

# **Human Genetic Mapping**

**K.H. Buetow**

**Fox Chase Cancer Center**

**voice: (215) 728-3152**

**FAX: (215) 728-3574**

**kh\_buetow@fccc.edu**

**introduction**

**generating a genetic map**

***principals - track meiotic transmission***

***variation detection strategies***

***analytic methods***

**status of the human genetic map**

**mapping a genetic trait**

***simple Mendelian trait***

***complex Mendelian trait***

***complex familial trait***

# **Genetics**

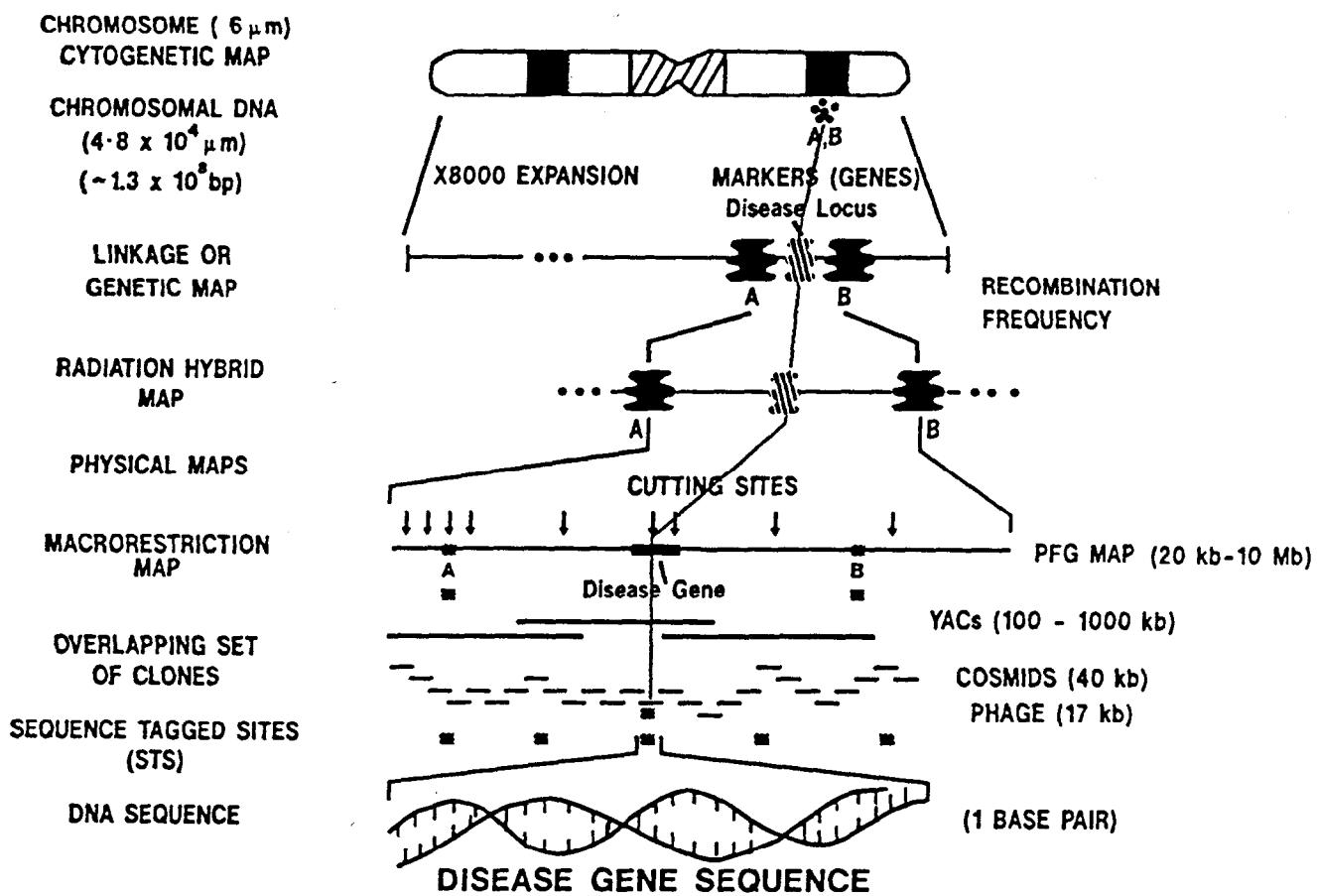
The biology of heredity,  
especially the study of  
***hereditary transmission***  
and *variation*

(American Heritage Dictionary)

---

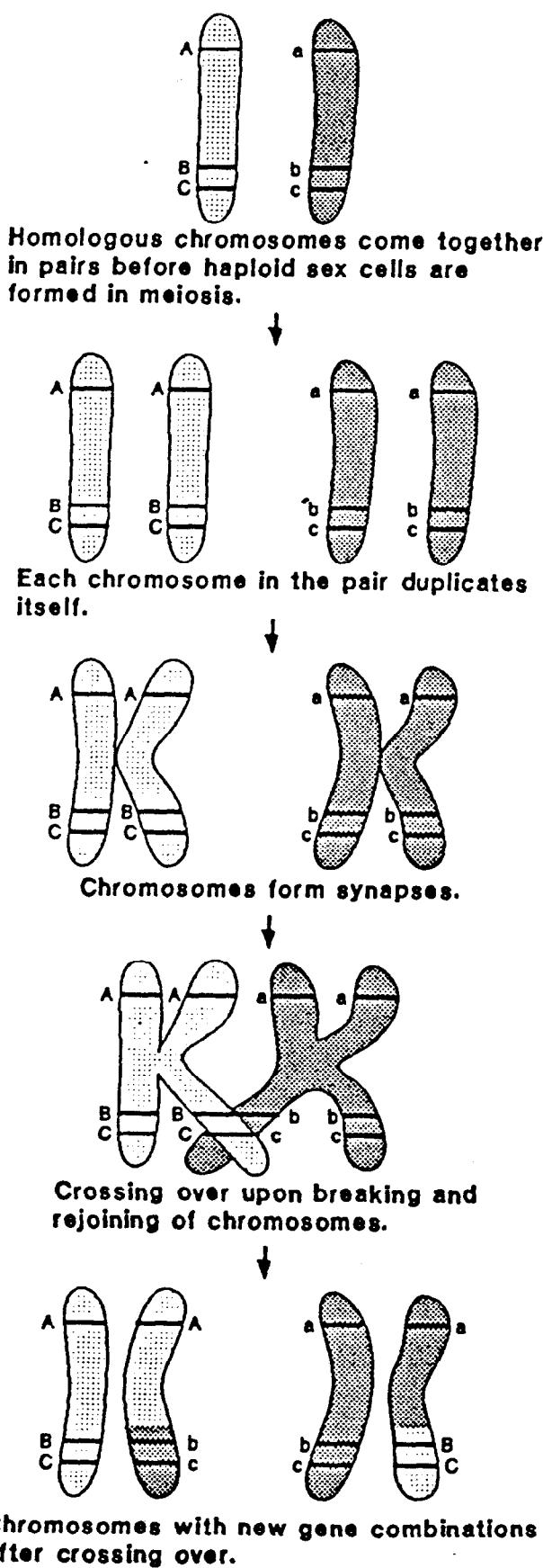
Genetic Epidemiology Group

Fox Chase Cancer Center



**Multiple Levels of Human Chromosome Mapping.** The line running vertically through the diagram represents the tracking of markers A and B through progressively more precise levels of mapping. In this way, investigators can follow a candidate disease gene from the coarsest to the finest map resolution, which is the DNA sequence. The cytogenetic map provides the lowest level of resolution, measuring the distance between chromosomal features (i.e., bands or breakpoints) visible under the light microscope. Chromosome banding can resolve features to about 5 Mb. The linkage or genetic map measures the recombination frequency between two linked markers, which can be genes or polymorphisms (A and B in this diagram.) Radiation hybrid maps are produced by breaking chromosomes with radiation and then identifying the fragment carrying the marker (the breakpoint); the resolution of these maps is comparable to that of linkage maps. At the next resolution level, macrorestriction fragments of 1 to 2 Mb are separated and the markers localized and mapped. Finer mapping resolution is provided by ordered libraries of yeast artificial chromosomes (YACs), which have insert sizes from 100 to 1000 kb. Ordered libraries of cosmids have smaller insert sizes, usually about 40 kb, and produce higher-resolution maps. The DNA base sequence is the highest-resolution map, with sequence tagged sites (STSs) used as unique reference points. (Figure provided by C. E. Hildebrand, LANL.)

**Figure 2-5.—Separation of Linked Genes by Crossing Over of Chromosomes During Meiosis**



SOURCE: Office of Technology Assessment, 1988.

## **alternative means of detecting genetic variation**

- restriction fragment length polymorphism (RFLP)**
- single strand confirmation polymorphism (SSCP)**
- denaturing gradient gel electrophoresis (DGGE)**
- PCR direct sequencing**

## simple tandem repeat polymorphism (STRP)

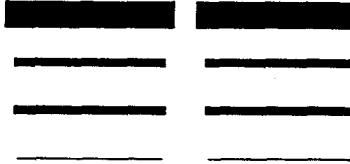
**CA<sub>10</sub>**      CACACACACACACACACA  
                  GTGTGTGTGTGTGTGTGTGT

