# Lecture 6:
# Linkage analysis in medical genetics

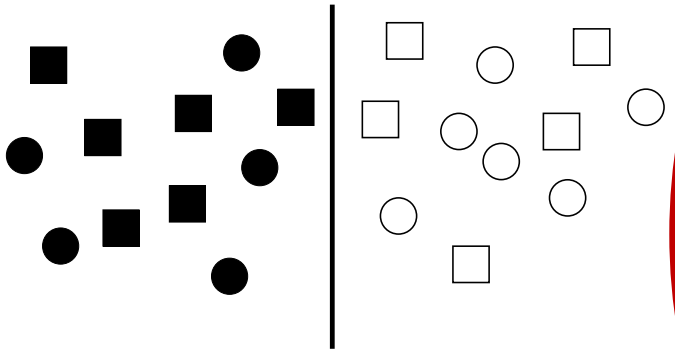
Magnus Dehli Vigeland

**Statistical methods in genetic relatedness and pedigree analysis**

NORBIS course, 6th – 10th of January 2020, Oslo

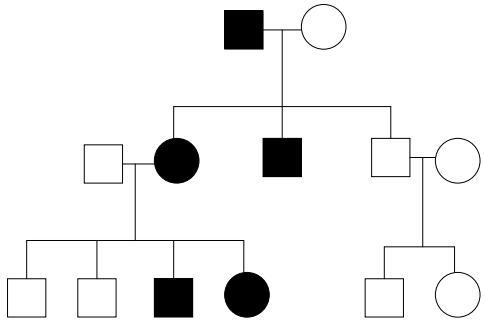
# Approaches to genetic mapping of disease

