HST.161 Molecular Biology and Genetics in Modern Medicine
Fall 2007

Harvard-MIT Division of Health Sciences and Technology
HST.161: Molecular Biology and Genetics in Modern Medicine, Fall 2007
Course Directors: Prof. Anne Giersch, Prof. David Housman

From Mendelian Inheritance Pattern to Human Disease Gene

- In each family which shows a Mendelian inheritance pattern there must be a site of DNA sequence variation which is responsible for the phenotype
- Challenge: How to locate that DNA sequence

Meiosis

- Reduces DNA content from diploid to haploid.
- Generates extraordinary level of genetic diversity in gametes.
- Errors during meiosis in chromosome segregation can have profound consequences for the fertilized embryo.

Image removed due to copyright restrictions.
Diagram of fertilization and the sex differentiation that results from X,Y and X,X chromosome pairings.

Accuracy of DNA Distribution in Meiosis is the Key to Genetic Linkage Tests

- Reduces DNA content from diploid to haploid.

- Generates extraordinary level of genetic diversity in gametes.

- Almost all live born human are the product of accurate meoises

Image removed due to copyright restrictions.
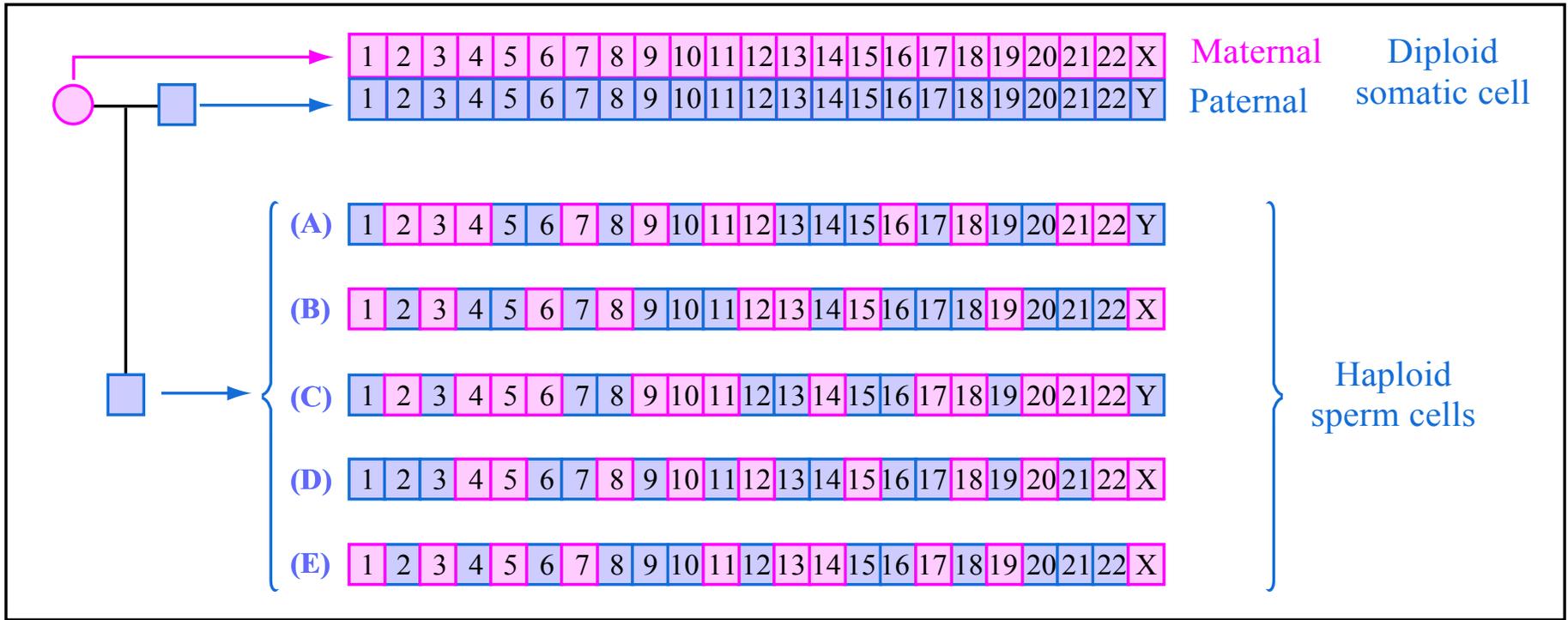Diagram of the meiosis and mitosis processes.

Figure by MIT OpenCourseWare.

Genetic Linkage

- In a diploid organism such as humans, there are two very similar but not necessarily precisely identical copies of every DNA sequence (referred to as alleles at a genetic locus). In an accurate meoisis, there is a 50/50 chance that a gamete will receive one of these two copies for every sequence in the genome.

- For DNA sequences located on different chromosomes the likelihood that two gametes will receive the same pair of sequences is (0.5 X 0.5)=0.25. However, if the two sequences are located extremely close together on the same chromosome (so that they are "linked" together) then the probability that the same pair of sequences will appear in a gamete will be close to 50%.

- This principle underlies a very powerful test to determine whether the site of genetic variation underlying a phenotype exhibiting a Mendelian pattern of inheritance in a family is at a specific location in the human genome.

Genetic Linkage

- Every DNA sequence has a 50/50 chance of being transmitted to the next generation at meiosis.

- Because chromosomes assort randomly at meiosis, if two genetic loci are on different chromosomes the two alternative alleles of each locus have a 50/50 chance of appearing together in a gamete.

- However, if two genetic loci are located close to each other on the same chromosome, then alleles of the two loci will travel together during meiosis unless there has been a crossover between them.

- Alleles at two loci which are extremely close together will travel together at meiosis (cosegregate) almost all the time.

