When Good DNA Goes Bad

Mark W. Perlin*

Cybergenetics Corp, 160 North Craig Street, Suite 210, Pittsburgh, PA 15213, USA

Abstract

DNA evidence is the forensic gold standard. However, the interpretation of this evidence can be challenging. Sophisticated mathematical computing can provide accurate and reliable interpretation of DNA mixtures that contain two or more individuals. But the reliability of human review of such data is less well established.

This paper explores what happens when good DNA data is badly interpreted. The genotype is used as a unifying concept, with mixture evidence containing multiple genotypes. Highly informative computer interpretation of mixtures is compared with less quantitative human approximation.

Seven cases are examined where, on the same data, human review gave appreciably different results than the computer's more thorough assessment. These interpretation differences can affect criminal justice outcomes, suggesting that sophisticated computing is needed to help people interpret challenging DNA mixture evidence.

Keywords: DNA; Mixture; Interpretation; Computer; Genotype information; Match; TrueAllele; CPI; Probability

Introduction

Some may think that DNA evidence is always perfect. As the forensic gold standard, maybe DNA mistakes are never made. With all the information right in front of us, perhaps we can immediately know the correct answer just by looking at the data. This situation is largely true for pristine reference samples, with their very small human error rate of 1 locus in 2,000 [1]. However, much DNA evidence is mixed or low-level, and so presents interpretation challenges [2].

Last year, the Washington Post reported that a state was reevaluating their DNA evidence in 375 cases [3]. The issue was DNA mixtures, evidence that contains more than one person. That article quoted a crime lab director who said, "Mixture cases are their own little nightmare. It gets really tricky in a hurry". The forensic DNA pioneer Dr. Peter Gill, of the former United Kingdom Forensic Science Service, has said, "If you show 10 colleagues a mixture, you will probably end up with 10 different answers" [4].

This paper explores when good DNA goes bad. We begin by describing the unifying genotype concept, and mixture evidence comprised of multiple genotypes. Highly informative computer interpretation of mixtures is then contrasted with less quantitative human approximation. We then examine DNA interpretation discrepancies in seven cases where both computer and human analyses were performed. We conclude that sophisticated computing is needed to help people interpret challenging DNA mixture evidence.

Methods

Genotype

A human genome packages DNA into 23 pairs of chromosomes. We can unravel the genetic alphabet from a chromosome's hundred million letters to zoom in to a small region of 100 to 500 letters. This genetic location, or "locus," could describe meaningful information like a functional gene, or, alternatively, its DNA letters could code for something more random. There are about 100,000 Short Tandem Repeat (STR) loci scattered across the genome, providing highly variable "junk DNA" useful for identifying people.

We generally inherit two DNA copies of an STR locus, one from our mother and one from our father. Each copy has a small word (say, 4 or 5 DNA letters) that is repeated some number (say, 10 or 20) of times. A larger number means a longer DNA sentence. Across a human population, there are about 10 or 20 different values of these varying sentences, called "alleles".

The genetic type (or "genotype") is the central concept in genetic identification [5]. At a locus, a person's genotype is their two inherited alleles. Each locus provides about 100 potential pairings, which is useful in genetic identity, where we want to distinguish between different people. Examining a person's genotype allele pair at 10 or 20 different loci can provide quintillions of different possibilities. A person's genotype is easy to measure in the molecular biology lab [6], and is relatively unique.

Mixture

The diagram illustrates the components of DNA mixture data, and its underlying genotypes (Figure 1). At this D18S51 locus, there are three people in the mixture. The x-axis shows increasing allele size, while the y-axis shows increasing allele amount. The major contributor (blue) contributes most of the DNA (taller bars), and has a genotype with allele pair 13,14. The second person (green) contributes less DNA (about a third of the height of the major), with a 16,18 allele pair. The far smaller amount (orange) of a third contributor has a 17,20 genotype. This schematic shows the true state of the world–the genotypes of each contributor, and their relative amounts of DNA.