Multifactorial disease



- Association analysis
- Case/control design
- Population based

Monogenic disease

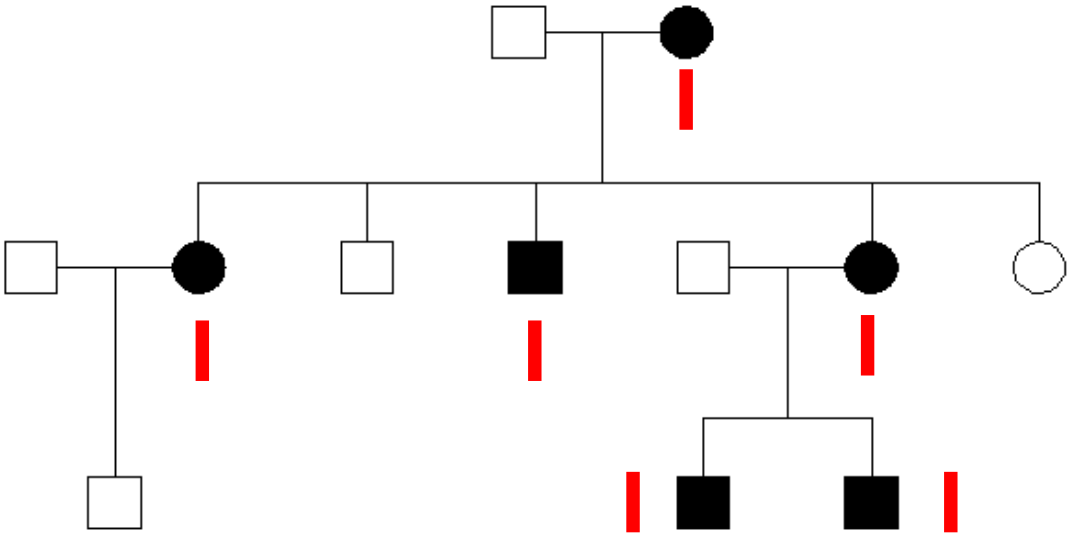


- Linkage analysis
- Family design

# Linkage analysis – the basic principle

Compare

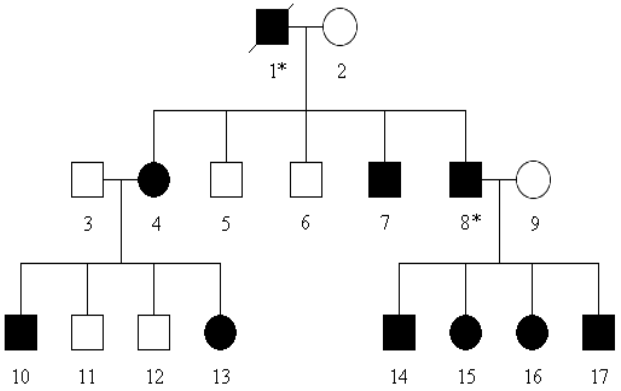
- the inheritance pattern of a disease
- the inheritance pattern of chromosomal regions



Goal: Identify the region harbouring the disease-causing locus

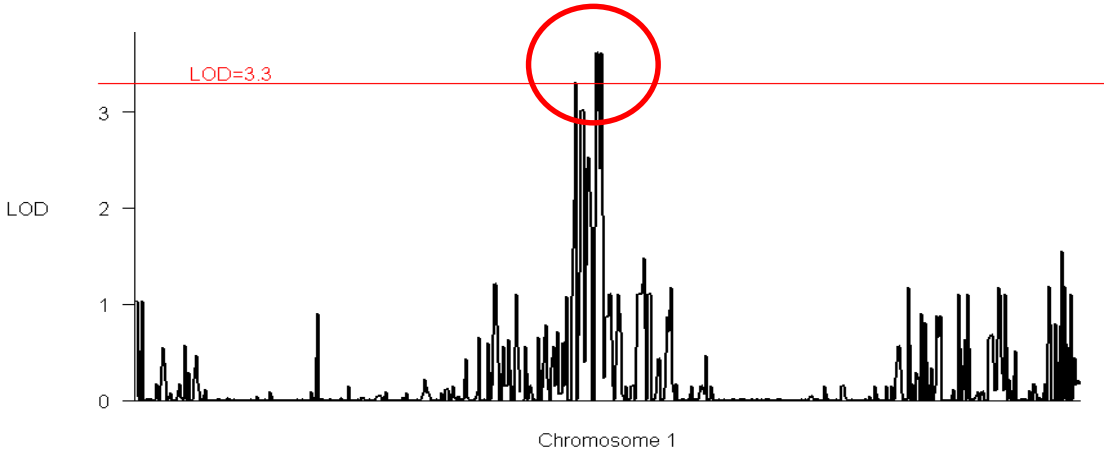
# Linkage analysis - traditional workflow

1. Collect (large) affected families




2. SNP genotyping

3. Parametric linkage analysis



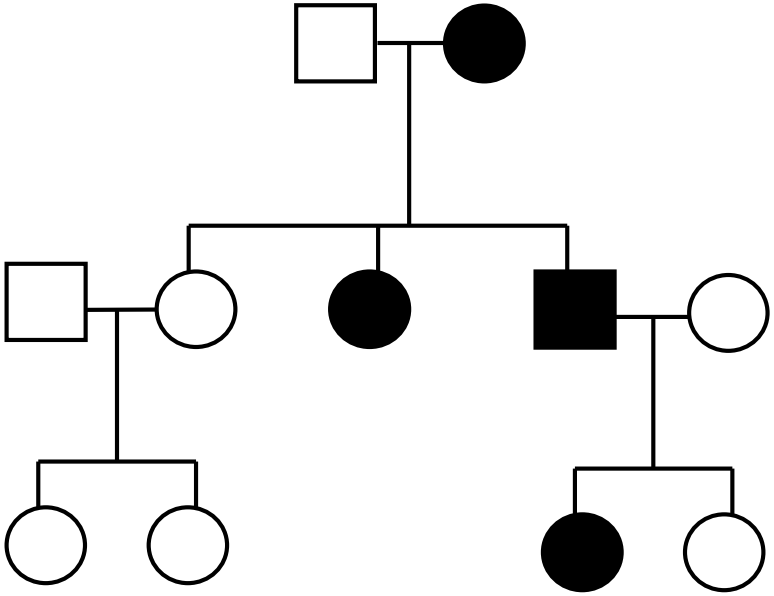

4. Sequence genes in linkage peak → identify causal mutation

# Inheritance patterns for monogenic disorders

(Part of the statistical model - needed in the likelihood computations)

