

Lectures 29-30: Statistical Evaluation of Genetic Linkage

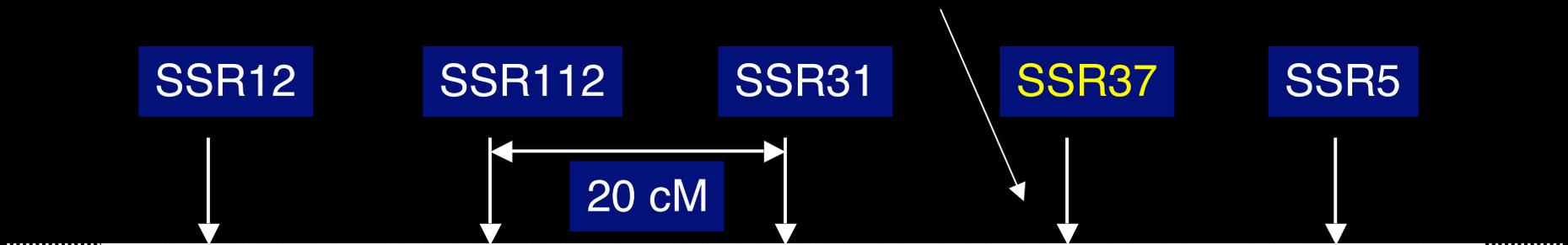
- Phase
- Lod scores

genetic linkage mapping

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 ?

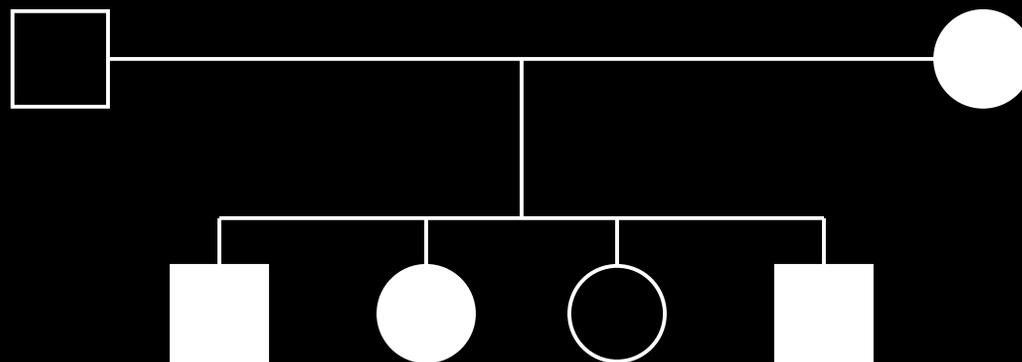


$$\text{LOD}_{0.06}(\text{family 1}) = \log_{10} (0.024/0.0039) = \log_{10} (6.25) = 0.796$$

Same for families #2 and #3:

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

Family #4:



Maternal alleles	HD SSR37	HD D	HD D	+	HD D
------------------	-------------	---------	---------	---	---------

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

$$= 1/2 (0.47 \times 0.47 \times 0.03 \times 0.47) + 1/2 (0.03 \times 0.03 \times 0.47 \times 0.03) = 0.0016$$

$$\text{LOD}_{0.06} (\text{family 4}) = \log_{10} (0.0016/0.0039) = \log_{10} (0.41) = -0.387$$

