

RESEARCH

Open Access



Demonstrating trustworthiness when collecting and sharing genomic data: public views across 22 countries

Richard Milne^{1,2*} , Katherine I. Morley^{3,4,5}, Mohamed A. Almarri^{6,7}, Shamim Anwer⁸, Jerome Atutornu¹, Elena E. Baranova⁹, Paul Bevan⁶, Maria Cerezo¹⁰, Yali Cong¹¹, Alessia Costa¹, Christine Critchley^{12,13}, Josepine Fernow¹⁴, Peter Goodhand¹⁵, Qurratulain Hasan^{16,17}, Aiko Hibino¹⁸, Gry Houeland¹⁴, Heidi C. Howard¹⁹, S. Zakir Hussain¹⁷, Charlotta Ingvaldstad Malmgren^{20,21}, Vera L. Izhevskaya²², Aleksandra Jędrzejak²³, Cao Jinhong²⁴, Megumi Kimura²⁵, Erika Kleiderman²⁶, Brandi Leach³, Keying Liu^{27,28}, Deborah Mascalzoni^{29,14}, Álvaro Mendes³⁰, Jusaku Minari³¹, Dianne Nicol¹³, Emilia Niemiec¹⁴, Christine Patch^{1,32}, Jack Pollard³, Barbara Prainsack^{33,34}, Marie Rivière³⁵, Lauren Robarts¹, Jonathan Roberts¹, Virginia Romano^{14,29}, Haytham A. Sheerah²⁷, James Smith⁶, Alexandra Soulier¹⁴, Claire Steed⁶, Vigdis Stefánsdóttir³⁶, Cornelia Tandre¹⁴, Adrian Thorogood²⁶, Torsten H. Voigt³⁷, Nan Wang¹¹, Anne V. West³⁸, Go Yoshizawa³⁹ and Anna Middleton^{1,40}

Abstract

Background: Public trust is central to the collection of genomic and health data and the sustainability of genomic research. To merit trust, those involved in collecting and sharing data need to demonstrate they are trustworthy. However, it is unclear what measures are most likely to demonstrate this.

Methods: We analyse the 'Your DNA, Your Say' online survey of public perspectives on genomic data sharing including responses from 36,268 individuals across 22 low-, middle- and high-income countries, gathered in 15 languages. We examine how participants perceived the relative value of measures to demonstrate the trustworthiness of those using donated DNA and/or medical information. We examine between-country variation and present a consolidated ranking of measures.

Results: Providing transparent information about who will benefit from data access was the most important measure to increase trust, endorsed by more than 50% of participants across 20 of 22 countries. It was followed by the option to withdraw data and transparency about who is using data and why. Variation was found for the importance of measures, notably information about sanctions for misuse of data—endorsed by 5% in India but almost 60% in Japan. A clustering analysis suggests alignment between some countries in the assessment of specific measures, such as the UK and Canada, Spain and Mexico and Portugal and Brazil. China and Russia are less

(Continued on next page)

* Correspondence: rm23@sanger.ac.uk

¹Society and Ethics Research Group, Wellcome Connecting Science, Wellcome Genome Campus, Cambridge CB10 1SA, UK

²Department of Public Health and Primary Care, University of Cambridge, Cambridge CB2 0SR, UK

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

(Continued from previous page)

closely aligned with other countries in terms of the value of the measures presented.

Conclusions: Our findings highlight the importance of transparency about data use and about the goals and potential benefits associated with data sharing, including to whom such benefits accrue. They show that members of the public value knowing what benefits accrue from the use of data. The study highlights the importance of locally sensitive measures to increase trust as genomic data sharing continues globally.

Background

The future of genomic medicine relies on the ability of researchers and clinicians to access large quantities of genomic and health data. The support of patients and the public for the collection and use of data is central to the success and sustainability of genomic research [1]. However, public willingness to share data and trust in the bodies responsible for the collection and sharing of genomic data varies between countries and between actors involved in the genomic data ecosystem [2, 3]. Trust in the for-profit research sector and governments, for example, is commonly lower than that in non-profit and clinical organisations [4–7]. In this paper, we present findings on public views of measures that may increase trust by ensuring or demonstrating the *trustworthiness* of the organisations, institutions and individuals working with genomic datasets.

The shift in focus from trust to trustworthiness is an important one, recognised in a growing body of work on genomic medicine [2, 8]. Trust involves a relationship between two actors with an expectation of an outcome [9]. Discussion of trust often places the emphasis on the one placing trust, whether a patient or a member of the public. However, these individuals are placing trust that another actor—for example a clinician, researcher or company—is motivated to act to pursue a particular goal [10]. It is here that trustworthiness is critical. If these actors are not trustworthy, trust is neither merited nor meaningful. Trust misplaced in this way has the potential to harm both those who place their trust and those who betray it, for example through long-term impact on reputation or the loss of future research opportunities.

An emphasis on trustworthiness moves the focus away from the public to those involved in collecting and using data and presents an opportunity for the latter to act to exhibit qualities that demonstrate that they are worthy of trust [11]. The meaning of trustworthiness in practice, however, remains unclear, including the activities or measures that show that those collecting and using data are worthy of trust [2, 12–14]. A number of features exhibited by ‘trustworthy’ systems for genomic data have been suggested. These include the importance of establishing shared values and common goals and motives between researchers and participants or members of the public involved in research. This may include

demonstrating a focus on the common good and the equitable distribution of risks and benefits [15]. It may also mean supporting research ethics measures such as informed consent with robust governance that is responsive to stakeholders, respectful, transparent, sustainable, audited and regularly assessed, potentially combined with legal protections [1, 16–19]. Further, work to embed research in relation to local values and goals may be particularly important in addressing potential distrust arising from historical discrimination in healthcare and research [10, 16].

This prior work identifies practices that are already in place in at least some large genomic data initiatives and those to which they and other initiatives might aspire—some that might be new, others that may involve refinement of existing activities. To date, however, discussions of the trustworthiness of genomic and health data initiatives and their relationship with public trust have rarely enabled comparisons between contexts, particularly between countries [3, 5, 20]. This makes it difficult to consider how transferable measures to establish trustworthiness might be and how they are differently valued in different research contexts. Such an international perspective is critical given the need for international data sharing and has the potential to support the development of policies for the sharing of genomic data through initiatives like the Global Alliance for Genomics and Health (GA4GH) [21].