# Autosomal dominant (AD) inheritance

- ~50% affected children of an affected mother or father
- Affected male:female ratio is ~1
- Can be inherited from mother or father to both sons and daughters



Penetrance values:

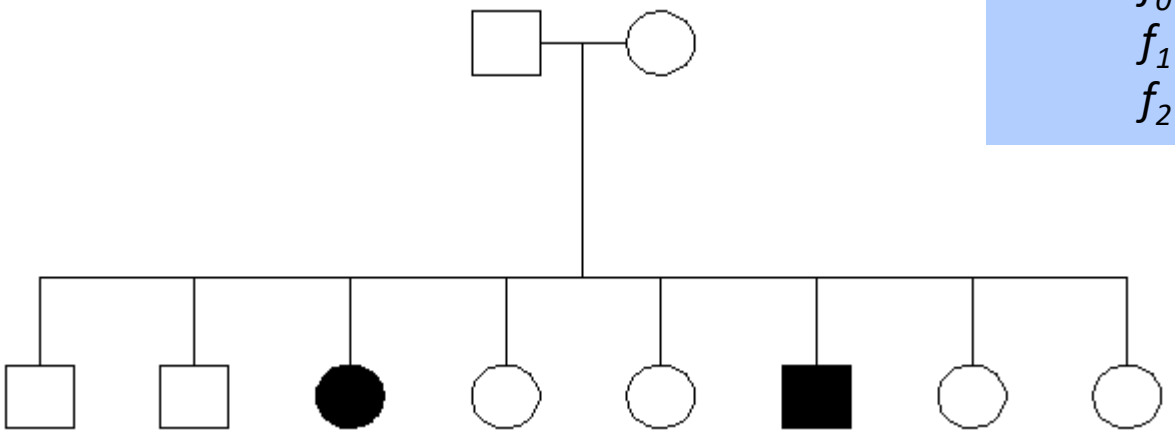
$f_0 = 0$

$f_1 = 1$

$f_2 = 1$

# Autosomal recessive (AR) inheritance

- Usually healthy parents
- ~25% of the children are affected
- Affected male:female ratio is ~1



Penetrance values:

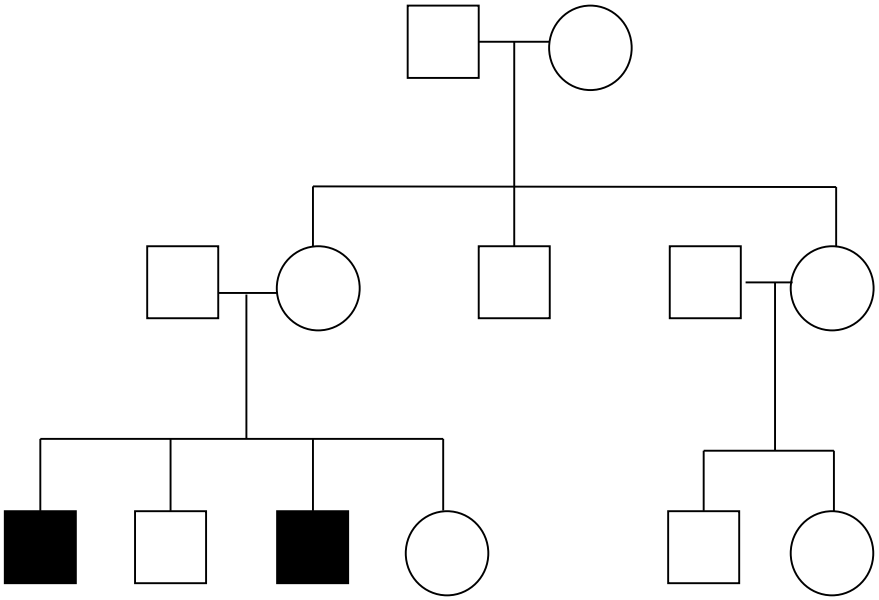
$f_0 = 0$

$f_1 = 0$

$f_2 = 1$

# X-linked recessive inheritance

- Only males are affected
- Usually inherited through healthy females



Penetrance values:

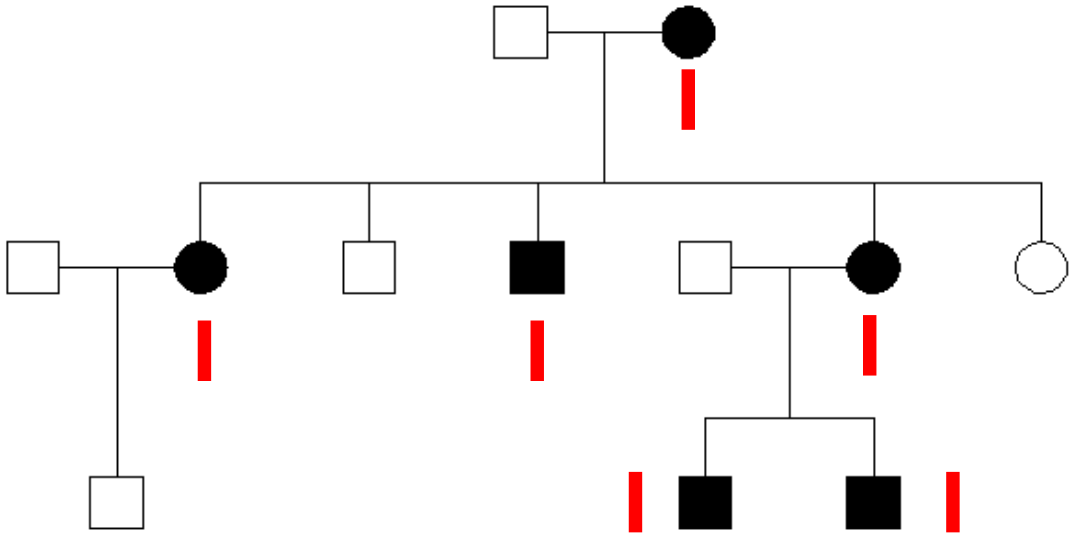
Females:

$f_0 = 0$
$f_1 = 0$
$f_2 = 1$

Males:

