

## VIEWPOINT



# Old and new challenges regarding comparable and viable data sharing in population-scale genomic research

Aviad Raz<sup>1</sup>✉, Jusaku Minari<sup>2</sup>, Kayo Takashima<sup>2</sup>, Hristina Gaydarska<sup>2</sup>, Yael Hashiloni-Dolev<sup>1</sup> and Ruth Horn<sup>3,4</sup>

© The Author(s), under exclusive licence to European Society of Human Genetics 2023

*European Journal of Human Genetics* (2023) 31:617–618; <https://doi.org/10.1038/s41431-023-01355-3>

Large-scale studies derived from genomic and health data repositories are at the forefront of personalized medicine. The US NIH Cancer Genome Atlas Project, for example, has generated maps of genomic changes in a large variety of cancer types and cases to help improve diagnosis and treatment based on the large-scale genome sequencing of patients. Another recent example is polygenic risk scoring (PRS), which, despite criticism that it should be carefully evaluated, will be used by the NHS for risk assessment as part of clinical decision-making, including access to screening. PRS, too, has become more widely available due to the development of population-level genetic studies, particularly genome-wide association studies (GWAS). Both cancer genomics and PRS demonstrate the need for and challenges involved in comparable and viable data sharing in population-scale genomic research. Here, we focus on genomic repositories sharing data among themselves, with researchers, and with donors. Despite the rapid expansion of large-scale national genomic repositories and efforts to create diverse datasets, genomic diversity, which is crucial in comparative research, is not easily attainable. Many barriers to sharing genomic data have already been identified, including the interrelated challenges of comparability, confidentiality, and viability.

First, let us consider comparability. While the research community has agreed that many more samples and genomic data are required, the diversity and limited availability of genomic data often create challenges in terms of comparability. As previous studies have shown by comparing biobank data from the UK, Japan, and Taiwan, PRS's prediction accuracy based on UK data was far lower in non-European populations. Indeed, it was 2.5-fold lower in East Asians and 4.9-fold lower in Africans, on average [1]. Similarly, the differences in accessible genomic data between populations have led, for example, to the development of Japanese oncogenetic panels for the Japanese population [2]. Transforming genomic data into scientific insights and knowledge requires appropriate quantifications and annotations from medical and clinical perspectives. Particularly when researchers use decentralized datasets internationally, while the processing of genomic data must be standardized, direct access to raw data is often avoided due to privacy concerns. However, these data repositories are still thought of as local and separate datasets [3]. In this regard, one of the key tasks is harmonizing the interpretation and reporting of disease-associated markers detected during sequencing, for example, cancer germline/

somatic mutations and their various categories of variants [4]. Moreover, there is an on-going struggle to integrate other health data from separate labs or hospitals with genomic data.

Second, to ensure the privacy and confidentiality of participants, data from national healthcare systems are rarely accessible outside of institutional boundaries. Even in digital form, individual genomic data are carefully protected sensitive data that, with effort, can be re-associated with specific individuals. Furthermore, biomedical researchers require not only access to genomic data but also additional medical and lifestyle information from individuals, creating further challenges regarding confidentiality and privacy.

Third, the continued viability and usefulness of repositories require significant resources [5]. While some biobanks and data repositories are solely supported by governmental funding, others depend on partnerships with commercial companies or have a “self-sustaining” business model. Because DNA donation is built on public trust and solidarity, when biobanks are connected with pharmaceutical companies as part of a joint venture, the impact of such public–private sector partnerships on public trust and the solidaristic ethos of “sharing while caring” should become a matter of concern [6].

The above-mentioned challenges involved in sharing genomic data can be demonstrated by looking at the UK's, Japan's, and Israel's biobanks. In these three population-based initiatives, DNA samples were collected from donors, and together, they could lead to increased comparability, data sharing, and viability. The UK Biobank Project sequenced the genomes of over half a million individuals. Designed to be representative of the general population of the UK, 94% of individuals whose data are in the UK biobank are “White” Europeans; in addition, donors, as compared to the general population, are older, better educated, wealthier, and generally healthier [7]. Japan's biobank is not diverse either. Given its home country's ethnic diversity, the Israeli “drop for research” biobank could be relatively more diverse than the other two biobanks, but no data are available regarding its actual ethnic composition. Together, these repositories are diverse, yet we are not familiar with any plans for collaboration among these three biobanks to increase comparability.

