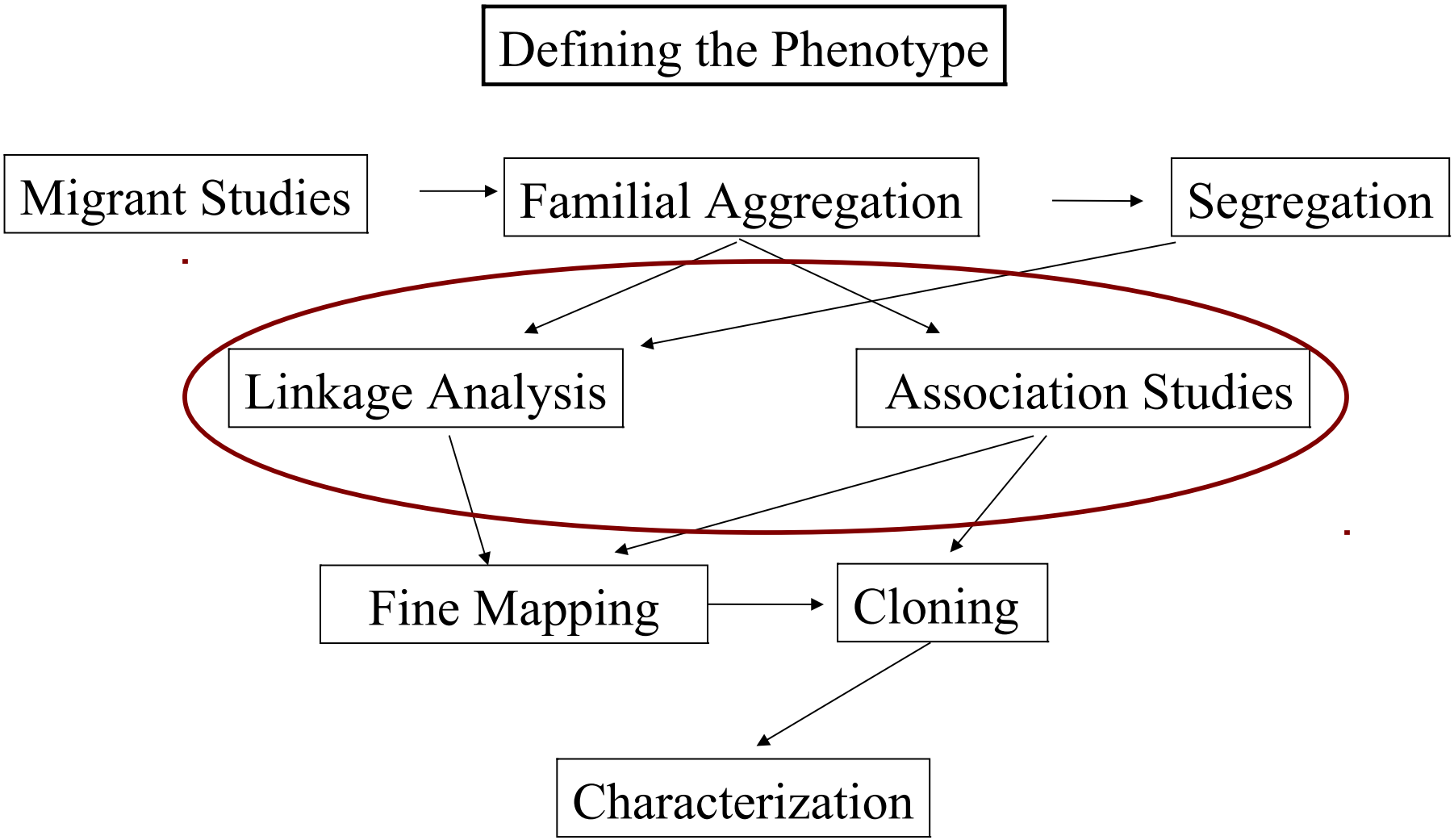
Family-Based Studies

“If you cannot get rid of the family skeleton, you may as well make it dance” (G.B. Shaw)

Outline

- Review last weeks homeworks
- **Linkage analysis, with PTC testing example**
  - **Background/Review**
  - Linkage analysis
- Family-based association tests
  - Overview
  - Trios/TDT
  - Discordant sibships

Review: Process of Genetic Epidemiology



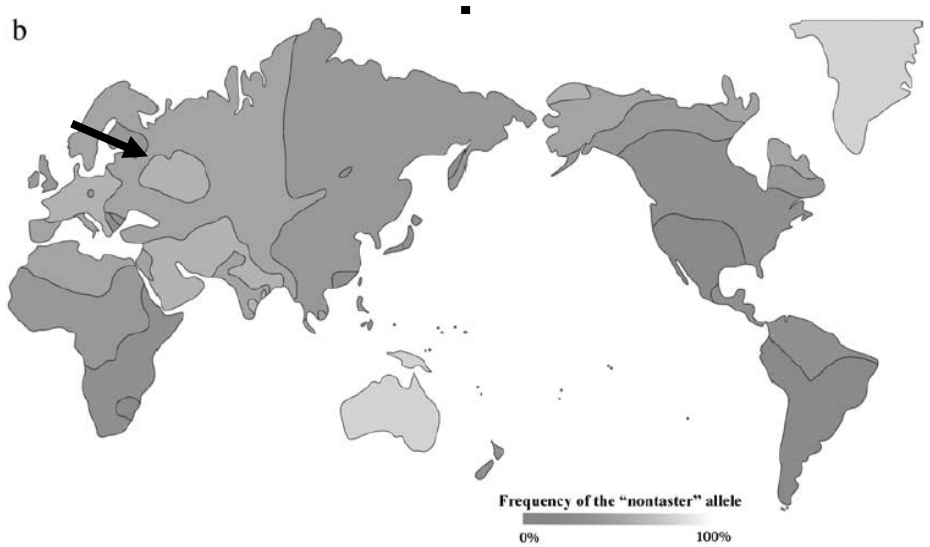
Review: Population stratification

- **Population stratification** is a form of confounding in genetic studies where a gene under study shows marked variation in allele frequency across subgroups of a population and these subgroups differ in their baseline risk of disease
- Previous e.g. of Pima & Papago Indians: Significant allele frequency difference *because of ancestry, not disease status*
- **Family-based tests** today tend to be **robust to this** issue (not require any special adjustment)

Outline

- Review last weeks homeworks
- **Linkage analysis, with PTC testing example**
  - Background/Review
  - **Linkage analysis**
- Family-based association tests
  - Overview
  - Trios/TDT
  - Discordant sibships

