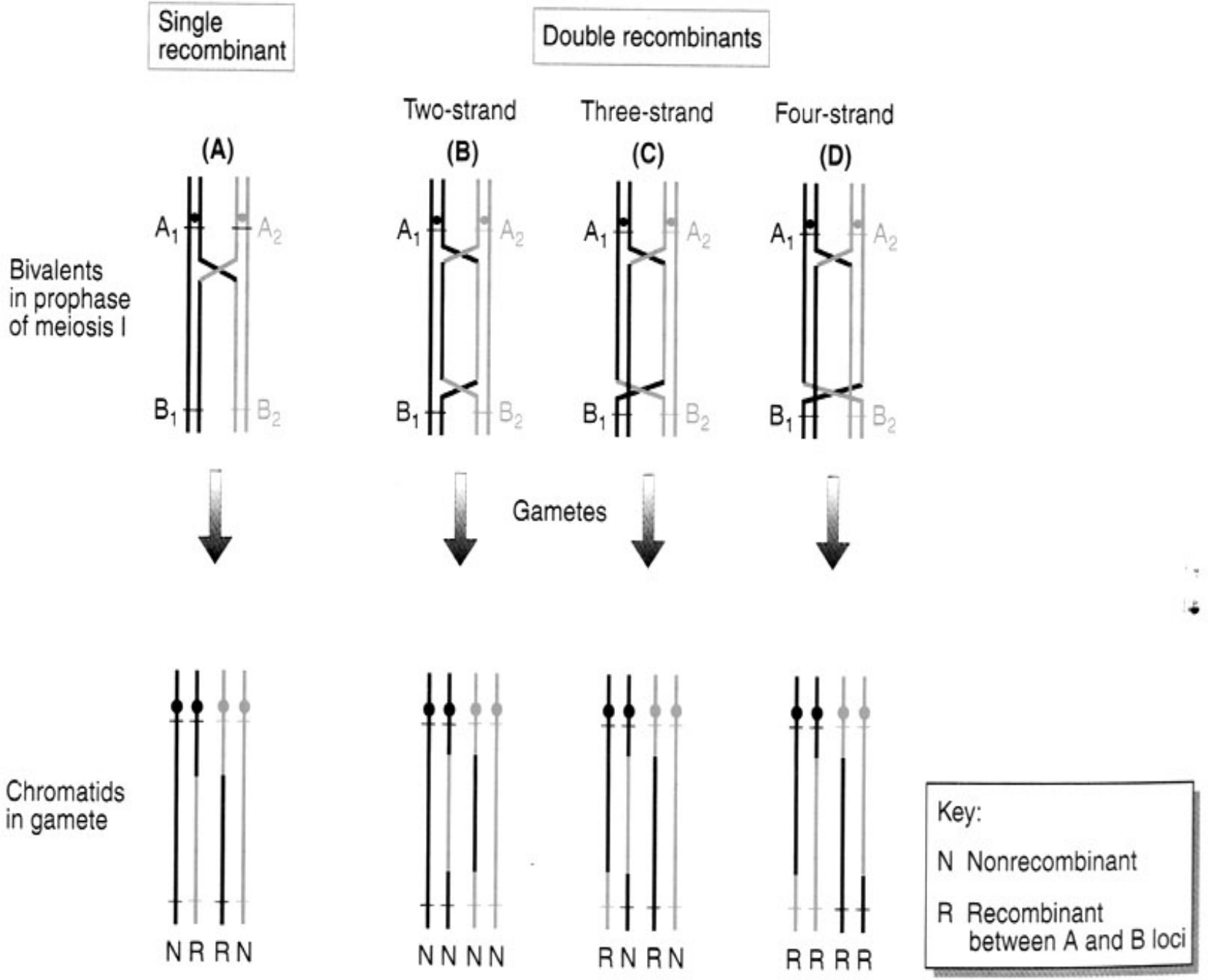


Figure 11.2: Single and double recombinants.

There are about 60 chiasmata in each male meiosis, corresponding to 30 crossovers per male germ cell

1 Morgan (M) is the genetic interval corresponding to 1 crossover in the genome; thus, the genetic length of the male genome is about 30 Morgans (or 3000 cM)

Gene (or marker) loci on the same chromosome are linked if their alleles stay together during transmission from parent to offspring significantly more often than not (i.e., if their recombination frequency theta (Θ) is < 0.5).

For small values of Θ (e.g., < 0.10), the recombination fraction in % is equivalent to the genetic distance in cM (5 % rec. --> 5 cM)

... but for Θ --> 0.5 , the genetic distance approaches ∞

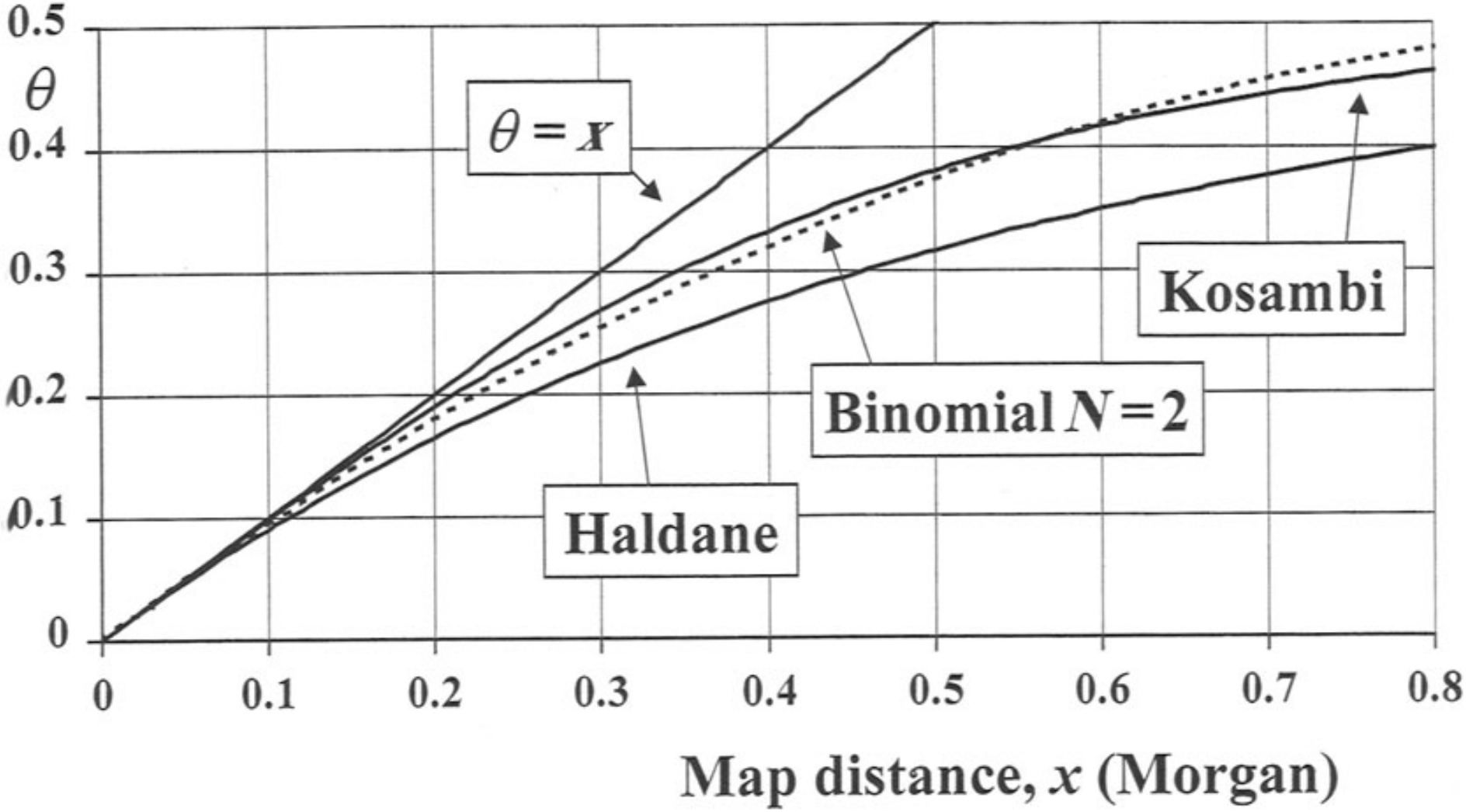


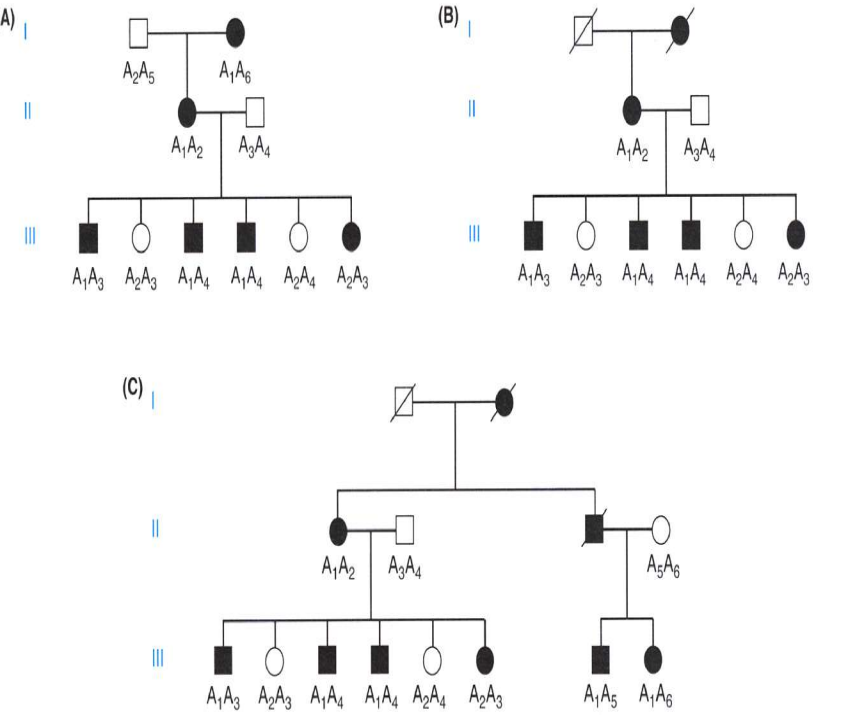
Figure 1.4. Graphs of several map functions.

The LOD (= log of odds) score:
a tool for the quantitiative assessment of linkage between disease genes and (other) genetic marker loci in families

Principle:
a) calculation of the likelihood that disease and marker allele co-segregate in the pattern observed in a family under the assumption that the two are linked (i.e., assuming that their true recombination frequency theta is < 0.5; e.g. 0.0, 0.1, 0.2, 0.3,....)

b) calculation of the likelihood of the segregation pattern under the assumption that the two loci are unlinked (theta = 0.5)

c) calculation of the log of (a/b) for any value of theta between 0.0 and 0.5 (where it becomes zero) --> **the LOD score curve**



Recognizing recombinants: three versions of a family with an autosomal dominant disease, typed for a marker A.

