

United States House of Representatives
Committee on the Judiciary
Subcommittee on Crime, Terrorism, and Homeland Security

Testimony of

Professor Henry T. Greely*

on H.R. 3361,

“Utilizing DNA Technology to Solve Cold Cases Act of 2011”

April 25, 2012

“The sins of the fathers are to be laid upon the children.”ⁱ

This biblical-sounding quotation is actually from *The Merchant of Venice* but what Shakespeare meant by it unclear, as he gives the line to the play’s fool. The Bible itself, at least in the King James version, does not use exactly this language, but in at least five places expresses similar sentiments about the Lord visiting the “iniquity” of the fathers on several generations of children.ⁱⁱ On the other hand, Ezekiel states “The son shall not bear the iniquity of the father, neither shall the father bear the iniquity of the son: the righteousness of the righteous shall be upon him, and the wickedness of the wicked shall be upon him.”ⁱⁱⁱ

In recent years new uses for forensic DNA matching have provoked similarly mixed reactions about the family connections and, perhaps not sin or iniquity, but crime. Our now “traditional,” but, in fact, less than 20 year old, forensic use of deoxyribonucleic acid (“DNA”) compares DNA profiles from crime scene DNA to either the profiles of particular suspects, or, through DNA databases, the profiles of people convicted of crimes – and, increasing, of people arrested for felonies or of non-U.S. nationals “detained” by the federal government. This method looks for a perfect or near-perfect match, indicating that the crime scene DNA almost certainly came from the suspect or from a person in the database (or from his identical twin). Family forensic DNA is a technique used when there is no perfect match, in the hope of generating investigative leads by seeing whether the crime scene DNA is likely to have come from a close genetic relative of a person in the database.

I was part of a group that published one of the first close analyses of family forensic DNA^{iv}, in 2006, and have continued to follow the issue. I believed then, and continue to believe now, that family forensic DNA, using our current technology, is a weak, inefficient, but occasionally useful method for generating investigative leads. I also believed, and continue to believe, that, although its use is disquieting, it raises no strong constitutional or other legal questions. It does raise a few policy problems, some, but not all, of which can be mitigated by regulating its use. Although it is not a panacea, the federal government should allow its careful use, but also should use the discussion of this technique to consider the future of forensic use of DNA. This bill, which combines a requirement that the Justice Department facilitate the technique’s use with discretion for the Attorney General to determine the proper ways to use it, is a good way to proceed.

I want to do five things in this testimony. First, I will explain how family forensic DNA works. Second, I will discuss its weaknesses as a law enforcement tool. Third, I will describe the possible legal and policy issues this tool raises and how they might (and might not) be mitigated. Fourth, I will discuss some possible ways to improve the effectiveness of the technique, though perhaps at the cost of exacerbating some of its problems. And, finally, I want to reflect on the trajectory of our use of forensic DNA and where that trajectory may eventually lead us.

How It Works

Each human has two complete human genomes, one inherited from his or her mother and one from the father. The information in each is contained in about 3.4 billion “base pairs” – molecules of adenine (A), cytosine (C), guanine (G), and thymine (T). Each A is paired with a T; each C is paired with G. Together, these base pairs form the “rungs” of the spiraling staircase that is DNA. Almost all this DNA is tucked away in the 46 chromosomes in the nuclei of our cells, 22 pairs of “autosomes,” cleverly named chromosomes 1 through 22, and two “sex” chromosomes, the X and Y-chromosomes. Men have one X chromosome, inherited from their mothers (who only have X chromosomes), and one Y chromosome, inherited from their fathers. Women have one X chromosome inherited from their mothers and a second X chromosome inherited from their fathers.

If we think of each base pair as a letter, the “book” that is each of our genomes is about 6.8 billion letters long. This is roughly the same length at *two* complete copies of F.2d – not of one volume, but from the first word of 1 F.2d through the end of 999 F.2d. The copy the human genome that each of us has is almost entirely identical to the copy found in any other human – we differ in only about one base pair in a thousand, so our genomes are roughly 99.9 percent identical. But, with 6.8 billion base pair, that 0.1 percent difference comes out to about 7 million differences.

