Genetic Mapping of Mendelian Traits

Source:

Human Molecular Genetics

By

Tom Strachan and Andrew Read

PKT

Two-point Mapping
Identifying recombinants in Pedigrees

- **Gathering information on the family history and drawing a family tree or pedigree. (Fig. 1)**
- **Genotyping each member of the family including affected and non-affected using genetic markers, like microsatellites, SNPs, etc. (Fig. 1, Table 1)**
- **Identifying recombinants and non- recombinants in the pedigree. (Fig. 1 & 2)**
- **Calculation of Recombination fraction (RF or the proportion of recombinant gametes between two loci, A and B). (Fig. 1)**
- **Appropriate statistical test is to be applied to see if RF is significantly different from 0.5 (i.e., null hypothesis of no linkage)**

Fig. 1: Pedigree, Genotyping and Identifying Recombinant and Non-Recombinants

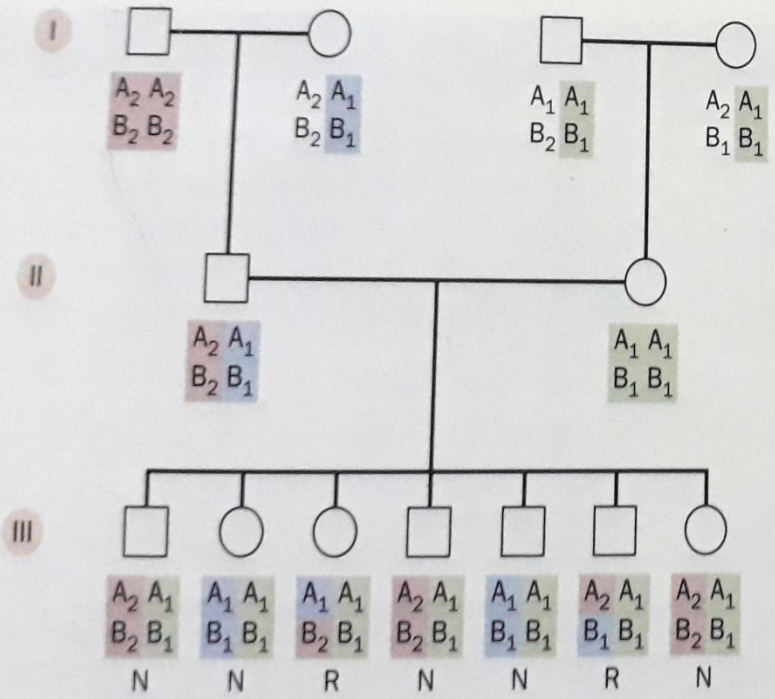

Figure 14.1 Recombinants and non-recombinants. In this family, there are two loci (*A* and *B*) at which alleles (A_1 and A_2, B_1 and B_2) are segregating. Colored boxes mark combinations of alleles that can be traced through the pedigree. In generation III, we can distinguish people who received non-recombinant (N; A_1B_1 or A_2B_2) or recombinant (R; A_1B_2 or A_2B_1) sperm from their father (II_1). Their mother (II_2) is homozygous at these two loci, and so we cannot identify which individuals in generation III developed from non-recombinant or recombinant oocytes.

Fig. 2: Meiotic Cross-over and the proportion of Recombinants and Non-Recombinants

