Unpacking NIST's Guidance On Genomic Data Cybersecurity

By **Christine Moundas, David Peloquin and Elana Bengualid** (January 9, 2024)

Christine Moundas

David Peloquin

Elana Bengualid

On Dec. 20, 2023, the National Institute of Standards and Technology National Cybersecurity Center of Excellence published its final internal report on cybersecurity of genomic data.[1]

The report notes that current risk management guidance does not adequately capture the unique cybersecurity and privacy concerns regarding the use of genomic data, particularly with respect to balancing access restrictions with the need to share such data.

Accordingly, the report highlights the specific privacy and cybersecurity concerns associated with the use of genomic data and, based on input from industry genomic stakeholders, government and academia, identifies significant gaps in current policy, regulations, legislation and guidance, as well as technology, for protecting genomic data.

The report concludes by proposing potential solutions to identified gaps and areas for further research. In this way, the report aims to assist organizations in protecting against misuse of genomic data and enabling secure collaborative innovations.

Increased Need for Risk Management Guidance

Genomic data — composed of information on deoxyribonucleic acid sequences, variants and gene activity — is heavily relied upon by researchers, government and private industry to evaluate how differences in DNA sequences affect health.

The field of genomic data science has grown rapidly, resulting in the increased generation and sharing of genomic data for research, often through big data collaborations that involve researchers from multiple institutions and countries. According to the NIH National Human Genome Research Institute, approximately 2 billion to 40 billion gigabytes of genomic data are generated each year from millions of people globally.[2]

In turn, as reflected in a 2022 executive order on advancing biotechnology and biomanufacturing innovation,[3] there has been emerging awareness of certain risks to the economy, biotechnology industry and individuals, as well as U.S. national security, resulting from privacy or cybersecurity incidents targeting genomic data.

Specifically, the report notes certain privacy risks for individuals inherent in the use of genomic data, including, "enabling intimidation for financial gain, discrimination based on disease risk, revelation of hidden consanguinity or phenotypes including health, emotional stability, mental capacity, appearance, and physical abilities."

In addition, the report explains that using a patient's genomic data for healthcare purposes may implicate certain concerns including "portability, chain-of-custody, reinterpretation of genomic data, and consent management" as well as harm resulting from theft or sabotage

of analytical processes or systems that govern the creation of precision medicine.

The report observes that current privacy and cybersecurity risk management guidance does not address the risks inherent in the use of genomic data.

First, much genomic data is created or maintained by entities such as research institutes that are not healthcare providers or life sciences companies that are neither covered entities nor business associates under the Health Insurance Portability and Accountability Act of 1996 and its implementing regulations, as amended, meaning that such data falls outside of HIPAA's robust privacy and security requirements.

Moreover, the federal regulations on research involving human subjects, known as the Common Rule, as well as state laws governing the collection, use and disclosure of genomic data are largely focused on privacy, and not security, requirements.

Specifically, the report describes the following significant gaps in current guidance, as identified by bioeconomy stakeholders during various 2022 workshops hosted by the National Cybersecurity Center of Excellence and through subsequent research:

- Practices across the lifecycle concerning genomic data generation;
- Safe and responsible sharing of genomic data;
- Monitoring the systems processing genomic data;
- Lack of specific guidance documents addressing the unique needs of genomic data processors; and
- Regulatory and policy gaps with respect to national security and privacy threats in the collection, storage, sharing and aggregation of human genomic data.

Potential Solutions to Gaps and Areas for Future Research

Accordingly, to bolster the privacy and security of genomic data, the report proposes the following:

- Existing guidance, such as the NIST Risk Management Framework,[4] Cybersecurity Framework[5] and Privacy Framework,[6] must be tailored to include specific and appropriate protections for genomic data;
- The NIST Privacy Framework Profile for Genomic Data, which is scheduled to be published in 2024, could clarify how to manage privacy risks associated with the aggregation, storage and processing of genomic data;
- The manufacturer usage description specification could improve sequencer security and reduce the likelihood of ransomware attacks as well as intellectual property or privacy loss from data exfiltration;

- Demonstration projects should be created to illustrate how to leverage secure cloud-based solutions to protect genomic data, as per the NIST Risk Management Framework, and how the use of federated homomorphic encryption could reduce the risk of loss of confidentiality or integrity caused by sharing genomic data; and

- Security guidelines or benchmarks for genomic sequencers could provide best cybersecurity practices, including with respect to improving supply chain security and cyber resiliency against future threats.

Lastly, the report identifies the following areas of future research:

- Methods for securely integrating genomic data with a patient's electronic health record while maintaining patient privacy and allowing for interoperability;

- Improving the precision of vulnerability scanners for software containers; and

- Technical solutions to solve the containment problem in genomic data for analysis methods not currently addressed by federated multiparty homomorphic encryption.