We provide an international perspective through an analysis of the ‘Your DNA, Your Say’ online survey, a study of public perspectives on genomic data sharing that draws on responses from 36,268 individuals across 22 low-, middle- and high-income countries, gathered in 15 languages. We have previously reported on variations in trust within and between these countries, and the relationship between trust and willingness to donate DNA and health information [6, 22]. Here we focus in detail on responses to a question asking participants, who would consider donating DNA or medical information, which measures would help them trust those with whom their data may be shared. Our aim with this question is to draw out how current and potential practices contribute to demonstrating the trustworthiness of data users. We analyse how participants perceived the relative value of these measures, how this varied across the 22

countries of the study, and provide a consolidated ranking of the measures. We then examine similarities and differences between countries, clustering countries that share similar perspectives on the value of these measures. This allows us to consider the implications of our findings for data sharing policies and their applicability across social, cultural and regulatory contexts.

Methods

Sample

Via the international network of researchers affiliated with the GA4GH, the research team invited social science, genetic counselling, bioethics and policy collaborators around the world to participate in conducting the 'Your DNA, Your Say' project, either supporting recruitment into the project and/or translating the survey. For all countries except Japan, Pakistan and India (see below), data were collected using a cross-sectional online survey with participants recruited via market research company Dynata. We aimed to recruit a sample that was as representative as possible of each country's population with regard to gender, age and education level. To this end, participant characteristics were monitored during recruitment to proactively include individuals from under-represented population subgroups. Sociodemographic characteristics of participants from each country are shown in Additional file 1: Table S1 and Additional file 1: Figures S1-S5.

In Japan, participants were recruited through a survey research company (Cross Marketing) using the same approach. In Pakistan and India, recruitment was conducted by market research companies (Foresight and Maction, respectively), and methods were varied to account for lower Internet access. In Pakistan, participants completed the questionnaire on a tablet at a central location rather than at home. In India, participants completed the questionnaire on tablets provided by field researchers. Completed surveys were gathered from Argentina, Australia, Belgium, Brazil, Canada, China, Egypt, France, Germany, India, Italy, Japan, Mexico, Pakistan, Poland, Portugal, Russia, Spain, Sweden, Switzerland, UK and the USA. Participants were paid a small financial reward (<£1) for participating, and due to the nature of the recruitment, there are no details on the non-response rate. The study methodology, design, recruitment strategy, limitations and process of data collection are described separately [23].

Measures

Our online survey can be accessed from www.YourDNAYourSay.org. It contains 29 questions; background information about the landscape of genomic research and data sharing is provided via nine films that

sit within the survey (see Fig. 1); no prior knowledge about genomics is required to participate.

In this paper, we analyse a question that asked participants what information would help them to trust the people asking them to donate DNA information and/or medical information. As shown in Table 1, this question allowed for a structured response, with participants selecting from a range of measures suggested in existing work on the ethics and governance of genomic data platforms. These measures include practices already in place in some, if not many, genomic data initiatives, and those that are more aspirational [24].

One feature of relevance to this question is how we presented information about motives associated with data donation and collection; in the survey films and text, we articulated how different actors may obtain and be motivated by disparate multiple and divergent benefits from the use of donated data. When asking whether participants would be willing to donate their DNA and medical information to medical doctors for use in making a diagnosis in another patient, they were told: 'Whilst there might be benefits to patients from this work, medical doctors might benefit too. For example, through getting more diagnoses for patients and therefore being better at their jobs or getting scientific publications'. When asking whether participants would be willing to donate their DNA and medical information for use by non-profit researchers doing research, for example on how DNA links to disease, they were told: 'There might be benefits to society from this work. But also, individual researchers and organisations might benefit too. For example, individual researchers could advance their career and organisations bring in new funding'. Then finally, when asking whether participants would be willing to donate their DNA and medical information for use by for-profit researchers doing research, for example, developing new medicines, they were told: 'There might be benefits to society from this work. But also, individual researchers and organisations might benefit too. For example, individual researchers might advance their career and companies make a profit'. Thus, we were explicit in giving examples of *how* access and use of data might be motivated by, and lead to, the accrual of personal and organisational, as well as societal and clinical benefit.

We treat countries as the unit of analysis rather than individuals, deriving country-level variables from data collected at the individual level. We use participant-provided information on country of residence and on which measures would help support trust. We focus on those who said that they would potentially be willing to donate DNA or medical information ($n=29,814$); participants who were not willing to donate were excluded from further analyses. The breakdown of refusals by

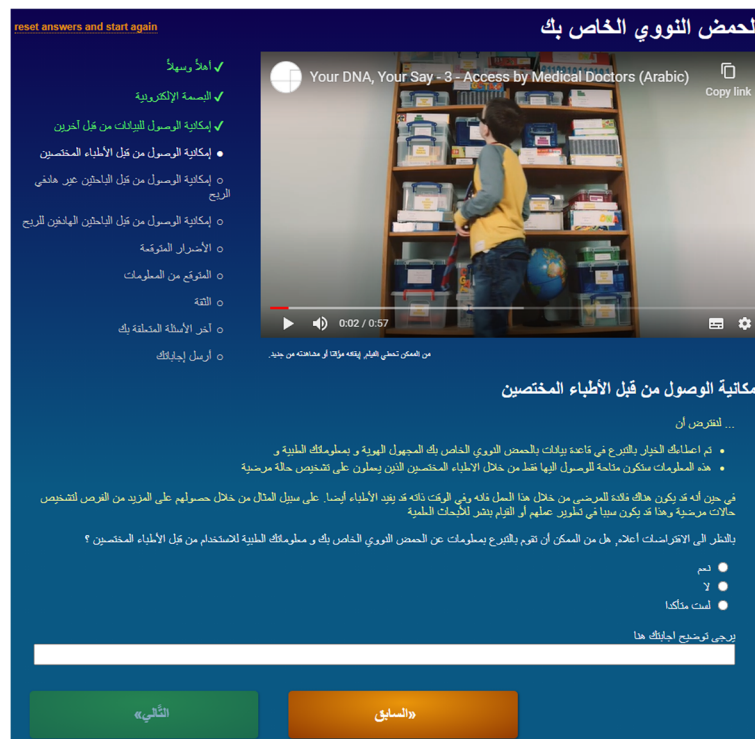


Fig. 1 The appearance of the online Your DNA, Your Say survey (Arabic version). The Your DNA, Your Say questionnaire was presented in 15 languages. Background information on genomic research and data sharing was provided by nine films

country is shown in Additional file 1: Table S2 (range 4% (India) to 33% (Japan); mean 16.6%).

Analysis

Country-level responses

We converted individual responses to this question into a country-level ranked list of these measures from most to least important. This ranking is the variable of interest. We also calculated a ‘global’ ranked list based on overall responses. This was compared with a consensus-generated list derived as part of the analysis (described below).