Actual STR mixture data (Figure 2) can look very similar to the true genotype illustration. Again, the x-axis measures allele size, and the y-axis allele quantity. The tallest allele peaks 13,14 (blue) are from

the first contributor. The medium-sized peaks 16,18 (green) are from the second contributor. The shortest peaks 17,20 (orange) are from the third contributor. This is high quality DNA data that can be highly informative when properly interpreted.

Human review

Unfortunately, such mixture data often goes largely unused in human review because it is not fully interpreted. Forensic protocols apply a "threshold" to the data that throws out quantitative information (Figure 3). Instead of using a mathematical model to explain the data, an analyst simply draws a line, whose height is set by their lab [7]. Below this threshold line, STR data is discarded. Above the threshold, peak heights are ignored and all alleles are treated as if they had the same DNA amount.

When data are not used, information can be lost. Quantitative data interpretation in this example supports a match to a person having a 16,18 genotype (green arrows). Yet once a threshold has been applied, human review would exclude this genotype as a contributor to the mixture.

Science makes inferences (such as genotypes) from observations. Data should be a fixed starting point that is never modified [8]. But DNA mixture protocols for human review typically apply a threshold that changes the STR peak height data. Some labs now use two thresholds [9], but that does not address the main issue: data should be used, not discarded.

Computer interpretation

Computer programs can conduct an objective and thorough examination of DNA mixture data. Cybergenetics TrueAllele® Casework is one such system, with validated reliability in sensitivity, specificity and reproducibility studies [10-12]. Originally developed over ten years ago, TrueAllele is now regularly used in criminal cases. TrueAllele mathematically models DNA mixture data, statistically examining virtually every possible solution.

The illustration shows one such proposed genotype solution (Figure 4). The computer has proposed a first major contributor with genotype 13,14 (blue), a lesser contributor having allele pair 16,18 (green), and a small amount of a 17,20 genotype contributor (orange). TrueAllele tries that genotype configuration, along with many others. Taller allele bars indicate that more DNA is present for a genotype. Small bars indicate a lesser contributor amount.

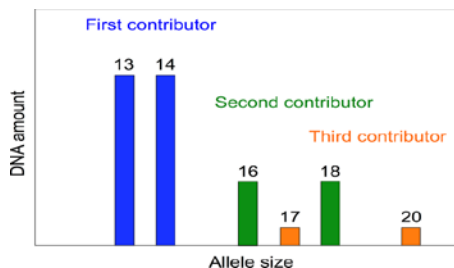The computer explores many possibilities. It proposes different

Figure 2: *Good DNA evidence data.* The electrophoretogram (EPG) data trace for this three contributor mixture shows one data peak for each allele. The x-axis gives allele size (in base pairs), while the y-axis measures peak heights on a relative fluorescence unit (RFU) scale. The allele pair for each genotype (blue, green, orange) has peaks with comparable heights. Each contributor corresponds to a genotype shown in Figure 1.

Figure 3: *Misinterpreted, data goes bad.* A threshold (red line) of height 150 RFU is applied to the quantitative STR locus data of Figure 2. This threshold renders the 13,14 genotype into equal all-or-none events, and eliminates from the data all trace of the other two genotypes.

Figure 4: *TrueAllele solves mixtures.* Quantitative computer modeling can explain the observed data. Better explanations of the data confer higher probabilities to their underlying genotypes. In the explanation shown, an allele pair is represented by two colored rectangles (blue, green, orange). The placement and heights of the allele pairs explain the observed data peaks well, and so these genotypes have a high probability.

