the genomic era and deliver the genomic dream!' (1). Advances in genomic technology mean it is now cheaper and easier than ever before to analyse one or many of a person's 20 000 genes, offering information to help diagnose, manage and treat genetic disease (2).

A defect in a single gene is thought to be the cause of the majority of 'rare' disease worldwide (3). By definition, this means that fewer than 5 people are affected per 10 000 (4), but as there are so many such conditions, collectively they are common (5). Estimates vary, but we think rare, single-gene conditions affect 1 in 17 of us (6).

Therefore, within our own social circles we are each likely to know (or be related to) someone with a serious, potentially life-threatening condition who could, in theory, receive benefit from

a genomic test. Unlocking the molecular basis of rare disease will also be the key to our understanding and treatment of common diseases (7) and we should not underestimate the contribution that genomic¹ knowledge makes to medicine in general (8).

1 Terminology is used very broadly in this review, e.g. 'genomic' could equally refer to either germline or somatic genomic data. 'Genetic' and 'genomic' both refer to the ability to search through DNA to understand and catalogue variation, when only one gene is explored 'genetic' is used, when many genes are explored then 'genomics' is used. We do not specify lab techniques on offer, but they may include, e.g. WGS, WES, microarray, SNP based testing etc. 'Genomic medicine' covers the healthcare discipline using genomic testing to enhance patient care; however, 'genomic medicine' also includes within it a backdrop of basic as well as clinical (both non-profit and for-profit) research that offer the pivotal datasets consulted in variant interpretation.

Received: January 24, 2018. Revised: March 5, 2018. Accepted: March 6, 2018

© The Author(s) 2018. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Projects such as the Deciphering Developmental Disorders (DDD) study (9), offering exome sequencing to children with severe developmental disorders, report that if a clinical exome was offered as a first line diagnostic test, >50% of these children would instantly receive a diagnosis (10). With advances in genomic technology, where relevant, it should be possible to identify the prime genetic cause for every rare disorder. What underpinned the success of the DDD project was the ability to match children at opposite ends of the UK (and now the world) to each other, using a database called DECIPHER (11). As each child's condition was uncommon—and for the doctors caring for that child, they may never have encountered a child with a similar condition before—the DECIPHER database afforded the opportunity to link children with the same genetic result and phenotype. This added to the credibility that the identified variant was indeed the cause of the child's condition.

What typifies the success of genomic medicine is thus collaboration (5,7,12). An integrated systems approach is required—with computers and databases that talk to each other, we need people to enter accurate data into these databases, we need legal systems to enable safe, protected data sharing and we need patients and research participants to contribute their data for use by others. We also need policy recognition that spans borders, jurisdiction and countries, acknowledging that global cooperation is paramount.

Genomic Medicine Depends on Large Genomic Datasets

The journey towards the translation of genomic technology into healthcare and the integration of genomic medicine within healthcare systems has definitely begun, indeed 'genomics is not tomorrow, it is here today' (1, p. 3). We have also reached a stage in history where Big Data and DNA go hand in hand (13).

The datasets required for variant interpretation involve genomic data from people of varying ages, ethnicities, differing stages of health and disease (14). There is global recognition that in order to avoid duplication of datasets and ensure quality assurance, data curation is paramount (15). We also need practical and ethical solutions for linking sequencing data with clinical phenotypes and networking computer systems to each other (14,16–18).

Genomic medicine is a discipline at the cutting edge of science, which relies heavily on input from both non-profit and for-profit research as well as participation with industry (19). Many of the databases that a clinical scientist will access in the NHS when doing genomic variant interpretation (e.g. DECIPHER or ExAC to list but two) are research databases. Within the scope of this review therefore, 'genomic medicine' blurs many boundaries between clinical and research practice. What is consistent, however, is that the evidence-base that guides genomic medicine is underpinned by the ability to consult large scale genomic datasets. And the data that sits in these databases at one point in time belonged to a person who chose (by virtue of a consent process or by condition of access to a health service) for their de-identified data to be donated and accessed by others.

