

Not Guilty,

CRIME LABS NEED NEW TOOLS TO INTERPRET DNA EVIDENCE.
A TEAM OF BU FORENSIC SCIENTISTS IS ON THE CASE.

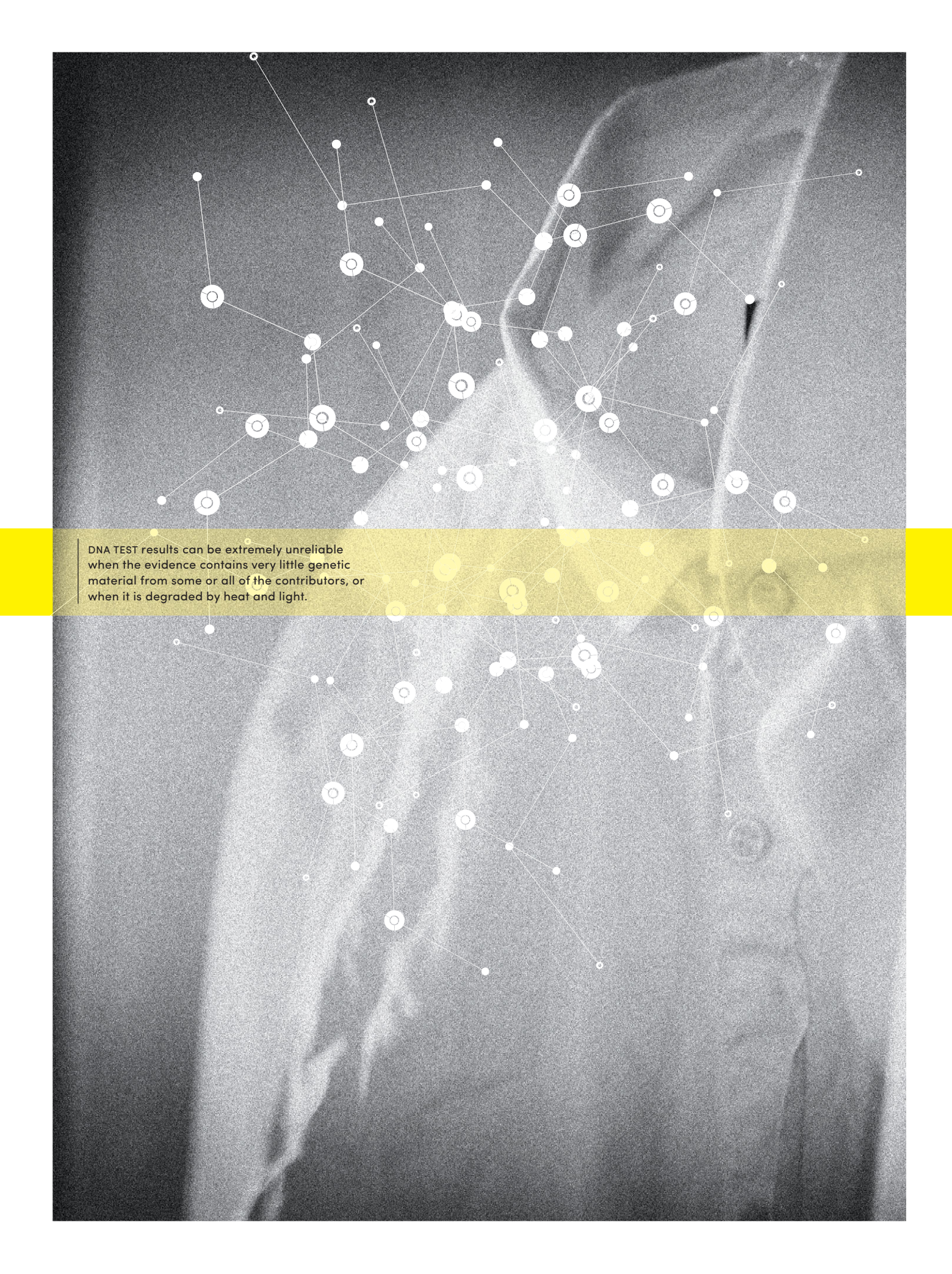
Probably

By Chris Berdik | *Photographs by Dan Aguirre*

44

BOSTONIA

Fall 2015



DNA TEST results can be extremely unreliable when the evidence contains very little genetic material from some or all of the contributors, or when it is degraded by heat and light.

