Overcoming DNA Stochastic Effects

2010 NEAFS & NEDIS Meeting
November, 2010
Manchester, VT

Mark W Perlman, MD, PhD
Cybergenetics, Pittsburgh, PA

Coping with uncertainty

- Reproducible DNA data exhibit stochastic variation
- Probability models can capture stochastic effects
- Genotypes inferred from uncertain data are probability distributions (e.g., CPI)

- Science expects us to account for stochastic effects
- The law expects us to testify within our certainty
- The likelihood ratio (LR) meets both demands
- The LR expresses genotype uncertainty, reflecting the underlying data uncertainty

PCR is a random process

PCR efficiency is not 100% efficient. A strand copies with probability $p$, and doesn't copy with probability $1-p$. 
STR peak is a random variable

\[ p = 80\%, n = 6 \]

One amplification

Another amplification

STR peak height measurement reflects probability distribution

Mean = 2,145
Stdev = 284

Relative peak certainty: coefficient of variation

Stdev = 12
Mean = 85
CV = 14%

\[ CV = \frac{\text{standard deviation}}{\text{mean value}} \]
Four times the peak height, gives twice the peak certainty

\[
\text{stdev} = 25 \\
\text{mean} = 350 \\
\text{CV} = 7\%
\]

STR peaks are random variables

To interpret quantitative data, some use a qualitative threshold

Over threshold, peaks are treated as allele events. Under threshold, alleles do not exist.
Thresholds introduce error

False allele exclusions

2010 SWGDAM Guidelines

When every peak is under threshold, all the data disappears.
No genotype, no match score, nothing left to say.
Higher false exclusion rate

SWGDAM 2010 – Mixtures

3.2.2. If a stochastic threshold based on peak height is not used in the evaluation of DNA typing results, the laboratory must establish alternative criteria (e.g., quantitation values or use of a probabilistic genotype approach) for addressing potential stochastic amplification. The criteria must be supported by empirical data and internal validation and must be documented in the standard operating procedures.

• higher peak threshold discards information
• probability modeling preserves information

Probability modeling

• Bayes Theorem: addresses scientific uncertainty
• uses likelihood function to update probability
• joint likelihood combines independent evidence

• likelihood: how well parameters explain the data
• STR data: must explain every peak (all rfu)
• likelihood gives probability at one peak: genotype allele prediction vs. peak height
• joint likelihood at all peaks:
  multiply together the individual peak likelihoods
Compare genotype pattern vs. data

Overcome stochastic effects

Computers can solve for genotype probabilities and other random variables, like peak variation. By modeling peak variation as just another parameter, computers can overcome DNA stochastic effects.

TrueAllele® Casework

- quantitative computer interpretation
- statistical search of probability model
- preserves all identification information
- objectively infer genotype, then match
- any number of mixture contributors
- stutter, imbalance, degraded DNA
- calculates uncertainty of every peak
- created in 1999, now in version 25
- used on 100,000 evidence samples
- available as product, service or both
Commonwealth v. Foley

<table>
<thead>
<tr>
<th>Score</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>13 thousand</td>
<td>obligation</td>
</tr>
<tr>
<td>23 million</td>
<td>allele</td>
</tr>
<tr>
<td>189 billion</td>
<td>TrueAllele</td>
</tr>
</tbody>
</table>

- peak threshold discards information
- probability modeling preserves information

Likelihood Comparison

Quantitative TrueAllele likelihood focuses probability on true genotype
Inclusion method likelihood disperses probability across incorrect allele pairs

SWGDAM 2010 – Mixtures

3.4.3.1. If composite profiles (i.e., generated by combining typing results obtained from multiple amplifications and/or injections) are used, the laboratory should establish guidelines for the generation of the composite result. When separate extracts from different locations on a given evidentiary item are combined prior to amplification, the resultant DNA profile is not considered a composite profile. Unless there is a reasonable expectation of samples originating from a common source (e.g., duplicate vaginal swabs or a bone), allelic data from separate extractions from different locations on a given evidentiary item should not be combined into a composite profile. The laboratory should establish guidelines for determining the suitability of developing composite profiles from such samples.

- joint likelihood function combines evidence
- probability modeling preserves information

Regina v. Broughton

- low template mixture
- three DNA contributors
- triplicate amplification
- post-PCR enhancement
- inconclusive human result
- TrueAllele interpretation

Regina v. Broughton

<table>
<thead>
<tr>
<th>Score</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>nothing</td>
<td>human inclusion</td>
</tr>
<tr>
<td>3.6 million</td>
<td>TrueAllele computer</td>
</tr>
</tbody>
</table>

- joint likelihood function combines evidence
- probability modeling preserves information
Preserve vs. Discard

- Peak threshold discards information - 70% of the time
- Quantitative likelihood preserves information - every time

Conclusions

- Quantitative data has stochastic effects
- Model data with joint likelihood function
- Probability modeling preserves information
- And can statistically combine DNA evidence

- Exact modeling of peak variation can replace inexact thresholds to scientifically overcome stochastic effects

Cybergenetics © 2007-2010
http://www.cybgen.com/information/presentations.shtml
perlin@cybgen.com