Consolidated ranking across countries

We used the top-*k* approach [25] to identify common top-ranking responses across countries and to generate a consolidated consensus rank for the measures. We calculated the consensus ranking for the measures using three algorithms (Borda, Markov Chain – majority rule, and the Order Explicit Algorithm) using the *TopKLists* package [26].

Correlation of responses between countries

The rank of the different measures was calculated for each country. We examined pair-wise correlations between all countries, estimated using Kendall’s tau-b

Table 1 YDYS survey question relating to measures that would help people to trust recipients of donated DNA/medical information

<p>Q: What information would help you to trust the people asking you to donate DNA information and/or medical information? (choose all that apply)</p>	<ul style="list-style-type: none"> • Transparent information about WHO will benefit from the data access • Transparent information about HOW others will benefit personally, professionally and commercially from the data access • A website that clearly explains the pros and cons of data access • The option to opt out of having your information accessed by other researchers • The option to withdraw your information in the future • Biographies and photos of the sorts of researchers who would access the data • Knowing exactly who is using your information, and for what purpose • The ability to access your own DNA information and/or medical information • Being able to communicate directly with gatekeepers of my DNA information and/or medical information • Details about the sanctions applicable if my data is misused by others • Other, please provide: • I would not donate my DNA information and/or medical information
--	--

coefficient, which permits estimation of correlations between non-parametric data with ties. In total, there are 231 unique pair-wise correlations; we used a Bonferroni correction to guide interpretation of significance of correlations. We used the *Superheat* package to visualise the pair-wise correlations and generate a cluster dendrogram illustrating similarities between countries [27].

Results

The percentage of participants endorsing each measure to increase trust in the recipients of donated data is shown for the overall sample and by country in Table 2 and Fig. 2. Measures are ordered by ranking in the overall sample.

The figure and table illustrate that there is substantial variability between countries in terms of the ranking of the different measures. This is shown further in the box-plots in Fig. 3.

For all but two countries, provision of transparent information about who will benefit from data access was endorsed by a majority (i.e. > 50%) of participants, including more than 70% of respondents in Egypt, Argentina, Portugal and Switzerland, although China was an extreme outlier with only 32% of respondents endorsing this. Overall, the option to withdraw information in the future was the second most endorsed measure across the sample, with the greatest number of respondents endorsing this option in Australia, Canada, Switzerland and the UK, but only 31% in Egypt.

Endorsement was most variable for details about sanctions for misuse of data, ranging from 5% in India to almost 60% in Japan, where it was the most chosen option. Direct interaction with gatekeepers was a divisive measure with many outliers both above (Australia, Russia, Portugal) and below (Italy, Japan, India) the average endorsement.

Countries varied in the number of measures that were endorsed by a majority of respondents. While five or more measures were endorsed by more than 50% of respondents in the UK, Portugal, Canada and Australia, only one—the ability to opt out of having information access by other researchers—was endorsed by an equivalent proportion of respondents in China.

Consolidated ranking

The consensus results from the top- k approach (Table 3) confirm the rank order obtained from the overall ranking based on percentage endorsement (from most to least important). The only exception is that the ability to access one's own DNA/or medical information is ranked one place above details about sanctions (as opposed to one rank lower based on percentage endorsement).

Correlation of responses between countries

The pair-wise correlation estimates for all country pairs are shown in Fig. 4. Pairwise correlations are shown in Table S3 for all estimates greater than or equal to 0.9. The two strongest correlations, with p values below the Bonferroni-corrected threshold of 0.0002, were between Mexico and Spain, and France and Poland (both at 0.98). The lowest correlation (data not shown) was between Russia and India, with an estimate of -0.16 .

Overall, the heatmap and dendrogram in Fig. 4 show groups of countries for which the rankings of these measures were very similar and that form specific clades. Similarity is not necessarily related to geographical proximity, as for example in the similarity of Spain and Mexico or the UK and Australia. Some countries, notably China and Russia and to some extent India, Egypt, and Japan have consistently lower correlations with most other countries and are linked only at higher levels of the dendrogram—with China and Russia forming a separate, distinct clade from the other 19 countries in the sample.

Discussion

Transparency about aspirations and access

Across countries, measures to demonstrate the trustworthiness of the actors asking for and sharing genomic and health data should focus on transparency about the potential benefits of research, to whom such benefits will accrue and how this will happen. That is, it is not just about describing the general promise and potential of research—such as a benefit to future patients—but also providing an outline of how these benefits will be realised through research. While existing work has suggested the importance of a clear common goal in building trusting relationships with potential data donors, our findings are the first to emphasise the global importance of being clear in this message [16, 17].

It is also important to note, however, that transparency is not limited to outlining expected societal benefits (e.g. finding treatment to a disease) and how they are to be realised. A recurrent feature of research on public trust has been the low levels of trust associated with for-profit or private sector actors, particularly in relation to financial profits for shareholders [5–7]. Yet the benefits motivating the use of research or clinical access to large datasets are not only pecuniary. A further step towards demonstrating trustworthiness, and 'respectful interactions' [17] around data would be for potential data users to reflect on and acknowledge their multiple interests in and motivations for collecting and using health data, and to be transparent about these. Such transparency may enable data donors to more clearly understand what motivates the use of data, and thus make a more confident

Table 2 Percentage of participants endorsing each measure proposed to help to trust recipients of donated DNA/medical information, overall and by country

Measure	Overall	Argentina	Australia	Belgium	Brazil	Canada	China	Egypt	France	Germany	India
Transparent information about WHO will benefit from the data access	61	74	68	58	69	63	32	75	59	57	61
The option to withdraw your information in the future	54	58	67	56	50	62	39	31	57	50	46
Knowing exactly who is using your information, and for what purpose	53	56	61	61	52	56	46	45	53	63	40
Transparent information about HOW others will benefit personally, professionally and commercially from the data access	49	57	62	35	51	54	38	52	41	37	56
The option to opt out of having your information accessed by other researchers	45	33	60	44	45	54	51	34	45	30	42
Details about the sanctions applicable if my data is misused by others	40	48	49	37	36	42	28	44	36	43	5
The ability to access your own DNA and/or medical information	38	40	53	51	24	47	31	28	43	37	27
A website that clearly explains the pros and cons of data access	37	38	49	32	35	40	38	38	28	29	56
Being able to communicate directly with gatekeepers of my DNA and/or medical information	27	30	37	29	30	32	24	32	28	28	13
Biographies and photos of the sorts of researchers who would access the data	21	21	17	14	21	20	34	28	19	15	31
Other	1	1	0	0	1	1	0	2	1	0	1