IN 2010 BOISE STATE UNIVERSITY FORENSIC biologist Greg Hampikian was contacted by the lawyer for Kerry Robinson, who is serving 20 years in prison after being convicted in 2002 of raping a woman in Moultrie, a small town in southern Georgia. The woman, who was raped by three men, identified only one of her attackers, a man named Tyrone White, and DNA analysis provided what court records called “essentially a conclusive match.”

As part of a plea bargain to reduce his sentence, White named Robinson as one of the rapists. But Robinson’s genetic variations—key markers used to identify the person the DNA belongs to—were such a poor match that as one forensic expert testified, up to 1,000 people among the 15,000 in Moultrie County could likely match the crime scene evidence to the same degree. Still, combined with White’s testimony, the prosecution was able to use the DNA evidence to convince a jury to convict Robinson.

Convictions like Robinson’s happen in large part because of the powerful influence DNA evidence has on a jury, and because of difficulties inherent in DNA analysis, particularly in cases involving multiple suspects and degraded samples.

When Hampikian and another forensic researcher sent the data from the trial evidence—a cheek swab from Robinson and DNA from the crime scene—to 17 accredited crime labs, only one agreed with the lab used by prosecutors. That lab had found that Robinson’s DNA shared some common genetic markers with the crime-scene evidence, meaning that he “could not be excluded” as a suspect. Of the other labs, 4 couldn’t conclude anything from the evidence, and 12 reported that Robinson *should be excluded* as a suspect. It’s important to note that these labs didn’t find differing numbers of shared genetic variations in the evidence. They just interpreted the strength of that evidence differently, and it’s the interpretation that matters in court.

“Errors in DNA forensics can be multiplied in the justice system,” says Hampikian. Often, he says, DNA is used to corroborate otherwise flimsy evidence. Robinson, for instance, claimed that White had named him because he suspected that Robinson had turned him in to the police. Just 2 of Robinson’s genetic markers were found in the evidence (compared with 11 of White’s), but because he had no corroborated alibi, that was enough for the lab to say he might have been at the crime scene.

Because of DNA’s vaunted reputation, Hampikian says, “all this weak evidence gets propped up by science.”

Robinson’s lawyer, Rodney Zell, says that a habeas petition, a claim of wrongful imprisonment, is pending in the court where Robinson was convicted.

Ironically, improvements in DNA technology, which is now 100 times as sensitive as it was at the dawn of DNA forensics in the 1980s, have made the science more problematic, at least in cases where genetic material comes from several people. Now, to fix the problem created by improved technology, forensic researchers must build even better technology. That’s what Catherine Grgicak (pronounced Ger-gi-chuk), a School of Medicine assistant professor of anatomy and neurobiology teaching in the Biomedical Forensic Sciences Program, hopes to do.

Backed by \$2.5 million in government funding from the US Department of Justice and Department of Defense, she and her team are developing software that could help crime labs unwind the genetic evidence that can identify the guilty without entangling the innocent.

“There are no national guidelines or standards saying that labs have to meet some critical threshold of a match statistic” to conclude that a suspect might have been at a crime scene, says Grgicak. Neither are there guidelines about when a DNA mixture is simply too complicated to analyze in the first place. Often, labs aren’t even certain how many people contributed to the jumble of DNA detected on a weapon or a victim’s clothing. Test results can be extremely unreliable when the evidence contains very little genetic material from some or all the contributors, and DNA is degraded by heat and light.

Existing software is so unreliable in so many cases that even the Innocence Project, a nonprofit legal organization that often relies on DNA testing to exonerate the wrongly convicted, doesn’t support the establishment of a standard match statistic that would exclude or include suspects of a crime. “We support national standards that are based on scientifically grounded and properly validated tools,” says Paul Cates, the organization’s communications director. “But it seems a little premature to be talking about those standards.”


Unfortunately for many defendants, such skepticism is largely absent among the general public. According to the Marshall Project, a nonprofit, nonpartisan news organization covering America’s criminal justice system, research conducted at the University of Nevada, Yale University, and Claremont McKenna College found that jurors rated DNA evidence 95 percent accurate and between 90 and 94 percent persuasive, depending on where the DNA was found.

DNA—GOLD STANDARD AND BULLETPROOF?

“People by and large are inordinately persuaded when they hear DNA because the public image is that DNA is the gold standard of identifying people, and it’s bulletproof,” says David Rossman, a School of Law professor and director of criminal law clinical programs. “There are a lot of places along the way where mistakes can creep in.”