- This principle underlies a very powerful test to determine whether the site of genetic variation underlying a phenotype exhibiting a Mendelian pattern of inheritance in a family is at a specific location in the human genome
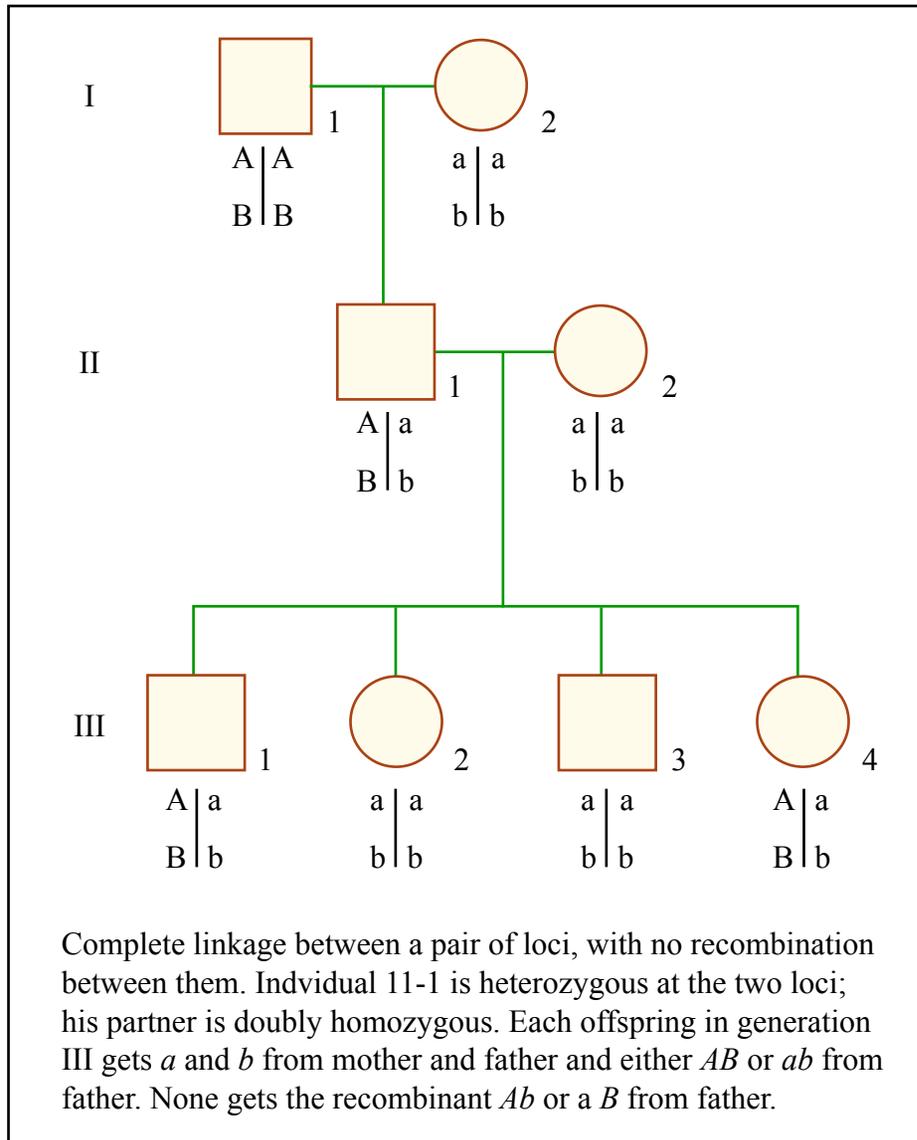
Complete linkage between a pair of loci, with no recombination between them. Indvidual 11-1 is heterozygous at the two loci; his partner is doubly homozygous. Each offspring in generation III gets *a* and *b* from mother and father and either *AB* or *ab* from father. None gets the recombinant *Ab* or a *B* from father.
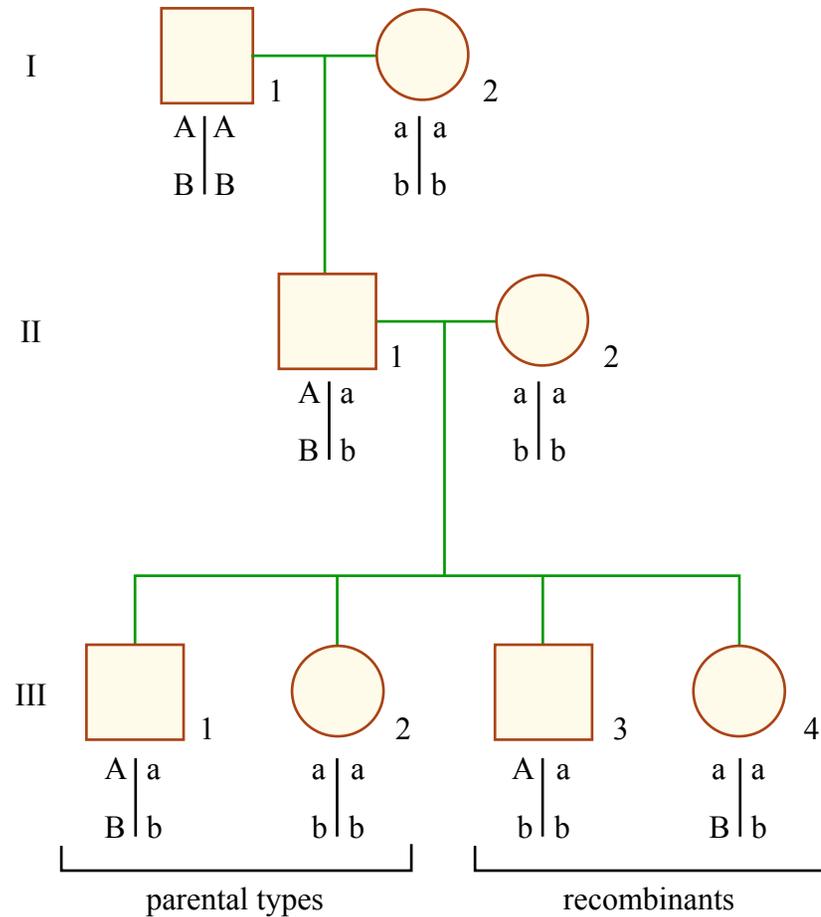
Figure by MIT OpenCourseWare.

Random segregation of alleles at two unlinked loci. Individual 11-1 is doubly heterozygous and produces four types of sperm: *AB*, *Ab*, *aB*, and *ab*. Each of these four combinations are represented in one of the offspring, resulting in equal numbers of parental (nonrecombinant) and recombinant offspring.

Figure by MIT OpenCourseWare.

Informative and uninformative meiosis

A meiosis is informative for linkage when we can identify whether or not the gamete is recombinant. Consider the male meiosis which produced the paternal contribution to the child in the four pedigrees below. We assume that the father has a dominant condition that he inherited along with marker allele A_1.

(A)

A_1A_1 A_2A_2

A_1A_2

(B)

A_1A_2 A_1A_2

A_1A_2

(C)

A_1A_2 A_1A_2

A_1A_1

(D)

A_1A_2 A_3A_4

A_1A_4
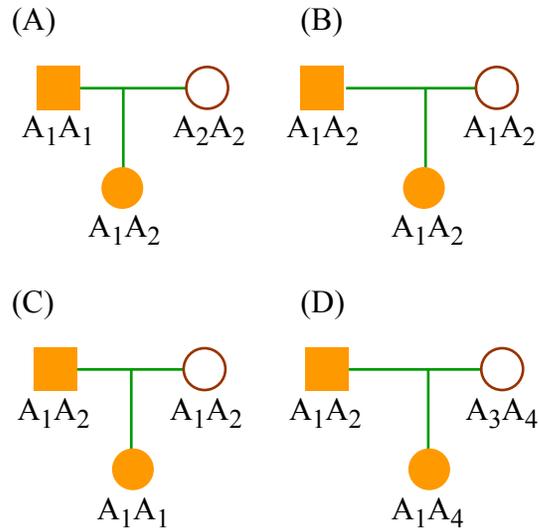
(A)   This meiosis is uninformative: the marker alleles in the homozygous father cannot be distinguished.