**CA<sub>5</sub>**      CACACACACA  
                  GTGTGTGTGT

110bps

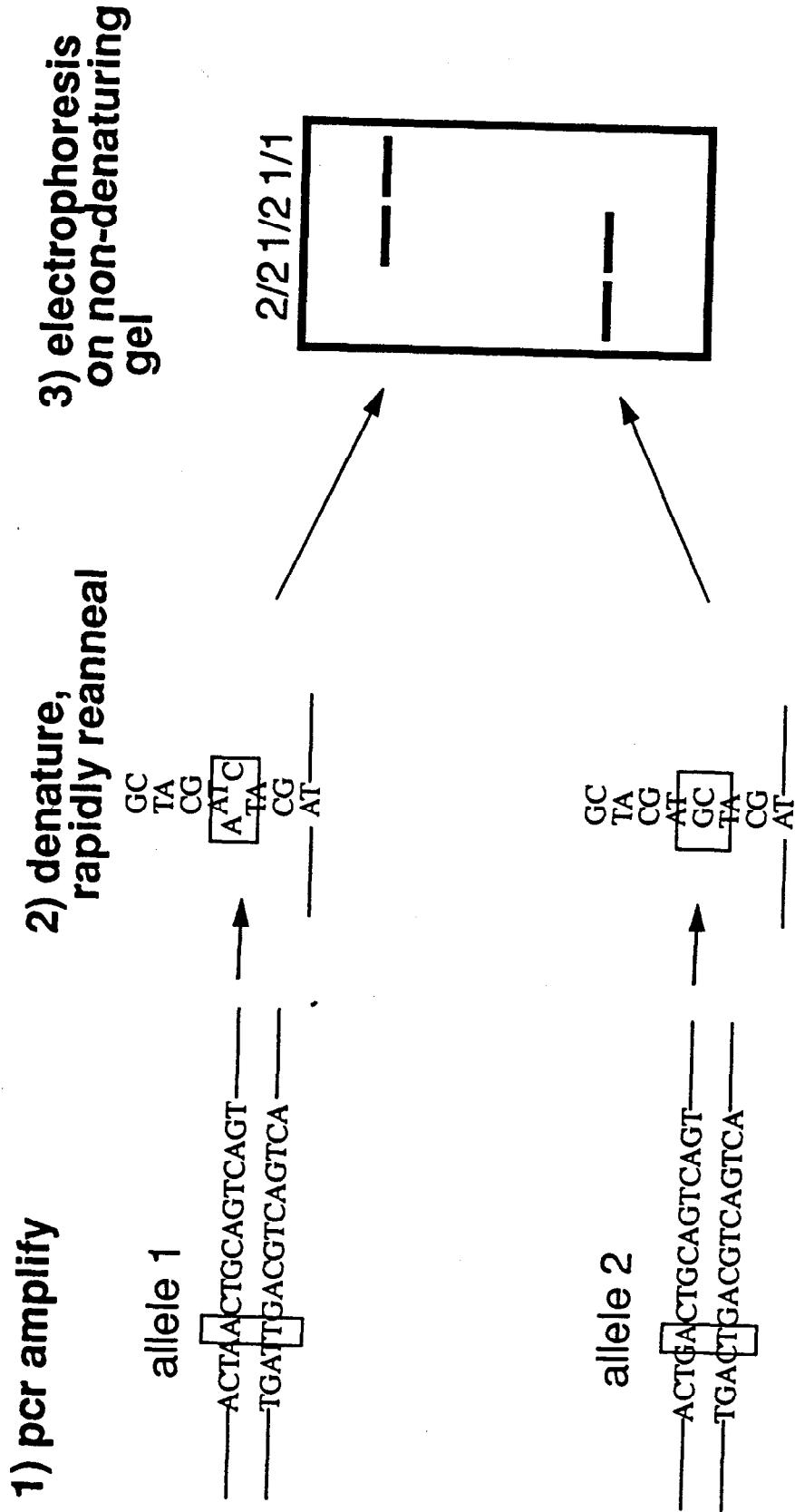


100bps



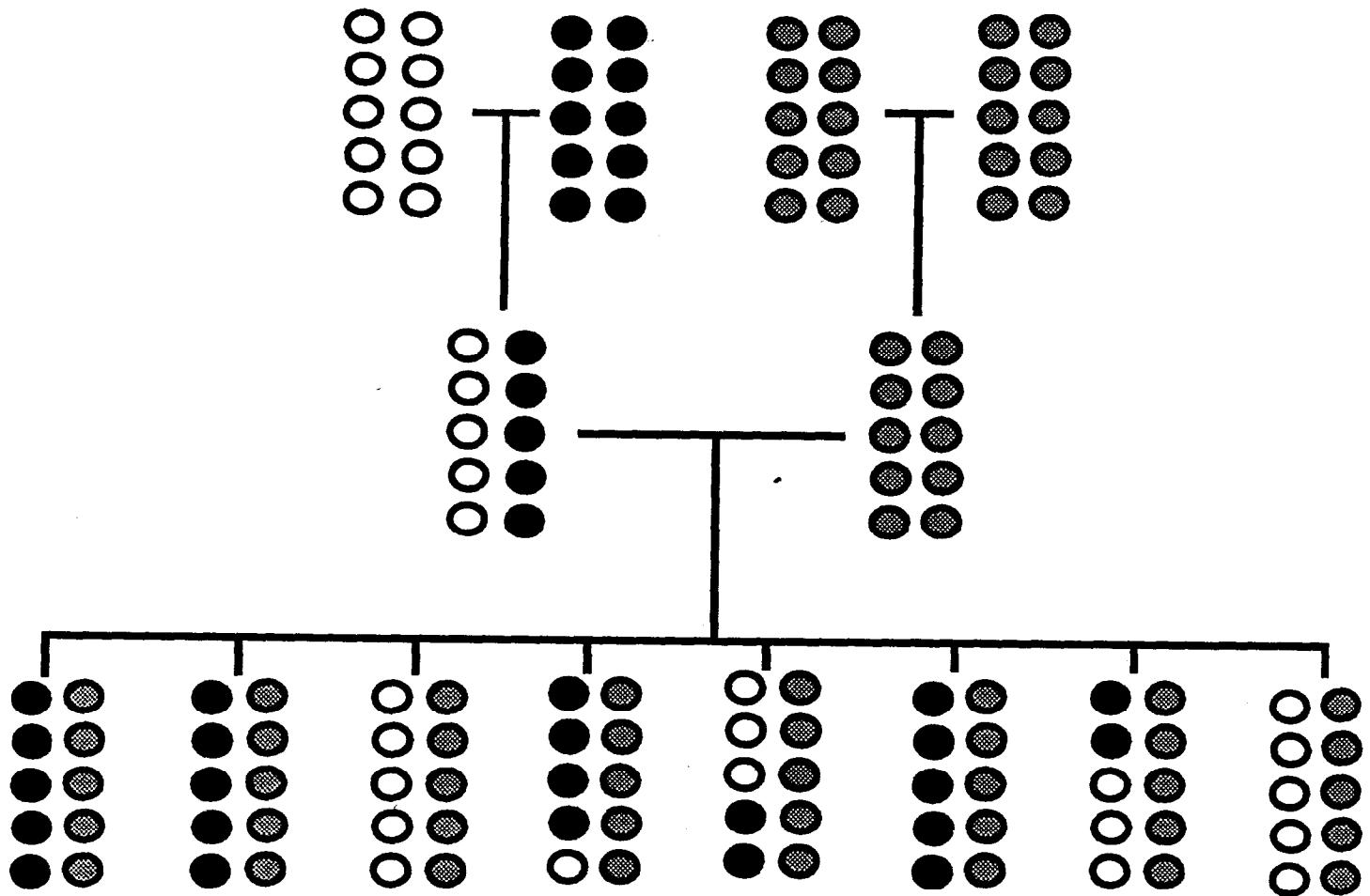
5/5    10/5    10/10

# Single strand conformation polymorphism (SSCP)

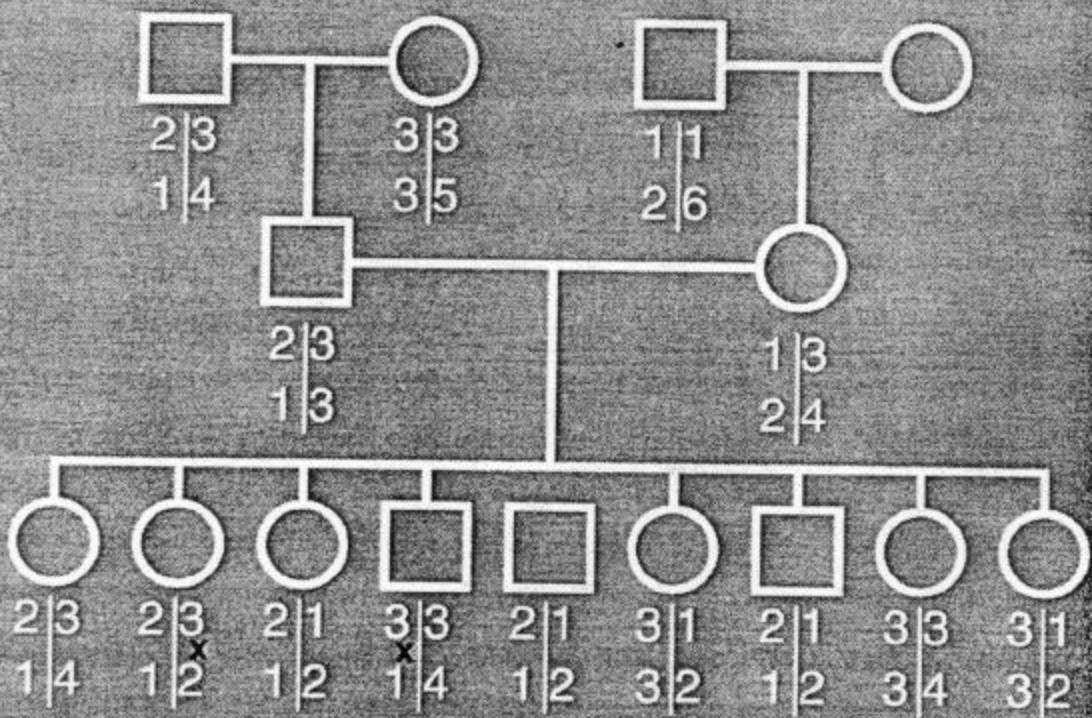


Genetic Epidemiology Group

Fox Chase Cancer Center



## CEPH PEDIGREE 1477

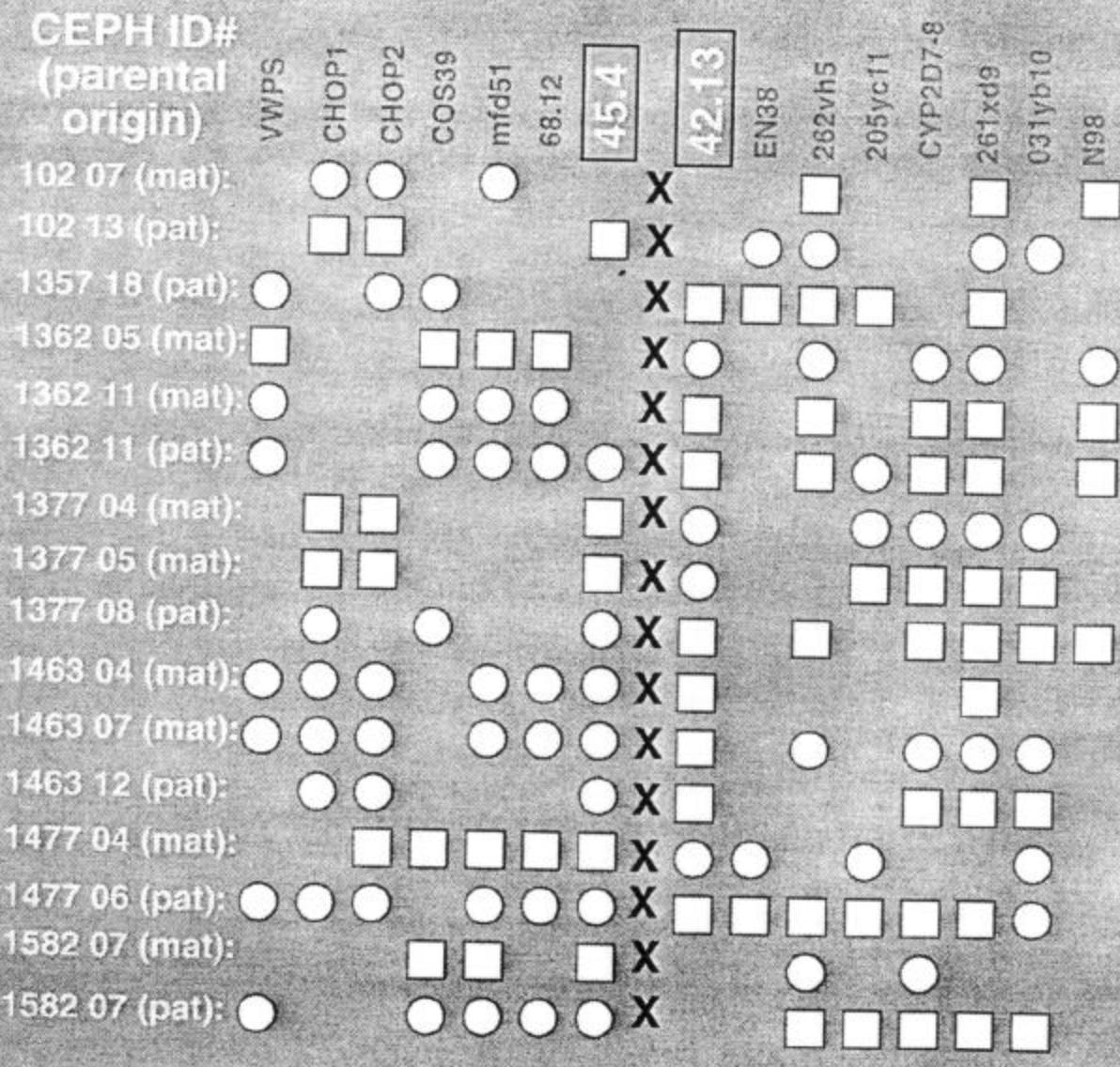


Cooperative Human Linkage Center

Fox Chase Cancer Center

# Recombinant Interval Database

recombinant haplotypes from index  
map for marker interval:

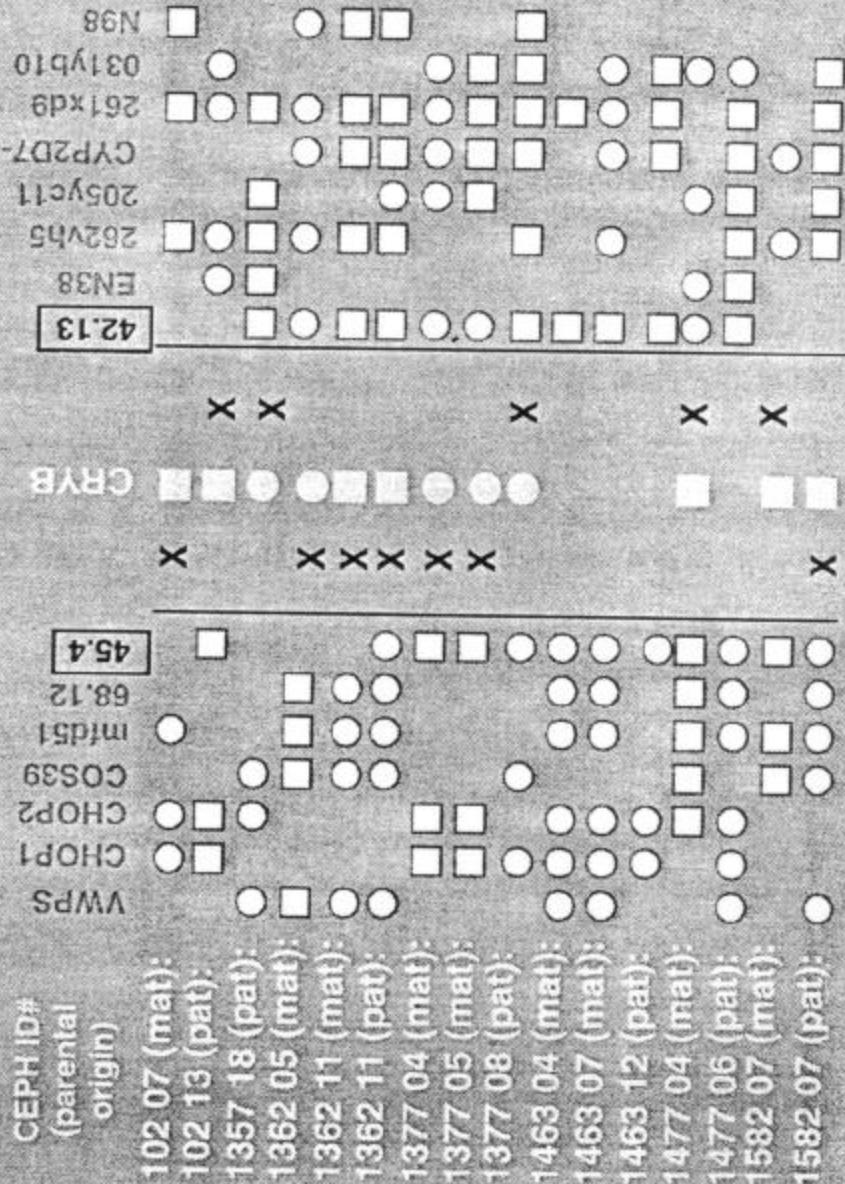


**ORIGIN:** □ grandpaternal  
● grandmaternal

Cooperative Human Linkage Center

Fox Chase Cancer Center

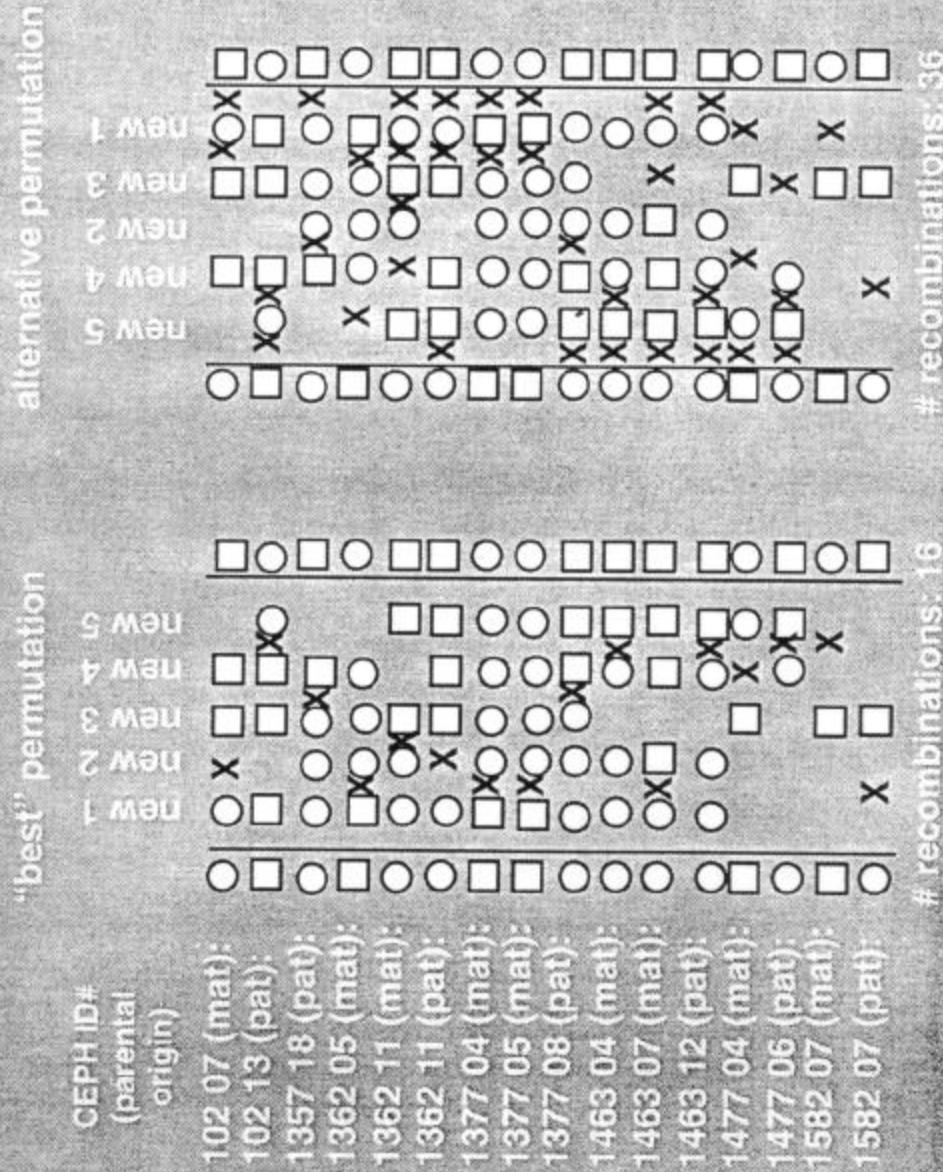
# Updated Reference Map Using Recombinant Interval Database



