

New software could make real-time DNA-authentication a reality

November 30, 2017

In the science-fiction movie *Gattaca*, visitors only clear security if a blood test and readout of their genetic profile matches the sample on file. Now, cheap DNA sequencers and custom software could make real-time DNA-authentication a reality.

Researchers at Columbia University and the New York Genome Center have developed a method to quickly and accurately identify people and cell lines from their DNA. The technology could have multiple applications, from identifying victims in a mass disaster to analyzing crime scenes. But its most immediate use could be to flag mislabeled or contaminated cell lines in cancer experiments, a major reason that studies are later invalidated. The discovery is described in the latest issue of *eLife*.

“Our method opens up new ways to use off-the-shelf technology to benefit society,” said the study’s senior author Yaniv Erlich, a computer science professor at Columbia Engineering, an adjunct core member at NYGC, and a member of Columbia’s Data Science Institute. “We’re especially excited about the potential to improve cell-authentication in cancer research and potentially speed up the discovery of new treatments.”

The software is designed to run on the MinION, an instrument the size of a credit card that pulls in strands of DNA through its microscopic pores and reads out sequences of nucleotides, or the DNA letters A, T, C, G. The device has made it possible for researchers to study bacteria and viruses in the field, but its high error-rate and large sequencing gaps have, until

now, limited its use on human cells with their billions of nucleotides.

In an innovative two-step process, the researchers outline a new way to use the \$1,000 MinION and the abundance of human genetic data now online to validate the identity of people and cells by their DNA with near-perfect accuracy. First, they use the MinION to sequence random strings of DNA, from which they select individual variants, which are nucleotides that vary from person to person and make them unique. Then, they use a Bayesian algorithm to randomly compare this mix of variants with corresponding variants in other genetic profiles on file. With each cross-check, the algorithm updates the likelihood of finding a match, rapidly narrowing the search.

Tests show the method can validate an individual's identity after cross-checking between 60 and 300 variants, the researchers report. Within minutes, it verified the identity of the study's lead author, Sophie Zaaier, a former member of NYGC and now a postdoctoral researcher at Cornell Tech.

To do this, the MinION matched the readout of Zaaier's genome, gleaned from a sample of cheek cells, with a reference profile stored among 31,000 other genomes on the public database, DNA.land. Erlich's identity was verified the same way, with initial sequencing done by Columbia students in the Ubiquitous Genomics class he and Zaaier taught in 2015.

They call their re-identification technique 'MinION sketching' which Zaaier compares to the brain's ability to make out a bird from a few telling features in an abstract Picasso line-drawing. The MinION's genetic 'sketch' of a cell-sample is compared to a growing database of sketches –similarly incomplete genetic profiles produced by at-home DNA-test kits like 23andMe and donated to science by consumers.

"Using our method, one needs only a few DNA reads to infer a match to an individual in the database," says Zaaier.

The most promising use for 'MinION sketching' may be as a cheap cell-authentication tool in experimental research, say scientists familiar with its capabilities. In the study, researchers quickly matched a strain of leukemic cells sequenced by the MinION against a reference file in the Cancer Cell Line Encyclopedia database, they report. When they tried contaminating the cells with other cultures, it correctly rejected a match if contamination levels climbed above 25 percent.

The use of misidentified or contaminated cell lines in medical research is blamed for as much as a third of the estimated \$28 billion spent each year on studies that can't be replicated, according to one recent study. In a 2014 essay in *Science*, the director of the National Institute of General Medical Sciences, Jon Lorsch, called for new policies and technologies to address the problem.

Lacking the expensive machinery needed to validate cell lines on their own, most researchers either skip validation or ship their cultures to specialized labs which can delay important findings and treatments. If an easier alternative were available, most researchers would use it, says Neville Sanjana, a core faculty member at NYGC and assistant professor at NYU's Department of Biology who works on skin and lung cancer cell lines and was not involved in the study.

"No one wants to waste time and reagents working on the wrong cells," he says. "At the right price, every lab will adopt this."

Source:

<http://news.columbia.edu/content/1791>