(B)   This meiosis is uninformative: the child could have inherited A_1 from father and A_2 from mother, or vice versa.

(C)   This meiosis is informative: the child inherited A_1 from the father.

(D)   This meiosis is informative: the child inherited A_1 from the father.

Figure by MIT OpenCourseWare.

Image removed due to copyright restrictions.
Illustration of the physical, actual crossing-over of alleles on chromosomes.

Chiasmata

- The two homologs for each chromosome (paternal and maternal) must form at least one chiasma during meiosis I
- The chiasmata are the sites of crossing over between homologs
- The chiasmata are crucial for the correct segregation of the two duplicated homologs to separate daughter nuclei because they play a key role in holding the maternal and paternal homologs together until the spindle separates them at anaphase I

Image removed due to copyright restrictions.
Illustration of chiasmata.

Images removed due to copyright restrictions.
Illustrations of single and double recombinants during and after meiosis crossover between maternal and paternal homolog chromosomes.

Measuring Recombination Distance

- Ratio of recombinant gametes to nonrecombinant gametes is a measure of recombination distance

- If recombinant gametes =1 % of total gametes then recombination distance between a and b is 1 centimorgan

Measuring Recombination Distance

- At larger distances proportion of Ab and aB is not precisely equal to number of centimorgans because of double recombinants

Dimensions of the human genome

- Whole genome haploid =3 billion base pairs
- Average chromosome=150 megabases
- Human genetic map=3000 centimorgans
- 1 megabase ~= 1 centimorgan
- Average human gene~= ~10 ,000 base pairs

Image removed due to copyright restrictions.
Graph of DNA recombination "hot spots"

- Recombination actually occurs at "hot spots" in the DNA.

- For two genes near each other on a chromosome, the probability of recombination in any given meiosis is extremely low.

- If two loci are on the same chromosome, they will be transmitted to gametes with some probability of recombination

- In the example at the right, because we know the genotypes of the paternal grandparents, we can establish "phase" with certainty in this family and identify the gametes carrying chromosomes transmitting A2B2 and A1B1 from the father as non-recombinant and the gametes transmitting A1B2 and A2B1 from the father as recombinant

- Note that the mother here is homozygous for both loci and her meiosis is therefore uninformative, but knowing which alleles she contributes to her children allows one to assign the alleles contributed by the father with certainty

Image removed due to copyright restrictions.
Pedigree color-coded for recombinant and nonrecombinant genotypes.
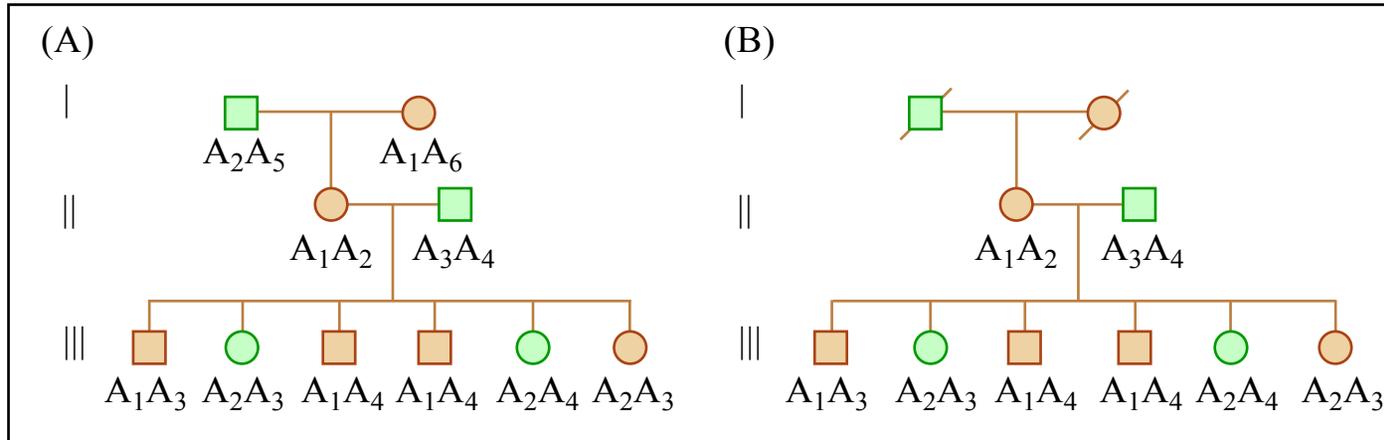
Determining Phase

Figure by MIT OpenCourseWare.

- In the example in panel A, we know that the disease gene, D was transmitted from the grandmother to the mother in a gamete which included allele A1 at the marker locus. We term this situation "phase known".

- In the example in panel B, that the disease gene, D could have been transmitted from the grandmother to the mother in a gamete which included either allele A1 or allele A2 at the marker locus. We term this situation "phase unknown".

The Same Phenotype Can Be Caused by Mutations in Different Genes in Different Families--This Situation is Termed Genetic Locus Heterogeneity

Linkage can be used to test whether a mutation has occurred at the same genetic locus in two different families in which the same clinical condition exhibits a Mendelian pattern of inheritance in each family

Example:Hypertrophic cardiomyopathy

# Genetic Linkage Can Be Used to Exclude Candidate Genes

- Test genetic markers which are very close to each candidate gene for cosegregation with the disease phenotype in a family

- Exclude candidate genes which show independent assortment with disease phenotype

How close does a genetic marker need to be?

- Marker within a few tens of kb will have virtually no recombination with a gene
- Even if 1 megabase away cosegregation between marker and gene is expected 99% of the time

Images removed due to copyright restrictions.
Diagrams showing HCM loci.

