

Open source DNA

31 August 2009



This is Dr. Eran Halperin of Tel Aviv University. Credit: AFTAU

A new mathematical tool from Dr. Eran Halperin of TAU's Blavatnik School of Computer Science aims to protect genetic privacy while giving genomic data to researchers.

In the chilling science fiction movie *Gattaca*, Ethan Hawke stars as a man with "inferior genes" who assumes another's genetic identity to escape a dead-end future. The 1997 film illustrates the very real fear swirling around today's [genome research](#) — fear that private genetic information could be used negatively against us.

Last year, after a published paper found serious security holes in the way DNA data is made publicly available, health institutes in the United States and across the world removed all [genetic data](#) from public access.

"Unfortunately, that knee-jerk response stymied potential breakthrough genetic research," says Dr. Eran Halperin of Tel Aviv University's Blavatnik School of Computer Sciences and Department of Molecular Microbiology and Biotechnology. He wants to put this valuable DNA information back in circulation, and has developed the tool to do it — safely.

Working with colleagues at the University of California in Berkeley, Dr. Halperin devised a [mathematical formula](#) that can be used to protect genetic privacy while giving researchers much of the raw data they need to do pioneering medical research. Reported in this month's issue of [Nature Genetics](#), the tool could keep millions of research dollars-worth of DNA information available to scientists.

New security to restart genetic research

"We've developed a mathematical formula and a software solution that ensures that malicious eyes will have a very low chance to identify individuals in any study," says Dr. Halperin, who is also affiliated with the International [Computer Science](#) Institute in Berkeley.

The mathematical formula that Dr. Halperin's team devised can determine which SNPs ? or small pieces of DNA - that differ from individual to individual in the human population ? are accessible to the public without revealing information about the participation of any individual in the study. Using computer software that implements the formula, the National Institutes of Health and similar institutes around the world can distribute important research data, but keep individual identities private.

"We've been able to determine how much of the DNA information one can reveal without compromising a person's identity," says Dr. Halperin. "This means the substantial effort invested in collecting this data will not have been in vain."

Why is this information so important? Genome association studies can find links in our genetic code for conditions like autism and predispositions for cancer. Armed with this information, individuals can avoid environmental influences that might bring on disease, and scientists can develop new gene-based diagnosis and treatment tools.

A new track for government policymakers

Examining SNP positions in our genetic code, Dr. Halperin and his colleagues demonstrated the statistical improbabilities of identifying individuals even when their complete genetic sequence is known. "We showed that even when SNPs across the entire genome are collected from several thousand people, using our solution the ability to detect the presence of any given individual is extremely limited," he says.

Dr. Halperin hopes his research will reverse the NIH policy, and he will provide access to the software so that researchers can use it to decide which genetic information can be safely loaded into a public database. He also hopes it will quell raging debates about DNA usage and privacy issues.

The Tel Aviv University-Berkeley research was done while Dr. Halperin was working with the International Computer Science Institute (ICSI), a non-profit research institute with close relations to the University of California (UC) and Tel Aviv University. Other coauthors of the study include Sriram Sankararaman, and Prof. Michael Jordan from UC, and Dr. Guillaume Obozinski from Willow, a joint research team between INRIA Rocquencourt, École Normale Supérieure de Paris and Centre National de la Recherche Scientifique.

Source: Tel Aviv University ([news](#) : [web](#))

APA citation: Open source DNA (2009, August 31) retrieved 23 September 2020 from <https://medicalxpress.com/news/2009-08-source-dna.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.