GMR
The Original

Ethical Challenges in Genomic Data Sharing and Patient Privacy in Precision Medicine

Dr. Daggupati Harith, Dr. Rajesh Kumar Bhola, Dr Sneh Hemantbhai Dudhia, Xu Xiang, Anubhav Bhalla, Thiyagarajan Sanjeevi, Uma Bhardwaj,

Assistant Professor, Department of Anaesthesiology and Critical Care Medicine, Mahatma Gandhi Medical College and Research Institute, Sri Balaji Vidyapeeth University, Puducherry, India, Email id- daggu.harith@gmail.com. Orchid id:0009-0003-6754-7748
Professor, Department of Pathology, IMS and SUM Hospital, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, Odisha, India, Email Id- rajeshkumarbhola@soa.ac.in, Orcid Id- 0000-0002-1601-7961
Professor, Department of Pharmacology, Parul Institute of Medical Sciences & Research, Parul University, Vadodara, Gujarat, India, Email Id- sneh.dudhia34654@paruluniversity.ac.in, Orcid Id- 0009-0004-8521-2439
Research Scholar, School of Business Management, Lincoln University College, Malaysia. Email ID: xiang.phdscholar@lincoln.edu.my
Centre of Research Impact and Outcome, Chitkara University, Rajpura- 140417, Punjab, India. anubhav.bhalla.orp@chitkara.edu.in https://orcid.org/0009-0005-7854-6075
Department of Medical Biotechnology, Aarupadai Veedu Medical College and Hospital (AVMC&H), Vinayaka Mission's Research Foundation (Deemed to be University), India thiyagarajan.sanjeevi@avmc.edu.in https://orcid.org/0000-0002-4006-7184
Professor, Department of Biotechnology and Microbiology, Noida International University, Uttar Pradesh, India. vc@niu.edu.in . 0000-0002-6414-9731

ABSTRACT

With precision medicine, the medical field has changed because medical interventions are personalized based on the specific genomic profile. The implementation of this paradigm shift depends on the huge aggregation and distribution of genomic data on worldwide research networks to gain statistical significance and clinical validity. Nevertheless, this model of open science poses a considerable amount of ethical tension in the case of patient privacy due to its collaborative characteristic. The main ethical issues related to genomic data sharing discussed in this paper are the following: First, the technical aspect of the matter that the genetic data is identifiable by nature, and hence the traditional methods of anonymization may not be enough to withstand the current re-identification attacks. Moreover, the drawbacks of the classical informed consent are also examined, with special emphasis on the issue of the tension between the broad consent toward further research and the autonomy of the participant. It is further discussed in terms of issues of data sovereignty and the possibility of worsening the health inequalities of underrepresented populations. This paper proposes a shift in approaches to data governance by assessing existing legal regulations, including GDPR and GINA, as well as considering recent technological privacy protection systems, including differential privacy and federated learning. Finally, it is noted that an effective, clear, and ethically-founded sharing of most protocols is a requirement in terms of keeping the population trustful and providing the sustainable development of precision medicine.

Keywords: *Precision Medicine, Genomic Privacy, Informed Consent, Data Sovereignty, Bioethics, Re-identification.*

INTRODUCTION

The introduction of high-throughput sequencing technologies into clinical practice has introduced the age of precision medicine, which is a paradigm shift that has replaced the concept of a "one-size-fits-all"

approach to treatment with interventions that are designed based on the distinctive genetic, environmental, and lifestyle characteristics of individual patients [7]. The essence of this revolution is the need to have large sets of data. To be able to differentiate between pathogenic and benign variations, clinicians and researchers need to compare the genomes of a particular individual with the enormous and varied populations. As a result, the exchange of genomic information between different countries and organizations has become the standard of present-day biomedical studies and pharmaceutical development [1][5].

Nonetheless, the fast pace of increasing data exchange has exceeded the creation of stronger ethical and legal protection, which has created a deep contradiction between scientific advancement and patient confidentiality [2][3]. In contrast to the traditional medical records, the genomic data is unique and unalterable by its nature; it is a biological identifier that is permanent, which discloses not only the disposition of a particular person but also that of their biological relatives [9][10]. The more the potential to cross-reference the de-identified research data with the public records and even the genealogy databases, the closer the promise of anonymity is regarded as a technical impossibility. The combination of big data and biology poses serious ethical concerns [4][6][8]. What can the medical community do to keep the principle of autonomy, as the future application of the donated data is unknown during the collection? How far does the quest for the common good by open science justify the risks of genetic discrimination that may arise or surveillance? Moreover, with the spread of the concept of precision medicine worldwide, the issue of data ownership and fairness of benefits distribution also emerges, especially concerning the marginalized population that has traditionally been either marginalized or victimized by the genome studies.