Forensic DNA uses those differences to say that crime scene DNA “matches” the DNA of a particular suspect. The chances that two different people (who are not identical twins) would have exactly the same DNA are infinitesimal. But with 6.8 billion base pairs, where should we look for differences? In the mid-1990s, the FBI decided to focus its identification efforts on 13 particular locations in the genome. These locations, known as “loci”, are often referred to as the CODIS loci, because the FBI uses them in its Combined Operating DNA Information System (CODIS).

The FBI chose thirteen loci where our genome “stutters.” These are short tandem repeats, sometimes called satellite tandem repeats. A CODIS locus might, for example, consist of a stretch of chromosome 8 where a four base pair sequence, say ATTG, repeats itself. On some copies of chromosome 8, there might be seven repeats; on others, three repeats; and on still others, twelve repeats. These thirteen CODIS loci are all found on the autosomes (chromosomes 1 through 22), so each of us has two copies of the each of those chromosomes, and so two copies of each locus – one inherited from our mother and one from our father. On one locus, for example, I might have five repeats on one chromosome and eight on another. On another, I might have six repeats on one and eleven on the other. My CODIS profile is thirteen pairs of numbers, two for each of the thirteen loci, where each number represents the number of times a sequence of bases repeats.

Those thirteen pairs of numbers are my “identity code,” because the chances that any human being (other than my identical twin), alive today or at any time during our species existence, shares the same thirteen pairs of numbers are very close to zero. Assume, for present purposes, that each of the thirteen loci has ten different sets of repeat lengths

(called alleles), each of which is found in ten percent of chromosomes. The chance that, at any locus, I would share both of my alleles (repeat lengths) with anyone else is about two in one hundred. Two percent is not a *very* low probability – but now extend that from one locus to thirteen loci. Two in one hundred becomes roughly 8,000 in 100 septillion, or about one in 10 sextillion – one in 10,000,000,000,000,000,000.

The actual percentages are calculated in a more accurate and complicated way, but this approach leads to courtroom testimony that the chances that some DNA came from someone other than the defendant (or his identical twin) are one in many trillions or even quadrillions. This is the power of DNA for identification and courts (and police, prosecutors, and defense counsel) have been using it with confidence for over 15 years.

The FBI did not have to choose these particular CODIS markers. The United Kingdom, which has an older and (as a proportion of its population) bigger database, uses ten loci, only some of which are used by the FBI. The FBI was looking for loci that were easy to analyze, using the technology of the mid-1990s, and that had a lot of variation across all humans. Many other short tandem repeats could have been used, as well as many other kinds of variation in the genome, but the CODIS markers work perfectly well for identification. When crime scene DNA is analyzed for its CODIS markers, the resulting profile can be compared to the CODIS profiles of suspects, or, through a computerized search, with the CODIS profiles of the roughly 10 million people whose profiles are in the FBI's Offender Database. A perfect match means it is almost certain that the crime scene DNA came from the person with the same recorded CODIS profile.

The Offender Database contains the CODIS profiles that Congress has authorized the FBI to collect and include, both from the federal judicial system and from state systems. The boundaries of the CODIS system have changed over the years, but they now include profiles from people whose DNA is authorized by federal or state law to be collected and put into such a database. These may be people convicted of various crimes – at this point, all felonies and some misdemeanors – or people arrested for felonies, or non-U.S. nationals detained under federal government authority. All the profiles must include the CODIS markers and states submitting profiles to CODIS have to meet various requirements. As of February 2012, the Offender Database in the National DNA Index in CODIS contained over 10,560,300 profiles. The FBI reported that the database had assisted over 166,700 prosecutions during its existence. This assistance had been provided when a profile determined from crime scene DNA had been checked against the CODIS Offender Database and a match had been found.

But what happens when a match is *not* found? Is the database then useless?

Note that in all the above discussion, I have excepted identical twins. Identical twins have the same genomes and hence the same CODIS markers. They are a special case of family forensic DNA – if crime scene DNA matches perfectly the profile of someone in the Offender Database, but that person could not have been the perpetrator (because, for example, he was in prison at the time of the crime), but he had an identical twin, that match could implicate the twin.

Most of us do not have identical twins, but we all have or had parents and many of us have siblings or children. Our genetic first-degree relatives – parents, siblings, or children – do not share all of our genetic variations (unless they are identical twins) but, on average, they share half of them. Two people randomly chosen from the population will, on average, share eight to nine of the 26 CODIS alleles; two first-degree relatives will, on average, share 15 to 17 of them. This is because relatives get their variations from the same people. Two genetic brothers *must* have inherited their CODIS markers from among their parents' markers. If, for one marker, one parent had six and eight repeats and the other parent had three and eleven repeats, the siblings must have either a six or an eight or a three or an eleven. On average, at any given locus, they will have identical markers 25 percent of the time, they will share one marker 50 percent of the time, and they will share neither marker 25 percent of the time.