Calculation of lod scores for the families in *Figure 11.4*

- Given that the loci are truly linked, with recombination fraction θ, the likelihood of a meiosis being non-recombinant is $1 - \theta$ and the likelihood of it being recombinant is θ.
- If the loci are in fact unlinked, the likelihood of a meiosis being either recombinant or nonrecombinant is 1/2.

Family A

There are five recombinants and one [redacted]recombinant.
The overall likelihood, given linkage, is $(1 - \theta)^5.\theta$
The likelihood given no linkage is $(1/2)^6$
The likelihood ratio is $(1 - \theta)^5.\theta / (1/2)^6$
The lod score, Z, is the logarithm of the likelihood ratio.

θ	0	0.1	0.2	0.3	0.4	0.5
Z	− infinity	0.577	0.623	0.509	0.299	0

Family B

II_1 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 nonrecombinant.
The overall likelihood is $\frac{1}{2}[(1 - \theta)^5.\theta / (1/2)^6] + \frac{1}{2}[(1 - \theta).\theta^5 / (1/2)^6]$. 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	− infinity	0.276	0.323	0.222	0.076	0

Family C

At this point nonmasochists turn to the computer.

Pedigree sizes (no. of informative meioses) required to 'prove' linkage:

co-segregation of two markers at one meiosis increases the likelihood ratio (odds) for linkage by a factor of 2, or the log of odds (LOD score) by 0.301

thus, co-segregation of two X-chromosomal markers from a grandfather to 2 daughters and to their 7 grandsons (7 informative meioses) yields a LOD score of 2.107

10 informative meioses are required to obtain statistically significant evidence for linkage between two autosomal loci (LOD score > 3)

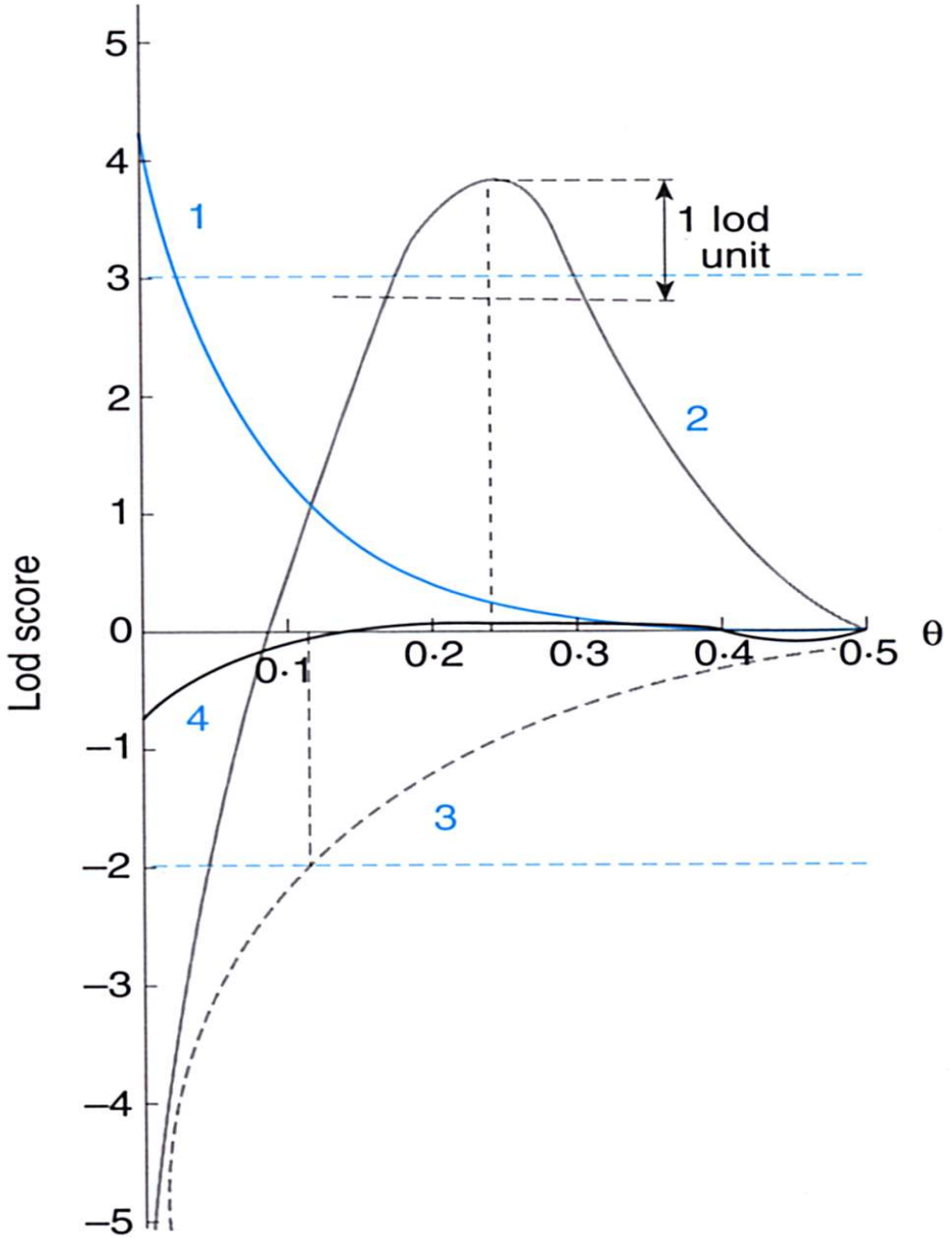


Figure 11.5: Lod score curves.

A LOD score of > 3 is considered as significant for linkage between autosomal genes and markers (for X-linked markers: LOD score > 2 !), and for a given recombination fraction Θ, linkage is excluded if the LOD score is < - 2

Rule of thumb for the relation between genetic and physical distances in the human genome:

1 centiMorgan (cM) = 10^6 basepairs (1 Megabase = 1 Mb)

(because the physical map of the human genome is roughly 3 billion bp long and its genetic length is about 3000 cM)

but: - does not apply everywhere in the genome
- in female meiosis, genetic map is larger (chiasmata are more frequent)

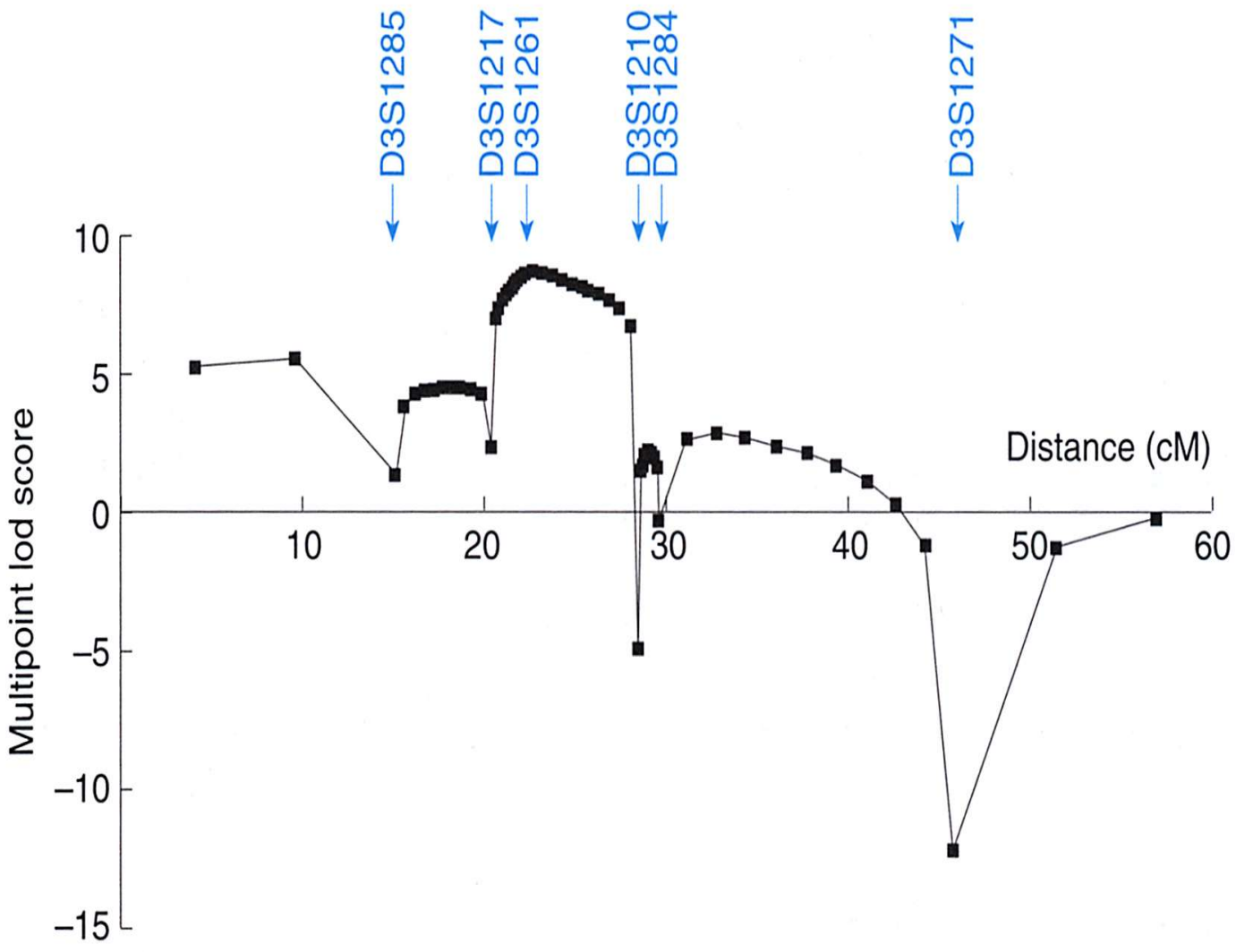


Figure 11.6: Multipoint mapping in man.

