

IT researchers break anonymity of gene databases

March 24 2017

DNA profiles can reveal a number of details about individuals, and even about their family members. There are laws in place that regulate the trade of gene data, which has become much simpler and cheaper to analyze today. However, these laws do not apply to an equally relevant type of genetic data, so-called microRNAs, even though these can also point to serious diseases. This means that anonymity needs to be strictly maintained in microRNA studies as well. Researchers from the Research Center for IT Security, CISP, have now been able to show that a few microRNA molecules are sufficient to draw conclusions about study participants. The computer scientists will be presenting their means of attack, and appropriate countermeasures, at the CeBIT computer fair in Hannover (Hall 6, Stand C47).

"Ever since American scientists successfully attacked a study in 2008, where just parts of the DNA were enough to identify participants, researchers have been debating if, and in what detail, we should be permitted to publish gene [data](#)," says Michael Backes, who is professor for Cryptography and IT Security at Saarland University, and the scientific director of the Competence Center for IT security, CISP. "Luckily we don't have that problem in Germany, that [health insurance companies](#) can ask for more money from someone who is sick," says Pascal Berrang, who is researching aspects of data privacy in [genetic data](#) as part of his doctoral studies at CISP. This is different in the United States, for instance, where there is already a flourishing trade in health data. Not even medical studies are safe, says Berrang.

The researchers from Saarbrücken, together with their colleagues Mathias Humbert and Praveen Manoharan, focused on analyzing data security issues for a specific kind of gene information, one that is now commonly used in medical research: microRNAs. These short molecules of ribonucleic acid have recently gained importance as new forms of biomarkers – biological identifiers that clearly indicate a patient's general health condition, or the presence of certain diseases, to physicians and researchers. MicroRNAs can therefore divulge even more details about a patient's condition than conventional DNA analysis, since the latter only yields the probability of the patient developing the disease in question. This aspect of microRNA analysis makes the Saarbrücken computer scientists' findings even more significant. Using two different attack techniques, they were able to break the anonymity of the test subjects in a microRNA study. "If the results were published, and a [health insurance](#) fund knew the microRNA profile of one of its members, it could deduce whether that patient was part of the study, and pinpoint individual diseases," Pascal Berrang says; that would be more than enough information.

The CISPA researchers have also been working on developing countermeasures. The main challenge was to maintain the anonymity of the data without making it unusable for [medical research](#) and diagnostics. These circumstances made two different strategies necessary. The first was to omit any telltale molecules of microRNA that were not relevant to the diagnosis; the other was to introduce additional random noise to the data, which helps to protect the identity of individual participants without distorting the overall tendency of the results. The second technique has become a commonly used tool for publishing statistical information, as it helps prevent the disclosure of identifying information; a principle experts call "differential privacy".

"Leaving telltale molecules out of the publication doesn't really help. Even if you published only ten molecules instead of a hundred, the

attack would still be feasible," says Pascal Berrang. The second intervention, the addition of random noise, did not prevent the attack, but just made work more difficult for the medical staff. For this reason, the CISPA researchers recommend that the data be randomized as little as possible, and that a sufficiently large number of participants take part in the trial. "This has several advantages: It increases the statistical relevance of the study, it requires less random noise, and the study is not as susceptible to these forms of attacks, because the more people take part, the more the individual blends into the greater crowd," says Berrang. In terms of specific numbers, he says: "Two hundred. At least 200 people in the study, and a bit of [random noise](#) in the data, that should be enough."

Provided by Saarland University

Citation: IT researchers break anonymity of gene databases (2017, March 24) retrieved 26 May 2024 from <https://phys.org/news/2017-03-anonymity-gene-databases.html>

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.