...And Speaking of HMMs...

GENOME SCIENCES SEMINARS
Presents:

Sean Eddy, Ph.D.
Janelia Farm Scientist
Howard Hughes Medical Institute

"HMMER3: A New Generation of Sequence Homology Search Software"

Wednesday, March 10, 2010
3:30pm – 4:30pm
Foege Auditorium, S060

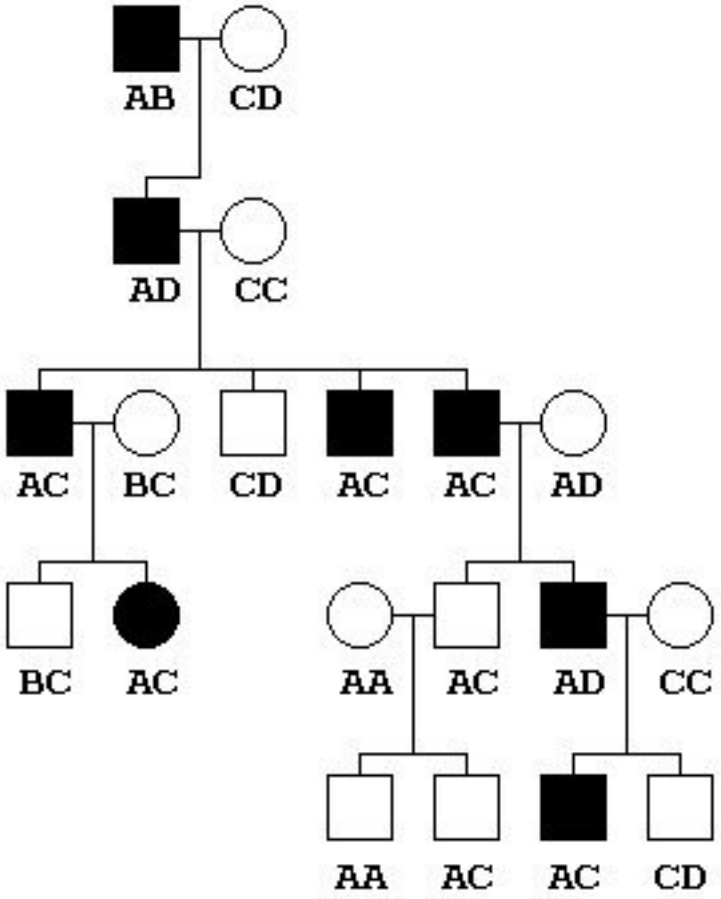
Refreshments will be served in the Foege Auditorium at 3:20PM.
Visit the Seminar website at http://www.gs.washington.edu Questions? Contact Carl

Lecture 19A — Pedigree-based Gene Mapping

- How to read a pedigree
- Transmission probabilities
- Lod scores

Thanks to Mary Kuhner for many slides

Reading a pedigree



- Squares are males, circles are females
- Shaded symbols are affected, half-shaded are carriers (none in example)

Modes of inheritance

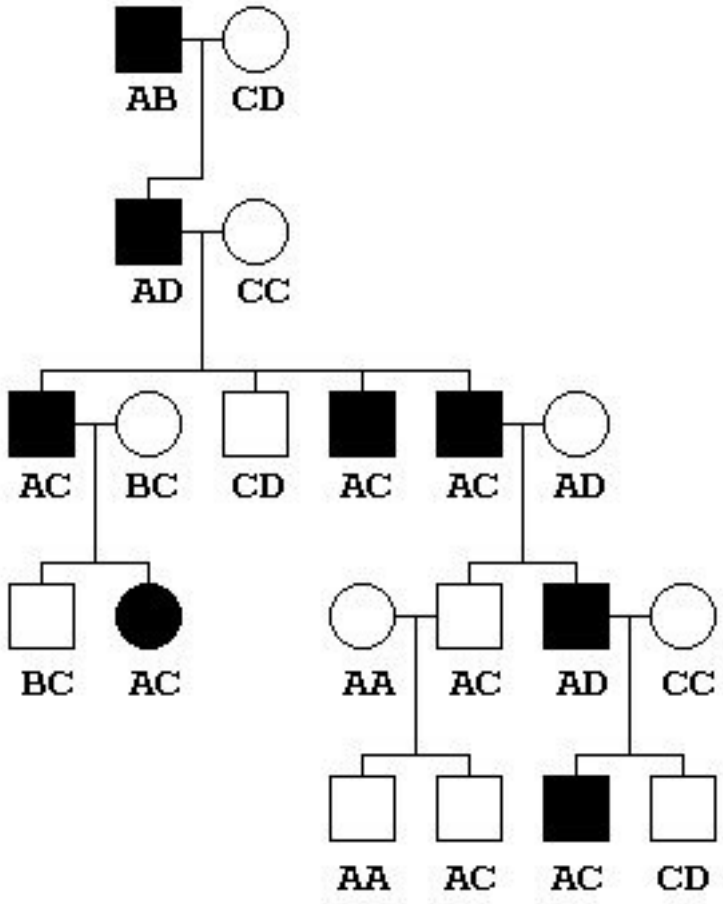
- Dominant–one gene copy leads to trait
- Recessive–two gene copies lead to trait
- Intermediate/Codominant–heterozygote is distinct
- Example: A/B/O blood group system
  - One gene, 3 alleles
  - A/B codominant, O is recessive

