The DNA Investigator™

Cybergenetics Newsletter Spring 2010

Validating DNA Mixture Interpretation Methods

Feature Article

Validating Interpretation Using Match Information

Every DNA laboratory wants to use reliable methods that can withstand scientific and legal scrutiny. Therefore, crime labs expend considerable resources validating every part of their DNA data generation process – extraction, quantification, amplification and detection. These validations ensure that they know the applicability and limitations of their DNA methods. The validation results are expressed using statistical measures of efficacy and reproducibility.

What about the interpretation of their DNA data to infer genotypes and produce match statistics? With unambiguous STR data, all interpretation methods do agree on a single definite genotype solution, and so proficiency testing suffices. But this certainty is not the case with complex DNA evidence, such as DNA mixtures with more than one contributor, or low amounts of DNA that might be damaged or degraded.

Uncertain DNA data leads to uncertain genotypes, where more than one allele pair possibility can account for the evidence. Different DNA mixture interpretation methods form different lists of these allele pair possibilities, and may assign them different probability values. Even using the same interpretation method, different analysts can infer different genotypes on the same data. How can we even compare these divergent solutions in order to validate our genotyping methods?

Fortunately, there is a standard statistical measure that reduces a complex genotype representation into a single number – the *match information*. And match information numbers can be easily compared to assess method reliability. Indeed, the match information is precisely what the police and courts demand as consumers of DNA evidence. It tells how much identification information is present in the DNA, and how well the evidence implicates or exonerates a suspect.

All DNA match statistics are "likelihood ratios" (LR), an information measure that gives the odds of identification after we have the DNA data relative to the identification odds before. Scientists use the "powers of 10" (or, logarithm) to report information. For example, a match score of a million (or 10^6) has an information value of 6 – the number of zeros in a million. To compute an inferred genotype's LR, we need to have on hand a contributor genotype and a reference population.

How *effective* is a DNA mixture interpretation method? Efficacy can be determined by looking at a set of mixture case items (actual or mock), and observing the average or range of match information. The *improvement* of one interpretation method over another can be measured by the differences in match information. For example, suppose that a newer method infers a genotype with match information of 12 (a trillion, or 10^{12}), but on the same DNA data an older method finds only 9 information units (a trillion, or 10^9). Then the new method gives an information improvement of 3 (10^{12-9} is 10^3, or a thousand).

How *reproducible* is an interpretation method? With ambiguous DNA data, two independent genotype interpretations are often not identical. However, their log(LR) information values are usually close. By calculating how close the duplicate interpretations are on a set of case items, we can find the "within-case" standard deviation that statistically measures reproducibility.

A validated DNA mixture interpretation method is a powerful asset. Its efficacy and reproducibility can be thoroughly understood, and introduced as supporting evidence in court. Objective comparisons can be made with other interpretation methods to learn which are most useful. These validation studies can tell us where weaker methods have less applicability, and where stronger methods may be essential for a just DNA identification.

www.cybgen.com

Testifying Tip
Presenting a Validated TrueAllele Match Result
Page 2

Case Study
TrueAllele Validation on 16 Case Mixture Items
Page 3

Tutorial
How to Validate Your Own Mixture Interpretation Method
Page 4

Free
TrueAllele DNA Case Assessment Details on
Page 4

Case Study

TrueAllele Validation on 16 Case Mixture Items

In collaboration with a well-regarded crime laboratory, we conducted a validation study of TrueAllele® mixture interpretation [Perlin & Duceman, 2010]. We collected 41 retrospective cases containing 368 items of evidence. We processed all of these items in duplicate using the TrueAllele system. We found complete concordance with human review for the single source samples, and so turned to the 86 mixture items. In this brief note, we describe the efficacy and reproducibility of computer interpretation of DNA mixtures.

On 8 two-person mixture items without a victim reference, using a suspect reference only for comparison, the lab reported a combined probability of inclusion (CPI) match score. We show the log(LR) match scores of the duplicate computer runs (blue, green) together with the log(CPI) of human review (orange) on the same item data (Figure 1).

