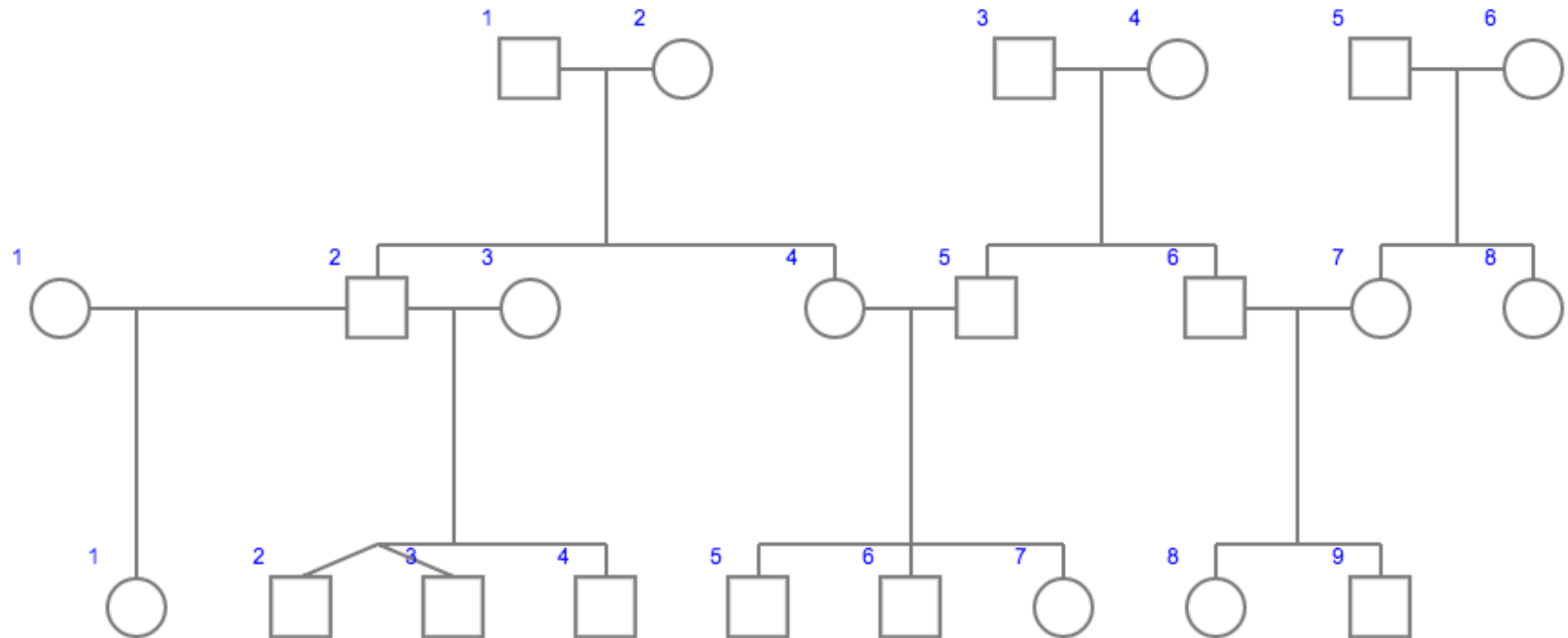
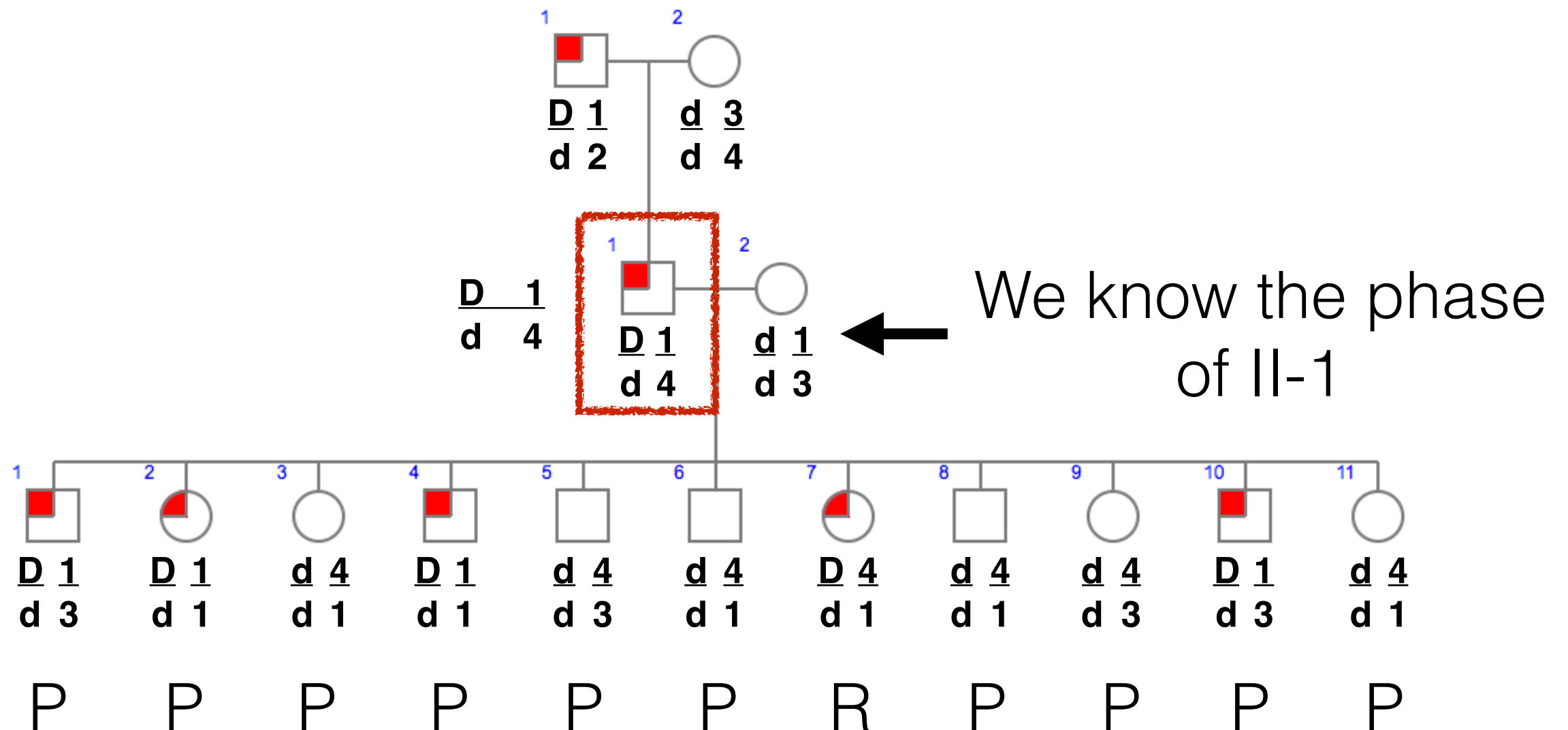


Bio393: Genetic Analysis

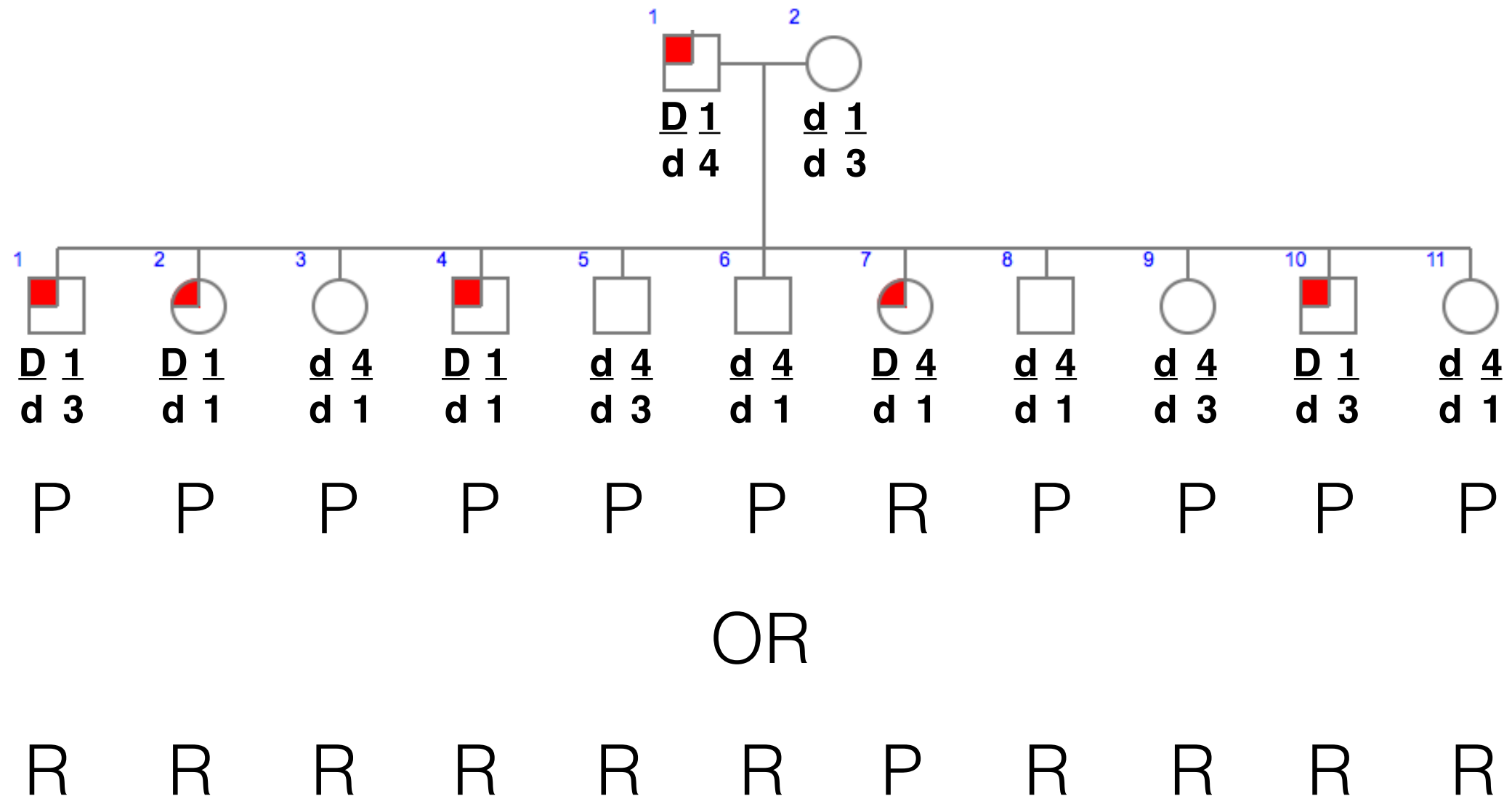
Linkage mapping in families



Linkage to genetic markers tells us where disease genes are



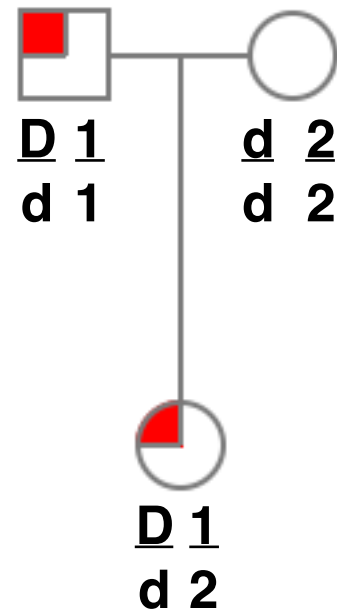
Linkage to genetic markers tells us where disease genes are



Sometimes, we don't know the phase of the parent, and both possibilities of phase are equally likely

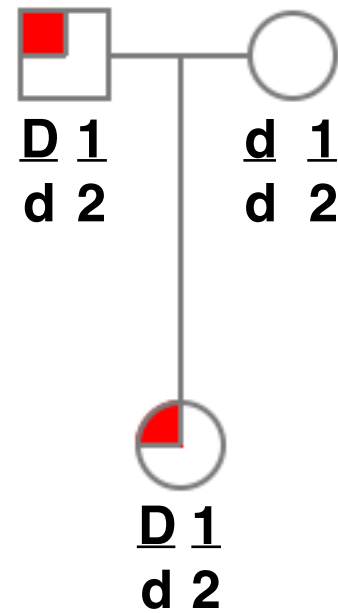
Some allelic combinations are non-informative and can not be included in mappings

Consider a dominant trait and a variant marker:



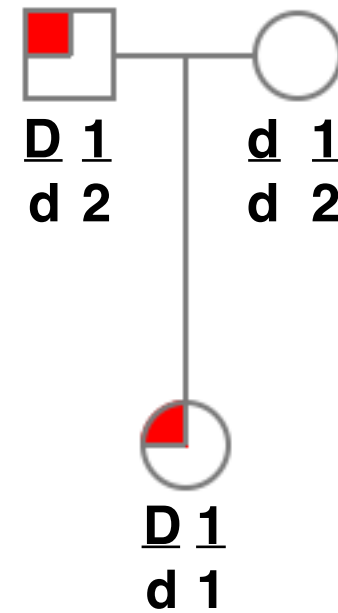
Uninformative

Which marker allele
is linked
to disease allele?



