

DNA COLD HIT CASES

MARCELLO DI BELLO

A suspect is typically identified by means of police investigations; this is the traditional, standard method. But more recently, identifications have also been carried out through database searches alone. Here is how it works: the investigators find some DNA traces at the crime scene, they create a DNA profile from the traces, and the resulting profile is then run through a database of existing profiles. If a DNA profile in the database shows a match, the individual associated with the matching profile becomes a suspect. Cases in which a suspect is identified through a database search are called cold-hit cases, and they are rising in number.

Cold-hit cases have triggered a controversy among legal scholars and forensic experts. Suppose the defendant is found to match and his DNA profile has an estimated frequency of 1 in 1 million; would it make any difference as to the significance of the match whether we are in a cold-hit case or in a standard case? There are two positions in the literature. The first asserts that in cold-hit cases a match has *less* significance; another asserts that it has the *same* significance. (Some even suggest that it has *more* significance.)

A parallelism can illustrate the claim that the match is less significant. In tossing a coin successively for ten times, a series of ten heads in a row is a quite rare and surprising event. It would be very surprising if, upon your very first attempt, you would get ten heads in a row. But now suppose you make 1,000 attempts, and on each attempt you toss the coin ten times successively. If among those 1,000 attempts you get at least one series of ten heads in a row, this result would not be so surprising any more. After all, getting ten heads in a row if you try enough times is practically certain. Now, what happens in a cold-hit case would seem to be very similar to what happens when you get ten heads in row after trying 1,000 times or more. A DNA profile is a very rare feature, so if you test only one individual and find a match, this is a very significant and surprising event. Yet, if you test 10 million people and find a match, the match becomes much less surprising. And this is what happens in cold-hit cases: you run the DNA profile against a database containing several million profiles, in the hope of finding a match. The analogy with the coin tossing case shows that, in cold-hit cases, the finding of a match is less surprising and thus less significant. To take this fact into account, the National Research Council suggests that a genetic profile's estimated frequency should be multiplied by the size of the database. This means that, for a database of size 1,000,000 and a profile frequency of $1/10,000,000$, the adjusted frequency will be $1/10$; this is a much less significant figure than the initial $1/10,000,000$. Note that if the database is of size n and the genetic frequency is $1/n$, the adjusted frequency will be 1; this essentially means that the DNA evidence match has no significance at all.

Let's now consider the claim that a match in a standard or cold-hit case has always the same significance. Here are two arguments for this claim. *First*, the significance of a match is determined by the frequency of the DNA profile for which the match is declared. If the frequency is the same in the cold-hit scenario and in the standard scenario, the match has the same significance. Similarly, the chance of getting a series of ten heads in a row does not change depending on whether you make one attempt or millions of attempts. For each attempt, the chance of getting ten heads in a row is very low—i.e. $0.5^{10} = 0.0009765625$. *Second*, if one had a world database containing the DNA profile of each individual, the finding of a (unique) match would certainly be extremely significant, contrary to the NRC approach.

(There are those who think that a match found in a cold-hit is more probative. The reason is that, since databases contain the profiles of previous offenders, the probability that a previ-

ous offender committed a crime is higher, hence a DNA match in a cold-hit case would be more probative than in standard cases.)

What is the reason for the disagreement among forensic experts and statisticians? There seem to be two probability values here. One is the probability of getting ten heads in a row on any given attempt; this probability does not change and it equals $0.5^{10} = 0.0009765625$. Another is the probability that, in a sufficiently larger number of attempts, we get ten heads in a row at least once. This probability reaches 1 as the number of attempts increases. In DNA evidence cases, the two probabilities would be: the probability of finding a genetic match on any given attempt (which equals the profile's estimated frequency); the probability of finding (at least) one match in a large database. Some have in mind the former probability and the NRC has in mind the latter probability. Recall that the NRC suggested to adjust the profile's frequency by the size of the database. Now, in a database of size 1,000,000, the probability of finding (at least) one individual with a profile having a frequency of $1/1,000,000$ equals one, as the NRC adjustment predicts.

As far as I can tell, in cold-hit cases, forensic experts are talking past each other. Some are interested in the probability of a match *on any given attempt to find a match*. Others are interested in the probability of a match *in a large database*. To be sure, forensic experts are genuinely disagreeing about which of the two probabilities is the more appropriate to assess the probative value of DNA evidence. The suggestion I want to make is that we cannot appreciate the probative value of DNA evidence in isolation, as merely an issue concerning the significance of a genetic match. Let me explain.

In standard cases, we have at least two sources of incriminating evidence: the police investigation that led to the identification of the suspect; and the DNA match. In cold-hit cases, instead, we lack the additional information coming from the police investigation. On the one hand, this difference should not affect the estimated profile's frequency, so those who think that there is no difference are partly right here. On the other, standard cases are different from cold-hit cases because in the former the police investigation was more thorough and extensive. The way to appreciate this difference is not, as the NRC does, by narrowly focusing on the significance of the genetic match in isolation from the rest of the incriminating evidence. In standard cases, the information collected by the police before the database search can contribute to support the prosecutor's overall case. In cold-hit cases, since the information coming from the police is missing, the overall prosecutor's case will be weaker. I conclude that, without focusing exclusively on the narrow question of a match's significance, the dispute among experts on cold-hit cases can disappear by having in sight the prosecutor's overall case.