BIOL 345 HUMAN GENETICS THEORY

LECTURE 6

GENETIC LINKAGE AND MAPPING

How does linkage mapping differ from physical mapping?

What can you use it for?

How do you make linkage maps?

Can you use the same methods for humans as for other organisms?

Are recessive diseases as easy to map as dominant diseases?

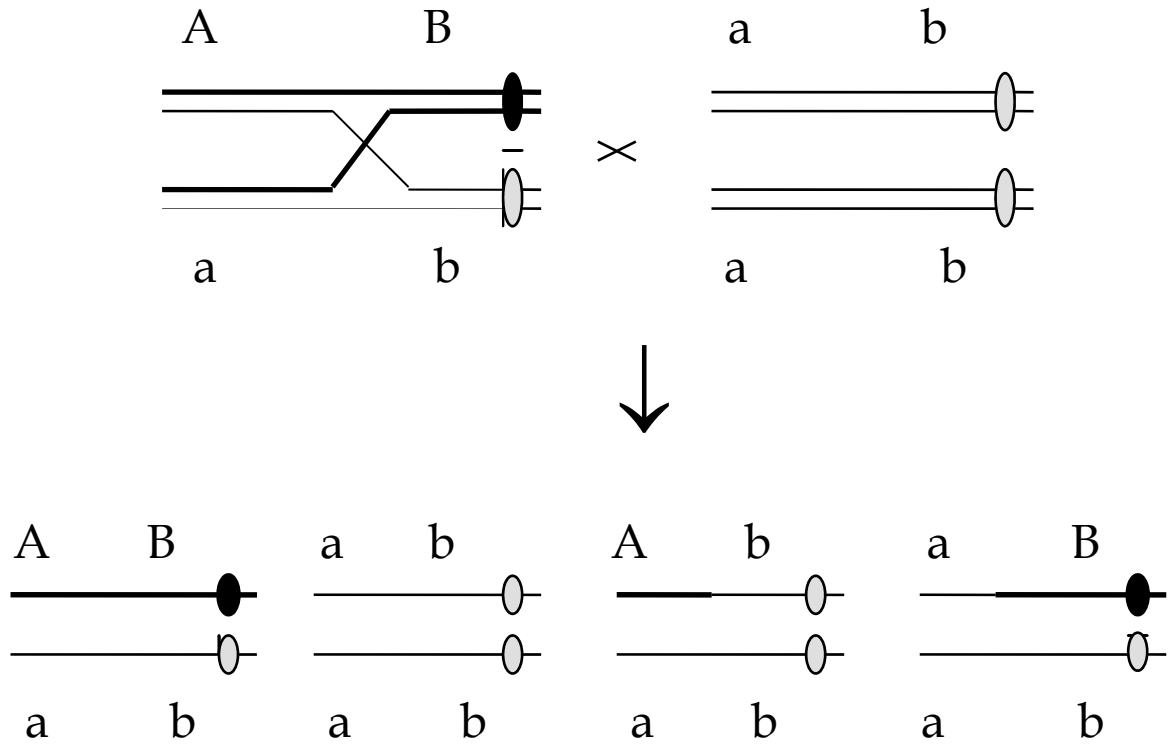
Basic principles of genetic linkage:

- Linkage results from the location of genes on chromosomes
- Gene loci on the same chromosome are inherited together (sometimes)
- Linkage is deduced from the progeny by observing the arrangement of alleles *(haplotype)*

Linkage phase: arrangement of alleles on each chromosome in a double heterozygote:

A B	A b
a b	a B
coupling	repulsion

Basic testcross for linkage:



PARENTALS RECOMBINANTS

Linkage is measured by Recombination Fraction θ

Also known as recombination rate or frequency

Proportion of recombinant gametes or recombinant offspring in a *testcross*

- e.g. 30 recombinants / 200 offspring θ = 0.15 or 15%

θ = r = 1% = 0.01 = 1 map unit ≈ 1 centiMorgan (1 cM)

Linkage maps

Differences in recombination rate between pairs of loci used to determine gene order in a linear map

The further apart 2 loci are on a chromosome, the greater the chance for a crossover between them.