DNA and Big Data

The sharing of generic health data within/between healthcare systems and research has already been happening for many years—and is integral to healthcare evaluation. The concept of

health data sharing is also, broadly speaking, acceptable to the public (20–23). However, we are now entering a new era of connectivity, with plans to link entire health systems, across countries, to each other (24); together with recognition of the importance of collaborating between non-profit and for-profit industries (25).

According to the European Commissioner for Health and Food Safety at the European Parliament: 'Big data has enormous potential to advance medical research, bring about greater innovation in healthcare, and improve the overall performance of health systems' (26, p. 1). The Million European Genomes Alliance (MEGA) aims to be operating by 2020, whereby existing and new research and clinical sequencing programmes connect and share data, crossing geographical borders and legal jurisdictions (24).

Every time a citizen uses a hospital service or participates in a biobank or research project, there is a new opportunity for a fresh set of health data to join the grid. The issue of data donation, access and sharing now appears to be relevant to more of us than ever before—even if we are not personally being asked to donate our data, we might be genetically related to someone who is, and so the decisions they make could be pertinent to us too (27).

The acceptance that one's personal health data will be used by others in the research endeavour is usually predicated on the condition that the data is 'anonymous'² and that careful steps have been taken to protect the donor's privacy. And yet, our genome is our most personal and unique identifier. Unless we have an identical twin, our sequence is the barcode that identifies us as 'us'. Even if only electronic sections of our genome are stored and shared, if we have something particularly rare, then there is always the risk that we can be identified from this alone (18,28–30). This becomes especially relevant if we have other information about ourselves connected to the genomic data (e.g. an unusual clinical characteristic) and also if we have other information in the public domain (e.g. our name appears in an ancestry database or we provide details about our health on Facebook). We must always assume, therefore, that there is a possibility we could be identified if our genomic data sits in a database.

Genomic data is traded, collated, re-categorized and literally bounced around the Internet every second of the day (18); when it has been accessed once and shared with others, it may be difficult to retract (18). Thus, it is negligent to imply that it would be easy or straightforward (or even possible) to withdraw an individuals' genomic data from every imprint on the Internet once it has been deposited and made available for access.

What will help to reassure us that the pros of data sharing outweigh the cons? 'People need to be satisfied that genomic medicine operates in their common interests, whilst protecting their individual privacy, and does not exploit some to benefit others. Protection of individual privacy cannot be absolute, nor can data ever be guaranteed as entirely secure, but there needs to be an understanding of the associated risks and reassurance that breaches are appropriately prosecuted' (31, Chap. 16, p. 4).

2 The terms used to describe the varying levels of anonymity are inconsistent (pseudonomised, de-identified, de-linked might be used interchangeably); very helpful work from the Global Alliance for Genomics and Health provide a clear lexicon of definitions.

Security Breaches

Policy makers and scholars alike recognise that publics³ and patients, need to know that their data is protected, safe and won't lead to personal harm (32). Shabani and Borry point out 'Adopting adequate privacy protections for genomic data has been endorsed by the establishment of the International Declaration on Human Genetic Data, which was issued...by UNESCO as complementary to its Universal Declaration on Human Genome and Human Rights' (33). Thus, as a principle, we should not be discriminated against, on the basis of our genomic information, for example, by employers or insurance companies. This latter point has been subject of considerable debate internationally (34) and indeed in many jurisdictions there are legal protections in place (35). Nevertheless, the *perception* and *fear* of discrimination is significant amongst many publics and so clearer messaging around this would be helpful.

Of particular interest in Europe is the application of the new General Data Protection Regulation and how it will be applied to genetic data (33,36). Here, significant fines will be imposed on the person/organization in control of the data if there is a security breach and individuals are identified.