How could this gold standard of forensic evidence become so tarnished? One reason is that our ability to detect DNA from a crime scene has outstripped our ability to make sense of it. Back 30 years ago, DNA science didn’t work well unless investigators were able to gather a lot of DNA from one person. But today’s tools enable investigators to swab more of the crime scene for genetic material—well beyond a bloody knife, to things like skin cells left on a computer keyboard or a doorknob.

“We have very sensitive techniques that give us these more complicated mixtures,” explains Robin Cotton, a MED associate professor and director of biomedical forensic sciences. “We need to be able to analyze this evidence. Otherwise, you just throw your hands up in the air and give up, which doesn’t do anybody any good.”

A black and white portrait of Catherine Grgicak. She has dark hair pulled back and is looking slightly to the right of the camera. She is wearing a dark blazer over a light-colored, vertically striped button-down shirt. A horizontal yellow bar is overlaid across the middle of the image, partially covering her face and the text. The text is positioned on the left side of this bar.

CATHERINE GRGICAK is developing software that could help crime labs identify the guilty without entangling the innocent.



GRGICAK says the first step in making sense of a DNA mixture is figuring out how many people contributed to it.

In fact, analyzing DNA mixtures has never been about achieving certainty. It's about partial matches, probabilities, big-time math, and a healthy dose of judgment calls by forensic scientists. Rossman sees an increasing potential for harm as the use of DNA forensics flourishes, thanks to diminishing costs and faster turnaround speeds. He points to the extensive use of DNA by the FBI, whose National DNA Index System (NDIS) already contains nearly 12 million offender profiles, more than 2 million arrestee profiles, and more than 600,000 forensic profiles.

To understand how DNA evidence can go wrong, it's helpful to start with what DNA fingerprinting actually entails. Forensic labs don't compare entire genomes. They examine tiny chunks of them, looking for commonalities at about 16 specific locations (the exact number varies depending on the kit used by the lab). At each location, there might be a few dozen possible genetic variations in the general population, and every person has two of them—one inherited from mom and one from dad. So, imagine that in DNA from a crime scene, each genetic location is a box containing Scrabble letter tiles representing variations. If each of these boxes contains just two letters, then the forensic scientist can assume the DNA is from just one person. They can compare that DNA fingerprint to the DNA from a suspect, knowing that it's almost impossible for two people (except for identical twins) to have perfect matches at every location.

But what if some of the boxes from our crime scene DNA contain not two letter tiles but six, while others contain

five, and a few contain seven? In this case, when the DNA is clearly from more than one person, forensic labs can no longer determine a match between the evidence and a suspect's DNA, but can only compute a likelihood ratio.

Typically, there are three basic conclusions a lab can make from DNA mixtures, depending on how many genetic variations a suspect's DNA and the crime scene evidence have in common: the suspect's DNA doesn't show up in the crime scene evidence; the suspect might have been at the crime scene based on commonalities between his DNA and the mixture; or the evidence is too complicated to analyze.

The odds of two people having a few genetic variations in common with DNA fingerprinting are pretty good. Imagine mixing Scrabble tiles for every letter of one person's first and last name in a hat. It's not hard to pull them out and match them to a single name. But add the tiles for two or three other people, and the number of names you can potentially spell skyrockets. So, with a DNA mixture, it's entirely possible to create false links between crime scene evidence and an innocent person who was nowhere near the crime.

This is where the two software programs that Grgicak and her team are developing come into play. One, called NOCI (NOC refers to number of contributors), uses statistical analysis to estimate the number of people whose DNA is part of the evidence—assigning a probability from one to five contributors. The other, called MATCHit, compares the DNA mixtures to the DNA from a suspect to compute a match statistic, known as a likelihood ratio, that this person contributed to the genetic mixture from the crime scene.

The team's goal is to combine both NOCI and MATCHit into a single tool. Grgicak hopes to do it by 2017.

VANISHINGLY SMALL PROBABILITY

During an interview in her office, Grgicak prints out two graphs showing analyzed DNA evidence from two mock crime scenes—the DNA is from real blood, but the blood is not from a crime. On the graphs, the variations at each genetic location (our hypothetical Scrabble tiles) show up as little spikes. In the evidence from a single DNA source, two spikes of nearly equal height poke up at distinct points for each of the 16 locations.

