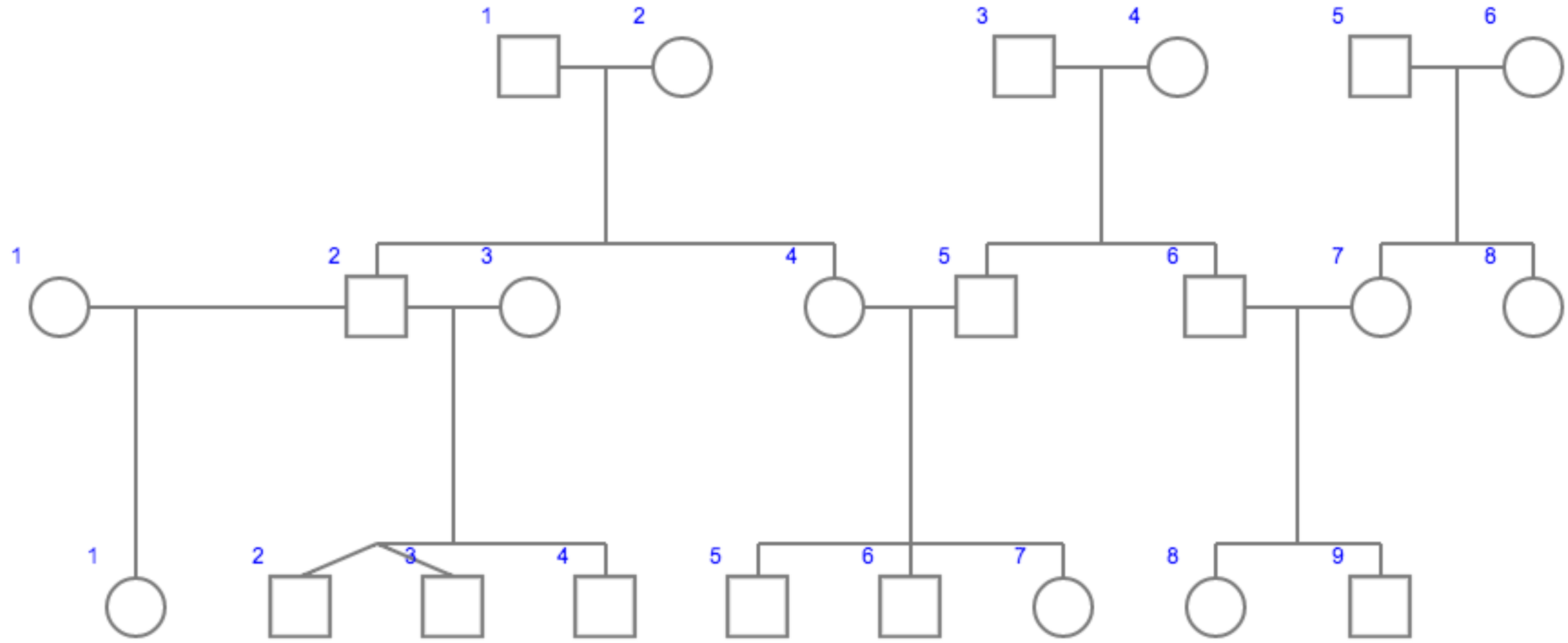