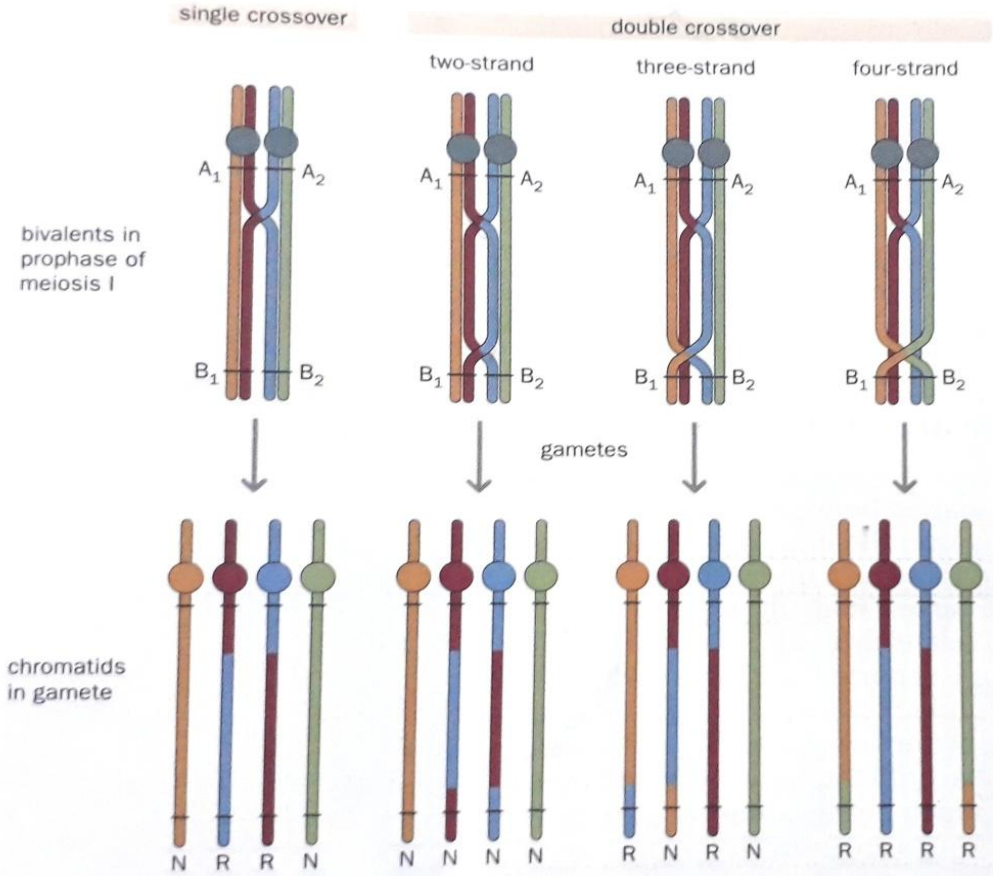

Figure 14.3 Single and double crossovers. The figure shows the chromosomal events that determine whether a gamete is recombinant or non-recombinant for the two loci *A* and *B*. Each crossover involves two of the four chromatids of the two synapsed homologous chromosomes. One chromosome carries alleles A_1 and B_1, the other alleles A_2 and B_2. Chromatids in the gametes labeled N carry a parental combination of alleles (A_1B_1 or A_2B_2). Chromatids labeled R carry a recombinant combination (A_1B_2 or A_2B_1). Note that recombinant and non-recombinant are defined only in relation to these two loci. For example, in the result of the three-strand double crossover, the second chromatid from the left has been involved in crossovers—but it is non-recombinant for loci *A* and *B* because it carries alleles A_1 and B_1, a parental combination. A single crossover generates two recombinant and two non-recombinant chromatids (50% recombinants). The three types of double crossover occur in random proportions, so the average effect of a double crossover is to give 50% recombinants.

Recombination fractions do not exceed 0.5, however great the distance between two loci

Table 1: Genetic Markers

A marker is any polymorphic Mendelian character that can be used to follow a chromosomal segment through a pedigree.