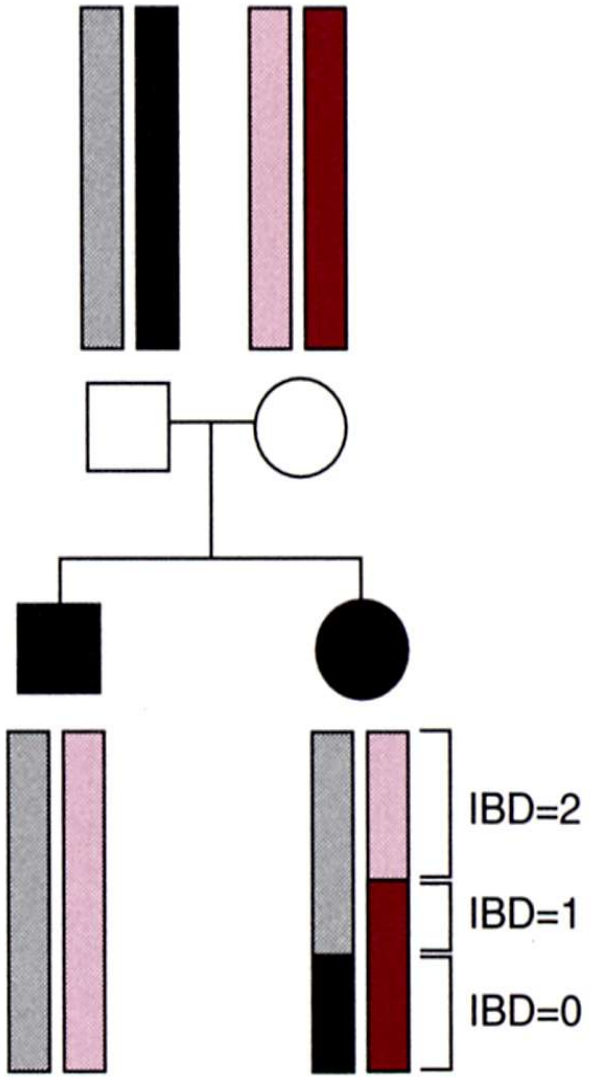


Figure 4.11. Sharing of alleles identical-by-descent in affected siblings. The parental chromosomes can be marked by the study of highly polymorphic markers at defined intervals; this leads to the unique identification of each homologous pair of parental chromosomes and to the identification of which segments were inherited by each offspring from a specific parent. Comparison of the offspring chromosomes identifies which segments are shared identical-by-descent (IBD) by the siblings. This sharing can be for both parental copies (2), one parental copy (1), or no sharing (0). By quantitating the degree of sharing at each chromosomal (genomic) site in many families one can identify those regions in which sharing is greater than expected, and thus likely to harbor a susceptibility allele. (Courtesy of Professor Aravinda Chakravarti, Department of Genetics, Case Western Reserve University.)

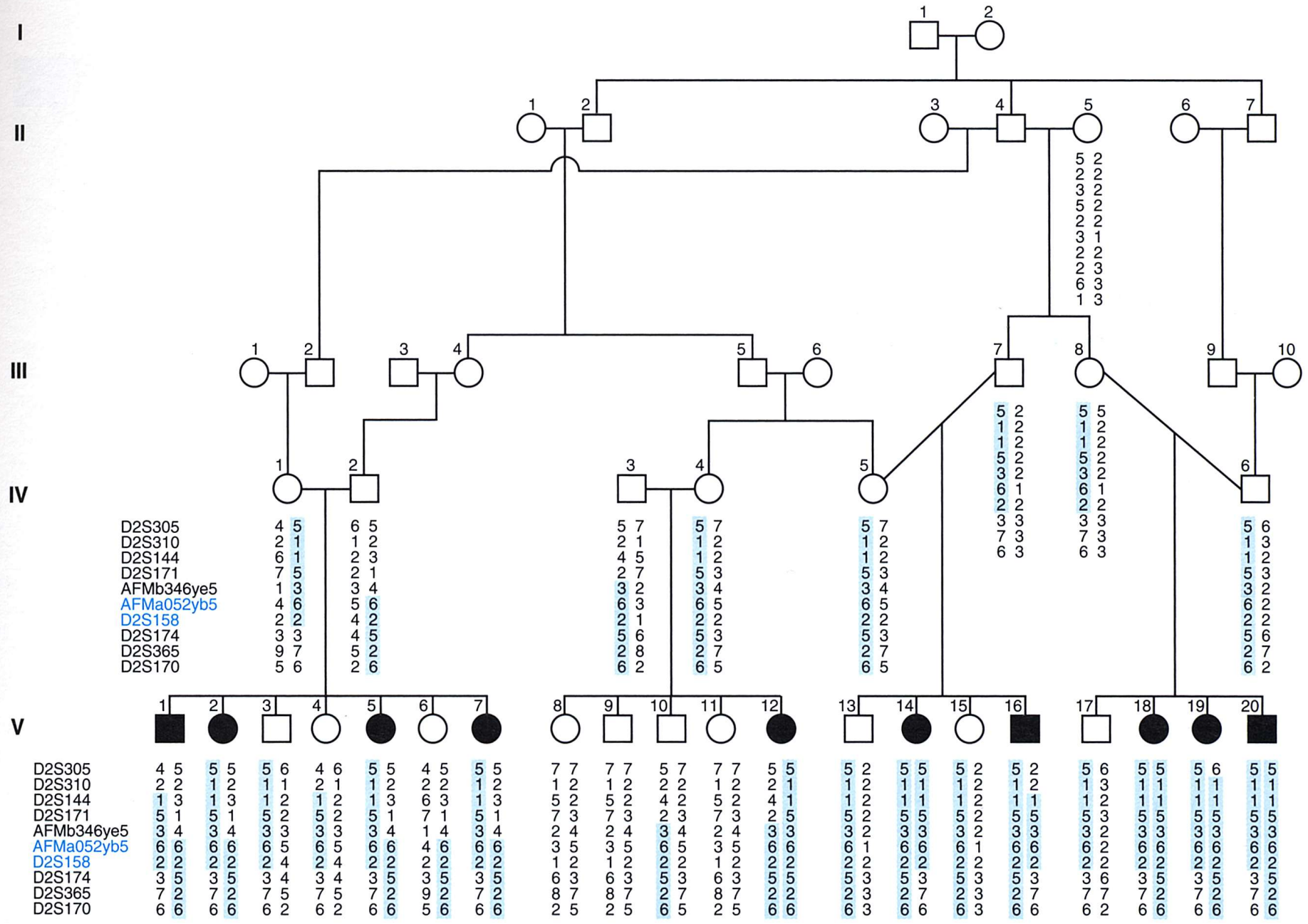


Figure 11.8 Autozygosity mapping.

Linkage disequilibrium ('Kopplungsungleichgewicht'):

genetic markers or gene defects are inherited as parts of chromosome segments which are limited by crossovers

if Θ is the probability of recombination separating two neighboring loci during one meiosis, $(1-\Theta)$ is their chance to stay together

For two descendants of the same common ancestor living n generations ago, this probability will be $(1-\Theta)^{2n}$

If this ancestor lived around 1550 (i.e., 22 generations ago), the chance of two closely linked markers ($\Theta = 0.01$) to stay together in both descendants would be $0.99^{44} = 64\%$!

Length of such evolutionarily conserved 'haplotypes' showing allelic associations depends on population history

Table 12.1: Allelic association in cystic fibrosis

Marker alleles	CF chromosomes	Normal chromosomes
X_1, K_1	3	49
X_1, K_2	147	19
X_2, K_1	8	70
X_2, K_2	8	25

Association is not necessarily due to linkage disequilibrium:

- marker could be directly responsible for the disease

- presence of associated factor might confer selective advantage to carrier of unlinked gene defect

- gene defect and associated marker might be confined to subset of the population (and be rare outside this subset) ('population stratification')

- association might be a statistical artefact (e.g., if n loci are tested, significance levels have to be raised accordingly)

- association due to linkage disequilibrium will only be observed if most disease-predisposing chromosomes are derived from common ancestor

Table 4.6. Frequency Distribution of Systolic Blood Pressure Determined by a Two-Locus Two-Allele Model[a]

	AA *1/4*	*Aa* *1/2*	*aa* *1/4*
BB 1/4	1/16 (*40*)	2/16 (*30*)	1/16 (*20*)
Bb 1/2	2/16 (*30*)	4/16 (*20*)	1/16 (*10*)
bb 1/4	1/16 (*20*)	2/16 (*10*)	1/16 (*0*)

[a] The numbers in parentheses indicate the increment (in mm Hg pressure) to the systolic blood pressure above a basal level of 100 mm Hg contributed by each genotype.

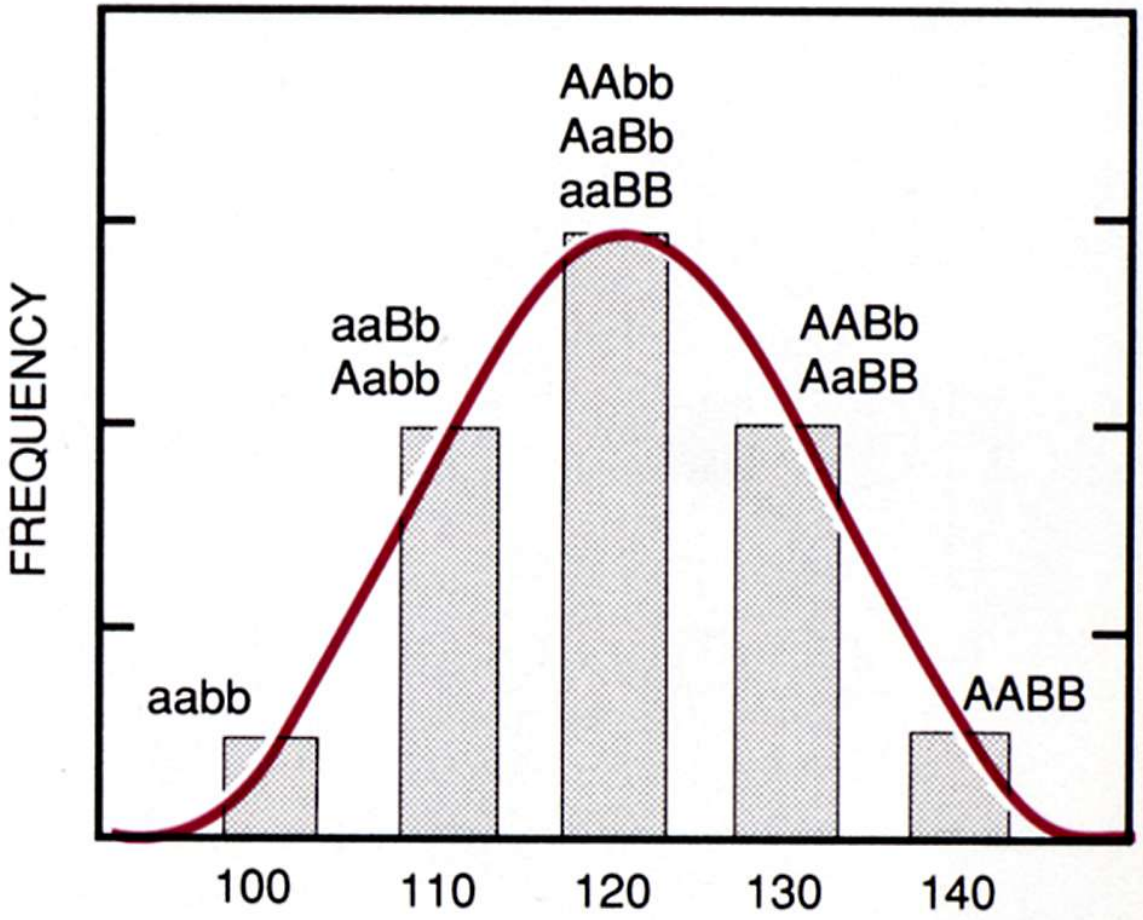


Figure 4.3. Frequency distribution of systolic blood pressure determined by a two-locus two-allele model. See text for explanation.