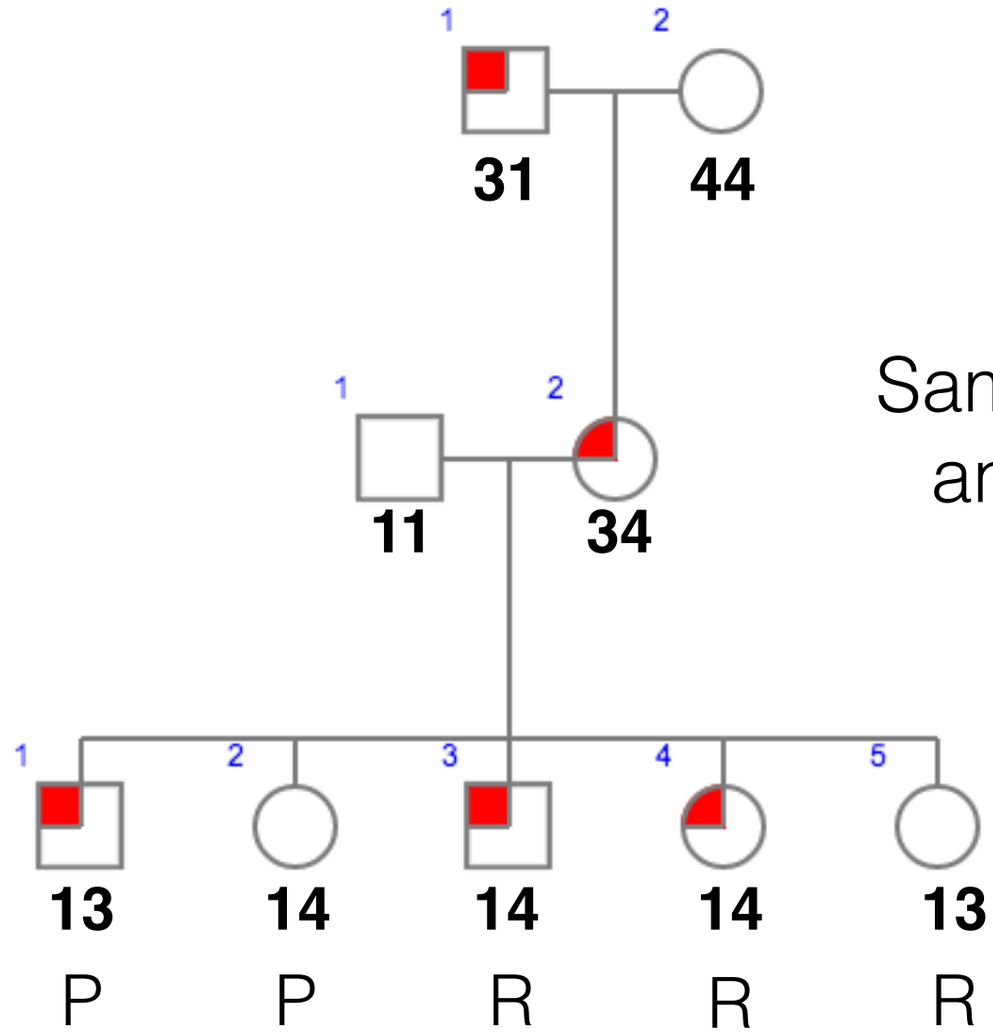
Linkage mapping in families