Figure 1: *DNA mixture genotypes.* The data schematic shows DNA amount (y-axis) versus allele size (x-axis) at an STR locus. Each contributor is shown in a different color (blue, green, orange). The allele pair of each contributor genotype is indicated by size location (x-axis), while the relative genotype amount is indicated by bar height (y-axis).
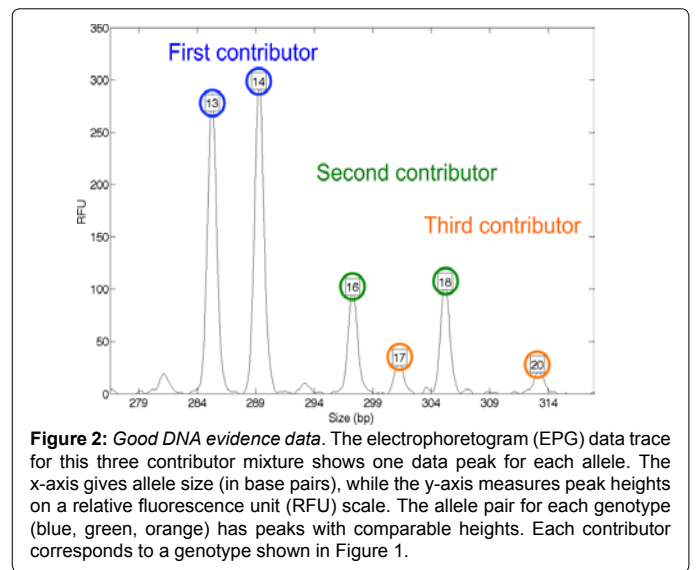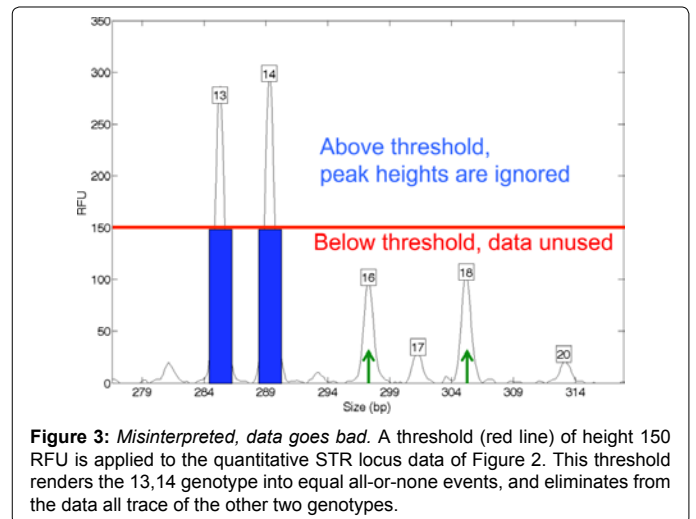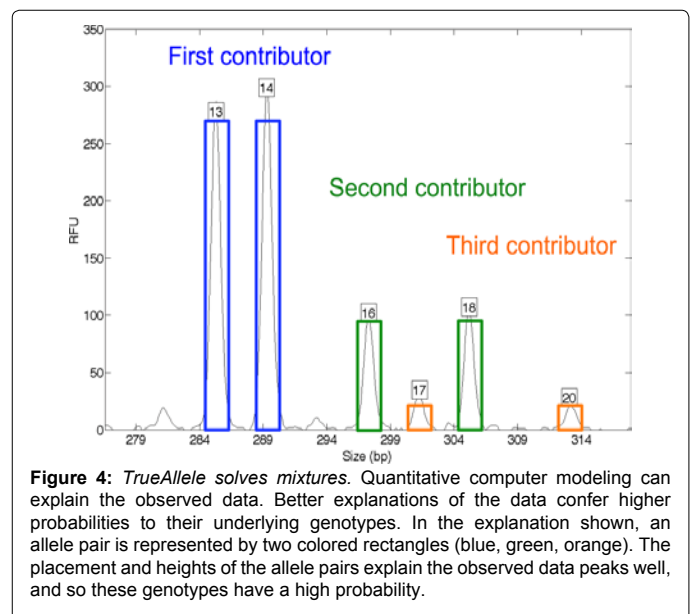
quantities of contributor DNA (bar height). It proposes alleles of different sizes (bar locations). Consider a first contributor genotype value of 5,6. Since there are no data peaks at those alleles, this hypothesis does not explain the data very well. When a proposed pattern has peak heights that do not fit the data, its proposed genotypes are assigned a low probability. Considering many thousands of possibilities through statistical search, those genotypes that better explain the data will receive higher probability, and poorer explanations of the data will get lower probability.