These are complex ethical issues that are discussed in this paper. It initially looks into the technical weaknesses of genomic privacy and the drawbacks of existing anonymization laws. It then analyses the changing environment of informed consent, the loopholes in the law regarding current privacy provisions, and the socioeconomic consequences of ownership of data. Lastly, the discussion suggests one way in which data stewardship could be structured, and that the future success of precision medicine requires a framework of governance in which transparency and trust in the community are equally important to scientific benefit.

The Privacy Landscape in Genomics

The fundamental problem of genomic data sharing is that the traditional distinction between the information that can be identified and the information that cannot is becoming blurred. The privacy in the majority of biomedical studies is preserved by excluding personal identifiers like names, addresses, social security numbers, etc. Nevertheless, these standard methods of de-identifying data are not effective in the case of the human genome, since this is a privacy paradox in which the outcomes of the research cannot be utilized without sharing the data, but sharing the data inevitably compromises the anonymity of the donor.

The Myth of Anonymization

Genomic information has been referred to as the final biological barcode. Since an individual's genetic make-up is a unique sequence (except with identical twins, the same sequence applies), it forms an irreversible and unalterable marker. It has been shown that even chopped-off or anonymized genomic information can be re-identified with the help of cross-referencing it with outside data sources. Another myth of anonymization is also broken by the fact that the information of the genome does not belong to the individual person; it is distributed among biological family members as well. Therefore, the act of an individual providing his/her data may unwillingly reveal the genetic predispositions and privacy of the unwilling family members.

Re-identification Risks and Linkage Attacks

The most common technical threat to the privacy of genomes is the linkage attack. This happens when an opponent takes a de-identified research dataset and links it with a publicly available dataset - e.g., a voter registration list or a consumer genealogy database (e.g., GEDmatch). Persons can be statistically identified with high probability by matching specific genetic markers or metadata.

The Genetic Picket Fence: Although an insignificant proportion of the population may post their DNA in open databases, most of the population can now be accessed via the so-called long-range familial searching.

Metadata Sensitivity: Privacy invasions do not usually happen due to the genetic sequence, only phenotypic data (age, zip code, rare disease status) needed to conduct precision medicine studies, but that forms a secondary inviolability point of identification.

Cybersecurity and Data Vulnerability

In addition to the theoretical dangers of re-identification, the centralized storage of genomic information in biobanks and cloud repositories makes it very valuable to attack. A stolen genome is a permanent violation, unlike a stolen credit card number, which may be canceled and a new one issued. Genetic stalking is a possibility, or there is the possibility of the sale of genetic profiles on the dark web, which is a new form of digital harm. The infrastructure to store this data should move beyond mere encryption as precision medicine becomes bigger, stronger, and more resilient structures that will resist external hacks as well as internal misuse of data.

Informed Consent and Autonomy

The conventional approach to informed consent, in which a participant signs an agreement to a particular study, with a one-time signature, is also becoming unsuitable, in the context of genomic research, to the needs of precision medicine. Data science is an iterative process such that a genome sequenced and stored can be used in hundreds of studies in several decades. This poses a basic contradiction between the principle of autonomy and the practical requirements of longitudinal research.

The Dilemma of Broad Consent

In order to overcome the logistical challenge of re-contacting participants each time a new study is done, most biobanks use broad consent. In this model, the participants will accept to be used in future unspecified research.

- Ethical Critique: Since critics claim that broad consent is not a real consent as the participant is unable to know the risk or nature of the research that is yet not designed.
- Commercial Interests: The participants should feel at ease with the idea that their information is sold to a pharmaceutical company to make a profit enterprise without their direct awareness.

Dynamic Consent: A Digital Solution

In an attempt to solve the flaws of broad consent, the dynamic consent concept has come up. The strategy makes use of digital platforms to have a continuous relationship between the researcher and the participant.

- Granular Control: The participants will have a secure application that will enable them to control their sharing choices such as permitting their data to be used in cancer research but not in behavioral genetics research.
- Real-Time Interaction: The model will convert the participant into an active partner and not a passive provider of data, as well as enable the withdrawal of consent or the revision of the personal information once the research environment changes.

Incidental Findings and the "Right Not to Know"

Genomic sequencing tends to have incidental discoveries made- some insight into the well-being of a patient that was not the intention of the test (e.g. finding out that a patient is at risk of Alzheimer when having his heart checked).

The Duty to Warn: Some ethical issues have been raised on whether a researcher has a moral duty to disclose such findings to the research participants, particularly when the condition is preventable or treatable

Individual Choice: Respectful autonomy involves the right of a participant to not know some of his or her genetic future. It is one of the most delicate issues in genomic ethics to strike a balance between the right and the clinical good of early intervention.