TABLE 14.1 THE DEVELOPMENT OF MARKERS FOR HUMAN GENETIC MAPPING

Type of marker	Dates used	No. of loci	Features
Blood groups	1910–1960	~20	May need fresh blood, and rare antisera. Genotype cannot always be inferred from phenotype because of dominance. No easy physical localization
Electrophoretic mobility variants of serum proteins	1960–1985	~30	May need fresh serum, and specialized assays. No easy physical localization. Often limited numbers of alleles and low frequency polymorphism
HLA tissue types	1970–	1	One set of closely linked loci, usually scored as a haplotype. Highly informative; many alleles with high or moderate frequency. Can only test for linkage to 6p21.3
DNA RFLPs	1975–	$>10^5$	Two-allele markers, maximum heterozygosity 0.5. Assayed previously by Southern blotting, now by PCR. Easy physical localization
DNA VNTRs (minisatellites)	1985–1990	$\sim10^4$	Many alleles, highly informative. Assayed by Southern blotting. Easy physical localization. Tend to cluster near ends of chromosomes
DNA VNTRs (microsatellites)	1990–	$\sim10^5$	Many alleles, highly informative. Can be assayed by automated multiplex PCR. Easy physical localization. Distributed throughout genome
DNA SNPs	2000–	$\sim10^7$	Two-allele markers. Densely distributed across the genome. Massively parallel technologies (microarrays, etc.) allow a sample to be assayed for thousands of SNPs in a single operation. More stable over evolutionary time than microsatellites

RFLP, restriction fragment length polymorphism; SNP, single nucleotide polymorphism; VNTR, variable number of tandem repeats.

Two-point Mapping

- **Since people heterozygous for two different diseases are rare, disease-disease mapping is not possible. Hence, disease-marker mapping may be more appropriate. We expect that a randomly selected individual may have a good chance of being heterozygous for a marker, which is sufficiently polymorphic.**

- **Scoring recombinants in a pedigree is not always simple; the meiosis may some time not be informative (as per selected markers) or even the pedigrees may not be in phase (the transmission of alleles can not be clearly traced through pedigree). Complex pedigrees make it difficult to accurately calculate RF.**

Fig. 3: Pedigrees may be informative or un-informative for linkage

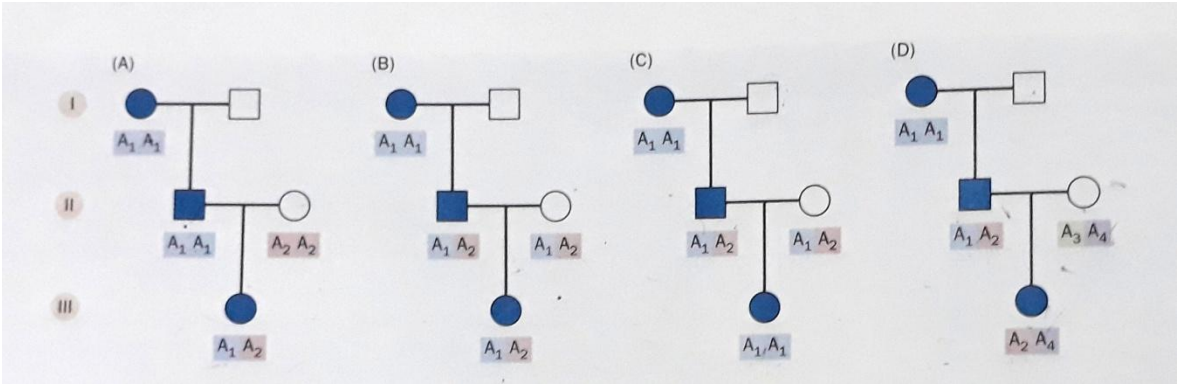

Figure 14.8 Informative and uninformative meioses. In each pedigree, individual II_1 has a dominant condition that we know he inherited along with marker allele A_1 because he inherited the condition from his mother, who is homozygous A_1A_1. Is the sperm that produced his daughter recombinant or non-recombinant between the loci for the condition and the marker? (A) This meiosis is uninformative: the father (II_1) is homozygous and so his marker alleles cannot be distinguished. (B) This meiosis is uninformative because both parents are heterozygous for alleles A_1 and A_2: the child could have inherited A_1 from the father and A_2 from the mother, or vice versa. (C) This meiosis is informative and non-recombinant: the child inherited A_1 from the father. (D) This meiosis is informative and recombinant: the child inherited A_2 from the father.