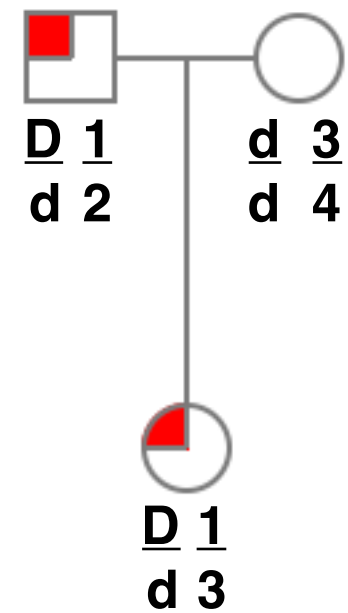
Uninformative

Which allele 1
is inherited
from Dad?



Informative

Each allele 1
must have come
from Mom and Dad

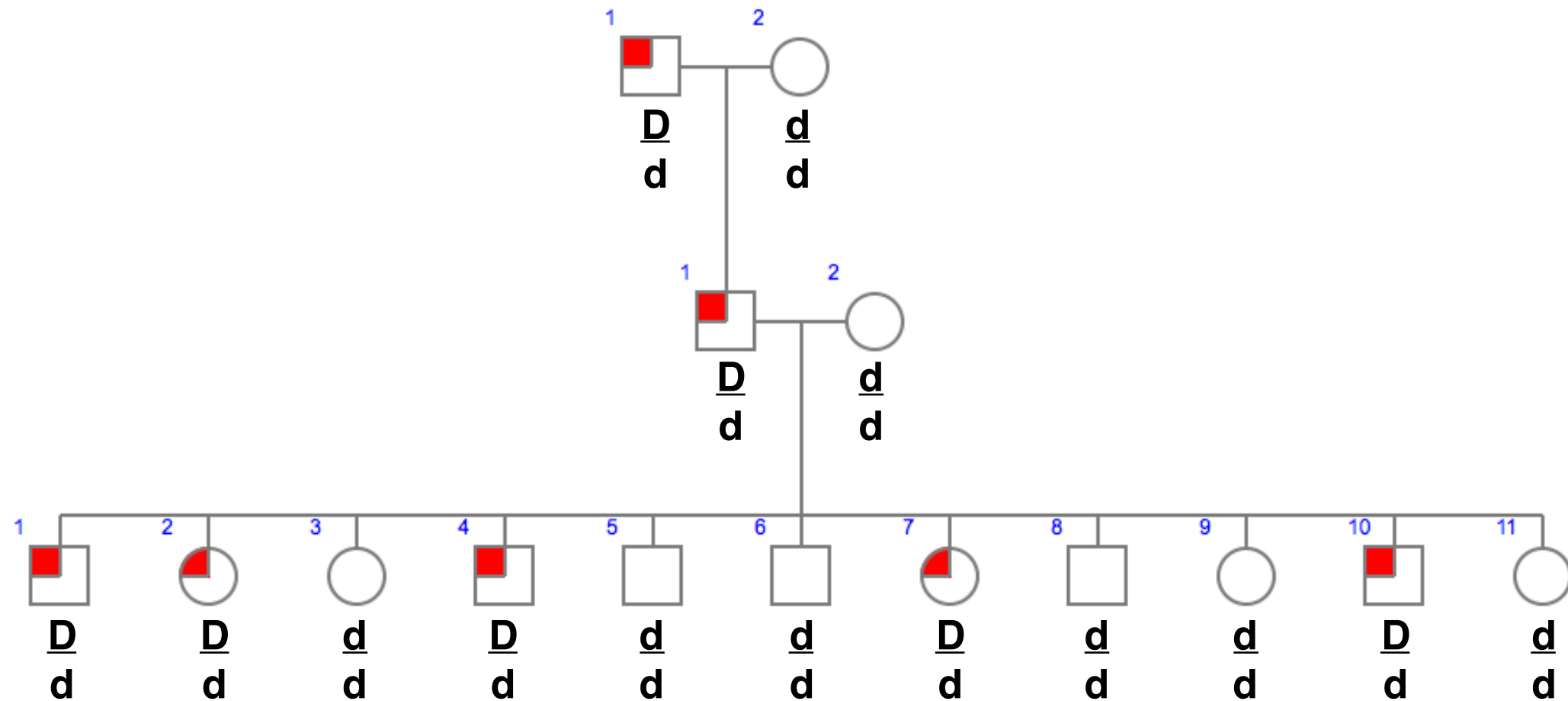


Informative

Different marker
alleles

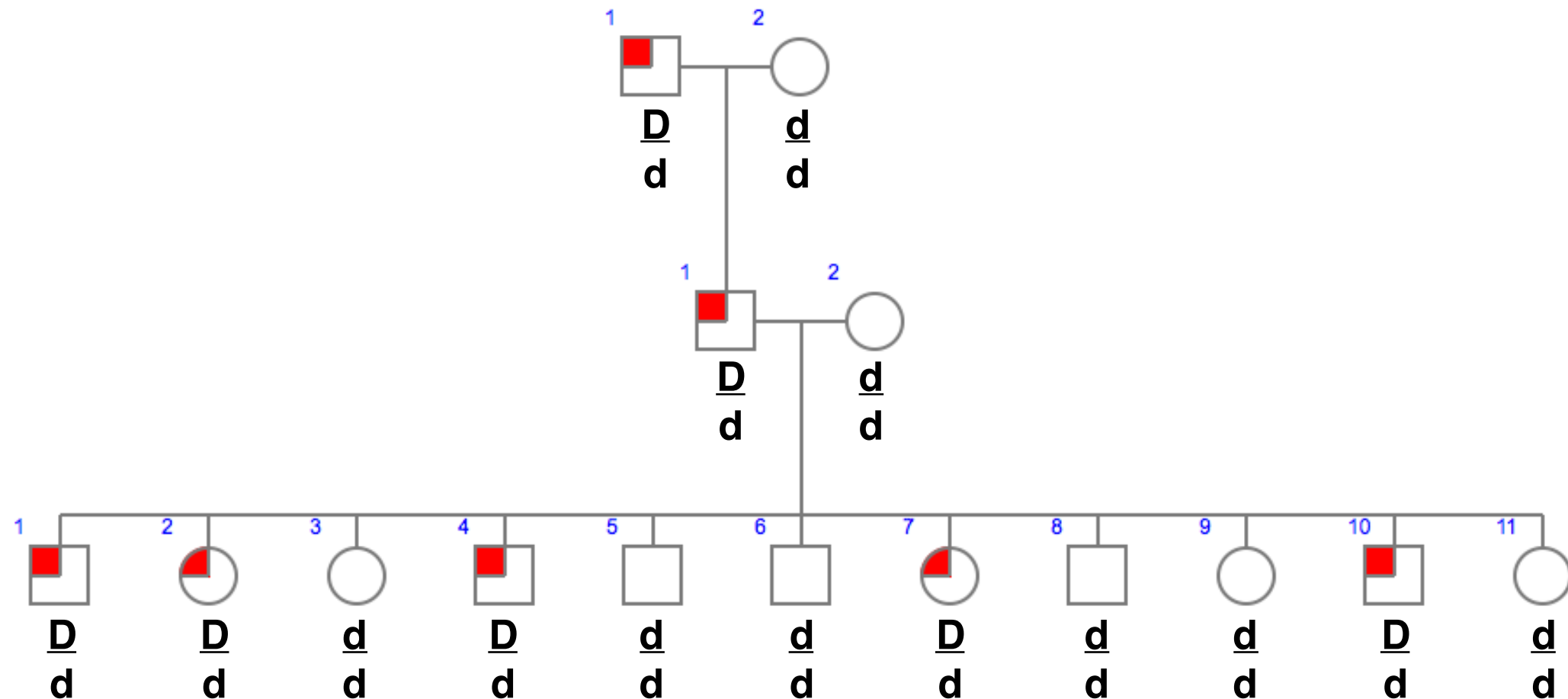
We want to determine if the daughter inherited a recombinant or parental chromosome

**Imagine you could genotype millions of markers
in each individual**



The goal is to measure linkage of many markers
to the disease-causing allele to map the gene

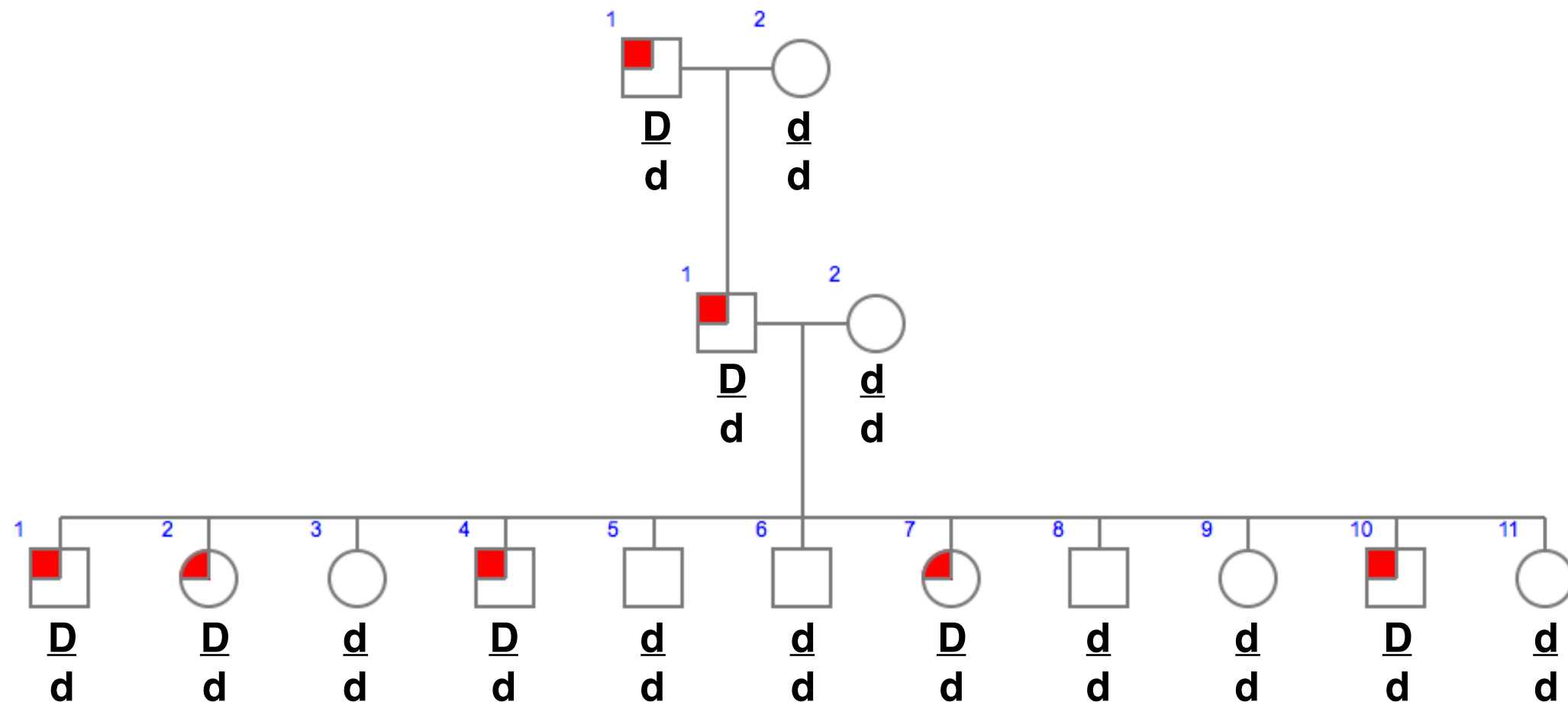
We want to measure how close each marker is to the disease-causing allele



$$\frac{\text{Number of recombinants}}{\text{Total progeny}} \times 100 = \text{Recombination frequency}$$

But we don't know who is a recombinant because we don't have true-breeding strains