Table 2 Percentage of participants endorsing each measure proposed to help to trust recipients of donated DNA/medical information, overall and by country (Continued)

Measure	Italy	Japan	Mexico	Pakistan	Poland	Portugal	Russia	Spain	Sweden	Switzerland	UK	USA
Transparent information about WHO will benefit from the data access	66	55	68	61	63	73	47	68	66	70	67	62
The option to withdraw your information in the future	48	55	60	51	58	59	49	58	53	62	61	51
Knowing exactly who is using your information, and for what purpose	45	56	57	47	52	60	62	52	53	63	53	45
Transparent information about HOW others will benefit personally, professionally and commercially from the data access	52	42	59	45	41	61	36	55	54	42	55	53
The option to opt out of having your information accessed by other researchers	34	37	38	44	52	51	37	37	25	49	53	44
Details about the sanctions applicable if my data is misused by others	29	58	45	22	38	49	50	37	36	42	41	29
The ability to access your own DNA and/or medical information	30	28	46	27	45	43	52	40	43	48	42	41
A website that clearly explains the pros and cons of data access	29	36	35	42	31	27	28	29	30	33	46	49
Being able to communicate directly with gatekeepers of my DNA and/or medical information	16	17	30	29	28	35	36	29	22	28	29	26
Biographies and photos of the sorts of researchers who would access the data	13	24	28	23	17	16	20	19	14	13	15	23
Other	1	2	1	3	0	0	1	0	1	0	1	1