Geographic Distribution of PTC Phenotype



Wooding Genetics 2006, adapted from Cavalli-Sforza 1994

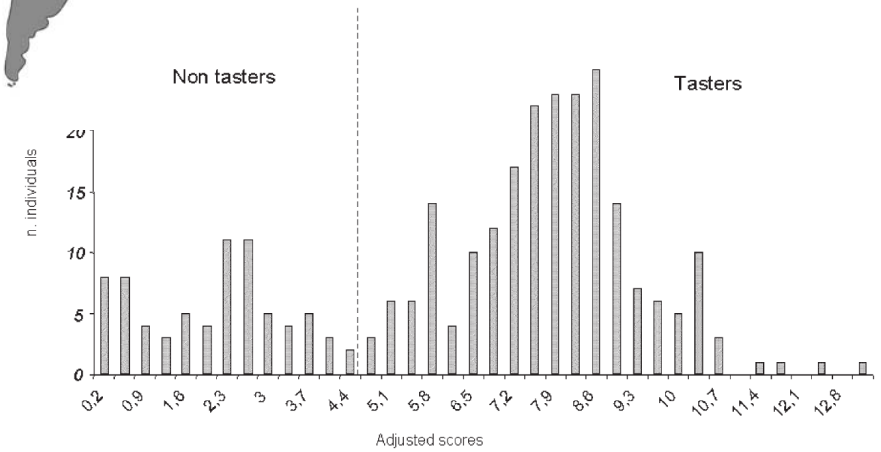


Figure 1 Distribution of PTC sensitivity in Talana. The *x*-axis indicates the PTC scores, adjusted for age and sex using the following formula: $n + [(a - a_M)/20] - 0.73$female, where n is the PTC tasted solution, a is the age of the individual and a_M is the average age of participants. The value of 0.73 was deducted in females since women are 0.73 dilution steps more sensitive than men at all ages. Distribution of corrected PTC scores showed significant bimodality with a cut-off value of 4.5 that separates the group of tasters from the group of non-tasters, with estimated means of 2.19 (SD = 1.12) and 7.99 (SD = 1.44), respectively.

Prodi, Chem. Senses, 2004

First PTC Family Study

	Number of Familie	Children	
		Can taste	Can not taste
Both parents can taste	40	90	16
One parent can taste, the other can n	51	80	37
Neither parent can taste	9	0	17

L. H. Snyder Science 1931

Types of Genetic Studies

Family Studies

Compare trait values across family members

Linkage Analysis

Compare trait values with inheritance patterns

Association

Compare trait values against genetic variants

Family Studies

Familial Relationships

- Twins
- Siblings
- Parents/offspring

Phenotype information

- Affected/Unaffected (Prostate Cancer)
- Quantitative measure (Blood Pressure)

No Genotype information required for some things (e.g., lecture 2)

Why do Family Studies?

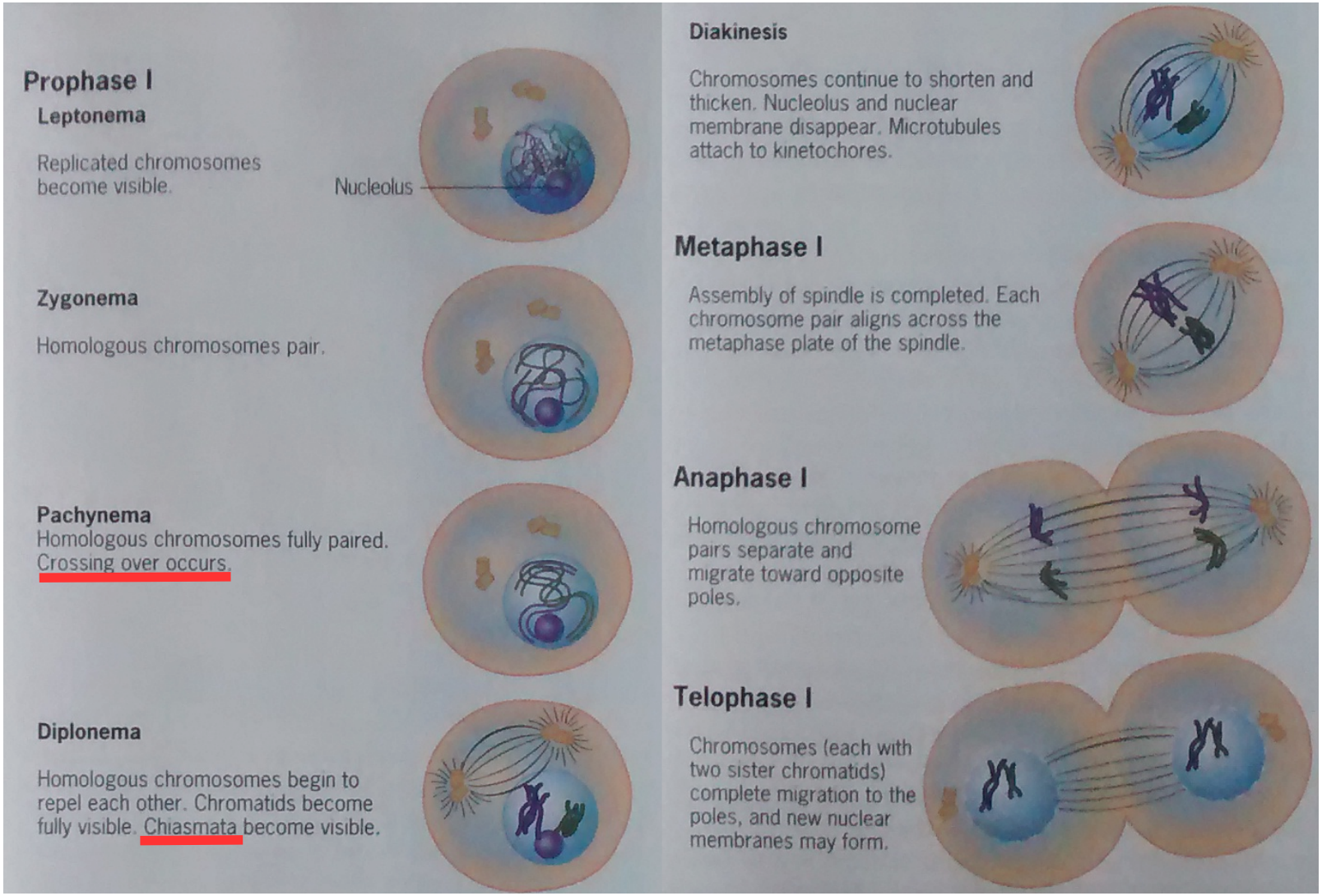
Is the trait genetic?