In fact, because their parents will sometimes have the same alleles – one parent has, say, five and seven repeats at one CODIS locus and the other has five and nine – siblings will, on average, share more than 13 alleles. In the European-American population, siblings will, on average, share both alleles at five CODIS loci, share one allele at seven CODIS loci, and share no alleles at one CODIS locus. Thus, on average, they will share 17 alleles.

The pattern for parent-child matches is a little different. Every child *must* have at least one allele from each genetic parent. If one compares the CODIS profile of a father and son, the son *must* have one of the father's alleles at each of the thirteen CODIS loci – because he got one of his two alleles at each locus from his father. Again, because the father and mother are likely to share some alleles, the actual average match between father and son will be more than 13 alleles. Among European-Americans, the average parent and child will match on about 15.7 alleles. This is fewer than the average siblings, but the parent-child pattern is distinctive; unlike the siblings, a parent-child pair *must* match at at least one of the two alleles at each locus.

This is the key to family forensic DNA. If crime scene DNA does not perfectly match the profiles of anyone in the Offender Database, it might match some of those profiles much more closely than one would expect. That might be a result of chance – or it might be the result of the crime scene coming from a close genetic relative of the person in the Offender Database. The close relatives of the person in the Offender Database could become leads, to be investigated to see if they might have been the source of the crime scene DNA. An interview might, for example, establish if the relative had a solid alibi or not. If enough evidence were collected to provide probable cause, the relative's DNA could be taken and directly tested to see if it matched the crime scene DNA. The partial, family match would no longer be relevant. The suspect's DNA profile either would or would not match the crime scene DNA profile; the family match would have only been the reason to investigate this person, it would not actually be evidence in court against him.

The British became using family forensic DNA as an investigative technique nearly a decade ago, with occasional success. At least two high profile American cases have used

variations on family forensic DNA. The Grim Sleeper case from Los Angeles is the purest example. The suspect was ultimately identified because the profile of the crime scene DNA bore a close resemblance to the DNA profile of his son, who was in the Offender Database as a result of his own run-ins with the law. The police interviewed the son, learned that his father had lived in the area of the crimes, and proceeded to investigate and ultimately arrest the father.

This use of family forensic DNA, the kind most commonly contemplated, basically asks the CODIS Offender Database, “are there people in the database whose DNA profiles indicate they are likely to be closely related to the person who left the crime scene DNA.” Unlike traditional CODIS searches, these will not turn up perfect matches, but only partial matches, but matches that are sufficiently good to raise an inference of a family relationship.

The term “partial match” needs to be used with care. “Partial match” has meaning in forensic DNA totally apart from family forensic DNA. In some cases, the crime scene DNA is degraded or damaged and not all 26 alleles can be derived from it. If, for example, only 20 alleles can be analyzed from the crime scene DNA and a suspect matches on those 20 alleles, this raises questions about what the odds are that the match is not coincidence. Family forensic DNA presupposes having all the alleles from the crime scene DNA but only having some match; “partial matching” has usually meant having only some of the alleles from the crime scene DNA but having all of those alleles matching. I believe there has been some confusion about whether state regulations governing partial matches were meant to apply to family forensic DNA.

The BTK case from Kansas provide a somewhat different example of using family forensic DNA. The police had plentiful crime scene DNA, but when they finally identified a suspect, they had no DNA from him. They got a court order to force a health clinic to provide a tissue sample from the suspect’s daughter. This was then checked to see if the crime scene DNA could have come from her father. When they concluded it could have, they got a DNA sample from the father, which matched the crime scene DNA. A guilty plea followed.

The Weakness of Family Forensic DNA

The biggest weakness of family forensic DNA is that, as an investigative technique, it is just not very good. It will almost always produce many false positives, people whose DNA profiles indicate that they *could* have family members who left the crime scene DNA but who did not. Additionally, the technique can produce false negatives, by not finding people whose close relatives actually did leave the crime scene DNA.