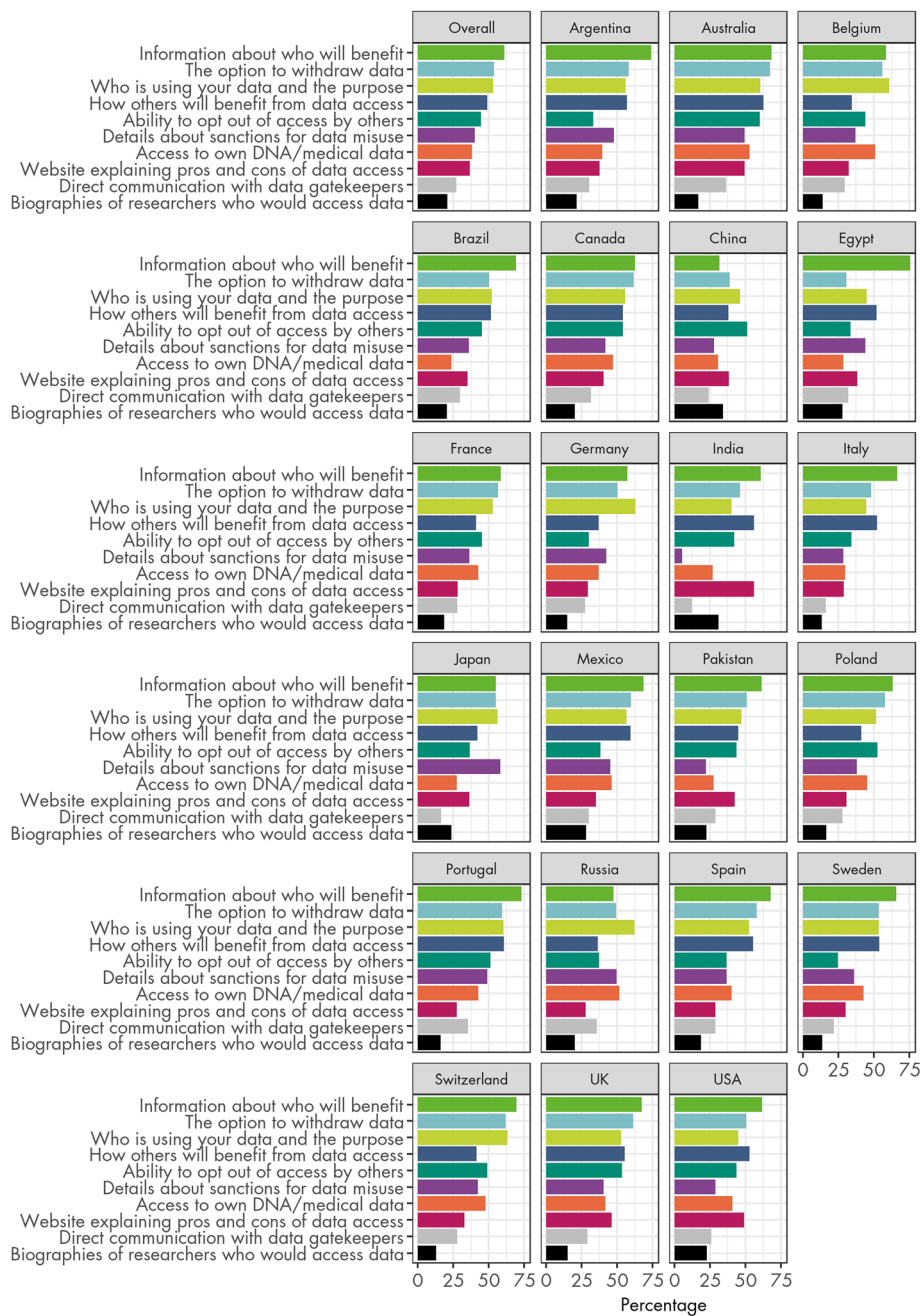


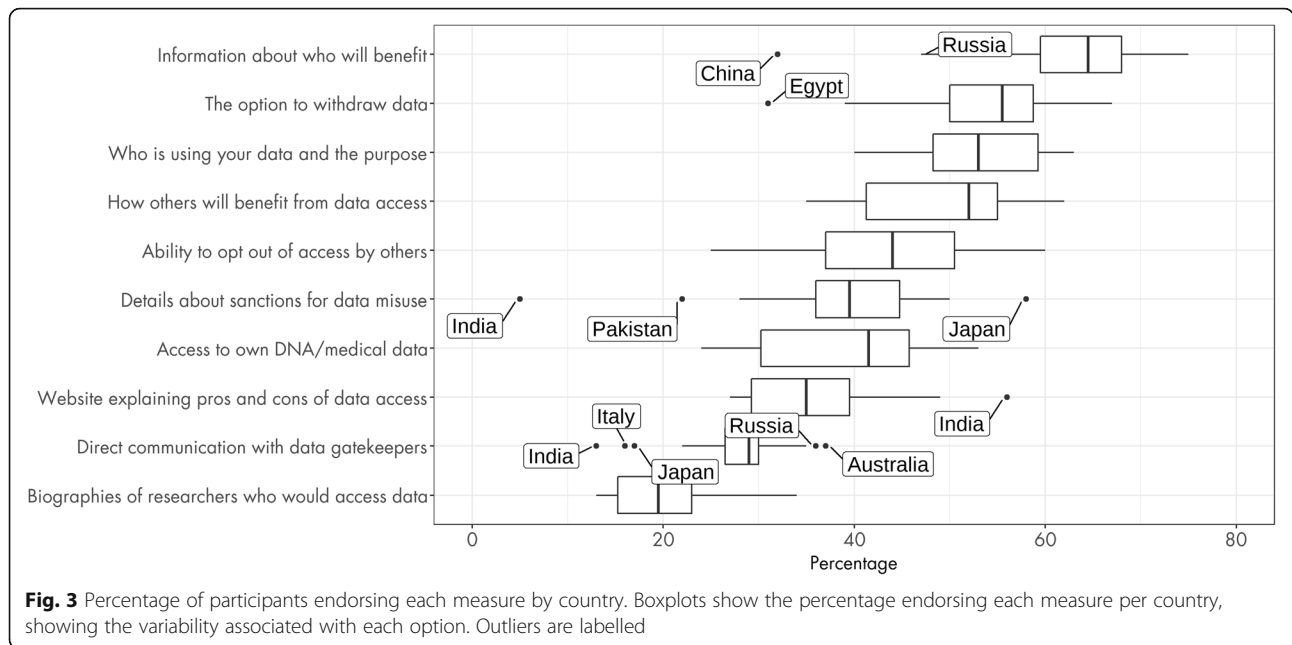
Fig. 2 Measures to help trust recipients of donated DNA/medical information. Percentage of participants endorsing each measure proposed to help to trust recipients of donated DNA/medical information, overall and by country

assessment of a data users’ motivation and the extent to which it reflects a desirable or common goal [10, 15].

A further role of transparency relates to the responsibility of those collecting and using genomic data to provide clear and accessible information on who is using data and for what purposes [28]. Our work suggests that such transparency is indeed an important feature of trustworthy data stewardship across the 22 countries studied. It also draws attention to the potential

importance of familiarity, not just with genetics—as we have highlighted elsewhere—but with the individuals or organisations responsible for genomic data [22, 29].

Our findings suggest, however, that the value of openness does not necessarily extend to individuals’ ability to access their own DNA and health data. Such direct access as a reciprocation of individual contributions was comparatively less important to participants as a measure to increase trust. Thus, while access to data may be



valued by some participants in light of personal interest and perceived personal utility, communicating the potential for such access is not necessarily likely to demonstrate the trustworthiness of actors using of genomic data.

Data rights and regulation

While the measures discussed above concentrate on the goals and roles of genomic research and transparency about access to data, the high ranking of withdrawal and opt outs from data use highlights the importance and value of demonstrating adherence to research ethics frameworks, including informed consent. The right to withdraw, for example, is protected and reinforced throughout clinical research ethics guidance as well as data protection law (e.g. General Data Protection

Table 3 Consolidated ranking of measures to increase trust based on the top-k approach

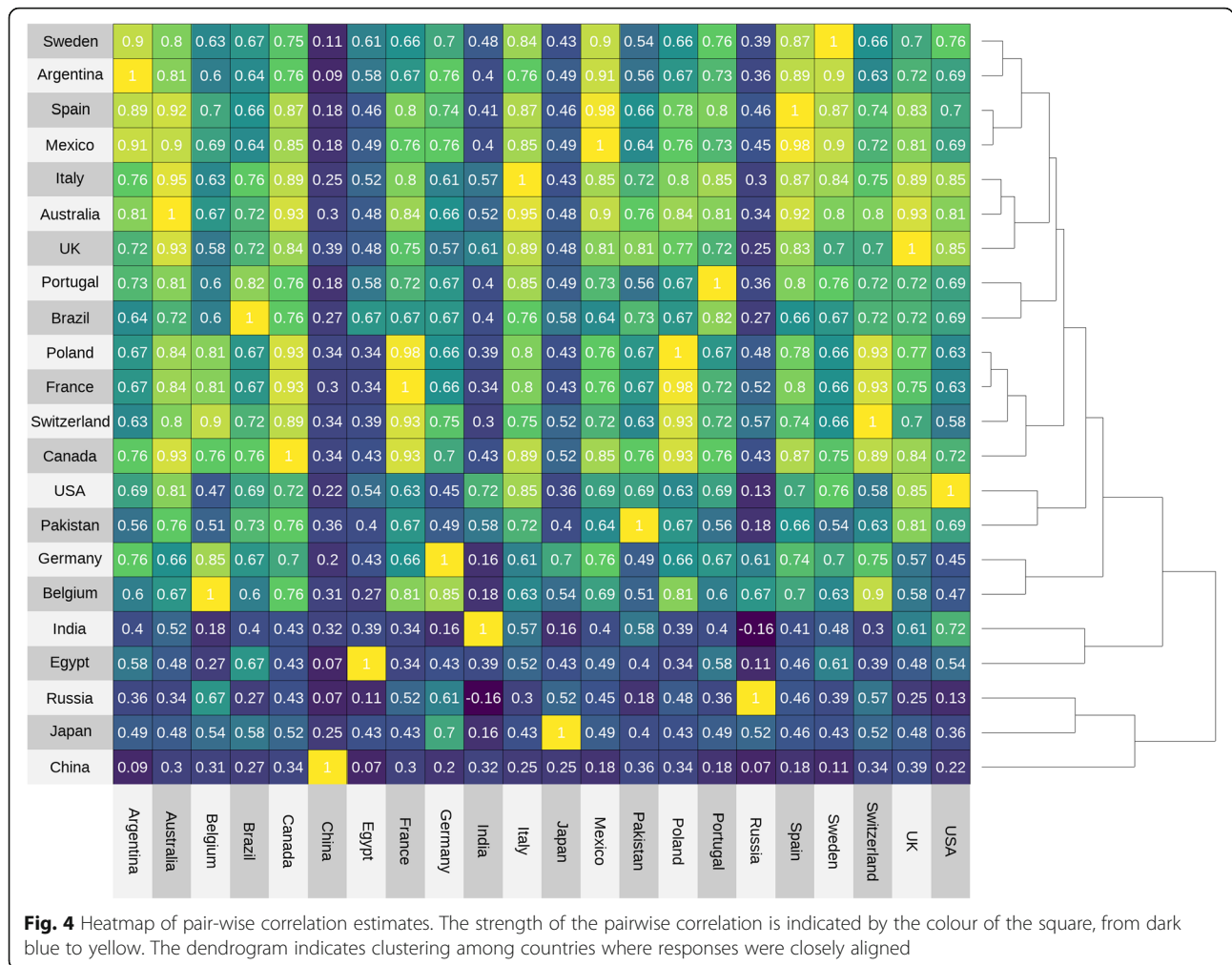
1. Transparent information about WHO will benefit from the data access
2. The option to withdraw your information in the future
3. Knowing exactly who is using your information, and for what purpose
4. Transparent information about HOW others will benefit personally, professionally and commercially from the data access
5. The option to opt out of having your information accessed by other researchers
6. The ability to access your own DNA and/or medical information
7. Details about the sanctions applicable if my data is misused by others
8. A website that clearly explains the pros and cons of data access
9. Being able to communicate directly with gatekeepers of my DNA and/or medical information
10. Biographies and photos of the sorts of researchers who would access the data