What is the mode
of transmission?

- Dominant
- Recessive
- Additive
- Polygenic (Multiple genes involved)

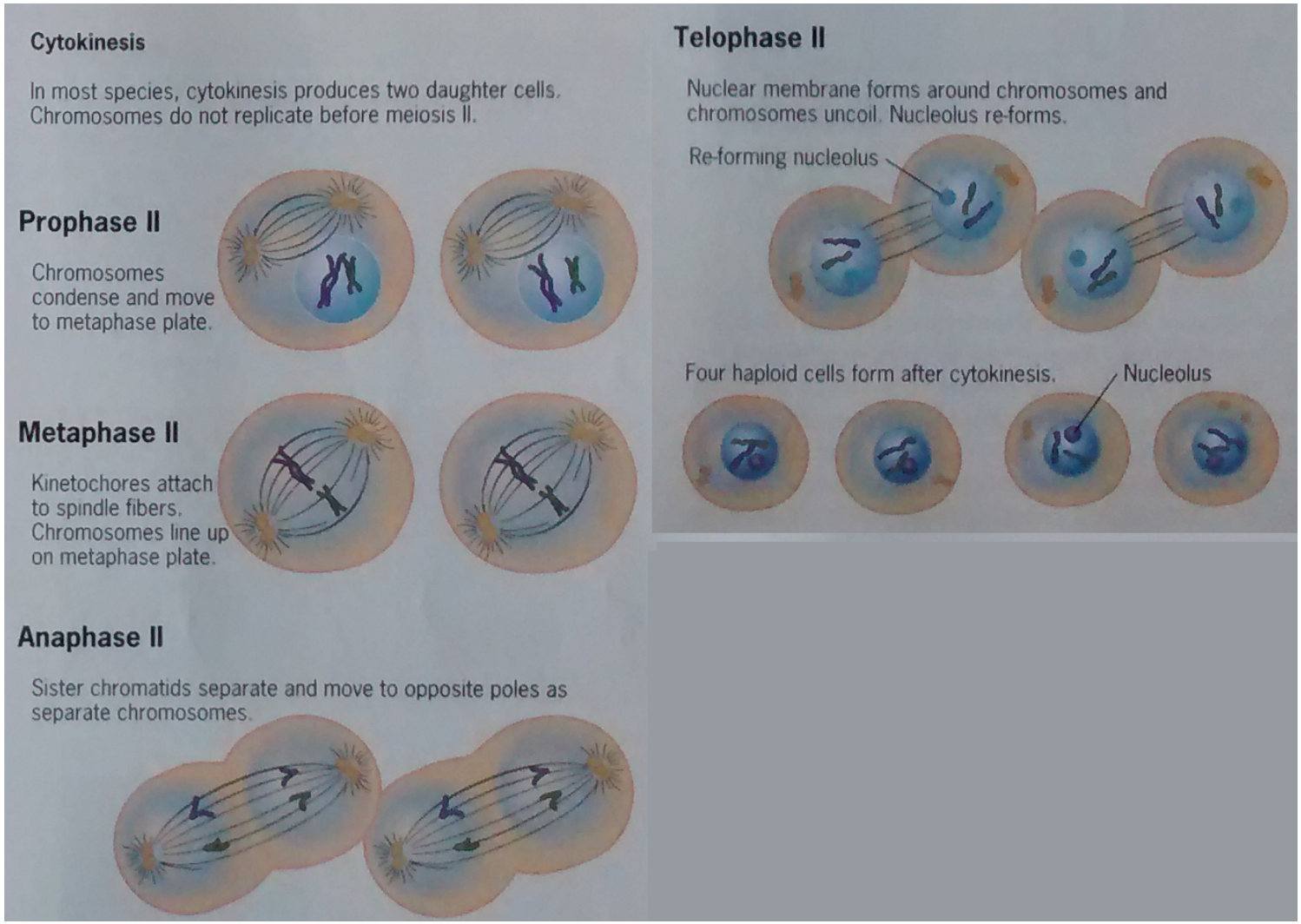
https://belkazan.files.wordpress.com/2012/04/stripe-shirts-mother-father-child-at-the-beach-boardwalk.png?w=620

Meiosis – Crossing over



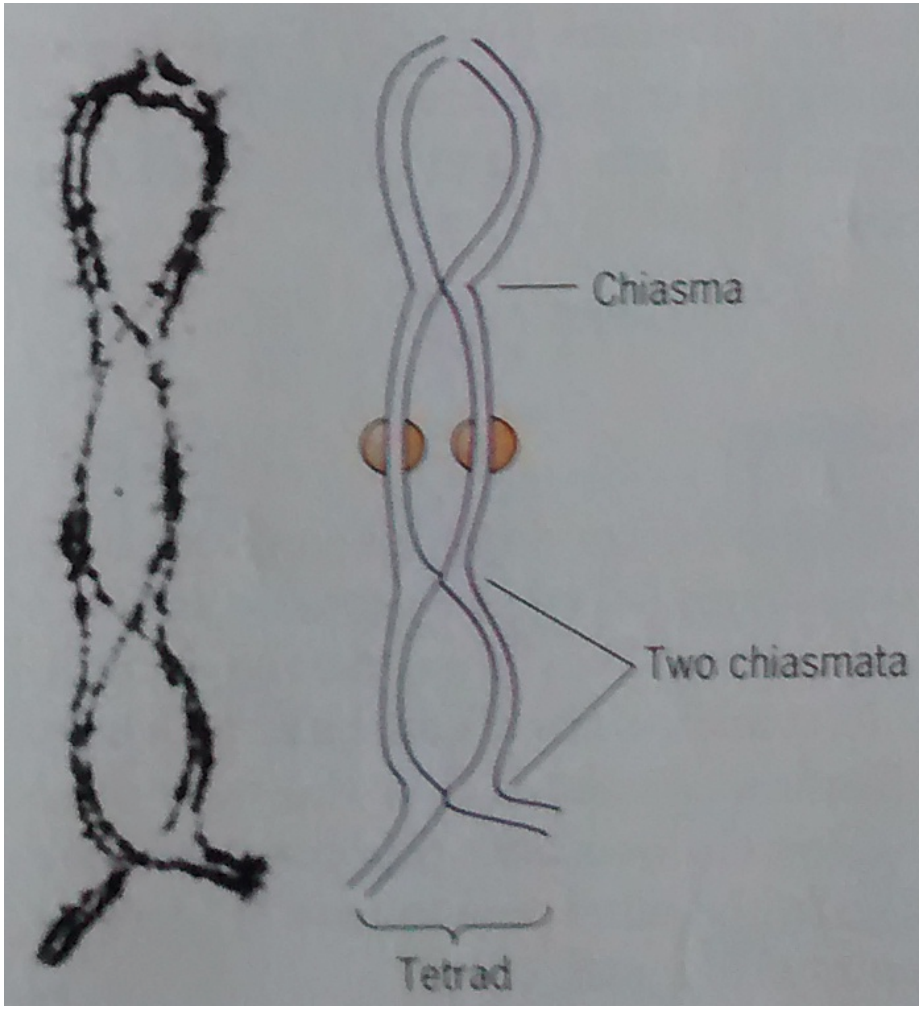
Snustad & Simmons, Principals of Genetics