AA or AO	$\Rightarrow$	"Type A"
BB or BO	$\Rightarrow$	"Type B"
AB	$\Rightarrow$	"Type A/B"
OO	$\Rightarrow$	"Type O"

Typical patterns of inheritance

- Recessive trait:
  - Skips generations
  - Shows up in both sides of the family tree
  - Two affected individuals have only affected offspring

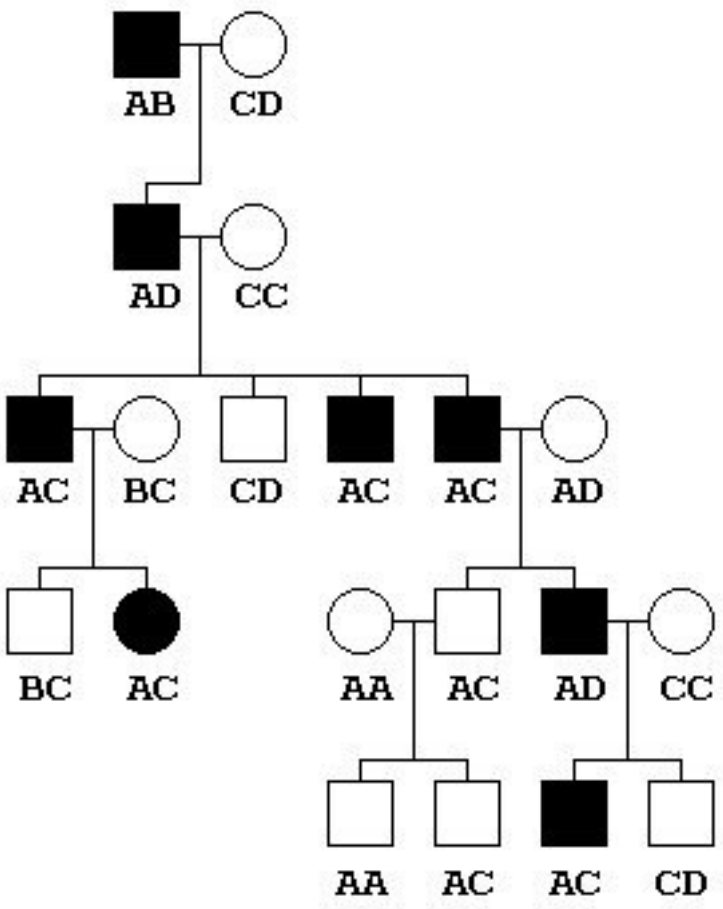
- Dominant trait:
  - Does not skip generations
  - Often in only one side of family tree
  - Two affected individuals may have unaffected offspring



Note the logic: e.g. "2 affected parents, unaffected child" $\Rightarrow$ not recessive, but "2 affected parents, all children affected" doesn't prove recessive.