What harm can actually come of being identified? In reality, 'identification' whilst undesirable, would likely consist of a name and/or other personal identifiers connected to a stretch of DNA code. On its own, and without access to specialist bioinformatics software, no clinical inferences are obvious. However, the personal stigma attached to this becomes apparent if clinical or health data are also revealed at the same time. But again, what actual harm can come of this? The harms can become reality, not from the data being known, per say, but from someone *doing* something with that data. In a court of law a claimant would have to provide tangible evidence of financial, or reputational damage in order to be able to sue for damages; 'embarrassment' on its own, is unlikely to be sufficient.

Application of the GDPR and case law on data security breaches will lead to further clarification on the legal responses to being identified. In the meantime, the Parliamentary Under-Secretary of State for Health published the UK government's response to the Caldicott review and concluded that 'Boosting cyber resilience, improving the response to data and cyber incidents and providing clarity on the handling of personal data remain an urgent priority for the health and care sector' (37).

We Need Empirical Attitude Data from Publics

In order to focus collective efforts on finding out what harms can come, if we are identified from our genomic data, we can usefully ask global publics what they would like to be protected from. Large scale attitude studies are needed, both quantitative and qualitative in method, to explore, from many different angles, what we as a society fear from being identified. Such studies are also useful for gauging the temperature on whether assumptions made in policy (e.g. that we should all donate our data for use in research) are likely to receive a public backlash—and we should never underestimate the power of public dissent to policy rolled out without proper consultation (38).

What is striking from the research that has been done to date on public attitudes is that people are not averse to contributing to research and participants feel a sense of being part of a

'public good' (39, 40), but they become understandably agitated when they are not consulted on the uses of their data (34) or are left with suspicions their contribution will lead to profits for big corporations or companies (41). Research from, for example Wellcome has shown that when time is taken to fully explain the necessity for a partnership between industry and healthcare and that medicines won't be developed without the cooperation of the pharmaceutical industry, publics are more accepting of their data being used in the for-profit industry (42).

Something in Return?

In recognition of the value that individual genomes have in the research endeavour, there have been calls for reciprocity, i.e. giving something back in return for data donation. This could be cash or individual results or raw data (43,44). The ethical issues relating to this are subject to increasing debate (45,46).

Assumed Altruism

The UN Declaration of Human Rights informs us that collectively we all have the right to benefit from science (47). But to benefit from science, in return we should contribute to science (48). Within the patient support groups for people with genetic conditions, we are encouraged to share our data 'it is important to gather information from as many patients as possible' (Genetic Alliance UK website on 'What is patient data and how is it used?'); we are also told that genomic data sharing is there for the 'global good' (14) and genomic data sharing is about 'all of us' (49).

In the 100 000 Genomes Projects, participants do not have a choice in this—participation in the project means that patient data will be accessed and shared with others (50). Thus, within the genomics world, assumptions are already being made about the 'social contract' (48): 'A commitment to open access, global genomic data and knowledge sharing reflects a call for a new social contract based on a principle of solidarity and a duty to act for the common good' (48). This implies that if we expect excellence in our healthcare then we should all be prepared for our data to be used within the evidence base that underpins that healthcare.

Particularly within research, a condition of participation is often that research data (which may include raw sequence data for example) will be deposited in a repository for secondary analysis by others (51,52). However, this brings ethical tensions as it is assumed that people *should* be altruistic with their contribution, drawing on the literature surrounding those who decline genetic testing shows us that people who choose to opt out of testing or showed dissent were seen by their family as difficult or 'lacking in character' (53). This suggests that when there is pressure to be altruistic, declining participation, may be difficult.