Fig. 4: Pedigrees may some time be complex enough (Phase unknown) to calculate RF

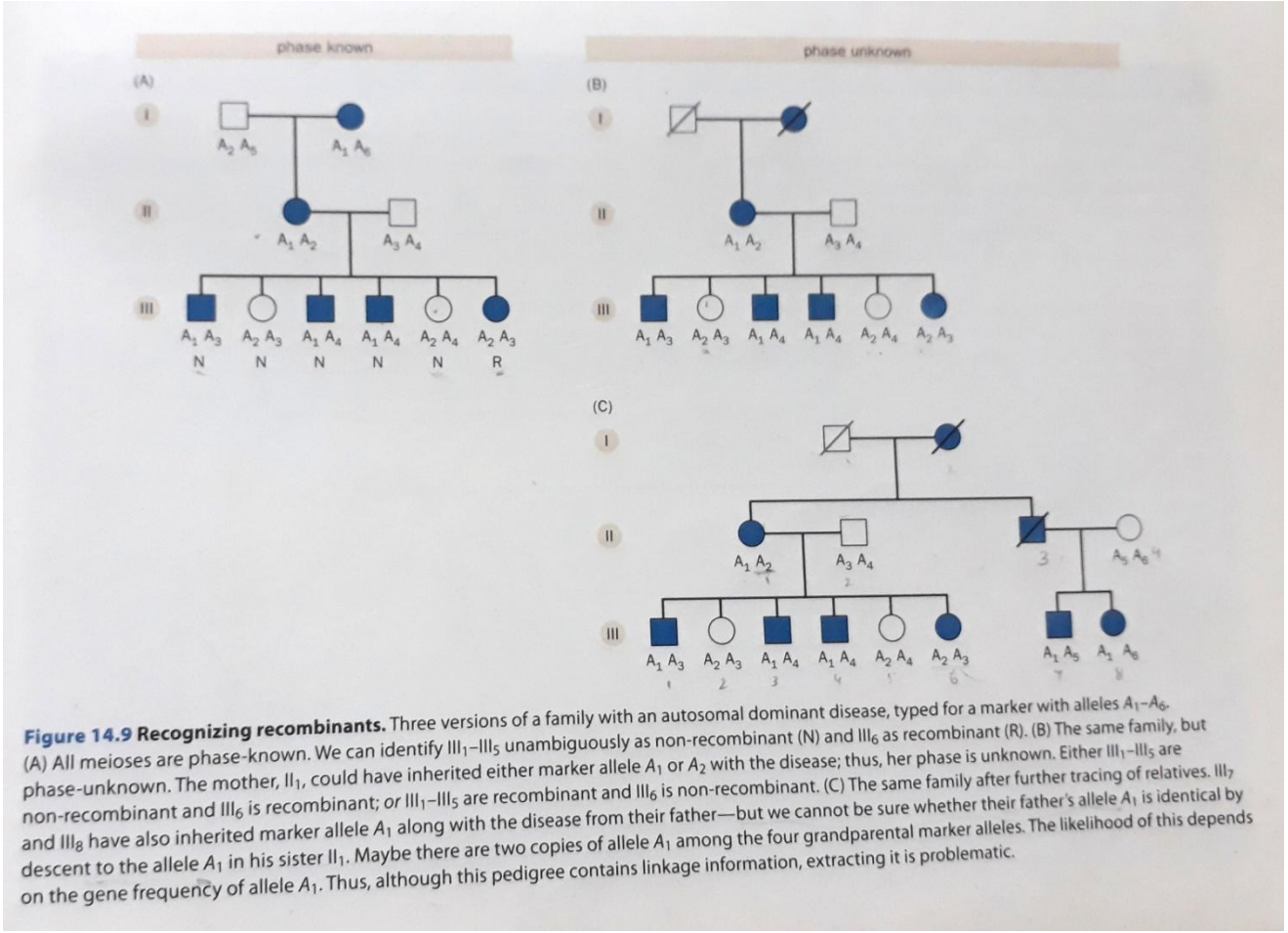

Figure 14.9 Recognizing recombinants. Three versions of a family with an autosomal dominant disease, typed for a marker with alleles A_1–A_6. (A) All meioses are phase-known. We can identify III_1–III_5 unambiguously as non-recombinant (N) and III_6 as recombinant (R). (B) The same family, but phase-unknown. The mother, II_1, could have inherited either marker allele A_1 or A_2 with the disease; thus, her phase is unknown. Either III_1–III_5 are non-recombinant and III_6 is recombinant; *or* III_1–III_5 are recombinant and III_6 is non-recombinant. (C) The same family after further tracing of relatives. III_7 and III_8 have also inherited marker allele A_1 along with the disease from their father—but we cannot be sure whether their father's allele A_1 is identical by descent to the allele A_1 in his sister II_1. Maybe there are two copies of allele A_1 among the four grandparental marker alleles. The likelihood of this depends on the gene frequency of allele A_1. Thus, although this pedigree contains linkage information, extracting it is problematic.