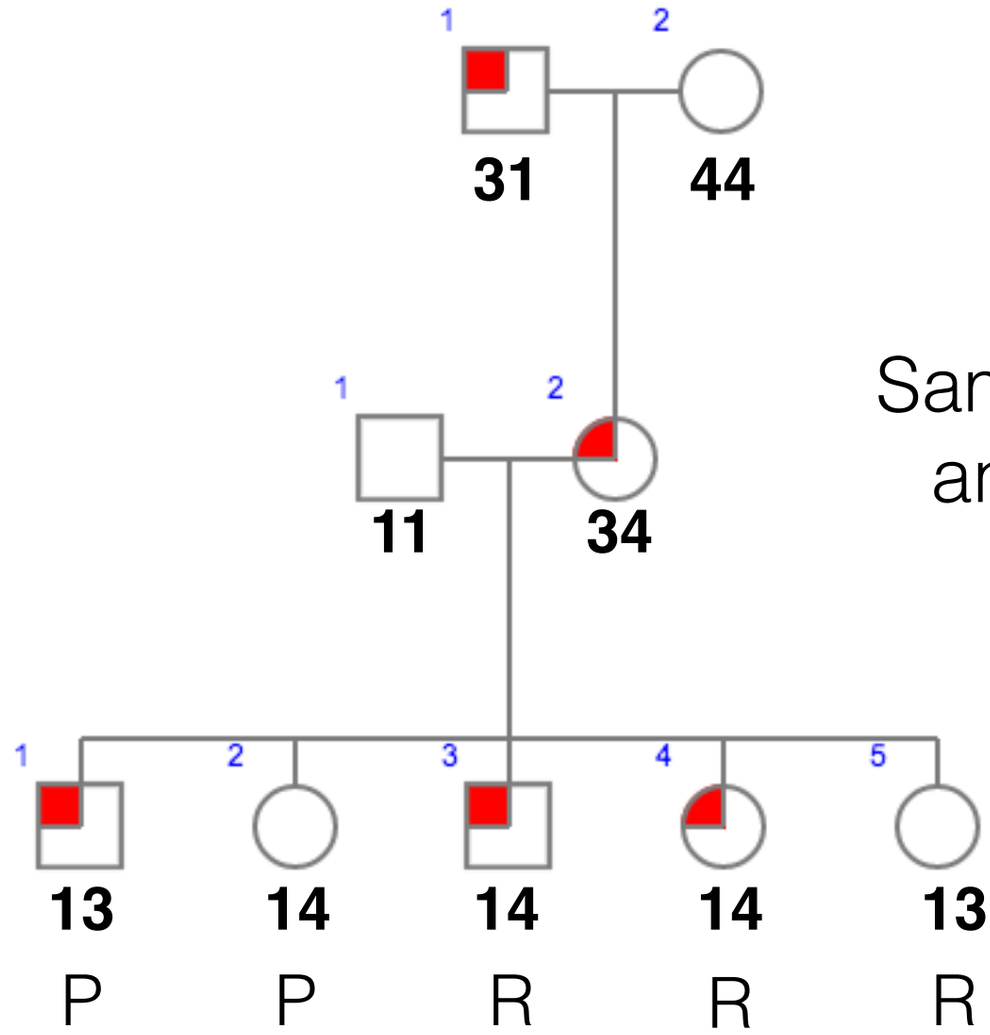
Let’s add another family



Same marker as before and we know phase

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

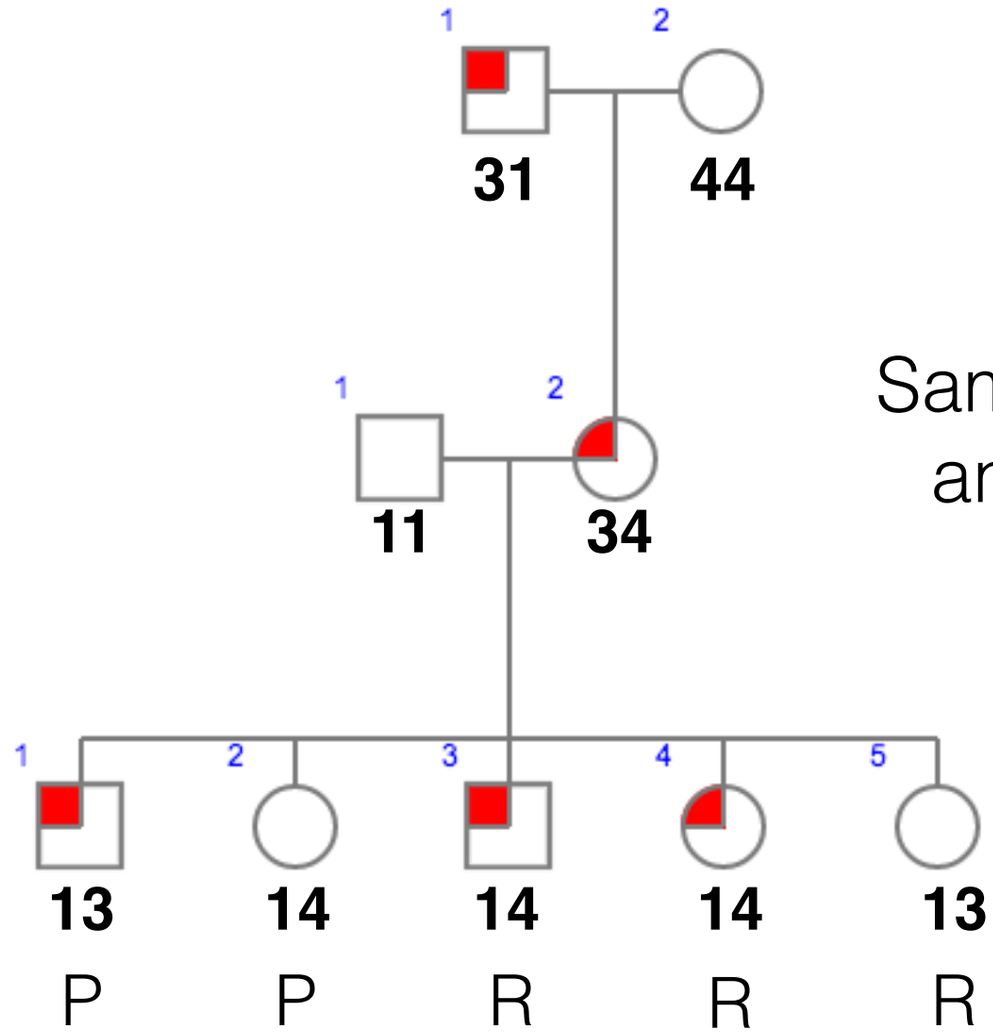
Let’s add another family



Same marker as before and we know phase

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