$f_0 = 0$
$f_1 = 1$

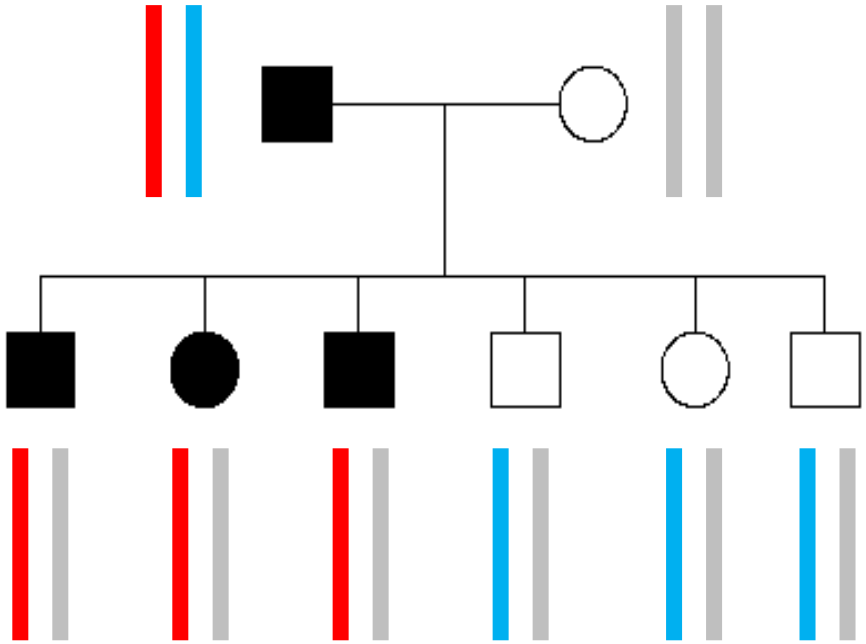
# Linkage analysis: Main challenges



- When faced with a region co-segregating with the disease: Is it by chance? (Evaluation of statistical significance)

- How to use markers to track inheritance patterns of chromosomal regions?

# Linkage and significance



Red haplotype fits perfectly with the dominant disease pattern. Coincidence?

Resembles tossing a coin which is either

- normal ↔ no linkage
- equal on both sides ↔ linkage

6 heads in a row!

Framework for evaluating significance:
**Statistical hypothesis testing**

*P-value = P(6 heads | normal coin)*
*= 0.015*

*Significant?*

# Hypothesis testing in linkage

- Hypotheses:

  $H_0$: $\theta = 0.5$ (no linkage)

  $H_A$: $\theta < 0.5$ (linkage)

θ = recombination rate between marker and disease

- For historical reasons the test statistic is

$$LOD = \log_{10} \frac{P(data \mid \theta = \hat{\theta})}{P(data \mid \theta = 0.5)}$$

LOD = "logarithm of the odds"

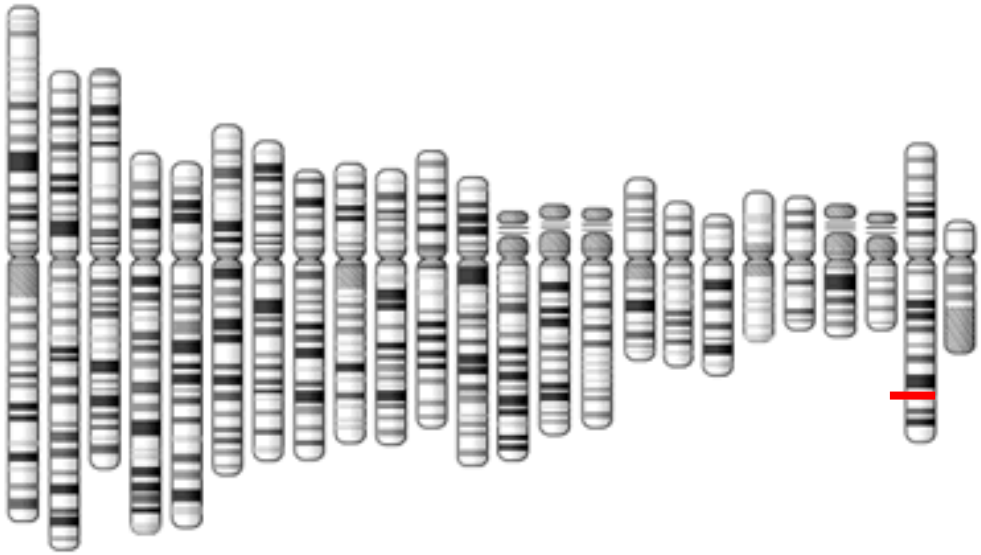
- Traditional significance thresholds:
  - Autosomal loci: LOD = 3 ($p \approx 0.0001$)
  - X-linked loci: LOD = 1.8

- Note: The "standard" significance level $\alpha = 0.05$ is not used in linkage analysis
- Reason: *a priori* low probability of $H_A$

$H_0$: $\theta = 0.5$ (no linkage)

$H_A$: $\theta < 0.5$ (linkage)

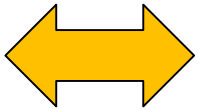
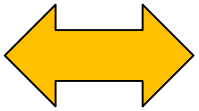
$P(H_A) \approx 1/50$