If A and a represent polymorphic sites (alleles) of a candidate gene for the genetic disease locus segregating in this family, that candidate gene would be excluded!
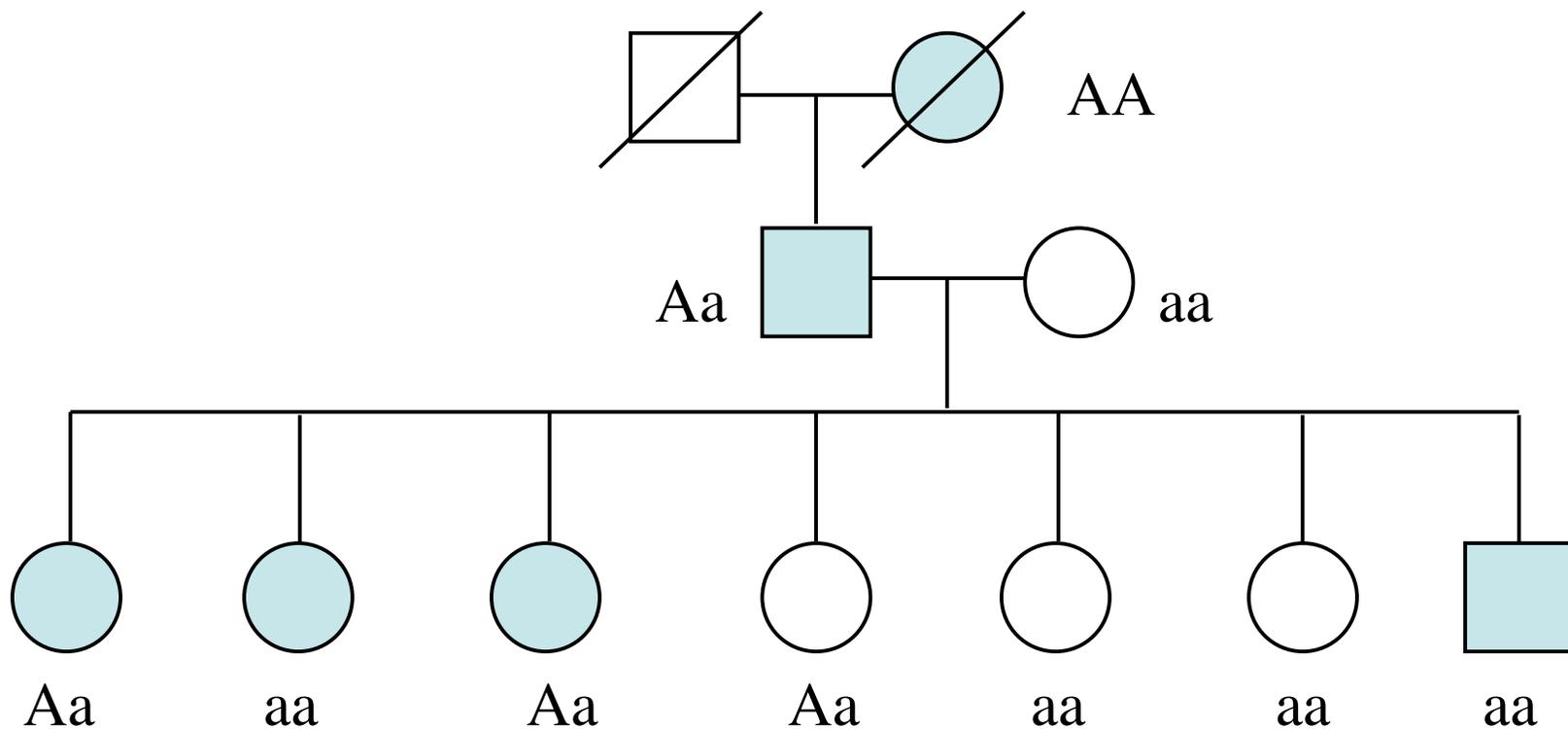
If A and a represent polymorphic sites (alleles) of a candidate gene for the genetic disease locus segregating in this family, then the data shown below would be sufficient to exclude that candidate gene!
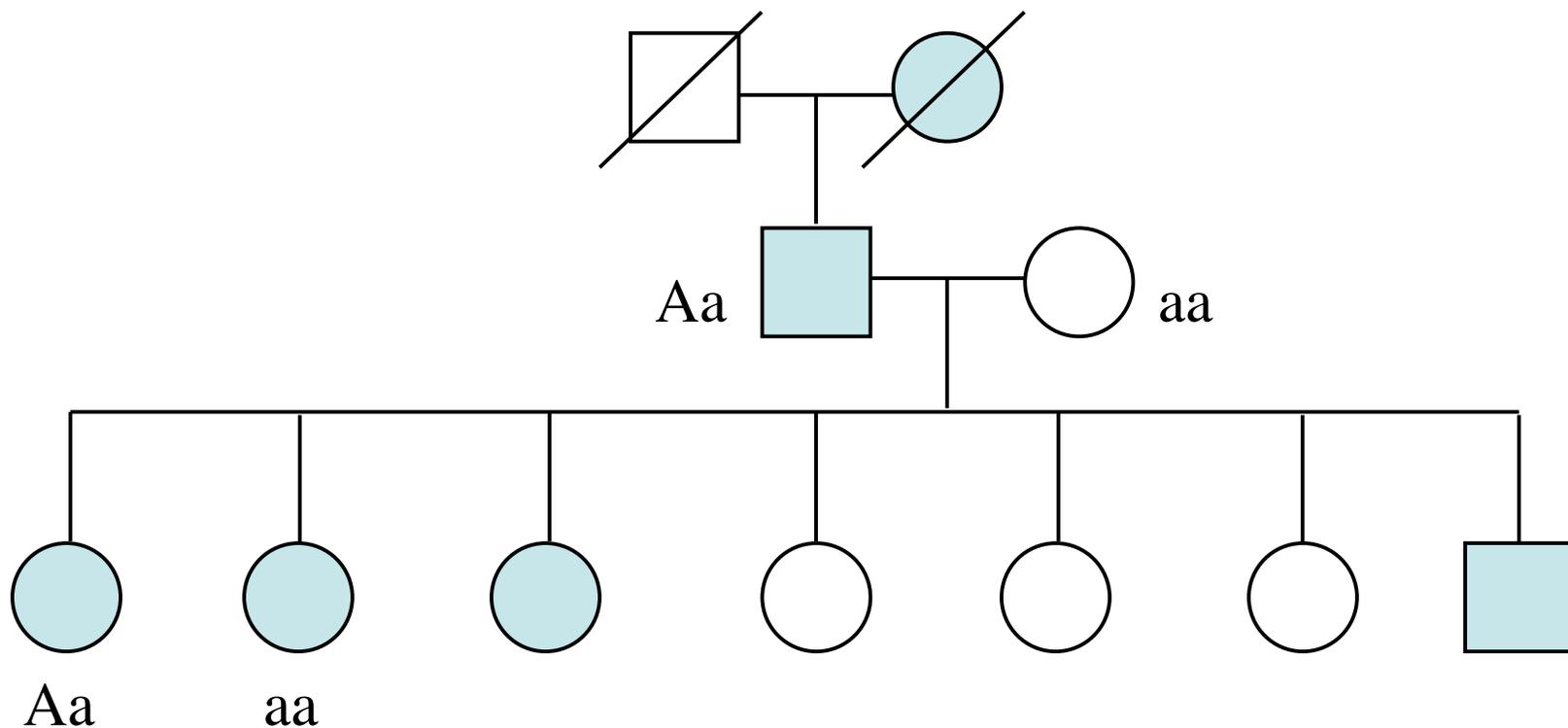
From Mendelian Inheritance Pattern to Human Disease Gene