Maximum value of θ is 0.5, independent assortment → genes far apart on a chromosome may not show linkage

N.B. Linkage is how genes <u>behave</u> together

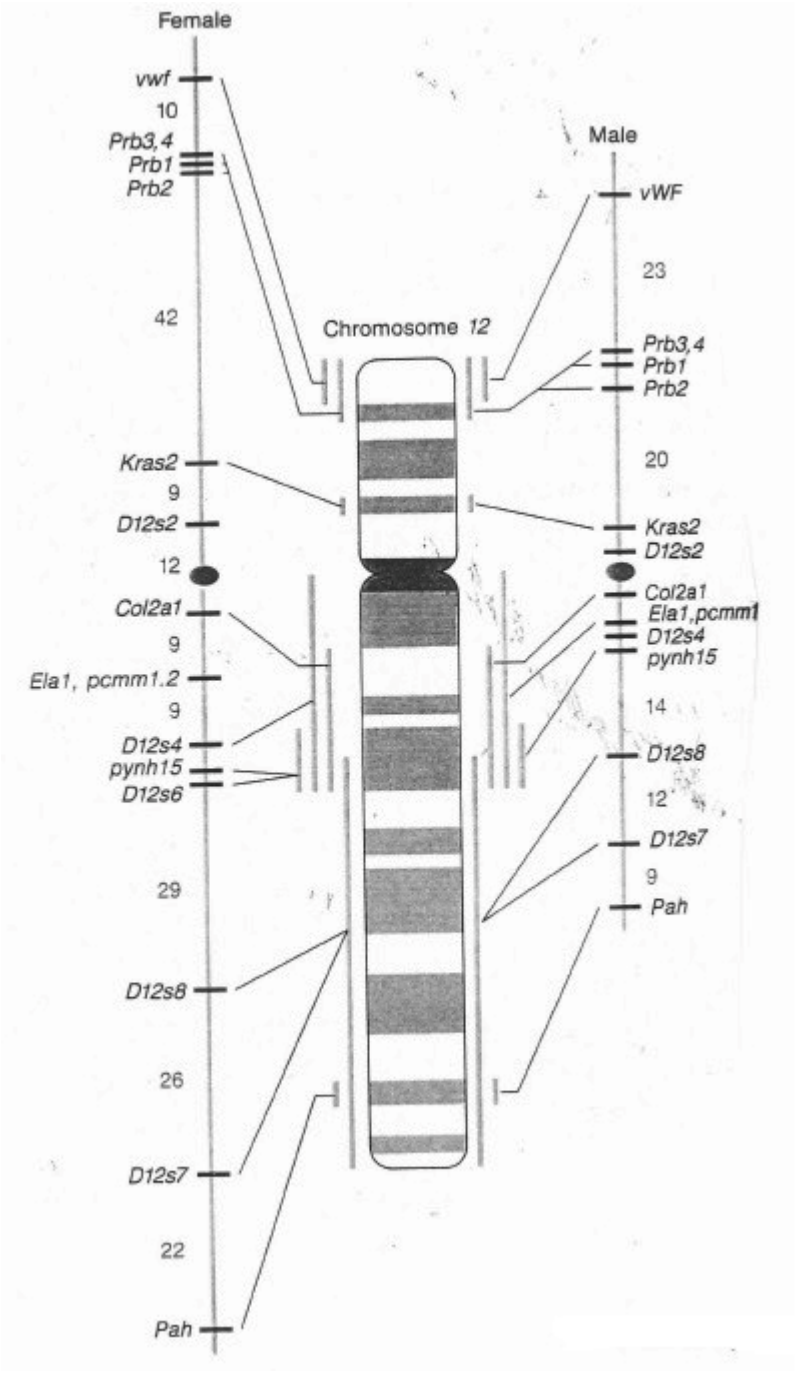
Genes located on same chromosome (irrespective of linkage are *syntenic*

Limitations on working out maps from 2 point data:
- small sample sizes → value of θ is only an estimate
- small errors in θ → map may not add up perfectly
- hard to work out order for very close or very far loci

Solution in experimental organisms is a 3 point testcross

Recombination is not constant along a chromosome

Human chromosome 12 linkage maps



- Recombination is not uniform along a chromosome: greater near telomeres - distribution of chiasmata?
- In general, greater recombination in females than males, except near telomeres

GENETIC MARKERS

Linkage can be used to localize a gene by assessing linkage with genetic markers of known position