We want to measure how close each marker is to the disease-causing allele



Likelihood of linkage between marker and disease-causing allele

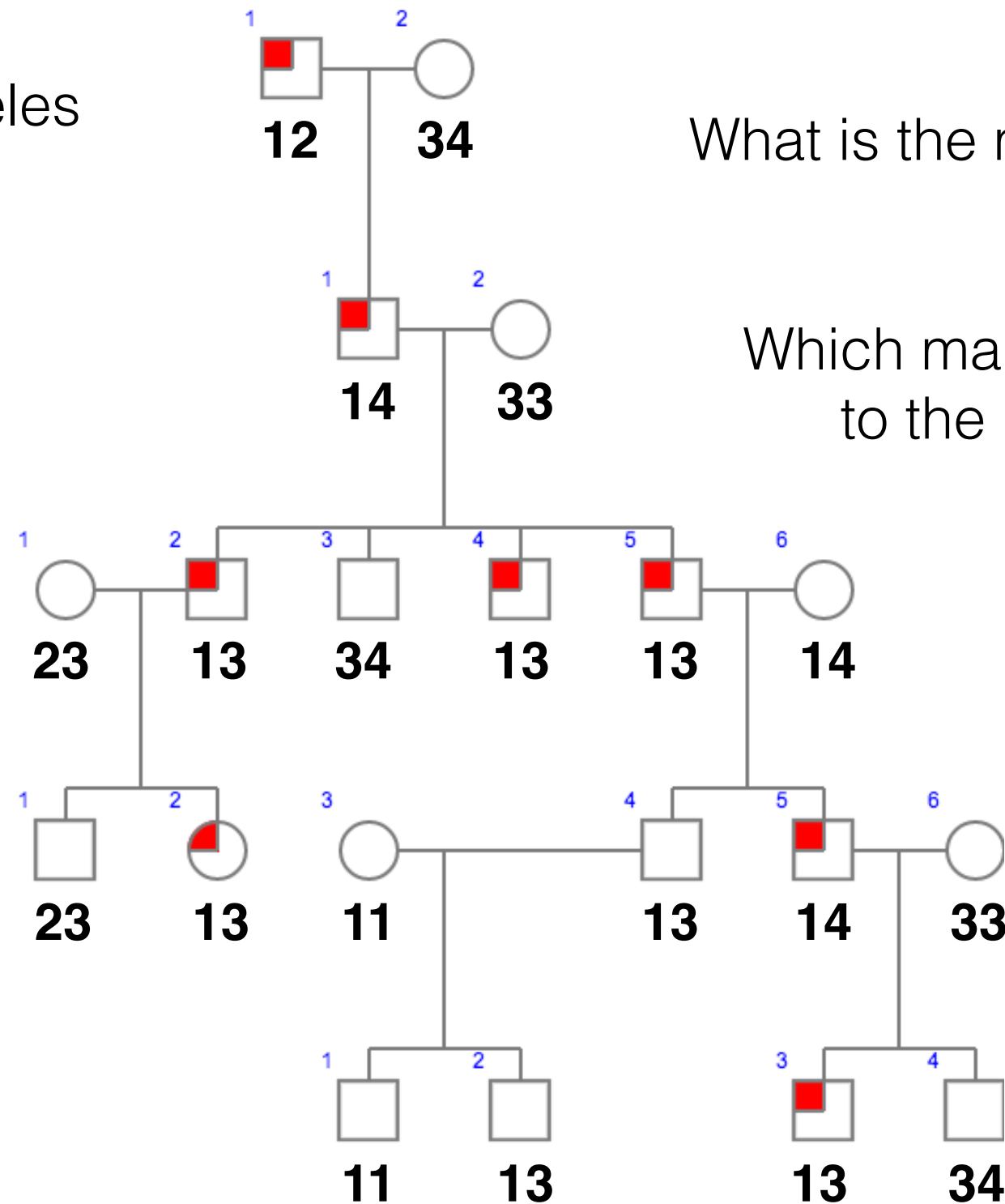
Likelihood of NO linkage between marker and disease-causing allele

The goal of linkage analysis is to identify a marker nearby the disease-causing allele

Marker with alleles
1, 2, 3, or 4

What is the mode of inheritance?

Which marker allele is linked
to the disease allele?



Linkage is reported as a log of the odds ratio or LOD

The odds ratio is a statistical method to measure association of some attribute with the presence or absence of another attribute.

Probability of pedigree under linkage versus no linkage

$$\text{LOD} = \log_{10} \frac{P(\text{pedigree with linkage})}{P(\text{pedigree with no linkage})}$$

LOD > 3 is good evidence of linkage
1 in 1,000 chance of data by random assortment

LOD < 0 means no linkage
and observed data are more likely by chance

How linked is our marker to the disease-causing allele?

Recombination frequency

Recombination frequency is written as θ

Percentage of recombinant gametes passed down

$\theta = 0$ is perfect linkage

$\theta = 0.5$ is no linkage

Linkage is reported as a log of the odds ratio or LOD

The odds ratio is a statistical method to measure association of some attribute with the presence or absence of another attribute.

Probability of pedigree under linkage versus no linkage

$$\text{LOD} = \log_{10} \frac{P(\text{pedigree with linkage})}{P(\text{pedigree with no linkage})}$$

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

$P(\text{data})$ = probability that a particular gamete was inherited

$P(\text{data} \mid \theta = 0.5)$ is independent assortment

Linkage is reported as a log of the odds ratio or LOD

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

$P(\text{data})$ = probability that a particular gamete was inherited

$$P(\text{data} \mid \theta) =$$

Probability that recombination did not occur for parental $1 - \theta$

Probability that recombination did occur for recombinant θ

Probability of phase of parent

Each individual is independent.
Use product rule to multiply probabilities.

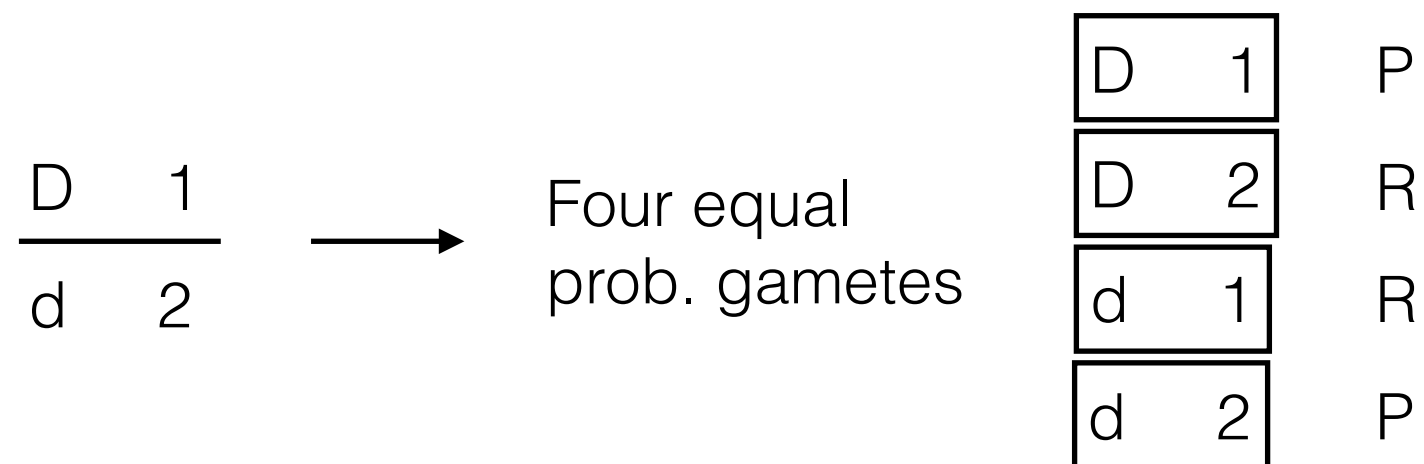
Linkage is reported as a log of the odds ratio or LOD

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

$P(\text{data})$ = probability that a particular gamete was inherited

$$P(\text{data} \mid \theta = 0.5)$$

Equal probability of two loci independently assorting and one gamete being passed down



Each individual is independent.
Use product rule to multiply probabilities.

Linkage is reported as a log of the odds ratio or LOD

Probability of pedigree under linkage versus no linkage

$$\text{LOD} = \log_{10} \frac{P(\text{pedigree with linkage})}{P(\text{pedigree with no linkage})}$$

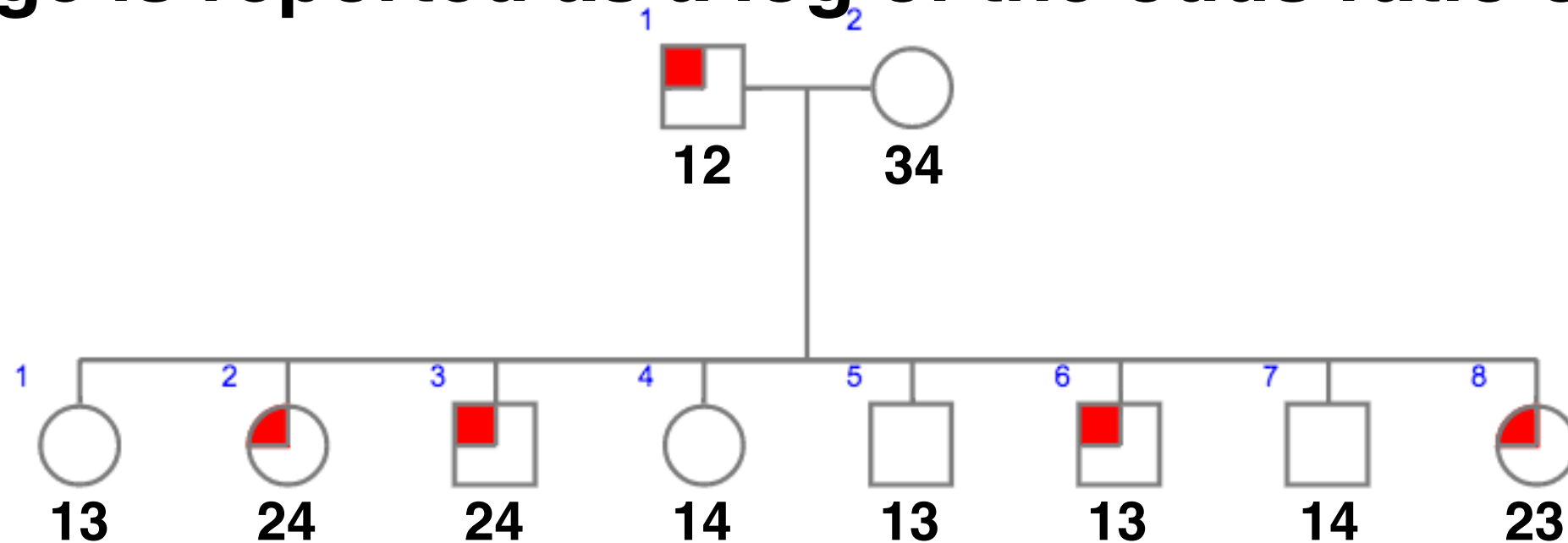
$$\text{LOD} = \log_{10} \frac{P(\text{data} \mid \theta)}{P(\text{data} \mid \theta = 0.5)}$$

$P(\text{data})$ = probability that a particular gamete was inherited

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

Use this equation when we know phase.

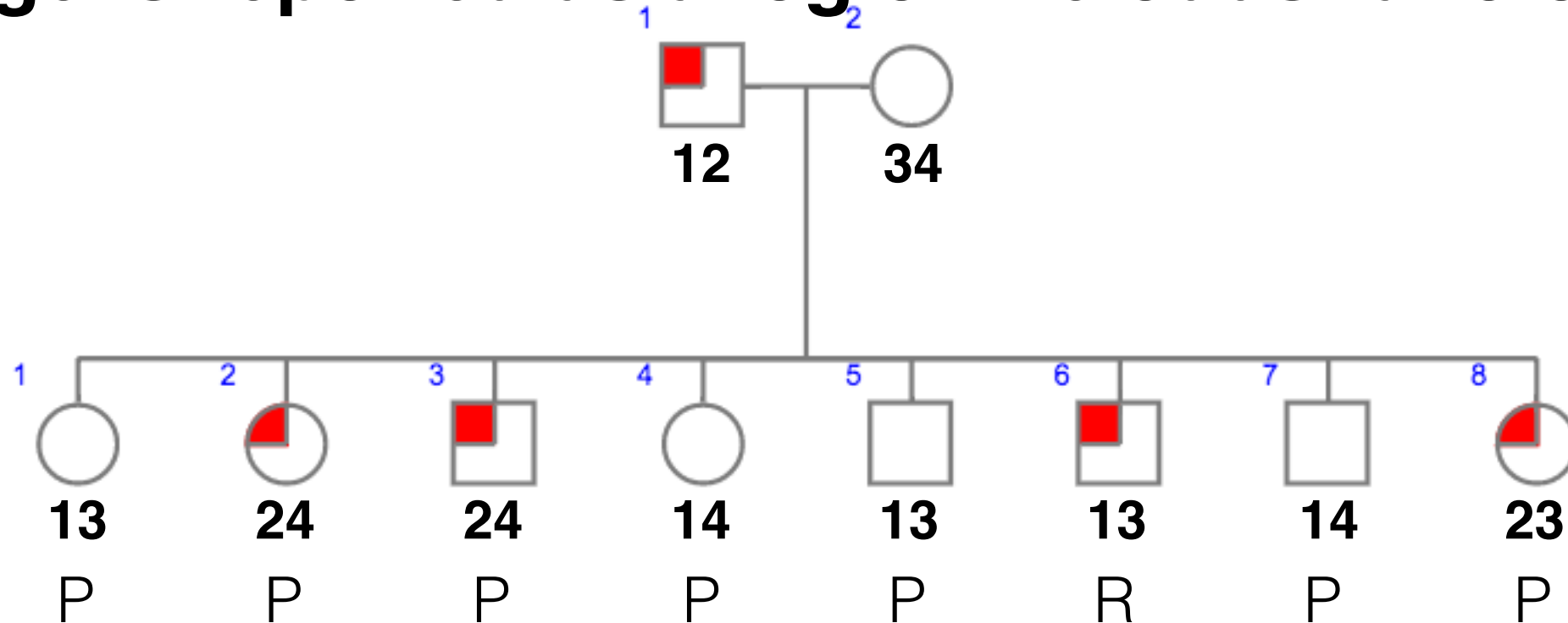
Linkage is reported as a log of the odds ratio or LOD



