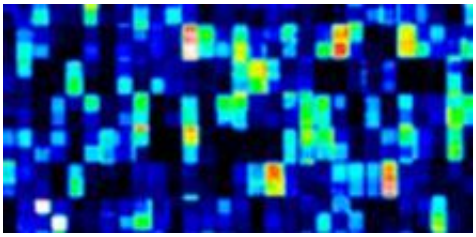
Researchers identify potential security hole in genomic data sharing network

October 29 2015



Sharing genomic information among researchers is critical to the advance of biomedical research. Yet genomic data contains identifiable information and, in the wrong hands, poses a risk to individual privacy. If someone had access to your genome sequence—either directly from your saliva or other tissues, or from a popular genomic information service—they could check to see if you appear in a database of people with certain medical conditions, such as heart disease, lung cancer or autism.

Work by a pair of researchers at the Stanford University School of Medicine makes that genomic data more secure. Suyash Shringarpure, PhD, a postdoctoral scholar in genetics, and Carlos Bustamante, PhD, a professor of genetics, have demonstrated a technique for hacking a network of global genomic databases and how to prevent it. They are working with investigators from the Global Alliance for Genomics and

Health on implementing preventive measures.

The work, to be published Oct. 29 in the *American Journal of Human Genetics*, also bears importantly on the larger question of how to analyze mixtures of genomes, such as those from different people at a crime scene.

A network of genomic data sets on servers, or beacons, organized by the National Institutes of Health-funded Global Alliance for Genomics and Health, allows researchers to look for a particular genetic variant in a multitude of genomic databases. The networking of genomic databases is part of a larger movement among researchers to share data. Identifying a gene of interest in a beacon tells researchers where to apply for more complete access to the data. A central assumption, though, is that the identities of those who donate their genomic data are sufficiently concealed.

"The beacon system is an elegant solution that allows investigators to 'ping' collections of genomes," said Bustamante. Investigators on the outside of a data set can ping and ask which data set has a particular mutation. "This allows people studying the same rare disease to find one another to collaborate."

Beacons' vulnerability

But many genomic data sets are specific to a condition or disease. A nefarious user who can find the match for an individual's genome in a heart disease beacon, for example, can infer that the individual—or a relative of that person—likely has heart disease. By pinging enough beacons in the network of beacons, the hacker could construct a limited profile of the individual. "Working with the Global Alliance for Genomics and Health, we've been able to demonstrate that vulnerability and, more importantly, how to put policy changes in place to minimize

the risk," said Bustamante.

To protect donors' identities, the organizers of the network, which is called the Beacon Project, have taken steps, such as encouraging beacon operators to "de-identify" individual genomes, so that names or other identifying information are not connected to the genome.

Despite such efforts, Shringarpure and Bustamante calculated that someone in possession of an individual's genome could locate that individual within the beacon network. For example, in a beacon containing the genomes of 1,000 individuals, the Stanford pair's approach could identify that individual or their relatives with just 5,000 queries.

Genomic information isn't completely covered by the federal law that protects health information, and the consequences for a person whose information is disclosed can be significant. For example, although the national Genetic Information Nondiscrimination Act prevents health insurers from denying someone coverage or raising someone's premiums because they have a particular genetic variant, the act does not apply to other forms of insurance, such as long-term care, disability or life insurance.

Approaches for better security

The Beacon Project has the potential to be enormously valuable to future genetic research. So plugging this security hole is as important to Shringarpure and Bustamante as to the Global Alliance for Genomics and Health. In their paper, the Stanford researchers suggest various approaches for making the information more secure, including banning anonymous researchers from querying the beacons; merging data sets to make it harder to identify the exact source of the data; requiring that users be approved; and limiting access in a beacon to a smaller region of

the genome.

Peter Goodhand, executive director of the Global Alliance for Genomics and Health, said, "We welcome the paper and look forward to ongoing interactions with the authors and others to ensure beacons provide maximum value while respecting privacy."

Goodhand also said that the organization's mitigation efforts, which adhere to the best practices outlined in its privacy and security policy, include aggregating data among multiple beacons to increase database size and obscure the database of origin; creating an information-budgeting system to track the rate at which information is revealed and to restrict access when the information disclosed exceeds a certain threshold; and introducing multiple tiers of secured access, including requiring users to be authorized for data access and to agree not to attempt specific risky scenarios.

Shringarpure and Bustamante are also interested in applying the technique described in their study to the area of DNA mixture interpretation, in which investigators seek to identify one DNA sequence in a mixture of many similar ones. The DNA mixing problem is relevant to forensics, studies of the microbiome and ecological studies. For example, Bustamante said, if a weapon used in a crime had DNA from several people on it, DNA mixture interpretation can help investigators pick out the DNA of a particular person, such as the suspect or the victim, revealing whether they touched the weapon. In fact, investigators could potentially use the same type of analysis used on the beacon network to look for individuals who may have touched a railing in a subway station or other public space.

Provided by Stanford University Medical Center