

NEWS

# Some genetic repositories may put private details at risk

BY JESSICA WRIGHT

5 NOVEMBER 2015

An online portal designed to give researchers easy access to genomic data while protecting the participants' privacy may still reveal some sensitive information. The findings, published last week in the *American Journal of Human Genetics*, highlight the delicate balance between the desire for scientific openness and the need to safeguard genetic information<sup>1</sup>.

The study focuses on 'beacons,' a set of servers that a group of researchers devised to deliver genetic information. Institutions can set up and maintain their own beacon, but all of the servers are accessible from a single portal.

This portal allows users to anonymously query more than 200 databases of sequences to see whether any individual in the databases carries a particular genetic variant. The beacons return only a 'yes' or 'no' answer for each variant.

A yes means only that at least one person among the hundreds or thousands of people in a beacon carries the variant. But someone with access to an individual's genome sequence could submit enough queries to find out whether that individual is in a particular database, the new study found.

For some databases, this information is unlikely to reveal anything more. But for other databases, it might signal, for example, that the individual or one of his or her family members has a condition such as autism.

## Beacon bits:

In the study, researchers outlined the calculations necessary to identify an individual in a beacon, and suggested ways to close this loophole. They calculated that it would take 5,000 variants to locate an individual in a database of 1,000 participants. They looked specifically at several

beacons, including one autism beacon that holds information from the **Simons Simplex Collection (SSC)**.

The SSC is a collection of families that have one child with autism and unaffected parents and siblings. The researchers estimate that it would take nearly 35,000 queries to find someone in the SSC. Identifying an individual would reveal that this person either has autism or has a family member with the condition.

Before publishing how to ‘hack’ the beacons, the researchers got in touch with the **Global Alliance for Genomics and Health**, which runs the beacon project. “We didn’t just want to unleash it and give people a secret decoder ring that they could use [to identify a participant],” says **Carlos Bustamante**, professor of genetics at Stanford University in California.

For example, after learning of the study’s results, the Simons Foundation, which funds the SSC (and *Spectrum*), has taken steps to ensure that it is no longer possible to use this method to link an individual to the SSC, says **Alex Lash**, chief informatics officer for the Simons Foundation Autism Research Initiative.

## **Mitigating risk:**

People have been struggling to balance privacy with open access to genetic information for years. A 2008 study showed that it was possible to identify individual participants in genetic studies that associate the frequency of a variant with a disorder<sup>2</sup>. As a result, researchers stopped making **these raw data publicly available**.

Researchers conceived of the beacon project as a way to mitigate these risks while still enabling open sharing of genetic data. The beacon network contains more than 200 datasets from 19 institutions. Four beacons contain individuals with a specific disorder: autism, heart disease, cancer or ulcerative colitis. Researchers searching for a particular variant can learn whether it is present or absent in each of these databases and then apply for further information with each institution.

It is possible to search the beacons without logging in, which allows researchers to engineer computer programs that automatically scan through thousands of variants. But this also opens the door to security problems.

“We think that anonymous pings are always going to be vulnerable because you don’t know the intentions of somebody,” Bustamante says. “We should have a system that people log onto and that tracks them in the same way as for hospital [medical records].”

But some researchers say it’s important to put the risks in perspective. The beacons reveal only whether somebody is a participant in a database. And it would take a lot of work to gain illicit access to somebody’s genetic information just to find this out.

“We need to put [the study] in the right context,” says **Yaniv Erlich**, assistant professor of computer science at Columbia University, who was not involved in the work. Erlich says the security concerns are minimal compared with privacy issues on the Internet in general. “If there was a beacon of drug users or pedophiles, you wouldn’t want people to know about you.”

## Safety in numbers:

In response to the study, the Global Alliance for Genomics and Health has created a conglomerate beacon to house information from databases that may include sensitive information.

Although aggregating the databases will not affect the amount of information available to researchers, it could take them longer to access it, says **David Haussler**, professor of biomolecular engineering at the University of California, Santa Cruz, and vice chair of the steering committee for the Global Alliance.

The conglomerate does not reveal which individual database a result is from, but at some point in the future, an institution may be able to choose to indicate that its database is included.

“Who is in it and how big it is will change as people decide to opt in or opt out,” says **Peter Goodhand**, executive director of the Global Alliance.

The SSC is now part of the conglomerate. “Before, a ‘yes’ answer [for an SSC variant] was meaningful, because you knew it was from a family with autism,” says Lash. “Now we’ve obscured it so that ‘yes’ answers are less meaningful, but ‘no’ answers are still meaningful.”

To follow up on a ‘yes’ answer, a user would have to contact the Global Alliance. The alliance would then verify that this user is a researcher and help them to gain access to the data.

People who manage databases should explain the risks of a beacon to the participants at the outset, says Haussler, because the the benefits still far outweigh the risks.

“If we do put in more hoops, more barriers, we have to make them easy and more automatic [to overcome],” he says. “Why do I buy a book on Amazon? Because it’s so damn easy. It’s one click.”

## REFERENCES:

1. Shringarpure S.S and C.D. Bustamante *Am. J. Hum. Genet.* Epub ahead of print (2015) [PubMed](#)
2. Homer N. *et al. PLoS Genet.* **4**, e1000167 (2008) [PubMed](#)