The paternal grandmother had the disease and her genotype was 24.
Her husband's genotype was 11 and he was not affected

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

Linkage is reported as a log of the odds ratio or LOD

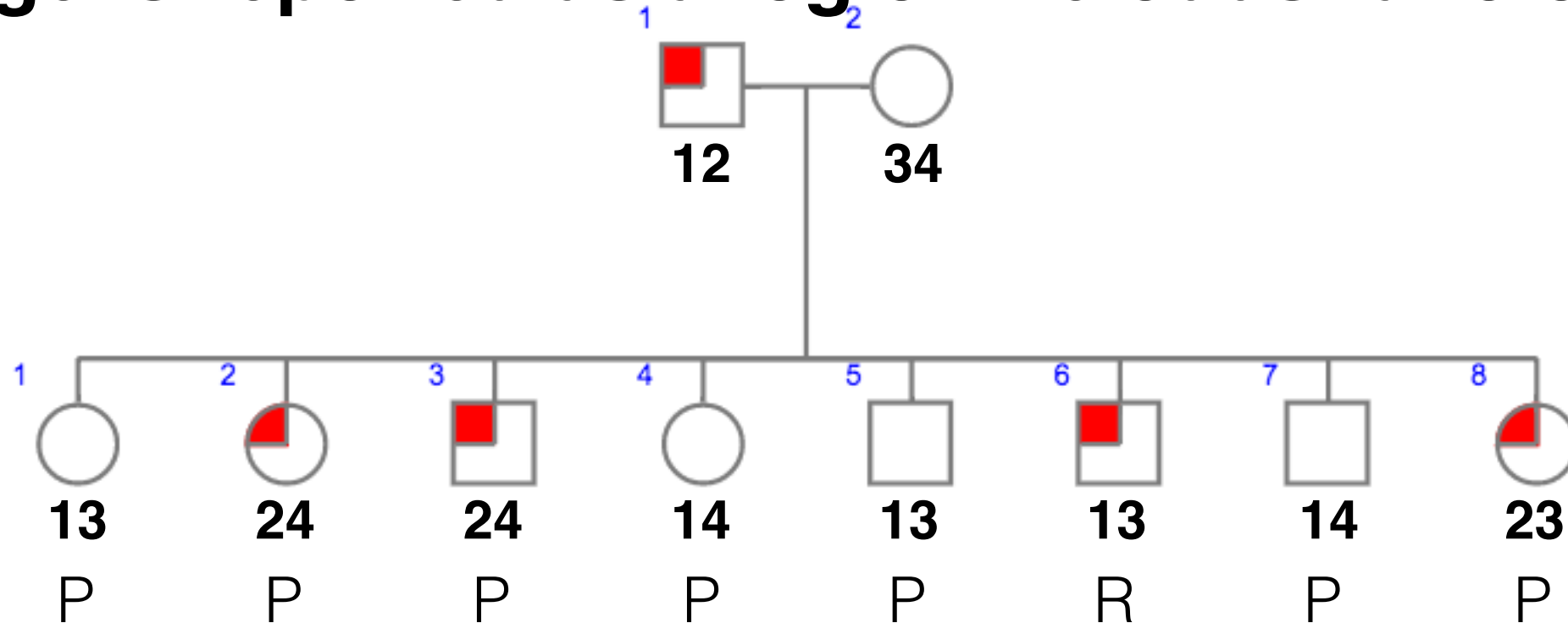


The paternal grandmother had the disease and her genotype was 24.
Her husband's genotype was 11 and he was not affected

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

1. Determine who is recombinant or parental

Linkage is reported as a log of the odds ratio or LOD

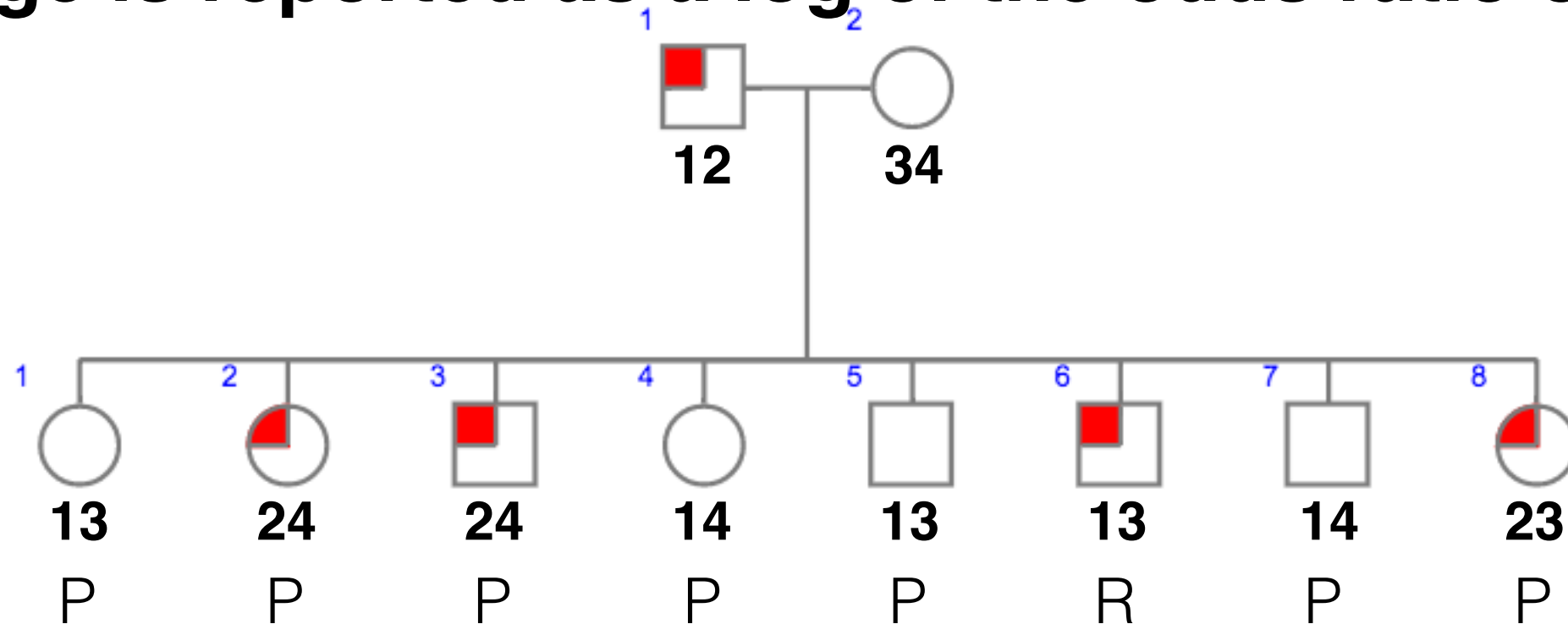


The paternal grandmother had the disease and her genotype was 24.
Her husband's genotype was 11 and he was not affected

$$\text{LOD} = \log_{10} \frac{(1 - \theta)^7 \times \theta^1}{0.5^{(7+1)}}$$

2. Put the P and R numbers in the equation

Linkage is reported as a log of the odds ratio or LOD



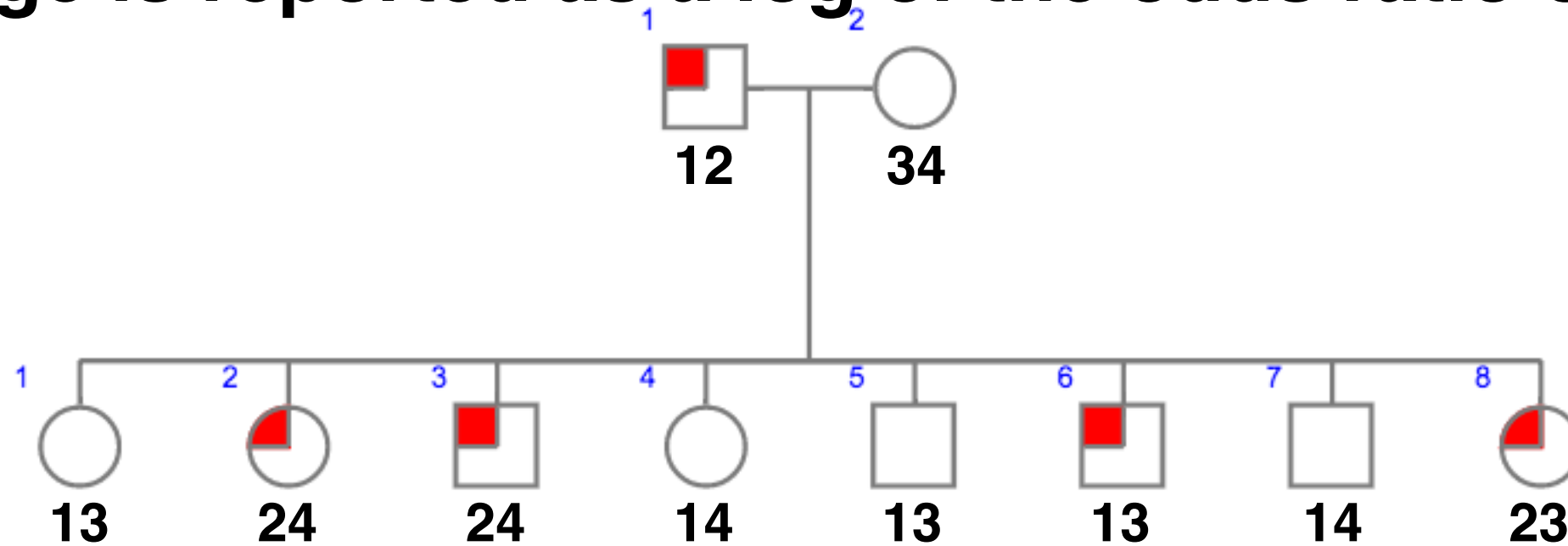
The paternal grandmother had the disease and her genotype was 24.
Her husband's genotype was 11 and he was not affected

$$\text{LOD} = \log_{10} \frac{(1 - 0.125)^7 \times 0.125^1}{0.5^{(7 + 1)}}$$

3. Put the recombination fraction in the equation

$$\theta = 0.125$$

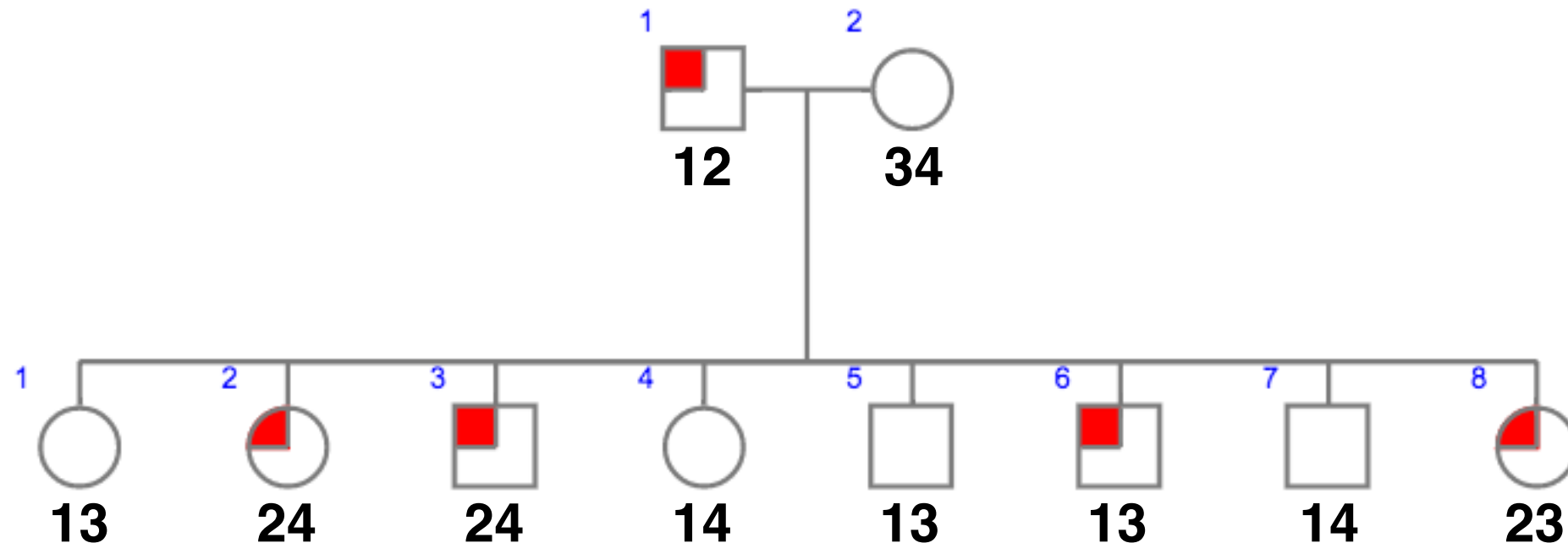
Linkage is reported as a log of the odds ratio or LOD



$$\text{LOD} = 1.1$$

$$\theta = 0.125$$

What happens when don't know the phase?