Cooperative Human Linkage Center

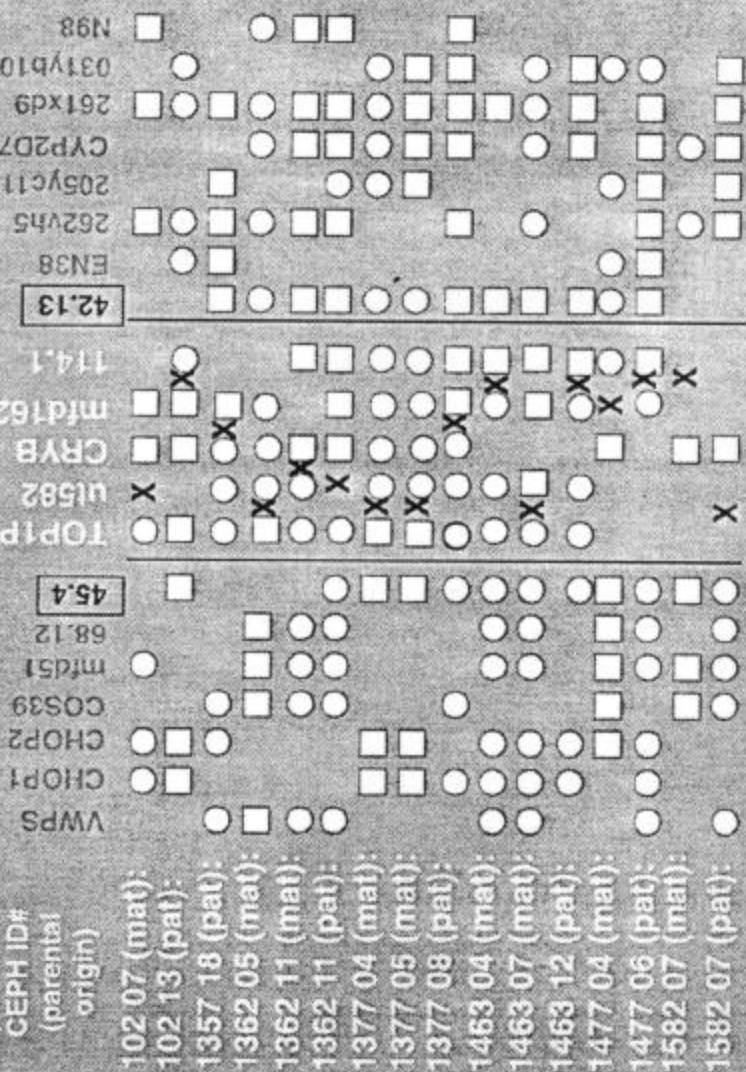
Fox Chase Cancer Center

# bin ordering by recombination minimization



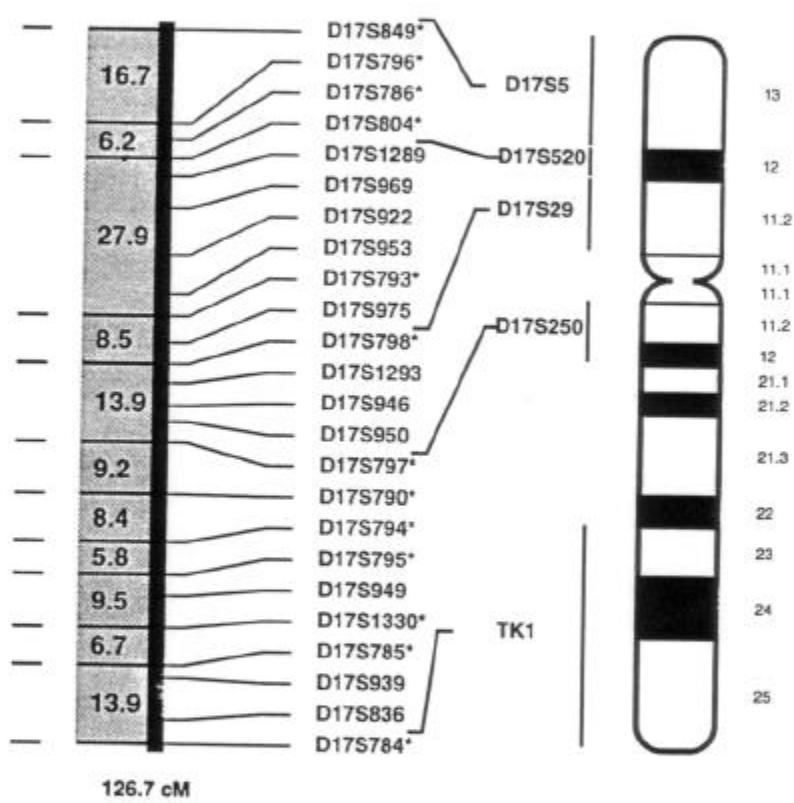
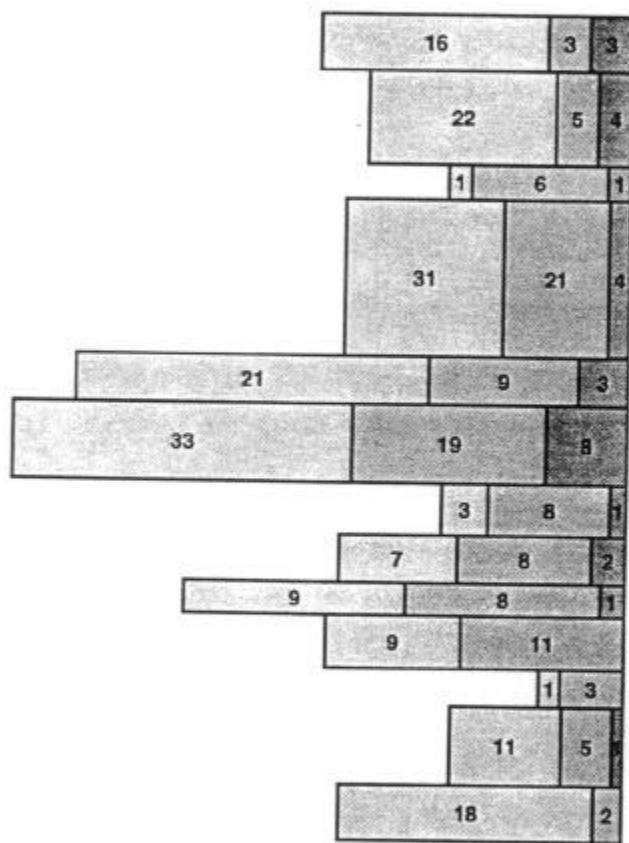
# recombinations: 16      # recombinations: 36

# Updated Reference Map Using Recombinant Interval Database

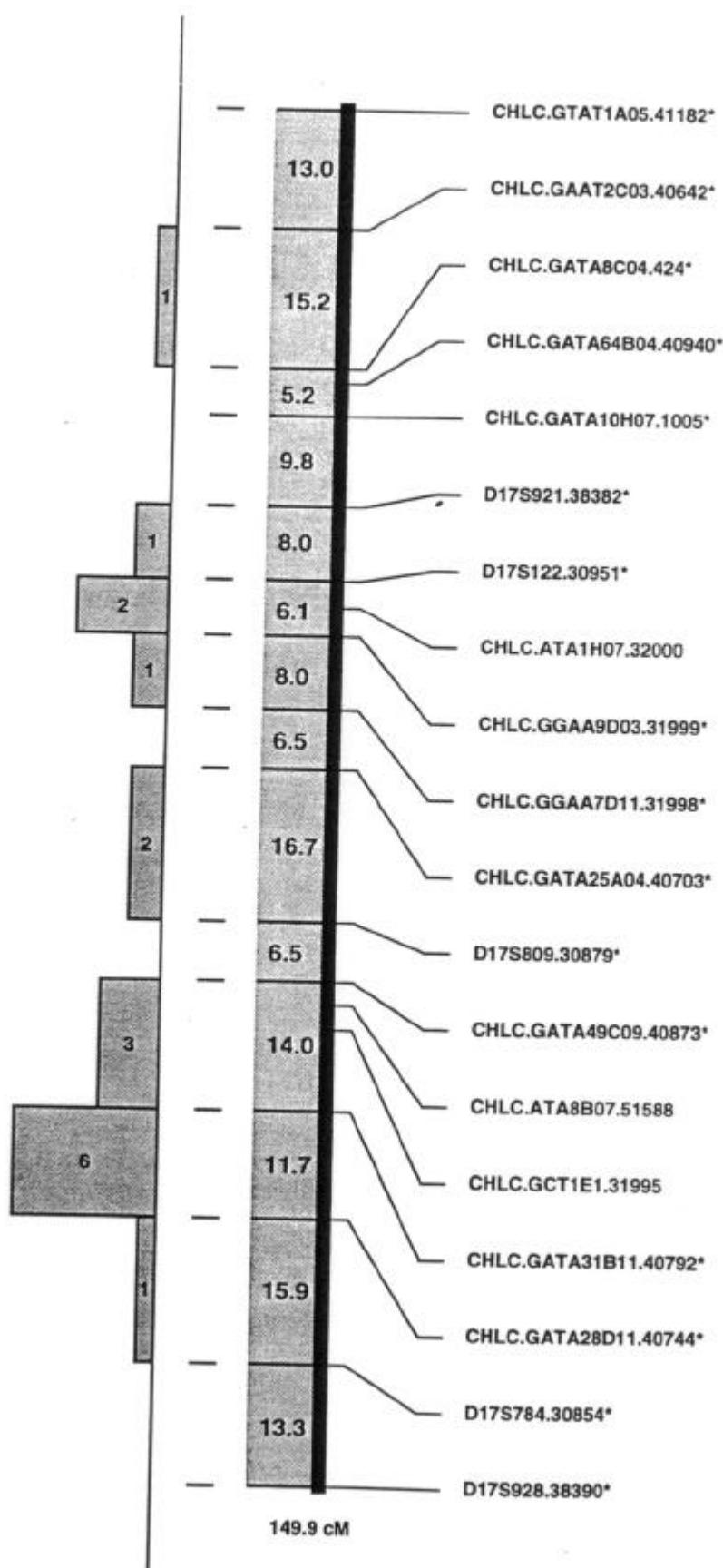


Cooperative Human Linkage Center

Fox Chase Cancer Center



# CHLC Marker Distribution in V6 Screening Map of Chromosome 17



$$z(x) = \log_{10} \left[ \frac{L(\text{pedigree given } \theta = x)}{L(\text{pedigree given } \theta = 0.5)} \right]$$

$$z(x) = \log_{10} \left[ \frac{\theta^R (1 - \theta)^{NR}}{(0.5)^R (0.5)^{NR}} \right]$$

$$z(0.05) = \log_{10} \left[ \frac{(0.05)^R (1 - 0.05)^{NR}}{(0.5)^N} \right]$$

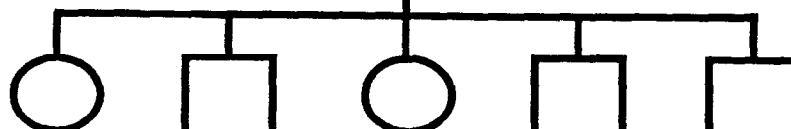
$$z(0.05) = \log_{10} \left[ \frac{(0.05)^1 (0.95)^7}{(0.5)^8} \right] = 0.9513$$

$$z(0.05) = 0.95.$$



genotype locus A: 1/1  
genotype locus B: 2/2

3/4  
3/4



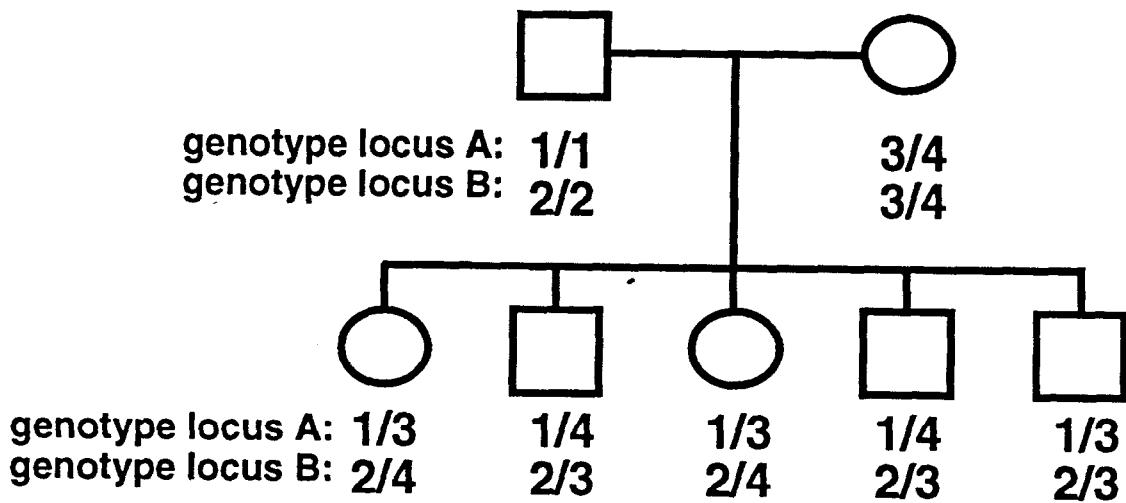
genotype locus A: 1/3  
genotype locus B: 2/4