- In each family which shows a Mendelian inheritance pattern there must be a site of DNA sequence variation which is responsible for the phenotype
- Challenge: How to locate that DNA sequence

Using Genetic Linkage to Find the Location of a Gene Causing a Clinical Phenotype

- A "genome wide scan" for genetic linkage between a series of DNA markers and the gene causing a clinical phenotype can be carried by testing one marker at a time for linkage to the gene causing the clinical phenotype

Using Genetic Linkage to Find the Location of a Gene Causing a Clinical Phenotype

- A statistical test for the likelihood of genetic linkage between the DNA marker and the gene causing the disease phenotype must be done--the LOD score

What Is a LOD Score?

- A LOD score is an odds ratio expressed in a logarithmic form

- Logarithmic form was used because when LOD scores were first developed calculators were not readily available

- LOD scores are used to evaluate the likelihood of linkage between two loci when a series of tests for linkage across the genome are carried out

What is an Odds Ratio?

- An odds ratio is the ratio of the odds that a given data set occurs if model #1 were correct compared to the odds that the same data set occurs if model #2 is correct

LOD Score is a Test for the Likelihood of Genetic Linkage

- A LOD score is a specific case of an odds ratio designed to test for linkage between two genes A and B

- The model in case #1 for the LOD score is that there is genetic linkage between A and B at a specific genetic map distance= Θ

- The model in case #2 is that A and B are unlinked

LOD Score

- The model for case #1 can be quite specific including parameters such as a phenocopy rate, penetrance values below 1, an age of onset correction curve, allele frequencies for each gene
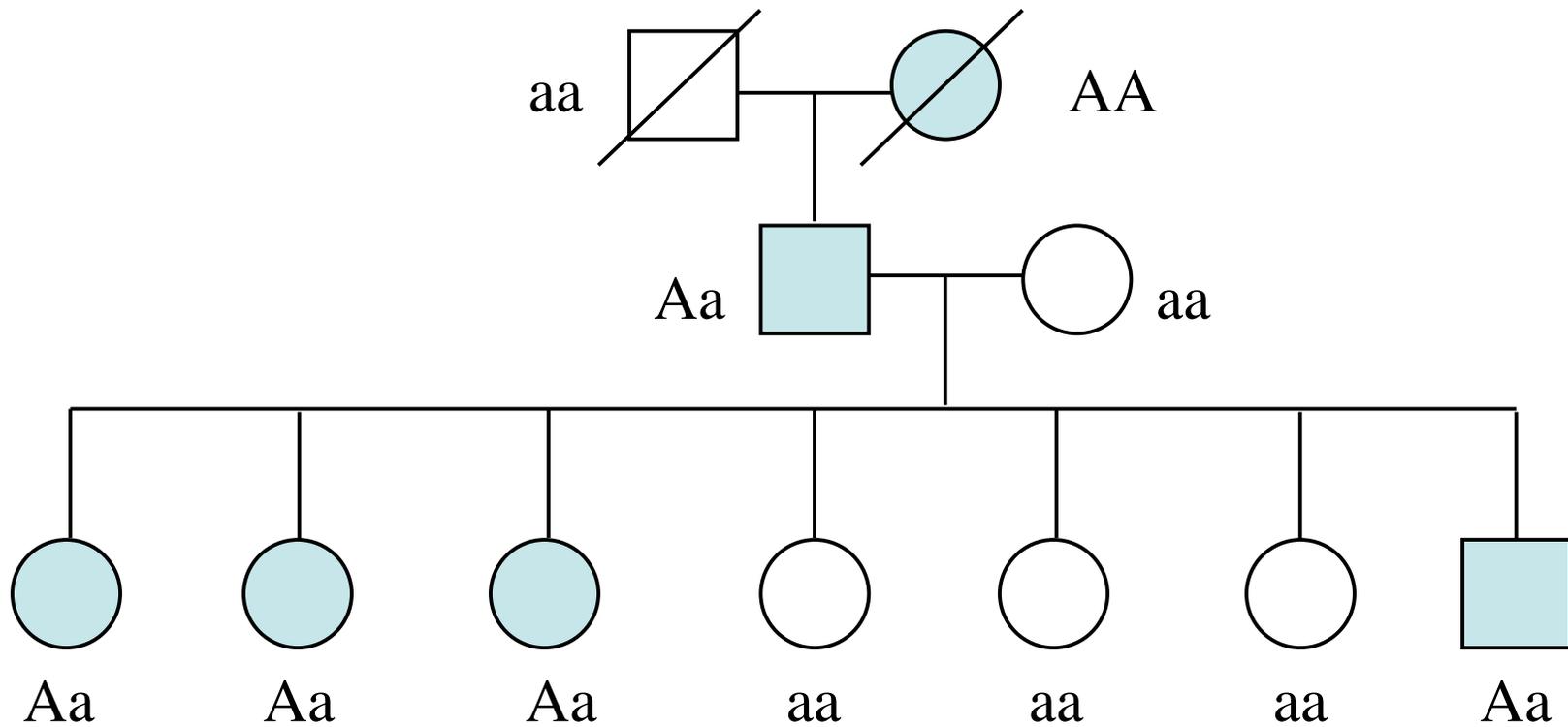
Parametric Method

- The specification of parameters in model #1 has led to the "LOD score method" being referred to as a "parametric method"