The false positive problem is large. Although, on average, parents and children will share about 15 to 16 alleles and unrelated people will share about 8.7 alleles, some unrelated people will share more than 8.7 alleles – some, in fact, will share more than 16 alleles. The larger the number of profiles in the database, the greater the chance of false positives.

Consider, for example, father-son matches. Each son *must* match his genetic father at at least one allele at each locus. What happens if one asks the CODIS system to identify everyone in the Offender Database who could be a parent (or child) of the source of the crime scene DNA – everyone who has at least one allele identical to the crime scene DNA at each of the 13 loci? In 2006, we calculated that the chance that crime scene DNA with an “average” set of alleles would be consistent with a parent-child relationship with a random profile from the Offender Database. We concluded that a DNA profile of average rarity would be a “parent-child” match to between 2,000 and 3,000 profiles in the Offender Database. When we made those calculations, the Offender Database had 2.75 million profiles; today it has over 10.5 million profiles. The average crime scene DNA should now produce 7 to 12 thousand possible “relatives.” All or all but one of them will be false positives. If the crime scene DNA has a particularly rare set of alleles, there may be no false positives; if it has the most common set of alleles, there may now be over 100,000 false positives. And as the Offender Database gets larger, these problems will only get worse.

Of course, one could cut down on false positives by tightening the requirement for a match. Instead of just requiring one match at every locus in order to raise suspicion of a parent/child match, one could require one match at every locus plus two matches at two or three loci. This is in line with the average number of matches expected, but it means that a true family match might be missed. If the parent (or child) of the actual source of the crime scene DNA is in the Offender Database, he might match at only 13 or 14 loci, not 15 or 16. The higher we set the bar, the fewer the false positives, but the greater the risk of false negatives.

This is even truer of sibling/sibling matches, as siblings do not have the same kind of minimum match as parents and children do. Two siblings could, in theory, match at every allele or match at none. If one set the standard for a possible sibling/sibling match so that it would have a false positive rate similar to that discussed above for parent/child matches, about 40 percent of actual sibling matches would be missed. If one set the false positive rate much lower, at, say, one in 100,000, leading to less than 100 false positives on average with today’s Offenders Database, one would miss about eighty percent of actual sibling/sibling matches.

As with most tests, there is an inevitable trade-off. The lower the rate of false positives, the higher rate of false negatives, and vice versa. But there is yet another problem with the accuracy of family forensic DNA. We have been talking about false negatives on the assumption that the source of crime scene DNA actually has a close relative in the Offender Database but that the comparison does not reveal the relationship. If the source of the crime scene DNA does *not* have a close relative in the Offender Database, the only positives that family forensic DNA could find would be false positives.

Family forensic DNA is, therefore, not a very good source of leads. It will usually throw up a vast number of possible suspects and, depending on where the line is drawn, it may well miss the actual perpetrator. It will almost always require substantial traditional police work to follow up the leads, work that, unlike a family search on CODIS, will eat up scarce police resources. It may be useful in high profile and difficult crimes, it may be

difficult in crimes where the crime scene DNA has a particularly set of variations, but it will not, at least as currently feasible, put a major dent into crime.

Issues with Family Forensic DNA

As set out in detail in our 2006 article, there seem to be no strong constitutional or other legal objections to the use of family forensic DNA. At first glance, it might seem to run afoul of the broad legal prohibition of “corruption of blood,” both in the Constitution (for the crime of treason) and in the constitutions and statutes of many states. But those prohibitions concern punishing innocent people for the crimes of their relatives, not of making people potential suspects based the crimes of their relatives.

If, in a line-up, the victim says “the mugger was not number 3, but he could have been his brother,” nothing prevents the police from investigating to see if “number 3” has a brother and his whereabouts at the time of the crime. Similarly, relatives of organized crime bosses are likely to be under increased suspicion of involvement in mob crimes. Family relationships are a clue that may properly lead to investigation. It feels “unseemly” to make someone a suspect based on the crimes of his relatives, but I see no good argument that it is unconstitutional, or even, in general, a bad idea.

This conclusion is particularly strong in the DNA context, where a false positive family connection can almost certainly *not* lead to a false conviction. Once a relative is identified, his DNA can be taken (voluntarily or, with probable cause, by legal action) and compared with the crime scene DNA. If he did not leave the crime scene DNA, no matter how closely the crime scene DNA matches that of his relative in the Offender Index, it cannot match his own DNA. The DNA evidence *must* exonerate the false positives.