Fig. 5: Calculation of Lod Score for different types of pedigrees (in Fig. 4).

The most likely RF is the value at which the lod score is maximum. To calculate overall linkage the lod scores can be added-up across families.

BOX 14.2 CALCULATION OF LOD SCORES

Let us take as an example a calculation of lod scores for the families in Figure 14.9A and B.

- Given that the loci are truly linked, with recombination fraction θ, the likelihood of a meiosis being recombinant is θ and the likelihood of its being non-recombinant is $1 - \theta$.
- If the loci are, in fact, unlinked, the likelihood of a meiosis being either recombinant or non-recombinant is 0.5.

Family A

There are five non-recombinants $(1 - \theta)$ and one recombinant (θ).
The overall likelihood, given linkage, is $(1 - \theta)^5 \times \theta$.
The likelihood, given no linkage, is $(0.5)^6$.
The likelihood ratio is $(1 - \theta)^5 \times \theta / (0.5)^6$.

θ	0	0.1	0.2	0.3	0.4	0.5
Z	$-\infty$	0.577	0.623	0.509	0.299	0

The lod score, Z, is the logarithm of the likelihood ratio.

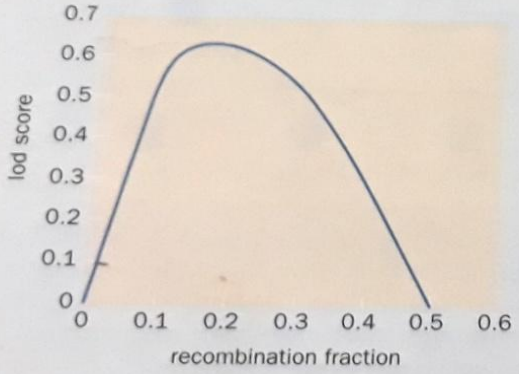

Figure 1 Lod score curve for family A.

Family B

The mother (II_1 in Figure 14.9B) is phase-unknown.
If she inherited A_1 with the disease, there are five non-recombinants and one recombinant.
If she inherited A_2 with the disease, there are five recombinants and one non-recombinant.
The overall likelihood is $\frac{1}{2}[(1 - \theta)^5 \times \theta/(0.5)^6] + \frac{1}{2}[(1 - \theta) \times \theta^5/(0.5)^6]$.
This allows for either possible phase, with equal prior probability.
The lod score, Z, is the logarithm of the likelihood ratio.

θ	0	0.1	0.2	0.3	0.4	0.5
Z	$-\infty$	0.276	0.323	0.222	0.076	0

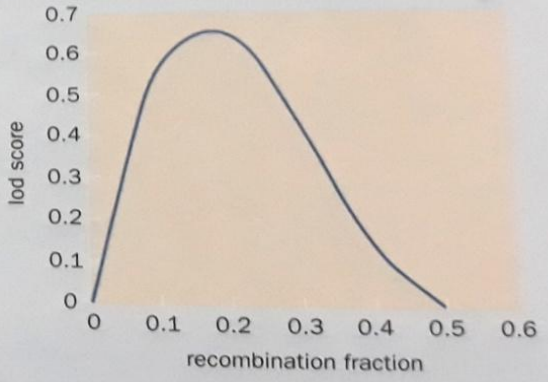

Figure 2 Lod score curve for family B.

Family C

At this point non-masochists turn to the computer.

Fig. 6: LOD (Z) Score curves.