LOD Score

- Model #2 in a LOD score calculation is always the case of A and B unlinked, hence the likelihood of co-segregation of alleles of A and B in each informative meiosis in model #2 is always 0.5
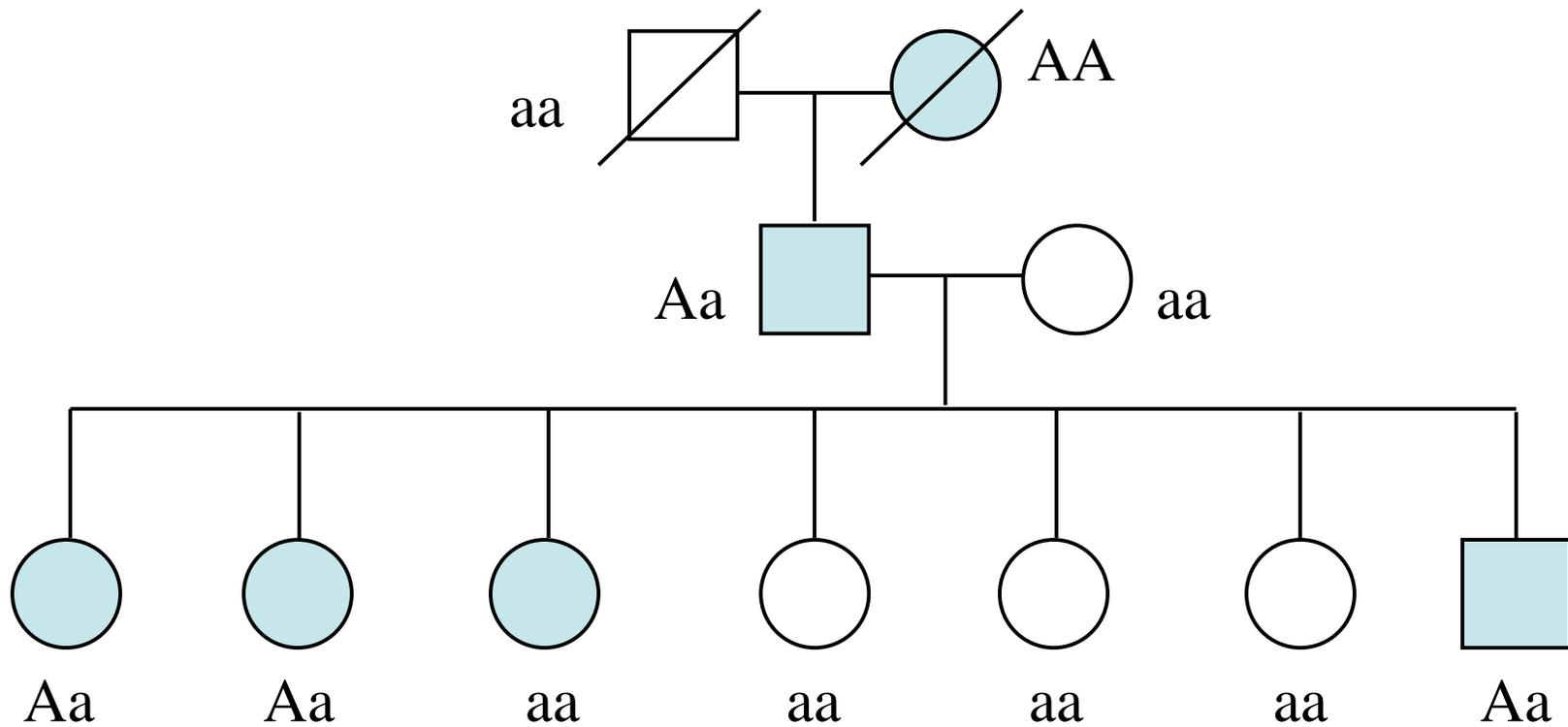
LOD Score

LOD Score

- To test for linkage at $\Theta = 0$ in this pedigree, we must calculate the odds that the data will occur if model #1 (linkage at $\Theta = 0$) compared to model #2 (no linkage)

- For model #1 the odds are 1 for each informative meioses that the data observed will be obtained

LOD Score

- For model #2 (no linkage) the odds are 1/2 for each informative meiosis that the data observed will be obtained

- The odds ratio will be $(1)^7/(0.5)^7 = 2^7$

- The logarithm of the odds ratio (LOD score) is approximately 2.1

LOD Score: An Obligate Recombinant

LOD Score

- To test for linkage at $\Theta = 0$ in the pedigree with these results, we must again calculate the odds that the data will occur if model #1(linkage at $\Theta = 0$) compared to model #2 (no linkage)

- For model #1 the odds are 1 for 6 of the 7 informative meioses that the data observed will be obtained, but for the seventh (recombinant) meiosis the odds that this event will occur according to the model are 0

LOD Score

- For model #2 (no linkage) the odds are 1/2 for each informative meiosis that the data observed will be obtained

- The odds ratio will be
  $(1)^6 X(0) /(0.5)^7 = 0$

- The logarithm of the odds ratio (LOD score) is now negative infinity

LOD Score

- Does this mean that the disease gene and the marker are not linked?
- Not necessarily, the analysis at $\Theta = 0$ simply excludes linkage at 0 recombination distance
- We can also calculated the LOD score when model #1 specifies a distance other than $\Theta = 0$ for the distance between genes
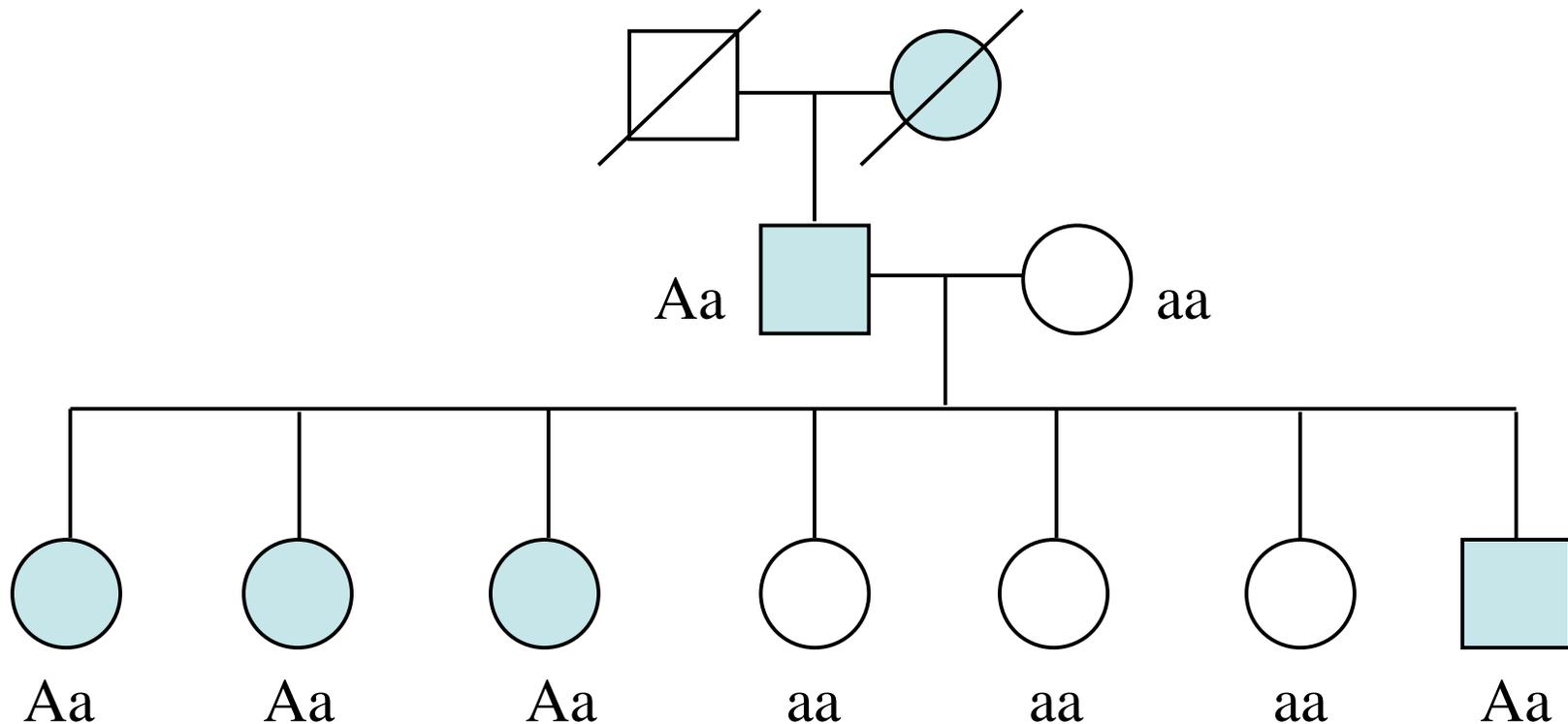
LOD Score

