

Social, Technical, and Ethical Challenges in Biomedical Data Privacy

Gamze Gürsoy

Yale University, New Haven, CT, 06511, USA

Email: gamze.gursoy@yale.edu

Bradley Malin

Vanderbilt University, Nashville, TN, 37203, USA

Email: b.malin@vumc.edu

Steven E. Brenner

University of California Berkeley, Berkeley, CA, 47420, USA

Email: brenner@compbio.berkeley.edu

With the advent of biotechnology, we are witnessing a paradigm shift in biology and medicine towards data intensive scientific discoveries. This brings a flood of physiological, contextual, and biological data at the individual and population level, and provides the perfect opportunity for integrating multidimensional information to study human variation and disease. This also creates new avenues in data breaches, necessitating careful studies of biomedical data privacy from a social, technical, and ethical standpoint. In this workshop, we explore the privacy issues related biomedical data use to advance human health. We will discuss the ethical and regulatory frames regarding privacy and implementations of such ethical considerations. We will also discuss technical approaches to ensuring privacy such as cryptographic techniques that allow responsible use and sharing of data.

An important lesson we have learned from the COVID19 pandemic is that data sharing among researchers, institutions, and even countries is essential and important for advancing biomedical research. We recognize that broad data sharing is a prerequisite for many important integrative analyses. However, improved technology, increased data availability, and a deeper understanding of the associations between different biological and clinical observations create new avenues for data breaches, necessitating a careful re-evaluation of privacy leakage in biological big data.

Studies of biological data privacy and security take attention of a diverse group of disciplines including genetics, biology, data science, computer science, law, and ethics. With the immense sensitivities and extremely identifying nature of biological data, especially genetic data (Erlich and Narayanan 2014; Gymrek et al. 2013; Hubaux, Katzenbeisser, and Malin 2017), we need to create avenues for interdisciplinary privacy discussions fueled by an understanding of the biology as well as ethics and technical challenges of use of such data. In part, the challenges in privacy of biomedical data use arise from a knowledge gap. There is a gap between the fields that generate and analyze personal biological data, the fields that establish theories and implementations for data privacy and

security, and the fields that study ethical and social implications of this data generation. By bringing ethicists and technology experts to the table, we aim to bridge this knowledge gap.

As the international scientific community, we acknowledge the benefit and importance of data sharing. Solutions such as access control can delay the access to the data, hence hampering the biomedical advances (Tryka et al. 2014; Fernandez-Orth, Lloret-Villas, and Rambla de Argila 2019). Moreover, the need to comply with different regulations in different geographies (e.g., HIPAA and GDPR), more biomedical data is increasingly being siloed behind firewalls. This highlights the importance of the ethical framework surrounding the data access solutions as well as the necessity of technical advances that prevent the privacy leakages while promoting data sharing (Knoppers and Beauvais 2021; Arellano et al. 2018).

This workshop represents various aspects of biological data use highlighting a number of ethical problems and technological solutions that need to be addressed to protect privacy of research participants and patients.

- Arellano, April Moreno, Wenrui Dai, Shuang Wang, Xiaoqian Jiang, and Lucila Ohno-Machado. 2018. "Privacy Policy and Technology in Biomedical Data Science." *Annual Review of Biomedical Data Science* 1 (1): 115–29.
- Erlich, Yaniv, and Arvind Narayanan. 2014. "Routes for Breaching and Protecting Genetic Privacy." *Nature Reviews. Genetics* 15 (6): 409–21.
- Fernandez-Orth, Dietmar, Audald Lloret-Villas, and Jordi Rambla de Argila. 2019. "European Genome-Phenome Archive (EGA) - Granular Solutions for the next 10 Years." In *2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE. <https://doi.org/10.1109/cbms.2019.00011>.
- Gymrek, Melissa, Amy L. McGuire, David Golan, Eran Halperin, and Yaniv Erlich. 2013. "Identifying Personal Genomes by Surname Inference." *Science (New York, N.Y.)* 339 (6117): 321–24.
- Hubaux, Jean-Pierre, Stefan Katzenbeisser, and Bradley Malin. 2017. "Genomic Data Privacy and Security: Where We Stand and Where We Are Heading." *IEEE Security & Privacy* 15 (5): 10–12.
- Knoppers, Bartha Maria, and Michael J. S. Beauvais. 2021. "Three Decades of Genetic Privacy: A Metaphoric Journey." *Human Molecular Genetics*, June. <https://doi.org/10.1093/hmg/ddab164>.
- Tryka, Kimberly A., Luning Hao, Anne Sturcke, Yumi Jin, Zhen Y. Wang, Lora Ziyabari, Moira Lee, et al. 2014. "NCBI's Database of Genotypes and Phenotypes: DbGaP." *Nucleic Acids Research* 42 (Database issue): D975-9.