LOD: The overall likelihood of the pedigree can be calculated based on the alternative assumptions that the loci are linked (RF=θ) or not linked (RF=0.5). The ratio of these two likelihoods gives the odds of linkage and log of the odd is the lod score.

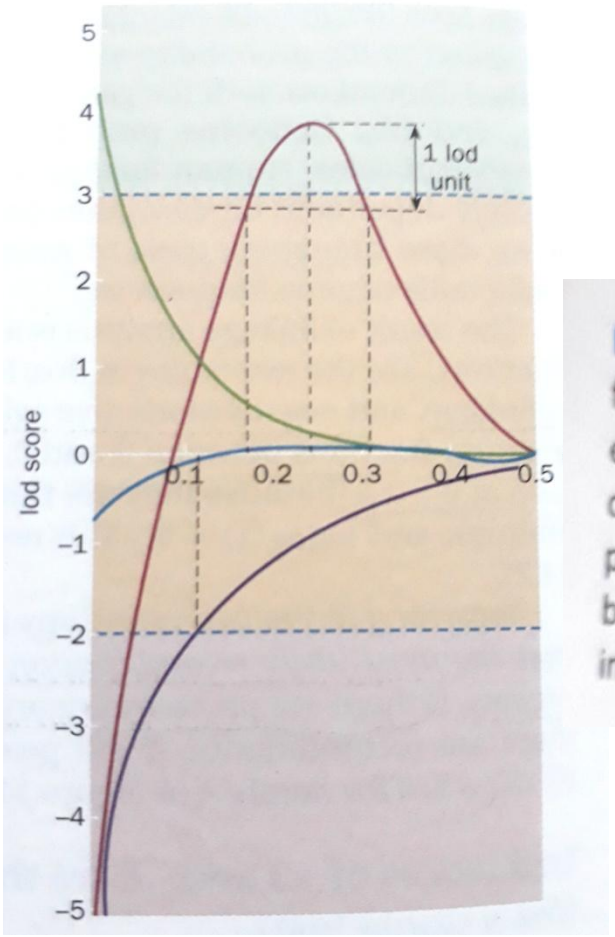

Figure 14.10 Lod score curves. Graphs of lod score against recombination fraction (θ) from a hypothetical set of linkage experiments. The green curve shows evidence of linkage ($Z > 3$) with no recombinants. The red curve shows evidence of linkage ($Z > 3$), with the most likely recombination fraction being 0.23. The purple curve shows linkage excluded ($Z < -2$) for recombination fractions below 0.12; inconclusive for larger recombination fractions. The blue curve is inconclusive at all recombination fractions.

LOD Score

- **Lod score of +3 and -2 are the criteria for linkage and exclusion, respectively.**
- **RF between 0-0.5 are meaningful and that all lod scores are zero at θ=0.5 (because they are then measuring the ratio of two identical probabilities and $\log_{10}(1)=0$**
- **Lod score is maximum at θ=0, if there are no recombinants**
- **Z will peak at the most likely RF for phase-known families, but difficult to know for phase unknown families.**

Threshold of Significance

- **Z=3.0 is the threshold for accepting linkage with 5% chance of error.**
- **Z values between -2 and +3 are inconclusive**
- **Generally, $p<0.05$ is used as the threshold of significance, but**
- **Z=3.0 corresponds to 1000:1 odds [$\log_{10}(1000)=3.0$]**
- **Such a stringent threshold is on account of the inherent improbability that the loci (marker and disease gene) chosen at random on 22 autosomes are linked, and even if they are , may be unlinked (separated enough). So, to rule out or accept such stringency is required.**
- **For X-linked characters the lod of 2.3 is required to consider linkage between a gene and a marker on the X-chromosome.**
- **Exclusion mapping: Negative lod score excludes a possible (less likely) candidate gene and increase the accuracy of reaching to right (more likely) one.**

Multi-point Mapping

- **Linkage analysis with more than two loci analysed simultaneously.**
- **This can establish chromosomal order of a set of linked loci (as three-point test cross). The rarest class is one with double cross-over.**
- **Ideally the whole genome is screened for linkage and the full data set would be used to calculate the likelihood at each location across the genome.**
- **Multipoint mapping helps limitation of limited informativeness of markers. Simultaneous linkage analysis of the disease with two or more markers gives the full information.**