We see that TrueAllele two unknown interpretation was more informative than CPI on the same data in every case. The computer (Table 1, two unknown column) had an average *efficacy* of $10^{13.26}$ (ten trillion) on these items, whereas human review (Table 1) averaged $10^{7.03}$ (ten million). The average per item information *improvement* was $10^{6.24}$ (= $10^{13.26}/10^{7.03}$, or one million). Duplicate computer runs (Table 1) showed a small within-group standard deviation 0.175, establishing *reproducibility*. We conclude that TrueAllele computer inference of two unknown genotypes was at least as reliable as CPI human review.

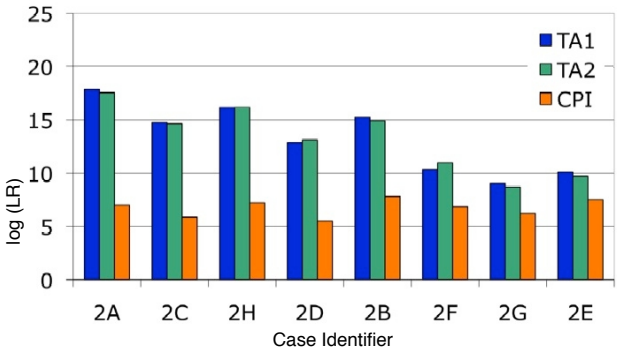

Figure 1. Match results of two unknown contributor cases.

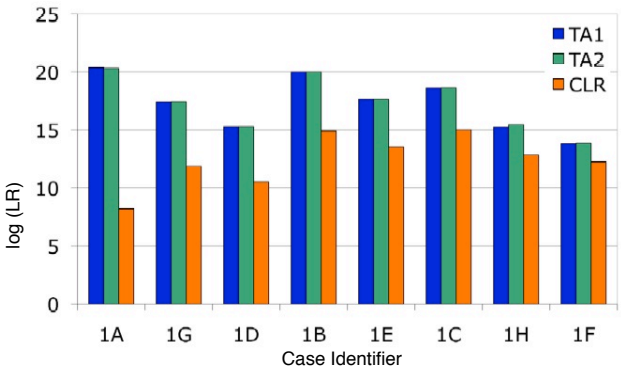

Figure 2. Match results of one unknown contributor cases.

How can we intuitively understand the magnitude of these results? Consider the improvement hypothesis that TrueAllele is more informative than human review. We can summarize our study data as our increased belief in this improvement having seen the 16 comparison results. The 8 independent two unknown inference items showed a total improvement of 8 x 6.24, or 50 log information units. The 8 one unknown experiments showed a total improvement of 8 x 4.67 equals 37 log information units.

Together, the likelihood ratio information gain in our improvement hypothesis was the product of these 16 independent results was 87 log information units, or 10^{87}. This vast number exceeds the number of electrons in the known universe (10^{80}). In other words, there is more evidence favoring an improvement by TrueAllele computer interpretation over human review than there is evidence for the existence of matter. We conclude that quantitative TrueAllele mixture interpretation is more informative than human review of the same DNA mixture evidence.

On 8 two-person mixture items having a victim reference, and a suspect reference for comparison, the lab reported a combined likelihood ratio (CLR) match score. We show the log(LR) match scores of the duplicate computer runs (blue, green) together with the log(CLR) of human review (orange) on the same mixture data (Figure 2).

We see that TrueAllele one unknown interpretation using a victim reference was more informative than CLR on the same data in every case. The computer (Table 1, one unknown column) had an average *efficacy* of $10^{17.33}$ (hundred quintillion) on these items, whereas human review (Table 1) averaged $10^{12.66}$ (trillion). This LR ratio represents a per-item information *improvement* of $10^{4.67}$ (=$10^{17.33}/10^{12.66}$, or fifty thousand). Duplicate computer runs (Table 1) showed a very small within-group standard deviation of 0.036, establishing *reproducibility*. We conclude that TrueAllele computer inference of one unknown genotypes on DNA mixture data using a victim reference was at least as reliable as CLR human review.

interpretation method	two unknown (without victim)	one unknown (with victim)
quantitative computer	13.26 (0.175) (ten trillion)	17.33 (0.036) (hundred quadrillion)
qualitative human	7.03 (ten million)	12.66 (five trillion)
improvement	6.24 (one million)	4.67 (fifty thousand)

Table 1. *Match information comparison*.

Testifying Tip

Presenting a Validated TrueAllele Match Result

A validated DNA interpretation method can be most desirable when testifying in court. Validations are also useful in a pretrial admissibility hearing. We discuss both settings.