In nature, there really is a true set of genotypes present in a DNA sample. But all we can know from observing data (through our models) is some inference about the possible states of the world, and their relative probabilities. A computer can consider a hundred thousand different scenarios, each one involving a hundred variables. A person cannot be that thorough.

Match

The TrueAllele computer objectively infers a genotype directly from the data, without using a known comparison reference. This approach eliminates examination bias [13]. The computer records an evidence genotype as a probability distribution. Afterwards, the computer can then make an objective comparison to a known reference genotype obtained from a suspect or some other person.

A DNA match statistic describes the strength of association between evidence and reference, relative to coincidence. The number is calculated from three genotypes–evidence, reference and population. The result is a likelihood ratio (LR), and can be stated as: "a match between the DNA evidence and a reference individual is (some number of times) more probable than coincidence" [14].

The logarithm (approximately the number of zeros after the first digit) of a LR is a standard measure of information. The log (LR) is also called the "weight of evidence" [15]. We use the logarithm of the match statistic to compare the information obtained by human and computer DNA interpretation methods.

Results

We now examine some DNA mixture cases, and compare computer and human interpretation results. These are all criminal cases on which I have reported or testified. Some of the details have been changed for anonymity.

Commonwealth v. Lyons

The evidence in this Pennsylvania homicide was an abandoned sweatshirt. The victim had been stabbed 27 times when she was trying to break off her relationship with parolee Glenn Lyons. The crime lab found a match statistic in the tens of thousands, or 4 zeros after the first digit. The prosecutor asked Cybergenetics to re-analyze the DNA mixture data. The TrueAllele computer produced a match statistic in the trillions, or 12 zeros.

The human match statistic was a million-fold lower than the more accurate computer result. This was partly due to DNA degradation. Degraded mixtures are harder for people to interpret because the mixture proportions change at every locus. However, a computer can mathematically model those locus mixture changes, and produce an accurate result.

We summarize the human–computer information comparison with a first point on a scatterplot (Figure 5, case 1). The x-axis gives

accurate identification information, as reproducibly measured by a validated statistical computer system. The y-axis gives the reported human estimate. For the Lyons case, we plot a point (12,4) with "12" (number of zeros in a trillion) for the true information x-coordinate, and "4" (the number of zeros in ten thousand) for the y-coordinate. Since match information is the logarithm (i.e. the number of zeros) of a match statistic, this log-log scatterplot provides an information comparison.

The combined probability of inclusion (CPI) is the most prevalent human mixture interpretation method [16]. But CPI can only give evidence in favor of an inclusion (or say nothing at all). Therefore, the y-axis for human review has only positive numbers–CPI cannot speak against an inclusion. Computer and other likelihood ratio methods [17], however, will quantify evidence with either positive or negative information.

Commonwealth v. Foley

Dr. John Yelenic was found in his living room lying in his own blood. He had been viciously slashed to death, with his blood splattered over unsigned divorce papers. Prior to the homicide, Dr. Yelenic's estranged wife Michelle had moved in with her boyfriend, Pennsylvania State Trooper Kevin Foley. There was virtually no physical evidence in this case, except for some trace DNA under the victim's fingernails. This Southwestern Pennsylvania murder has been documented in books and videos [18].

Foley was the first case where sophisticated statistical computing was ever used to thoroughly assess a DNA mixture, with results introduced in court as criminal evidence. My recent book chapter about this landmark computer interpretation of DNA evidence spans the crime, pretrial hearing, admissibility, trial, conviction, sentencing, appeal, and appellate precedent [19].