- To calculate a LOD score for $\Theta = 0.1$ for the same results, we can calculate the odds ratio as follows

- For each informative meiosis in which the result was co-segregation of marker and disease gene the likelihood of the data occurring according to the model is 0.9.

- For the recombinant meiosis, the likelihood of the data occurring according to the model is 0.1.

- The likelihood of the entire dataset occurring according to the model is $(0.9)^6 X (0.1)^1$

LOD Score

- The likelihood of the entire dataset occurring according to the alternative model (no linkage) is $(0.5)^7$

- The odds ratio will be $(0.9)^6 \times (0.1) / (0.5)^7 = 6.8$

- The logarithm of the odds ratio (LOD score) is now = 0.83 at $\Theta = 0.1$

If the genotypes of the grandparents are not known, then the PHASE of the marker with respect to the disease gene is not known. (i.e. we knew A is linked to the disease gene in model 1 in the previous data, but without the grandparental genotypes it is equally possible that a is linked to the disease gene in model 1). Not knowing the PHASE will lower the LOD score at $Q = 0$ by 0.3 compared to the same family with the phase known because we consider the two possible cases A linked to the diseases gene and a linked to the disease gene to have equal probability so we must sum the odds ratios for the two cases and divide by 2. The odds ratio for the case A linked to the disease gene at will be $(1)7/(0.5) 7 = 2^7$. While the odds ratio for the case a linked to the disease gene will be 0. The final odds ratio will therefore be $2^7 + 0 / 2 = 2^6$

LOD Scores and Multiple Test Correction

- For a single statistical test, a p value less than 0.05 is usually considered statistically significant
- However, if more than one test is performed, a correction must be made for the fact that a series of independent tests will be made

LOD Scores and Multiple Test Correction

- To test a set of markers which cover all human chromosomes, we estimate that approximately 50 independent tests are performed
- To achieve statistical significance equal to p less than 0.05, we multiply this value by 1/50 (0.02) to correct for these multiple tests

LOD Scores and Multiple Test Correction

- An odds ratio of 0.001 (LOD score=3) is therefore chosen as a threshhold for statistical significance in a test for linkage between 2 loci in the human genome

This pedigree will generate a very high positive LOD Score

Two approaches for collecting families for linkage analysis. The successful linkage of Huntington disease to a polymorphic marker on chromosome 4 relied in large part on a single large Venezuelan pedigree, a small part of which is shown in A. The successful linkage of cystic fibrosis to a polymorphic marker on chromosome 7, however, relied on a collection of many smaller families, some of which are shown in B.
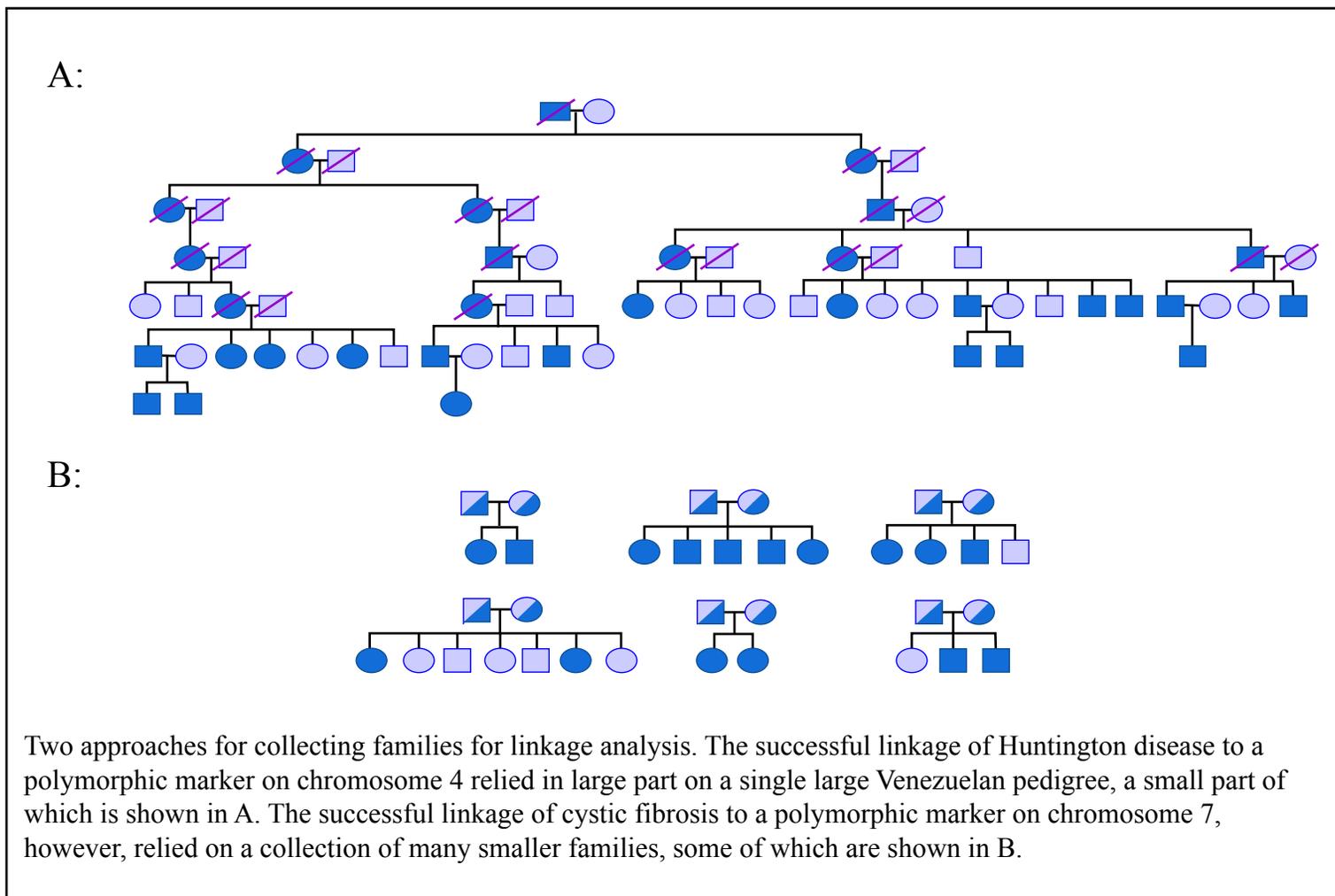
The top pedigree will generate a very high positive LOD score with a closely linked marker.

