


Is blockchain the breakthrough we are looking for to facilitate genomic data sharing? The European Union perspective

Digital Health
Volume 8: 1–3
© The Author(s) 2022
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20552076221114225
journals.sagepub.com/home/dhj


Fidelia Cascini¹ , Flavia Beccia¹, Francesco A Causio¹ , Andrea Gentili¹, Andriy Melnyk¹, Stefania Boccia^{1,2} and Walter Ricciardi^{1,3}

Abstract

The recent progress of genomics research is providing unprecedented insight into human genetic variance, susceptibility to disease and risk stratification. Current trends predict that a massive amount of genomic data will be produced in the upcoming years which, when coupled with the fast-paced development of the field, will create new social, ethical, and legal challenges. In the complex legislative environment of the European Union, genomic data sharing policies will have to weigh the benefits of scientific discovery against the ethical risks posed by the act of sharing sensitive data. In this complex, interconnected environment, blockchain provides a unique and novel solution to accountability, traceability, and transparency issues regarding genomic data sharing. Implementing a distributed ledger technology-based database could empower both patients and citizens to responsibly use genomic data pertaining to them because it allows for a higher degree of control over the recipients of their data and their uses. The blockchain technology will engage both data owners and policymakers to address the multiple issues of genomic data sharing and allow us to redefine the way we look at genomics.

Keywords

Genomics, blockchain, data, tokenization, distributed ledger, European Union

Submission date: 30 May 2022; Acceptance date: 30 June 2022

Main Piece

Genomics enables researchers to use powerful computational and statistical methods to decode the functional information hidden in DNA sequences.¹ Progress from the analysis of whole-genome sequencing data provided unprecedented insights into human genetic variance, disease, evolution, and migration patterns.² Estimates based on the current growth trends predict that genetics research will generate between 2 and 40 exabytes (million terabytes) of data within the next decade.³

This amount of data has the potential to transform health care, revolutionize the understanding of human health and disease, and lead to a personalized approach to prevention and care. Big data allows for an in-depth assessment of health outcomes based on patient characteristics; genomic biomarkers are increasingly used to stratify disease outcomes and susceptibility into subgroups that reflect underlying heterogeneity and potential response to

interventions.⁴ For example, one study compared the clinical radiologists' original decisions with a big data system's cancer predictions based on a biopsy that confirmed breast cancer. The big data approach results had 5.7% fewer false positives and 9.4% fewer false negatives than radiologists' results.⁵ Another example regards the screening for diabetic retinopathy. A deep learning system, fuelled by genetic data, demonstrated enhanced health care decision-making with automated image analysis

¹Section of Hygiene and Public Health, Department of Life Sciences and Public Health, Università Cattolica del Sacro Cuore, Roma, Italy

²Department of Woman and Child Health and Public Health-Public Health Area, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy

³Istituti Clinici Scientifici Maugeri, Pavia, Italy

Corresponding author:

Francesco A Causio, Section of Hygiene and Public Health, Università Cattolica del Sacro Cuore, 00168 Roma, Italy.

Email: francescoandrea.causio01@icatt.it



and accelerated the process of manual evaluation of image data to screen for eye disease in diabetics.⁶ Genomics can improve individual and population health outcomes, but making genomic data widely available creates ethical, legal and social challenges.

In February 2021, the EU kicked off the Joint Action “Towards a European Health Data Space” (TEHDAS), which will further develop the principles regarding secondary use of data (including genomic), promote health data sharing between member states, support scientific research and foster digital health services in which clear liability rules are implemented and enforced. Additionally, many national governments and supranational organizations are collaborating to make genomic data sharing easier and more feasible. In the EU, the “1+ Million Genomes” (1 + MG) initiative brings together 22 EU countries, plus the United Kingdom and Norway, to ensure secure, federated access to genomic data, while also keeping it legally and ethically safe.