Meiosis – Crossing over



Snustad & Simmons, Principals of Genetics

Crossing over – Chiasmata



Snustad & Simmons, Principals of Genetics

Idea behind linkage

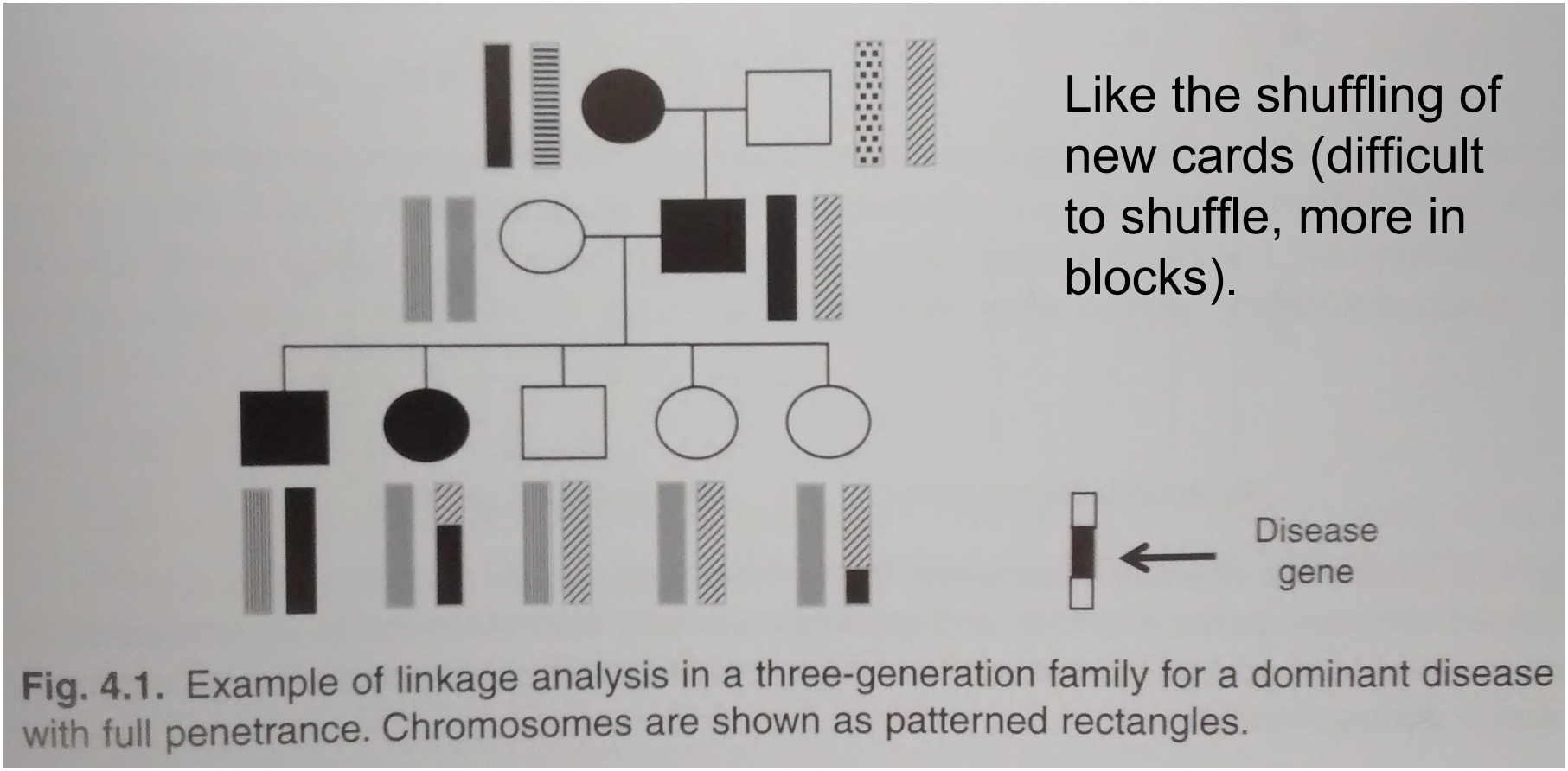
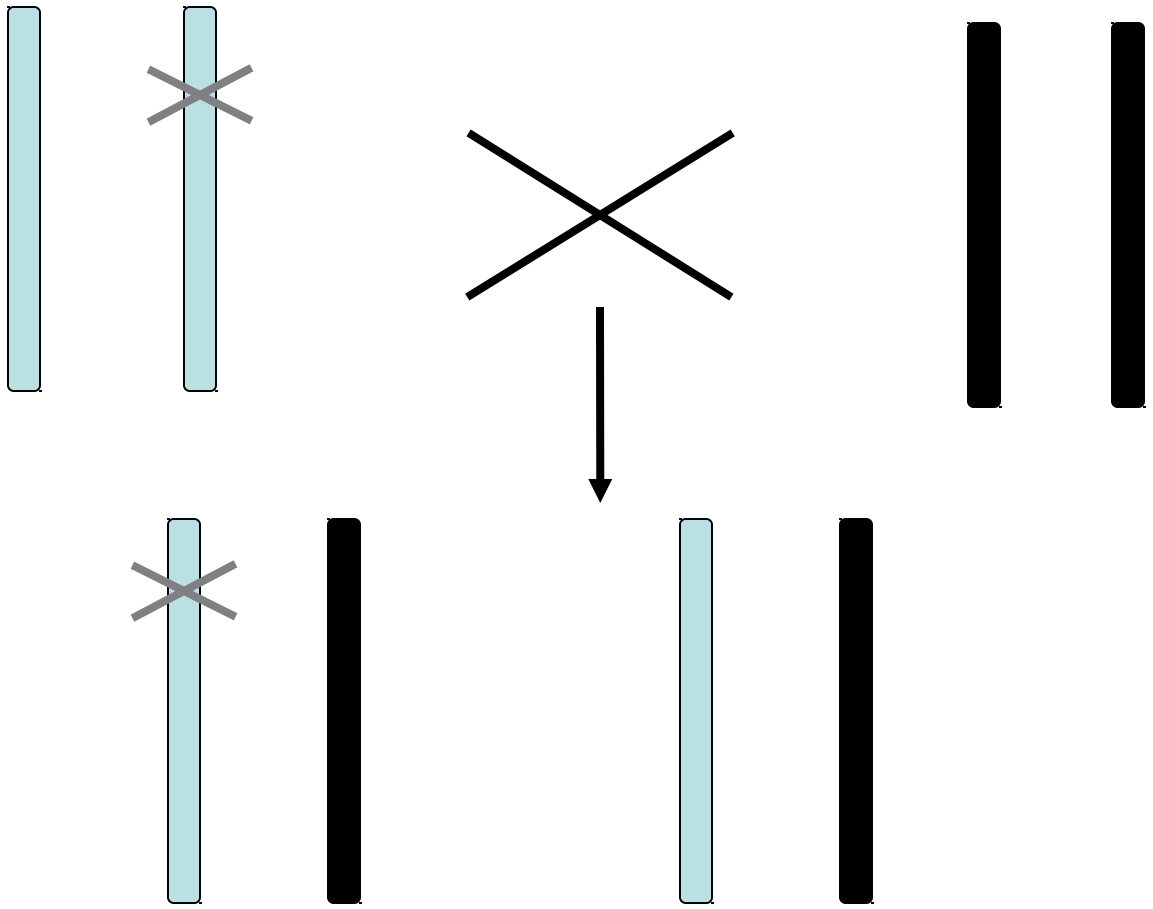