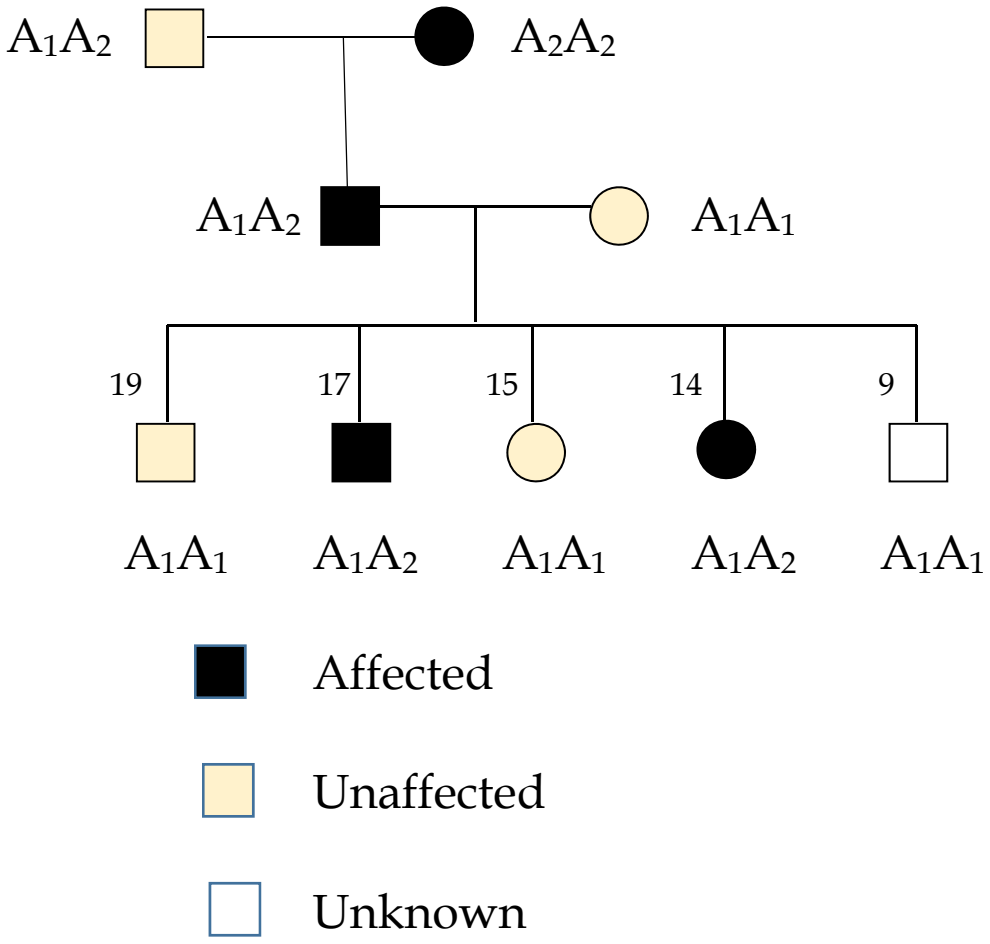
TYPE	DETECTION
morphological, behavioural (e.g. eye colour, taste)	inspection, responses
antigens (blood groups, HLA)	immunological
chromosome heteromorphisms	cytogenetic
protein, enzyme variants (serum proteins, isozymes etc)	electrophoresis
RFLPS	Southern blotting
VNTRs	Southern blotting
minisatellites	Southern blotting
microsatellites	PCR
SNPs	DNA sequencing microarrays

Factors determining their usefulness:

1. Informativeness
2. Number
3. Ease of use
4. Expense

Using recombination fraction as a probability

eg Late onset muscle weakness – onset at puberty
Genetic marker linked at $\theta = 0.15$



Will III.5 develop the disease?

TESTING FOR LINKAGE

A strain of mice homozygous for dominant alleles *AABB* is crossed to a strain recessive at both loci *aabb*. The F1 is backcrossed to the recessive parent, producing 115 *AaBb*, 88 Aabb, 86 *aaBb*, and 111 *aabb* offspring. Are these loci linked?

One might be tempted to answer by saying independent assortment predicts a 1:1:1:1,
for which χ^2_3 =

$$\frac{(115-100)^2}{100} + \frac{(88-100)^2}{100} + \frac{(86-100)^2}{100} + \frac{(111-100)^2}{100}$$

$$= \quad 6.86, \ P > 0.05$$

However, this can be broken down to test 3 hypotheses, i.e.

	χ^2_1	p	
A:a = 1:1	0.09	>0.05	
B:b = 1:1	0.01	> 0.05	
AB + ab : Ab + aB = 1:1	6.76	<0.05	NR : R

Total 6.86

LINKAGE ANALYSIS IN HUMAN FAMILIES

Problems in estimating linkage arise from:

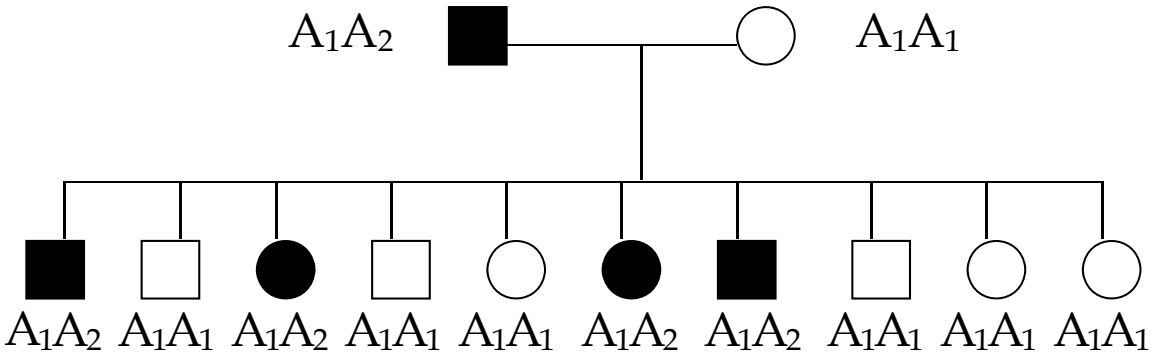
1. family size:
    - generally inadequate to get accurate estimate of θ
    - How do you pool data between families?
2. family structure:
    - often lacking information on phase
    - How do you combine information between phase known and phase unknown families?

