CASEWORK VALIDATION OF TRUEALLELE MIXTURE INTERPRETATION

Mark W. Perlin, PhD, MD, PhD[1], Erin Turo[1] Barry W. Duceman, PhD[2],

[1]Cybergenetics, Pittsburgh, PA
[2]New York State Police Forensic Identification Center, Albany, NY

Interpreting DNA mixtures can be challenging. With the advent of statistical computing, one can reproducibly infer consistent, highly informative results. Such reliable mixture inference is critical for the admissibility of scientific evidence. This paper establishes the efficacy of TrueAllele® computer mixture interpretation by comparing it with currently used methods on adjudicated mixture cases, also showing the reproducibility of the computer's results.

The key mixture interpretation task is inferring a questioned genotype of an unknown contributor. Different mixture interpretation methods infer such genotypes up to probability. These are list-based inclusion methods, such as:
• CPI, a qualitative approach that does not use the victim genotype, and
• CLR, which does use the victim genotype.

TrueAllele is a statistical computer approach that infers genotypes by hypothesizing all feasible solutions, comparing these with observed STR peak height data, and assigning higher probabilities to genotype hypotheses that better fit the data. Here we used:
• TA1, which uses a victim profile to infer one unknown contributor, and
• TA2 which does not, instead inferring two unknown contributors.

The Likelihood Ratio (LR) is the generally accepted forensic science measure of match rarity. The LR gives the probability of a match between the evidence genotype and a suspect, relative to a match with a random person. The data-inferred evidence genotypes above (CPI, CLR, TA1, TA2) each automatically produce a LR match statistic when their genotype is substituted into a generic LR match formula.

The efficacy of TrueAllele was determined by comparing its LR match information to other methods. In particular, the LR logarithm (order of magnitude, or number of zeros after the decimal place) was determined on eight adjudicated cases for the two unknown TA2 computer method, and compared with that of the reported CPI value. Whereas the average log(LR) information for CPI was 6 (million to one), the average match information on these same cases with TA2 was 12 (trillion to one). This shows a six order of magnitude improvement using TA2 relative to CPI.

We also assessed the relative efficacy when the victim profile was known and just one unknown contributor was inferred. The average log(LR) match information on eight adjudicated CLR cases was 12 (trillion to one). The average TA1 match information on these same cases was 17, a five order of magnitude improvement. We therefore conclude that for both one and two unknown contributors, the TrueAllele mixture interpretation method is more informative than the CPI and CLR match statistics.

Reproducibility was measured on these 16 mixture cases by obtaining duplicate computer solutions for each case. The average match information difference between the two independent solutions was considerably less than one log(LR) unit.

We conclude from this study that the TrueAllele system provides reliable mixture interpretation. Specifically, when inferring either one or two unknown contributor genotypes, TrueAllele is effective relative to current methods. Moreover, we have quantified the reproducibility of its match information. TrueAllele Casework has already been admitted into evidence in Frye jurisdictions. This validation study (efficacy and reproducibility) establishes TrueAllele's reliability under the additional prongs of Daubert.