Socializing the Benefits of Genomics for Society

What would help here are clear explanations as to the risks and benefits of data donation, access and subsequent sharing. Plus, a global public engagement strategy delivered at a high enough level to have an international reach. It is insufficient to rely on the charitable sector, individual healthcare systems, research projects or even funding bodies to deliver this strategy. Involvement of the World Health Organization and United Nations may be a good start. Whilst it is completely unrealistic to believe that we should achieve some sort of public 'world

3 We use the terms 'publics' instead of 'public' to respect that we (public) are not a singular, homogenous group of people, but a diverse mixture of many individuals and communities.

consensus' on the principle of data donation and sharing, before we can continue, we should at least bring the concepts of a collective altruism into societal discussion and social consciousness. Despite calls for 'public education' in genetics (54), work in the science communication and public engagement literature suggests publics respond to more social and sociable ways of engagement (55) to increase their science capital (56).

At the heart of any future engagement about genomic data access needs to be an exploration of trust (38,57,58), together with a clear explanation of the legal sanctions in place to protect us. We need to feel reassured that our contribution will bring no personal harm to us or our relatives and that we need to understand how our data will contribute to understanding human health. We also need to feel that we have a choice about participation.

Global Alliance for Genomics and health

The Global Alliance for Genomics and Health (GA4GH) is a non-profit organization that unites >500 healthcare providers, universities, research institutes, funders, patient advocacy groups and industry all with the 'shared ideal that sees maximizing the public good as a chief priority of genomic innovation in health' (59). Guided by Human Rights legislation (60) the GA4GH Framework 'provides guidance for the responsible sharing of human genomic and health-related data, including personal health data and other types of data that may have predictive power in relation to health... [and is based on]... human rights of privacy, non-discrimination and procedural fairness' (60).

Historically, researchers have very unhelpfully worked in silos (61), with a resistance to share their data for use by other researchers. Other researchers have wanted to share their data but have had concerns about how best to do this ethically (18). There is now agreement (52,62) that a parochial way of working, which also applies to industry (63), should not be tolerated. Indeed, many funders now require their researchers, as a condition of funding, to deposit raw research data into repositories to that it can be opened up for secondary analysis (52,64). The GA4GH aims to provide the ethical and regulatory infrastructure, IT solutions and practical support to enable data to be shared globally so that we as a society feel comfortable and protected that our altruism is indeed beneficial to others as opposed to harmful to ourselves.

Conclusion

Genomic medicine involves the analysis of a person's genome and the use of this to inform diagnosis, disease management and, where possible, personalized treatment, tailor-made to the individual's germline and/or somatic genomic profile. This is the vision. And yet, 'To realise the full impact of genomic medicine, genomic and clinical data must be interoperable across traditional geographic, jurisdictional, sectoral and domain boundaries. Extremely large and diverse data sets are needed to provide a context for interpretation of genetic sequences. No single country or institution can achieve the necessary scale and diversity alone. Data must be shared' (65, p. 104).

Whilst it is logistically possible to create computing systems that talk to each other and aggregate datasets ready to reveal novel correlations, the bottom line is that this will only happen if *people* (whether they be scientists, clinicians, patients, research participants, policy makers, politicians, law makers) support the principle that we should be donating, accessing and sharing our DNA data in this way. And in order to make the

most sense of genomics, given the geographical and ancestral variation between us, such *people* need to be most of us. It is thus time for global leadership to enable us as a society to be part of the conversation about the use of our data. We need the concepts of genomic data access to be socialized for us so that we can trust that our participation will not harm us but moreover will be good for us and society. We also need a collective buy to the concept of altruism. All of us need to decide, are we in or are we out?

Acknowledgements

We would also like to thank the following people from GA4GH for their encouragement and infrastructure support: Peter Goodhand, Julia Wilson, Bartha Knoppers, Adrian Thorogood, Erika Kleiderman, Ted Dove. Finally, we would like to acknowledge in input from staff at the Wellcome Genome Campus for supporting the administration and project management for the Your DNA, Your Say study (Lauren Farley).

Conflict of Interest statement. None declared.