1/4  
2/3

1/3  
2/4

1/4  
2/3

1/3  
2/3



probability if linkage phase is:

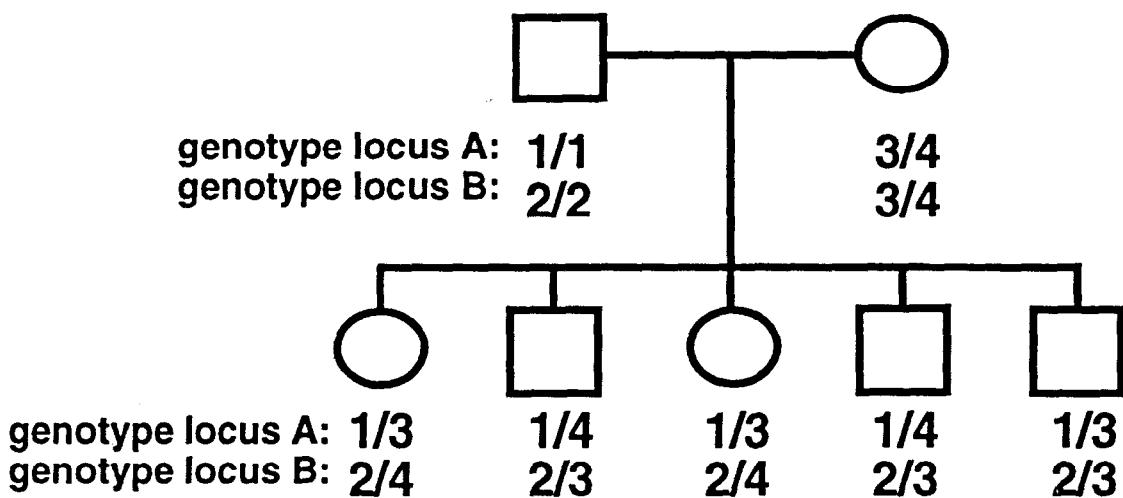
"coupling"

$$\begin{array}{c|c} & 3|4 \\ " & 3|4 \end{array}$$

"repulsion"

$$\begin{array}{c|c} 3|4 \\ 4|3 \end{array}$$

	$\frac{3}{3}   \frac{1}{2}(1-\theta)$	$\frac{1}{2}\theta$
mother provides gamete:	$\frac{4}{4}   \frac{1}{2}(1-\theta)$	$\frac{1}{2}\theta$
	$\frac{3}{4}   \frac{1}{2}\theta$	$\frac{1}{2}(1-\theta)$
	$\frac{4}{3}   \frac{1}{2}\theta$	$\frac{1}{2}(1-\theta)$



"coupling"

3	4
3	4

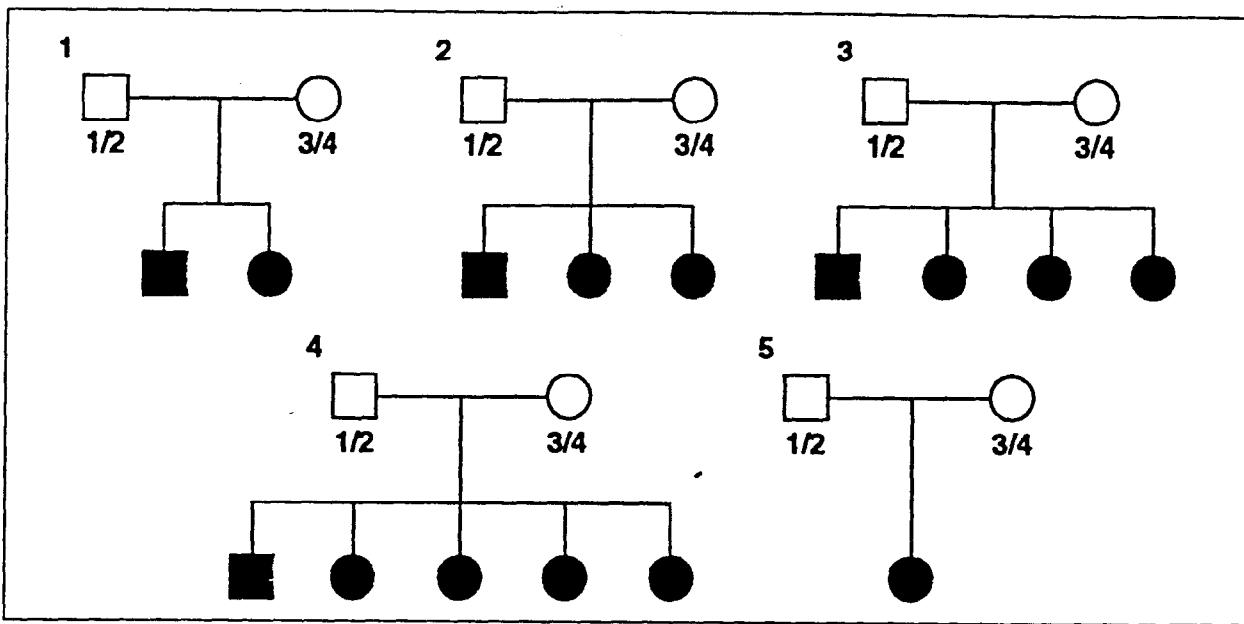
"repulsion"

3	4
4	3

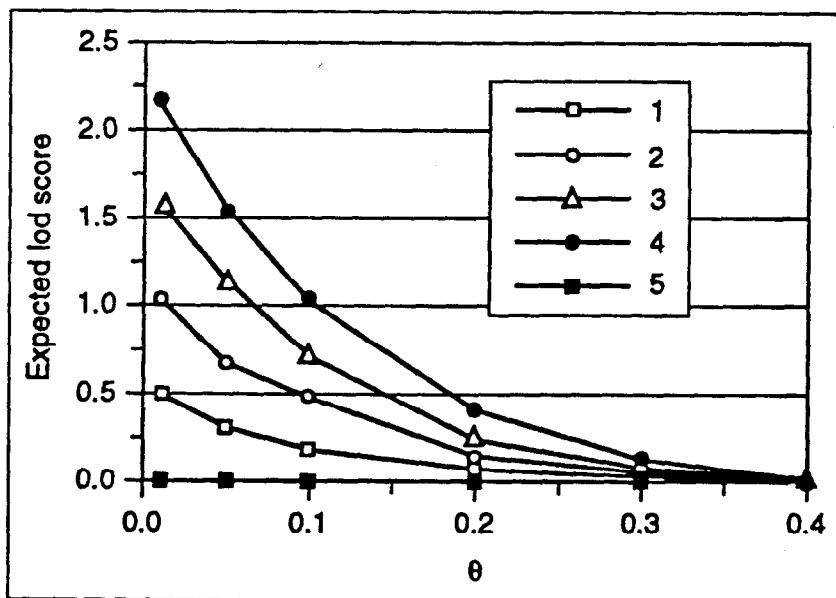
$$\begin{aligned}
 L(\theta) &= [(1/2\theta)4(1/2(1-\theta)) + (1/2(1-\theta))4(1/2\theta)] / \\
 &\quad 2 \\
 &= 1/64 [\theta 4(1-\theta) + \theta(1-\theta)4]
 \end{aligned}$$

$$\begin{aligned}
 L(\theta = 0.2) &= 1/64 [0.00128 + 0.08192] \\
 &= 0.0013
 \end{aligned}$$

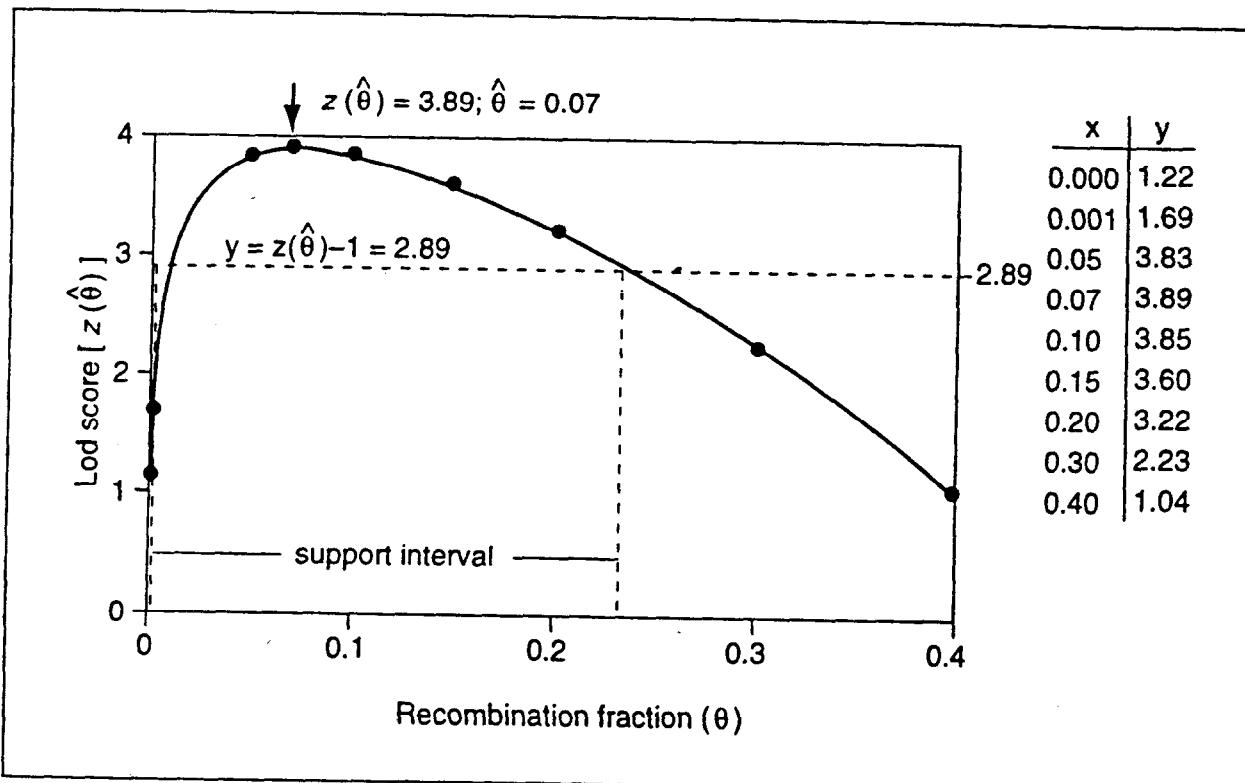
$$\begin{aligned}
 Z &= \log_{10} [L(\theta = 0.2) - L(\theta = 0.50)] \\
 &= 0.124
 \end{aligned}$$



**Figure 1.2.6** Nuclear families segregating for an autosomal recessive disease. The father has a 1/2 marker genotype, while the mother has a 3/4 marker genotype. Figure 1.2.7 shows the expected lod scores for each family.



**Figure 1.2.7** The ex- pected lod score as a function of the recombi- nation fraction ( $\theta$ ) and pedi- gree structure (shown in Fig. 1.2.6).



**Figure 1.7.10** Graph of two-point lod score (y axis) versus recombination fraction (x axis) using values for marker 3 from Table 1.7.3. The horizontal line drawn across the graph is drawn at  $y = z(\hat{\theta}) - 1.0$  and is used in constructing the one lod unit support interval for the maximum likelihood estimate of the recombination fraction ( $\hat{\theta}$ ).