The chance of false conviction, however, is not the only cost to being falsely identified as a suspect. Being interviewed by the police will often be a time-consuming and stressful experience, even for people who know they are innocent. The family suspect may not seriously risk false conviction, but neither will he be compensated for the time, anxiety, and possible embarrassment the investigation causes.

Three other issues deserve mention: the possible revelation of family secrets, possible unfairness to groups that are relatively genetically homogenous, and possible unfairness to groups that are disproportionately represented in the Offender Database.

Family forensic DNA is using possible family relationships to look for suspects. By looking at genetic evidence for family relationships, though, the technique could reveal facts about those relationships that are unwelcome, unknown, or both. These facts are most likely to involve so-called “false paternity” – the situation where a child’s genetic father is not, as a result of adoption, sperm donation, or other sexual partners, the man accepted as the genetic father. (The “preferred” term for this is “misattributed parentage,” but “false maternity” is, for understandable reasons, quite rare.)

It is easy to find geneticists who will say that in various genetic studies, five to ten percent of children have “unexpected” genetic fathers. There is a real dearth of actual

published evidence on the frequency of false paternity and some of the published evidence points to much lower rates. It does seem likely, though, that the rate is high enough to be non-trivial – and disconcerting to men who think they are genetic fathers.

It is possible that family forensic DNA could reveal cases of false paternity. If, for example, crime scene DNA is consistent with the suspect being the son of a particular person in the Offender Database, the “offender” could be asked about his children and his sons could then be questioned. If the interrogation included a DNA sample, its analysis might show definitively that there can be no parental link between the two men. Analysis of the CODIS markers could not rule out, definitively, the possibility that two people were siblings, but could make that result extremely unlikely.

This information might, or might not, already be known, or suspected, by those involved. If the investigators do not reveal it, it seems no concrete harm would be done, though people could still be understandably upset that the government learned either secret or previously unknown about their family connections. Alternatively, if the family members subsequently discovered the government had this information, they might complain that they were not told; there are, for example, some potential medical benefits to having an accurate understanding of one’s family history. If the investigators did reveal the information, though, the chances of disruptions of the family ties – and perhaps even of violence directed against the mother – seem quite real. There seems to be no investigative reason to disclose the results; if the son or brother presumably was ruled out as a suspect by the DNA analysis whether or not he was related to the person in the Offender Database. Prohibiting, or greatly limiting, the dissemination of this kind of family relationship information seems proper.

Second, some populations are more closely genetically homogenous than others. A small and relatively isolated Native American tribe or a group of immigrants from one community in, say, Southeast Asia, for example, is likely to have much closer family relationships, and hence much more genetic similarity, than, say, “European-Americans” or even “Irish Americans.” If the crime scene DNA came from a member of such a population, a higher percentage of people from that group who are in the Offender Database will be indicated as possibly related to the source of the crime scene DNA. Law enforcement should be aware of this bias in the method and treat the community sensitively.

Finally, and, to my mind, most importantly, the results of family forensic DNA searches of the CODIS Offender Database will be skewed in the same way that database is skewed. Most notably, African-Americans are convicted of felonies at roughly three times the rate of their roughly 13 percent share of the population. One can debate endlessly the reasons for this disproportion; for present purposes it is enough that it exists. The result is that, on average, a higher percentage of the African-American population is likely to be closely related to someone in the Offender Database than of most other American populations. The people identified as potential suspects by this method are therefore much likelier to be African-American than people randomly chosen from the population. This could be seen as unfair “special surveillance” of the African-

American population, and particularly of innocent members of the population whose only suspicious action is to share DNA variations with someone in the Offender Database.

At the same time, African-Americans are already likely to be suspects at a disproportionate rate, for whatever reasons lie behind the conviction disproportion. And much of the crime committed by African-Americans victimizes other African-Americans. Still, widespread use of family forensic DNA, with its vast number of false positives bringing under suspicion many innocent people, could well be seen by many African-Americans as another “racist” action by the American criminal justice system. Although these concerns about family forensic DNA do not seem to me to rise to the level of a possible constitutional violation, the public reaction could still be real and problematic.

