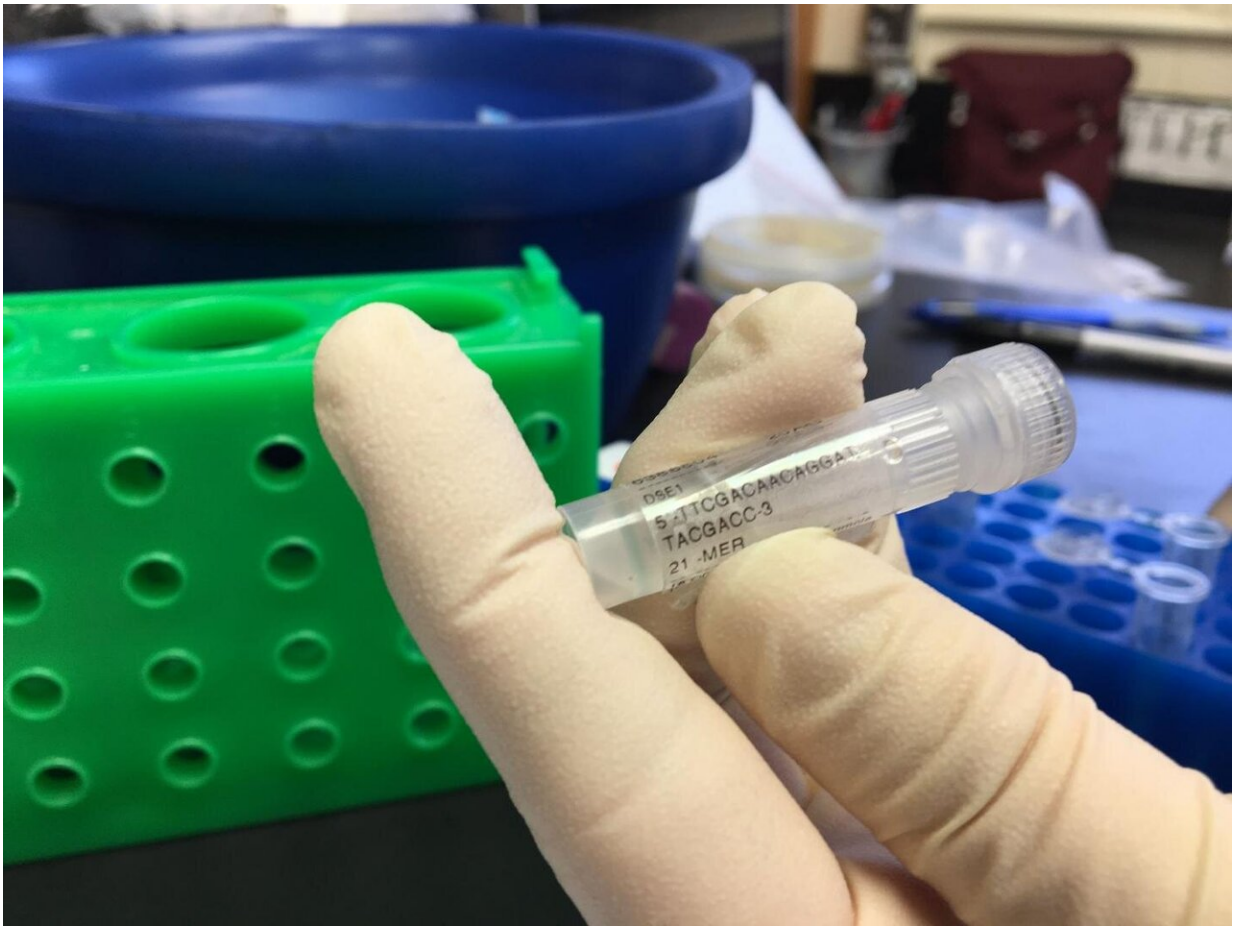
Genetic barcodes can ensure authentic DNA fingerprints

May 21 2020, by Ken Kingery



Secret primers, as the one shown in this photo, could be used to reveal genetic 'barcodes' added to DNA samples taken in the field to ensure they arrived at the laboratory unaltered. Credit: Mohamed Ibrahim, Duke University

Engineers at Duke University and the New York University's Tandon School of Engineering have demonstrated a method for ensuring that an increasingly popular method of genetic identification called "DNA fingerprinting" remains secure against inadvertent mistakes or malicious attacks in the field.