Landmarks in human linkage analysis

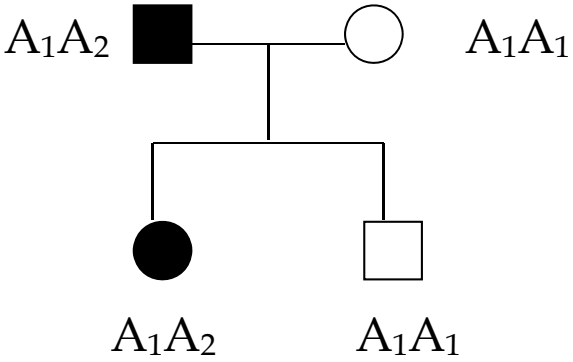
Haldane & Smith (1947) probability ratio test – the probability of seeing the family if the genes were linked compared to the probability if they were unlinked

Barnard (1949): log of probability ratio
lod (log odds) score

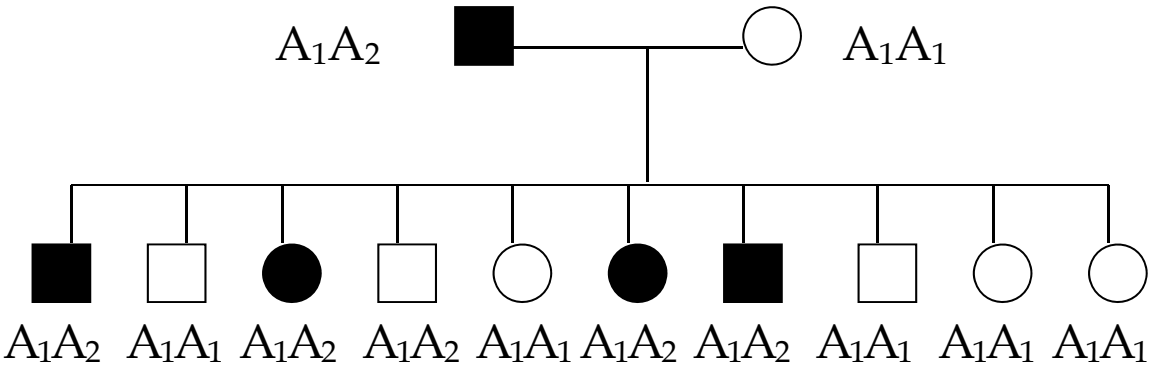
Morton (1955) sequential test using probability ratio