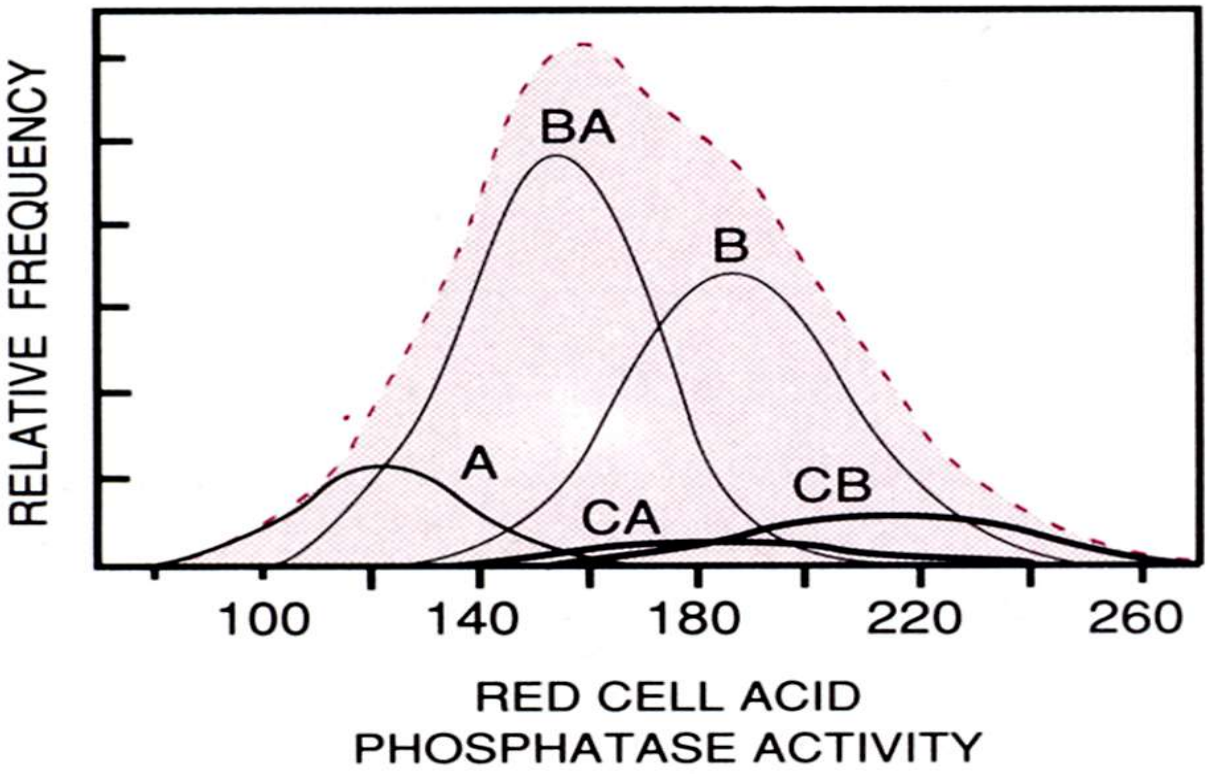


Figure 4.4. Distribution of red cell acid phosphatase activities in the general population (*broken red line*) and in individuals with the separate phenotypes. The *solid* curves are constructed from the data on the different phenotypes as found in the British population. (From Harris H. The principles of human biochemical genetics. 3rd ed. Amsterdam: Elsevier/North-Holland, 1980:186.)

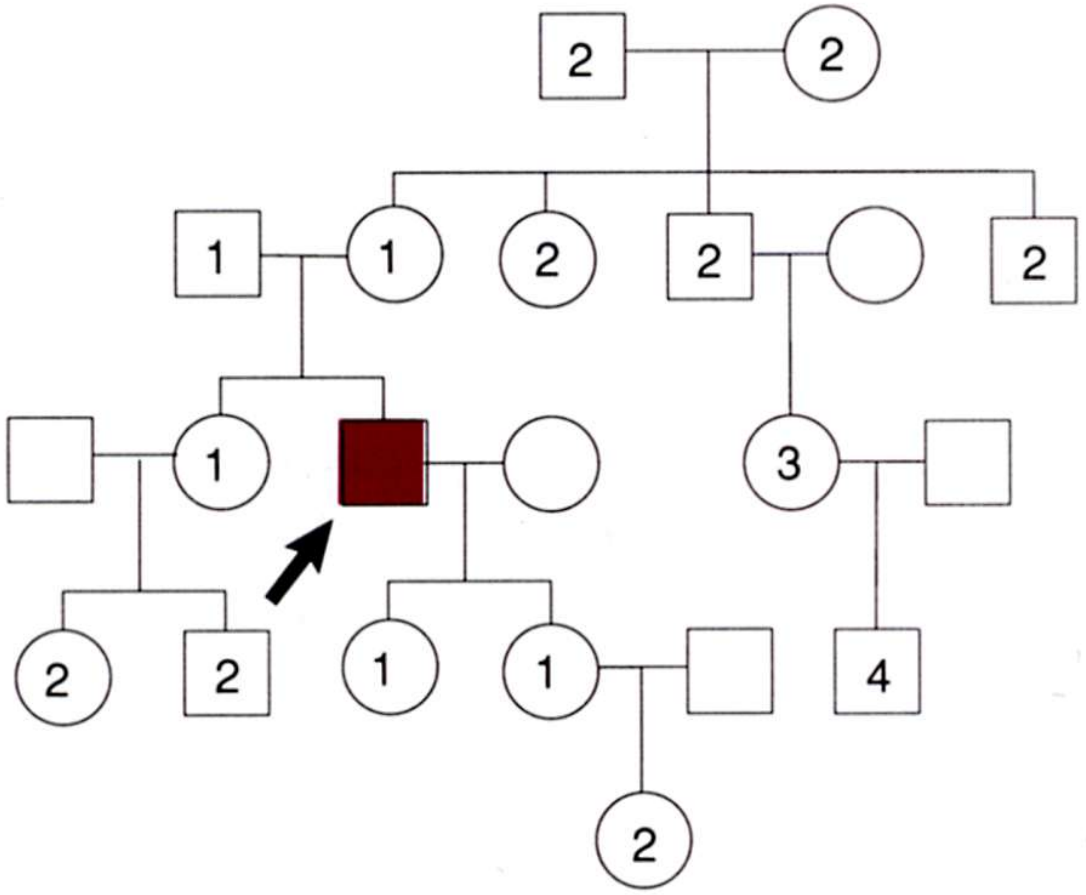


Figure 4.5. Pedigree showing degree of relationship. In this figure, the *numbers inside the symbols* indicate the degree of relationship to the proband (*red symbol*).

Table 4.7. Correlation of Fingertip Ridge Counts among Relatives Compared with Expectations Based on the Proportion of Shared Genes[a]

Relationship	*Observed Correlation*	*Expected Correlation*
Monozygotic twins	0.95 ± 0.07	1.00
Dizygotic twins	0.49 ± 0.08	0.50
Siblings	0.50 ± 0.04	0.50
Parent-child	0.48 ± 0.04	0.50
Spouses	0.05 ± 0.07	0.00

[a] From Carter CO: Genetics of common disorders. Br Med Bull 25:52–57, 1969.

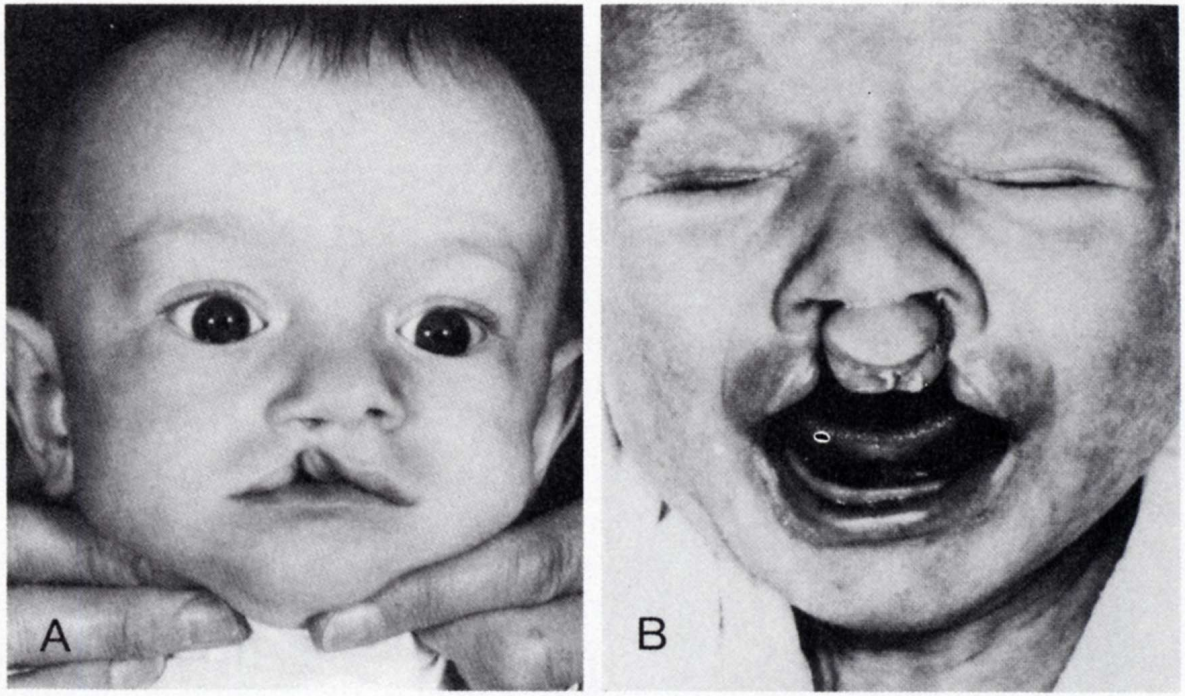


Figure 4.6. Children with cleft lip ± cleft palate. **A.** Child with unilateral cleft lip; **B.** Child with bilateral cleft lip and cleft palate. (From Ross RB, Johnson MC. Cleft lip and palate. Baltimore: Williams & Wilkins, 1972:131, 141.)

