“I get a sinking feeling when I hear a client has been fingered by a DNA test,” a defense lawyer recently told us. “Seems there’s not much I can do but negotiate a guilty plea.” Promoters of forensic DNA testing have done a good job selling the public, and even many criminal defense lawyers, on the idea that DNA tests provide a unique and infallible identification. DNA evidence has sent thousands of people to prison and, in recent years, has played a vital role in exonerating men who were falsely convicted. Even former critics of DNA testing, like Barry Scheck, are widely quoted attesting to the reliability of the DNA evidence in their cases. It is easy to assume that any past problems with DNA evidence have been worked out and that the tests are now unassailable.

BY WILLIAM C. THOMPSON, SIMON FORD, TRAVIS DOOM, MICHAEL RAYMER AND DAN KRANE
The problem with this assumption is that it ignores case-to-case variations in the nature and quality of DNA evidence. Although DNA technology has indeed improved since it was first used just 15 years ago, and the tests have the potential to produce powerful and convincing results, that potential is not realized in every case. Even when the reliability and admisissibility of the underlying test is well established, there is no guarantee that a test will produce reliable results every time it is used. In our experience there often are case-specific issues and problems that greatly affect the quality and relevance of DNA test results. In those situations, DNA evidence is far less probative than it might initially appear.

The criminal justice system presently does a poor job of distinguishing unassailably powerful DNA evidence from weak, misleading DNA evidence. The fault for that serious lapse lies partly with those defense lawyers who fail to evaluate the DNA evidence adequately in their cases. This article describes the steps that a defense lawyer should take in cases that turn on DNA evidence, you must first understand the table of alleles.

Our focus here is on the most widely used form of DNA testing, which examines genetic variants called short tandem repeats, or STR's. Our STRs. The Genetic Analyzer then measures their length and thereby determines which alleles are present.

By examining the DNA profiles, one can tell whether each suspect could or could not have been the source of the blood. Suspects 1, 2, and 4 are ruled out as possible sources because they have different alleles than the blood at one or more loci. However, Suspect 3 has exactly the same alleles at every locus, which indicates he could have been the source of the blood. In a case like this, the lab report will typically say that Suspects 1, 2, and 4 are “excluded” as possible sources of the blood, and that Suspect 3 “matches” or is “included” as a possible donor.

One of the loci analyzed is called amelogenin (Amel) and is used for typing the sex of a contributor to a sample. Males have X and Y versions of the alleles at that locus; females have only the X because they inherit two copies of the X chromosome. All of the profiles shown in Figure 1 appear to be of males.

Lab reports generally also contain estimates of the statistical frequency of the matching profiles in various reference populations (which are intended

---

**FIGURE 1: TABLE OF ALLELES**

Which suspect is a possible source of the blood? Only one of the four suspects has a DNA profile that matches the DNA profile observed in the blood sample.

<table>
<thead>
<tr>
<th>BloodSample</th>
<th>VWA</th>
<th>FGA</th>
<th>Amel</th>
<th>D8S1179</th>
<th>D21S11</th>
<th>D18S51</th>
<th>D5S818</th>
<th>D13S317</th>
<th>D7S820</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suspect 1</td>
<td>16,16</td>
<td>13,15</td>
<td>21,24</td>
<td>XY</td>
<td>12,13</td>
<td>27,30</td>
<td>15,14</td>
<td>10,11</td>
<td>9,12</td>
</tr>
<tr>
<td>Suspect 1</td>
<td>16,18</td>
<td>13,16</td>
<td>21,24</td>
<td>XY</td>
<td>12,14</td>
<td>27,28</td>
<td>15,17</td>
<td>11,12</td>
<td>8,11</td>
</tr>
<tr>
<td>Suspect 2</td>
<td>15,15</td>
<td>18,18</td>
<td>19,23,2</td>
<td>XY</td>
<td>13,15</td>
<td>29,30</td>
<td>17,17</td>
<td>11</td>
<td>8,9</td>
</tr>
<tr>
<td>Suspect 3</td>
<td>16,16</td>
<td>13,15</td>
<td>21,24</td>
<td>XY</td>
<td>12,13</td>
<td>27,30</td>
<td>15,14</td>
<td>10,11</td>
<td>9,12</td>
</tr>
<tr>
<td>Suspect 4</td>
<td>16,16</td>
<td>16,17</td>
<td>19,24</td>
<td>XY</td>
<td>14</td>
<td>30,30</td>
<td>12,16</td>
<td>10,11</td>
<td>9,10</td>
</tr>
</tbody>
</table>
to represent major racial and ethnic groups). Crime labs compute these estimates by determining the frequency of each allele in a sample population, and then compounding the individual frequencies by multiplying them together. If 10% (1 in 10) of Caucasian Americans are known to exhibit the 14 allele at the first locus (D3S1358) and 20% (1 in 5) are known to have the 15 allele, then the frequency of the pair of alleles would be estimated as $2 \times 0.10 \times 0.20 = 0.04$, or 4% among Caucasian Americans. The frequencies at each locus are simply multiplied together (sometimes with a minor modification meant to take into account the possibility of under-represented ethnic groups), producing frequency estimates for the overall profile that can be staggeringly small: often on the order of 1 in a billion to 1 in a quintillion, or even less. Needless to say, such evidence can be very impressive.