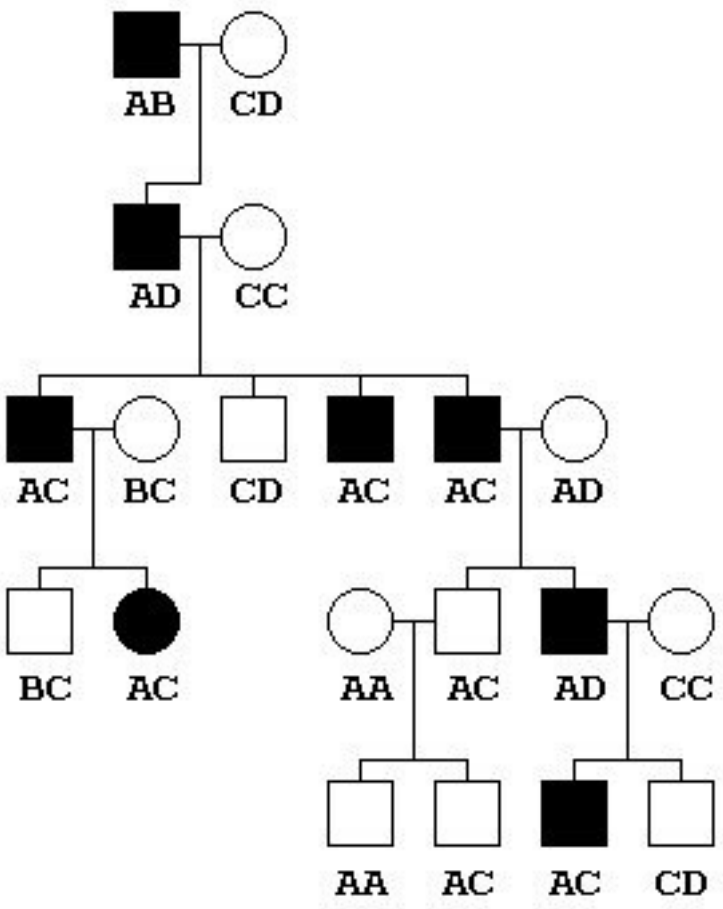
Analyzing a pedigree with marker data, Ex. 1

- Letters denote alleles at a specific marker locus, perhaps near the disease allele
- Try to identify the chromosome carrying the disease trait
- Trace it through the pedigree



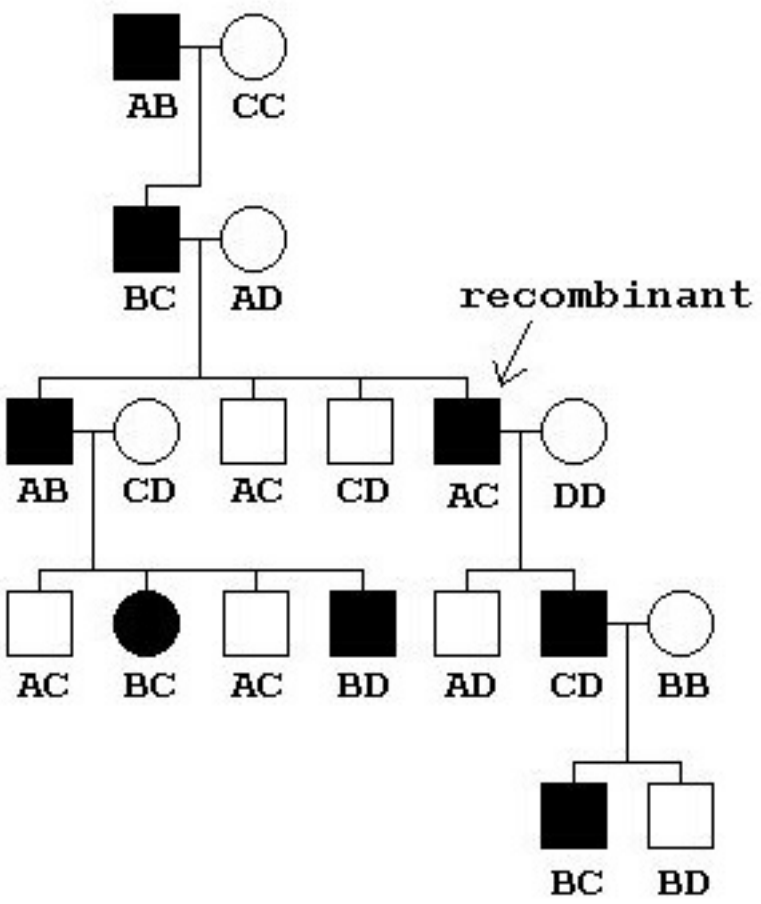
Analyzing a pedigree with marker data, Ex. 1

- Letters denote alleles at a specific marker locus, perhaps near the disease allele
- Try to identify the chromosome carrying the disease trait
- Trace it through the pedigree
- In this pedigree, disease assorts with A throughout



