How to find genetic determinants of naturally varying traits?

Mapping a disease locus

Fig. 11.A

(Autosomal dominant)

phenotype (variation in locus 1)

marker genotype (variation in locus 2)

<table>
<thead>
<tr>
<th>Genotype for marker</th>
<th>Genotype for marker</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1 D A2 d</td>
<td>A1 d A2 d</td>
</tr>
</tbody>
</table>

Fig. 11.A
Mapping a disease locus

![Diagram of mapping a disease locus](image)

LOD scores

<table>
<thead>
<tr>
<th>r</th>
<th>odds</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>12.244</td>
</tr>
<tr>
<td>0.2</td>
<td>10.737</td>
</tr>
<tr>
<td>0.3</td>
<td>6.325</td>
</tr>
<tr>
<td>0.4</td>
<td>2.867</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
</tr>
</tbody>
</table>

\[
\text{Odds} = \frac{P(\text{pedigree} | r) - P(\text{pedigree} | r = 0.5)}{0.5(\text{total # meioses})}
\]

A computational search through many r values

![Graph showing LOD scores vs recombination fraction](image)

observed RF is single best estimate, 1/8 = 0.125.

A computational search through many r values

![Graph showing LOD scores vs recombination fraction](image)

observed RF is single best estimate, 1/8 = 0.125.

But we already knew that. What’s the point?
More realistic situation: in dad, phase of alleles unknown

Dad phase unknown

\[
P(\text{pedigree}|r) = \frac{1}{2}(1-r)^n \cdot r^k + \frac{1}{2}(1-r)^n \cdot r^k
\]

\(k = \# \text{ recomb}, n = \# \text{ non-recomb}\)

Dad phase unknown

\[
P(\text{pedigree}|r) = \frac{1}{2}(1-r)^n \cdot r^k + \frac{1}{2}(1-r)^n \cdot r^k
\]
Dad phase unknown

odds ratio = \frac{1/2[(1-r)^3 \cdot r^3] + 1/2[(1-r)^3 \cdot r^3]}{0.5(\text{total # meioses})}

Now there are two k’s, one for each phase. We could ask for observed r; would be 1/8 or 7/8.

What single r value best explains the data?

Now you really need the computational search.

maximum likelihood
r = 0.13
Now you really need the computational search.

Maximization method was invented to map mammalian diseases in complex pedigrees.

Cystic fibrosis mapping, 1985

What does this mean?
Cystic fibrosis mapping, 1985

The molecular basis of the human serum paraoxonase activity polymorphism

(Metabolizes insecticide)

Cystic fibrosis mapping, 1985

via somatic cell hybrid mapping

(562x43) 6
Cystic fibrosis mapping, 1985

Best model is \( r = 0 \): what does this mean?

How did they get 27 kids?

Combining families
Combining families

How to get an overall estimate of probability of linkage?
A. Multiply odds together
B. Add odds together
C. Take the largest odds
D. Take the average odds

Modern genetic scans

How do you know which marker to test?
Modern genetic scans

Genotype 1000’s of markers for each individual; test each marker at various r’s across all individuals

Fig. 11.17
What does the “max” in “max LOD score” refer to?
A. The strongest-linking marker
B. The most probable recombination fraction
C. The most severe phenotype

Remember?

Max LOD score is the one from the best r value
Modern genetic scans

What is the simplest explanation for so many tall black lines around Chr 13?
A. Multiple markers in the region, which makes LOD higher
B. Multiple markers are all linked to a single disease mutation
C. Multiple mutations on Chr 13 cause the disease
D. Higher LOD is counted by the number of linking markers

Modern genetic scans

(Smooth curve = inferred genotype at positions between markers)
Modern genetic scans

But...

No Major Schizophrenia Locus Detected on Chromosome 1q in a Large Multicenter Sample

Why would an experiment fail to observe linkage?

Marker density matters

Try to minimize genotyping cost.

But if the only marker you test is >50 cM away, will get no linkage.
Number of families matters

If low number of patients, no statistical significance. Tune in next lecture for more about this.

Improper statistics

Can make noise look like a fabulously significant linkage peak.
Age of onset in breast cancer

Familial breast cancer is heterogeneous. Only early-onset families show linkage.
A landmark:  BRCA1

December 13, 1998

Some Genetic Pieces Are Falling Into Place In Breast Cancer Puzzle

BY STEVE JORDAN

BREAST cancer is a complex disease that strikes for years, as one mutation after another happens deep in a breast cell and gradually turns all tumor cells into genetic monsters.

Now scientists report significant progress in understanding one of the important steps in the early stages of the disease: how the driver genes assemble the machinery in the cell to form the disease.

"The puzzle is to understand the sequence of mutations that occur in breast cancer," said Dr. William M. Nelson, a medical geneticist at the University of Washington, Seattle. "But when you understand that it's a very complex, do it improve the chances of predicting which patients are at risk?"

In a report on the current state of the breast cancer research, the University of California at Berkeley and the Mayo Clinic revised its statistical model of the disease's genetic causes.

"We think that one of the key steps is to identify the type of mutations that occur in these genes," said Dr. Nelson. "But when you understand that it's a very complex, you improve the chances of predicting which patients are at risk?"

In a report on the current state of the breast cancer research, the University of California at Berkeley and the Mayo Clinic revised its statistical model of the disease's genetic causes.

"We think that one of the key steps is to identify the type of mutations that occur in these genes," said Dr. Nelson. "But when you understand that it's a very complex, you improve the chances of predicting which patients are at risk?"

Bring a coin and a calculator to next class.