A pretrial Daubert hearing seeks to establish the testability and error rate of a scientific method, while a Frye hearing focuses on the use of generally accepted methods. A validation study for your DNA interpretation method (which is distinct from your laboratory data validation) can be very helpful here. The efficacy and reproducibility establish what identification information one can expect to elicit from your DNA data using your method.

For example, suppose that you are using an inclusion method to interpret your DNA mixture. Then the *efficacy* (Figure 1, orange bars) establishes the range of log(LR) information expected from your data. Alternatively, if you are using quantitative computer interpretation (Figure 1, blue and green bars), then the court might expect significantly more informative match scores.

The *reproducibility* (Figure 2, blue and green bars) shows the reliability of your method on repeated interpretations. The small variation shown (less than 0.036) establishes that a method produces similar information across independent data interpretations.

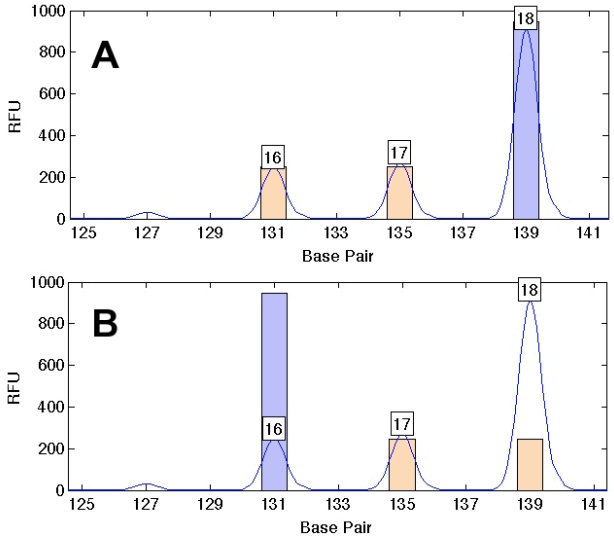

Figure 3. *Mixture Interpretation.* **(A)** *quantitative,* **(B)** *qualitative.*

On direct examination at trial, you can explain to the jury how your interpretation method infers genotypes (e.g., by inclusion or quantitative addition). A genotype allele pair forms an intuitive visual pattern that you can compare with the DNA data peaks to justify your inference. For example, the TrueAllele Explain interface lets you do "what if" analyses to illustrate allele pair combinations that fit the data well (Figure 3A, quantitative interpretation) or poorly (Figure 3B, qualitative interpretation) with mixture data. Since "a better fit's more likely it," the first allele pair has a higher likelihood of explaining the quantitative data.

C1 vs. G (CYB)

File Signature Statement Summary Calculation

C1 vs. G
23-Mar-2010

The LR calculation assumes two unknown contributors in the evidence relative to a CYB human population.
The match rarity between the evidence and suspect is 111 quintillion.

The joint LR is approximately 111 quintillion.
The log(LR) information is 20.04.

locus	allele pair	Q	R	S	LR	log(LR)
CSF1PO	12, 12	0.429	0.0894	1	4.797	0.681
D13S317	9, 13	0.975	0.0159	1	61.352	1.788
D16S539	9, 12	0.999	0.0832	1	12.012	1.080
D18S51	13, 15	1	0.0288	1	34.699	1.540
D21S11	30, 31	1	0.0345	1	28.963	1.462
D3S1358	16, 17	0.997	0.1128	1	8.839	0.946
D5S818	12, 12	0.871	0.1181	1	7.377	0.868
D7S820	10, 10	0.330	0.0920	1	3.586	0.555
D8S1179	8, 11	1	0.0012	1	828.507	2.918
FGA	21, 22	1	0.0511	1	19.550	1.291
Penta_D	12, 14	1	0.0106	1	93.943	1.973
Penta_E	7, 14	0.981	0.0184	1	53.389	1.727
TH01	9, 9.3	0.996	0.0586	1	16.998	1.230
TPOX	8, 8	0.931	0.1746	1	5.332	0.727
vWA	15, 18	1	0.0549	1	18.215	1.260

Figure 4. *TrueAllele Simple Report.*

The match report is a computer calculation that automatically follows from your inferred genotype. In the case report, I like to include a spreadsheet that explains the LR computation. At trial, there are simple LR approximations that may be easier to explain. For example, the TrueAllele simple report lists the match information at every locus (Figure 4), with the LR calculated as the (posterior to prior) probability ratio of inferred genotype to population prevalence at the allele pair that matches the suspect. This "information gain" makes it easy to explain the LR – just point to the before and after evidence genotype probabilities and state their ratio.