Let’s add another family

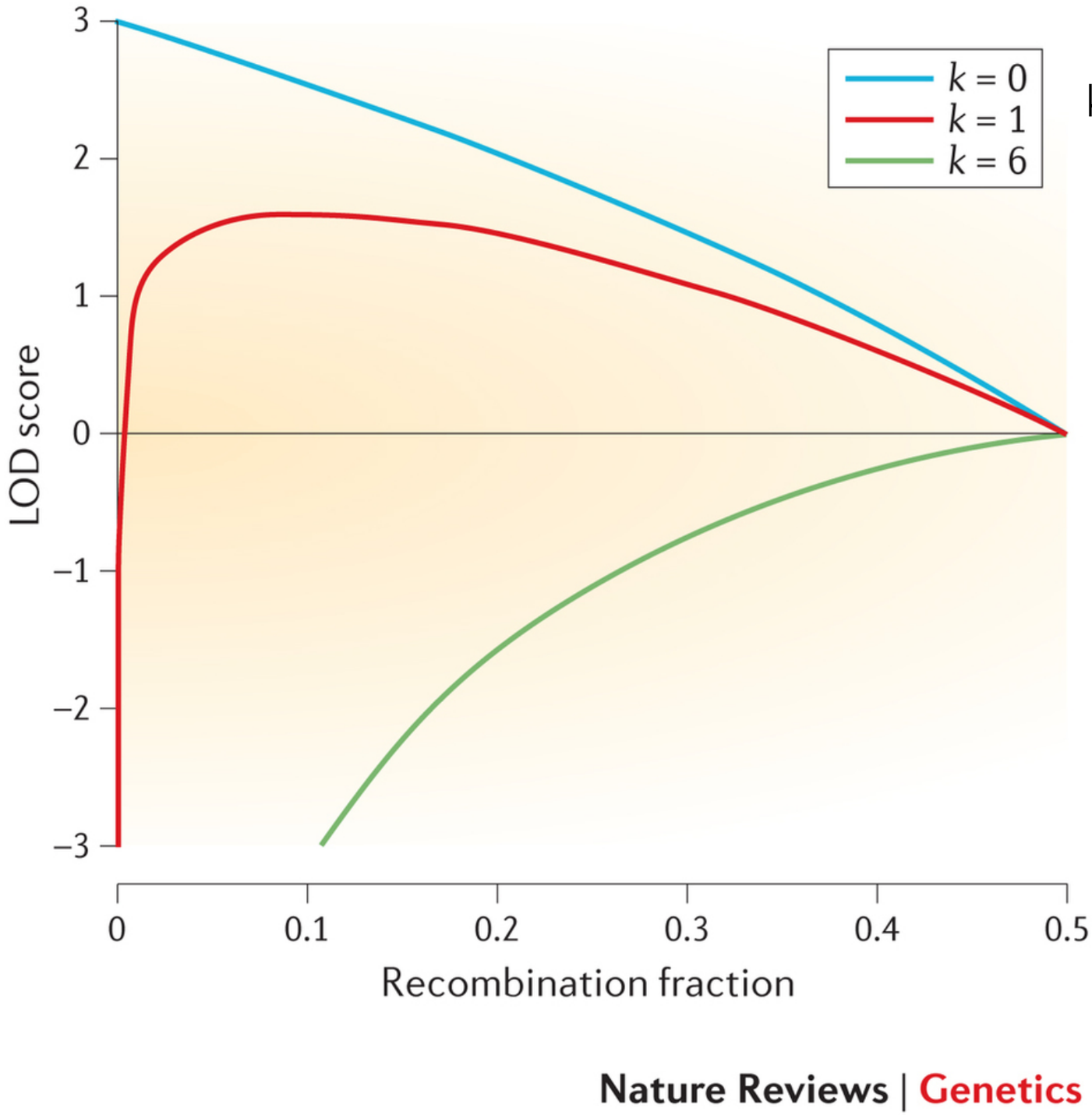


Same marker as before and we know phase

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?



k = # of recombinants out of 10 individuals

Ott *et al.* Nature Reviews Genetics 2015

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.