The DNA under Dr. Yelenic's fingernails was a mixture that was 93% from the victim, plus 7% of an unknown contributor. The lab inclusion statistic was in the tens of thousands, or 4 zeros, and was thus not the strongest of evidence. On the same STR data, the TrueAllele computer separated out the two genotypes, and found a match of the minor 7% contributor to Kevin Foley in the hundreds of billions, or 11 zeros. Trooper Foley was found guilty, and sentenced to life in prison.

We add a second point (11, 4) to the log-log information plot for the Foley DNA mixture (Figure 5, case 2). The blue 45° "iso-information"
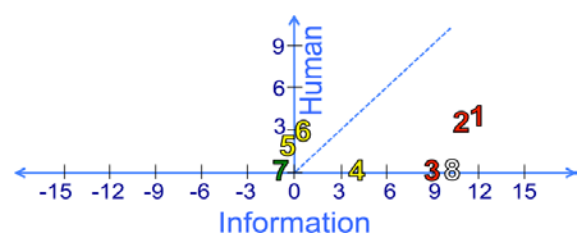

Figure 5: *TrueAllele vs. human interpretation.* This scatterplot compares human match results (y-axis) versus the true computer-determined match information (x-axis). Both axes express the logarithm (i.e., order of magnitude) of a match statistic. Each number refers to a case described in the Results section. The cases are color coded as homicide (red), rape (yellow), weapon (green) and exculpatory (white). The 45° blue dashed iso-information line is where the scatterplot points would reside if human and computer methods produced the same DNA match statistic.

line shows where human and computer match statistics would equal each other. Below that line, towards the bottom right, human review has lost DNA match information.

People may give significantly lower match statistics when discarding quantitative information. Since STR data peak heights are not equal, thresholds can discard considerable information about genotype quantity. While some may think reporting a lower match statistic is "conservative", we shall soon see that this is not always true [20].

The Queen v. Shivers

The 2009 Massereene Barracks attack disrupted the ten year Northern Ireland peace process. Hooded Real IRA gunmen emerged from a car in front of an army barracks, where young unarmed troops (who were headed for Afghanistan the next day) were collecting a pizza delivery. The gunmen shot at everything in sight, killing two young soldiers, Mark Quinsey and Patrick Azimkar.

Forensic science provided all the evidence [21]. One item was a mobile phone found in the burned out getaway car. The phone was used to talk about the attack right after it happened. Brian Shivers' DNA was left on that phone.

Human review of this DNA was inconclusive. That is, there was no match information, or 0 zeros. The computer produced a match statistic of a billion, or 9 zeros. The human method's interpretation failure was due to the very small amount of DNA in the mixture. In 2012, Mr. Shivers was convicted and sentenced to 25 years in prison. The third scatterplot point (9, 0) lies well below the iso-information line, showing that human review can lose considerable DNA identification information (Figure 5, case 3).

State v. Diggins

The fourth case is from the Southern United States. Mr. Booker Diggins was sentenced to life in prison for a brutal rape. Two years ago, the Innocence Project in New York heralded ABO blood typing results that had not been brought out at trial as "bulletproof evidence" that Mr. Diggins could not have possibly committed the crime. Defense negotiations began with the district attorney for his release from prison.

The state then found a mislaid rape kit in an attic. The defense wanted to analyze this new DNA evidence in order to help find the true perpetrator, so they sent the kit off to a private testing laboratory. The state lab retained some of the evidence (e.g. a tampon string), but it was not interpretable by human review. So the lab sent their STR data to Cybergenetics for computer analysis.

Despite the small amount of DNA, TrueAllele found a match to Mr. Diggins in the tens of thousands, or 4 zeros. He was not released from prison. A later defense report on the DNA rape kit evidence agreed with Cybergenetics TrueAllele findings. Mr. Diggins is still serving his life sentence. This was an "almost innocence" case, though not quite. In the scatterplot, we see again how at point (4, 0) the computer reported a match statistic, but people could not (Figure 5, case 4).