Fig. 4.1. Example of linkage analysis in a three-generation family for a dominant disease with full penetrance. Chromosomes are shown as patterned rectangles.

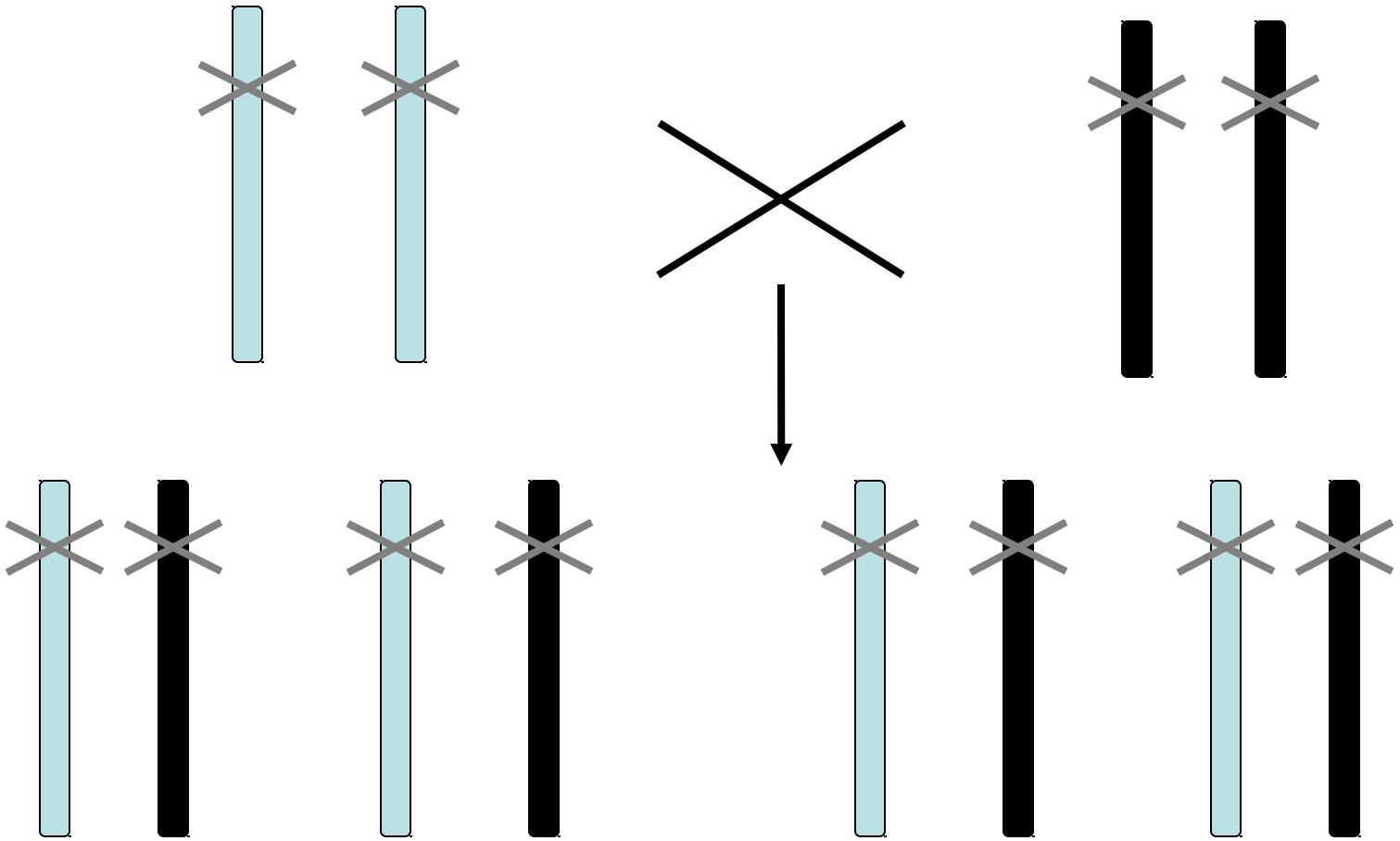