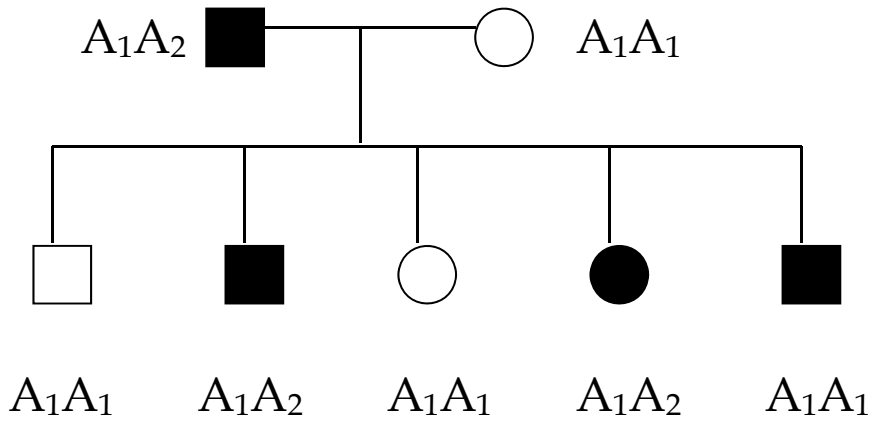
No recombinants



No recombinants



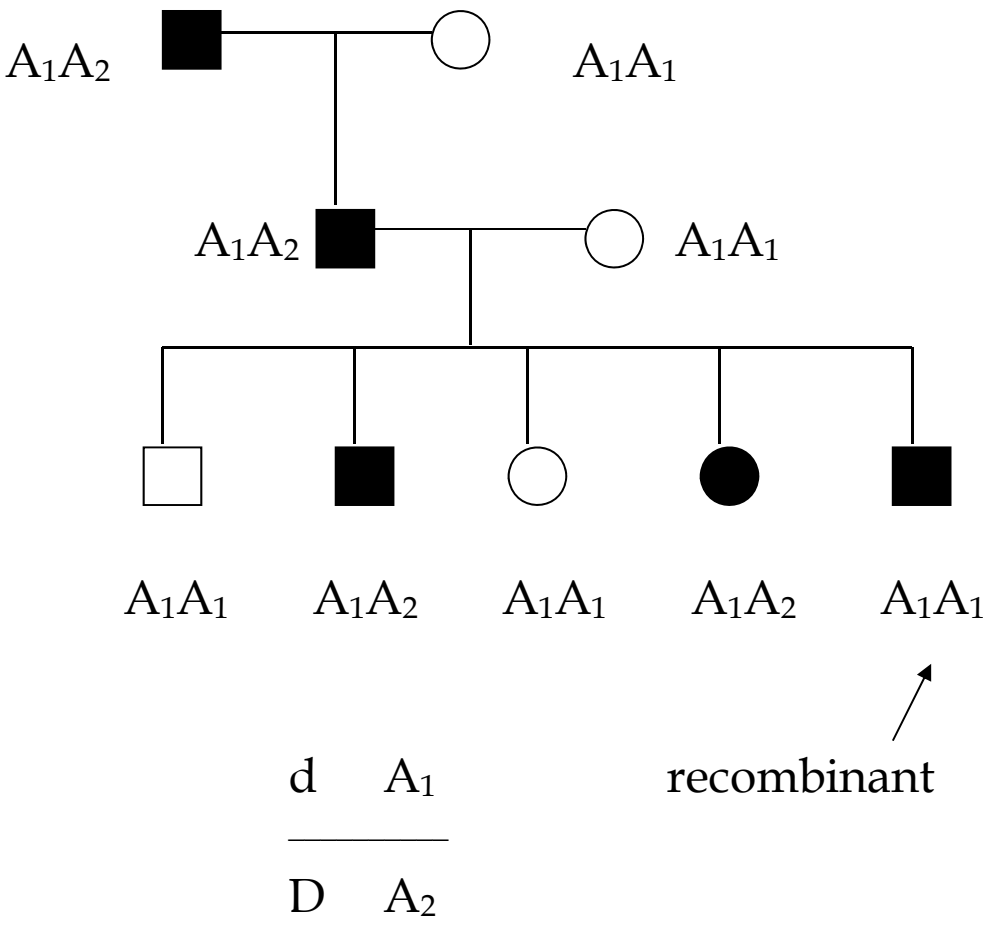
1/10 recombinants



D	A_1	1
d	A_2	0

d	A_1	2
D	A_2	2

Either 4 recombinants and 1 non-recombinant
Or 1 recombinant and 4 non-recombinants


A1A2
A1A1
A1A2
A1A1
A1A1
A1A2
A1A1
A1A2
A1A1
recombinant
d A1
D A2

LOD SCORES

lod = log odds = log probability ratio

$$\text{lod} = \log_{10} \frac{\text{Prob of pedigree if linked, } \theta = x}{\text{Prob of pedigree if unlinked, } \theta = {}^1/_2}$$

Example:

3 generations, phase known

A_2 D 4nr : 1r prob = $\theta (1 - \theta)^4$

2 generations, phase unknown

A_2 D 4nr : 1r prob = $\theta (1 - \theta)^4$

A_1 D 1nr : 4r prob = $\theta^4 (1 - \theta)$

$$\text{Average probability} = \frac{\theta (1 - \theta)^4 + \theta^4 (1 - \theta)}{2}$$

Then substitute values for θ and divide by probability of seeing the family if genes are unlinked ie $\theta = {}^1/_2$

N.B. Have to work out probability expression for each family

Lods (Z)

3 generation family, phase known 4nr : 1 r

θ	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.50
	- ∞	0.12	0.32	0.40	**0.42**	0.40	0.36	0.0

2 generation family, phase unknown 4 : 1

θ	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.50
	- ∞	-0.19	0.02	0.10	**0.12**	0.11	0.09	0.0

3 generation family, phase known 5nr : 0 r

θ	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.50
	1.50	1.40	1.28	1.15	1.02	0.88	0.73	0.0