When the estimated frequency of the shared profile is very low, some labs will simply state “to a scientific certainty” that the samples sharing that profile are from the same person. For example, the FBI laboratory will claim two samples are from the same person if the estimated frequency of the shared profile among unrelated individuals is below one in 260 billion. Other labs use different cut off values for making identity claims. All of the cut-off values are arbitrary: there is no scientific reason for setting the cut off at any particular level just as there is no formally recognized way of being “scientifically certain” about anything. Moreover, these identity claims can be misleading because they imply that there could be no alternative explanation for the “match,” such as laboratory error, and they ignore the fact that close relatives are far more likely to have matching profiles than unrelated individuals. They can also be misleading in that the DNA tests themselves are powerless to provide any insight into the circumstances under which the sample was deposited and are generally unable to determine the type of tissue that was involved.

**Looking behind the lab report: Are the laboratory’s conclusions fully supported by the test results?**

Many defense lawyers simply accept lab reports at face value without looking behind them to see whether the actual test results fully support the laboratory’s conclusions. This can be a serious mistake.

In our experience, examination of the underlying laboratory data frequently reveals limitations or problems that would not be apparent from the laboratory report, such as inconsistencies between purportedly “matching” profiles, evidence of additional unreported contributors to evidentiary samples, errors in statistical computations and unreported problems with experimental controls that raise doubts about the validity of the results. Yet forensic DNA analysts tell us that they receive discovery requests from defense lawyers in only 10-15% of cases in which their tests incriminate a suspect.

Although current DNA tests rely heavily on computer-automated equipment, the interpretation of the results often requires subjective judgment. When faced with an ambiguous situation, where the call could go either way, crime lab analysts frequently slant their interpretations in ways that support prosecution theories.

Part of the problem is that forensic scientists refuse to take appropriate steps to “blind” themselves to the government’s expected (or desired) outcome when interpreting test results. We often see indications, in the laboratory notes themselves, that the analysts are familiar with facts of their cases, including information that has nothing to do with genetic testing, and that they are acutely aware of which results will help or hurt the prosecution team. A DNA analyst in one case wrote:

“Suspect-known crip gang member — keeps ‘skating’ on charges—never serves time. This robbery he gets hit in head with bar stool — left blood trail. [Detective] Miller wants to connect this...
In another case, where the defense lawyer had suggested that another individual besides the defendant had been involved in the crime, and might have left DNA, the DNA laboratory notes include the notation: “Death penalty case. Need to eliminate [other individual] as a possible suspect.”

It is well known that people tend to see what they expect (and desire) to see when they evaluate ambiguous data. This tendency can cause analysts to unintentionally slant their interpretations in a manner consistent with prosecution theories of the case. Furthermore, some analysts appear to rely on non-genetic evidence to help them interpret DNA test results. When one of us questioned an analyst’s interpretation of a problematic case, the analyst defended her position by saying: “I know I am right — they found the victim’s purse in [the defendant’s] apartment.” Backwards reasoning of this type (i.e., “we know the defendant is guilty, so the DNA evidence must be incriminating”) is another factor that can cause analysts to slant their reports in a manner that supports police theories of the case. Hence, it is vital that defense counsel look behind the laboratory report to determine whether the DNA to scene w/DNA ...

Figure 3 shows the results for three loci that were labeled with blue dye. The position of the peaks on the graph (how far left or right) indicates how long it took the allele to pass through the capillary, which indicates the length of the underlying DNA fragment. Backwards reasoning of this type (i.e., “we know the defendant is guilty, so the DNA evidence must be incriminating”) is another factor that can cause analysts to slant their reports in a manner that supports police theories of the case. Hence, it is vital that defense counsel look behind the laboratory report to determine whether the DNA evidence must be incriminating.