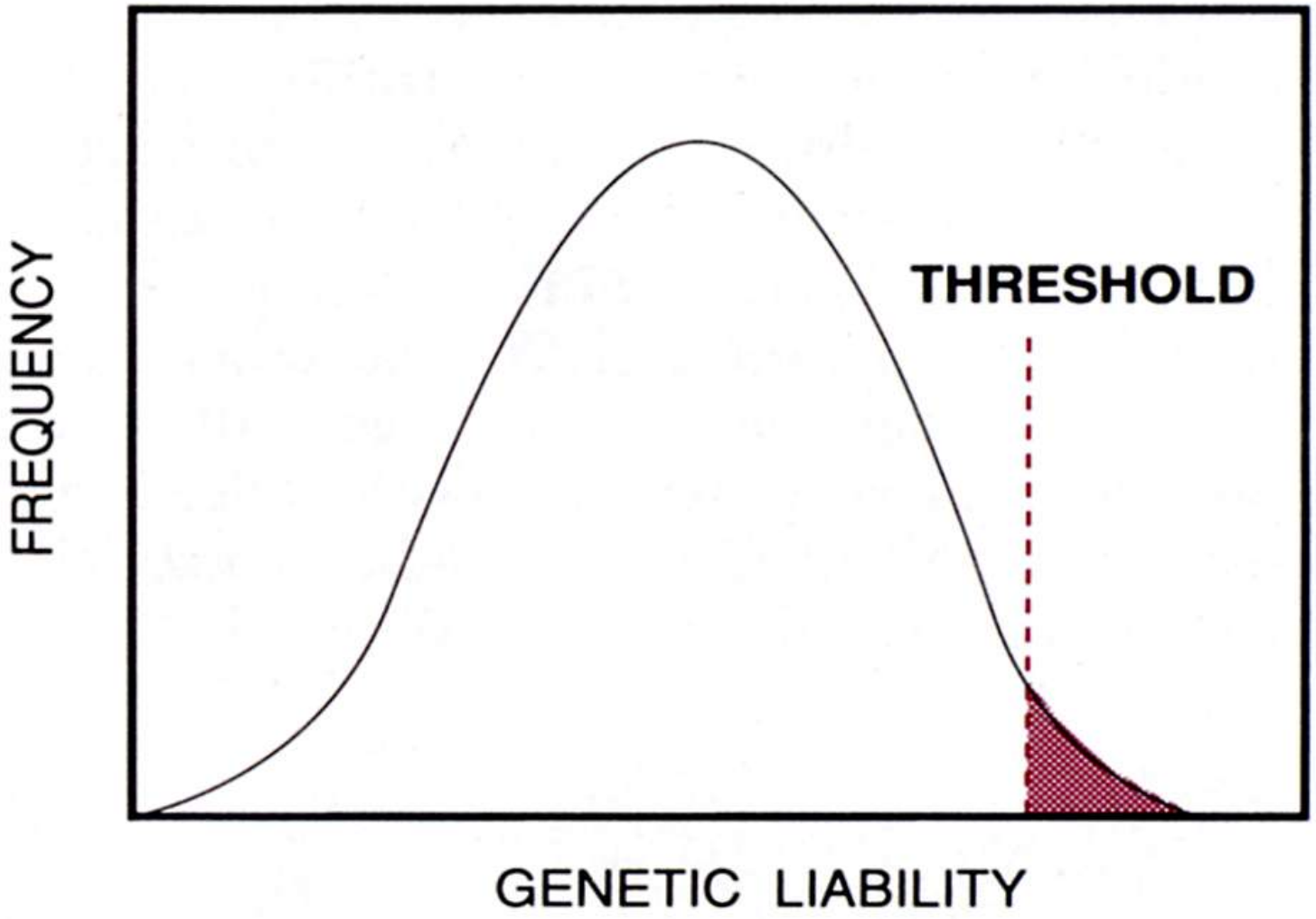
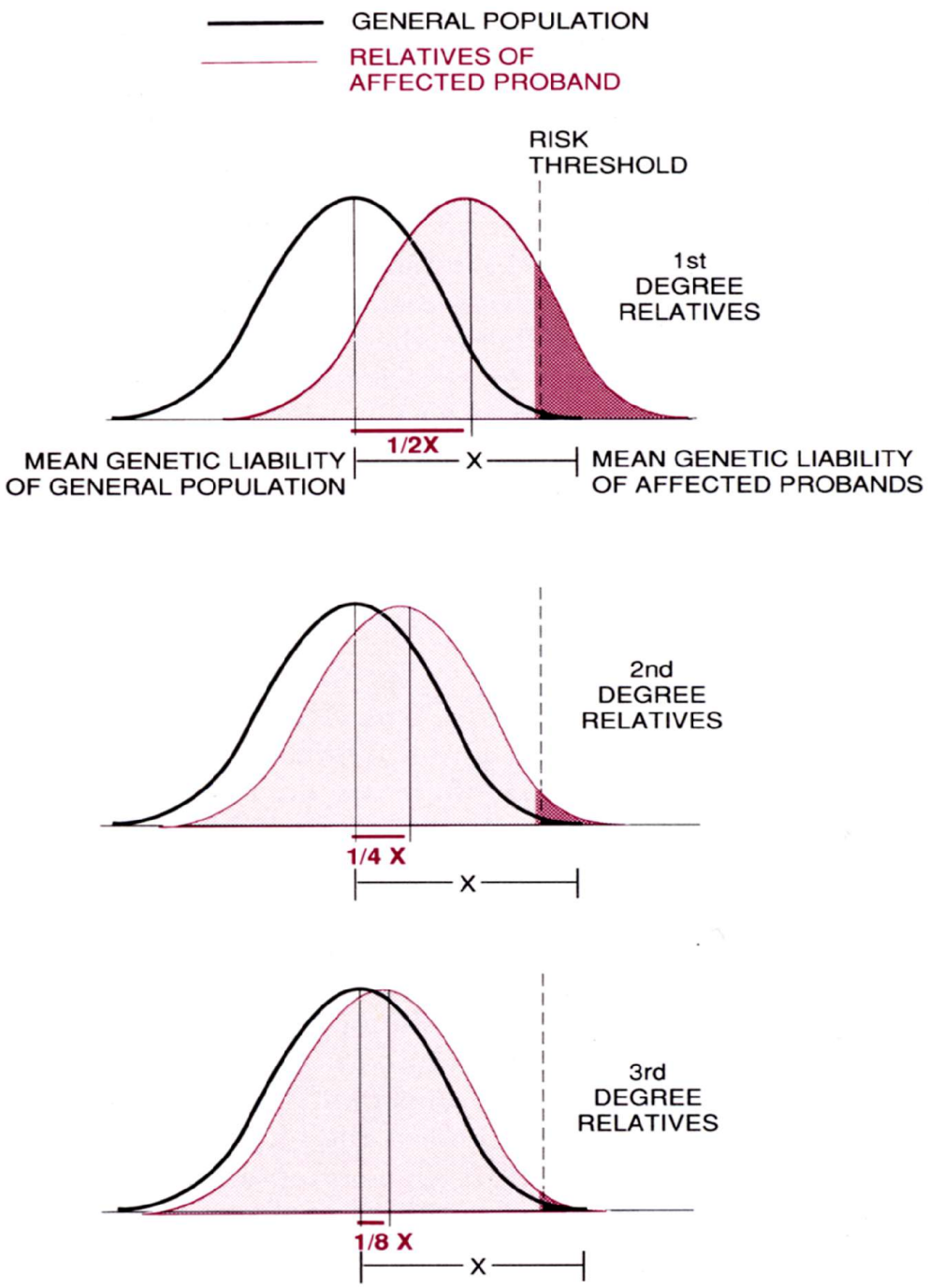


Figure 4.7. Threshold model of multifactorial inheritance. See text for details.

Figure 4.8. Multifactorial threshold model: distribution of genetically determined liability among relatives. The distribution of genetic liability in relatives of an affected proband is indicated by the *lightly shaded red area below* the *red curve*. *X* is the difference in mean genetic liability between affected probands and the general population. See text for details. (From Carter CO. Multifactorial genetic disease. Hosp Pract 1970;5:45–59.)



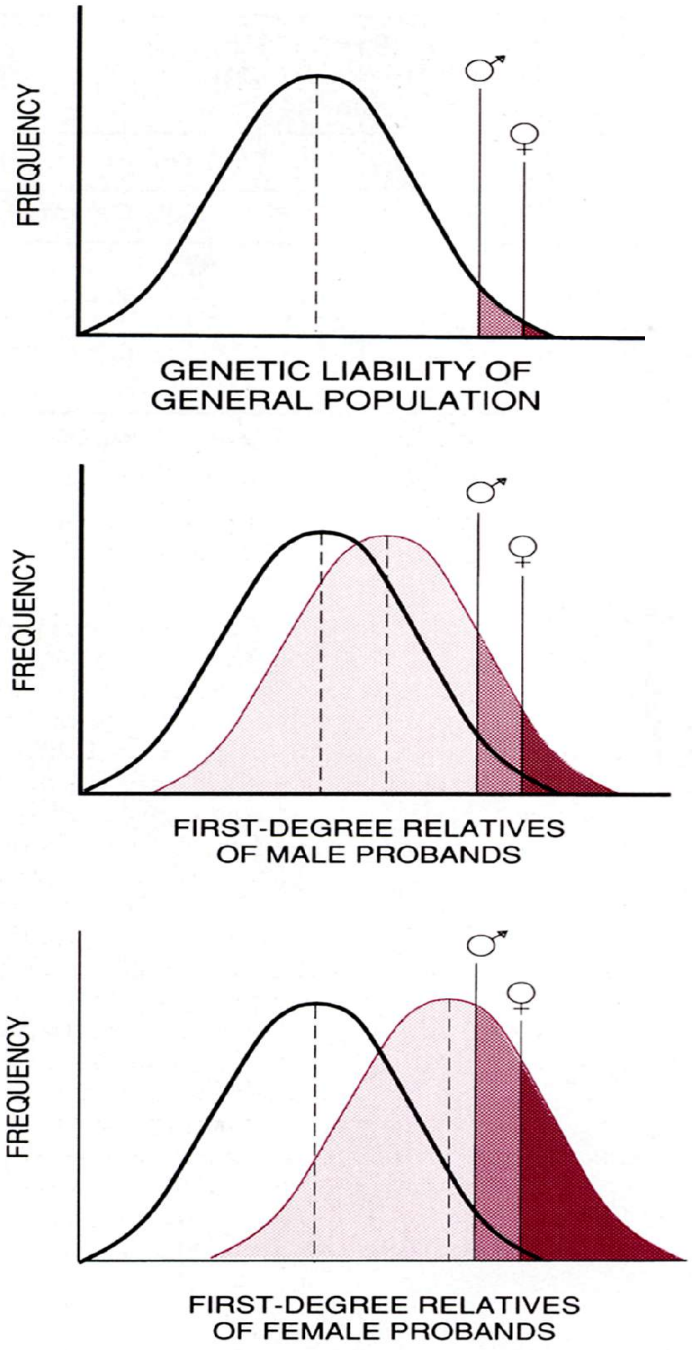


Figure 4.10. Multifactorial threshold model: explanation for sex differences in incidence of pyloric stenosis in probands and their relatives. The risk thresholds for males and females are indicated by the *solid vertical lines*. The distribution of genetic liability in relatives of male and female probands is indicated by the *red shaded areas below* the *red curve*. Affected individuals are indicated by the *darker red areas*. See text for details. (Redrawn from Thompson M. Genetics in medicine. 4th ed. Philadelphia: WB Saunders, 1986:217.)