On cross-examination, you may be asked why there is more than one match statistic. These statistics can arise from using different interpretation methods and different ethnic populations for allele frequencies. It is easy to educate the jury about how some DNA interpretation methods use more of the data than others. For example, less informative inclusion uses neither quantitative peak heights nor the victim profile (Figure 1, orange), while more informative computer interpretation can use both (Figure 2, blue and green.) Therefore, one expects a million or billion times more match information from better use of the same DNA evidence data.

Tutorial

How to Validate Your Own Mixture Interpretation Method

To validate your own (human or computer based) mixture interpretation method, you begin with these materials:

- a set of DNA items
- duplicate interpretations of each item done independently using the same method
- a suspect reference for comparison on each item
- one or more population databases

Each DNA item's interpretation corresponds to a genotype. With data uncertainty, these genotypes may contain multiple allele pairs (having positive probabilities) that can explain the data. This genotype (probability distribution) summarizes how well the interpretation method has preserved the data's identification information.

The likelihood ratio (LR) quantifies how much identification information was gained through the inferred genotype. More informative interpretation methods capture more of the data in their genotype, and produce higher LRs. Every DNA match statistic between evidence and suspect relative to a population is a LR:

- CPI. Using your Popstats computed CPE percentage value, the CPI LR is calculated as $LR_{CPI} = 100\% / (100\% - CPE)$
- CLR. The Popstats program gives you a LR in the proper form when you subtract out the victim genotype.
- CMP or RMP. The coancestry-adjusted single source conditional (or unadjusted random) match probability is the reciprocal of the LR. Just use the LR ratio 1/CMP.
- LR. TrueAllele's Report interface provides LR values and reports letting you choose populations and set coancestry coefficients. To be conservative, use the lowest match score across the different ethnic population databases.

To measure the *efficacy* of your interpretation method, simply take the average of all your $\log_{10}$ (LR) scores for every item. The logarithm is used because it lets us add together match information ($\log(LR_1 \times LR_2) = \log(LR_1) + \log(LR_2)$), and descriptive statistics work best in additive units.

To measure the *reproducibility* of your interpretation method, form the within-group average of log(LR) scores for one item. Then calculate the within-group squared deviation between each interpretation's log(LR) value and the item average. Add these squared deviations together for all interpretations of all items, and divide by the total number of interpretations. The square root of this number is the conventional within-group standard deviation that describes reproducibility [Perlin, 2006].

You have now validated your DNA interpretation method. When challenged in court, you can produce your spreadsheet, and quantitatively show the *reliability* of your method. The *efficacy* value tells how much identification information your method extracts from typical data, while the *reproducibility* value gives the expected variation. Your validation results can help justify why the match scores you report in a case are entirely expected, given your representative DNA lab data and validated interpretation method.

References

Perlin MW. Scientific validation of mixture interpretation methods. Promega's Seventeenth International Symposium on Human Identification, 2006 Oct 10-12; Nashville, TN. 2006.

Perlin MW, Duceman BW. Casework validation of genetic calculator mixture interpretation (A77). AAFS 62nd Annual Scientific Meeting, 2010 February 22-27; Seattle, WA. American Academy of Forensic Sciences; 2010. p. 62-3.

Conference talk with live audio
http://www.cybgen.com/information/AAFS2010validate.shtml

Challenging DNA Evidence? Try TrueAllele Casework service – FREE*

Some DNA seems inconclusive or has a low match statistic. But TrueAllele is a million times more informative. Let Cybergenetics run TrueAllele on your DNA data. We will give you a verbal report of your data results.

This "free first look" offer is valid through August 1, 2010. Please contact Cybergenetics at expert@cybgen.com to send your data.

*Limited to one case per crime lab, legal practice or other entity. No more than five samples per case. Cybergenetics reserves the right to use anonymized case data and results from this offer for scientific and educational purposes.