Regulation), but presents specific challenges in the case of long-term genomics research [1, 30]. Our findings suggest the value of efforts to reinforce and protect this right, even as data are de-identified and shared, making it clear to data donors that this right exists, but also to be open about its limits and the difficulties associated with tracking shared de-identified data.

The comparatively lesser importance attributed to sanctions for data misuse suggests that regulation and enforcement measures to prevent misuse and exploitation may be important to prevent a loss of trust but make a smaller contribution to demonstrating trustworthiness [14]. Respondents across all countries were less likely to value being able to communicate directly with the gatekeepers of genomic and health data collections, or to be able to access websites about data sharing or see profiles of researchers. The first is interesting, given the importance placed by policy-makers on identifying those responsible for data in regulatory interventions such as European Union’s General Data Protection Regulation (GDPR). The latter suggests that, a website might be of informational value to some, but that efforts to build trust may need more proactive engagement with data donors. This concurs with work that suggests details about individual data users may have little value to data donors without knowledge of why that individual is trustworthy, returning discussion to the measures outlined above [31].

Cross-country consistency and variation

While the overall picture provided by both the consolidated ranking and the individual country rankings



grouped in Figs. 2 and 3 are consistent, our data show variation between countries in their views of measures to enhance trust. The option of communicating with gatekeepers, for example, was selected by a far higher proportion of respondents in Australia, Portugal and Russia than in Japan, Italy or India. Variation is also seen among countries that share, to an extent, legal and decision-making frameworks relevant to genomic data, such as member states of the European Union—although all ten European countries in the sample fall within the same broad clade of 17 countries in Fig. 4, each often has greatest similarity with countries outside the bloc. At a European level, these findings may contribute to the exploration of national variation in the implementation of regulation and legal safeguards derived from the GDPR [32].

Further patterns of consistency and divergence can be seen in the detail of the dendrogram. Some results appear intuitive while others are more unexpected. For example, the clade in which the UK and Australia cluster may be anticipated given their (partly) shared histories,

and similar governance and social healthcare systems. The connection between France and Switzerland might also be expected, particularly among Francophone respondents. Other results are less expected—it is interesting that the USA is not as closely associated with some countries as might be anticipated given geographic proximity and/or shared histories, such as Canada or the UK. The clade that brings together India and Egypt with Japan, Russia and China is also suggestive of interesting directions for further investigation.

As a whole, the patterns of consistency and variation shown here provide nuance to discussions of trustworthiness and present challenges to those developing standards and governance models to facilitate the international sharing of genomic and health data. Most significantly, they suggest that while it is important to work to establish codes of conduct and demonstrate shared values and goals between researchers and data donors, conclusions from such work can only be tentatively extended across national settings. They further point to

the need for detailed comparative work, including qualitative studies, to understand how and why trustworthiness can be demonstrated by the individuals and institutions using genomic and health data.

Limitations

The limitations of the study and design have been published separately [24]. As an exploratory cross-sectional online survey, the study is limited in that it captures intended behaviour at a single time point. Three particular limitations are important to note. Firstly, our analysis is limited to those who would be willing to consider donating their DNA or health information for research. While this includes the majority of the sample, it cannot tell us which measures to increase trust may be more valued by those who definitely will not donate. Second, although the survey was translated and back-translated, nuances of language and culture may affect how participants interpret the options presented. Finally, measures deemed as important in this study, while important and likely necessary, are unlikely to be sufficient on their own to assure potential donors of DNA and health data of the trustworthiness of actors involved in collecting, using and sharing data.

Conclusions

The analysis of responses across the 22 countries included in the Your DNA, Your Say survey suggests practical findings related to demonstrating the trustworthiness of genomic data initiatives, and directions for further research to explore global public perspectives in more detail. They show the importance of research ethics principles related to the right to withdraw, but suggest that legal and regulatory controls may be more important in preventing the loss of trust than in rebuilding it. While highlighting significant variation between countries, they emphasise the importance of transparency about data uses, but particularly about the goals of data collection, and the potential benefits for patients and society, and for data users themselves.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13073-021-00903-0>.

Additional file 1. Additional tables and figures related to the Your DNA, Your Say sample and responses in pdf format.

Acknowledgements

This work was supported by Wellcome grant [206194] to Society and Ethics Research, Connecting Science, Wellcome Genome Campus. We would also like to acknowledge the input of Noémie Aubert Bonn, Universiteit Hasselt, Belgium. In memory of Professor Christine Critchley.

Authors' contributions

RM conceived and drafted the paper and interpreted the data. AM conceived and designed the study and substantively revised the work. HH and EN co-designed the survey and study concepts and revised the work. KM, JP and BL analysed and interpreted the data. PB, JS and CS created the Your DNA, Your Say research platform. All other authors were involved in the acquisition or analysis of data and revised the work. All authors approved the submitted version

Funding

This work was supported by Wellcome Trust grant [206194] to Society and Ethics Research, Connecting Science, Wellcome Genome Campus.

Availability of data and materials

The full dataset is published at <https://societyandethicsresearch.wellcomeconnectingscience.org/project/your-dna-your-say> and available, without restriction, for anyone to access, download and analyse.