The technique relies on introducing genetic "barcodes" to DNA samples as they are collected and securely sending information crucial to identifying these barcodes to technicians in the laboratory. The system shows one way to guarantee that a sample taken in the field, transported to a lab and processed for genetic identification is genuine.

The results appear online on May 14 in the journal *IEEE Transactions on Information Forensics and Security*.

"If you think about conventional encryption techniques, like security for a smartphone, there's usually a passcode that only one person knows," said Mohamed Ibrahim, a system-on-chip design engineer at Intel Corporation and recent Duke electrical and computer engineering Ph.D. graduate. "Our idea is to inject non-harmful material into genetic samples immediately when they are collected in the field that act as a similar password. This would ensure that the samples are authentic when they reach the processing stage."

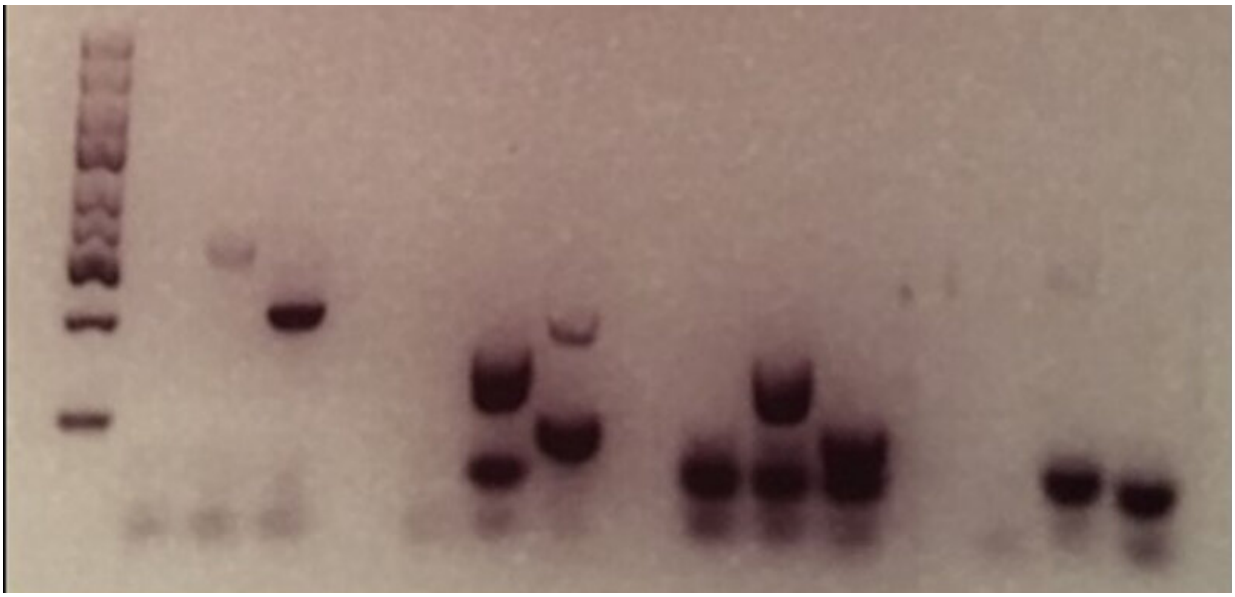
DNA fingerprinting is a method of identifying a specific person, organism or disease based on only a small amount of genetic material. While about 99.9 percent of the DNA between two unrelated humans is the same, that still leaves around three million base pairs that are different. And within that potentially identifying dataset, certain short segments of DNA sequences are much more likely than others to vary in composition from person to person.