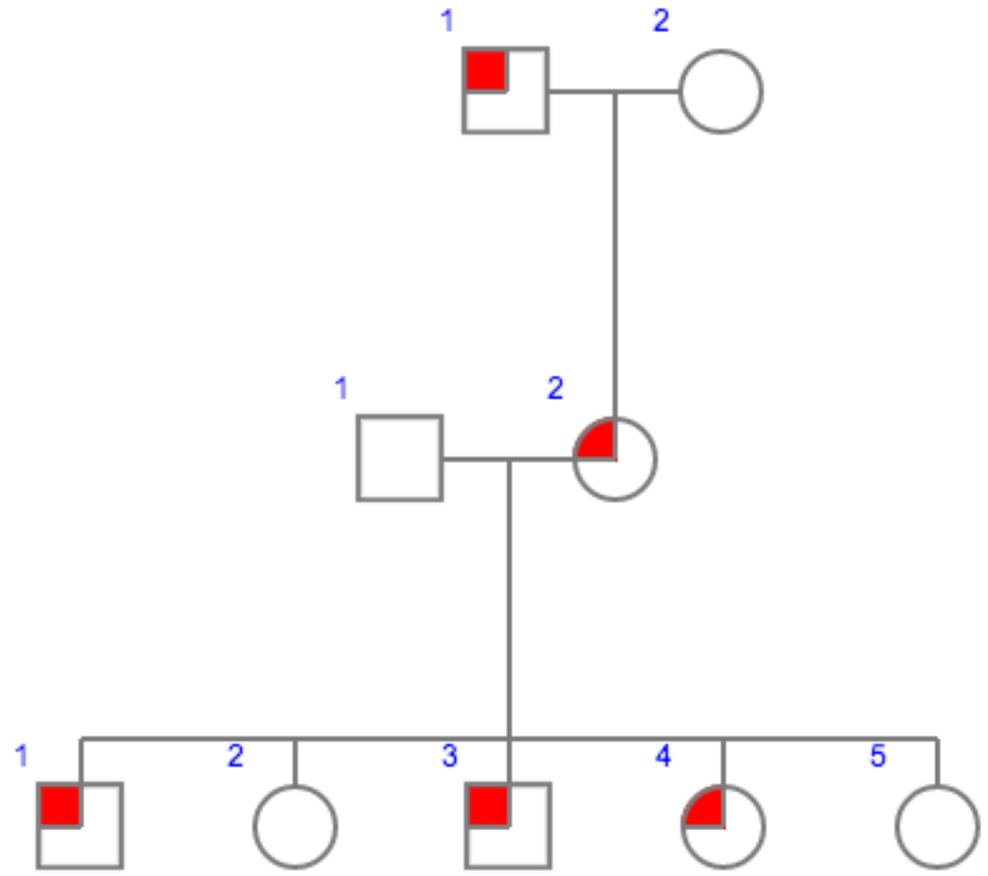
1. **Determine informativeness**
2. **Assess phase**
3. **Calculate LOD in family**
4. **Add families (at the same theta)**