Table 4.11. Proportion of Children Affected with Pyloric Stenosis[a]

Proband	Children	
	Sons	Daughters
	%	
Father	5.5	2.4
Mother	19.4	7.3
Population incidence	0.5	0.1

[a] From Carter CO: Genetics of common disorders. Br Med Bull 25:52–57, 1969.

Table 4.10. Family Patterns in Some Common Congenital Malformations[a]

Malformation	Incidence in General Population	Incidence Relative to General Population			
		Monozygotic Twins	First Degree Relatives	Second Degree Relatives	Third Degree Relatives
Cleft lip (± cleft palate)	0.001	×400	×40	×7	×3
Club foot	0.001	×300	×25	×5	×2
Neural tube defects	0.002		×8		×2
Congenital dislocation of hip (females only)	0.002	×200	×25	×3	×2
Congenital pyloric stenosis (males only)	0.005	×80	×10	×5	×1.5

[a] From Carter CO: Genetics of common disorders. Br Med Bull 25:52–57, 1969 and Smith DW, Aase JM: Polygenic inheritance of certain common malformations. J Pediatr 76:653–659, 1970.

Table 4.8. Family Studies of the Incidence of Cleft Lip (± Cleft Palate)[a]

Relatives	*Pecentage of Relatives Affected*	*Incidence Relative to General Population*
First degree		
Sibs	4.1	× 40
Children	3.5	× 35
Second degree		
Aunts and uncles	0.7	× 7
Nephews and nieces	0.8	× 8
Third degree		
First cousins	0.3	× 3

[a] From Carter CO: Genetics of common disorders. Br Med Bull 25:52–57, 1969.

Table 4.9. Concordance among Monozygotic and Dizygotic Twins for Common Malformations and Diseases

Trait	*Concordance*	
	MZ	*DZ*
	%	
Cleft lip ± cleft palate	40	4
Pyloric stenosis	22	2
Schizophrenia	46	14
Insulin-dependent diabetes mellitus	30	6

Table 15.1: Risk of schizophrenia among relatives of schizophrenics: pooled results of several studies

Relative	No. at risk[a]	Risk, %	λ[b]
Parents	8020	5.6	7
Sibs	9920.7	10.1	12.6
Sibs, one parent affected	623.5	16.7	20.8
Offspring	1577.3	12.8	16
Offspring, both parents affected	134	46.3	58
Half-sibs	499.5	4.2	5.2
Uncles, aunts, nephews, nieces	6386.5	2.8	3.5
Grandchildren	739.5	3.7	4.6
Cousins	1600.5	2.4	3

[a]Numbers at risk are corrected to allow for the fact that some at-risk relatives were below or only just within the age of risk for schizophrenia (say, 15–35 years).

[b]λ Values are calculated assuming a population incidence of 0.8%.

Data assembled by McGuffin (1984).

Table 15.2: Twin studies in schizophrenia

Study	Concordant MZ pairs	Concordant DZ pairs
Kringlen, 1968	14/55 (21/55)	4–10%
Fischer, 1969	5/21 (10/21)	10–19%
Tienari, 1975	3/20 (5/16)	3/42
Farmer, 1987	6/16 (10/20)	1/21 (4/31)
Onstad, 1991	8/24	1/28

The numbers show pairwise concordances, i.e. counts of the number of concordant (+/+) and discordant (+/–) pairs ascertained through an affected proband. Figures in brackets are obtained using a wider definition of affected, including borderline, phenotypes. Concordances can also be calculated probandwise, counting a pair twice if both were probands. This gives higher values for the MZ concordance. Probandwise concordances are thought to be more comparable with other measures of family clustering. Only the studies of Onstad and Farmer use the current standard diagnostic criteria, DSM-III. For references, see Onstad *et al.* (1991) and Fischer *et al.* (1969).

Table 15.3: An adoption study in schizophrenia

	Schizophrenia cases among biological relatives	Schizophrenia cases among adoptive relatives
Index cases (47 chronic schizophrenic adoptees)	44/279 (15.8%)	2/111 (1.8%)
Control adoptees (matched for age, sex, social status of adoptive family and number of years institutionalized)	5/234 (2.1%)	2/117 (1.7%)

The study involved 14 427 adopted persons aged 20–40 years in Denmark, 47 of whom were diagnosed as chronic schizophrenic. The 47 were matched with 47 nonschizophrenic control subjects from the same set of adoptees. Data of Kety *et al.* (1994).

Table 15.4: Complex segregation analysis

Model	d	t	q	H	z	x	χ^2	p
Mixed	1.00	7.51	9.6×10^{-6}		0.01	0.15		
Sporadic							334	$<1 \times 10^{-5}$
Polygenic				1.00	1.00		78	$<1 \times 10^{-5}$
Major recessive locus	0.00	8.22	3.8×10^{-3}				35	$<1 \times 10^{-5}$
Major dominant locus	1.00	7.56	1.2×10^{-5}			0.19	2.8	0.42

Data are for families ascertained through a proband with long-segment Hirschsprung disease. Parameters that can be varied are t (the difference in liability between people homozygous for the low-susceptibility and the high-susceptibility alleles of a major susceptibility gene, measured in units of standard deviation of liability), d (the degree of dominance of any major disease allele), q (the gene frequency of any major disease allele), H (the proportion of total variance in liability which is due to polygenic inheritance, in adults), z (the ratio of heritability in children to heritability in adults) and x (the proportion of cases due to new mutation). A single major locus encoding dominant susceptibility explains the data as well as a general model in which a mix of all mechanisms is allowed. Data of Badner *et al.*, 1990.

Table 15.5: A recessive gene for attending medical school?

Model	d	t	q	H	χ^2	p
Mixed	0.087	4.04	0.089	0.008		
Sporadic					163	$<1 \times 10^{-5}$
Polygenic				0.845	14.4	<0.005
Major recessive locus	0.00	7.62	0.88		0.11	N.S.

Data of McGuffin and Huckle (1990) from a survey of medical students and their families. Meaning of symbols as in *Table 15.4*. 'Affected' is defined as attending medical school. The analysis appears to support recessive inheritance, since this accounts for the data equally well as the unrestricted model. The point of this work is to illustrate how analysis of family data can produce spurious results if shared family environment is ignored (see text).

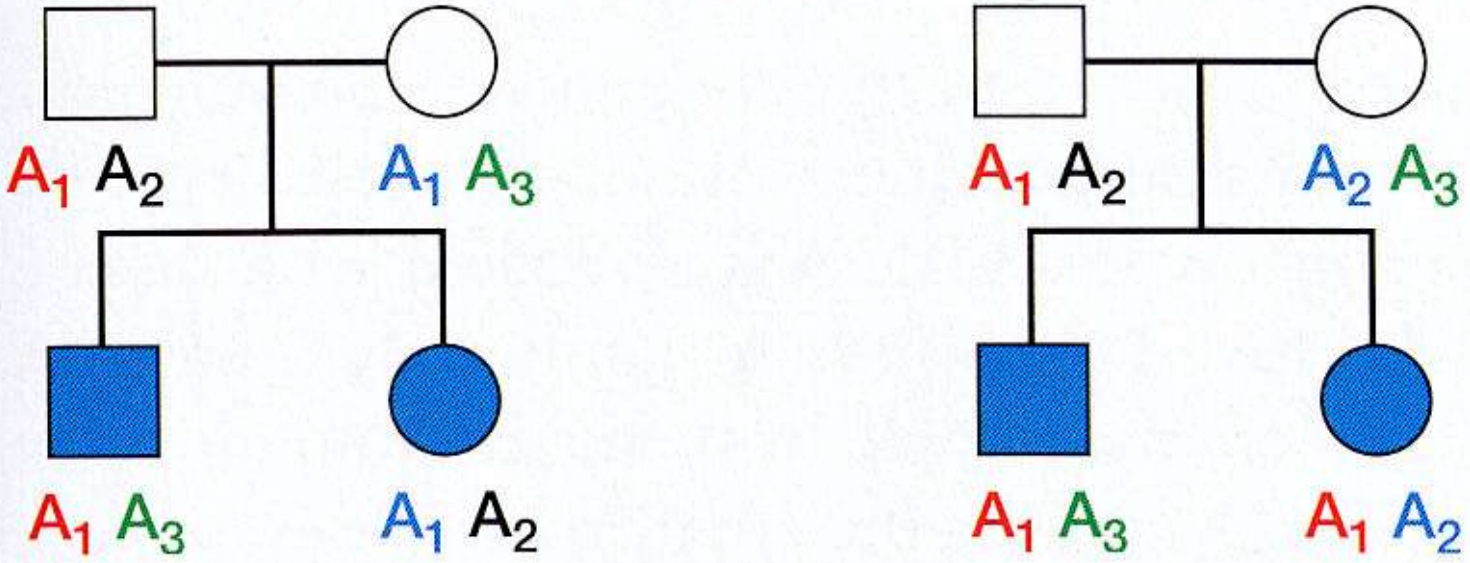


Figure 15.2: Identity by state (IBS) and identity by descent (IBD).

Both sib pairs share allele A_1. The first sib pair have two independent copies of A_1 (IBS but not IBD); the second sib pair share copies of the same paternal A_1 allele (IBD). The difference is only apparent if the parental genotypes are known.

