

New file type improves genomic data sharing while maintaining participant privacy

October 17 2018







Gamze Gursoy, Ph.D., Yale University. Credit: Gamze Gursoy

Based on an analysis of data leakages and opportunities to prevent the potential misuse of genetic information, researchers have developed a new file format for functional genomics data that enables data sharing while protecting the personal information of research participants. The findings were presented at the American Society of Human Genetics (ASHG) 2018 Annual Meeting in San Diego, Calif.

Functional genomics is the study of how the genome functions in the body, such as how genes are regulated, are expressed into proteins, and interact with proteins to affect cellular functions in disease and health. Gamze Gursoy, Ph.D., postdoctoral research associate at the Yale University Computational Biology and Bioinformatics Program, and her colleagues set out to identify weaknesses in current functional genomics data files and processes and to find practical fixes.

"As functional genomics technology is still emerging, the data resulting from this research has not been well-studied by privacy researchers," said Dr. Gursoy. Previous analyses have shown that in certain cases, it is possible to trace de-identified functional genomics data back to the individual participant, a concept known as data leakage. Through a series of tests in the past few years, Dr. Gursoy and her colleagues measured the amount of variant information leaked in gene expression and functional genomics experiments involving different data types, and the extent to which this information could be mapped back to individuals.

"Just like genetic data, this data comes from real individuals, and we wanted to raise awareness that there could be leakages. At the same time, we want to democratize access to data and avoid bureaucratic hurdles,"



she said. To accomplish this goal, the researchers developed ways to measure leakage from raw functional genomics data and a <u>file format</u> to reduce the leakage in a targeted way.

Notably, the format they developed is easily layered onto <u>genetic data</u> file types already in common use, such as sequence alignment mapping and binary alignment mapping. Dr. Gursoy hopes its ease of use encourages more researchers to make their findings available through the proper channels.

"We want to balance participant privacy with flow of scientific information," said Dr. Gursoy. "If researchers restrict their data completely, scientific discovery stops."

Dr. Gursoy is now working with existing data repositories, such as ENCODE. She emphasized that privacy protection is a continuous effort that does not stop with this one file format; it's also about educating the public.

"Genomic privacy is very unique," said Dr. Gursoy. "Genetic data can be used to link people to their disease status in certain databases. While there are laws in place like the Genetic Information Nondiscrimination Act, people are unaware that insurance companies cannot use your <u>genetic information</u> to refuse coverage."

Dr. Gursoy hopes that this file type will be adopted more widely, leading to more collaboration in the field and fewer hurdles to reproducing research. She continues to work on methods to provide research data in a timely manner while keeping information secure.

Provided by American Society of Human Genetics



Citation: New file type improves genomic data sharing while maintaining participant privacy (2018, October 17) retrieved 25 April 2024 from <u>https://techxplore.com/news/2018-10-genomic-privacy.html</u>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.