Do we know the phase of I-1?

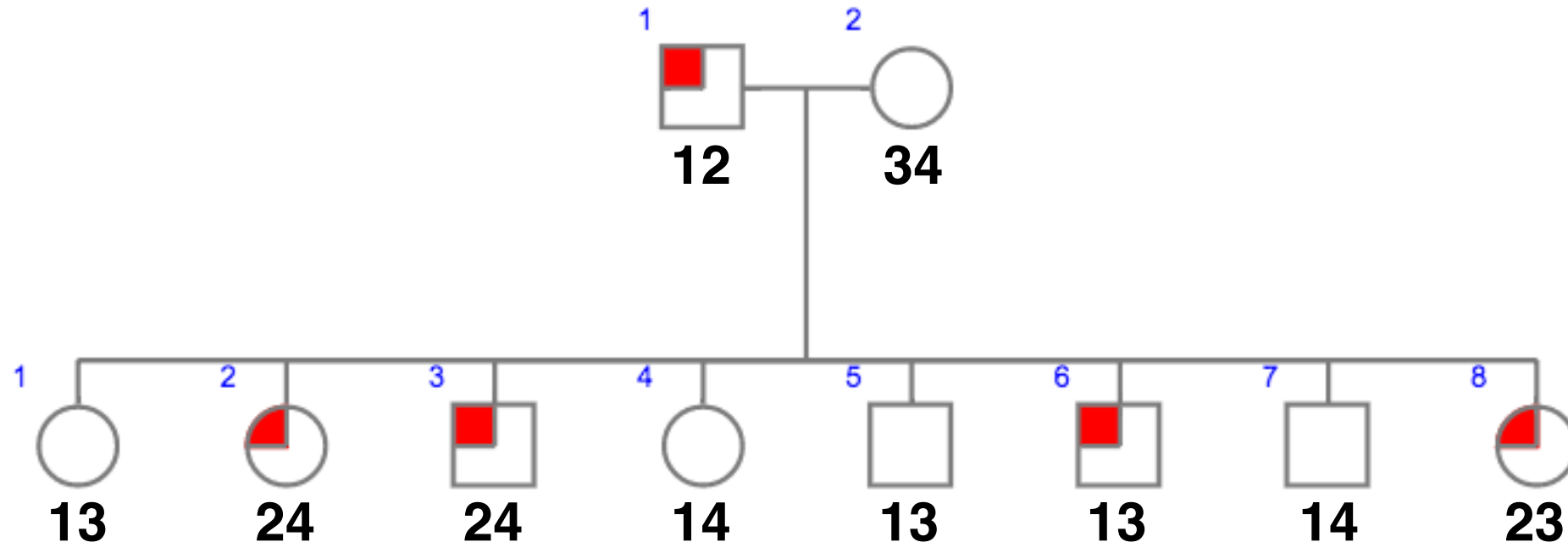
NO

Every child has an equal chance of being a recombinant or parental

Difference between being informative and knowing phase

Informative = Parent heterozygous at each of two loci

What happens when don't know the phase?



With unknown phase of parent, the LOD equation is...

$$\text{LOD} = \log_{10} \frac{\frac{1}{2}((1 - \theta)^{\text{NR}} \times \theta^{\text{R}}) + \frac{1}{2}((1 - \theta)^{\text{NR}} \times \theta^{\text{R}})}{(0.5(\text{NR} + \text{R}))}$$

1/2 chance for each phase

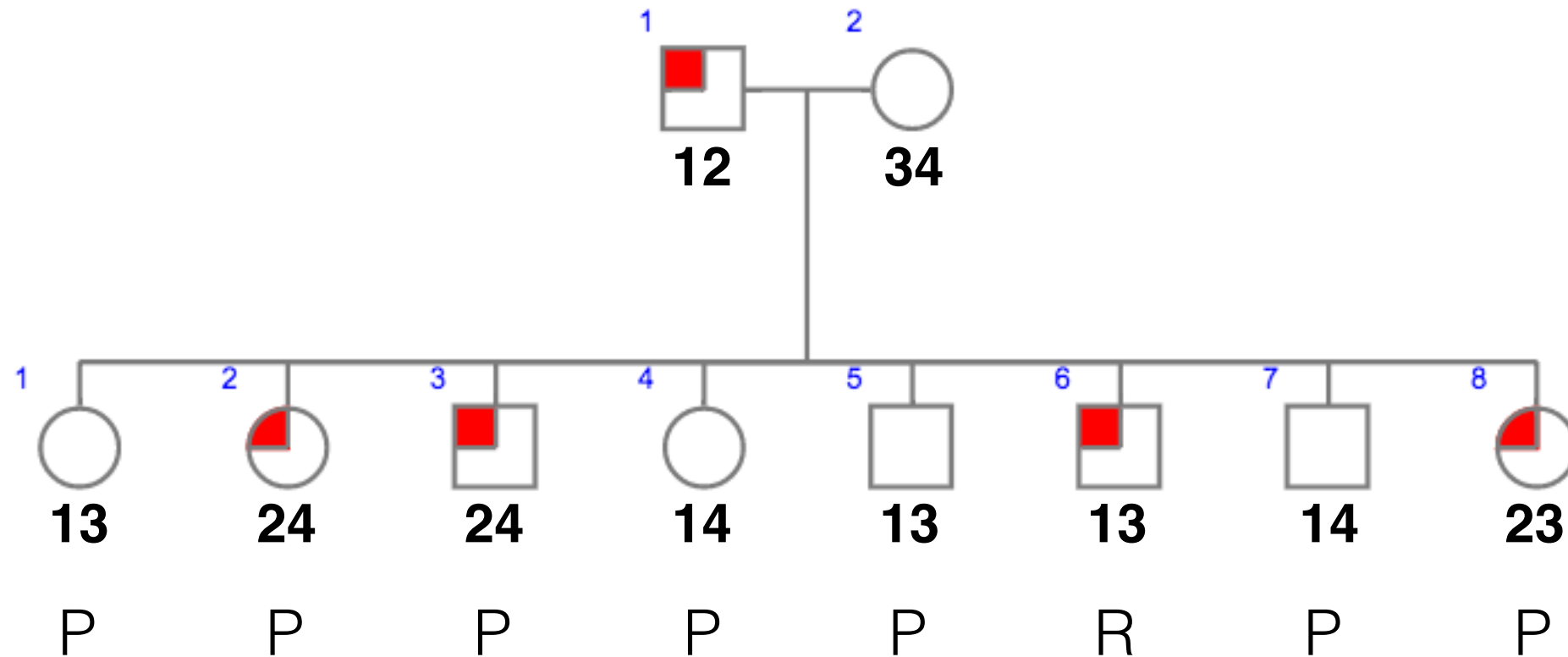
Phase 1

Phase 2

$$\frac{D}{d} \frac{2}{1}$$

$$\frac{D}{d} \frac{1}{2}$$

What happens when don't know the phase?



$$\text{LOD} = \log_{10} \frac{\frac{1}{2}((1 - \theta)^{\text{NR}} \times \theta^{\text{R}}) + \frac{1}{2}((1 - \theta)^{\text{NR}} \times \theta^{\text{R}})}{(0.5(\text{NR} + \text{R}))}$$

½ chance for each phase

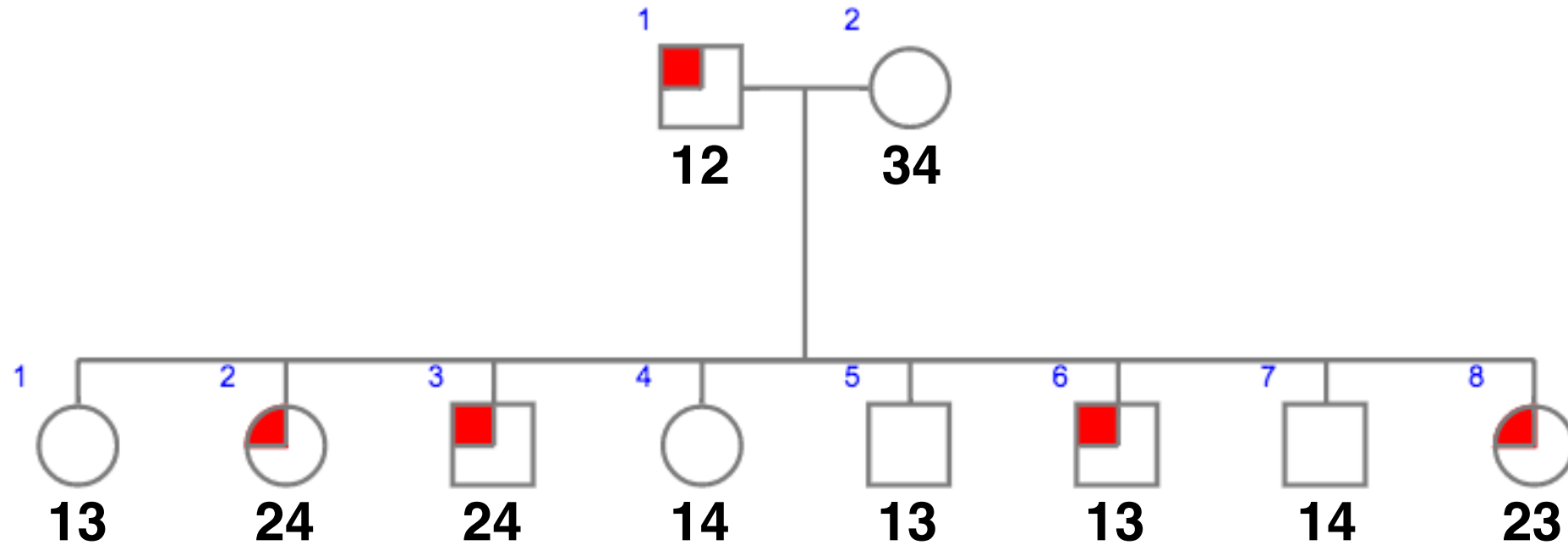
Phase 1

$$\frac{D}{d} \quad \frac{2}{1}$$

Phase 2

$$\frac{D}{d} \quad \frac{1}{2}$$

What happens when don't know the phase?



Phase 1 P P P P P R P P

$$\text{LOD} = \log_{10} \frac{\frac{1}{2}((1 - \theta)^7 \times \theta^1) + \frac{1}{2}((1 - \theta)^{\text{NR}} \times \theta^{\text{R}})}{(0.5^{(8)})}$$

1/2 chance for each phase

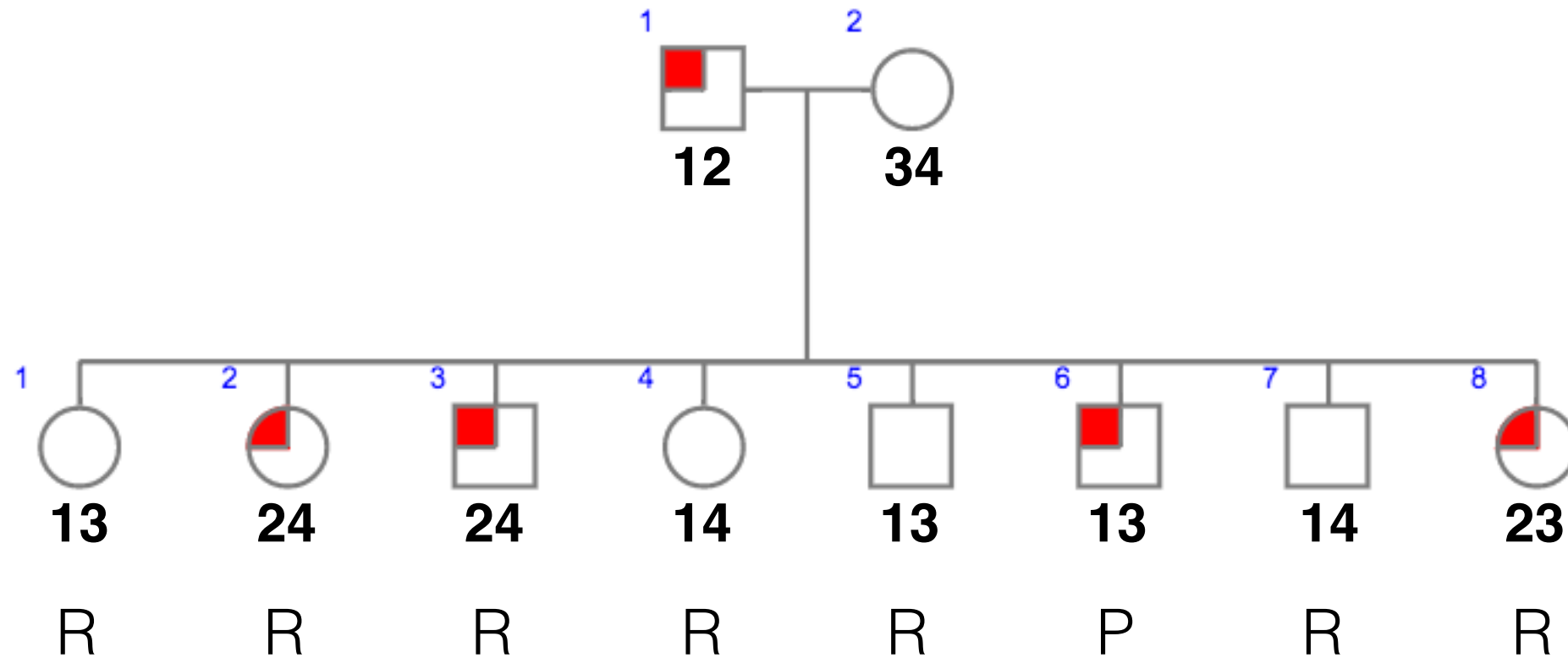
Phase 1

$$\frac{D}{d} \frac{2}{1}$$

Phase 2

$$\frac{D}{d} \frac{1}{2}$$

What happens when don't know the phase?