Funding

This work was supported by Global Alliance for Genomics and Health, with their funding delivered via Wellcome Trust (GA4GH grant, with thanks to Audrey Duncansen). This work has also been supported by Wellcome Trust grant (206194) delivered to A.M. via Connecting Science, Wellcome Genome Campus. Funding to pay the Open Access publication charges for this article was provided by Wellcome grant [206194].

References

1. Davies, S.C. (2017) *Annual Report of the Chief Medical Officer 2016, Generation Genome*. Department of Health, London.
2. Di Sanzo, M., Cipolloni, L., Borro, M., La Russa, R., Santurro, A., Scopetti, M. et al. (2017) Clinical applications of personalized medicine: a new paradigm and challenge. *Curr. Pharm. Biotechnol.*, **18**, 194–203.
3. Chial, H. (2008) Rare genetic disorders: learning about genetic disease through gene mapping, SNPs, and microarray data. *Nat. Educ.*, **1**, 192.
4. Richter, T., Nestler-Parr, S., Babela, R., Khan, Z.M., Tesoro, T., Molsen, E. and Hughes, D.A. (2015) Rare disease terminology and definitions: a systematic global review—report of the ISPOR rare disease special interest group. *Value Health*, **18**, 906–914.
5. Boycott, K.M., Rath, A., Chong, J.X., Hartley, T., Alkuraya, F.S., Baynam, G. et al. (2017) International cooperation to enable the diagnosis of all rare genetic diseases. *Am. J. Hum. Genet.*, **100**, 695–705.
6. Caulfield, M., Thomas, E., Turnbull, C., Scott, R., Rendon, A., Jones, L., Lawson, K., Murugaesu, N., Craig, C., Prapa, M. et al. (2017) 100000 Genomes Project. In: Davies SC, *Annual Report of the Chief Medical Officer 2016*. Department of Health, Genomics London.
7. Julkowska, D., Austin, C.P., Cutillo, C.M., Gancberg, D., Hager, C., Halftermeyer, J., Jonker, A.H., Lau, L.P.L., Norstedt, I., Rath, A. et al. (2017) The importance of international collaboration for rare diseases research: a European perspective. *Gene Therapy*, **24**, 562–571.