Equity, Inclusion, and Data Sovereignty

Although the sharing of genomic data is expected to serve humanity, the principle is commonly undermined through the imbalanced systems of collecting and controlling data. Precision medicine is a danger of a luxury of the global North that could create health disparities if the information that is laid on its foundation is not something that represents all human diversity.

The Diversity Gap and Clinical Bias

Much of the current genomic database is based on populations of European origin, which makes them biased in terms of ancestry, having a direct ethical impact on clinical care. The resultant lack of diversity means that there are high risks of misdiagnosis with genetic variants deemed benign in one group being misdiagnosed as pathogenic in a situation where there is a dearth of reference databases. As an illustration, studies have shown that people of African or Asian origin have a higher likelihood of a false positive diagnosis of some cardiac diseases because their respective genetic backgrounds are not recorded in the clinical data. This creates inequity, as any treatment worked out on a narrow genetic profile can be less effective or even harmful to underrepresented populations, which further entrenches any underlying health disparities.

Indigenous Data Sovereignty and Self-Determination

To most of the Indigenous communities, genetic information is not just biological but is a sacred cultural resource related to ancestry and land. Genomics history is characterised by biocolonialism, in which the samples were stolen without proper consent or subsequent benefit, as in the case of the Havasupai Tribe. Indigenous Data Sovereignty holds that such communities should be allowed to control the collection, ownership and use of data of their individuals. It has resulted in the creation of certain models of governance, including CARE Principles (Collective Benefit, Authority to Control, Responsibility and Ethics). These principles are an ethical counter-proportion that is required by the open-science principles of FAIR, so that the accessibility of data does not compromise community self-determination.

Benefit-Sharing and the Challenge of Exploitation

The continued commercialization of genomic findings and the ensuing exploitation is a constant ethical issue. As a low-income or marginalised population is harvested on data, both the expensive therapies created out of that data are not affordable to those communities that constitute the same communities that also form the source of the information. International initiatives such as the Nagoya Protocol aim at mitigating the benefit of fair and equitable distribution in case of genetic resource utilization. Nevertheless, genuine equity in precision medicine is only possible in a radical way of transformation of extractive-based research models to collaborative relationships of ethical reciprocity. In these models, the data donors will receive tangible medical and economic advantages, and this will ensure that the byproducts of genomic research are shared fairly and equally among the entire global population.

Legal and Policy Frameworks

The management of genomic data is regarded by the disjointed world environment in which legal settings usually cannot adapt to the swift biotechnological improvements. Nowadays, the most noticeable regulatory

pillars are the General Data Protection Regulation (GDPR) in Europe and the Health Insurance Portability and Accountability Act (HIPAA) and the Genetic Information Nondiscrimination Act (GINA) in the United States. Although the GDPR offers a very strong structure by categorizing genetic data as one of the special categories of sensitive data that the highest levels of protection are required, it contains some exceptions in terms of scientific research that may result in the fact that one jurisdiction will have better protection than another. On the other hand, GINA provides essential safeguards against genetic discrimination by both health insurers and employers in the United States, but with enormous gaps that leave life insurance, disability insurance, and long-term care coverage largely unaddressed, people are likely to avoid the idea of engaging in genomic research due to the risk of having their finances negatively impacted in the future. In addition to these set laws, there is an increasing realization that the current policies are not enough to accommodate the specifics of digital biology. The majority of the existing frameworks are based on the notion of the so-called de-identification being a legal safe harbor, yet, as it was shown in previous paragraphs, the technicalities of genomic uniqueness render such legal definitions less and less relevant. Moreover, due to the nature of precision medicine that is global, there is almost always a certain flow of data across countries, and in most cases, it finds itself in a jurisdiction where the privacy legislation is much weaker. Such regulatory arbitrage is dangerous to the participants who could agree to participate in research under the rules, only to have their data treated under different rules. To overcome these weaknesses an urgent international harmonization of the genomic laws is needed that are not based on the compliance framework but one that is entrusted to active, transparent oversight and accountability systems that will safeguard the individual irrespective of the locality in which the data is stored.

Proposed Mitigations and Future Directions

The solution to the ethical conflicts involving the issue of genomic data sharing involves a shift towards the administrative protection mechanisms to a more resilient and privacy-friendly technological and governance framework. The scientific community is also considering the use of novel approaches that would enable the derivation of clinical information without necessarily revealing raw, identifiable genetic sequences, as the constraints of the traditional anonymization process become more apparent.