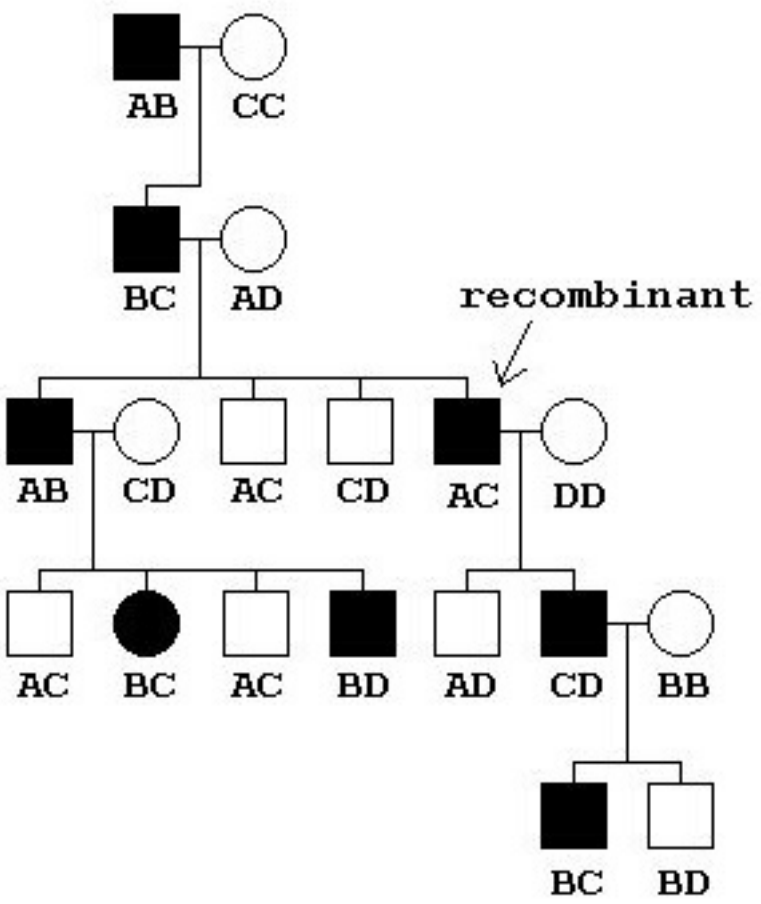
Analyzing a pedigree with marker data, Ex. 2

- Letters denote alleles at a specific marker locus, perhaps near the disease allele
- Try to identify the chromosome carrying the disease trait
- Trace it through the pedigree
- *In this pedigree, there has been a recombination*



Analyzing a pedigree with marker data, Ex. 2

- Letters denote alleles at a specific marker locus, perhaps near the disease allele
- Try to identify the chromosome carrying the disease trait
- Trace it through the pedigree
- *In this pedigree, there has been a recombination*
- And frequency of such events is related to the distance between marker and disease loci.



Recombination frequency

- Written as θ
- Percentage of transmissions in which a (newly) recombinant chromosome was transmitted
- $\theta = 0$ is perfect linkage
- $\theta = 0.5$ is no linkage

Lod score

- Lod == "Log of Odds"
- Lod score measures probability of pedigree under linkage versus no linkage hypotheses
- Normally computed using log_{10} (base 10 log)

$$Lod = \log_{10} \frac{P(\text{data} \mid \theta)}{P(\text{data} \mid \theta = 0.5)}$$

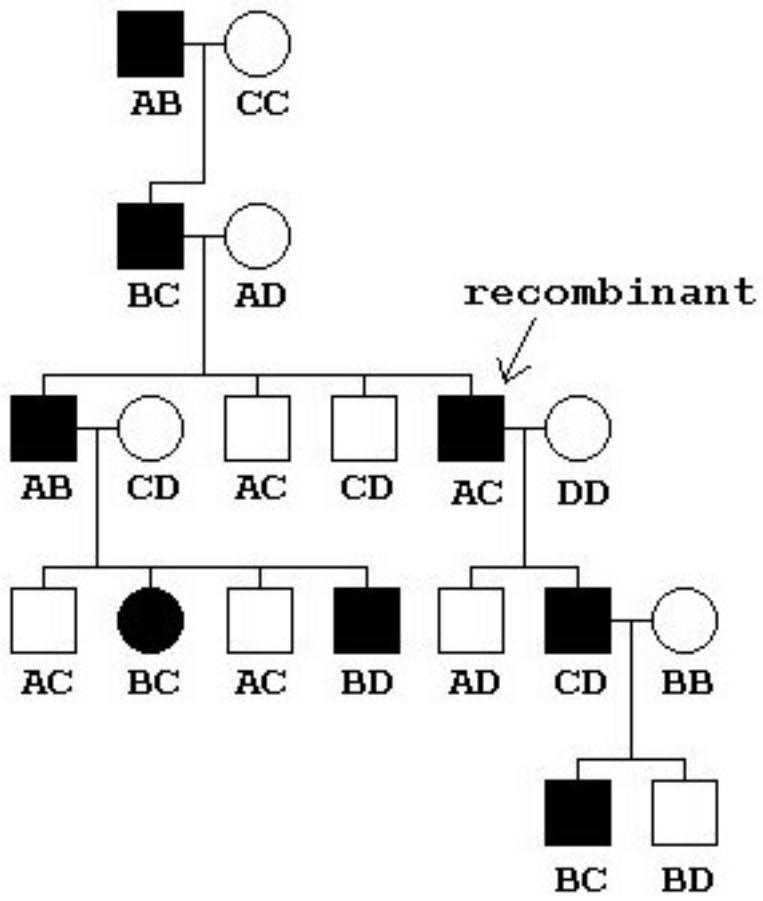
$$Lod = \log_{10} \frac{(1-\theta)^{NR} \times \theta^{R}}{0.5^{(NR+R)}}$$