Rather than sequencing a person's entire genome, which still costs more

than $1,000 a pop, scientists can target a handful of these short sequences for identification. In DNA fingerprinting, a technique called polymerase chain reaction (PCR) replicates the genetic sequences at these sites repeatedly so that they can easily be read. Based on the specific combinations of nucleic acids at these various sites, genetic samples can be matched to their sources. While it may seem like data from hundreds of these sites would be needed to make a definitive match, they vary so much from person to person that the Federal Bureau of Investigations currently recommends that only 13 are necessary.

As the popularity of this technique and the PCR technology underlying it increases, multiple companies are in a race to simplify the process and create cheaper solutions. And as these devices become smaller, more complex and more automated, it may create more opportunities to attack the process. Recent studies suggest that these opportunities raise unprecedented security concerns, creating a whole new category of potential weaknesses that has been dubbed "cyberbiosecurity threats."

"Researchers have identified a diverse array of cyberbiosecurity threats over the past few years," said Krishnendu Chakrabarty, the John Cocke Distinguished Professor of Electrical and Computer Engineering at Duke. "Our main goal is to become a part of the community trying to address these threats by focusing on one of the most vulnerable time periods, which is before a sample even gets to the lab."



A DNA test that reveals the presence or absence of the genetic barcode added to a DNA sample after being taken in the field. The first set of columns has no barcodes because no primers were used. The second set of columns display the barcodes the laboratory technicians are looking for. The third set of columns matches the second set, indicating a sample is authentic, while the fourth set of columns does not match, indicating something has gone awry with the samples. Credit: Duke University

In their new paper, Chakrabarty, Ibrahim, Tung-Che Liang, a current doctoral student in Chakrabarty's laboratory, Ramesh Karri, professor of electrical and computer engineering at NYU Tandon, and Kristin Scott, adjunct assistant professor of molecular genetics and microbiology at Duke, demonstrate the usefulness of a genetic barcode to ensure samples taken in the field are not swapped or otherwise tampered with on their way to the laboratory.

Relying on the genetic PCR expertise of Scott and working out of her laboratory, the researchers first added two short stretches of synthetic

DNA to genetic samples bound for DNA fingerprinting. Because they are synthetic, they can be made in almost any combination of the four available DNA base pairs imaginable. And at 280 and 190 base pairs each, the chances of correctly guessing the genetic combination is vanishingly small.

"We analyzed the conditions an adversary needs to satisfy to undermine the barcoding system," said Karri, who is a co-founder and co-chair of the NYU Center for Cyber Security. "These conditions are related to the physical characteristics of the molecular barcode. By linking these conditions back to how barcodes are generated and the expansiveness of the search space, we show that the probability that an adversary can discover the barcode is negligibly low."

Meanwhile, the primers needed to amplify each barcode are sent securely to the technicians in the laboratory. For a PCR machine to repeatedly copy a specific segment of DNA, it first must know how that sequence starts and ends. Primers provide that information, and without it, an attacker would have no chance of amplifying the correct barcodes.

Once the technicians finish an initial PCR run with the samples and primers provided, the two barcodes appear as peaks or lines in the resulting genetic data, depending on the method being used to identify them. To make sure the samples are authentic and not tampered with, the technicians must simply make sure these two barcodes appear as expected.

"When the right primers are used to unlock a barcode, you should get a positive result," said Ibrahim. "If you don't, then that means that the sample is not genuine. Some sort of switching or alteration has occurred."

While this system currently relies on information being transmitted

securely to the laboratory, the researchers say there are ways that the genetic barcodes could be streamlined into the technology. For example, the barcodes could be correlated in some way to the samples being sent, and technicians could look up the correct primers to use from a database. With the proper hardware, the idea could also conceivably be translated into the chips running the DNA analysis themselves.

"This work is an excellent example of multiple disciplines coming together to develop new solutions to existing real-world problems," said Scott.

More information: Mohamed Ibrahim et al, Molecular Barcoding as a Defense against Benchtop Biochemical Attacks on DNA Fingerprinting and Information Forensics, *IEEE Transactions on Information Forensics and Security* (2020). DOI: 10.1109/TIFS.2020.2994742

Provided by Duke University