Combined lods for 3 families

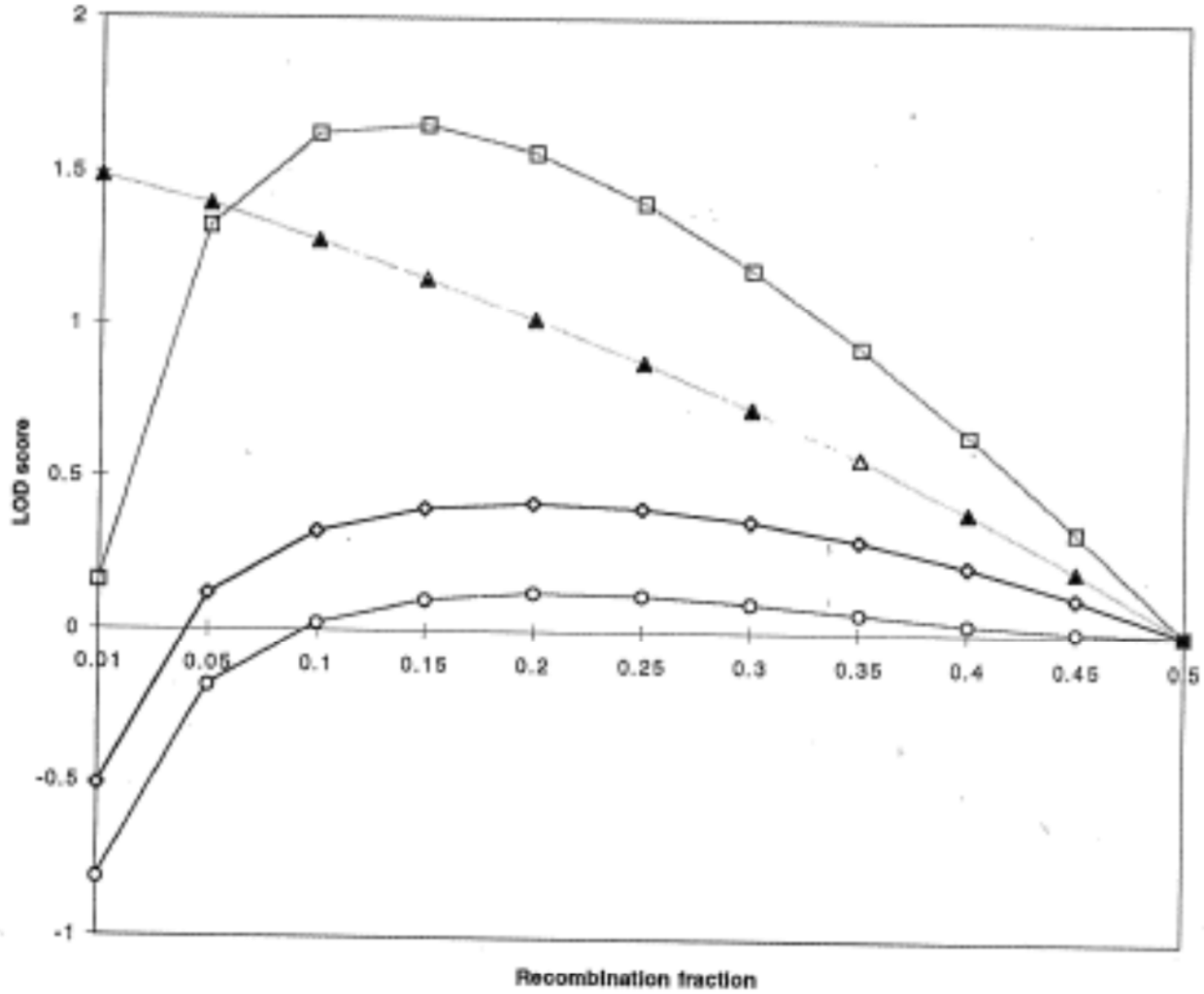
θ	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.50
	- ∞	1.33	1.62	**1.65**	1.56	1.39	1.18	0.0