Commonwealth v. Doe

This Eastern United States rape case has a mixture of three people on a glove. The defendant clearly contributed to the mixture, as did a victim from another crime. The lab thought that the victim from this rape was also included, based on human review of the fuzzy mixture data. Their inclusion statistic was very low though, around 100, or 2

zeros. However, the computer-inferred information did not show statistical support for a match to the rape victim.

The fifth (0, 2) point on the log-log scatterplot has a computer x-value of 0 (no information), but a human y-value of 2 (log of 100). CPI gives all references the same match statistic–100 in this case. The new point falls on the left side of the iso-information line. By not making use of all the data, human review might incorrectly include the victim, which would falsely implicate the defendant (Figure 5, case 5).
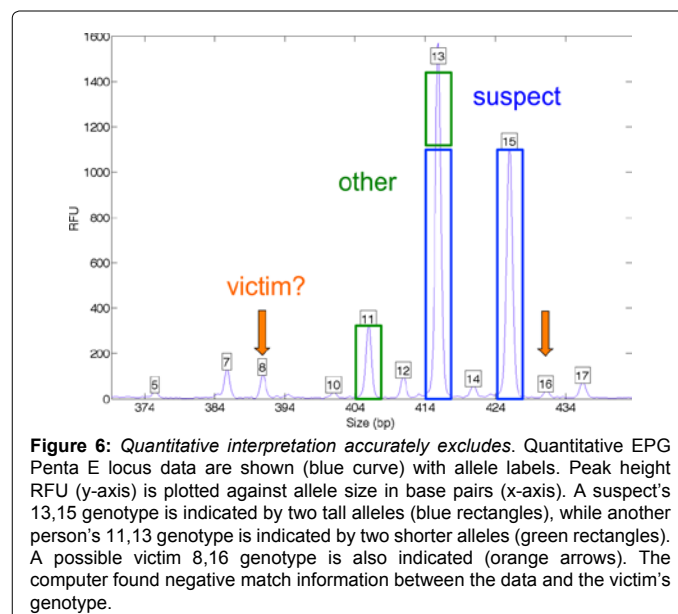
With threshold-based interpretation methods, only loci that support a conclusion are reported. Reasons can be given (e. g. allele drop out), but data weighing against an inclusion are not used [20]. In this case, the favorable loci only added up to a CPI statistic of a hundred.

The quantitative DNA mixture data (Figure 6) shows how information can be lost. The STR peak height versus allele length EPG data at this Penta E locus is shown (blue curve). The suspect's major genotype highlights the tall allele pair 13,15 (blue rectangles). The other victim (from a different crime) has a minor genotype that is highlighted in the shorter 11,13 allele pair (green rectangles). The victim of this crime has genotype 8,16 (orange arrows).

The computer evaluated all triple genotype combinations and weightings, not just the one hypothesis shown in Figure 6. TrueAllele considered the quantitative evidence data, proposing patterns generated from virtually all possible genotypes, contributor quantities and other variables. The computer worked on this problem for over a day, testing a hundred thousand possibilities.

At this locus, TrueAllele found dozens of allele pair possibilities for a third contributor genotype, most of them quite unlikely. The victim's 8,16 value had very little probability in the evidence genotype. TrueAllele determined that the negative information in the evidence argued *against* a DNA match to the victim.

The lab used a simpler approach, drawing a threshold line across the locus data to form allele lists. Peaks that fell under the line were discarded, while those over the line were treated as if they all shared

Figure 6: *Quantitative interpretation accurately excludes.* Quantitative EPG Penta E locus data are shown (blue curve) with allele labels. Peak height RFU (y-axis) is plotted against allele size in base pairs (x-axis). A suspect's 13,15 genotype is indicated by two tall alleles (blue rectangles), while another person's 11,13 genotype is indicated by two shorter alleles (green rectangles). A possible victim 8,16 genotype is also indicated (orange arrows). The computer found negative match information between the data and the victim's genotype.

the same peak height. Ultimately, this locus was not used in the CPI statistic, since the absence of victim alleles was attributed to drop out. Human review included 13 of the 15 loci, and assigned a CPI statistic of 100 (2 zeros) to every one of the three references.