Austin, Genetic Epidemiology

Recall: Mutation and Meiosis



Recall: E.g., Recessive trait

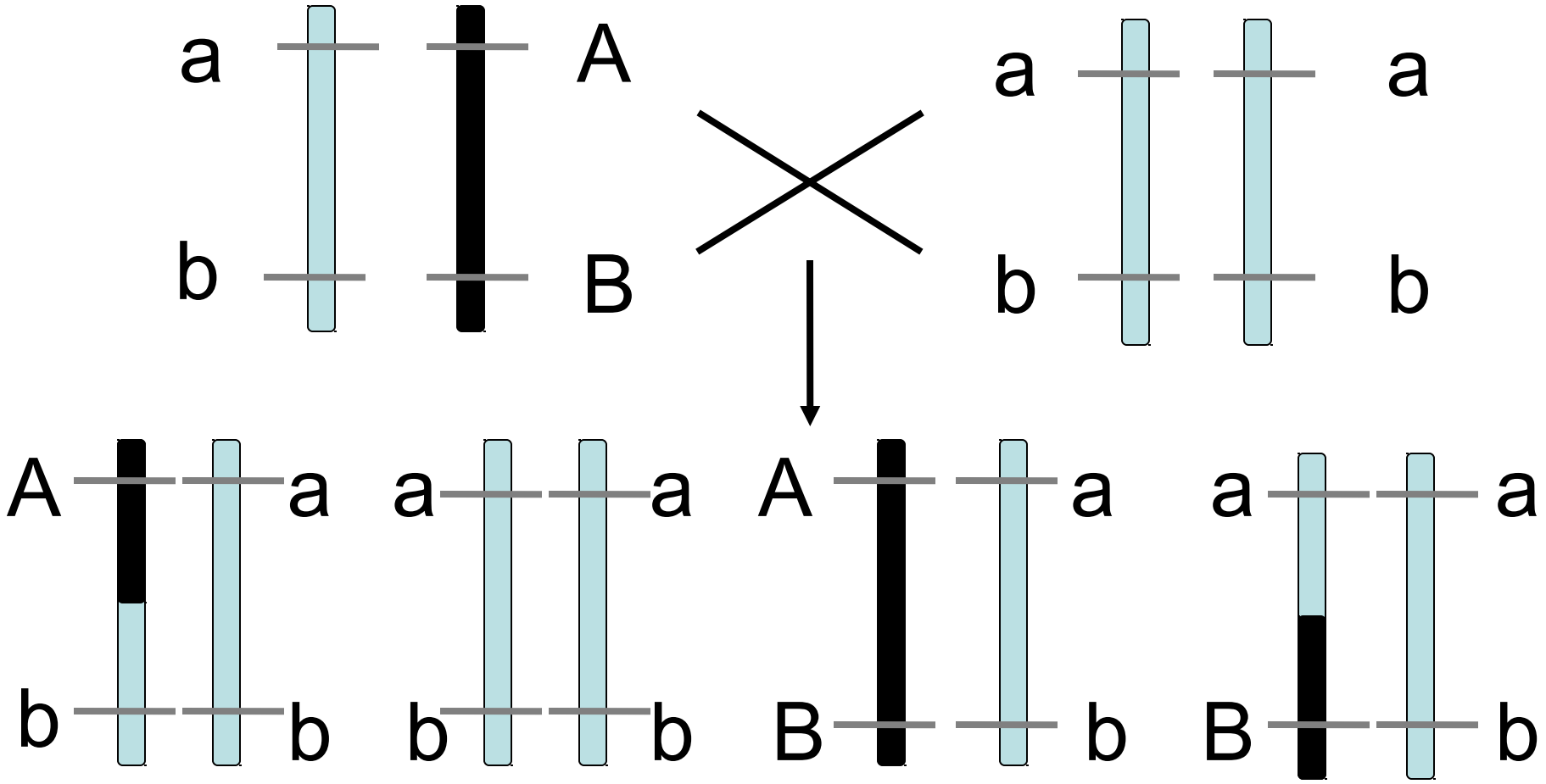
(Know the trait, know the genotype at the disease loci)