- R & NR = numbers of recombinants/non-recombinants observed, resp.

Lod score example

- $R = 1$
- $NR = 11$
- $\hat{\theta} = 1/12$

$$
\begin{aligned}
Lod &= \log_{10} \frac{(1-\theta)^{NR} \times \theta^{R}}{0.5^{(NR+R)}} \\
&= \log_{10} \frac{(11/12)^{11} \times (1/12)^{1}}{(.5)^{12}} \\
&= 2.117505
\end{aligned}
$$

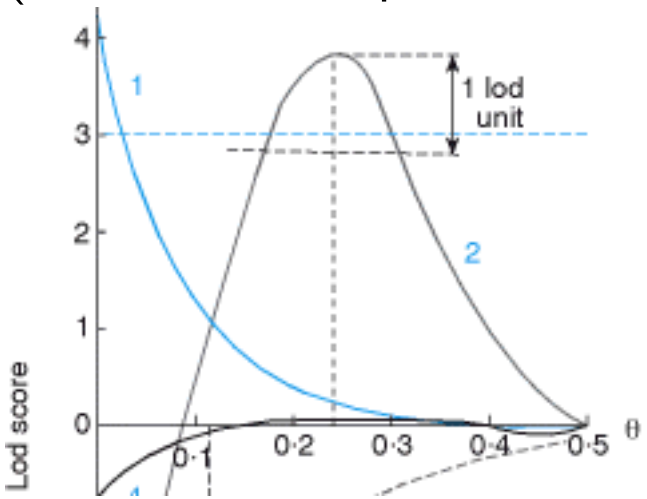


Lod score

- Lod scores can be added across families
- Value greater than 3.0 considered to show linkage
- (This is a 1 in 1000 chance–conservative but allows for multiple tests)
- Value less than -2.0 shows non-linkage (100:1 against)

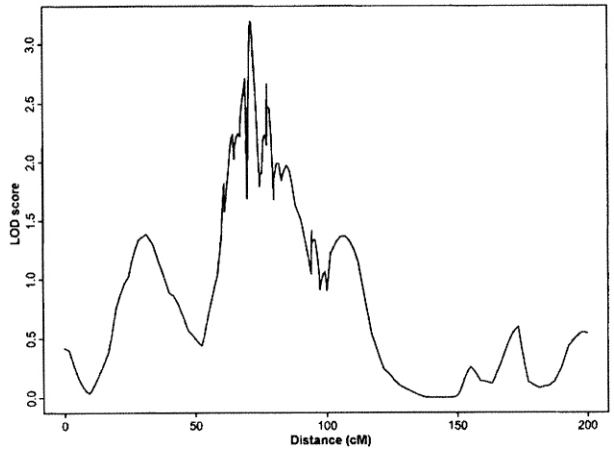
Complex Pedigrees

- When individuals are ambiguous, can sum over possibilities
- MCMC (Markov chain Monte Carlo) can be used here
- Compute Lod score for different values of θ
- Plot as a curve: maximum is most likely recombination distance. (Another example of *maximum likelihood parameter estimation.)*



from Strachan & Read Human Molecular Genetics 2 http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=hmg

Lod score curve



- More than one marker makes a better map
- Multiple densely placed markers give the most accurate map

Family linkage studies

- Advantages:
  - Reduced chance of disease heterogeneity within a family
  - Clear observation of recombinations
- Disadvantages:
  - Suitable large families rare
  - Seldom locates gene more closely than 5 cM
  - Very difficult for late-onset diseases