Possible Improvements in the Effectiveness of Family Forensic DNA

The biggest problems of family forensic DNA stem from its inaccuracy. It is likely to be throw up so many possible family connections that its use will often impose costs, in police time and in the costs to innocent family members of being, even briefly, suspects, as to limit its use to only very unusual cases. These would be cases where the rarity of some of the alleles in the crime scene DNA greatly limits the number of “hits” or where the difficulty and importance of solving the crime justifies spending great resources. This inaccuracy can be combated, in ways both mundane and scientific, though these solutions raise their own problems.

One problem in implementing family forensic DNA is the need to find out whether someone in the Offender Database who is identified as a possible relative of the source of the crime scene DNA in fact has any relatives who might have been that source. This will typically involve finding and interviewing the “offender,” as well as hoping for his cooperation. This step could be eased if a computerized record existed of the relatives of everyone in the Offender Database. A simple questionnaire at time of conviction or arrest could provide such information and then a database search could quickly narrow down the possible family connections to only those with relatives – and could give priority to investigating the families of those who have relatives of the expected sex, age, and geographic location to have been involved in the crime. The problem is that it seems hard to justify asking a newly arrested or convicted person about his relatives and even harder to make a case that answering such questions should be required.

The technical approaches are more promising, but they, too, have problems. CODIS is just not a very good system for determining family relationships. With only 26 alleles, the chances are fairly good that some non-relatives will randomly match the crime scene DNA on enough alleles to signal a possible family relationship. That chance grows with the Offender Database. This is the fundamental cause of the vast number of false positives with this technique.

Using more alleles can make the process much more accurate. Our 2006 paper calculated that by adding 20 more loci similar to the existing CODIS markers, the chances of a false parent/child match would be about one in 200 million, reducing the number of false positives from hundreds or thousands to a handful or fewer.

California's implementation of family forensic DNA uses a similar expansion of alleles to narrow the number of false positives. It requires the authorities to check Y chromosome markers from the offender and the crime scene DNA and only authorizes proceeding to investigate the family match if the Y chromosome markers also match. Men inherit their Y chromosomes from their fathers. If two people have identical sets of markers on their Y chromosomes, they are very likely to share an ancestor in their paternal line. They might be father and son, brother and brother, or cousins who are both the sons of brothers. They may also be more distantly related, but the Y chromosome is sufficiently variable in human populations that exact Y chromosome matches will be rare. The existence of a Y chromosome match does not itself indicate guilt – innocent brothers will share the same Y chromosome – but use of Y chromosome matching will pare down the number of leads enormously, again reducing the number of false positives. This both improves the efficiency of the process for the police and cuts the number of innocent people who will be, however briefly, suspects.

The alleles examined on the Y chromosome share with the CODIS loci the virtue of having no known (or likely) medical or physical consequences. They seem to be so-called “junk” DNA, useful only for identification. One problem with the Y chromosome is that it is only found in men. Neither crime scene DNA from a woman nor the DNA of any women in the Offender Database could be checked against the Y chromosome. As over 90 percent of convicted felons are male, this is a concern, but not a huge one. And other parts of the genome that are similarly variable to the Y chromosome could be checked from women.

A bigger problem with using Y chromosome matching as part of family forensic DNA, though, is that the Y chromosome alleles have not been analyzed for the 10.5 million people already in the Offender Database. To do that analysis would require either re-analyzing the saved DNA sample the “offender” earlier provided – if it was saved – or acquiring a new sample. The costs of doing that for over ten million people, or even of finding many of them, would be quite high. On the other hand, one could do it a case at a time, seeking to analyze only the Y chromosomes of those “offenders” picked out by the family forensic analysis. This requires DNA from those “offenders” to be readily available or to be easy to re-acquire. It is hard to see a justification for forcing an offender to provide another DNA sample to investigate a crime that, as the result of the lack of an exact match, we *know* he cannot have committed. It might be possible to obtain a search warrant requiring a new DNA sample, but the more positive family matches there are, and, as a result, the lower the chance that any one of the offenders involved in those positive matches is actually related to the source of the crime scene DNA, the harder the case would seem for show probable cause.