8. Boyle, E.A., Li, Y.I. and Pritchard, J.K. (2017) An expanded view of complex traits: from polygenic to omnigenic. *Cell*, **169**, 1177–1186.
9. Wright, C.F., Fitzgerald, T.W., Jones, W.D., Clayton, S., McRae, J.F., van Kogelenberg, M., King, D.A., Ambridge, K., Barrett, D.M., Bayzina, T. et al. (2014) Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. *Lancet*, **385**, 1305–1314.
10. Wright, C.F., McRae, J.F., Clayton, S., Gallone, G., Aitken, S., Fitzgerald, T.W., Jones, P., Prigmore, E., Rajan, D., Lord, J. et al. (2018) Making new genetic diagnoses with old data: iterative reanalysis and reporting from genome-wide data in 1133 families with developmental disorders. *Genet. Med.*, doi: 10.1038/gim.2017.246. [Epub ahead of print]
11. Bragin, E., Chatzimichali, E.A., Wright, C.F., Hurles, M.E., Firth, H.V., Bevan, A.P. and Swaminathan, G.J. (2014) DECIPHER: database for the interpretation of phenotype-linked plausibly pathogenic sequence and copy-number variation. *Nucleic Acids Res.*, **42**, D993–d1000.
12. Austin, C.P., Cutillo, C.M., Lau, L.P.L., Jonker, A.H., Rath, A., Julkowska, D., Thomson, D., Terry, S.F., de Montleau, B., Ardigo, D. et al. (2018) Future of rare diseases research 2017–2027: an IRDiRC perspective. *Clin. Trans. Sci.*, **11**, 21–27.
13. Shi, X. and Wu, X. (2017) An overview of human genetic privacy. *Ann. N. Y. Acad. Sci.*, **1387**, 61–72.
14. Rotimi, C. and Mulder, N. (2014) Data acquisition and data/knowledge sharing in global genomic studies. *App. Trans. Genomics*, **3**, 109–110.
15. Zou, D., Ma, L., Yu, J. and Zhang, Z. (2015) Biological databases for human research. *Genomics Proteomics Bioinformatics*, **13**, 55–63.
16. He, K.Y., Ge, D. and He, M.M. (2017) Big data analytics for genomic medicine. *Int. J. Mol. Sci.*, **18**, 412.
17. Roy, S., LaFramboise, W.A., Nikiforov, Y.E., Nikiforova, M.N., Routbort, M.J., Pfeifer, J., Nagarajan, R., Carter, A.B. and Pantanowitz, L. (2016) Next-generation sequencing informatics: challenges and strategies for implementation in a clinical environment. *Arch. Path. Lab. Med.*, **140**, 958–975.
18. Kaye, J., Heeney, C., Hawkins, N., de Vries, J. and Boddington, P. (2009) Data sharing in genomics: re-shaping scientific practice. *Nat. Rev. Genet.*, **10**, 331–335.
19. UK Government. Strategy for UK life sciences: one year on. 2012. Accessed from: <https://www.gov.uk/government/publications/strategy-for-uk-life-sciences-one-year-on>.
20. Grande, D., Mitra, N., Shah, A., Wan, F. and Asch, D.A. (2013) Public preferences about secondary uses of electronic health information. *JAMA Int. Med.*, **173**, 1798–1806.
21. Weitzman, E.R., Kaci, L. and Mandl, K.D. (2010) Sharing medical data for health research: the early personal health record experience. *J. Med. Int. Res.*, **12**, e14.
22. Weitzman, E.R., Kelemen, S., Kaci, L. and Mandl, K.D. (2012) Willingness to share personal health record data for care improvement and public health: a survey of experienced personal health record users. *BMC Med. Inform. Decis. Mak.*, **12**, 39.
23. Aitken, M., de St Jorre, J., Pagliari, C., Jepson, R. and Cunningham-Burley, S. (2016) Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies. *BMC Med. Ethics*, **17**, 73.
24. Horgan, D., Romao, M. and Hastings, R. (2017) Pulling the strands together: MEGA steps to drive european genomics and personalised medicine. *Biomed. Hub*, **2**, 169–179.
25. Marx, V. (2015) The DNA of a nation. *Nature*, **524**, 503–505.