TrueAllele instead separated the mixture into its component genotypes. Using all 15 loci, the computer found a match to the suspect of a quintillion (18 zeros), and a match to the victim of another crime also at a quintillion (18 zeros). However, TrueAllele found no statistical support for a match to the victim of this crime. As a result, the prosecutor dropped charges against the defendant in this case.

Commonwealth v. Brown

In this sixth case, a single Y-chromosome was found on an alleged rape victim's underwear, quite likely through adventitious transfer. The rape kit, and all other DNA evidence, was negative for the defendant. All five possible male suspects in the house that night shared the same Y chromosome, since they all descended from the same paternal grandfather.

The Y-chromosome DNA could not distinguish between any of the likely suspects. Therefore, the DNA did not actually identify the defendant. There was no probative information (x=0), but the human-reported Y-STR match statistic was in the thousands (y=3). We record this (0, 3) point on the log-log plot (Figure 5, case 6). Following a pretrial hearing, the judge precluded this irrelevant DNA evidence from testimony.

State v. Smith

A man was accused of aiming his gun at five policemen, not shooting at them, and then fleeing and "dropping" the gun 35 feet away. He is currently incarcerated. Seeking post-conviction relief, his defense attorney asked Cybergenetics to assess the DNA mixture data from the handgun.

TrueAllele inferred genotypes from the data, none of which matched the convict. The computer found no statistical support for a match; the log (LR) value was -1. Human review was also uninformative for a match (zero information). This finding adds the seventh point (-1, 0) to the log-log scatterplot (Figure 5, case 7).

From the handgun, one of the nonmatching genotypes had an expected LR value of a billion, or 9 zeros. This potentially exculpatory evidence, not reported by the state lab, adds the eighth point (9, 0) to our log-log scatterplot (Figure 5, case 8). Could a DNA database be used to find the true person who was holding the gun?

Unfortunately not, since good data has gone bad. Government DNA index systems, designed for reference samples, do not work well with mixtures. At best, just 20% of mixtures (the easiest ones) are uploaded to a United States DNA index system. These government DNA databases use allele lists that rob evidence of probative power, instead of using probabilistic genotypes that can preserve DNA identification information [11]. Government rules block mixture evidence upload, in order to avoid false positive errors [22] that might misidentify people in their database systems.

A better solution is to use a genotype DNA database [22]. Then identification information from DNA evidence can be preserved as a probabilistic genotype, and compared with offender profiles to produce a true LR match statistic. Government agencies may eventually adopt information solutions that can better identify criminals and reduce crime. Such information-preserving DNA mixture databases are currently available from the private sector. For example, Cybergenetics TrueAllele genotype database was used in the World Trade Center mass disaster for identifying victim remains [23,24].

Conclusion

Good DNA mixture data can help make identifications. From this data, computers can determine accurate match information [25]. TrueAllele validation studies have been conducted on both synthesized data and casework mixture samples, comparing computer and human review. TrueAllele computer mixture interpretation is thorough, objective, sensitive, specific, and reproducible; human review methods are not.

In the cases discussed in this paper, approximate human statistics did not correlate well with more exact computer solutions. The scatterplot points did not fall along the iso-information line (Figure 5). Each case has its own reasons for this disconnect between the actual data information and less accurate human results.

Scientific truth does not depend on whether it is presented by the prosecution or the defense. Courts may want to more fully account for current discrepancies in DNA mixture data interpretation. Computers such as TrueAllele are needed to help with challenging mixture evidence [26], so that good DNA data does not go bad.

References

1. Kadash K, Kozlowski BE, Biega LA, Duceman BW (2004) Validation study of the TrueAllele® automated data review system. J Forensic Sci 49: 1-8.