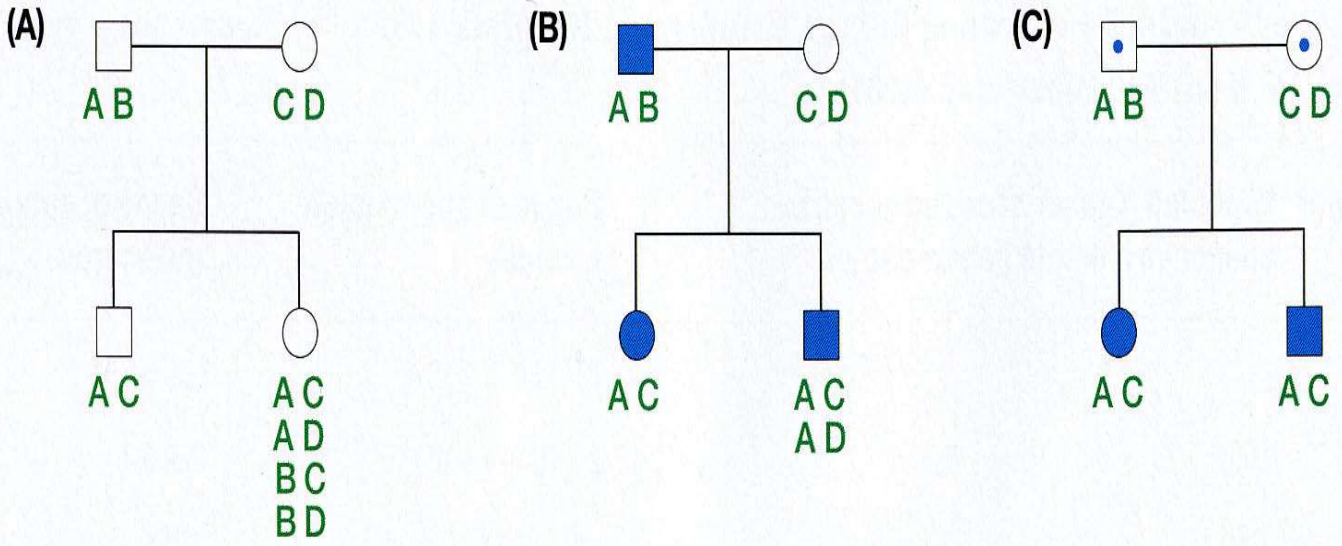


Figure 15.3: Affected sib pair analysis.

(A) By random segregation sib pairs share 0, 1 or 2 parental haplotypes ¼, ½ and ¼ of the time, respectively. **(B)** Pairs of sibs who are both affected by a dominant condition share either one or two copies of the relevant parental chromosomal segment. **(C)** Pairs of sibs who are both affected by a recessive condition necessarily share both parental haplotypes for the relevant chromosomal segment. Above-random haplotype sharing by affected sib pairs identifies chromosomal segments containing susceptibility genes.

Table 15.6: Suggested criteria for reporting linkage (Lander and Kruglyak 1995). The figures for *p* values and lod scores are from Altmüller *et al.* (2001).

Category of linkage	Expected number of occurrences by chance in a whole genome scan	Range of approximate *p* values	Range of approximate lod scores
Suggestive	1	7×10^{-4}–3×10^{-5}	2.2–3.5
Significant	0.05	2×10^{-5}–4×10^{-7}	3.6–5.3
Highly significant	0.001	$\leq 3 \times 10^{-7}$	≥ 5.4
Confirmed	0.01 in a search of a candidate region that gave significant linkage in a previous independent study		

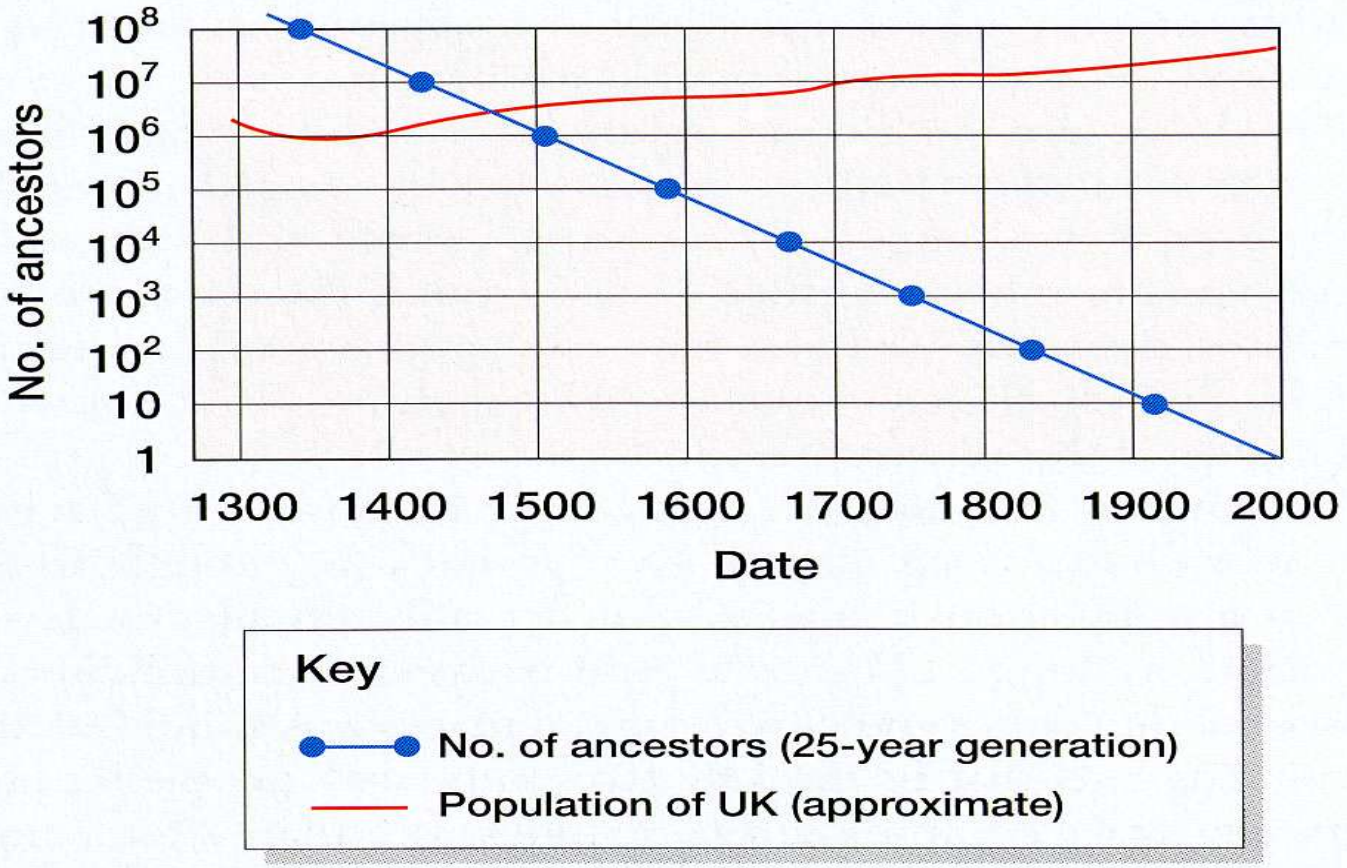


Figure 15.4: Merging into the gene pool.

A fully-outbred person has 2^n ancestors n generations ago. If the UK population were fully outbred, two 'unrelated' present-day people would share all the same ancestors in 1500. In reality, of course, the population is not fully outbred, and the two people would have strongly overlapping but not identical pools of ancestors in 1500.

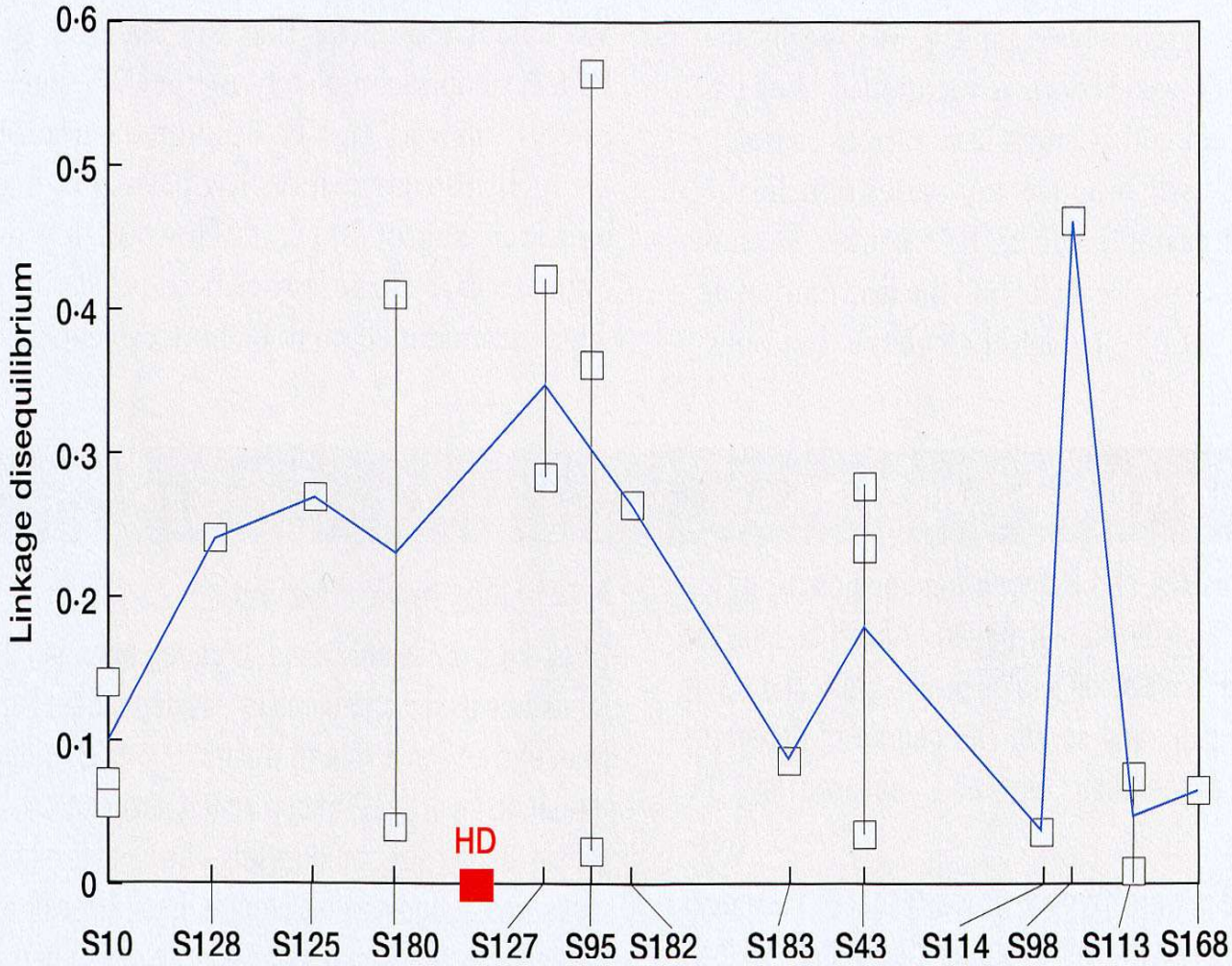


Figure 15.5: Linkage disequilibrium around the Huntington disease locus.