One could also use other technical solutions. A common tool for genetics and genomics research, with some commercial uses, is the so-called “SNP chip.” This device allows the operator to determine, cheaply and quickly, which base (A, C, G, or T) a person carries at locations known as “single nucleotide polymorphisms” (“SNPs”), where substantial percentages of the population carry different bases. These SNP chips can quite easily examine hundreds of thousands or even millions of these SNPs. While the chance that two unrelated people might share 13 of the 26 CODIS alleles by chance is not

necessarily small, the chance that two unrelated people would share 300,000 out of 600,000 SNPs is vanishingly small. SNP chips could determine the existence of a wide range of relationship, not just first-degree relationships like parent/child or sib/sib, but uncle/nephew, cousin/cousin, and others. SNP chips could easily replace the CODIS loci entirely.

This solution, though, also has problems. It shares one with the Y chromosome tactic – it would require re-analyzing DNA from the entire 10.5 million person Offender Database in order to use it to search that database. But it has another problem. Unlike the CODIS loci or the commonly analyzed Y chromosome markers, many of these SNPs are associated with particular diseases or other genetic traits. Doing a SNP analysis for forensic purposes does raise all the privacy questions that are avoided when the genetic variations being used seem to be useful only for identification.

Improving the efficiency of family forensic DNA is both possible and, if the method is to be used at all, valuable both to police and to innocent potential suspects. If this bill passes, the Attorney General, in promulgating regulations, should give serious consideration to these ways to minimize false positives. But each of them poses serious challenges.

The Trajectory of Forensic DNA

I cannot leave this topic without noting the trajectory of forensic DNA use. Governments initially required DNA samples from people convicted of the most serious felonies, usually murder and sexual assaults. Then they began to require DNA samples from people convicted of less serious felonies or of serious misdemeanors or from juveniles found delinquent for reasons that would, had they been adults, been felonies or serious misdemeanors. More recently, first states and then the federal government required DNA samples from people charged with felonies, whether or not they were then, or ever, convicted. (The constitutionality of these statutes under the Fourth Amendment continues to be debated in federal and state courts across the country.) Federal legislation now authorizes the mandatory collection of DNA from non-U.S. persons “detained” under the government’s authority, whether or not charged with a felony or any crime. And just last month, the State of New York passed legislation requiring DNA samples from people convicted of most misdemeanors.

The trajectory has clearly been to collect more and more DNA from people with decreasingly serious involvement with the criminal justice system. Advocates have argued, and most judges have agreed, that people with those connections to the criminal justice system have forfeited some of their rights as a result of their convictions, arrests, or detainments. Family forensic DNA is a technique that uses the DNA provided under those statutes to extend the reach of forensic DNA to people who have not necessarily had *any* contact with the criminal justice system, let alone conviction or charge – people whose only link is that they are related to people who were convicted or arrested or detained. It is a logical and scientifically useful outgrowth of the earlier collections, not, I think, a planned consequence of those databases but a clever way to use them to solve more crimes, based on the reality that genetic variations run in families. This bill would

take that informal and almost accidental growth and give it the force of law, providing a legislative endorsement of the extension of forensic DNA to catch people who had no prior record of conviction, arrest, or detention.

This makes sense as a way to catch more criminals and its costs to the innocent are low. But if we really want to maximize the value of forensic DNA, why stop with (the usually innocent) first degree relatives of those convicted, arrested, or detained? The logical size for a forensic DNA database, at least once forensic DNA is cut loose from its mooring to an individual's involvement in the criminal justice system, is universal. A truly universal forensic DNA database would make family searching obsolete – the family members you might find would already be in the database. And it would also end the ways family searching discriminates against people whose family members were convicted, arrested, or detained.

In fact, an unplanned and impromptu version of such a universal database may be on its way. The cost of genomic analysis, and even of sequencing a person's entire genome, has been falling dramatically. The medical value of that information has been increasing steadily, although, unfortunately, not as dramatically. Within a decade scores, if not hundreds, of millions of Americans will have substantial genomic information in their clinical electronic health records, information that will be perfectly useful for identification – and that is only a court order away from the scrutiny of the government (or, in some cases, private litigants).

A universal DNA forensic database seems to me politically impossible today. At any time, such a database would be fraught with concerns about privacy and misuse. (I would note that restricting such a database to genetic information useful only for identification and not for any other purpose, unlike the information in medical records, would be a useful way to handle some of those concerns.) Whether such a database could be justified as a matter of policy would depend crucially on the protections that came with it. Whether a mandatory universal database could be justified constitutionally is another thorny question; my guess is that it would not be upheld as a mandate but might be upheld as a condition to participation in some governmental program for which definite identification is useful, like a driver's license, Social Security, or Medicare.