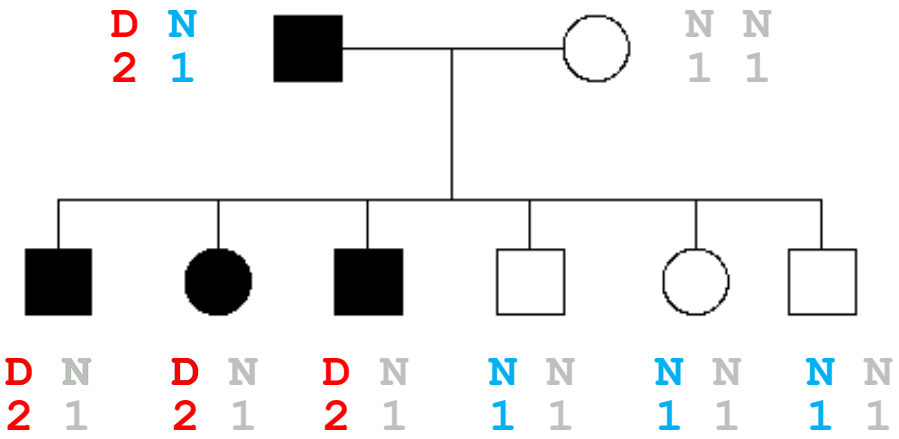
© Clinical Tools, Inc.

# Linkage analysis with markers: Key idea

Treat the disease locus as a marker

Close markers  Rarely any crossovers between them  Low recombination rate

# Linkage between marker and disease locus




- 6 informative meioses
- 0 recombinants

→ LOD = 6 log2 = 1.8

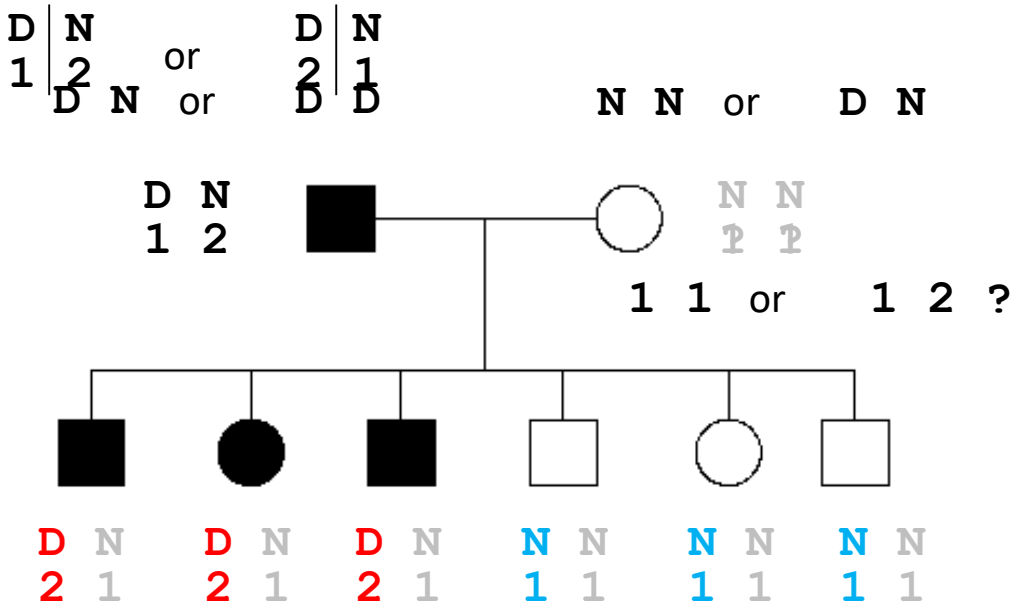
- Disease locus:
  - alleles **D** and **N**
- Marker locus:
  - alleles 1 and 2
- Model:
  - dominant
  - fully penetrant
  - no phenocopies
  - very rare