Declarations

Ethics approval and consent to participate

The online survey is fully anonymous. Participants are informed that their consent is given when they choose to click off the landing page and start answering the questions. On the landing page, the purpose of the project is explained as well as what participation involves, participants have a choice at any stage within the survey, to stop answering the questions and withdraw. The online project is physically based at the Wellcome Genome Campus with all data collected and stored in encrypted files at the Wellcome Sanger Institute in Cambridge. As part of the conditions of research delivery at this research institution, the project passed ethical review by the Human Materials and Data Management Committee of the Wellcome Sanger Institute (Registration Number: 16/029) as well as legal review to ensure that it was compliant with ethical and legal standards for participant involvement, data collection and storage. This ethics approval was sufficient to cover recruitment into the online survey for most of the collaborators attached to the project; with the exception of Australia, whereby the University of Tasmania required an additional local IRB process to be completed plus their own separate consent form adding onto the landing page of the survey for Australian participants only. The study was approved by the Tasmanian Social Sciences Human Research Ethics Committee on the 5th of July 2017, reference number H0016682. This research conformed to the Declaration of Helsinki.

Consent for publication

Not applicable

Competing interests

KM, DM, QH and SA declare no competing interests associated with their affiliations. The remaining authors declare no competing interests.

Author details

¹Society and Ethics Research Group, Wellcome Connecting Science, Wellcome Genome Campus, Cambridge CB10 1SA, UK. ²Department of Public Health and Primary Care, University of Cambridge, Cambridge CB2 0SR, UK. ³RAND Europe, Cambridge CB4 1YG, UK. ⁴Institute of Psychiatry, Psychology & Neuroscience, King's College London, London SE5 8AF, UK. ⁵Centre for Epidemiology and Biostatistics, Melbourne School of Global and Population Health, The University of Melbourne, Melbourne 3010, Australia. ⁶Wellcome Sanger Institute, Cambridge CB10 1SA, UK. ⁷Department of Forensic Science and Criminology, Dubai Police GHQ, Dubai, United Arab Emirates. ⁸Keynote IAS, New Delhi 110060, India. ⁹Russian Medical Academy of Continuous Professional Education, Moscow 119049, Russia. ¹⁰EMBL-EBI, Wellcome Genome Campus, Cambridge CB10 1SA, UK. ¹¹Medical Ethics Program, Peking University Health Science Center, Beijing 100191, China. ¹²Department of Psychological Sciences, Swinburne University of Technology, Melbourne 3122, Australia. ¹³Centre for Law and Genetics, University of Tasmania, Hobart 7001, Australia. ¹⁴Centre for Research Ethics & Bioethics (CRB), Uppsala University, SE-751 22 Uppsala, Sweden. ¹⁵Ontario Institute for Cancer Research, MaRS Centre, Toronto M5G 0A3, Canada. ¹⁶Department of Genetics & Molecular Medicine, Kamineni Hospitals, Hyderabad 500 068, India. ¹⁷SAAZ Genetics, Hyderabad 500033, India.

¹⁸Faculty of Humanities and Social Sciences, Hiroasaki University, Hiroasaki 036-8560, Japan. ¹⁹Medical Ethics, Lund Universitet, Sölvegatan, 19 Lund, Sweden. ²⁰Department of Public Health and Caring Scienc, Uppsala University, 751 22 Uppsala, Sweden. ²¹Department of Molecular Medicine and Surgery, Karolinska Institutet, 171 76 Solna, Sweden. ²²Research Centre for Medical Genetics, Moscow 115522, Russia. ²³Independent Scholar, Warsaw, Poland. ²⁴Department of Epidemiology and Biostatistics, School of Health Sciences, Wuhan University, Wuhan 430071, China. ²⁵Institute of Innovation Research, Hitotsubashi University, Tokyo 186-8603, Japan. ²⁶Centre of Genomics and Policy, McGill University, Montreal H3A 0G1, Canada. ²⁷Public Health, Department of Social Medicine, Osaka University Graduate School of Medicine, Osaka 565-0871, Japan. ²⁸School of Public Health, Peking University Health Science Center, Beijing 100191, China. ²⁹EURAC, Institute of Biomedicine, 39100 Bolzano, Italy. ³⁰UniGENE and CGPP – Centre for Predictive and Preventive Genetics, IBMC – Institute for Molecular and Cell Biology, i3S – Instituto de Investigação e Inovação em Saúde, Universidade do Porto, 4200-135 Porto, Portugal. ³¹Uehiro Research Division for iPS Cell Ethics, Center for iPS Cell Research and Application (CIRA), Kyoto University, Kyoto 606-8507, Japan. ³²Genomics England, Queen Mary University of London, London EC1M 6BQ, UK. ³³Department of Political Science, University of Vienna, 1010 Vienna, Austria. ³⁴Department of Global Health & Social Medicine, King's College London, London WC2R 2LS, UK. ³⁵DILTEC, Sorbonne Nouvelle, 75005 Paris, France. ³⁶Landsþítali, the National University Hospital of Iceland, 101 Reykjavík, Iceland. ³⁷Institute of Sociology, RWTH Aachen University, 52062 Aachen, Germany. ³⁸Indiana University Maurer School of Law, Bloomington 47405, USA. ³⁹Work Research Institute (AFI), Oslo Metropolitan University, 0130 Oslo, Norway. ⁴⁰Faculty of Education, University of Cambridge, Cambridge CB2 8PQ, UK.