Privacy-Preserving Technologies

The rise of federated learning and differential privacy is a massive move in the way genomic research can be performed. Federated learning enables the researcher to use all decentralized servers to train analytical models; the data is not exchanged and is only safeguarded by the security measures of the institution where it was created, with the model parameters (the learned patterns) only being exchanged. This method of taking the code to the data helps in eradicating the risks that are involved in massive data transfers. In addition to this, differential privacy injects mathematical noise into datasets, so that it could be guaranteed that the overall statistical results are going to be correct, but the contribution of an individual cannot be detected. These instruments offer a technical firewall that builds the usefulness of big data with the requirement of participant anonymity.

Blockchain for Secure Data Stewardship

The blockchain technology provides a good platform to handle the problems of transparency and consent. With the help of a decentralized and immutable registry, genomic repositories can give the participants a clear house of access and usage of their data by whom. Smart contracts will be able to implement the consent preferences and make sure that only the researchers who fulfill certain ethical standards pre-set by the donor will be allowed to release data. This is a paradigm shift where the ownership of data is often hard to legally enforce, to data stewardship, where the responsibility of the information that is well managed throughout its lifecycle is of primary concern and the management is responsible and can be traced.

Cultivating Public Trust Through Transparency

Finally, the necessity of the public trust cannot be substituted by the technical one completely. The use of open communication and active participation of participants in the research process should be considered the major directions of precision medicine in the future. Creating the research models of participant-led and appropriate protocols of the clinical results return can turn the donors of the data into participants of the process. The sphere of precision medicine has the chance to guarantee the sustainability and acceptance of the technology in the long run because of the creation of an environment in which transparency is the norm, and ethical protection is embedded into the technological framework.

CONCLUSION

The success of precision medicine will depend on the contradiction of sharing such fundamentally personal genomic information to promote worldwide science. Even though this combination of large datasets is crucial to clinical accuracy and personalized treatment, it brings a lot of ethical tension in terms of privacy loss of patients, inefficiency of classical informed consent, and the potential of inadvertently disadvantaging disadvantaged groups. To address these tensions, the medical community needs to cease to extract data models toward a structure of ethical stewardship. With the combination of privacy-saving technologies such as federated learning with the principles of inclusivity and a strong legal framework, genomic research can be promoted and, at the same time, individual autonomy and the trust of the population can be ensured.

REFERENCE

[1] Frizzo-Barker, J., Chow-White, P. A., Charters, A., & Ha, D. (2016). Genomic big data and privacy: challenges and opportunities for precision medicine. *Computer Supported Cooperative Work (CSCW)*, *25*(2), 115-136.

[2] Erdmann, A., Rehmann-Sutter, C., & Bozzaro, C. (2021). Patients' and professionals' views related to ethical issues in precision medicine: a mixed research synthesis. *BMC medical ethics*, *22*(1), 116.

[3] Ahmed, L., Constantinidou, A., & Chatzittofis, A. (2023). Patients' perspectives related to ethical issues and risks in precision medicine: a systematic review. *Frontiers in Medicine*, *10*, 1215663.

[4] Winkler, E. C., & Knoppers, B. M. (2022, September). Ethical challenges of precision cancer medicine. In *Seminars in cancer biology* (Vol. 84, pp. 263-270). Academic Press.

[5] Azencott, C. A. (2018). Machine learning and genomics: precision medicine versus patient privacy. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *376*(2128), 20170350.

[6] Alvarez, M. J. R., Griessler, E., & Starkbaum, J. (2022). Ethical, legal and social aspects of precision medicine. In *Precision Medicine in Clinical Practice* (pp. 179-196). Singapore: Springer Nature Singapore.

[7] Ormond, K. E., & Cho, M. K. (2014). Translating personalized medicine using new genetic technologies in clinical practice: the ethical issues. *Personalized medicine*, *11*(2), 211-222.

[8] Schaefer, G. O., Tai, E. S., & Sun, S. (2019). Precision medicine and big data: The application of an ethics framework for big data in health and research. *Asian Bioethics Review*, *11*(3), 275-288.

[9] Sadulla, S. (2025). Hybrid-HAVA: Enhancing AI Security in Healthcare with GAN-Based Adversarial Filtering. *Frontiers in Computational Science and Engineering*, 35-43.

[10] de Groot, N. F. (2024). A contextual integrity approach to genomic information: what bioethics can learn from big data ethics. *Medicine, Health Care and Philosophy*, *27*(3), 367-379.

[11] Tolwinski, K., Madden, D., & Wagner, J. K. (2022). A genetic imaginary for social justice: The rhetoric of genomic medicine as exceptional, ordinary, and ethical obligation. *SSM-Qualitative Research in Health*, *2*, 100120.

[12]