26. Andriukaitis, V. (2017) Personalised medicine in Europe. *Biomed. Hub.*, **2**, 1–40.
27. Roberts, J. and Middleton, A. (2017) Genetics in the 21st century: implications for patients, consumers and citizens. *F1000Research*, **6**, 2020.
28. Wjst, M. (2010) Caught you: threats to confidentiality due to the public release of large-scale genetic data sets. *BMC. Med. Ethics*, **11**, 21.
29. Malin, B. and Sweeney, L. (2004) How (not) to protect genomic data privacy in a distributed network: using trail re-identification to evaluate and design anonymity protection systems. *J. Biomed. Inform.*, **37**, 179–192.
30. Shringarpure Suyash, S. and Bustamante Carlos, D. (2015) Privacy risks from genomic data-sharing beacons. *Am. J. Hum. Genet.*, **97**, 631–646.
31. Lucassen, A., Montgomery, J. and Parker, M. (2017) Ethics and the social contract for genomics in the NHS. In: Davies SC, *Annual Report of the Chief Medical Officer 2016, Generation Genome*. Department of Health, London.
32. Parker, M. Ethics and new technologies using patient data Presentation given to Wellcome's work on Understanding Patient Data Wellcome; 2017. Accessed from: [https://understandingpatientdata.org.uk/sites/default/files/2017-08/Parker Patient Data.pdf](https://understandingpatientdata.org.uk/sites/default/files/2017-08/Parker%20Patient%20Data.pdf).
33. Shabani, M. and Borry, P. (2018) Rules for processing genetic data for research purposes in view of the new EU General Data Protection Regulation. *Eur. J. Hum. Genet.*, **26**, 149–156.
34. McCormack, P., Kole, A., Gainotti, S., Mascalcioni, D., Molster, C., Lochmuller, H. and Woods, S. (2016) 'You should at least ask'. The expectations, hopes and fears of rare disease patients on large-scale data and biomaterial sharing for genomics research. *Eur. J. Hum. Genet.*, **24**, 1403–1408.
35. Bale, M. (2017) Genomic information and insurance. In: Davies SC, *Annual Report of the Chief Medical Officer 2016, Generation Genome*. Department of Health, London.
36. Chassang, G. (2017) The impact of the EU general data protection regulation on scientific research. *Ecancelmedscience*, **11**, 709.
37. Lord O'Shaughnessy. NHS Technology: Written statement - HLWS41 2017 [updated 12 July 17. Available from: <http://www.parliament.uk/business/publications/written-questions-answers-statements/written-statement/Lords/2017-07-12/HLWS41/>].
38. Carter, P., Laurie, G.T. and Dixon-Woods, M. (2015) The social licence for research: why care.data ran into trouble. *J. Med. Ethics*, **41**, 404–409.
39. Garrison, N.A., Sathe, N.A., Antommaria, A.H., Holm, I.A., Sanderson, S.C., Smith, M.E., McPheeters, M.L. and Clayton, E.W. (2016) A systematic literature review of individuals' perspectives on broad consent and data sharing in the United States. *Genet. Med.*, **18**, 663–671.
40. Dixon-Woods, M. and Tarrant, C. (2009) Why do people cooperate with medical research? Findings from three studies. *Soc. Sci. Med.*, **68**, 2215–2222.
41. Trinidad, S.B., Fullerton, S.M., Bares, J.M., Jarvik, G.P., Larson, E.B. and Burke, W. (2010) Genomic research and wide data sharing: views of prospective participants. *Genet. Med.*, **12**, 486–495.
42. Wellcome. (2016) *The One-Way Mirror: Public Attitudes to Commercial Access to Health Data*. Ipsos MORI report prepared for the Wellcome Trust, London.
43. Wright, C.F., Middleton, A., Barrett, J.C., Firth, H.V., FitzPatrick, D.R., Hurles, M. and Parker, M. (2017) Returning genome