ProfilerPlus™ uses “primers” to identify the relevant STR-DNA segments and then “amplifies” (replicates) these segments using a process called polymerase chain reaction (PCR). Each locus is “labeled” with a colored dye (either blue, yellow or green). The Genetic Analyzer measures the length of the DNA segments by using an electrical current to impel them through a narrow capillary tube, wherein the shorter fragments move more quickly than the longer fragments. Under laser light, the colored dyes produce florescent light, signaling the presence of DNA. A computer-operated camera detects the light as the fragments reach the end of the capillary. The “peaks” on the electropherogram record these flashes of light. Based on the color of the light, and the time it took the DNA to pass through the capillary, a series of computer programs determines which alleles are present at each locus.

Figure 2 show the results for three loci that were labeled with blue dye. The position of the peaks on the graph (how far left or right) indicates how long it took the allele to pass through the capillary, which indicates the length of the underlying DNA fragment. From this, the computer program infers which allele is represented and generates the appropriate label.

The height of the peaks corresponds to the quantity of DNA present. The unit of measurement for peak heights is the RFU, or “relative fluorescent unit,” which reflects the intensity of the fluorescent light detected by the computer-operated camera. Peaks representing alleles from the same person are expected to have roughly the same heights measured in RFUs throughout a given sample, although peak height imbalances occasionally occur.
lab's conclusions are well supported, and whether there is more to the story than the report tells.

Behind the Table of Alleles Detected (Figure 1) is a set of computer-generated graphs called electropherograms that display the test results. When evaluating STR evidence, a defense lawyer should always examine the electropherograms because they sometimes reveal unreported ambiguities and, fairly frequently, evidence of additional, unknown contributors. The electropherograms shown in Figure 2 display the results for the crime scene blood and four suspects discussed above at three of the ten loci summarized in Figure 1.

The “peaks” in the electropherograms indicate the presence of human DNA. The peaks on the left side of the graphs represent alleles at locus D3S1358; those in the center represent alleles at locus vWA; and those on the right represent alleles at locus FGA. The numbers under each peak are computer-generated labels that indicate which allele each peak represents and how high the peak is relative to the baseline.

By examining the electropherograms in Figure 2, one can readily see that the computerized system detected two alleles in the blood from the crime scene at locus D3S1358. These are alleles 14 and 15, which are reported in the Table of Alleles (Figure 1). The other alleles reported in the allele chart (Figure 1) can also be seen. Our initial examination of these electropherograms reveals no obvious problems of interpretation in this case.

However, other cases are not so clear cut. Consider the electropherogram in Figure 3, which shows the DNA test results that purportedly “matched” a defendant to a saliva
sample taken from the breast of an alleges sexual assault victim. Although the laboratory report stated that the same alleles were found in both samples at these three loci, close examination of the electropherograms supports a significantly different conclusion. There are two additional “peaks” in the saliva sample that the laboratory failed to report — a peak labeled “12” (indicating allele 12) at locus D3S1358, and a peak labeled “OL Allele” (indicating a possible “off-ladder,” or unclassified, allele) at locus FGA. The laboratory decided to ignore these two peaks and never mentioned them in its report. A defense lawyer who failed to examine the underlying test results would never have known about them. However, they clearly complicate the interpretation of the evidence — raising the possibility, for example, that the DNA on the breast swab is from a person with alleles 12 and 17 at locus D3S1358, rather than just allele 17, which would exclude the defendant as a possible contributor.

Sources of ambiguity in STR interpretation

A number of factors can introduce ambiguity into STR evidence, leaving the results open to alternative interpretations. To competently represent an individual incriminated by DNA evidence, defense counsel must uncover these ambiguities, when they exist, understand their implications, and explain them to the trier-of-fact.

Mixtures. One of the most common complications in the analysis of DNA evidence is the presence of DNA from multiple sources. A sample that contains DNA from two or more individuals is referred to as a mixture. A single person is expected to contribute at most two alleles for each locus. If more than two peaks are visible at any locus, there is strong reason to believe that the sample is a mixture.

By their very nature mixtures are difficult to interpret. The number of contributors is often unclear. Although the presence of three or more alleles at any locus signals the presence of more than one contributor, it often is difficult to tell whether the sample originated from two, three, or even more individuals because the various contributors may share many alleles. If alleles 14, 15 and 18 are observed at a locus, they could be from two individuals, A and B, where A contributed 15 and B contributed 14, 18. Alternatively, A could have contributed 14, 15 while B contributed 15, 18, and so on. There might also be three contributors. For example A could have contributed 14, 15, while B contributed 15, 18 and C contributed 15. Many other combinations are also consistent with the findings. A study of one database of 649 individuals found over 5 million three-way combinations of individuals that would have shown four or fewer alleles across all 12 commonly tested STR loci.5

