

# Gene matches could aid science, but raise privacy concerns

May 16 2017, by Nathan Collins

---



A new study showing how to link distinct sets of genetic data could aid police and scientists but may raise legal questions about the use of forensic DNA.  
Credit: Getty Images

How much could one really figure out about a person from 13 tiny snippets of DNA? At first glance, not much – in the world of genetics, 13 is tiny. But a new study suggests it may be enough to infer hundreds of thousands more markers, potentially revealing a wealth of genetic information, Stanford biologists report May 15 in *Proceedings of the National Academy of Sciences*.

Those results may help foster scientific collaborations and aid researchers working with degraded or incomplete DNA samples, such as those collected from wildlife or [archaeological sites](#), said Noah Rosenberg, a professor of biology and the new paper's senior author.

But the ability to infer so much on the basis of so little information raises privacy concerns as well, Rosenberg said. It also suggests that the assumptions about genetics that underlie a number of recent legal arguments, including the Supreme Court's decision to uphold a controversial law concerning the collection of forensic DNA, may not be quite correct.

## **Patterns in our genes**

The new findings are based on two sets of genetic data from 872 human genomes. The first comprised just 13 markers that until this year were the basis of the FBI's forensic genetic marker set, the Combined DNA Index System, or CODIS. (The system was recently upgraded to include seven additional markers, bringing the total to 20.)

The second, much broader dataset included 642,563 [genetic markers](#) that did not overlap with the first set. The question was, how well could Rosenberg and his team match a person's record in one dataset to their record in the other? Put differently, how well could they predict the second set of genetic markers based solely on the first, forensic set?

Pretty well, actually. Rosenberg and team found there were strong enough patterns in our DNA – or at least in the DNA of the diverse set of people they studied – that they could match upward of 90 percent of the records. If they added in 17 more forensic markers, bringing the total to 30, they could match more than 99 percent of the records in the two datasets – meaning that with the right combination of databases, it may be possible to infer a wealth of [genetic information](#) based on a very small set of markers.

## **Is prediction the enemy of privacy?**

The team's conclusions suggest there may be flaws in the way law enforcement officials, courts and businesses that conduct genetic tests have thought about genetic privacy. Previously, it had been assumed that forensic DNA collections were only useful for matching DNA samples to names already in a database – that is, for placing a suspect at a crime scene – and fundamentally could not reveal any information beyond identity matches.

That assumption was a key element in the Supreme Court's 2013 decision in *Maryland v. King*, which upheld the state's practice of retaining DNA from anyone who'd been arrested there. Since the CODIS markers could not be used to infer private health data or other traits, the majority argued, the benefits of recording them from anyone even suspected of a crime outweighed those suspects' privacy concerns.

Similarly, genetic testing company 23andMe argued in a blog post last year that their data was unlikely to be useful to police. Their data relies on a different set of markers than those used for forensic analysis, so, they argued, it was very difficult to connect police records with theirs.

Such arguments may need to be reconsidered, Rosenberg said, because when the same person is included in more than one genetic database, it

may be possible to infer genetic traits from CODIS data or to find matches across different sets of DNA markers.

## **The upside**

Privacy and legal issues aside, "there are several other places where this result is useful," Rosenberg said.

"The approach we are using dates back to the 1960s, when computer scientists and statisticians were first trying to figure out how to link records from the same people in different government, medical or corporate databases," said Michael Edge, a recent PhD graduate and lead author on the paper. "It is interesting to see that the same type of problem arises in so many contexts in genetics."

One issue is backward compatibility of the forensic marker system, which is what drew the team to the problem in the first place. The problems forensic geneticists face are often harder than simply matching profiles – for example, determining whether one person's DNA is present in a mixture of several people's DNA left on a doorknob at a crime scene.

With just 13 or 20 genetic markers, there is a substantial risk of false positive matches, Rosenberg said. Using larger and more modern marker sets would reduce false positive rates, but that introduces another problem: It might not be possible to check for matches against decades of profiles collected with the 13 markers that have been used to date.

The new results, Rosenberg said, give a proof of principle that it may be possible to develop a forensic genetic system with new marker sets and still be able to test for matches against databases assembled with the earlier CODIS markers.

The findings may also help scientists fill in missing details from DNA samples when they do not know if they are sampling the same individual. For example, when wildlife biologists want to study DNA from elusive animals, they cannot always take a blood sample. Instead, they often rely on hair or scat samples, and it can be difficult to tell if the same animal has been sampled multiple times. The same is true when sampling DNA from ancient bones recovered at archaeological sites. In both cases, Rosenberg said, the new results suggest that some of the missing genetic details could be filled in.

And then there's "a scenario that's happened to me at least twice," Rosenberg said: data sharing. It is not uncommon for collaborators from two labs to want to share different kinds of data on the same people – except it is not so simple as just sending over the data when the two datasets might have samples that are shared in common but that are labeled differently. In principle, Rosenberg said, the results of the new study could be used to match entries across datasets, making research collaborations that much easier.

**More information:** Michael D. Edge et al. Linkage disequilibrium matches forensic genetic records to disjoint genomic marker sets, *Proceedings of the National Academy of Sciences* (2017). [DOI: 10.1073/pnas.1619944114](https://doi.org/10.1073/pnas.1619944114)

Provided by Stanford University

Citation: Gene matches could aid science, but raise privacy concerns (2017, May 16) retrieved 9 April 2024 from <https://phys.org/news/2017-05-gene-aid-science-privacy.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.