Suffolk County TrueAllele® Validation

Suffolk County Crime Laboratory
Cybergenetics, Corp.

New York State Commission on Forensic Science
DNA Subcommittee
Meeting on 20 May 2011

Introduction

The purpose of this study is to determine the relative efficacy and reproducibility of Cybergenetics TrueAllele® computer inferred genotype matches, and the suitability of TrueAllele for routine use in laboratory casework. The human-classified "inconclusive" samples are of particular interest. A key question is to what extent the computer can obtain informative DNA match results from such previously uninformative evidence items.

Materials and Methods

The Suffolk County TrueAllele validation study was done on 67 Suffolk crime cases that covered a broad range of crime categories (Table 1), including 29 rapes or other sexual assaults. The computer interpreted 122 mixture items (Table 2), primarily comprising 87 two contributor mixtures. More information about each item is given in sheets 1 and 2 of spreadsheet file Item_Summary.xls.

The Suffolk County laboratory generated data using the Identifiler STR panel, size separating the fluorescently labeled amplicons on an ABI 3130 DNA sequencer. TrueAllele was used to analyze the electronic .fsa data files, and upload quality checked quantified peak size data into a Suffolk project TrueAllele database. Following request generation in the VUIer™ interface, TrueAllele interpreted the evidence. The number of assumed contributors was estimated from the data; when this number of was uncertain, the case item was processed multiple times, with each run assuming a different number of contributors. The TrueAllele database was copied and reset, with reinterpretation of all requests producing duplicate genotype results.

TrueAllele objectively inferred genotypes from the electronic evidence data, without access to any suspect reference genotypes. TrueAllele then compared these inferred evidence genotypes with the suspect references to compute a likelihood ratio (LR) match statistic, relative to three standard ethnic populations (Black, Caucasian, Hispanic). The smallest LR was reported, following the lab's practice. A coancestry coefficient of 1% was used throughout.

The $\log_{10}$(LR) is a standard additive measure of information, used throughout the study. Efficacy is determined by the average log(LR) value, relative efficacy by the average log(LR) difference between two methods, and reproducibility by within-item standard deviation based on multiple interpretations of the same item using the same method (1).

Human interpretation of the data was done by the Suffolk County laboratory using threshold-based methods, and recorded in a case file. These threshold-based methods were Random Match Probability (RMP) using a theta value of 3%, and Combined Probability of Inclusion (CPI). Both of these match scores are LRs, and so can be directly compared with the computer results to assess how much information is preserved from the evidence data.

Results

The detailed results are recorded in spreadsheet file Validation.xls, following the column legend provided in document file Descriptions.pdf. We focus here on the results described in spreadsheet file Match_Comparison.xls; please see Descriptions.doc for column descriptions.

The suspect match information is summarized (Table 3), stratified by how much log(LR) identification information the laboratory obtained from the evidence items using human review methods. Of the 52 computer-scored evidence items, 27 (52%) were deemed "inconclusive" by human review and produced no match score. On this majority of cases, the computer found an average log(LR) match efficacy of 10.02 (10.5 billion).

As the lab-assigned informativeness of the mixture item increased (from under a thousand LR to over a billion), the average log(LR) efficacy of computer review increases. The within-item standard deviation increased as well, with an outlier of 0.744 in the "< billion" category due to a small sample size of 3 items that included an item having discrepant log(LR) information values of 7.09 and 4.67.

Discrepant LR values are important in probabilistic genotype interpretation of DNA evidence because they can identify outlier computer results. In casework practice, the item would be processed for a third time in TrueAllele to determine the true genotype and LR match information. Most often, the lower score represents an outlier statistical genotype computation due to processing variation. As with PCR experiments in the molecular biology laboratory, random processes sometimes go awry. In such cases, outlier detection is needed, and, through recomputation on the same data, use the two genotypes having (the typically higher) concordant LR values.

By definition, the human review results based on peak threshold interpretation follow the same pattern, with information increasing in each category. In the "uninformative" category, the information discrepancy is 10 log units. Under a billion LR, the average loss of information in Suffolk-reported items is about 5 log units. Only with straightforward data (having an LR over a billion) does human review achieve parity with computer review, with both approaches having average LRs of over a quadrillion.

The overall results on the 52 items of evidence showed an average log(LR) information efficacy of 10.92 for TrueAllele interpretation, and 4.23 (accounting for "inconclusives") for threshold-based methods. The average information difference of 6.69 represents a loss of 4.9 million for the typical mixture item, which is consistent with previous mixture comparison studies (1).

The log(LR) information comparison results for victim matches (Match_Comparison.xls, sheet Victim) follows a similar pattern of information preservation and loss. Again, a majority of mixture items (14 of 27) were considered "inconclusive" by human review, but yielded an informative TrueAllele result, with an average information gain of 11.51 (324 billion).

There were 70 items for which no reference was available for comparison (Match_Comparison.xls, sheet NoRef). No match score was produced on 42 items (Match_Comparison.xls, sheet NoMatch). For some items, human review data yielded a match rarity data statistic, whereas the computer could exclude an individual by finding a large negative log(LR) score (Match_Comparison.xls, sheet RSUI).