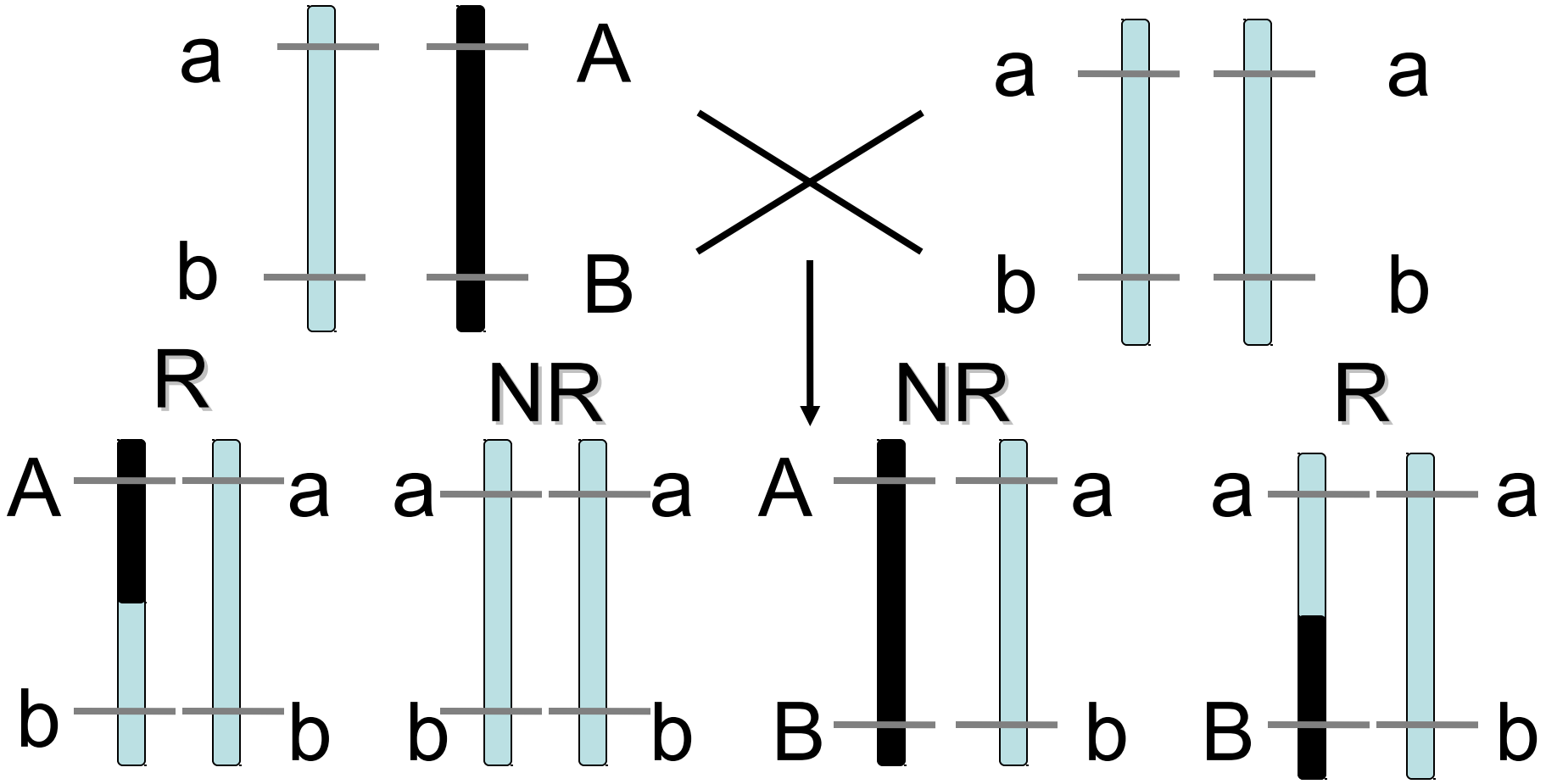
Linkage Analysis

- Narrow down position of disease gene
- No biological knowledge needed
- Genetic markers (not disease gene)
- Recombination

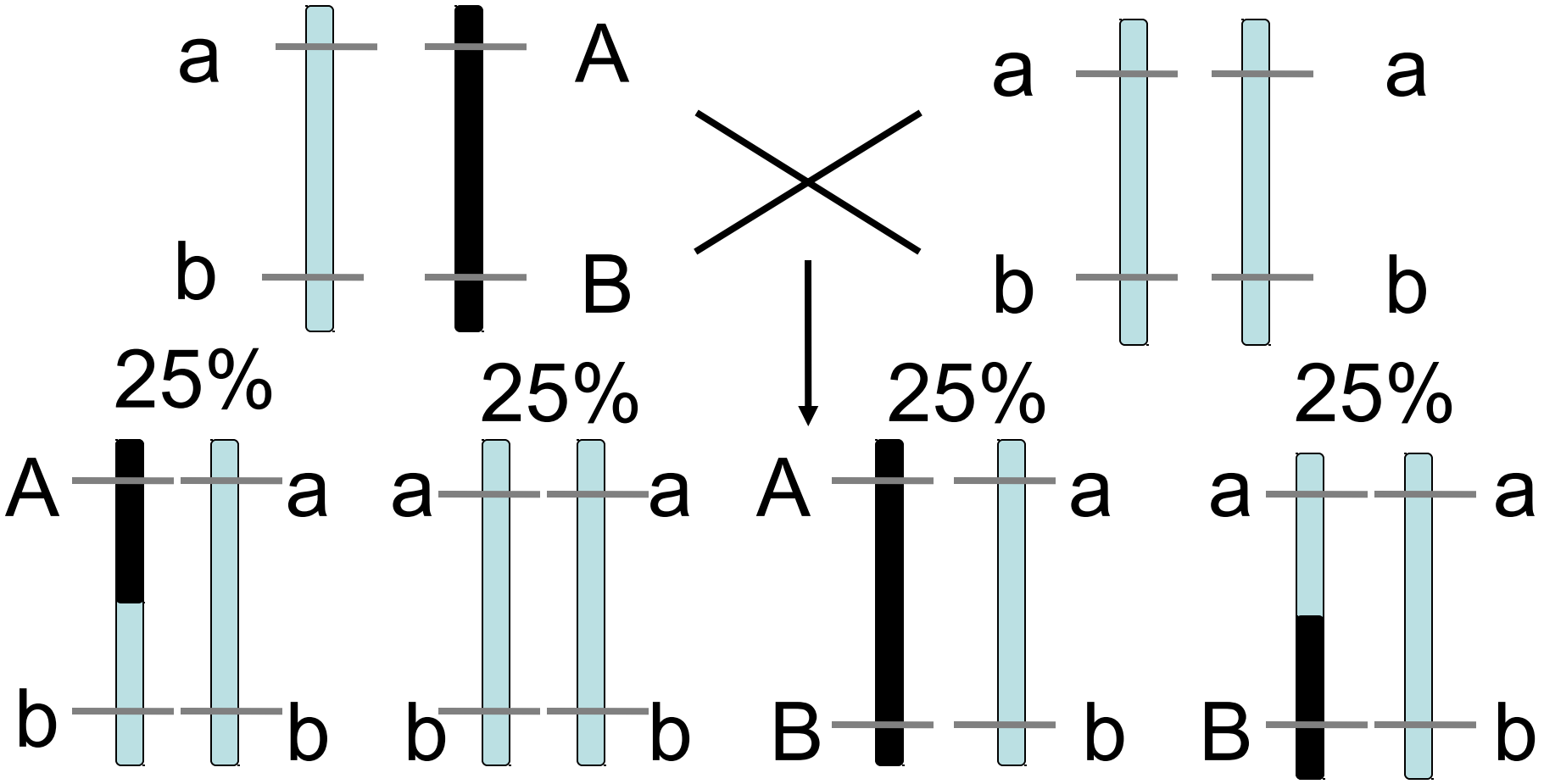
Recombination – was there recombination?