Because they are lods of probabilities, lods are <u>additive</u>

lod > 3 → linked

lod < -2 → not linked

-2 < lod < 3 → wait; collect more families

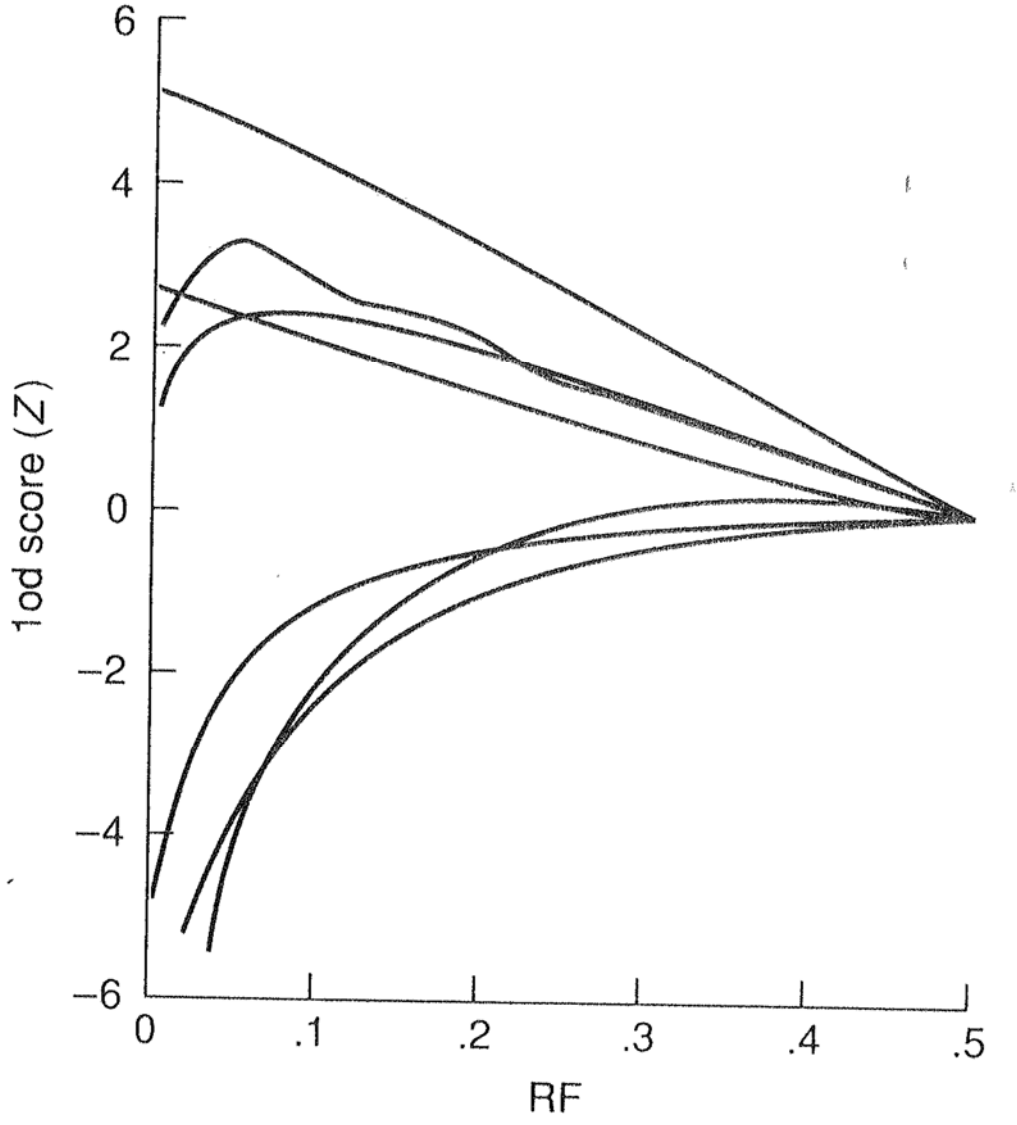

2
1.5
1
0.5
0
-0.5
-1
LOD score
0.01
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5
Recombination fraction
4 nr : 1 r (3 Gen)
4nr : 1r (2 Gen)
5 nr : 0 r (3 Gen)
Total

Genetic heterogeneity

Before pooling pedigrees, check for:

- consistency (e.g. paternity)
- segregation

Morton (1956) Rh and elliptocytosis



Adapted from Morton (1956) *Am J Hum Genet.* **8**: 80–96

Problem: in the presence of locus heterogeneity, how to tell which gene is being transmitted in a moderately small family?

The problem of gene order

A---B $\theta \approx 0.11$
B---C $\theta \approx 0.03$
A---C $\theta \approx 0.12$

Most probable gene order is A----B----C BUT

- estimates of θ have wide confidence limits
- smaller the dataset, wider the limits

The solution: multipoint mapping programs

Linkage analysis with autosomal recessive diseases

Linkage analysis is much harder for recessive diseases because:

- Genotype of normal offspring unknown
  $\therefore$ less information available than for dominant genes
- Often smaller families

The solution:

- Look for consanguineous pedigrees with rare recessive diseases

Autozygosity mapping [homozygous identical by descent]