**Table 2****VWS Linkage with Iq Markers**

VWS vs.	$\theta_m = \theta_f$							$Z_{\max}$	$\hat{\theta}$	$Z(m,f)$	$\hat{\theta}_m$	$\hat{\theta}_f$
	0	.001	.05	.10	.20	.30	.40					
REN.....	7.15	8.22	8.85	8.05	6.03	3.79	1.61	9.09	.02	9.28	0	.03
D1S53.....	-.94	1.12	3.83	3.79	3.11	2.18	1.14	3.87	.07	4.73	0	.18
CR1.....	-3.95	-1.47	3.14	3.43	2.87	1.83	.74	3.43	.10	3.43	.09	.10
D1S58.....	-.08	1.13	2.83	2.82	2.21	1.34	.49	2.88	.07	3.16	.12	0
D1S52.....	0.57	1.31	2.71	2.69	2.19	1.47	.66	2.74	.07	2.99	0	.11
CR2.....	-.22	.76	2.40	2.44	1.93	1.21	.53	2.46	.08	2.74	0	.12
DAF.....	-2.14	-1.86	1.35	1.92	1.81	1.18	.50	1.99	.13	2.02	.09	.15
D1S65.....	-.90	-.43	1.82	1.95	1.58	.10	.42	1.95	.09	1.97	.13	.06
LAMB2.....	-6.27	-4.10	-.36	.54	1.13	1.01	.56	1.14	.22	1.17	.26	.19

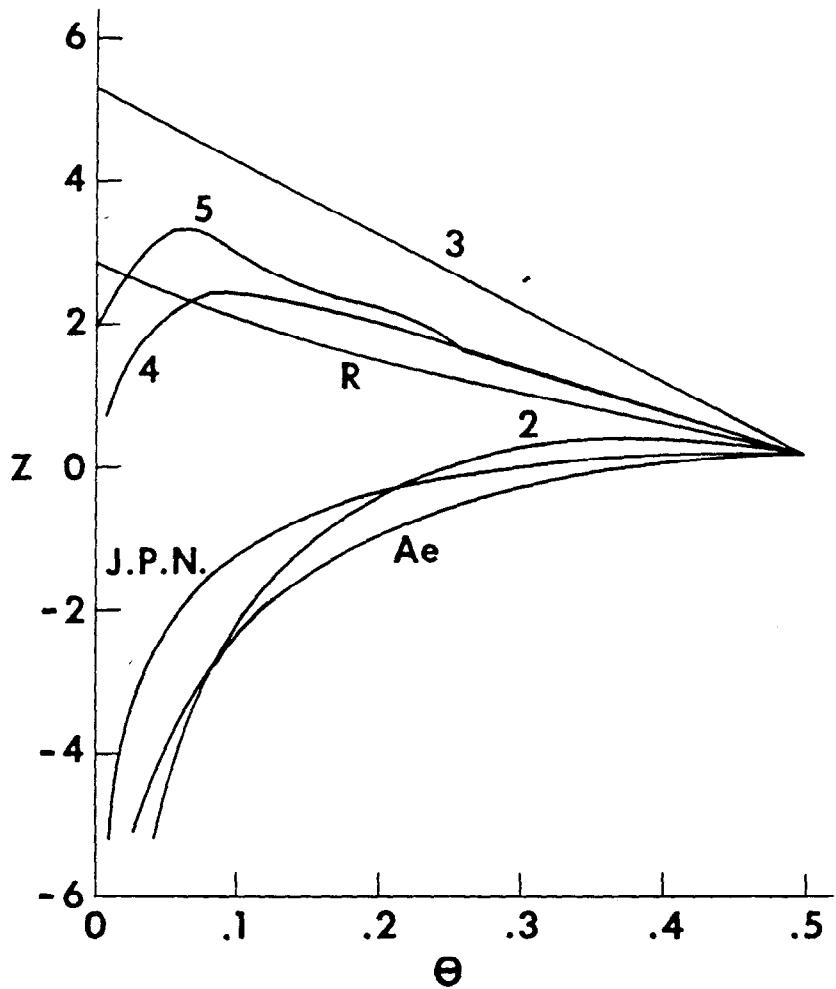


Fig. 6.2.1: Linkage of elliptocytosis with *RH* in 4 of 7 families. At a recombination value of  $\theta = .05$ , the lod score  $Z$  is greater than 2 in families 3, 4, 5, and R (linkage) and less than 2 in families 2, Ae, and JPN (no linkage).

**REN-VWS Linkage by Family**

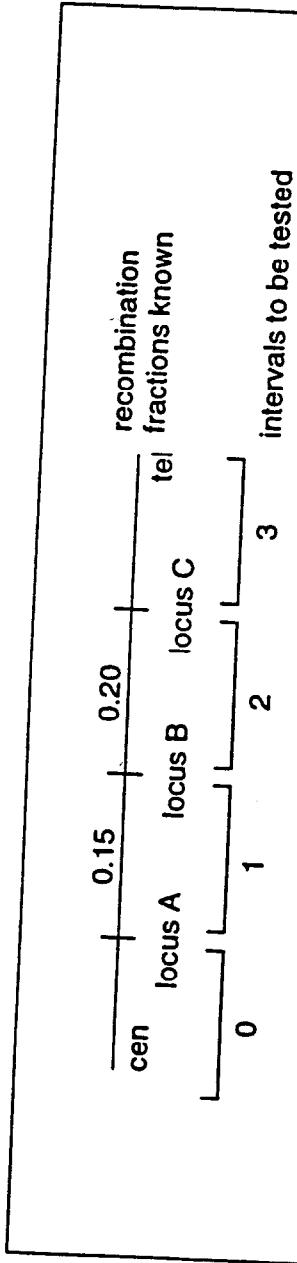
FAMILY	$\theta_m = \theta_j$					$Z_{max}$	$\hat{\theta}$	$Z(m, f)$	$\hat{\theta}_m$	$\hat{\theta}_j$
	0	.001	.05	.10	.20					
VWS 1 .....	2.77	2.77	2.47	2.17	1.54	.91	.34	2.77	0	2.77
VWS 2 .....	1.68	1.68	1.55	1.41	1.06	.68	.31	1.68	0	1.69
VWS 3 .....	1.39	1.39	1.25	1.10	.79	.48	.21	1.39	0	.10
VWS 4 .....	-2.04	-.96	.56	.70	.67	.56	.27	.72	.14	0
VWS 5 .....	.81	.81	.73	.65	.48	.31	.15	.81	0	.35
VWS 6 .....	<u>2.54</u>	<u>2.54</u>	<u>2.29</u>	<u>2.02</u>	<u>1.48</u>	<u>.91</u>	<u>.34</u>	<u>2.54</u>	<u>0</u>	<u>0</u>
Overall .....	7.15	8.23	8.85	8.05	6.02	3.85	1.62	9.91	.02	10.40
										.03

$$\chi^2 = 2 (\ln 10) \left[ \sum_1^i z_i(\theta_i) - z\theta \right]$$

**Table 1.1.2 Correspondence between Mode of Inheritance, Clinical Variation, and Genetic Heterogeneity for Selected Mendelian Disorders**

Disorder	Mode(s) of inheritance <sup>a</sup>	Clinical variation		Heterogeneity	
		Familial	Nonfamilial	Allelic	Locus
Alzheimer disease	AD	Y	Y	?	Y
Charcot-Marie-Tooth disease	AD, AR, XD, XR	Y	Y	?	Y
Cystic fibrosis	AR	Y	Y	Y	N
Duchenne muscular dystrophy	AR, XR	N	N	Y	Y
Ehlers-Danlos syndrome I, II, and V	AD, XR	Y	Y	Y	N
Huntington disease	AD	Y	Y	Y	N
Marfan syndrome	AD	Y	Y	Y	N
Neurofibromatosis	AD	N	Y	Y	N
Tuberous sclerosis	AD	N	Y	?	Y
Usher syndrome, type I	AR	N	N	?	Y
Wilson disease	AR	Y	Y	?	N

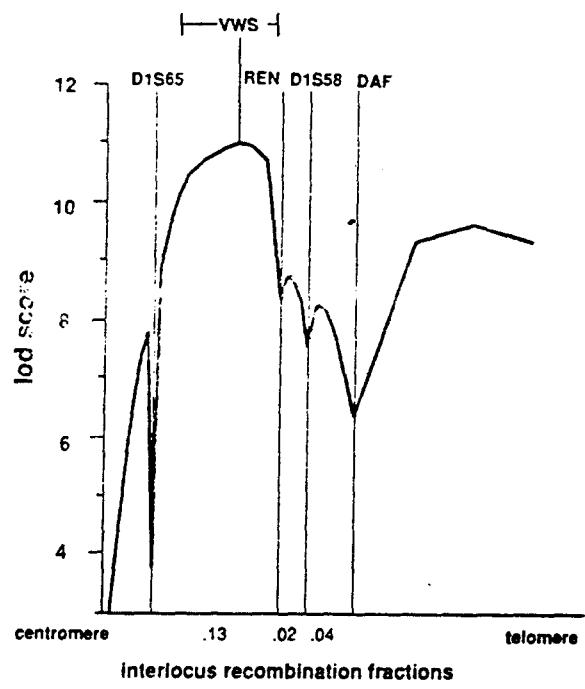
<sup>a</sup>Abbreviations: AD, autosomal dominant; AR, autosomal recessive; N, no; XD, X-linked dominant; XR, X-linked recessive; Y, yes; ?, unknown.



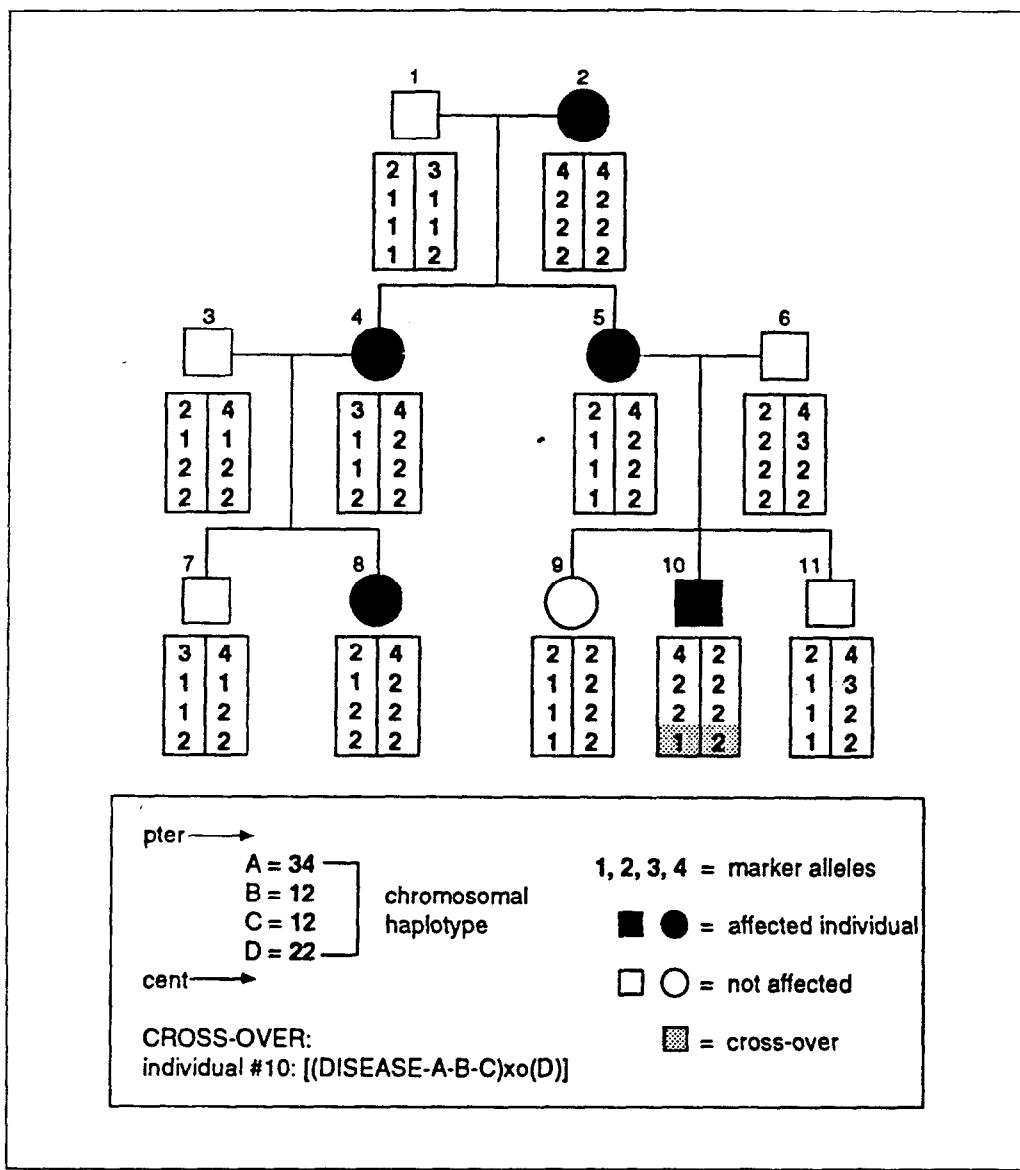
**Figure 1.7.1** Genetic map showing recombination frequencies between markers A and B and between markers B and C, along with designation of each of the four intervals to be tested in a multipoint linkage analysis.

$$\text{multipoint lod score} = \log_{10} \frac{L(\text{Ped} \mid \theta_{A-B}, \theta_{B-C}, \theta_{\text{disease}=x})}{L(\text{Ped} \mid \theta_{A-B}, \theta_{B-C}, \theta_{\text{disease}=0.50})}$$

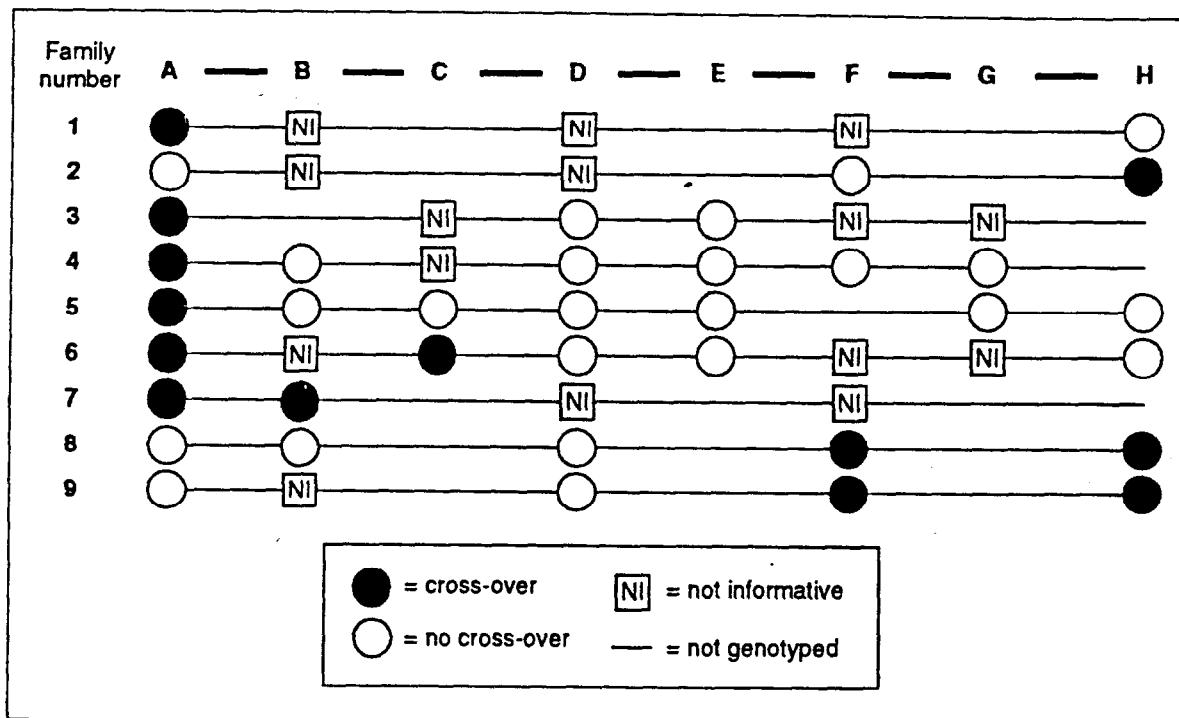
$$\text{multipoint lod score} = \log_{10} \frac{L(\text{Ped} \mid \theta_{\text{disease}-A=0.05}, \theta_{A-B=0.15}, \theta_{B-C=0.20})}{L(\text{Ped} \mid \theta_{\text{disease}-A=0.50}, \theta_{A-B=0.15}, \theta_{B-C=0.20})}$$



**Figure 2** Results of location-score analysis of VWS for region surrounding REN. Brackets flanking VWS indicate 1-lod support interval. Anchor-map distances are presented as frequency of recombination for each interval.