However, the EU has a complex legislative environment regarding data protection. Article Eight of the EU Charter of fundamental rights stipulates the right of EU citizens to the protection of personal data, fair processing for specified purposes, and the necessity of consent.⁷ Every individual has the right to access and rectify personal data concerning them that has been collected. The 2016 General Data Protection Regulation (GDPR) further strengthens individuals’ fundamental rights in the digital age. Under Article Four of the GDPR, genetic data is included in the catalogue of special categories of data and requires explicit consent for its use and processing.⁸ Thus, in the EU, the discussion of genomic data sharing is still open and emerging technologies, such as blockchain, will play a key role in addressing priority issues.

Data sharing policies must weigh the scientific discovery benefits against the ethical risks posed to the people contributing their data and delicately balance the appropriate level of shareability (for the benefit of the community) and individual protection. Although some consent procedures for genomic data sharing already exist, full genomic data ownership is a central focus because it allows individuals to personalize the degree to which their data is shared and the recipients of such data.⁹

Even though publicly shared data are fully anonymized, they may still be sensitive to breaches via different routes, such as cross-referencing different sources or identity tracking. A blockchain is a database containing information on transactions that occurred within a given network, to be simultaneously used and shared within a large decentralized, publicly accessible network.¹⁰ Its name is related to the blocks of transaction data being chained together by incorporating a digital fingerprint of each block in the following one. Consequently, each block is tightly linked to the previous and the following one, making it impossible to alter one without affecting the others as well.¹¹ Once the block is filled with data, it is timestamped and

chained onto the previous block, hence making the data chained together in chronological order.

The distributed ledger technology is the technological infrastructure and protocol that allows simultaneous access, validation, and record of blockchain systems plus updates them in an immutable manner across a network that is spread across multiple entities or locations.¹² Distributed networks eliminate the need for a central authority to keep a check against manipulation, while also guaranteeing that information is stored in a secure and accurate manner using cryptography. In addition, once it is stored, the information cannot be edited or modified, and it becomes governed by the rules of the network. The very nature of such a decentralized ledger makes it immune to cyberattacks, as hackers would need to attack all the copies stored across the network at the same time. This also avoids the centralization of knowledge, as the individual blocks do not have access to information stored in other blocks. In centralized forms of data storage, instead, companies or entities in charge of the database have full control over the data, hence raising doubts about data security and reliability. This kind of network has unlimited uses: the most famous and controversial one is Bitcoin.¹³ But in the field of data sharing, blockchain opens a world of possibilities, especially when the type of data being handled is particularly sensitive, as in the case of genomic data.

A successful case study is represented by the Estonian application of blockchain technology to health care. Starting in 2016, the Estonian government applied blockchain to store medical records with immediate, positive results: it eased the sharing of data among authorized parties and bridged traditional data silos, it decreased medical care costs through better insurance claims and, finally, improved data mining through the immutable records maintained by the blockchain.¹⁴ Similarly, a prospective cohort study investigated the use of a blockchain-authenticated platform for genomic data sharing, together with de-identified electronic health records and imaging data, yielding encouraging results. The investigated blockchain-based pilot infrastructure was not significantly different in the completeness of available data when compared to the gold standard data-sharing platforms but it was more robust when recording certain data elements in the registry format, making them more accurate for analyses that aim to further personalized medicine.¹⁵

People’s motivation to share their own genomic data can range depending on several variables, including people’s familiarity with the concepts of DNA and their trust in the recipient.¹⁶ Ensuring data owners feel empowered about the choice of who gets to use their data and for what purposes is a key issue to address in order to increase data sharing for secondary use and promote the use of genomic big data for disease outcome stratification and accurate prediction of response to pharmaceutical or clinical interventions.¹⁷ By using a blockchain-based sharing

platform, all information regarding data access and management is transparent and allows the user to be aware of all stakeholders using the data at any time, for any purpose. The blockchain's innate accountability and traceability of transactions are helpful in clarifying how, when, and how many times a user's genomic data is transferred.

In addition, using blockchain-based services allows tokenization of genomic data, allowing users to exert a form of control over sharing their own data. Young businesses, such as Genecoin, built a marketplace-like system where consumers can trade their private genomic data for non-monetary tokens that users can later exchange for goods and services. This resulted in increased user satisfaction, increased interest in how data was used and users being more prone to sharing their data when asked by different entities from the one storing the data.¹⁸

Although blockchain is still a relatively new technology, it could be a real game-changer in the field of genomics and beyond. Blockchain gives stakeholders novel tools to simultaneously protect sensible data and empower data owners to exert full control over their use, as it allows a case-by-case consent evaluation for each data access query. Progress in the field of personalized medicine will urge policymakers to address the multiple issues of data sharing and cement blockchain as a promising technology to redefine the way we look at genomics.