While all three biobanks maintain participant privacy and obtain broad consent, they differ in terms of their policies on data sharing with donors. For instance, the Tohoku Medical Megabank (TMM) Project in Japan, with more than 150,000 participant samples, carefully returns individual genomic alongside genetic counseling

<sup>1</sup>Department of Sociology and Anthropology, Ben-Gurion University of the Negev, Beer-Sheba, Israel. <sup>2</sup>Uehiro Research Division for iPS Cell Ethics, Center for iPS Cell Research and Application (CiRA), Kyoto University, Kyoto, Japan. <sup>3</sup>The Ethox Centre and Wellcome Centre for Ethics and Humanities, Nuffield Department of Population Health, University of Oxford, Oxford, UK. <sup>4</sup>Ethics of Medicine, Faculty of Medicine, University of Augsburg, Augsburg, Germany. ✉email: [aviadrz@bgu.ac.il](mailto:aviadrz@bgu.ac.il)

Received: 21 March 2023 Accepted: 22 March 2023

Published online: 31 March 2023

[8]. With 150,000 donor samples, the Israeli biobank's policy is to return only actionable incidental genetic findings to donors [9]. The UK biobank will return only nongenetic findings measured at enrollment [10]. Finally, in terms of viability, while the UK and Japanese biobanks can have relatively secure funding through independent non-commercial funding bodies and governmental funding agencies, the Israeli biobank, which is operated by an HMO (Maccabi), adopts a "self-sustaining" business model that depends on cost recovery through user fees.

The challenges we described are expected to intensify in the near future as biobanks and cancer sequencing are merging. Comparability and data sharing are becoming even more crucial as AI algorithms are increasingly used for GWAS. The challenge posed by the variety of biobanks' consent models may lead to adding dynamic re-consent, in addition to the original broad consent used, perhaps due to future coordination [11]. Federated Health Data Networks (FHDNs) have recently been proposed to facilitate the sharing of sensitive health data across healthcare institutions as well as regional and national borders [12]. In this model, a series of decentralized, interconnected nodes allows data to be queried by other nodes in the network without the data leaving the node it is located at. As opposed to data sharing, transfer, or pooling, FHDNs facilitate data access or data visiting, meaning that queries and algorithms, which are increasingly used to analyze large genomic datasets, can be sent and applied to the pseudonymized data. Genomic repositories must explore these new technological options, along with appropriate public engagement and genomic promotion programs, to increase their comparability, data sharing, and viability.

## REFERENCES

- Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet.* 2019;51:584–91.
- Nagahashi M, Shimada Y, Ichikawa H, Hitoshi K, Kazuaki T, Shujiro O, et al. Next generation sequencing-based gene panel tests for the management of solid tumors. *Cancer Sci.* 2019;110:6–15.
- Carress H, Lawson DJ, Elhaik E. Population genetic considerations for using biobanks as international resources in the pandemic era and beyond. *BMC Genomics.* 2021;22:351.
- Li MM, Datto M, Duncavage EJ, Shashikant K, Neal I, Somak R, et al. Standards and guidelines for the interpretation and reporting of sequence variants in cancer. *J Mol Diagn* 2017;19:4–23.
- Chalmers D, Nicol D, Kaye J, Bell J, Campbell A, Calvin W, et al. Has the biobank bubble burst? Withstanding the challenges for sustainable biobanking in the digital era. *BMC Med Ethics.* 2016;17:39.
- Horn R, Kerasidou A. Sharing whilst caring: solidarity and public trust in a data-driven healthcare system. *BMC Med Ethics.* 2020;21:110.
- Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* 2015;12:e1001779.
- Kawame H, Fukushima A, Fuse N, Nagami F, Suzuki Y, Sakurai-Yageta M, et al. The return of individual genomic results to research participants: design and pilot study of Tohoku Medical Megabank Project. *J Hum Genet.* 2022;67:9–17.
- Raz A, Hashiloni-Dolev Y. "Donating with eyes shut": attitudes regarding DNA donation to a large-scale biobank in Israel. *N Genet Soc.* 2022;41:47–65.
- Wolf SM, Crock BN, Van Ness B, Lawrenz F, Kahn JP, Beskow LM, et al. Managing incidental findings and research results in genomic research involving biobanks and archived data sets. *Genet Med.* 2012;14:361–84.
- Teare HJA, Prictor M, Kaye J. Reflections on dynamic consent in biomedical research: the story so far. *Eur J Hum Genet.* 2021;29:649–56.
- Hallock H, Marshall SE, 't Hoen PAC, Nygård JF, Hoorne B, Fox C, et al. Federated networks for distributed analysis of health data. *Front Public Heal.* 2021;9:712569.

## AUTHOR CONTRIBUTIONS

Original draft preparation: AR; conceptualization: JM and AR; review and editing of the draft: AR, JM, KT, HG, YH, RH; All authors have read and agreed to the published version of the manuscript.

## FUNDING

AR and JM are grateful for the funding provided by the JSPS – ISF Joint Program, grant 62/22, "Biobanks for genomic medicine in Israel and Japan: An analysis of ethics and policy".

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Correspondence** and requests for materials should be addressed to Aviad Raz.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.