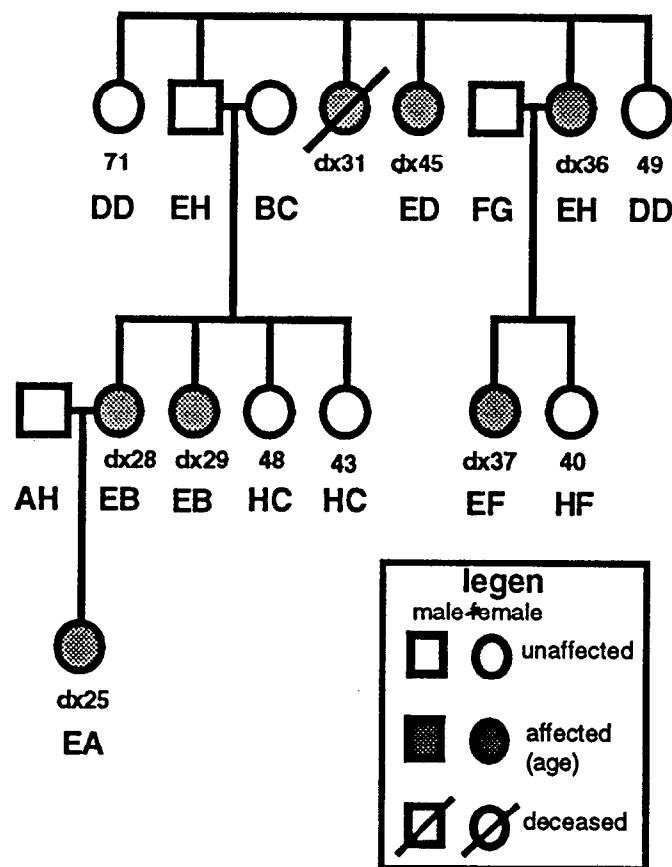


**Figure 1.4.6** Portion of a large disease pedigree depicting a key cross-over individual. Individual 10 is crossing over for marker D. Thus marker D defines a distal flanking marker for the disease in question, and the sequence of loci must be disease-A-B-C-(cross-over)-D.

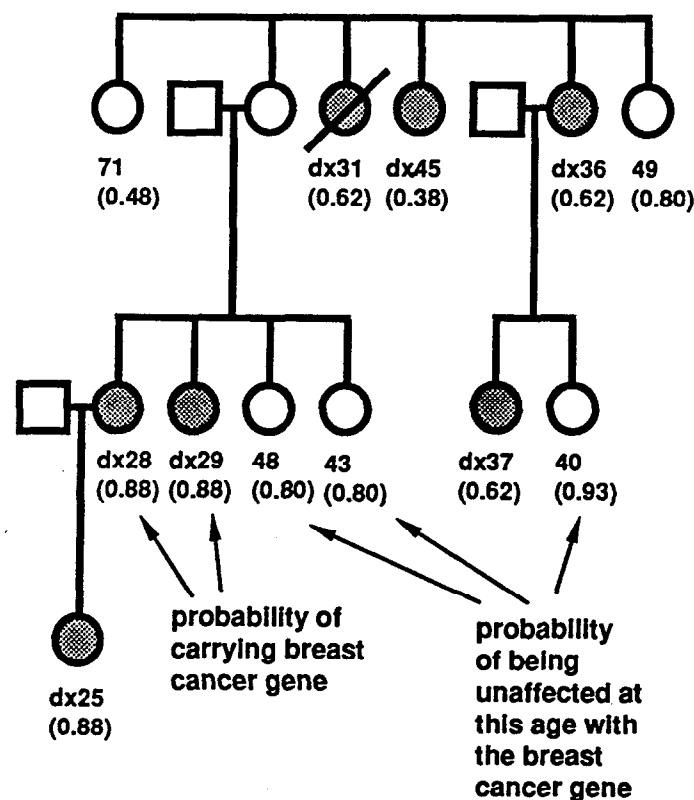


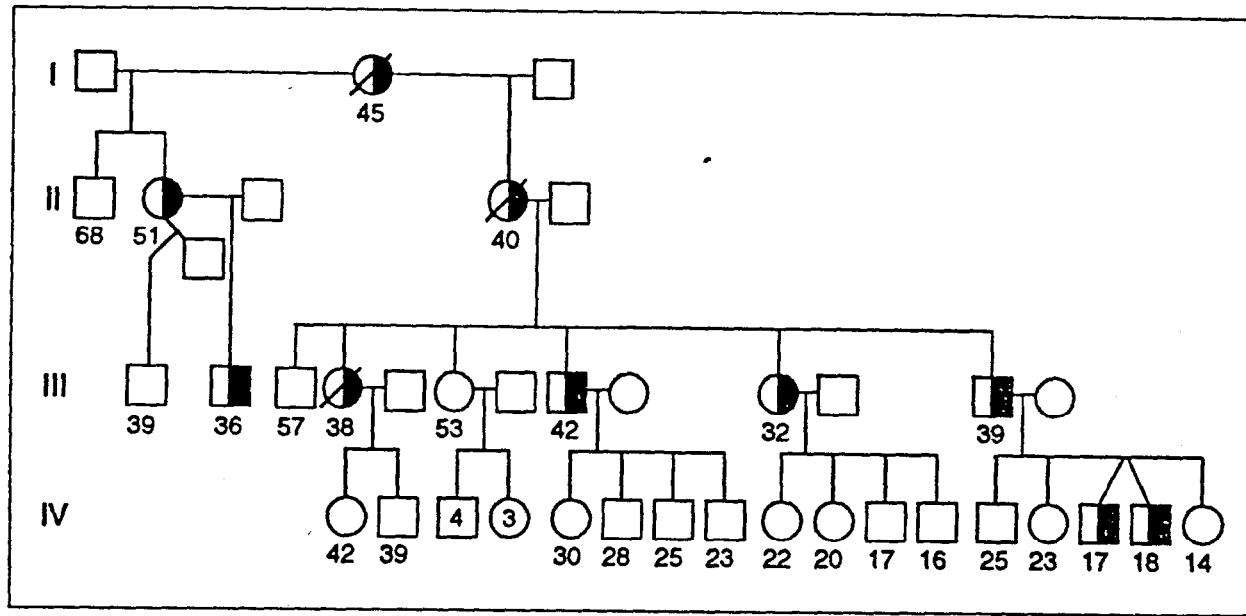
**Figure 1.4.7** Method for summarizing all cross-over individuals in all family data used in a study. The most likely region for the disease gene is between markers C and F, the region where no cross-overs are found.

## MCK-Breast Cancer Family 1 (D17S74)

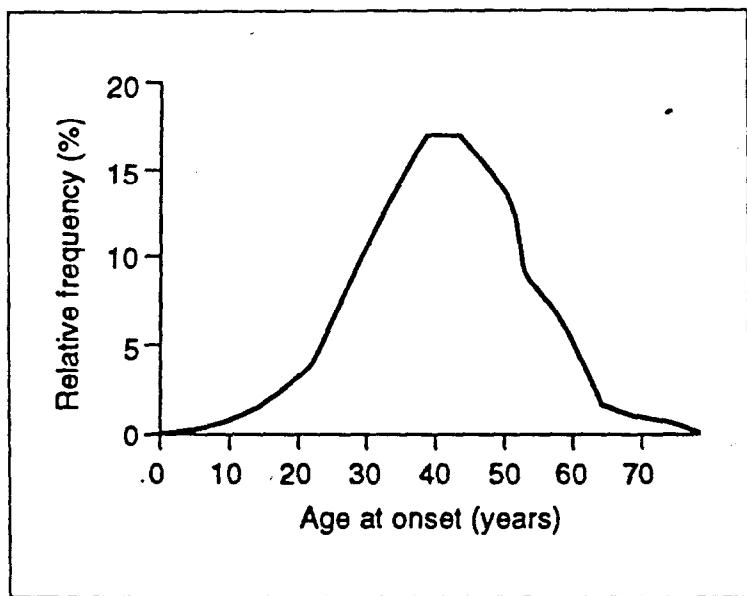


# MCK-Breast Cancer Family 1





**Figure 1.1.3** Pedigree of family with Huntington disease showing age-dependent penetrance. Few persons in youngest generation are affected because most HD gene carriers among them are not old enough to manifest symptoms. A half-shaded symbol indicates the person is affected. A slash through the symbol indicates that the person is deceased. Numerals within symbols represent multiple unaffected individuals of the same sex; numerals below symbols represent age at onset (for affected individuals) or age at last exam (for unaffected individuals).



**Figure 1.1.2** Distribution of age at onset in Huntington disease ( $n = 610$ ). Percentage of total cases (relative frequency) is plotted against age at which they occur.  
From Farrer, 1985.

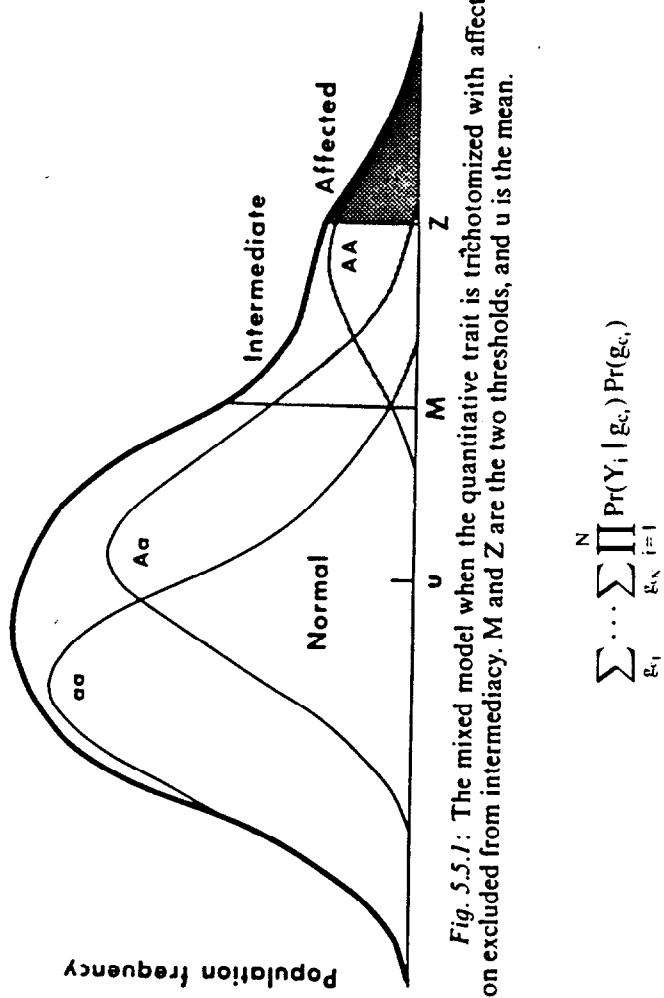
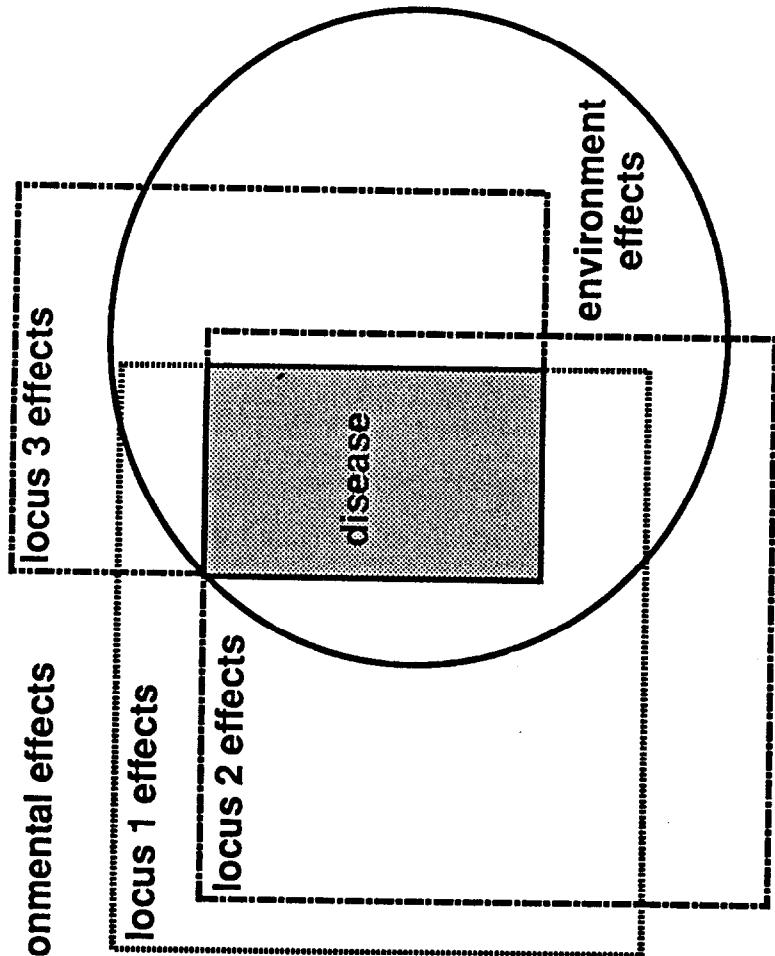
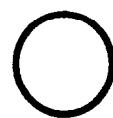


Fig. 5.5.1: The mixed model when the quantitative trait is trichotomized with affection excluded from intermediacy. M and Z are the two thresholds, and u is the mean.