Conclusions

With the advent of statistical computing at the dawn of the computer age (2), science has been able to solve complex problems through probability models. This Markov chain Monte Carlo (MCMC) approach (3) endeavors to fully explain the data by having a computer repeatedly explore many possible solutions thousands or millions of times, yielding results by statistical sampling. Data uncertainty is represented through the standard scientific language of probability, with data variation easily modeled by computer (4). Scientific reporting of identification information is captured in the LR, specifically designed for describing information change under uncertainty (5).

Probabilistic genotype methods have been used in forensic STR mixture interpretation for many years (6-9). These objective methods use of all the quantitative data (10-12), rather than some threshold-based data reduction to all-or-none putative allele events. Forensic workers now use MCMC computation to reliably solve DNA mixture problems (13, 14). In this way, probability modeling conforms to the accepted practices and procedures of 21st century scientific thought, where computers are used to tame uncertainty by explaining it. The recent 2010 SWGDAM guidelines accept probabilistic genotypes as a valid way of interpreting mixtures (paragraph 3.2.2).

This TrueAllele validation study demonstrates that computers can reliably interpret DNA evidence items, including mixtures. The computer more than doubled the number of suspect matches, from the initial human review total of 25 up to final count of 52. This suggests that half of informative DNA mixture evidence may be discarded by current interpretation practices. The average information gain using computer review was over a million-fold. Moreover, the genotype reproducibility showed a standard deviation of under half a log(LR) information unit.

Forensic DNA identification is an information science, in a world where computers are expected to do the heavy lifting for people. The American public pays for DNA programs in order to ensure public safety by catching criminals, and it expects the best possible practices to be used by crime labs in order to protect society. This study establishes the clear superiority of computer probabilistic genotyping based on quantitative DNA data over the current human threshold approach. It is time for forensics to conform to the rest of science, and adopt more powerful computer methods that can solve and prevent crime.

Table 1. The crime type and number of cases reviewed in this study.

Type of Crime	**Quantity**
arson	1
assault	5
burglary	12
criminal possession of a controlled substance	1
criminal possession of a weapon	3
endangering the welfare of a child	1
larceny	6
rape	24
other sexual assault	5
reckless discharge of a firearm	1
robbery	8
Total	**67**

Table 2. The number of contributors contained in each mixture item, and the quantity of each type of mixture item.

Number of contributors	Quantity
2	87
3	32
4	3
Total number of items	**122**

Table 3. Summary of the suspect match information results, as stratified by the laboratory's information classification. Information is measured by log(LR).

Information	N	Efficacy	Reproducibility	Laboratory	Method	Difference
uninformative	27	10.02	0.510	0.00	None	10.02
< thousand	5	5.83	0.337	1.20	CPI	4.63
< million	7	9.21	0.316	3.91	CPI	5.30
< billion	3	12.39	0.744	7.46	CPI	4.93
> billion	10	16.65	0.125	16.40	RMP	0.24
Total	*52*	*10.92*	*0.441*	*4.23*	*Human*	*6.69*

References

1. Perlin MW, Legler MM, Spencer CE, Smith JL, Allan WP, Belrose JL, et al. Validating TrueAllele® DNA mixture interpretation. Journal of Forensic Sciences. 2011;56(November):in press.
2. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. J Chem Phys. 1953;21(6):1087–92.
3. Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. Biometrika. 1970;57:97-109.
4. Gelfand AE, Smith AFM. Sampling-based approaches to calculating marginal densities. J Amer Statist Assoc. 1990;85:398-409.
5. Good IJ. Probability and the Weighing of Evidence. London: Griffin, 1950.
6. Cowell RG, Lauritzen SL, Mortera J. A gamma bayesian network for DNA mixture analysis. Bayesian Anal. 2007;2(2):333-48.
7. Mortera J, Dawid AP, Lauritzen SL. Probabilistic expert systems for DNA mixture profiling. Theoretical Population Biology. 2003;63:191-205.
8. Perlin MW. Simple reporting of complex DNA evidence: automated computer interpretation. Promega's Fourteenth International Symposium on Human Identification, 2003; Phoenix, AZ. 2003.
9. Tvedebrink T, Eriksen PS, Mogensen HS, Morlin N. Identifying contributors of DNA mixtures by means of quantitative information of STR typing. J Comput Biol. 2011;18(In press).
10. Evett IW, Gill P, Lambert JA. Taking account of peak areas when interpreting mixed DNA profiles. J Forensic Sci. 1998;43(1):62-9.
11. Gill P, Sparkes R, Pinchin R, Clayton TM, Whitaker JP, Buckleton J. Interpreting simple STR mixtures using allele peak area. Forensic Sci Int. 1998;91:41-53.
12. Perlin MW, Szabady B. Linear mixture analysis: a mathematical approach to resolving mixed DNA samples. J Forensic Sci. 2001;46(6):1372-7.
13. Curran J. A MCMC method for resolving two person mixtures. Sci Justice. 2008;48(4):168-77.
14. Perlin MW, Sinelnikov A. An information gap in DNA evidence interpretation. PLoS ONE. 2009;4(12):e8327.