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

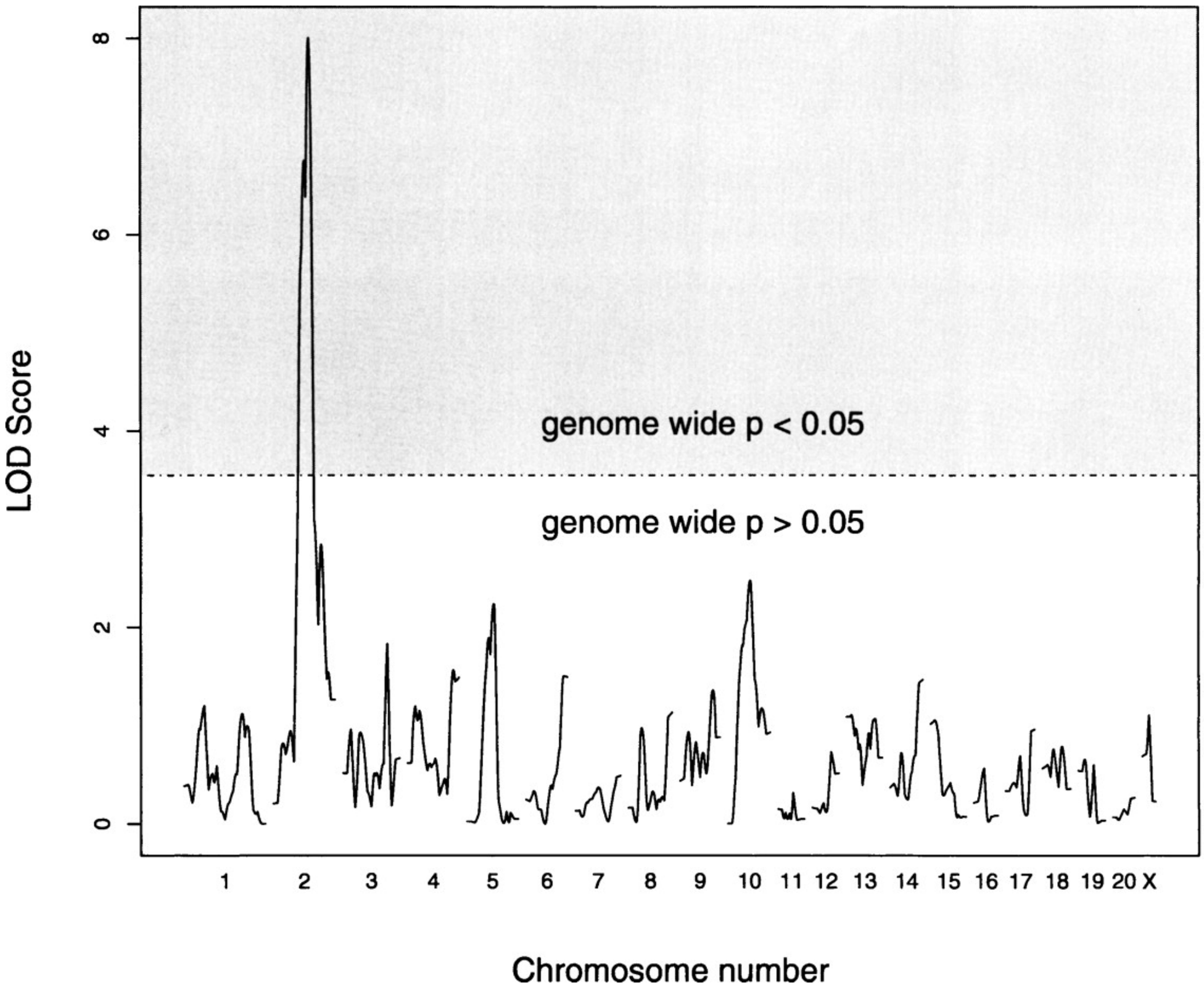
Negatives:
- Large families are rare
- Different families with the “same” disease could have different genetic causes
- Mapping resolution is 5 cM or 5 megabase pairs
- Difficult for late-onset diseases

