Open Access DNA Database

Whose DNA is it Anyway?

Duquesne University
March, 2013
Pittsburgh, PA

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

DNA Mixture Data
Quantitative peak heights at a locus

Genotype Inference
Computer-based probabilistic genotyping

Explain the peak pattern
Thorough & objective

Victim's allele pair
Another person's allele pair

Allele Pair
5% 15, 15
90% 15, 16
5% 15, 17
15, 18
16, 16
16, 17
16, 18
17, 17
17, 18
18, 18
Identification Information

Likelihood Ratio
Explaining all the data under two competing hypotheses

Matching genotype probability

Evidence 90%
Coincidence 9%

\[ LR = \frac{90}{9} = 10 \]

\[ \log(LR) = \log(10) = 1 \text{ ban} \]

Data Summary for CPI

Over threshold, peaks are labeled as allele events

All-or-none allele events

Threshold

Allele Pair
5\%15, 15
9\%15, 16
15\%15, 17
9\%15, 18
5\%16, 16
15\%16, 17
9\%16, 18
13\%17, 17
15\%17, 18
5\%18, 18

Information Loss

Combined probability of inclusion
CPI explains less of the data (no peak heights or model)

Matching genotype probability

Evidence 9%
Coincidence 9%

\[ LR = \frac{9}{9} = 1 \]

\[ \log(LR) = \log(1) = 0 \text{ ban} \]
Mixture Information Study

2 & 3 person DNA mixtures

- 16 cases
- 31 evidence items
- 41 genotype matches

<table>
<thead>
<tr>
<th>Crime</th>
<th>Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>homicide (7)</td>
<td>clothing (12)</td>
</tr>
<tr>
<td>sexual assault (5)</td>
<td>weapon (6)</td>
</tr>
<tr>
<td>assault (2)</td>
<td>vehicle (5)</td>
</tr>
<tr>
<td>death investigation (1)</td>
<td>skin swab (3)</td>
</tr>
<tr>
<td>robbery (1)</td>
<td>vaginal swab (3)</td>
</tr>
<tr>
<td></td>
<td>fingernail (1)</td>
</tr>
<tr>
<td></td>
<td>rectal swab (1)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Information (per locus)</th>
</tr>
</thead>
<tbody>
<tr>
<td>log(LR) Count</td>
</tr>
<tr>
<td>LR computer information</td>
</tr>
<tr>
<td>CPI human match statistic</td>
</tr>
<tr>
<td>0.746 (0.590) 0.489 0.746 (0.590)</td>
</tr>
<tr>
<td>N = 517</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CPI Statistic (per locus)</th>
</tr>
</thead>
<tbody>
<tr>
<td>log(LR) Count</td>
</tr>
<tr>
<td>LR computer information</td>
</tr>
<tr>
<td>CPI human match statistic</td>
</tr>
<tr>
<td>0.000 (0.615) 0.746 (0.590)</td>
</tr>
<tr>
<td>N = 517</td>
</tr>
</tbody>
</table>
Joint Statistic Distribution

- Imaginative
- Overstated
- Same Information
- Conservative

Joint Statistic Distribution

\[ r = 0.376 \]
\[ r^2 = 0.141 \]

Investigative DNA Database

<table>
<thead>
<tr>
<th>probablistic genotypes</th>
<th>reference genotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMEL0</td>
<td>1 2</td>
</tr>
<tr>
<td>CSF1PO</td>
<td>12 12</td>
</tr>
<tr>
<td>D13S317</td>
<td>9 12</td>
</tr>
<tr>
<td>D16S539</td>
<td>9 12</td>
</tr>
<tr>
<td>D18S51</td>
<td>13 15</td>
</tr>
<tr>
<td>D21S11</td>
<td>30 31</td>
</tr>
<tr>
<td>D31S150</td>
<td>16 17</td>
</tr>
<tr>
<td>D5S818</td>
<td>12 12</td>
</tr>
<tr>
<td>D7S8020</td>
<td>10 10</td>
</tr>
<tr>
<td>D8S1179</td>
<td>8 11</td>
</tr>
<tr>
<td>POLA</td>
<td>21 22</td>
</tr>
<tr>
<td>Penta D</td>
<td>12 14</td>
</tr>
<tr>
<td>Penta E</td>
<td>7 14</td>
</tr>
<tr>
<td>TH01</td>
<td>9 6 3</td>
</tr>
<tr>
<td>TPOX</td>
<td>8 6</td>
</tr>
<tr>
<td>VWA</td>
<td>15 18</td>
</tr>
</tbody>
</table>

LR match statistic \[10^{17.5}\]
Investigative Applications

- evidence-to-suspect, solve cold cases
- evidence-to-evidence, connect serial crime
- evidence-to-kinship, find missing people
- kinship-to-reference, conduct familial search
- remains-to-missing, identify disaster victims


Sensitive

Two contributors, victim known
Infer one probabilistic genotype

Measure DNA match information
log(LR) 0 10 20 17.7

Specific

Compare with 1,000 random genotypes

-30 -20 -10 0 10 20 -23.9 17.7
Allele List Database

<table>
<thead>
<tr>
<th>Locus</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSF1PO</td>
<td>12</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D2S1308</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D5S810</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D7S800</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D8S1179</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FGA</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TH01</td>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TPOX</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VWA</td>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

7 of 13 loci hit

moderate stringency

Information Comparison

<table>
<thead>
<tr>
<th>Type of DNA database</th>
<th>Sensitive?</th>
<th>Specific?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probabilistic genotype</td>
<td>LR average is about a quadrillion</td>
<td>False positive rate &lt; 0.01%</td>
</tr>
<tr>
<td>Allele list (moderate stringency)</td>
<td>Upload fails about 1/3 of the time</td>
<td>Hits 5%-25% of the DNA database</td>
</tr>
</tbody>
</table>

State v. Sullivan

Location: Western United States
Crime: Assault with weapon
Evidence: Gun
True information: -1, +9
Human guess: 0 (inconclusive)
Reason: Mixture
Outcome: Still in prison
DNA database exoneration?

Post-conviction test
Information < 0

The right person?
Information ~ 9

10 million offenders
CODIS ineffective
Any alternative?

State v. Diggins

Location: Southern United States
Crime: Rape
Evidence: Tampon string
True information: 4
Human guess: 0 (inconclusive)
Reason: Low DNA amount
Outcome: Still serving life sentence

DNA database identification?

Computer-inferred evidence genotype

Anonymized offender suspect genotypes
Is everything being done that could be done?

Government monopoly
• offender genotypes
• database access
• uses old technology
• not highly effective

Balancing public safety v. personal privacy
Access to anonymized DNA database

Reducing rape by DNA prevention

Learn More
DNA Identification Science
• Newsletters
gentle introduction to ideas
• Courses
for lawyers and scientists
• Presentations
handouts, movies, transcripts
• Publications
abstracts, manuscripts

www.cybgen.com/information