- sequences to research participants: policy and practice [version 1; referees: 2 approved]. *Wellcome. Open. Res.*, **2**, 15.
44. Thorogood, A., Bobe, J., Prainsack, B., Middleton, A., Scott, E., Nelson, S., Corpas, M., Bonhomme, N., Rodriguez, L.L. and Murtagh, M. (2018) APPLaUD: access for patients and participants to individual level uninterpreted genomic data. *Hum. Genomics*, **12**, 7.
 45. Dove, E.S. and Ozdemir, V. (2015) What role for law, human rights, and bioethics in an age of big data, consortia science, and consortia ethics? The importance of trustworthiness. *Laws*, **4**, 515–540.
 46. Parker, M. and Bull, S. (2015) Sharing public health research data: toward the development of ethical data-sharing practice in low- and middle-income settings. *J. Empir. Res. Hum. Res. Ethics*, **10**, 217–224.
 47. Universal declaration of human rights, (1948). Accessed from: <http://www.un.org/en/universal-declaration-human-rights/>.
 48. Carol Isaacson, B. (2014) Global genomic knowledge sharing: a call for affirmative action. *App. Transl. Genomics*, **3**, 79–81.
 49. Aungst, H., Fishman, J.R. and McGowan, M.L. (2017) Participatory genomic research, ethical issues from the bottom up to the top down. *Ann. Rev. Genomics Hum. Genet.*, **18**, 357–367.
 50. Caulfield, M.D.J., Dennys, M., Elbahy, L., Fowler, T., Hill, S., Hubbard, T., Jostins, L., Maltby, N., Mahon-Pearson, J., McVean, G. et al. The 100 000 Genomes Project Protocol 2017 [updated January 2017. Available from: <https://www.genomicsengland.co.uk/library-and-resources/>].
 51. Francis, L.P. (2014) Genomic knowledge sharing: a review of the ethical and legal issues. *App. Trans. Genomics*, **3**, 111–115.
 52. van Schaik, T.A., Kovalevskaya, N.V., Protopapas, E., Wahid, H. and Nielsen, F.G.G. (2014) The need to redefine genomic data sharing: a focus on data accessibility. *App. Trans. Genomics*, **3**, 100–104.
 53. Cowley, L. (2016) What can we learn from patients' ethical thinking about the right 'not to know' in genomics? Lessons from cancer genetic testing for genetic counselling. *Bioethics*, **30**, 628–635.
 54. Dougherty, M.J., Lontok, K.S., Donigan, K. and McInerney, J.D. (2014) The critical challenge of educating the public about genetics. *Curr. Genet. Med. Rep.*, **2**, 48–55.
 55. Parry, V. and Middleton, A. (2017) Socialising the genome. *Lancet*, **389**, 1603–1604.
 56. House of Commons Science and Technology Committee. (2017) *Science Communication and Engagement*. House of Commons, London. Accessed from: <https://publications.parliament.uk/pa/cm201617/cmselect/cmsctech/162/162.pdf>.
 57. Sanderson, S.C., Brothers, K.B., Mercaldo, N.D., Clayton, E.W., Antommara, A.H.M., Aufox, S.A., Brilliant, M.H., Campos, D., Carrell, D.S., Connolly, J. et al. (2017) Public attitudes toward consent and data sharing in biobank research: a large multi-site experimental survey in the US. *Am. J. Hum. Genet.*, **100**, 414–427.
 58. Dame Fiona Caldicott. (2013) *Information: To Share or Not to Share? the Information Governance Review*. Department of Health. Crown copyright, London.
 59. Rahimzadeh, V., Dyke, S.O. and Knoppers, B.M. (2016) An international framework for data sharing: moving forward with the global alliance for genomics and health. *Biopreserv. Biobank*, **14**, 256–259.
 60. GA4GH. Framework for Responsible Sharing of Genomic and Health-Related Data Global Alliance for Genomics and Health 2014 [updated 12 September 2014. Available from: <https://www.ga4gh.org/ga4gh toolkit/regulatoryandethics/framework-for-responsible-sharing-genomic-and-health-related-data/>].
 61. Lochmuller, H., Torrent, I.F.J., Le Cam, Y., Jonker, A.H., Lau, L.P., Baynam, G., Kaufmann, P., Dawkins, H.J., Lasko, P., Austin, C.P. et al. (2017) The international rare diseases research consortium: policies and guidelines to maximize impact. *Eur. J. Hum. Genet.*, **25**, 1293–1302.
 62. Chalmers, D.R.C., Nicol, D. and Otlowski, M.F. (2014) To share or not to share is the question. *App. Trans. Genomics*, **3**, 116–119.
 63. Angrist, M. and Cook-Deegan, R. (2014) Distributing the future: the weak justifications for keeping human genomic databases secret and the challenges and opportunities in reverse engineering them. *App. Trans. Genomics*, **3**, 124–127.
 64. Budin-Ljosne, I., Isaeva, J., Knoppers, B.M., Tasse, A.M., Shen, H.Y., McCarthy, M.I., ENGAGE Consortium and Harris, J.R. (2014) Data sharing in large research consortia: experiences and recommendations from ENGAGE. *Eur. J. Hum. Genet.*, **22**, 317–321.
 65. Scollen, S., Page, A. and Wilson, J. (2017) From the data on many, precision medicine for 'one': the case for widespread genomic data sharing. *Biomed. Hub*, **2**, 21.