S10, S125 etc. are shorthand for the DNA markers *D4S10, D4S125* etc., shown in their map positions relative to the HD locus. The total distance represented is 2500kb. For some loci, several different RFLPs exist, which sometimes show very different allelic association, for example marker S95 (see text). From Krawczak and Schmidtke (1998) *DNA Fingerprinting,* 2nd edn. BIOS Scientific Publishers, Oxford.

Box 15.3: The transmission disequilibrium test (TDT) to determine whether marker allele M_1 is associated with a disease.

1) Affected probands are ascertained
2) The probands and their parents are typed for the marker
3) Those parents who are heterozygous for marker allele M_1 are selected. They may or may not be affected

Let **a** be the number of times a heterozygous parent transmits M_1 to the affected offspring, and **b** be the number of times the other allele is transmitted.

The TDT test statistic is $(\mathbf{a}-\mathbf{b})^2 / (\mathbf{a}+\mathbf{b})$. This has a χ^2 distribution with 1 degree of freedom, provided the numbers are reasonably large.

Box 15.4: Sample sizes needed to find a disease susceptibility locus by a whole genome scan using either affected sib pairs (ASP) or the transmission disequilibrium test (TDT).

Risch and Merikangas (1996) calculated the sample sizes needed to distinguish a genetic effect from the null hypothesis with power $(1-\beta)$ and significance level α. This Box summarizes their formulae and equations, but the original paper should be consulted for the derivations and for details.

A standard piece of statistics tells us that the sample size M required is given by $(Z_\alpha - \sigma Z_{1-\beta})^2 / \mu^2$, where Z refers to the standard normal deviate. The mean μ and variance σ^2 are calculated as functions of the susceptibility allele frequency (p) and the relative risk γ conferred by one copy of the susceptibility allele. The model assumes that the relative risk for a person carrying two susceptibility alleles is γ^2; that the marker used is always informative; and that there is no recombination with the susceptibility locus.

For ASP, the expected allele sharing at the susceptibility locus is given by $Y = (1+w)/(2+w)$, where $w = [pq(\gamma-1)^2]/(p\gamma + q)$. $\mu = 2Y-1$ and $\sigma^2 = 4Y(1-Y)$. The genome-wide threshold of significance (probability of a false positive anywhere in the genome = 0.05; testing for sharing IBD) requires a lod score of 3.6, corresponding to $\alpha = 3 \times 10^{-5}$, and $Z_\alpha = 4.014$. For 80% power to detect an effect, $1-\beta = 0.2$ and $Z_{1-\beta} = -0.84$.

For the TDT, the probability that a parent will be heterozygous for the allele in question is $h = pq(\gamma+1)/(p\gamma+q)$. P(trA), the probability that such a heterozygous parent will transmit the high-risk allele to the affected child, is $= \gamma/(1+\gamma)$. $\mu = \sqrt{h}(\gamma-1)/(\gamma+1)$, and $\sigma^2 = 1-[h(\gamma-1)^2/(\gamma+1)^2]$. As discussed above, for an ultimate genome screen involving 1 000 000 tests, $\alpha = 5 \times 10^{-8}$, $Z_\alpha = 5.33$ and, as before, $Z_{1-\beta} = -0.84$.

In Table 15.7 the Z_α, $Z_{1-\beta}$, μ and σ^2 values are used to calculate sample sizes by substituting in the formula $M = (Z_\alpha - \sigma Z_{1-\beta})^2 / \mu^2$. For the TDT, the answer is halved because each parent–child trio allows two tests, one on each parent.

Table 15.7: Sample sizes for 80% power to detect significant linkage or association in a genome-wide search

		ASP analysis		TDT analysis	
γ	p	Y	N-ASP	P(trA)	N-TDT
5	0.01	0.534	2530	0.830	747
	0.1	0.634	161	0.830	108
	0.5	0.591	355	0.830	83
3	0.01	0.509	33797	0.750	1960
	0.1	0.556	953	0.750	251
	0.5	0.556	953	0.750	150
2	0.1	0.518	9167	0.667	696
	0.5	0.526	4254	0.667	340
1.5	0.1	0.505	115537	0.600	2219
	0.5	0.510	30660	0.600	950
1.2	0.1	0.501	3951997	0.545	11868
	0.5	0.502	696099	0.545	4606

γ is the relative risk for individuals of genotype Aa compared to aa; p is the frequency of the A susceptibility allele. For affected sib pair (ASP) analysis, Y is the expected allele sharing and N-ASP the number of pairs required for significance, based on IBD testing ($\alpha = 3 \times 10^{-5}$). For transmission disequilibrium testing (TDT), P(trA) is the probability that an Aa parent will transmit A to an affected child, and N-TDT is the number of parent-child trios required for significance. After Risch and Merikangas (1996).

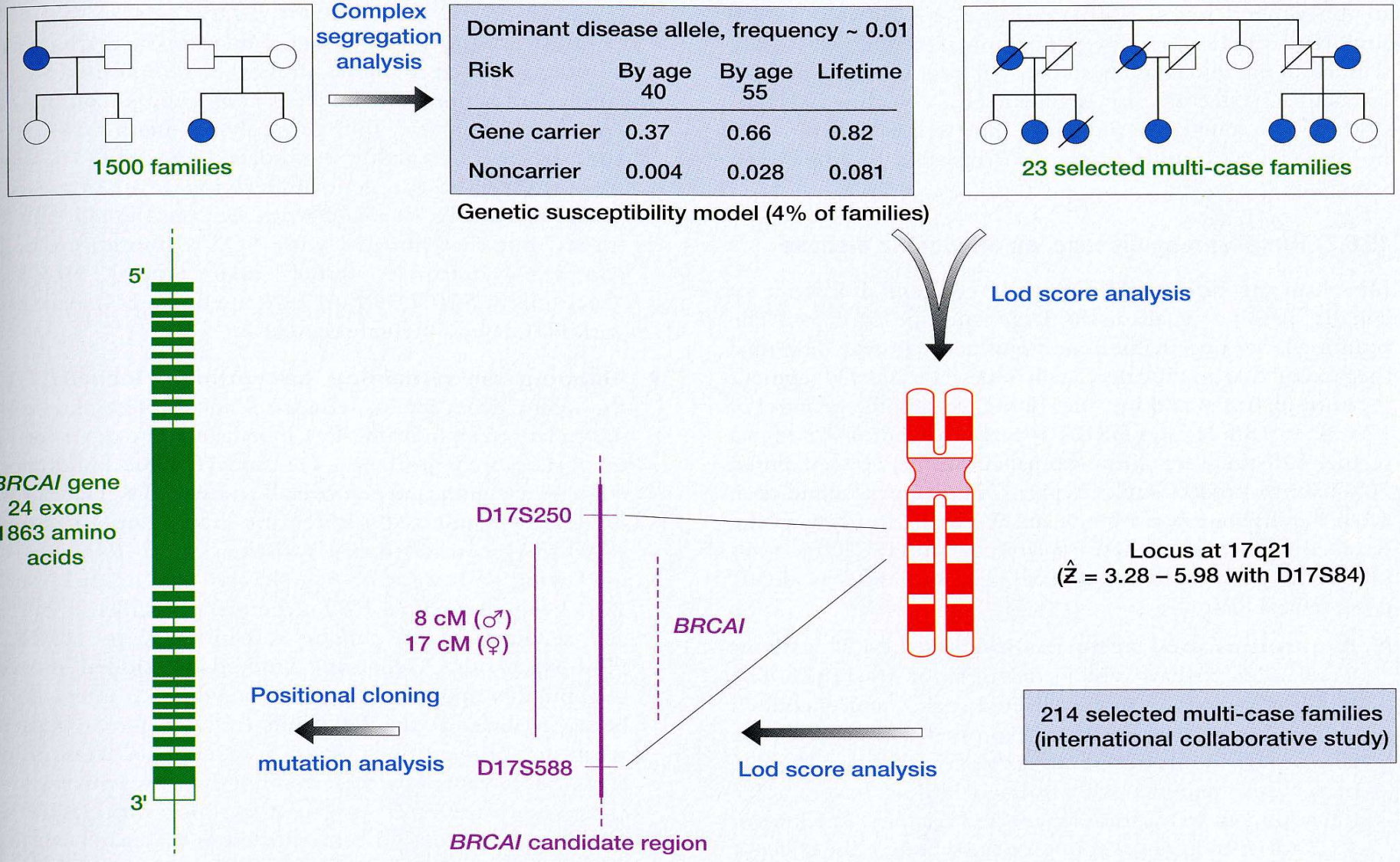


Figure 15.7: How the *BRCA1* gene was found.

Standard positional cloning successfully identified a gene conferring susceptibility to a common disease – but only to a Mendelian subset of the disease. *BRCA1* appears to have little role in the common, sporadic breast cancer.

Table 15.8: Clinical classification of diabetes

Type 1 diabetes	Type 2 diabetes	MODY
Juvenile onset	Maturity onset (> 40 years)	Juvenile onset
0.4% of UK population	6% of US population	Rare
Requires insulin	Usually controllable by oral hypoglycemics	As type 2 diabetes
No obesity	Strong association with obesity	No obesity
Familial: MZ concordance 30% sib risk 6–10%	Familial: MZ twin concordance 40–100% sib risk 30% (maybe subclinical)	Familial: autosomal dominant?
Associated with HLA-DR3 and DR4	No HLA association	No HLA association

MODY (maturity onset diabetes of the young) is an uncommon Mendelian form for which various genes have been identified (see MIM 600496). Additionally, diabetes can be part of a number of uncommon syndromes.

Table 15.9: Main Type 1 diabetes susceptibility loci suggested by affected sib pair (ASP) or transmission disequilibrium (TDT) analysis

Locus	MIM No.	Location	Status
IDDM1	222100	6p21	$\lambda_s = 3.1$; determinant is *HLA-DQB*
IDDM2	125852	11p15	$\lambda_s = 1.3$; determinant is a VNTR upstream of *INS* gene
IDDM4	600319	11q13	$\lambda_s = 1.6$; significant linkage in combined results of three screens (596 families)
IDDM5	600320	6q24–q27	$\lambda_s = 1.2$; observed in four studies
IDDM6	601941	18q21	ASP and TDT evidence in one study.
IDDM7	600321	2q31-q33	$\lambda_s = 1.3$; seen in three ASP studies. Linkage disequilibrium with candidate gene *CTLA4,*
IDDM12	600388	2q33	$p = 5 \times 10^{-5}$ but only in some populations.
IDDM8	600883	6q25–q27	$\lambda_s = 1.8$; not clearly distinct from *IDDM5.*
IDDM10	601942	10p11-q11	ASP and TDT data from three studies
IDDM13	601318	2q34	Same as IDDM7 and/or 12?
IDDM15	601666	6q21	Confirmed (though hard to separate from HLA effect)

Data from the OMIM entries and papers cited therein; λ_s values are from Luo *et al.* (1995).