The bottom set of pedigrees can produce a high positive LOD score only if the same genetic locus (but not necessarily the same mutation) is responsible for the disease phenotype in all families

Finding the site of mutation causing the disease

- Once linkage is established between genetic markers and a disease gene, it is useful to narrow down the interval in which to search for the mutation causing the disease

Finding the site of mutation causing the disease

- Meioses which show recombination between marker and disease gene are now helpful in locating disease gene

Recombination Events Define Location of Disease Gene
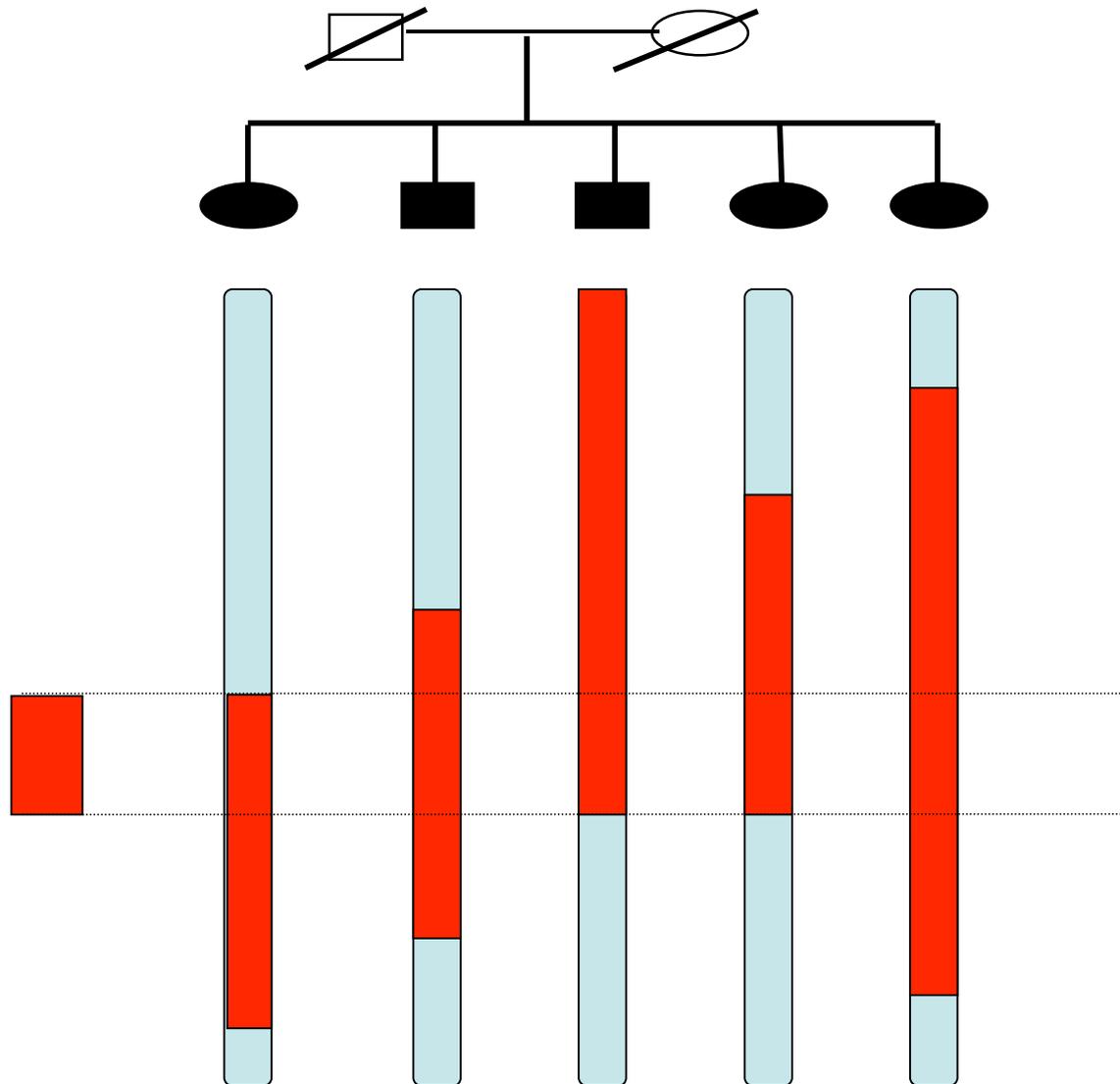
Many Genetic Markers in a Chromosomal Region
are Used to Bracket the Disease Gene as Closely as Possible

Image removed due to copyright restrictions.
Autozygosity mapping.

How many informative meioses are required to give good localization of the disease gene?

- 10 informative meioses may be enough to establish linkage
- BUT resolution will be at approximately 13.3 cM =approximately 13 megabases
- 25 informative meioses--resolution=5.6 cM approximately 5.6 megabases
- 100 informative meioses--resolution= 1.4 cM approximately 1.4 megabases

Once a Chromosomal Position for Gene is Established the Interval Must Be Searched Systemically for a Mutation which Causes Disease

- Candidate genes within the interval must be sequenced in affected family members to search for a DNA sequence difference which explains the inheritance of the pathological phenotype in that family