Some laboratories try to determine which alleles go with which contributor based on peak heights. They assume that the taller peaks (which generally indicate larger quantities of DNA at the start of the analysis) are associated with a “primary” contributor and the shorter peaks with a “secondary” contributor. In Figure 4, for example, a laboratory analyst might conclude that allele 15 in the left locus, and alleles 10 and 13 in the right locus are associated with a primary contributor while alleles 14 and 18 in the left locus, and allele 12 in the right locus are associated with a secondary contributor. But these inferences are often problematic because a variety of factors, other than the quantity of DNA present, can affect peak height. Moreover, labs are often inconsistent in the way they make such inferences, treating peak heights as a reliable indicator of DNA quantity when doing so supports the government’s case, and treating them as unreliable when it does not.

These interpretive ambiguities make it difficult, and sometimes impossible, to estimate the statistical likelihood that a randomly chosen individual will be “included” (or, could not be “excluded”) as a possible contributor to a mixed sample. Defense lawyers should look carefully at the way in which laboratories compute statistical estimates in mixture cases because these estimates often are based on debatable assumptions that are unfavorable to the defendant.

Degradation. As samples age, DNA like any chemical begins to break down (or degrade). This process occurs slowly if the samples are carefully preserved but can occur rapidly when the samples are exposed for even a short time to unfavorable conditions, such as warmth, moisture or sunlight.

Degradation skews the relationship between peak heights and the quantity of DNA present. Generally, degradation produces a downward slope across the electropherograms in the height of peaks because degradation is more likely to interfere with the detection of longer sequences of repeated DNA (the alleles on the right side of the electro-
Degraded samples can be difficult to type. The process of degradation can reduce the height of some peaks, making them too low to be distinguished reliably from background “noise” in the data, or making them disappear entirely, while other peaks from the same sample can still be scored. In mixed samples, it may be impossible to determine whether the alleles of one or more contributors have become undetectable at some loci. Often analysts simply guess whether all alleles have been detected or not, which renders their conclusions speculative and leaves the results open to a variety of alternative interpretations. Further, the two or more biological samples that make up a mixture may show different levels of degradation, perhaps due to their having been deposited at different times or due to differences in the protection offered by different cell types. Such possibilities make the interpretation of degraded mixed sample particularly prone to subjective (unscientific) interpretation.

Allelic Dropout. In some instances, an STR test will detect only one of the two alleles from a particular contributor at a particular locus. Generally this occurs when the quantity of DNA is relatively low, either because the sample is limited or because the DNA it contains is degraded, and hence the test is near its threshold of sensitivity. The potential for allelic dropout complicates the process of interpretation because analysts must decide whether a mismatch between two profiles reflects a true genetic difference or simply the failure of the test to detect all of the alleles in one of the samples.

Figure 6 shows three additional loci from the case shown in Figure 3, in which a defendant’s profile was “matched” to the profile of a saliva sample from a woman’s breast. The laboratory reported that the DNA profile of the saliva sample shown in Figure 6 was consistent with the defendant’s profile, despite the absence of the defendant’s 10 allele at locus D13S317 because it assumed that the 10 allele had “dropped out.” However, the occurrence of “allelic dropout” is cannot be independently verified — the only evidence that this phenomenon occurred is the “inconsistency” that it purports to explain.

Obviously, there is another possible interpretation that is more favorable for this defendant — i.e., that police arrested the wrong man.

Spurious Peaks. An additional complication in STR interpretation is that electropherograms often exhibit spurious peaks that do not indicate the presence of DNA. These extra peaks are referred to as “technical artifacts” and are produced by unavoidable imperfections of the DNA analysis process. The most common artifacts are stutter, noise and pull-up.

Stutter peaks are small peaks that occur immediately before (and, less frequently, after) a real peak. Stutter occurs as a by-product of the process used to amplify DNA from evidence samples. In samples known to be from a single source, stutter is identifiable by its size and position. However, it is sometimes difficult to distinguish stutter bands from a secondary contributor in samples that contain (or might contain) DNA from more than one person.

“Noise” is the term used to describe small background peaks that occur along the baseline in all samples. A wide variety of factors (including air bubbles,
urea crystals, and sample contamination) can create small random flashes that occasionally may be large enough to be confused with an actual peak or to mask actual peaks.

Pull-up (sometimes referred to as bleed-through) represents a failure of the analysis software to discriminate between the different dye colors used during the generation of the test results. A signal from a locus labeled with blue dye, for example, might mistakenly be interpreted as a yellow or green signal, thereby creating false peaks at the yellow or green loci. Pull-up can usually be identified through careful analysis of the position of peaks across the color spectrum, but there is a danger that pull-up will go unrecognized, particularly when the result it produces is consistent with what the analyst expected or wanted to find.