Author contributions: FC conceptualized and wrote (review and editing) the paper. FB wrote (original draft) the paper. FAC searched the literature, conceptualized and wrote (original draft) the paper. AG conceptualized and wrote (review and editing) the paper. AM searched the literature, conceptualized and wrote (original draft) the paper. SB revised the paper critically for important intellectual content. WR conceptualized and wrote (review and editing) the paper.

Conflicts of interest: The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding: The authors received no financial support for the research, authorship and/or publication of this article.

Guarantor: FC

ORCID iDs: Fidelity Cascini  <https://orcid.org/0000-0001-6499-0734>

Francesco A Causio  <https://orcid.org/0000-0002-6545-3196>

References

1. Taher L, Narlikar L and Ovcharenko I. Identification and computational analysis of gene regulatory elements. *Cold Spring Harb Protoc* 2015; 2015: pdb.top083642.
2. Wu D, Dou J, Chai X, et al. Large-Scale whole-genome sequencing of three diverse Asian populations in Singapore. *Cell* 2019; 179: 736–749.e15.
3. Stephens ZD, Lee SY, Faghri F, et al. Big data: astronomical or genomics? *PLoS Biol* 2015; 13: e1002195.
4. Houry MJ, Armstrong GL, Bunnell RE, et al. The intersection of genomics and big data with public health: opportunities for precision public health. *PLoS Med* 2020; 17: e1003373.
5. McKinney SM, Sieniek M, Godbole V, et al. International evaluation of an AI system for breast cancer screening. *Nature* 2020; 577: 89–94.
6. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal Fundus photographs. *JAMA* 2016; 316: 2402–2410.
7. Charter of Fundamental Rights of the European Union. *Official Journal of the European Communities*. Published online 2016.
8. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance) - Publications Office of the EU. Accessed May 16, 2022. <https://op.europa.eu/en/publication-detail/-/publication/3e485e15-11bd-11e6-ba9a-01aa75ed71a1>
9. O'Neill O. Autonomy and Trust in Bioethics. *Autonomy and Trust in Bioethics*. Published online April 18, 2002. <https://doi.org/10.1017/CBO9780511606250>
10. Gordon WJ and Catalini C. Blockchain technology for healthcare: facilitating the transition to patient-driven interoperability. *Comput Struct Biotechnol J* 2018; 16: 224–230.
11. Catalini C and Gans JS. Some simple economics of the blockchain. *SSRN Electron J*. Published online April 20, 2019. doi:10.2139/SSRN.2874598
12. Ribitzky R, St Clair J, Houlding DI, et al. Pragmatic, interdisciplinary perspectives on blockchain and distributed ledger technology: paving the future for healthcare. *Blockchain Healthc Today* 2018; 1: 2–3. doi:10.30953/bhty.v1.24.
13. Nakamoto S. Bitcoin: A Peer-to-Peer Electronic Cash System. Accessed May 16, 2022. www.bitcoin.org
14. Heston TF. A Case Study in Blockchain Healthcare Innovation. Published online 2017. Accessed May 16, 2022. <https://www.pubpub.org/pub/medrec>.
15. Glicksberg BS, Burns S, Currie R, et al. Blockchain-Authenticated Sharing of Genomic and Clinical Outcomes Data of Patients With Cancer: A Prospective Cohort Study. *J Med Internet Res* 2020; 22: e16810. doi: 10.2196/16810.
16. Middleton A, Milne R, Almarri MA, 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: 743–752.
17. Clayton EW, Evans BJ, Hazel JW, et al. The law of genetic privacy: applications, implications, and limitations. *J Law Biosci* 2019; 6: 1–36.
18. Thiebes S, Schlesner M, Brors B, et al. Distributed ledger technology in genomics: a call for Europe. *Eur J Hum Genet* 2019; 28: 139–140.