$$\text{LOD} = \log_{10} \frac{\frac{1}{2}((1 - \theta)^7 \times \theta^1) + \frac{1}{2}((1 - \theta)^1 \times \theta^7)}{(0.5^{(8)})}$$

1/2 chance for each phase

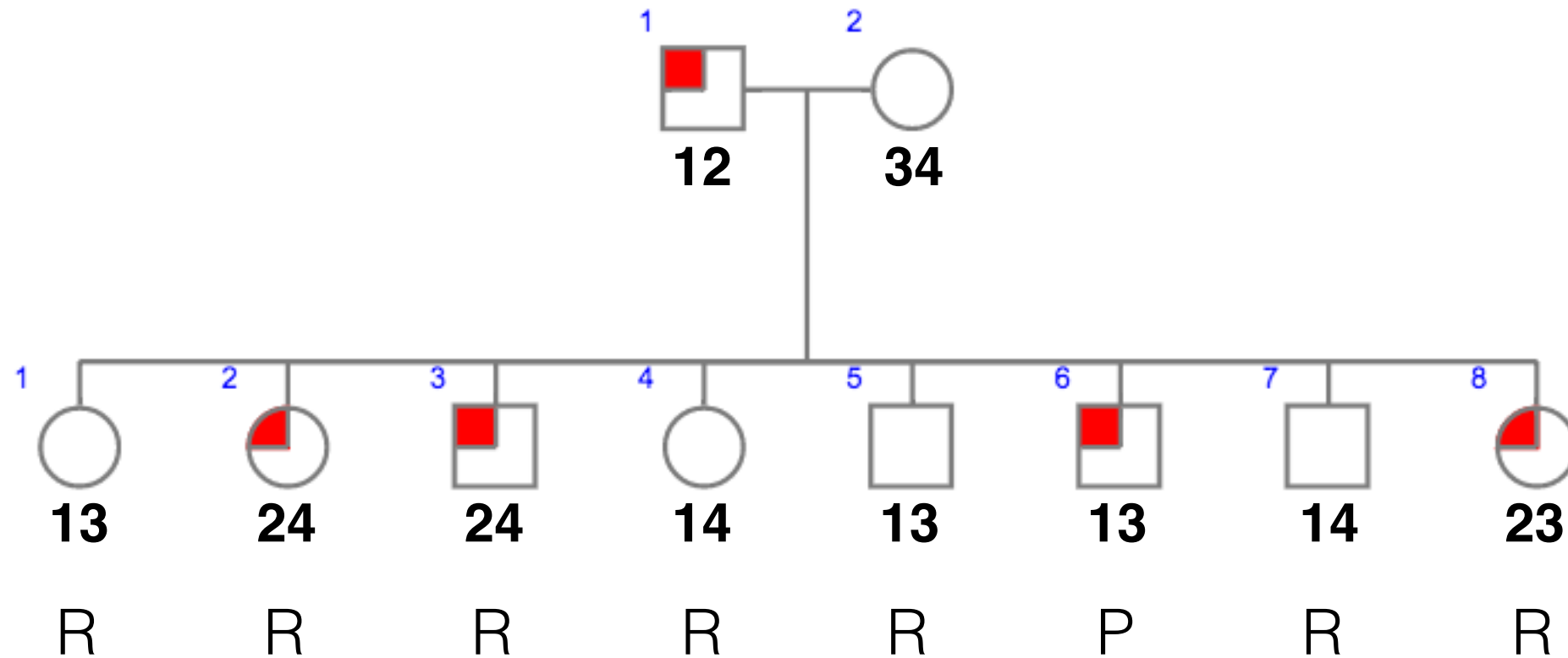
Phase 1

$$\frac{D}{d} \frac{2}{1}$$

Phase 2

$$\frac{D}{d} \frac{1}{2}$$

What happens when don't know the phase?



$$\text{LOD} = \log_{10} \frac{\frac{1}{2}((1 - 0.125)^7 \times 0.125^1) + \frac{1}{2}((1 - 0.125)^1 \times 0.125^7)}{(0.5^{(8)})}$$

$\frac{1}{2}$ chance for each phase

$$\theta = 0.125$$

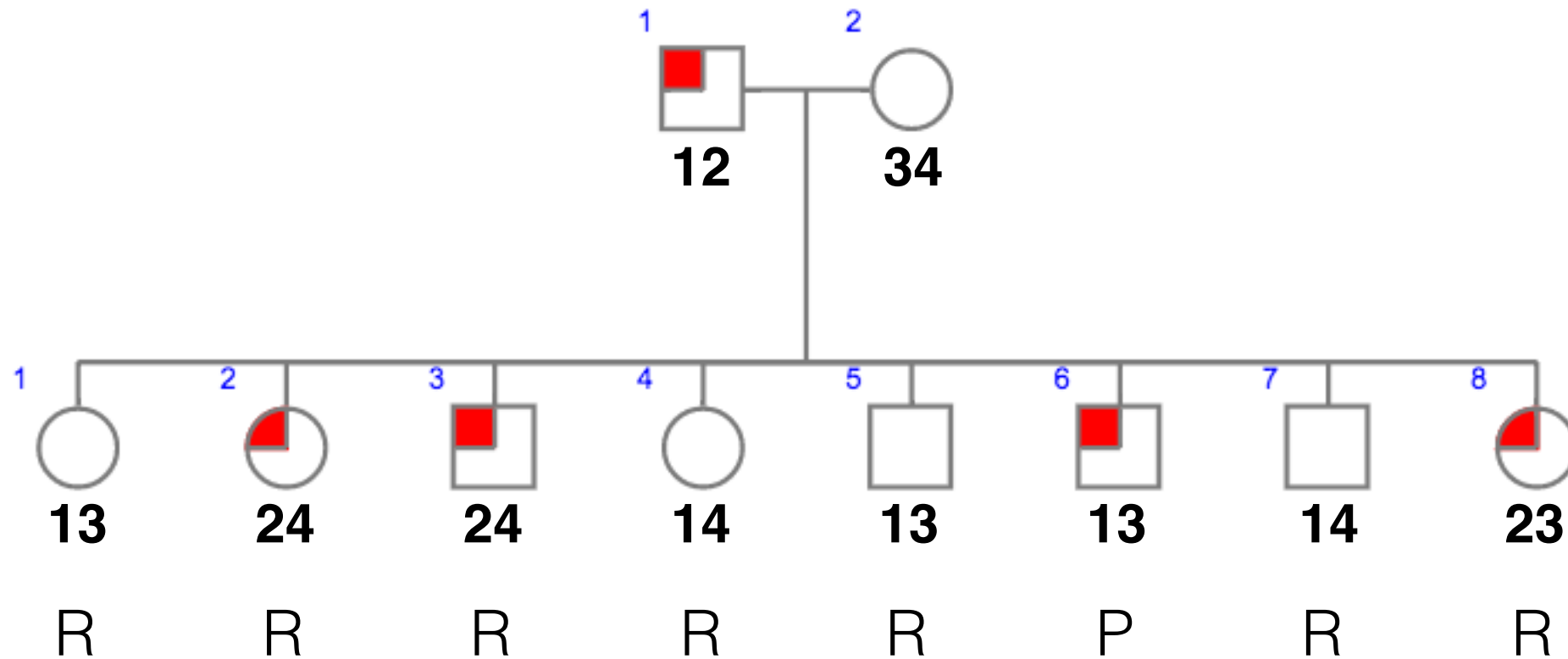
Phase 1

$$\frac{D}{d} \frac{2}{1}$$

Phase 2

$$\frac{D}{d} \frac{1}{2}$$

What happens when don't know the phase?



$$\text{LOD} = 0.79$$

$$\theta = 0.125$$

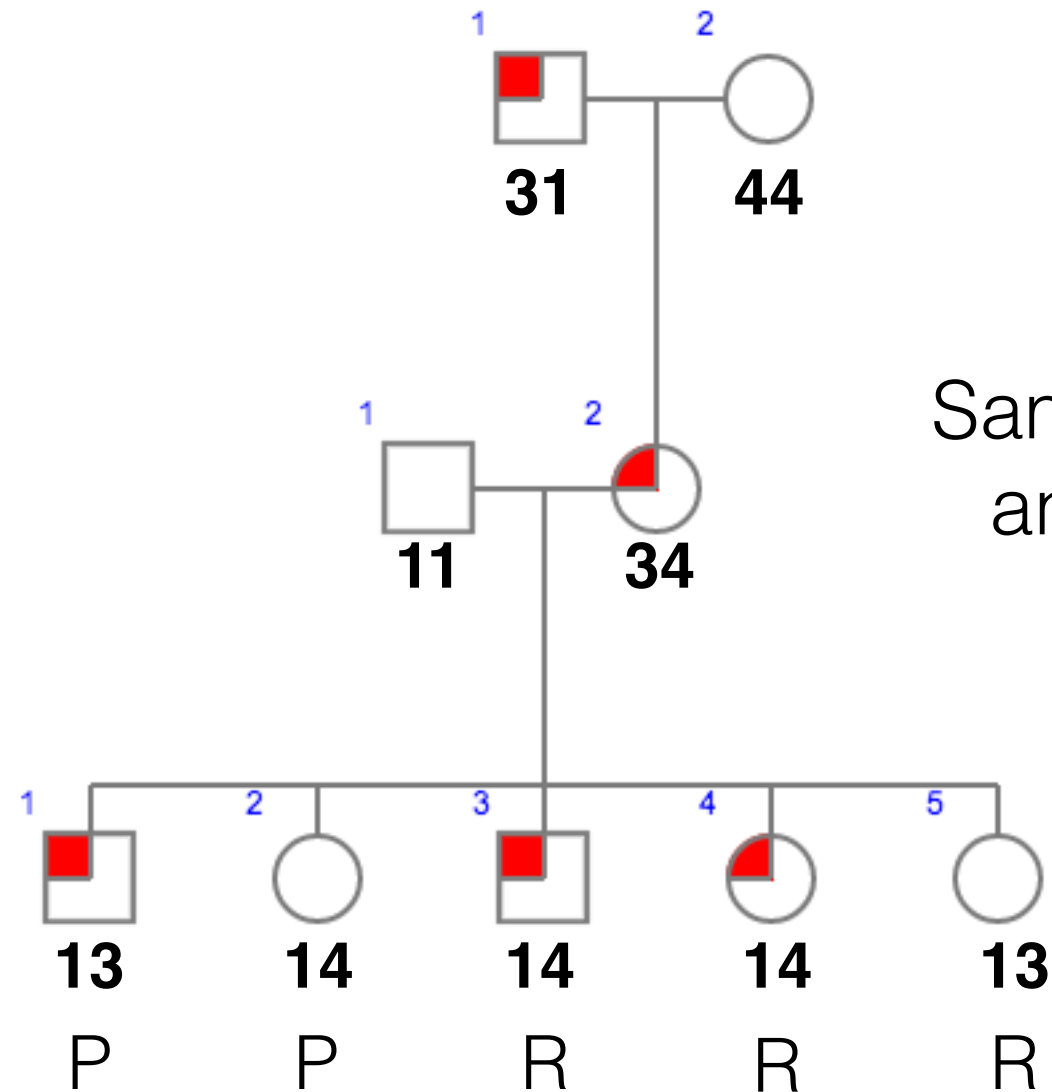
Without phase information, LOD scores decrease

Some properties of LOD scores

LOD scores from independent families can be added
(product rule with logarithms)