Identifiability of Genomic Data

One area in which the report may cause some confusion for organizations subject to HIPAA as a covered entity or business associate is with respect to its discussion of deidentification of genomic data.

The report notes that it is inappropriate to deidentify genomic data under HIPAA's safe harbor method, suggesting that such data is inherently identifiable.[7]

While this remains a debated issue, in other contexts, the U.S. federal government has taken the position that genomic data are not inherently identifiable. For instance, guidance issued by the National Institutes of Health strongly suggests that genomic data may be deidentified pursuant to HIPAA's safe harbor method. [8]

Additionally, in the preamble to its 2015 revisions to the Common Rule, the federal agencies that have adopted the Common Rule acknowledged that genomic data is not "identifiable private information" within the meaning of the Common Rule unless it is accompanied by additional personal data that would readily ascertain the identity of the individual.[9]

Revisions to the Common Rule that were issued in 2017 called for the Common Rule agencies to issue further guidance on the identifiability of data generated by particular types of technologies, such as whole genome sequencing, by 2020.[10]

To date, such guidance has not been issued. As such, the report is not authoritative with respect to its assessment of the treatment of genomic data under the current U.S. regulatory framework, specifically with respect to the identifiability of such information.

Looking Ahead

Given the strong recommendations set forth in the report, stakeholders in this space should stay abreast of potential developments regarding the privacy and cybersecurity measures necessary to safeguard genomic data.

In particular, because many entities that conduct research using genomic data or that utilize such data for certain secondary purposes are not regulated by HIPAA, and because state laws are quite varied and typically focus on the privacy and not security of genomic data, to the extent the report's suggestions are implemented, it will provide a helpful rubric for the industry.

Christine Moundas and David Peloquin are partners, and Elana Bengualid is an associate, at Ropes & Gray LLP.

[1] Nat'l Inst. of Standards and Tech., NIST IR 8432: Cybersecurity of Genomic Data (Dec. 2023), available at https://nvlpubs.nist.gov/nistpubs/ir/2023/NIST.IR.8432.pdf.

[2] Nat'l Human Genome Research Ins., Fact Sheet: Genomic Dat Science, https://www.genome.gov/about-genomics/fact-sheets/Genomic-Data-Science#:~:text=As%20biomedical%20research%20projects%20and,data%20now%20generated%20each%20year (last visited Jan. 3, 2024).

[3] Executive Order on Advancing Biotechnology and Biomanufacturing Innovation for Sustainable, Safe, and Secure American Bioeconomy (Sept. 12, 2022), https://www.whitehouse.gov/briefing-room/presidential-actions/2022/09/12/executive-order-on-advancing-biotechnology-and-biomanufacturing-innovation-for-a-sustainable-safe-and-secure-american-bioeconomy/.

[4] Nat'l Inst. of Standards and Tech., NIST Special Publication 800-37: Risk Management Framework for Information Systems and Organizations (Dec. 18), available at https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-37r2.pdf.

[5] Nat'l Inst. of Standards and Tech., Public Draft: The NIST Cybersecurity Framework 2.0 (Aug. 8, 2023), https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.29.ipd.pdf.

[6] Nat'l Inst. of Standards and Tech., NIST Privacy Framework: A Tool for Improving Privacy Through Enterprise Risk Management, Version 1.0 (Jan. 16, 2020), https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.01162020.pdf.

[7] See Report at 19, 22; 45 C.F.R. § 164.514(b)(2).

[8] The NIH Genomic Data Sharing Policy requires that any large-scale genomic data generated in NIH-funded research and submitted to public repositories be de-identified pursuant to the HIPAA Safe Harbor Method. Nat'l Insts .of Health, NIH Genomic Data Sharing Policy (Aug. 27, 2014), https://grants.nih.gov/grants/guide/notice-files/not-od-14-

124.html. Additionally, the NIH webpage on privacy in genomics states that "NIH houses several databases where researchers can share de-identified genomic data" and explains that "[i]n 2013, as required by the passage of the Genetic Information Nondiscrimination Act, the Privacy Rule was modified to establish that genetic information is considered PHI" and is thus protected by the Privacy Rule to the extent that such information is individually identifiable, as "there are no [HIPAA] restrictions on the use or disclosure of PHI that has been de-identified. Privacy in Genomics, Nat'l Insts. of Health, https://www.genome.gov/about-genomics/policy-issues/Privacy (last visited Nov. 6, 2021).

[9] 80 Fed. Reg. 53,933, 53,943, (Sept. 8, 2015).

[10] 45 C.F.R. § 46.102(e)(7)(ii).