Multipoint Disease-Marker Mapping

- **Multipoint mapping is useful in constructing marker framework maps.**
- **There are n!/2 possible order of n markers and current maps have hundreds of markers per chromosome.**
- **Markers then can be typed by PCR, can be used as sequence tagged sites and physically localized, either by database searching or experimentally using radiation hybrids.**
- **The result is a physically anchored marker framework.**
- **Linkmap or Genehunter program can help reach the disease gene.**
- **The result is a curve of lod scores against map position. When the curve stays below-2 across a region, the disease locus is excluded (Fig. 7). The highest peak marks the most likely location.**

Fig. 7: Multi-point Mapping

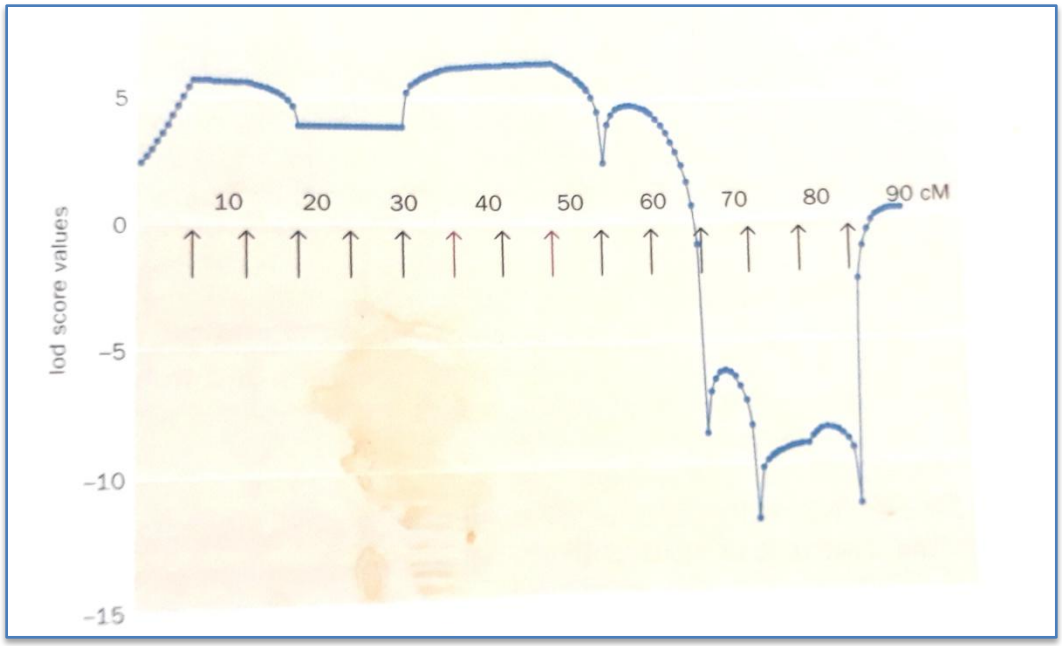

Figure 14.11 Multipoint mapping in humans. The data from Table 14.3 on linkage between migraine with aura and markers from chromosome 15, analyzed by a multipoint mapping program (Simwalk). The genetic map of markers is taken as fixed; the disease locus is placed successively at each point on the genetic marker framework (marker positions are shown by black arrows; colored arrows indicate the position of two candidate genes) and the multipoint lod score is calculated for each possible location of the disease locus. Lod scores dip to strongly negative values near the position of loci that show recombinants with the disease. The highest peak marks the most likely location; odds in favor of this location are measured by the degree to which the highest peak overtops its rivals. The peak lod score is 6.54; that is, a tenfold stronger likelihood than the maximum of 5.56 obtained by two-point analysis. There were no recombinants within this region in this data set, so the candidate region could not be narrowed down further. [From Russo L, Mariotti P, Sangiorgi E et al. (2005) *Am. J. Hum. Genet.* 76, 327–333. With permission from Elsevier.]