Two random strangers could easily share one or two of these spikes, but the probability of more than one person's DNA matching every spike is vanishingly small. However, the chance of a false identification grows substantially when the genetic evidence is from multiple people, as it is in the second piece of mock evidence. This graph shows a DNA mixture from 5 people, and each of the 16 locations has from 4 to 7 spikes of varying heights. Because people often share a few genetic variations, it's possible that some of the spikes represent DNA from more than one person. Plus, several low spikes suggest that at least one person contributed only a trace of genetic material to this evidence—possibly so little that his genetic markers at other locations weren't even detected by the test.

Grgicak points to one location with seven spikes. Maybe the first two spikes are from the same person, or maybe it's the first and the third, or the second and the fourth.

"It becomes a game of combinations," she says, and those combinations multiply quickly, especially when looking for a few shared genetic markers. Pretty soon, lots of innocent people could appear to be linked to the crime scene.

The first step to making sense of a DNA mixture, she explains, is to figure out how many people contributed to it. That number is the basis for nearly every other conclusion about the evidence. The old way to estimate it is to count the maximum number of spikes at any genetic location, divide by two, and round up. There were up to seven spikes in Grgicak's mock evidence DNA mixture, so a forensic scientist using the old formula would conclude that *at least four* people contributed to it.

"It's one thing to report that the *minimum* number of contributors is four, but it's another thing to use that number in the calculation of a match statistic," she says. Recall that there were actually five contributors.

The murkiness of mixture analysis compelled Grgicak and collaborators at Rutgers University and the Massachusetts Institute of Technology to spend years developing NOCI—computational algorithms that could sort through all the possible combinations of DNA spikes in a piece of evidence, taking into account their prevalence in the general population, to determine the likelihood that the genetic material came from one, two, three, four, or five people.

In testing using mock evidence, NOCI might conclude that one mixture is 99.9 percent likely to have two contributors, for instance. Or it might estimate a 35 percent likelihood of three contributors and a 65 percent likelihood

of four contributors. In these studies, Grgicak's team designates any probability over one percent as a possible answer to the number of DNA contributors.

In September 2014, the Department of Defense awarded Grgicak's lab a \$1.7 million contract to turn their NOCI prototype into something ready to be adopted by forensic labs nationwide.

The ultimate goal, of course, is to increase the certainty that a suspect's DNA is or isn't part of the crime scene evidence. To that end, in January 2015, the Department of Justice awarded Grgicak and her collaborators \$800,000 to develop MATCHit. The prototype is a bare-bones computer software program asking for the numbers that the algorithm will crunch, including the number of contributors, and how common every DNA variation is in the general population, according to a database such as the one compiled by the National Institute of Standards and Technology (NIST). In addition to generating a match statistic between the suspect and the crime scene evidence, the program also yields a common statistical measure called a p value to indicate how likely it is that a random person's DNA would have a match statistic as strong (or stronger) than the suspect's. The range of p values goes from zero to one. The closer it gets to zero, the more robust the match statistic becomes.

As with NOCI, the question with MATCHit is: where does a forensic lab draw the line in interpreting these probabilities? So far, Grgicak's research shows that the match statistics of noncontributors—innocent people—never have a p value below .01, no matter how complex the crime scene mixture. They have tested MATCHit using DNA mixtures of one, two, and three people (their goal is five), and so far, it's performed well. "We know, at least from our own early tests of MATCHit, that we have not falsely included individuals using that threshold," says Grgicak, "and that's the most important thing."

While Grgicak's software is being tested, other software programs have been in use for years. Michael Coble, a NIST forensic biologist, says he has seen an explosion of probabilistic genotyping software programs in the past four or five years. "They all work differently and they all have different assumptions, but they all try to give a statistical likelihood ratio," says Coble. "The question is, what does that likelihood ratio mean? You can give the same sample to 100 labs and get 100 answers."

Coble says guidelines or standards for both methodology and interpretation of data would be a major step forward, and researchers at NIST have been studying the pros and cons of some commonly used software programs. And while Coble is too unfamiliar with NOCI and MATCHit to suggest that Grgicak's work could pave the way to a national standard, he does see merit in the concept and progress in the software.

"I'm optimistic that moving forward some of these programs could be quite helpful," he says. "And I can see how it would be beneficial for the judicial system to have this kind of safeguard. It would cut down on a lot of the subjectivity and bias that people are beholden to at the moment." ■