# Complicating factors in practice

*In principle* computing LOD scores is very simple:

- Count meioses/recombinations
- Use binomial probabilities

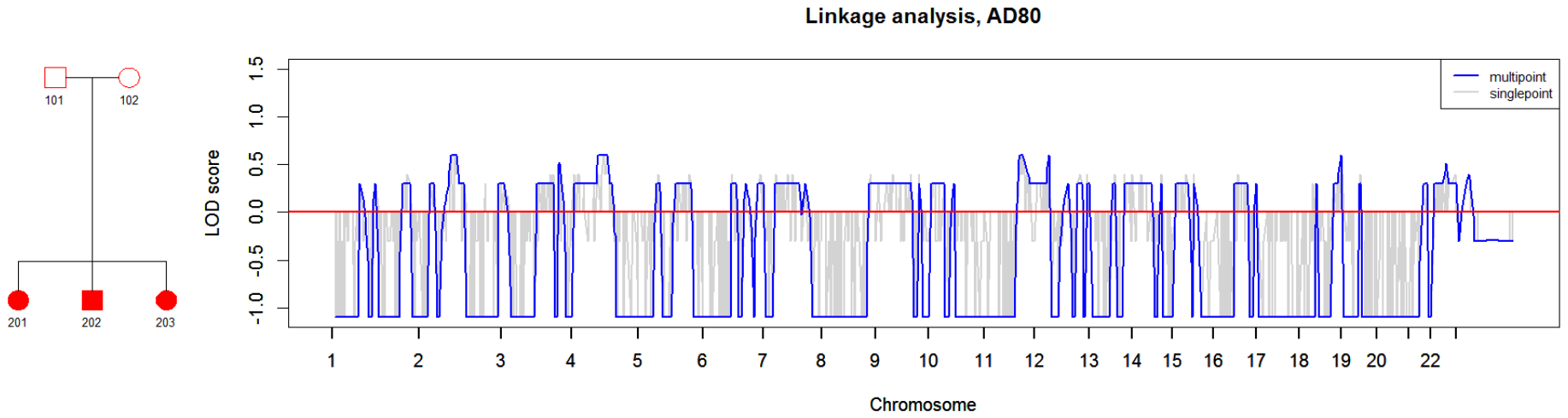
In practice: Impossible to do by hand!



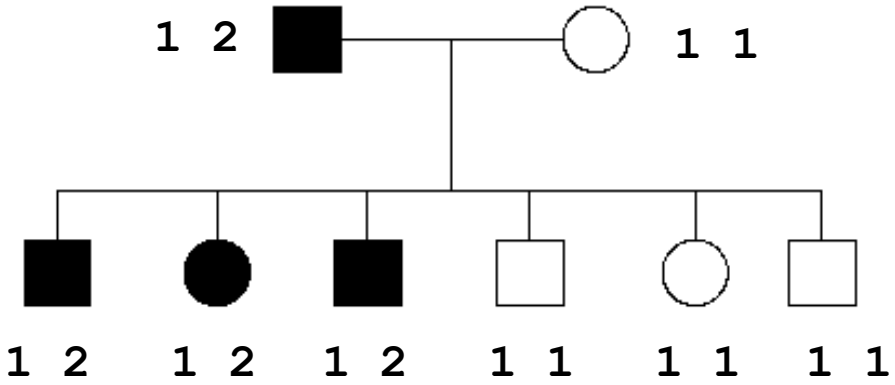

- Unknown phase
- Missing genotypes
- Uncertain genotype at disease locus

# Multipoint analysis

- Combines data from consecutive markers.
  - More powerful than single-point
  - Lander-Green algorithm (hidden Markov model)
  - Very computer intensive! Software: MERLIN
  - Paramlink can run MERLIN and postprocess the results



# LOD scores in R/paramlink




```
>   # Set AD model
>   x = setModel(x, model = 1, dfreq = 0.0001)

>   # Compute lods
>   lod(x)
```

That's it - now you try for yourself!