Lectures 29-30: Statistical Evaluation of Genetic Linkage

• Phase
• Lod scores
We genotype the six members of the family for SSRs scattered throughout the genome (which spans 3300 cM)—

one SSR must be within 10 cM of the Huntington's gene: HD?
$$\text{LOD}_{0.06} \text{(family 1)} = \log_{10} \left( \frac{0.024}{0.0039} \right) = \log_{10} (6.25) = 0.796$$

Same for families #2 and #3:

$$\sum \text{LOD}_{0.06} \text{ (families 1, 2, 3)} = 3 \times 0.796 = 2.388$$

Family #4:

<table>
<thead>
<tr>
<th>Maternal alleles</th>
<th>HD SSR37</th>
<th>HD</th>
<th>HD</th>
<th>+</th>
<th>HD</th>
</tr>
</thead>
<tbody>
<tr>
<td>HD D</td>
<td>D D D D</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$$P \text{ if linked at } 0.06 = \frac{1}{2} (P \text{ if phase 1}) + \frac{1}{2} (P \text{ if phase 2})$$

$$= \frac{1}{2} (0.47 \times 0.47 \times 0.03 \times 0.47) + \frac{1}{2} (0.03 \times 0.03 \times 0.47 \times 0.03) = 0.0016$$

$$\text{LOD}_{0.06} \text{ (family 4)} = \log_{10} \left( \frac{0.0016}{0.0039} \right) = \log_{10} (0.41) = -0.387$$
\[ \sum \text{LOD}_{0.06} (\text{families 1, 2, 3, 4}) = 2.388 - 0.387 = 2.001 \]

Still not sufficient to publish. What to do?

1. It's tempting to ignore family 4 — to declare it to be irrelevant for some reason or another.
   
   But this would not be an acceptable solution.

2. Calculate LOD scores for other \( \theta \) values?
3. Get more families — always a good idea

4. Determine phase in affected parents

In each of the four families, we were uncertain about phase, and our LOD calculations embodied those uncertainties.

Family #4:

<table>
<thead>
<tr>
<th>Phase</th>
<th>HD</th>
<th>SSR37</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:</td>
<td>HD</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>B</td>
</tr>
<tr>
<td>2:</td>
<td>HD</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>D</td>
</tr>
</tbody>
</table>

two possible arrangements of alleles on mother's chromosomes
Typing the maternal grandparents for SSR37:

Family #4:

SSR37

A
B
C
D
E

Now we can deduce the phase in the mother:

Family #4:

SSR37

A
B
C
D
E

<table>
<thead>
<tr>
<th>Locus:</th>
<th>HD</th>
<th>SSR37</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase 1:</td>
<td>HD</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>B</td>
</tr>
<tr>
<td>Phase 2:</td>
<td>HD</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>D</td>
</tr>
</tbody>
</table>
Here is a more realistic version of the genotypic information we might obtain:

Family #4:

SSR37

A
B
C
D
E

dead
refused consent

or
inferred
Before we had written:

\[ P \text{ if linked at 0.06} = \frac{1}{2} (P \text{ if phase 1}) + \frac{1}{2} (P \text{ if phase 2}) \]

\[ = \frac{1}{2} (0.47 \times 0.47 \times 0.03 \times 0.47) + \frac{1}{2} (0.03 \times 0.03 \times 0.47 \times 0.03) = 0.0016 \]

But we now know that

\[ P \text{ if linked at 0.06} = \frac{1}{2} (P \text{ if phase 1}) + \frac{1}{2} (P \text{ if phase 2}) \]

\[ = \frac{1}{2} (0.47 \times 0.47 \times 0.03 \times 0.47) + \frac{1}{2} (0.03 \times 0.03 \times 0.47 \times 0.03) = 0.0032 \]

\[ \text{LOD}_{0.06}^{\text{family 4}} = \log_{10} \left( \frac{0.0032}{0.0039} \right) = \log_{10} (0.82) = -0.086 \]

We can sum the LOD\(_{0.06}\) scores for all four families:

\[ \sum \text{LOD}_{0.06}^{\text{family 1, 2, 3, 4}} = 2.388 - 0.086 = 2.302 \]
Overall effect of determining phase in all four families:

Add increment of $\log_{10}(2) = 0.301$ to each family’s LOD score.

\[
\sum \text{LOD}_{0.06} \text{ (families 1,2,3,4: all phased)} = \\
\sum \text{LOD}_{0.06} \text{ (families 1,2,3,4: unphased)} + 4 \log_{10} (2) \\
= 2.001 + 4 \times (0.301) = 3.205
\]

Publish!

What if we had not been able to obtain samples from any grandparents?

Try more markers
Search for SSR marker showing no recombination with HD: Where to look?

Marker showing no recombination with HD

\[ \Sigma \text{LOD}_0 \text{ (families 1,2,3,4: unphased)} = 4 \times 0.903 = 3.609 \]

Very strong conclusion!!
Fig. 5.2. Large pedigree from Venezuela with Huntington disease. A, B, C. Three different “alleles” of a DNA polymorphism. The Huntington gene is transmitted together with “allele” C. One individual, VI,5 (arrow) has so far been unaffected. She will most likely be affected later (See text). (From Gusella et al. 1983 [19])