Determining phase increases the LOD score

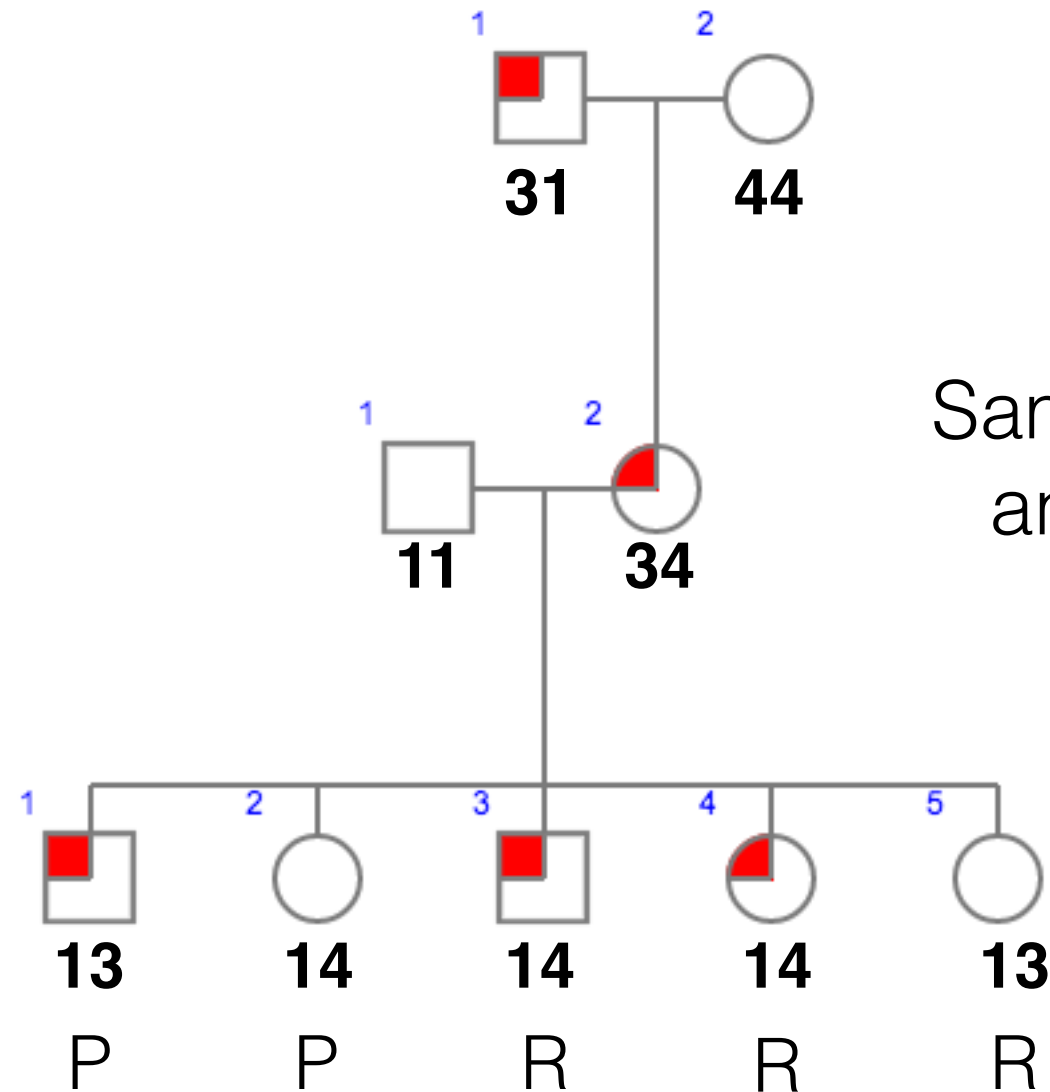
Let's add another family



Same marker as before
and we know phase

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

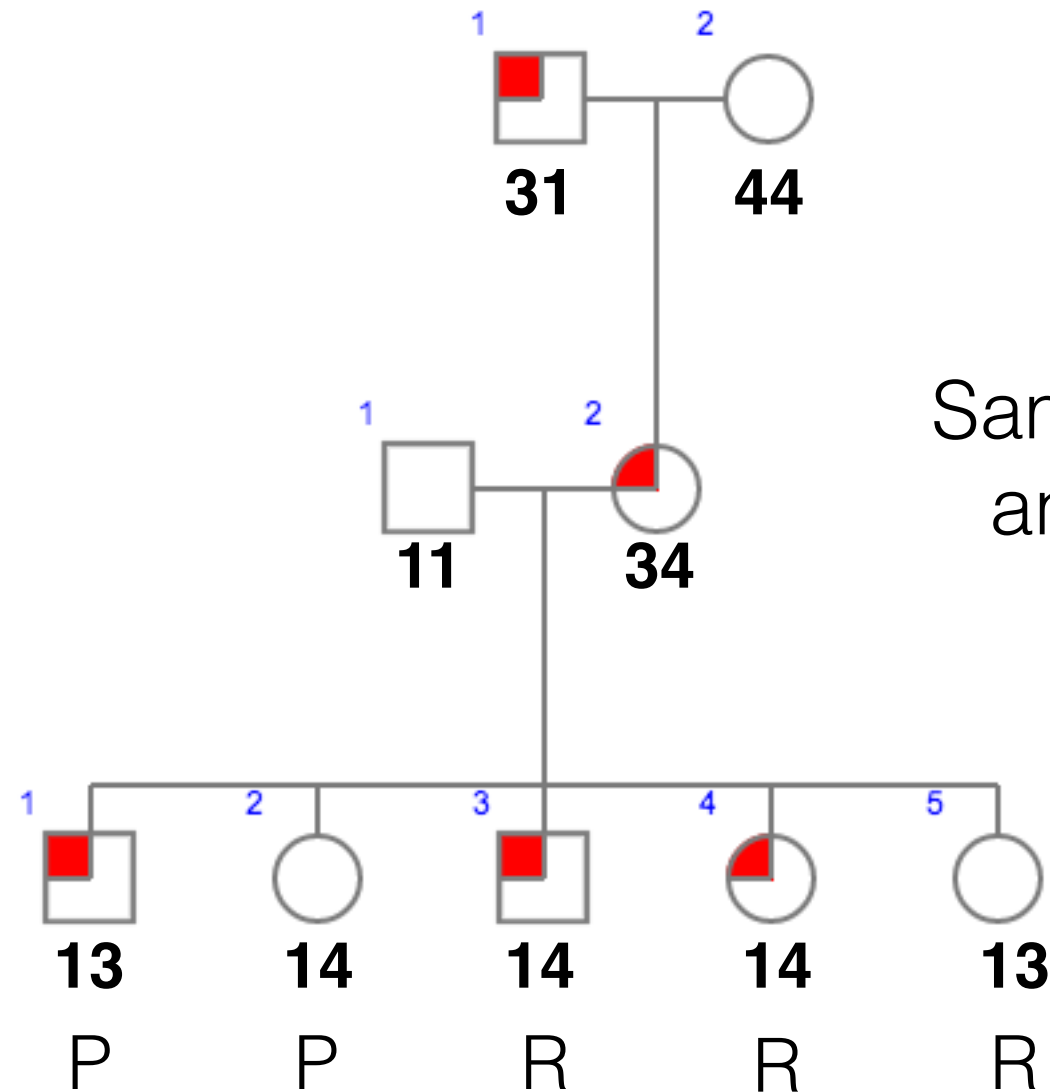
Let's add another family



Same marker as before
and we know phase

$$\text{LOD} = \log_{10} \frac{(1 - 0.125)^2 \times 0.125^3}{0.5(2 + 3)}$$

Let's add another family



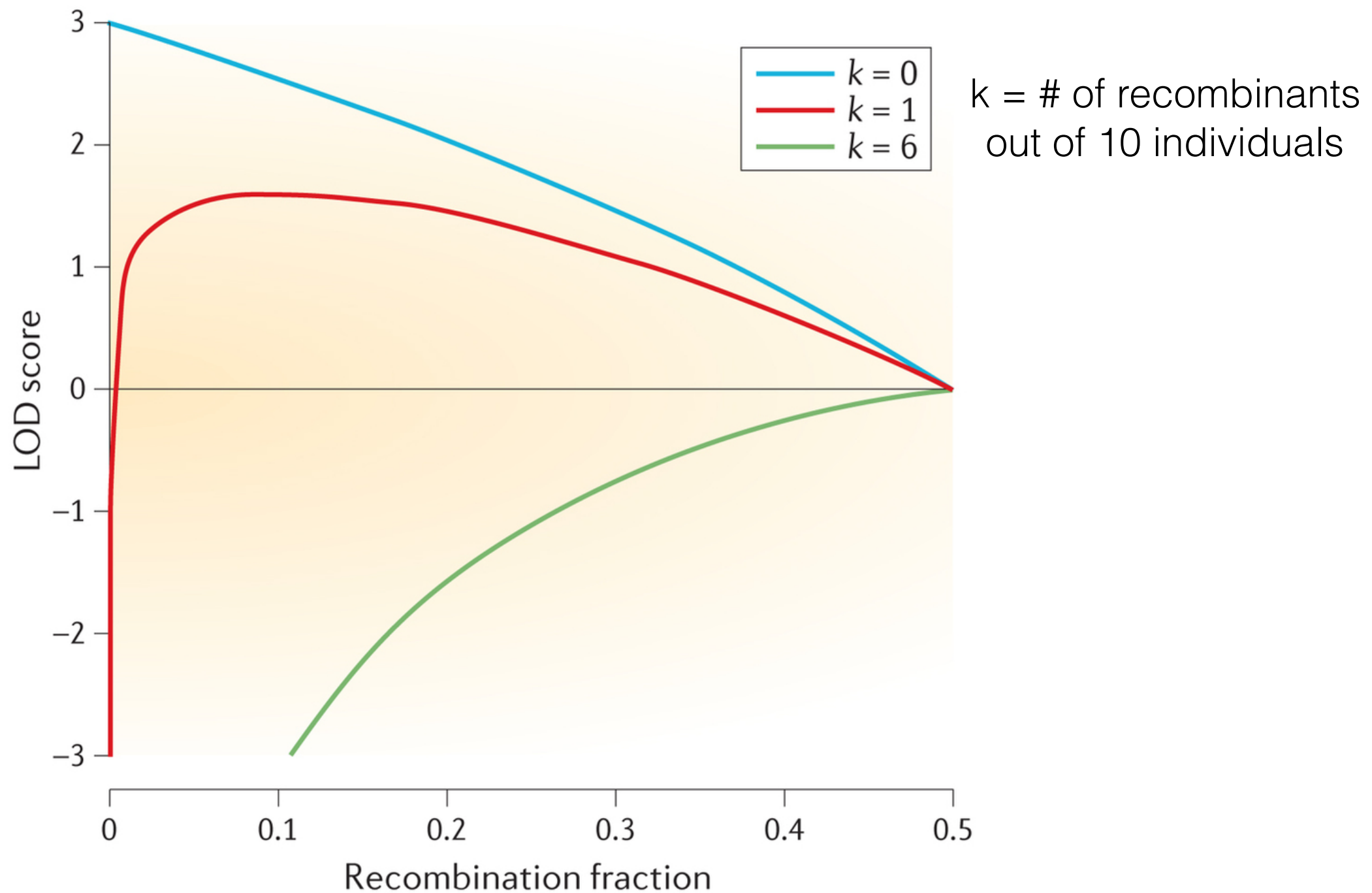
Same marker as before
and we know phase

$$\text{LOD} = -0.1159$$

$$\theta = 0.125$$

Change theta to 0.6? No, 0.5 is unlinked

What if we try all possible thetas between 0 and 0.5?



Nature Reviews | **Genetics**

Some properties of LOD scores

LOD scores from independent families can be added
(product rule with logarithms)

Determining phase increases the LOD score

Within one family...

some individuals will have phase known,
some individuals will not have defined phase,
some individuals will be uninformative.

- 1. Determine informativeness**
- 2. Assess phase**
- 3. Calculate LOD in family**
- 4. Add families (at the same theta)**

The good and the bad of family-based linkage analysis

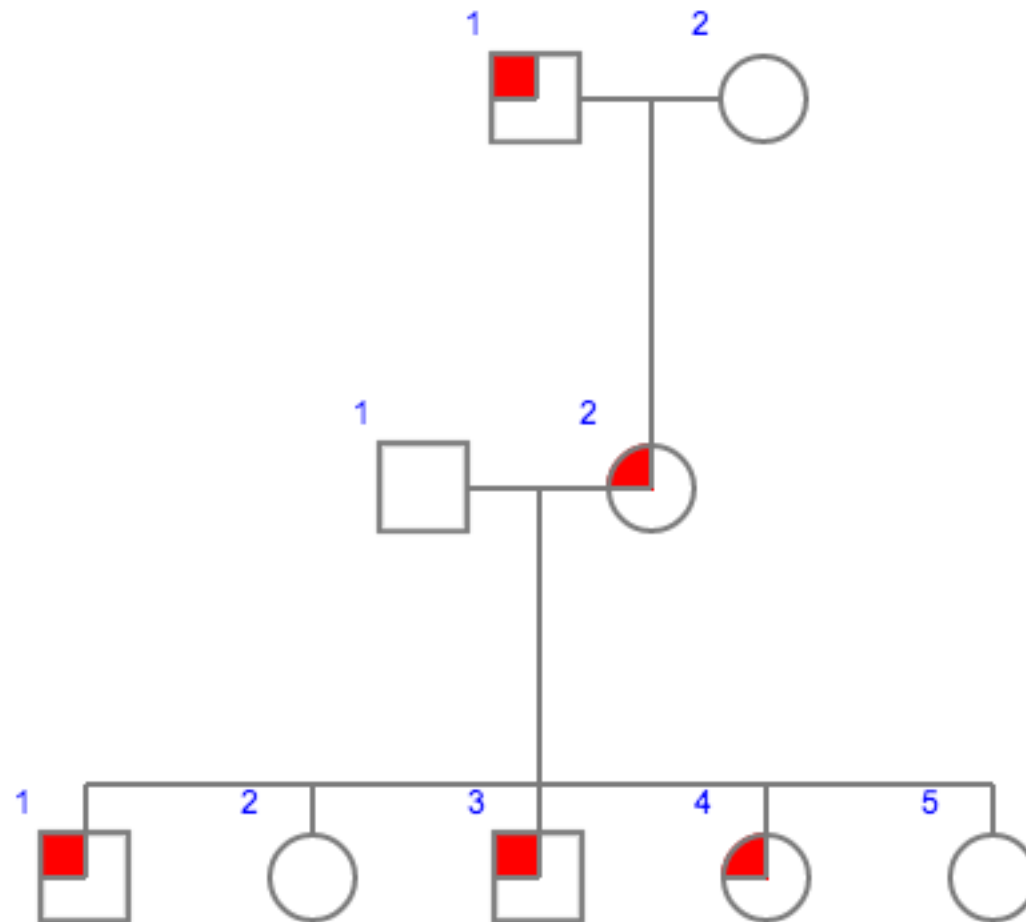
Positives:

- Less allelic heterogeneity in families
- Clearly tell recombination events
- Powerful method to find rare variant effects

Negatives:

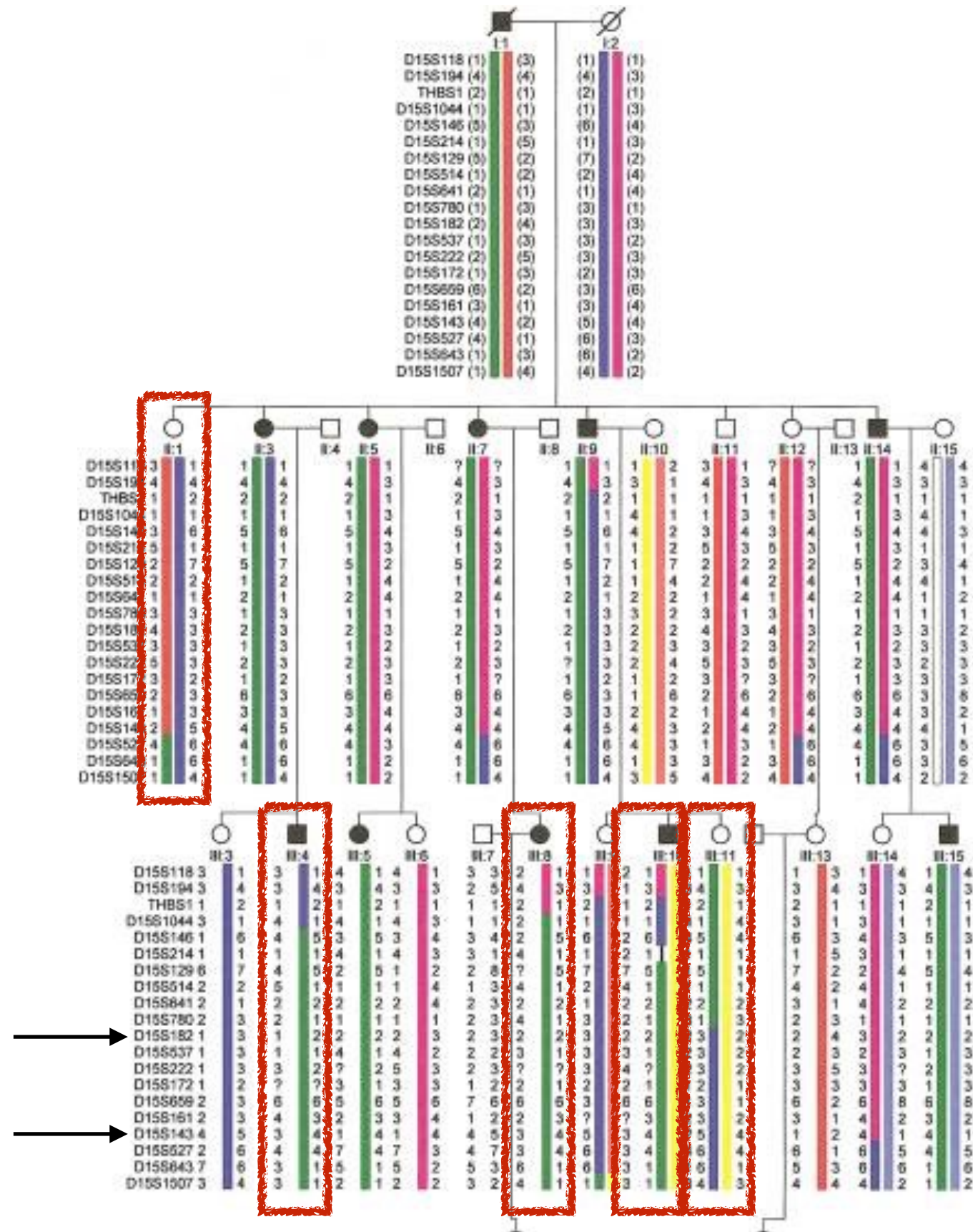
- Large families are rare
- Different families with the “same” disease could have different genetic causes
- Mapping resolution is 5 cM or 5 megabase pairs
- Difficult for late-onset diseases

With whole-exome and whole-genome sequencing, family linkage analysis gets even more powerful

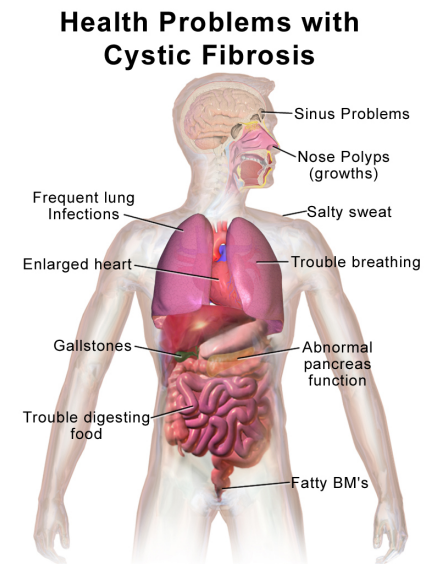
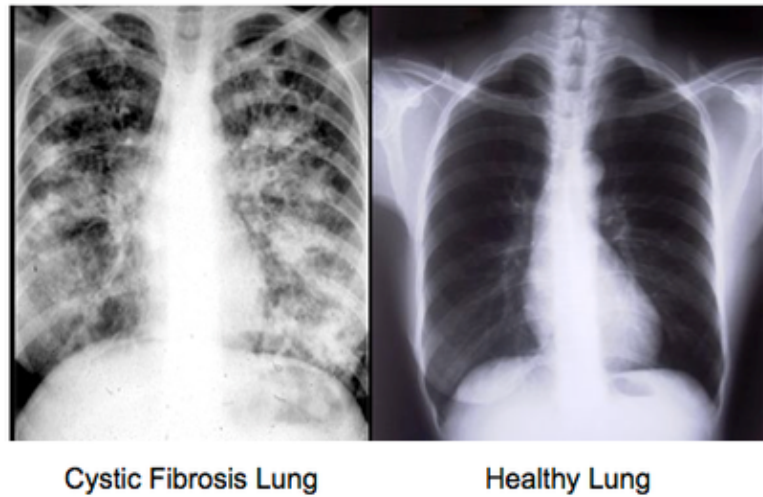


- Look at “all” markers in the genome simultaneously
- Dominant disorders mean look for heterozygous mutations linked to shared blocks of variants (haplotypes)
- X-linked lets you focus on the X chromosome
- Variants should be private to the family and deleterious

Linked markers on chromosomes allow for haplotype mapping



What about cystic fibrosis?



1. Autosomal recessive disorder
2. Not caused by chromosomal aberrations or meiotic NDJ
3. Mapped to chromosome 7
4. Mutations in CF gene are null or hypomorphs
5. Compound heterozygosity (failure to complement) is common
6. No known epistatic genes to CF gene
7. Genetic enhancers are known (immune modulatory genes)
8. No genetic suppressors are known yet
9. Cell autonomous action of CF gene
10. CF caused by mutations in CFTR

Cystic fibrosis locus defined by a genetically linked polymorphic DNA marker

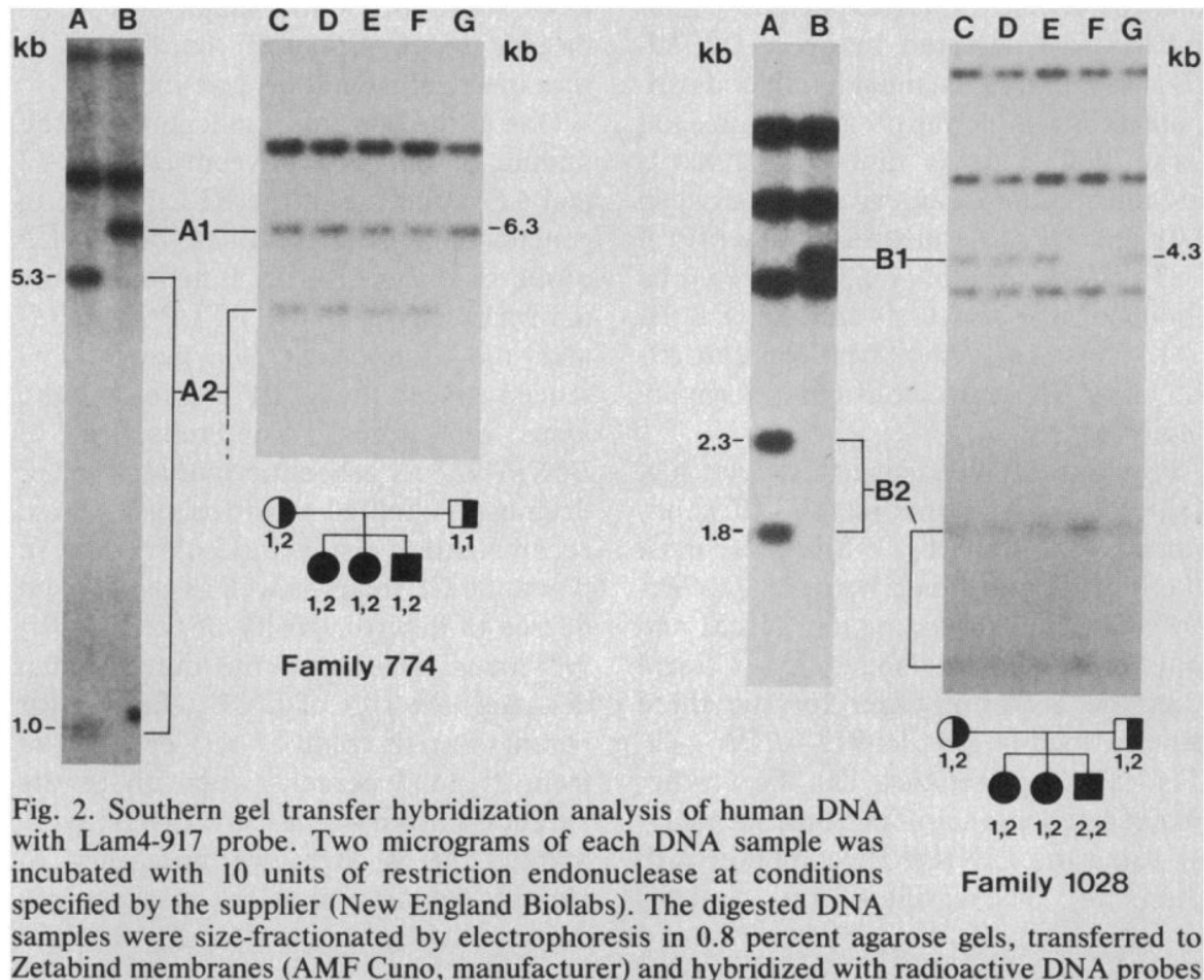


Table 1. Linkage relationships of *D0CRI-917—CF* and *D0CRI-917—PON*.

Loci	Number of informative families	LOD (z) scores at recombinant fractions (θ) of:								
		0.01	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40
D0CRI-917—CF	36 (Can)	−6.02	0.98	2.84	3.20	2.96	2.44	1.79	1.13	0.55
	3 (HGCMR)	0.14	0.69	0.79	0.75	0.66	0.53	0.39	0.25	0.12
	39 (Total)	−5.88	1.67	3.63	3.95	3.62	2.97	2.18	1.38	0.67
D0CRI-917—PON	11 (Can)	4.27	5.01	4.78	4.28	3.66	2.97	2.25	1.51	0.81

Cystic fibrosis locus defined by a genetically linked polymorphic DNA marker

Abstract. A polymorphic DNA marker has been found genetically linked, in a set of 39 human families, to an autosomal recessive gene that causes cystic fibrosis (CF), a disease affecting one in 2000 Caucasian children. The DNA marker (called D0CRI-917) is also linked to the PON locus, which by independent evidence is linked to the CF locus. The best estimates of the genetic distances are 5 centimorgans between the DNA marker and PON and 15 centimorgans between the DNA marker and the CF locus, meaning that the location of the disease gene has been narrowed to about 1 percent of the human genome (about 30 million base pairs). Although the data are consistent with the interpretation that a single locus causes cystic fibrosis, the possibility of genetic heterogeneity remains. The discovery of a linked DNA polymorphism is the first step in molecular analysis of the CF gene and its causative role in the disease.