$$\sum_{\mathcal{E}_1} \cdots \sum_{\mathcal{E}_N} \prod_{i=1}^N \Pr(Y_i | \mathcal{E}_{c_i}) \Pr(\mathcal{E}_{c_i})$$

## Underlying etiology of a complex phenotype

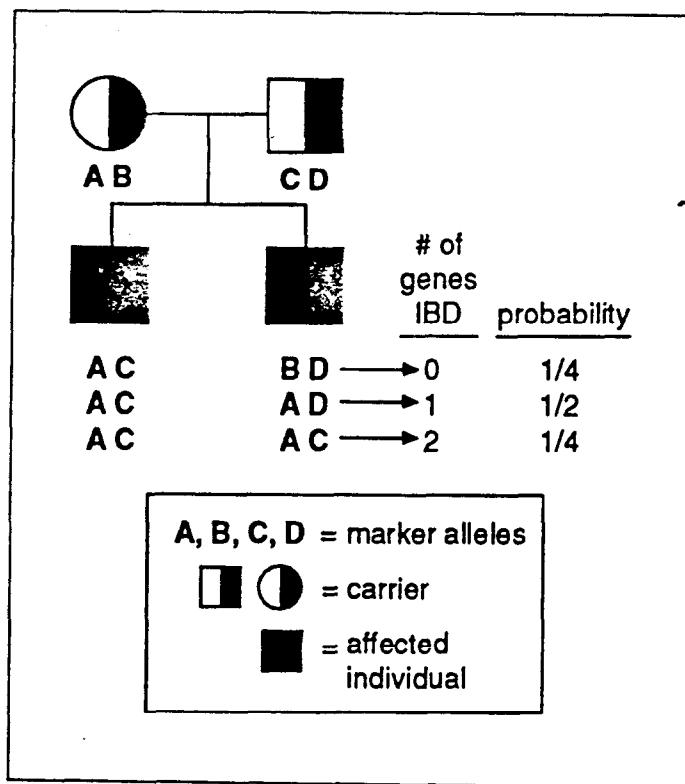


**Table 1.1.3 Some Disorders Caused by Gene-Environment Interaction**

Phenotype	Genotype	Environmental risk factor	Mechanism
Alcoholism	Alcohol dehydrogenase deficiency	Alcohol consumption	Genotype exacerbates effect of risk factor
Emphysema	$\alpha$ -1 antitrypsin deficiency	Smoking	Genotype and risk factor independently confer susceptibility
Heart disease	Familial hypercholesterolemia	High-cholesterol diet; lack of exercise	Genotype and risk factor independently confer susceptibility
Hemolytic anemia	G6PD deficiency	Fava bean consumption	Both genotype and risk factor required for expression
Malignant hypothermia	Malignant hypothermia	Anesthesia	Both genotype and risk factor required for expression
Skin cancer	Xeroderma pigmentosum	UV exposure	Genotype exacerbates effect of risk factor

## **PREMISE:**

In a disease with a non-zero fraction attributable to genetic factors, regions of the genome that are responsible for the disease should show non-random patterns of variation



**Figure 1.4.8** Identity-by-descent (IBD) relationship. IBD relationships form the basis for the sib-pair (SP) linkage test. Shared alleles and their associated probabilities are given.

## Kinship ( $\phi_{ij}$ ):

For individuals i and j in the population, the probability that a gene chosen randomly from one is identical by descent with one chosen randomly from the other

For pedigree analysis Weeks and Lange proposed (AJHG 42:315-326, 1988):

$$T = \frac{Z - E(Z)}{\sqrt{\text{var}(Z)}}$$

where:

$$Z = \sum_{i < j} Z_{ij}$$

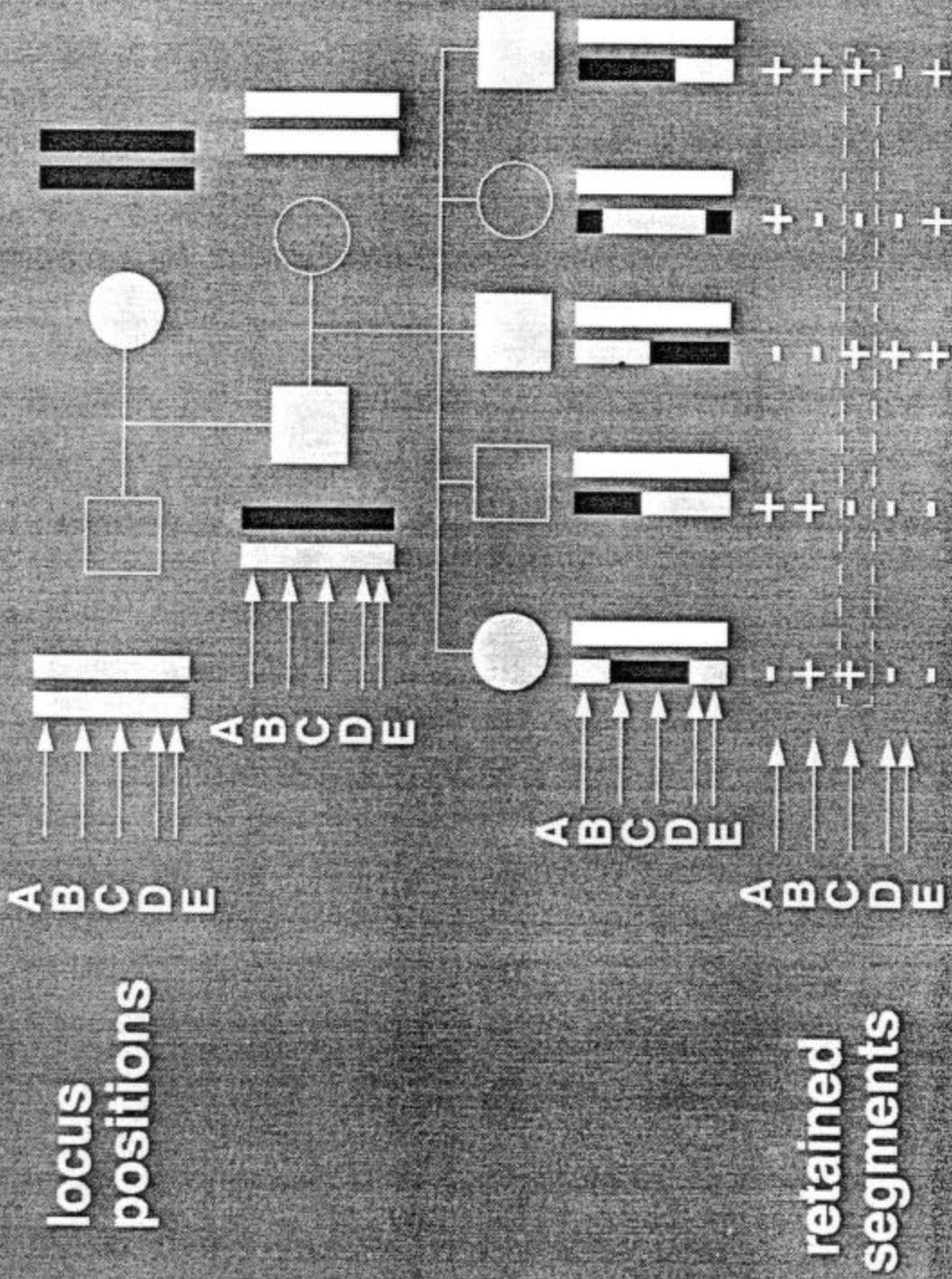
$$\begin{aligned} Z_{ij} = & 1/4\delta(G_{ix}, G_{jx})f(p_{Gix}) + 1/4\delta(G_{ix}, G_{jy})f(p_{Gix}) \\ & + 1/4\delta(G_{iy}, G_{jx})f(p_{Gi y}) + 1/4\delta(G_{iy}, G_{jy})f(p_{Gi y}) \end{aligned}$$

$$E(Z) = \sum_{i < j} E(Z_{ij})$$

$$E(Z_{ij}) = \varphi_{ij} \sum_{k=1}^n p_k f(p_k) + (1 - \varphi_{ij}) \sum_{k=1}^n p_k^2 f(p_k)$$

$$\text{var}(Z) = \sum_{i < j} \sum_{k < l} E(Z_{ij} Z_{kl}) - E(Z)^2$$

for all pairs of affected individuals i and j



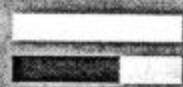
Genetic Epidemiology Group

Fox Chase Cancer Center

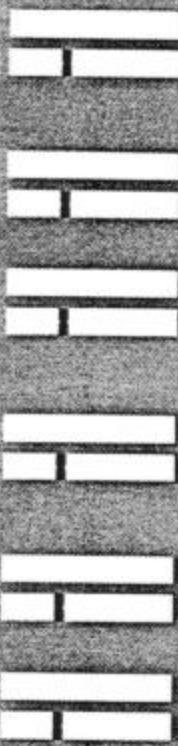
founder  
generation:



after One  
generation:

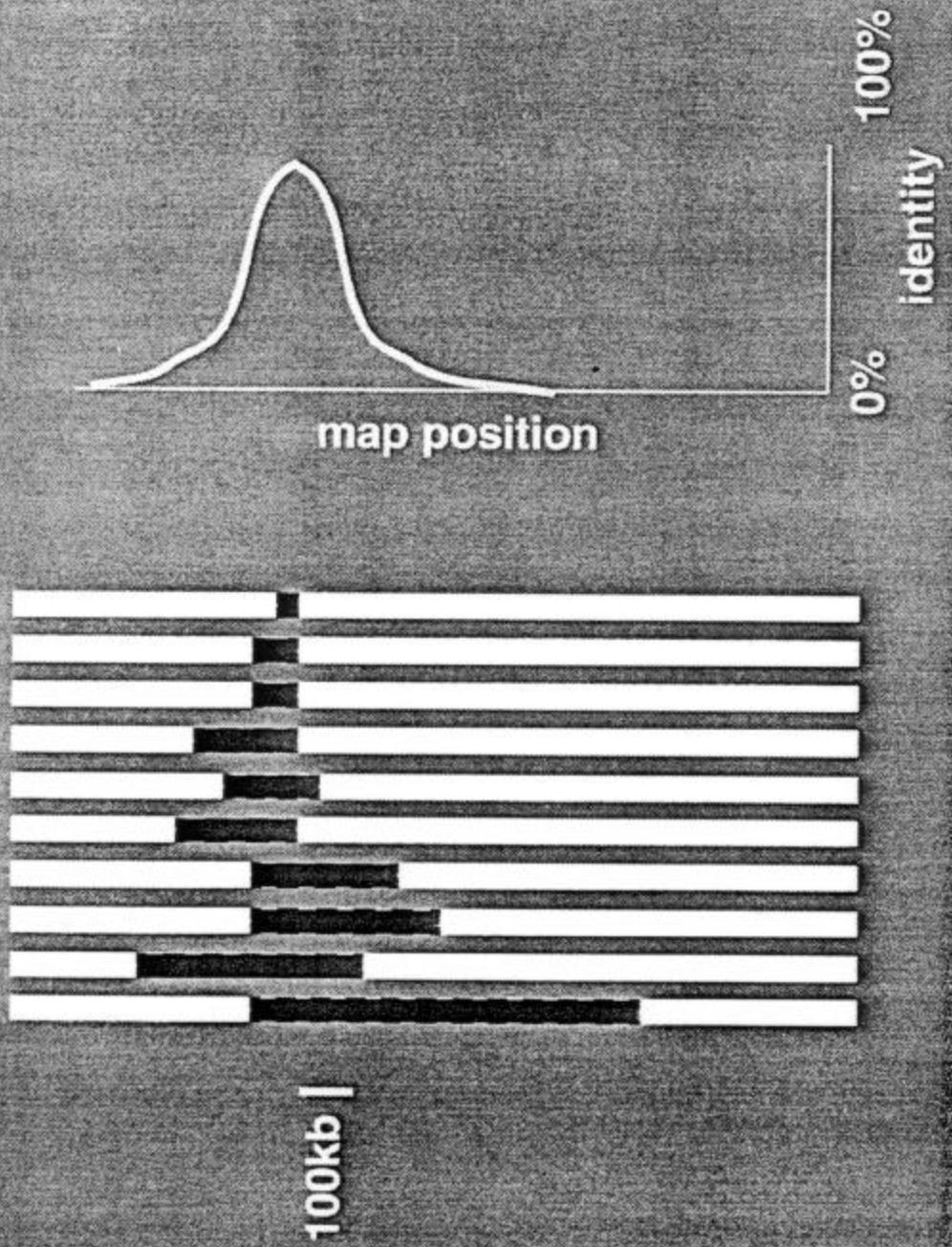


after many  
generations:

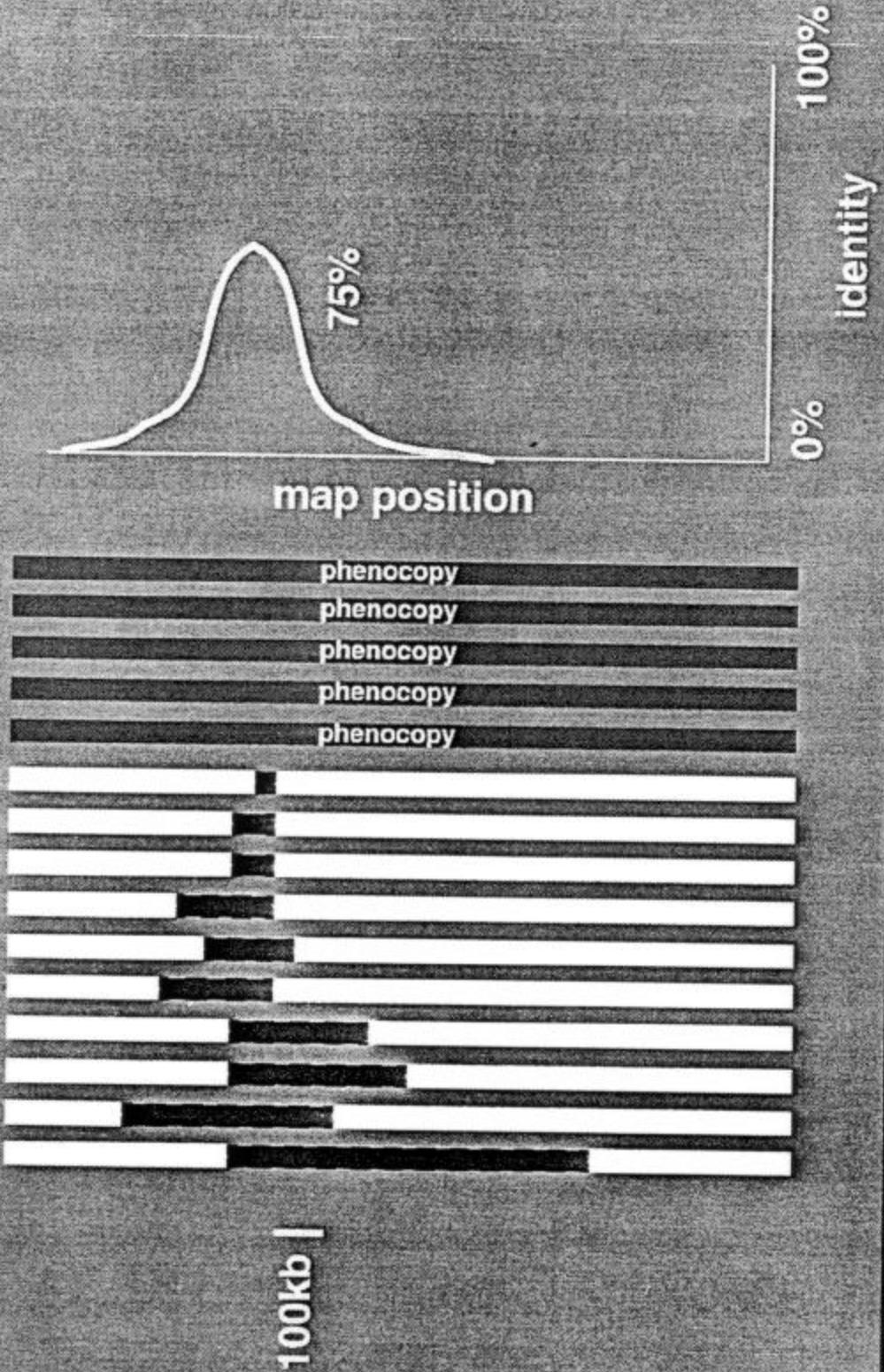


---

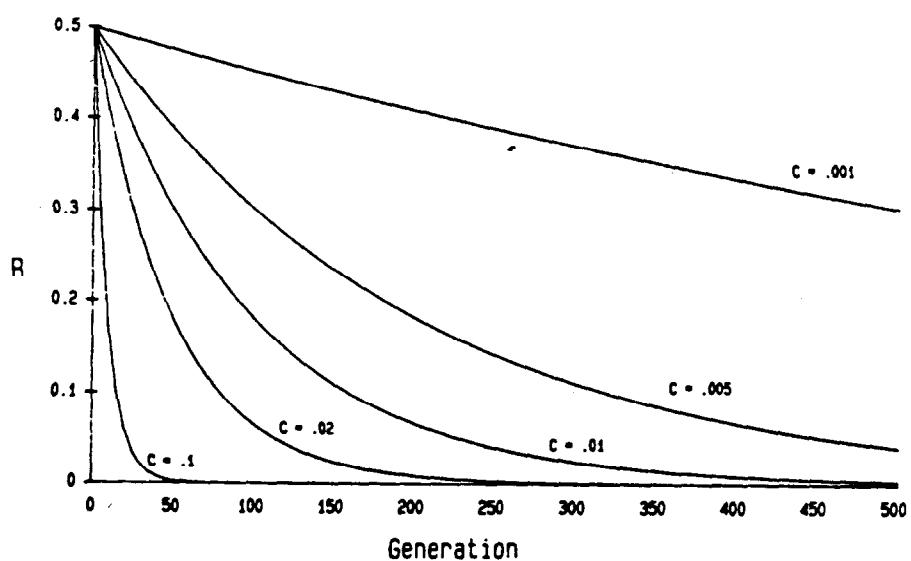
Cooperative Human Linkage Center  
Fox Chase Cancer Center



Cooperative Human Linkage Center  
Fox Chase Cancer Center

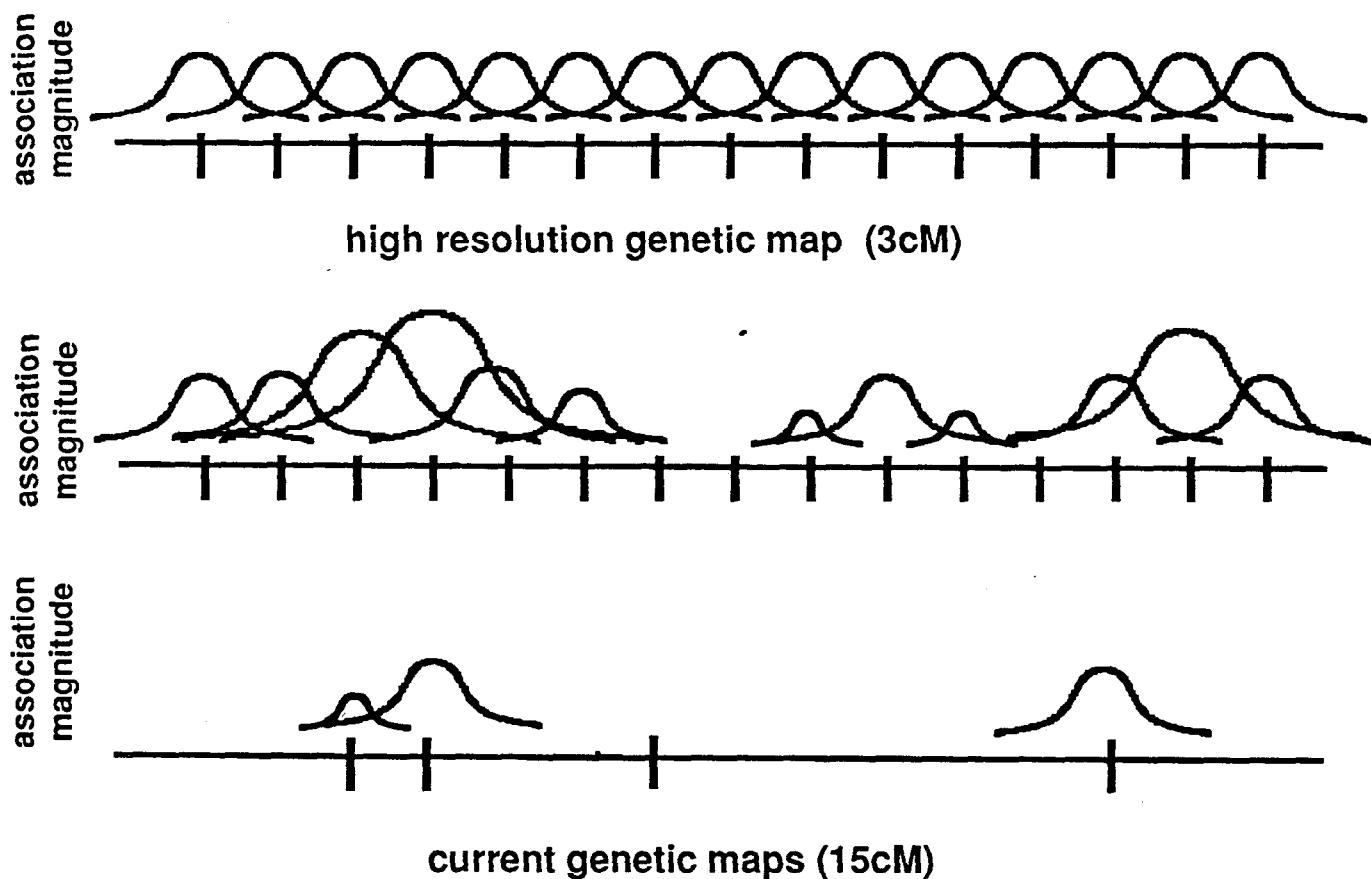


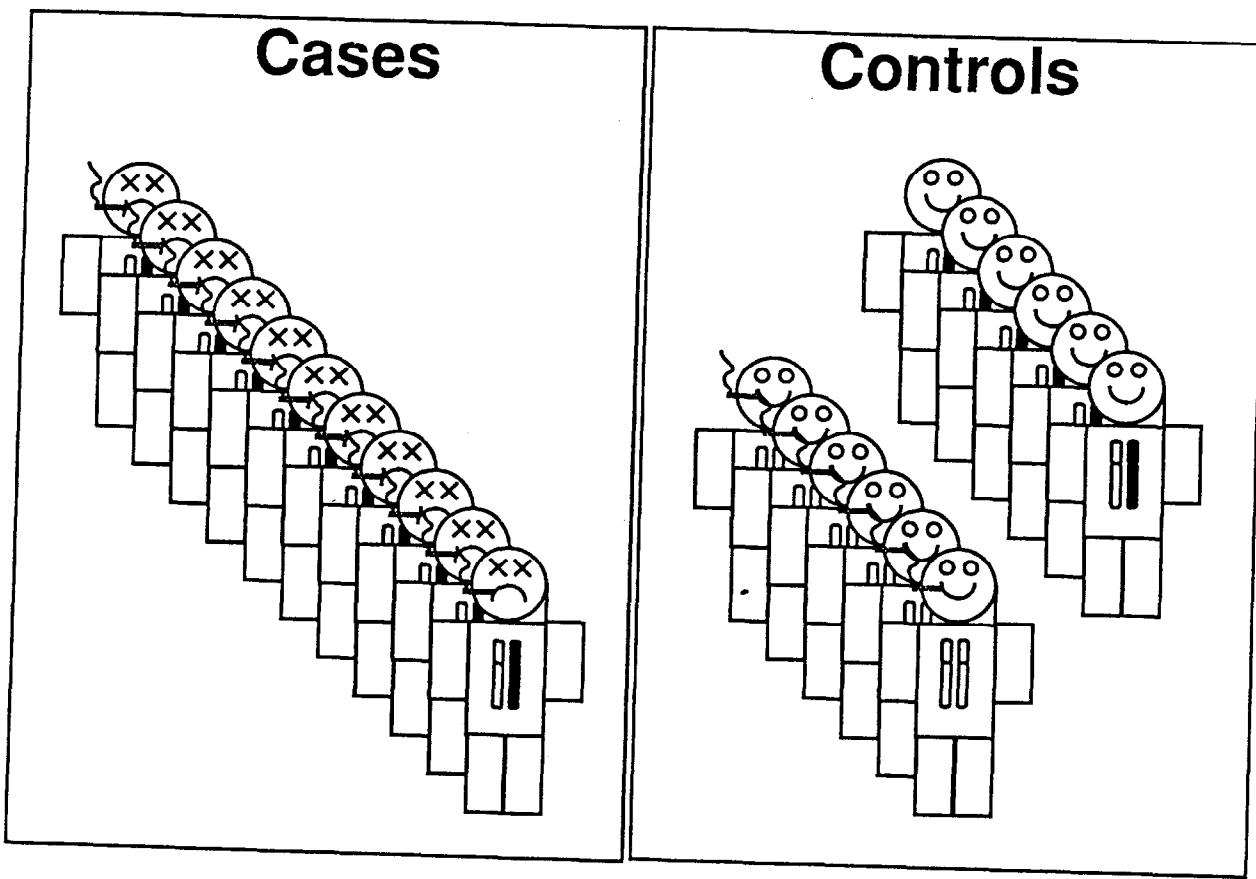
Cooperative Human Linkage Center  
Fox Chase Cancer Center



**FIGURE 1.** Decay of linkage disequilibrium ( $R$ ) with time as a function of recombination ( $c$ ).

## genome search by association walking





█ 50%

█ 50%

█ 75%

█ 25%

**Table 5–19.** The association between transforming growth factor alpha alleles and the risk of cleft lip and palate

	Genotypes			
	$C_1C_1$	$C_1C_2$	$C_2C_2$	Total
Cases	59	17	2	78
Controls	89	8	1	98
Odds ratio	1.0	3.2	3.0	

$C_1$  allele is 3.0 kbp restriction fragment length polymorphism (RFLP) segment, while the  $C_2$  allele is 2.7 kbp RFLP segment using Taq1 restriction enzyme analysis of the transforming growth factor alpha (TGFA) gene probe.

Data adapted from Arding et al. (1989).

# **Human Genetic Mapping**

## **Electronic Address List**

### **General**

<http://linkage.cpmc.columbia.edu>

<http://lenti.med.umn.edu/linkage/linkage.html>

### **Reference Maps and Data**

<http://www.chlc.org>

[http://gopher.genethon.fr/genethon\\_en.html](http://gopher.genethon.fr/genethon_en.html)

<http://www.cephb.fr>

<http://www.ncbi.nlm.nih.gov>

<http://gdbwww.gdb.org>

### **Trait Mapping Software**

#### **LINKAGE ANALYSIS (comprehensive)**

<http://linkage.cpmc.columbia.edu/software.html>

#### **FASTLINK version of LINKAGE PACKAGE**

<http://www.cs.rice.edu/~scaffer/fastlink.html>

#### **HETEROGENEITY ANALYSIS**

<ftp://linkage.cpmc.columbia.edu/software/homog>

#### **APM ANALYSIS**

<ftp://watson.hgen.pitt.edu/pub>

#### **LINKAGE DISEQUILIBRIUM**

<ftp://linkage.cpmc.columbia.edu/software/diseq>

#### **SIMLINK SOFTWARE (for power calculations)**

<http://www.sph.umich.edu/group/statgen/software>

## Bibliography

Genetic Mapping. In: Current Protocols in Human Genetics, Chapter 1, Dracopoli NC, Haines JL, Korf BR, Morton CC, Seidman CE, Seidman JG, Moir DT, Smith D, eds, Current Protocols Pub, Brooklyn, NY, pp 1.0.1-1.7.93, 1995.

Khoury MJ, Beaty TH, Cohen BH. Fundamentals of Genetic Epidemiology. Oxford Univ. Press, New York, 1993.

Morton NE. Outline of Genetic Epidemiology. Karger, Basel; New York, 1982.

Ott, J. Analysis of human genetic linkage. Rev. ed., The Johns Hopkins University Press, Baltimore, MD, 1991.

Terwilliger JD and Ott J. Handbook of Human Genetic Linkage. Johns Hopkins Univ. Press, Baltimore, 1994.