2. Buckleton JS, Triggs CM, Walsh SJ (2004) Forensic DNA Evidence Interpretation. Boca Raton, FL: CRC Press.

3. Flaherty MP (2011) Virginia reevaluates DNA evidence in 375 cases. Washington Post. Washington, DC.

4. Gill P (2005) Bridging the gap between science and law. 1st International Human Identification E-Symposium (HumID 2005). Forensic Institute.

5. Perlin MW (2012) DNA Identification Science. In: Wecht CH, editor. Forensic Sciences. Albany, NY: LexisNexis® Matthew Bender®. pp. Chapter 37C.

6. Butler JM (2005) Forensic DNA Typing: Biology, Technology, and Genetics of STR Markers. New York: Academic Press.

7. SWGDAM (2000) Short Tandem Repeat (STR) interpretation guidelines (Scientific Working Group on DNA Analysis Methods). Forensic Sci Commun (FBI) 2.

8. Gelman A, Carlin JB, Stern HS, Rubin D (1995) Bayesian Data Analysis. Boca Raton, FL: Chapman & Hall/CRC.

9. SWGDAM (2010) Interpretation guidelines for autosomal STR typing by forensic DNA testing laboratories.

10. Perlin MW, Belrose JL, Duceman BW (2013) New York State TrueAllele® Casework validation study. J Forensic Sciences 58 (6): in press, 2013.

11. Perlin MW, Legler MM, Spencer CE, Smith JL, Allan WP, et al. (2011) Validating TrueAllele® DNA mixture interpretation. J Forensic Sci 56: 1430-1447.

12. Perlin MW, Sinelnikov A (2009) An information gap in DNA evidence

interpretation. PLoS ONE 4: e8327.

13. Thompson WC (2009) Painting the target around the matching profile: the Texas sharpshooter fallacy in forensic DNA interpretation. Law, Probability and Risk 8: 257-276.

14. Perlin MW (2010) Explaining the likelihood ratio in DNA mixture interpretation. Promega's Twenty First International Symposium on Human Identification. San Antonio, TX.

15. Good IJ (1950) Probability and the Weighing of Evidence. London: Griffin.

16. Devlin B (1993) Forensic inference from genetic markers. Stat Methods Med Res 2: 241-262.

17. Aitken CG, Taroni F (2004) Statistics and the Evaluation of Evidence for Forensic Scientists. Chicester, UK: John Wiley & Sons.

18. Smith C (2011) Dying for Love: The True Story of a Millionaire Dentist, his Unfaithful Wife, and the Affair that Ended in Murder. New York, NY: St. Martin's True Crime.

19. Perlin MW (2012) The Blairsville slaying and the dawn of DNA computing. In: Niapas A, editor. Death Needs Answers: The Cold-Blooded Murder of Dr John Yelenic. New Kensington, PA: Grelin Press.

20. Curran JM, Buckleton J (2010) Inclusion probabilities and dropout. J Forensic Sci 55: 1171-1173.

21. Perlin MW, Galloway J (2012) Computer DNA evidence interpretation in the Real IRA Massereene terrorist attack. Evidence Technology Magazine 10: 20-23.

22. Perlin MW (2012) Investigative DNA databases that preserve identification information. American Academy of Forensic Sciences 64th Annual Meeting. Atlanta, GA: AAFS.

23. Perlin MW (2007) Mass casualty identification through DNA analysis: overview, problems and pitfalls. In: Okoye MI, Wecht CH, editors. Forensic Investigation and Management of Mass Disasters. Tucson, AZ: Lawyers & Judges Publishing Co. pp. 23-30.

24. Perlin MW (2007) Identifying human remains using TrueAllele® technology. In: Okoye MI, Wecht CH, editors. Forensic Investigation and Management of Mass Disasters. Tucson, AZ: Lawyers & Judges Publishing Co. pp. 31-38.

25. Perlin MW (2012) Forensic science in the information age. Forensic Magazine 9: 17-21.

26. Perlin MW (2012) Easy reporting of hard DNA: computer comfort in the courtroom. Forensic Magazine 9: 32-37.