Received: 20 October 2020 Accepted: 4 May 2021

Published online: 25 May 2021

References

- Kaye J. The tension between data sharing and the protection of privacy in genomics research. *Annu Rev Genomics Hum Genet.* 2012;13(1):415–31. <https://doi.org/10.1146/annurev-genom-082410-101454>.
- Clayton EW, Halverson CM, Sathe NA, Malin BA. A systematic literature review of individuals' perspectives on privacy and genetic information in the United States. *PLoS One.* 2018;13(10). <https://doi.org/10.1371/journal.pone.0204417>.
- Ghafur S, Dael JV, Leis M, Darzi A, Sheikh A. Public perceptions on data sharing: key insights from the UK and the USA. *Lancet Digit Health.* 2020;0(0). [https://doi.org/10.1016/S2589-7500\(20\)30161-8](https://doi.org/10.1016/S2589-7500(20)30161-8).
- Critchley C, Nicol D, Otlowski M. The impact of commercialisation and genetic data sharing arrangements on public trust and the intention to participate in biobank research. *Public Health Genomics.* 2015;18(3):160–72. <https://doi.org/10.1159/000375441>.
- Gaskell G, Gottweis H, Starkbaum J, Gerber MM, Broerse J, Gottweis U, et al. Publics and biobanks: Pan-European diversity and the challenge of responsible innovation. *Eur J Hum Genet.* 2013;21(1):14–20. <https://doi.org/10.1038/ejhg.2012.104>.
- Milne R, Morley KI, Howard H, et al. Trust in genomic data sharing among members of the general public in the UK, USA, Canada and Australia. *Hum Genet.* 2019;138(11):1237–46. <https://doi.org/10.1007/s00439-019-02062-0>.
- Voigt TH, Holtz V, Niemiec E, Howard HC, Middleton A, Prainsack B. Willingness to donate genomic and other medical data: results from Germany. *Eur J Hum Genet.* 2020;28(8):1000–9. <https://doi.org/10.1038/s41431-020-0611-2>.
- Dheensa S, Samuel G, Lucassen AM, Farsides B. Towards a national genomics medicine service: the challenges facing clinical-research hybrid practices and the case of the 100 000 genomes project. *J Med Ethics.* 2018;44(6):397–403. <https://doi.org/10.1136/medethics-2017-104588>.
- Hardin R. *Trust and Trustworthiness.* New York: Russell Sage Foundation; 2002.
- Baier A. *Trust and Antitrust.* Ethics. 1986;96(2):231–60. <https://doi.org/10.1086/292745>.
- O'Doherty KC, Burgess MM, Edwards K, Gallagher RP, Hawkins AK, Kaye J, et al. From consent to institutions: Designing adaptive governance for genomic biobanks. *Soc Sci Med.* 2011;73(3):367–74. <https://doi.org/10.1016/j.socscimed.2011.05.046>.
- Nicol D, Critchley C, McWhirter R, Whitton T. Understanding public reactions to commercialization of biobanks and use of biobank resources. *Soc Sci Med.* 2016;162:79–87. <https://doi.org/10.1016/j.socscimed.2016.06.028>.
- Burke W, Beskow LM, Trinidad SB, Fullerton SM, Brelford K. Informed consent in translational genomics: insufficient without trustworthy governance. *J Law Med Ethics.* 2018;46(1):79–86. <https://doi.org/10.1177/1073110518766023>.
- Sheehan M, Friesen P, Balmer A, Cheeks C, Davidson S, Devereux J, et al. Trust, trustworthiness and sharing patient data for research. *J Med Ethics.* 2020: medethics-2019-106048. <https://doi.org/10.1136/medethics-2019-106048>.
- Deverka PA, Gilmore D, Richmond J, Smith Z, Mangrum R, Koenig BA, et al. Hopeful and concerned: public input on building a trustworthy medical information commons. *J Law Med Ethics.* 2019;47(1):70–87. <https://doi.org/10.1177/1073110519840486>.
- Kraft SA, Cho MK, Gillespie K, Halley M, Varsava N, Ormond KE, et al. Beyond consent: building trusting relationships with diverse populations in precision medicine research. *Am J Bioeth.* 2018;18(4):3–20. <https://doi.org/10.1080/15265161.2018.1431322>.
- Xafis V, Schaefer GO, Labude MK, Brassington I, Ballantyne A, Lim HY, et al. An ethics framework for big data in health and research. *ABR.* 2019;11(3): 227–54. <https://doi.org/10.1007/s41649-019-00099-x>.
- Nicol D, Eckstein L, Bentzen HB, Borry P, Burgess M, Burke W, et al. Consent insufficient for data release. *Science.* 2019;364(6439):445–6. <https://doi.org/10.1126/science.aax0892>.
- Dixon-Woods M, Ashcroft RE. Regulation and the social licence for medical research. *Med Health Care Philos.* 2008;11(4):381–91. <https://doi.org/10.1007/s11019-008-9152-0>.
- Sheikh Z, Hoeyer K. "That is why I have trust": unpacking what 'trust' means to participants in international genetic research in Pakistan and Denmark. *Med Health Care Philos.* 2018;21(2):169–79. <https://doi.org/10.1007/s11019-017-9795-9>.
- Stark Z, Dolman L, Manolio TA, Ozenberger B, Hill SL, Caulfield MJ, et al. Integrating genomics into healthcare: a global responsibility. *Am J Hum Genet.* 2019;104(1):13–20. <https://doi.org/10.1016/j.ajhg.2018.11.014>.
- Middleton A, Milne R, Almarri M, Anwer S, Atutornu J, et al. Global public perceptions of genomic data sharing: what shapes the willingness to donate DNA and health data? *Am J Hum Genet.* 2020;107(4):743–52. <https://doi.org/10.1016/j.ajhg.2020.08.023>.
- Middleton A, Your DNA. *Your Say.* New Bioeth. 2017;23(1):74–80. <https://doi.org/10.1080/20502877.2017.1314890>.
- Middleton A, Niemiec E, Prainsack B, Bobe J, Farley L, Steed C, et al. 'Your DNA, Your Say': global survey gathering attitudes toward genomics: design, delivery and methods. *Personalized Med.* 2018;15(4):311–8. <https://doi.org/10.2217/pme-2018-0032>.
- Hall P, Schimek MG. Moderate-deviation-based inference for random degeneration in paired rank lists. *J Am Stat Assoc.* 2012;107(498):661–72. <https://doi.org/10.1080/01621459.2012.682539>.
- Schimek MG, Budinská E, Kugler KG, Švendová V, Ding J, Lin S. TopKLists: a comprehensive R package for statistical inference, stochastic aggregation, and visualization of multiple omics ranked lists. *Stat Appl Genet Mol Biol.* 2015;14(3):311–6. <https://doi.org/10.1515/sagmb-2014-0093>.
- Barter RL, Yu B. Superheat: An R package for creating beautiful and extendable heatmaps for visualizing complex data. *J Comput Graphical Stat.* 2018;27(4):910–22. <https://doi.org/10.1080/10618600.2018.1473780>.
- Byrd JB, Greene AC, Prasad DV, Jiang X, Greene CS. Responsible, practical genomic data sharing that accelerates research. *Nat Rev Genet.* 2020;10(1):1–15. <https://doi.org/10.1038/s41576-020-0257-5>.
- Passmore SR, Jamison AM, Hancock GR, Abdelwadoud M, Mullins CD, Rogers TB, et al. "I'm a Little More Trusting": components of trustworthiness in the decision to participate in genomics research for African Americans. *PHG.* 2019;22(5-6):215–26. <https://doi.org/10.1159/000505271>.
- Staunton C, Slokenberga S, Mascalzoni D. The GDPR and the research exemption: considerations on the necessary safeguards for research biobanks. *Eur J Hum Genet.* 2019;27(8):1159–67. <https://doi.org/10.1038/s41431-019-0386-5>.
- Ipsos MORI. *The One-Way Mirror: Public Attitudes to Commercial Access to Health Data.* London: The Wellcome Trust; 2016.
- Molnár-Gábor F, Korbel JO. Genomic data sharing in Europe is stumbling—Could a code of conduct prevent its fall? *EMBO Mol Med.* 2020;12(3):e11421. <https://doi.org/10.15252/emmm.201911421>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.