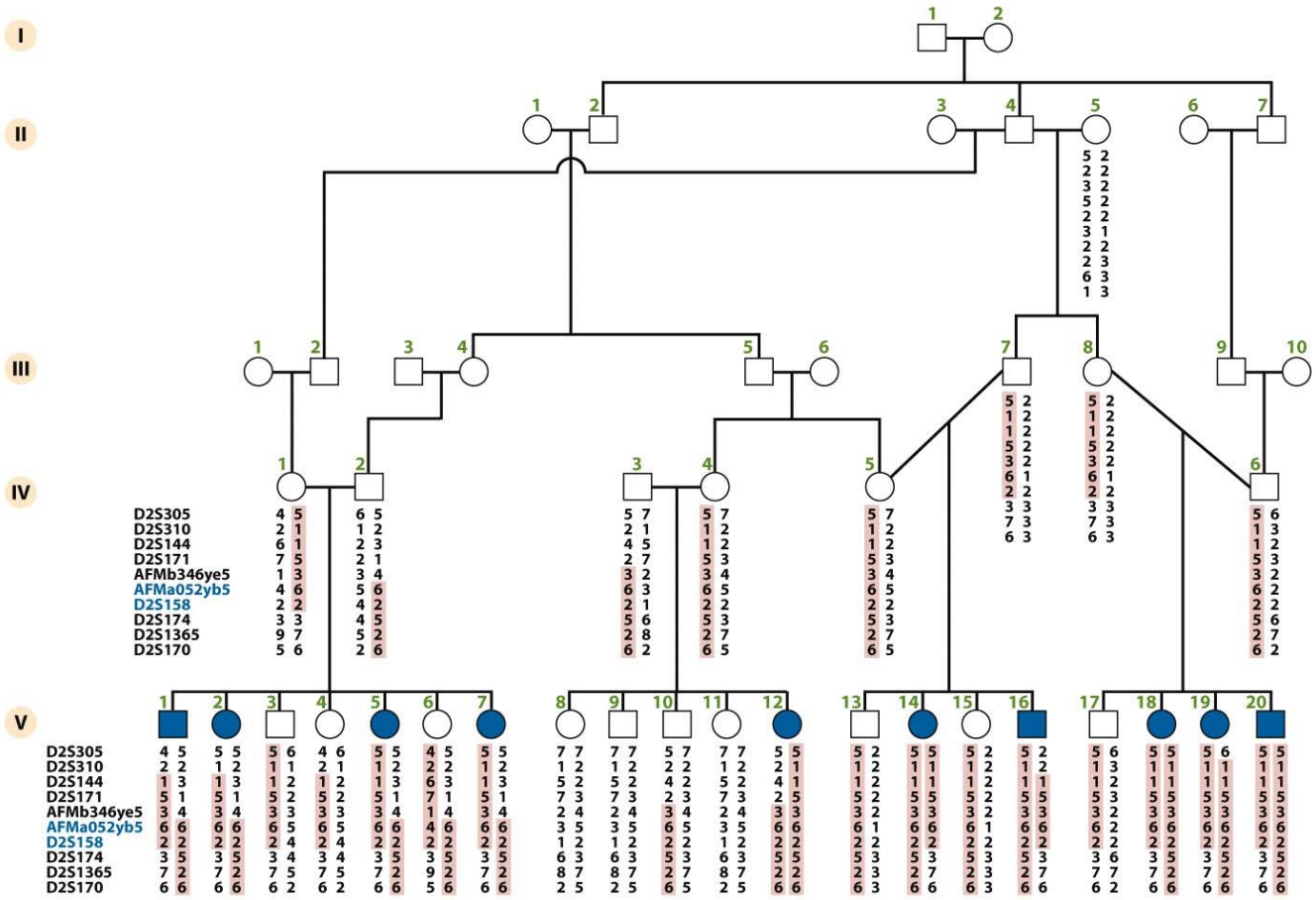


Figure 14.13 Human Molecular Genetics, 4ed. (© Garland Science)

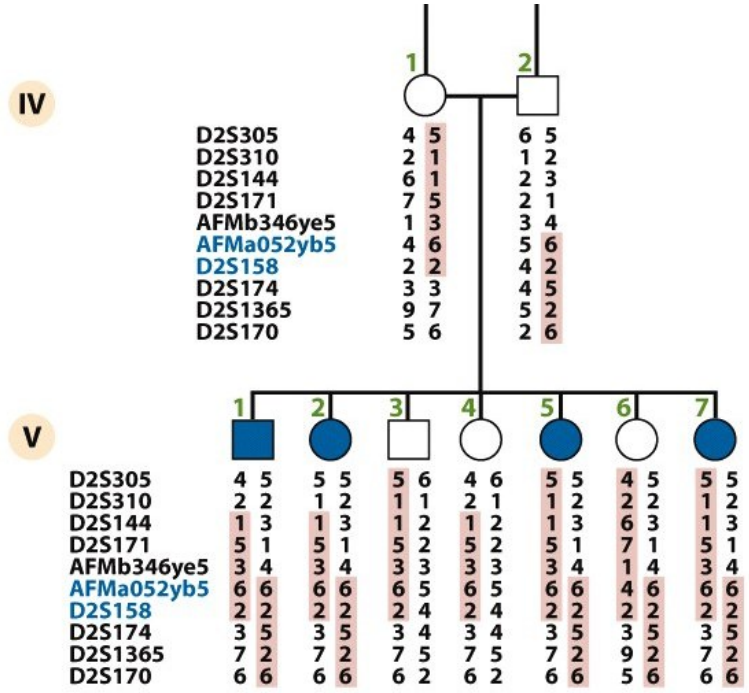


Figure 14.13 Human Molecular Genetics, 4ed. (© Garland Science)

Roscioli et al (2006) Mutations in the gene encoding the PML nuclear body protein Sp110 are associated with immunodeficiency and hepatic veno-occlusive disease. *Nature Genetics* **38**: 620 - 622