Although many technical artifacts are clearly identifiable, standards for determining whether a peak is a true peak or a technical artifact are often rather subjective, leaving room for disagreement among experts. Furthermore, analysts often appear inconsistent across cases in how they apply interpretive
standards — accepting that a signal is a “true peak” more readily when it is consistent with the expected result than when it is not. Hence, these interpretations need to be examined carefully.

Spikes, blobs and other false peaks. A number of different technical phenomena can affect genetic analyzers, causing spurious results called “artifacts” to appear in the electropherograms. Spikes are narrow peaks usually attributed to fluctuation in voltage or the presence of minute air bubbles in the capillary. Spikes are usually seen in the same position in all four colors. Blobs are false peaks thought to arise when some colored dye becomes detached from the DNA and gets picked up by the detector. Blobs are usually wider than real peaks and are typically only seen in one color. The “OL Allele” shown in Figure 8 below may be a blob.

Spikes and blobs are not reproducible, which means that if the sample is run through the genetic analyzer again these artifacts should not re-appear in the same place. Hence, the correct way to confirm that a questionable peak is an artifact is to rerun the sample. However, analysts, to save time, often simply rely on their “professional experience” to decide which results are spurious and which are real. This practice can be problematic because no generally accepted objective criteria have yet been established to discriminate between artifacts and real peaks (other than retesting).

Threshold Issues: Short Peaks, “Weak” Alleles. When the quantity of DNA being analyzed is very low (as indicated by low peak-heights) the genetic analyzer may fail to detect the entire profile of a contributor. Furthermore, it may be difficult to distinguish true low-level peaks from technical artifacts. Consequently, most forensic laboratories have established peak-height thresholds for “scoring” alleles. Only if the peak-height (expressed in RFU) exceeds a standard value will it be counted.

There are no generally accepted thresholds for how high peaks must be to qualify as a “true allele.” Applied Biosystems, Inc., which sells the most widely used system for STR typing (the ABI Prism 310 Genetic Analyzer™ with the ProfilerPlus™ system) recommends a peak-height threshold of 150 RFU, saying that peaks below this level must be
interpreted with caution. However, many crime laboratories that use the ABI system have set lower thresholds (down to 40 RFU in some instances). And crime laboratories sometimes apply their standards in an inconsistent manner from case to case or even within a single case. Hence, a defendant may be convicted in one case based on “peaks” that would not be counted in another case, or by another lab. And in some cases there may be unreported peaks, just below the threshold, that would change the interpretation of the case if considered.

Finding and evaluating low-level peaks can be difficult because labs can set their analytic software to ignore peaks below a specified level and can print out electropherograms in a manner that fails to identify low-level alleles. The best way to assess low-level alleles is to obtain copies of the electronic data files produced by the genetic analyzer and have them re-analyzed by an expert who has access to the analytic software.

Figure 9 shows electropherograms from a rape/homicide case. The defendant admitted having intercourse with the victim, but contended another man had subsequently raped and killed her. The crime lab reported finding only the defendant’s profile in vaginal samples from the victim; the lab report stated that the second man was “excluded” as a possible source of the semen collected from the victim’s body. However, a review of the electronic data by a defense expert revealed low-level alleles (peaks) consistent with those of the second man, which significantly helped the defense case. Notice how these low-level alleles are obscured in the upper electropherogram, while the Y-axis. These low peaks are revealed in the lower electropherogram, where the defense expert set the software to ignore peaks below the threshold, that would change the interpretation of the case if considered.

William C. Thompson, J.D., Ph.D., is a professor in the Department of Criminology, Law & Society at the University of California, Irvine and a member of the California Bar. He has been a member of the NACDL DNA Strikeforce and has represented a number of defendants in cases that turned on DNA evidence.

William C. Thompson
Department of Criminology, Law & Society
University of California, Irvine
949-824-6156
Fax: 561-679-0890
Email: wcthomps@uci.edu

Simon Ford, Ph.D. was trained in molecular biology and biochemistry. He is President of Lexigen Science and Law Consultants, a firm that specializes in providing advice to lawyers about genetic evidence. He frequently consults on DNA cases and has conducted workshops for a number of agencies on computer analysis of STR test results.

Simon Ford
Lexigen Science & Law Consultants, Inc.
2261 Market St, Box 302
San Francisco, CA 94114
(415) 865-0600
Email: sford@lexigen.com