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



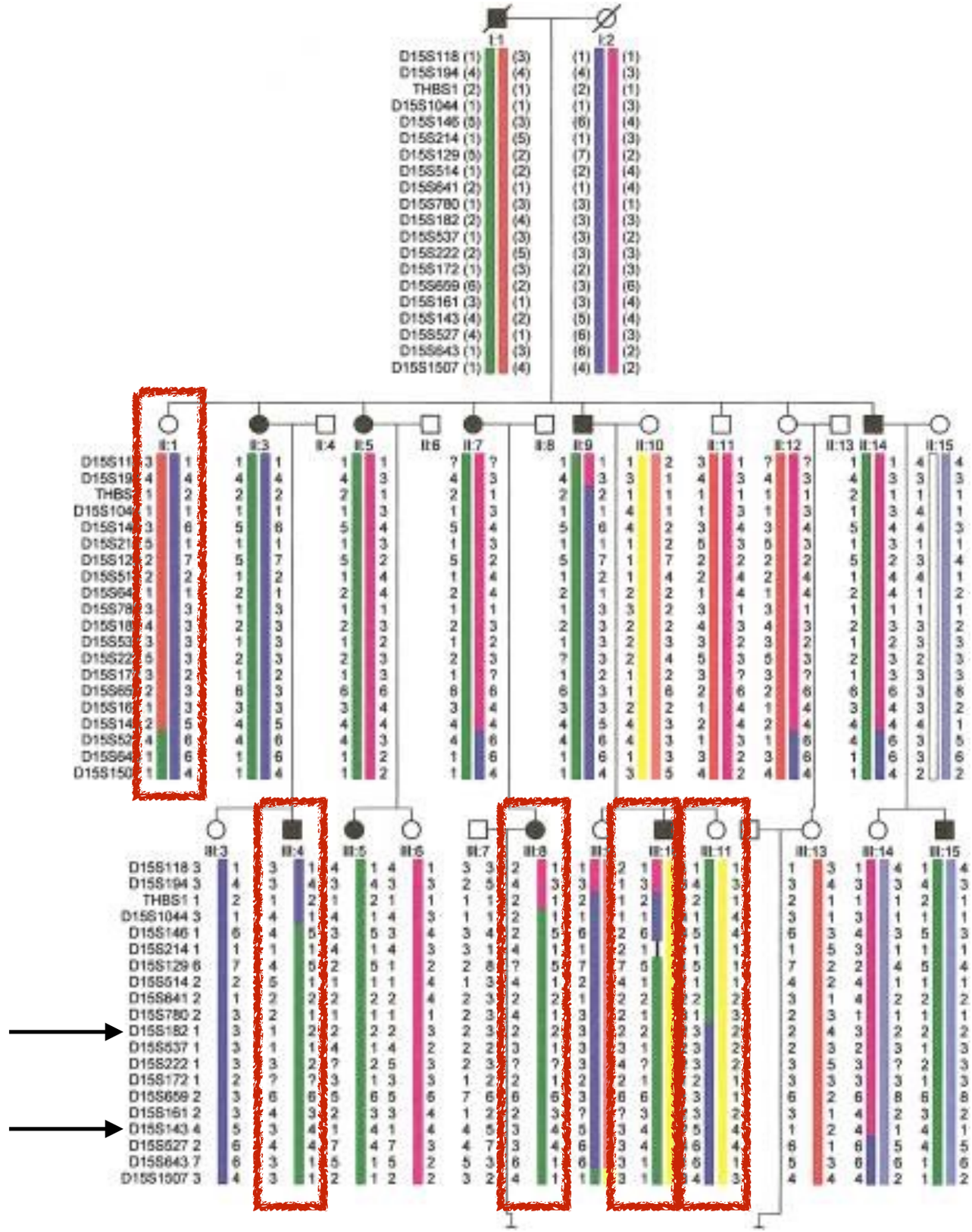
- Look at “all” markers in the genome simultaneously
- Dominant disorders mean look for heterozygous mutations linked to shared blocks of variants (haplotypes)
- X-linked lets you focus on the X chromosome
- Variants should be private to the family and deleterious

Genome-wide linkage mapping data

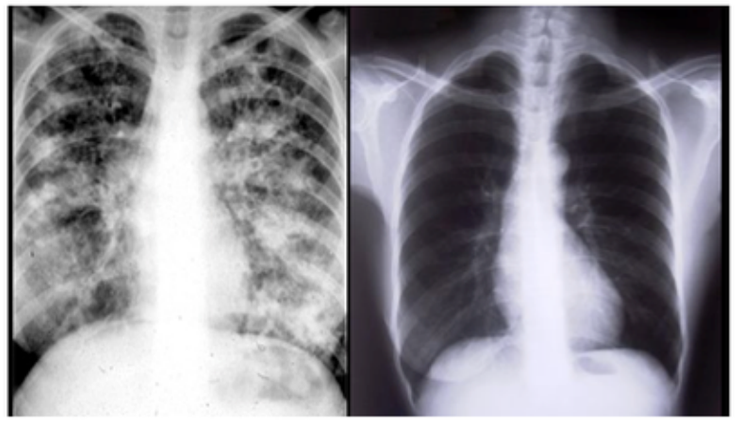


Lystig *Genetics* 2003

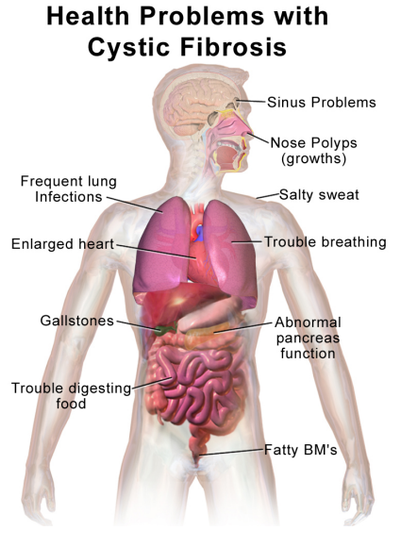
Linked markers on chromosomes allow for haplotype mapping



What about cystic fibrosis?