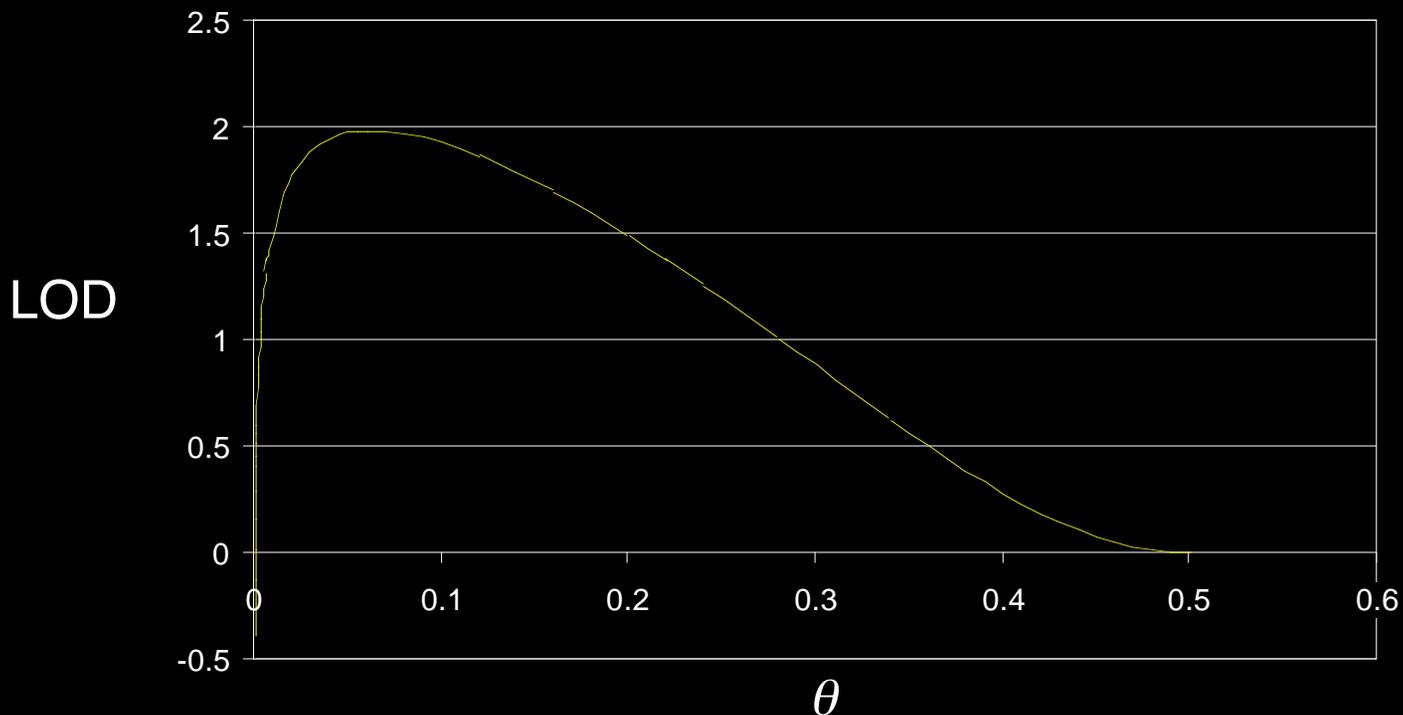
$$\Sigma \text{LOD}_{0.06} (\text{families 1, 2, 3, 4}) = 2.388 - 0.387 = 2.001$$

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 θ 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.

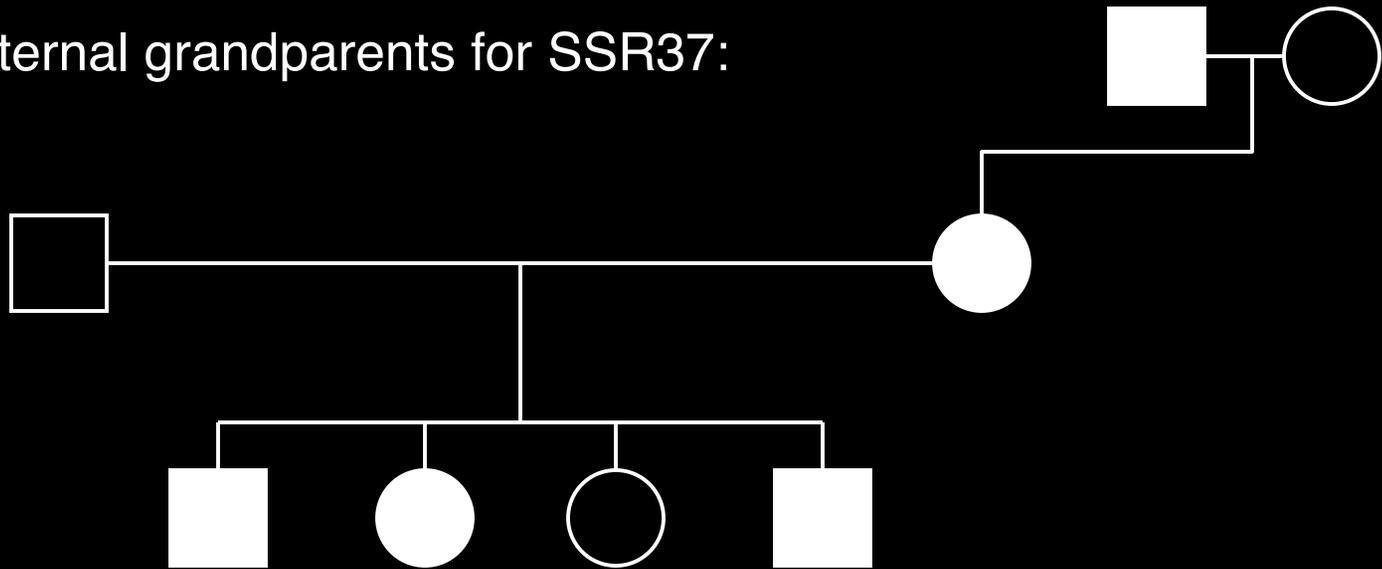
Family #4:

	HD	SSR37
Phase 1:	HD	D
	+	B
Phase 2:	HD	B
	+	D

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:

Locus:	HD	SSR37
Phase 1:	HD	D
	+	B
Phase 2:	HD	B
	+	D

Before we had written:

$$\begin{aligned} P \text{ if linked at } 0.06 &= 1/2 (P \text{ if phase 1}) + 1/2 (P \text{ if phase 2}) \\ &= 1/2 (0.47 \times 0.47 \times 0.03 \times 0.47) + 1/2 (0.03 \times 0.03 \times 0.47 \times 0.03) = 0.0016 \end{aligned}$$

But we now know that phase 1 was correct:

$$\begin{aligned} P \text{ if linked at } 0.06 &= 1/2 (P \text{ if phase 1}) + 1/2 (P \text{ if phase 2}) \\ &= 1/2 (0.47 \times 0.47 \times 0.03 \times 0.47) + 1/2 (0.03 \times 0.03 \times 0.47 \times 0.03) = 0.0032 \end{aligned}$$

$$\text{LOD}_{0.06}(\text{family 4}) = \log_{10} (0.0032/0.0039) = \log_{10} (0.82) = -0.086$$

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

$$\sum \text{LOD}_{0.06}(\text{family 1, 2, 3, 4}) = 2.388 - 0.086 = 2.302$$

phase known

Overall effect of determining phase in all four families:

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

$\Sigma \text{LOD}_{0.06}$ (families 1,2,3,4: all phased) =

$\Sigma \text{LOD}_{0.06}$ (families 1,2,3,4: unphased) + $4 \log_{10}(2)$

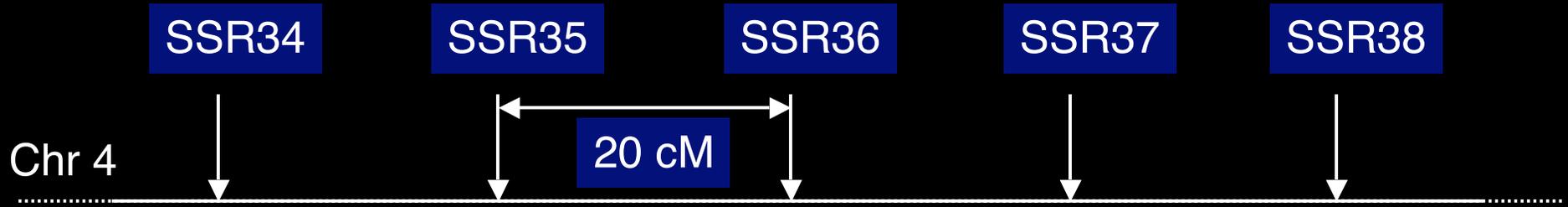
= $2.001 + 4(0.301) = 3.205$

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

or

HD

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

Very strong conclusion!!