Foreign countries and political leaders have toyed with the idea of a universal DNA database, including the democratic government of Portugal and the former prime minister of the United Kingdom, Tony Blair. It is *not* a question for this subcommittee today, or, I suspect, any day soon. But endorsing the use of family forensic DNA and using DNA to make suspects of people with no prior personal connection to the criminal justice system takes us one logical step toward a universal DNA database – and would make the day when that discussion is necessary draw nearer.

Conclusion

I support H.R. 3361, although with reservations. It is not a panacea. It will not solve a large number of crimes, but it will solve some crimes, at some cost to the public in convenience, in privacy, and in their presumed innocence. If managed well by the

Attorney General's regulations, that (small) cost seems to me likely to be a cost that is likely to be outweighed by the technique's (also fairly small) benefits. The technique should be used responsibly and ways to improve it – for the benefit of both the police and the public – should be explored and debated. But this bill should also make us think about where we want the use of forensic DNA techniques and databases to go. If all the bill does is to spark a realistic discussion of that question, that alone may make it worthwhile.

* Deane F. and Kate Edelman Johnson Professor of Law; Professor, by courtesy, of Genetics; Director, Center for Law and the Biosciences; Stanford University. The views expressed in this testimony are not necessarily those of Stanford University – or anyone else. I would like to thank my co-authors from our 2006 paper, Drs. Joanna Mountain, Daniel Riordan, and Nanibáa Garrison.

ⁱ The Merchant of Venice, Act III, Scene 5, line 1.

ⁱⁱ See Exodus, 20:5, Exodus 32:7, Numbers 14:18, Deuteronomy 5:9, and Jeremiah 32:18.

ⁱⁱⁱ Ezekiel 18:20.

^{iv} Henry T. Greely, Daniel P. Riordan, Nanibaa' A. Garrison, Joanna L. Mountain, *Family Ties: The Use of DNA Offender Databases to Catch Offenders' Kin*, JOURNAL OF LAW, MEDICINE & ETHICS, 34:248-262 (Summer 2006). Much of the analysis in this testimony is drawn from that article, although my conclusions are not necessarily shared by my co-authors on that paper.

Other particularly useful articles on this topic include Frederick H. Bieber, Charles H. Brenner & David Lazer, *Finding Criminals Through DNA of Their Relatives*, 312 SCIENCE 1315-16 (2006); Sonia M. Suter, *All in the Family: Privacy and DNA Familial Searching*, 23 HARV. J. LAW & TECH. 309 (2010); Erin Murphy, *Relative Doubt: Familial Searches of DNA Databases*, 109 MICH. L. REV. 291 (2010); and Natalie Ram, *Fortuity and Forensic Familial Identification*, 63 STAN. L. REV. 751 (2011). Bieber, et al., is the other early discussion of the issue. Suter and Murphy take a more negative view of the technique than I do and are particularly worth reading; Ram provides some actual data about different state policies on the method.

Other published legal articles and notes on the topic include Lina Alexandra Hogan, *Note: Fourth Amendment – Guilt by Relation: If Your Brother Is Convicted of a Crime, You Too May Do Time*, 30 W. NEW ENG. L. REV. 543 (2008); Kimberly A. Wah, *Note and Comment: A New Investigative Lead: Familial Searching as an Effective Crime-Fighting Tool*, 29 WHITTIER L. REV. 909 (2008); Jules Epstein, “Genetic Surveillance” – *The Bogeyman Response to Familial DNA Investigations*, 2009 U. ILL. J. L. TECH. & POL’Y 141; Jessica D. Gabel, *Probable Cause from Probable Bonds: A Genetic Tattle Tale Based on Familial DNA*, 21 HASTINGS WOMEN’S L.J. 3 (2010); Brett Mares, *A Chip Off the Old Block: Familial DNA Searches and the African American Community*, 29 LAW & INEQ. 395 (2011); Amanda Paddock, *It's All Relative: Familial DNA Testing and the Fourth Amendment*, 12 MINN. J. L. SCI. & TECH. 851 (2011); Mary McCarthy, *Am I My Brother's Keeper?: Familial DNA Searches in the Twenty-First Century*, 86 NOTRE DAME L. REV. 381 (2011); and Jenny Choi, *California and the Future of Partial Match DNA Investigations*, 39 HASTINGS CONST. L.Q. 713 (2012).