Cystic Fibrosis Lung Healthy Lung



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

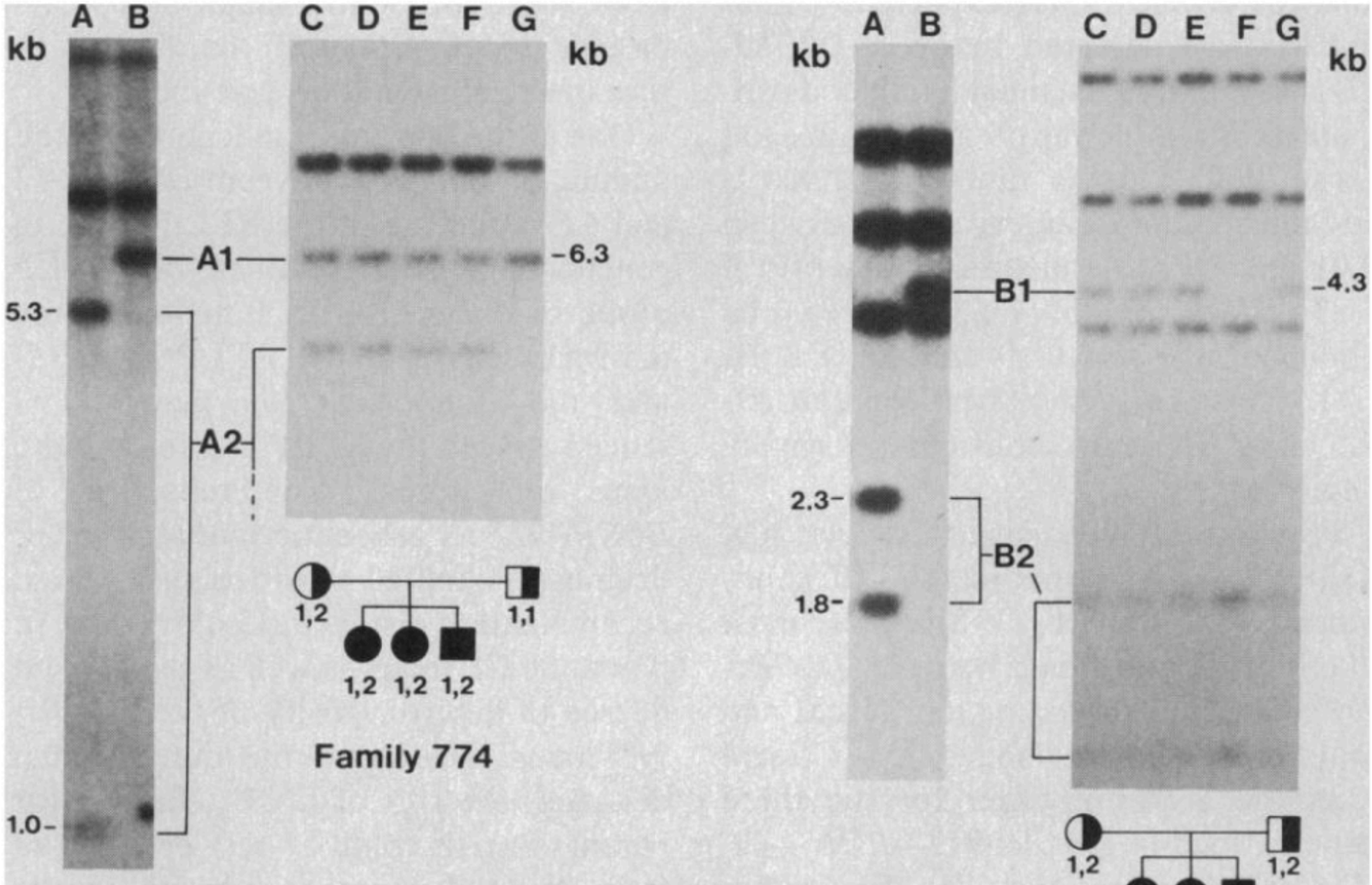


Fig. 2. Southern gel transfer hybridization analysis of human DNA with Lam4-917 probe. Two micrograms of each DNA sample was incubated with 10 units of restriction endonuclease at conditions specified by the supplier (New England Biolabs). The digested DNA samples were size-fractionated by electrophoresis in 0.8 percent agarose gels, transferred to Zetabind membranes (AMF Cuno, manufacturer) and hybridized with radioactive DNA probes

Tsui and Buchwald *Science* 1985

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.*

Tsui and Buchwald *Science* 1985