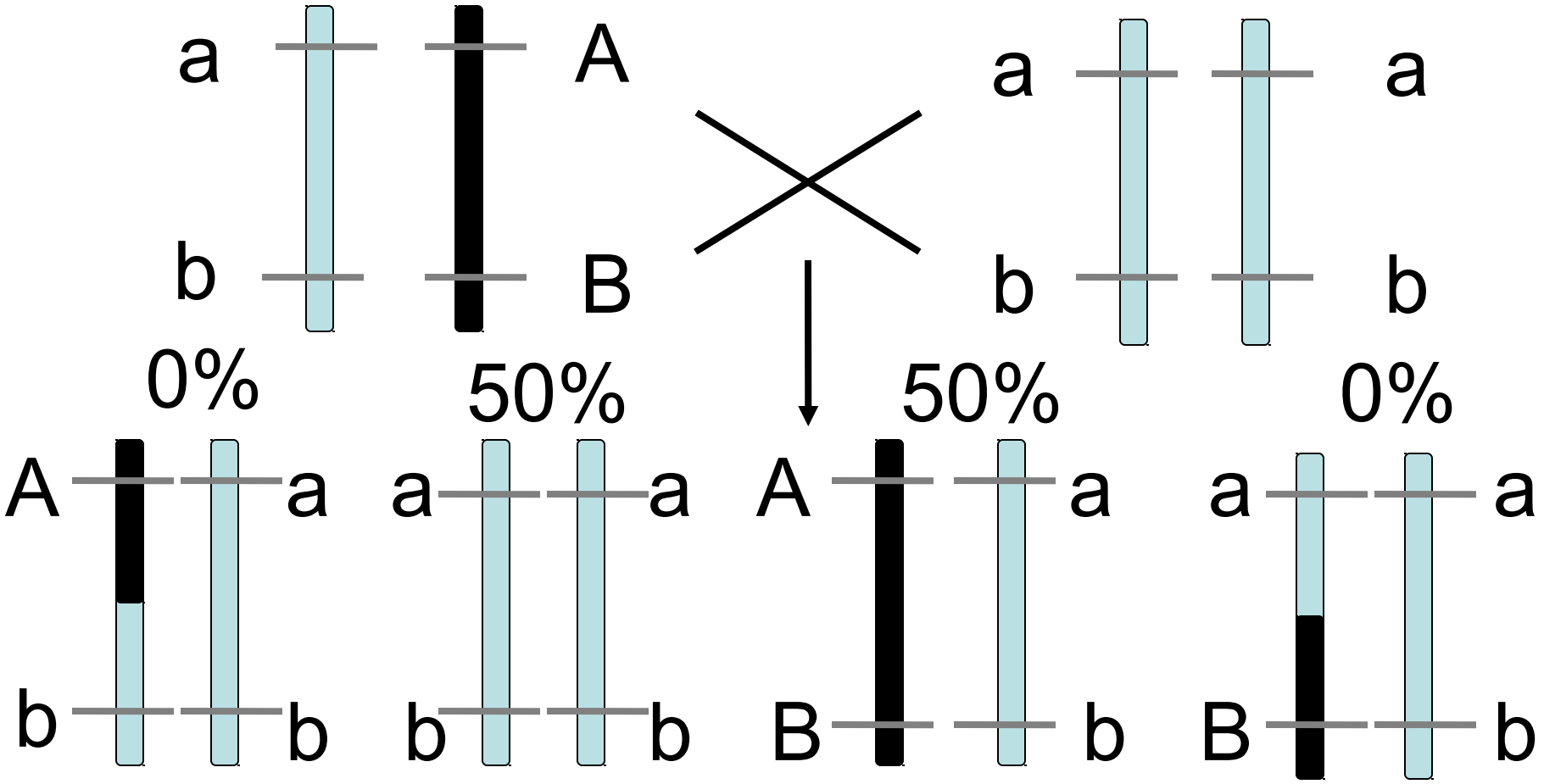
Recombination– was there recombination?



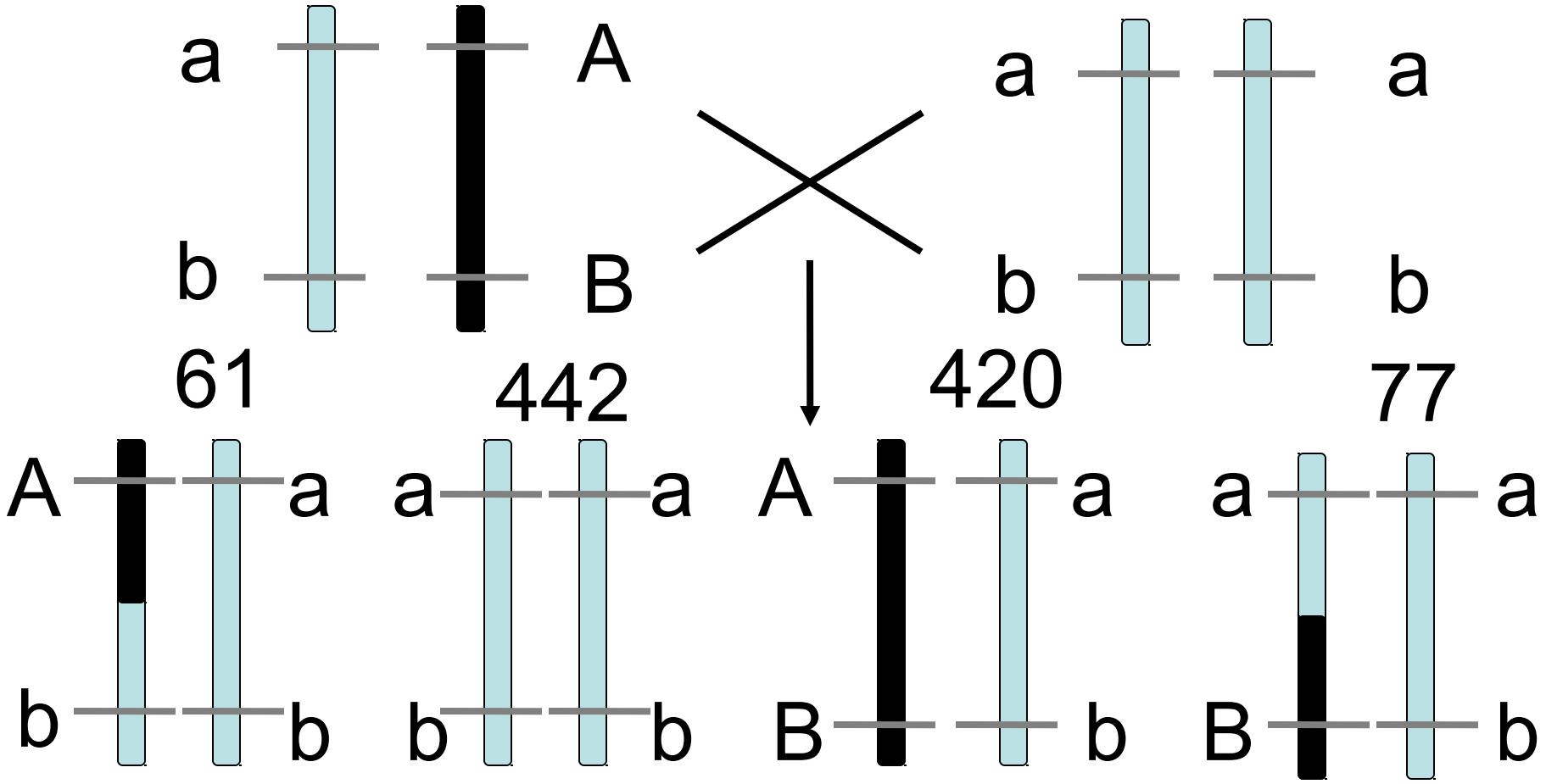
Independent Assortment



No recombination



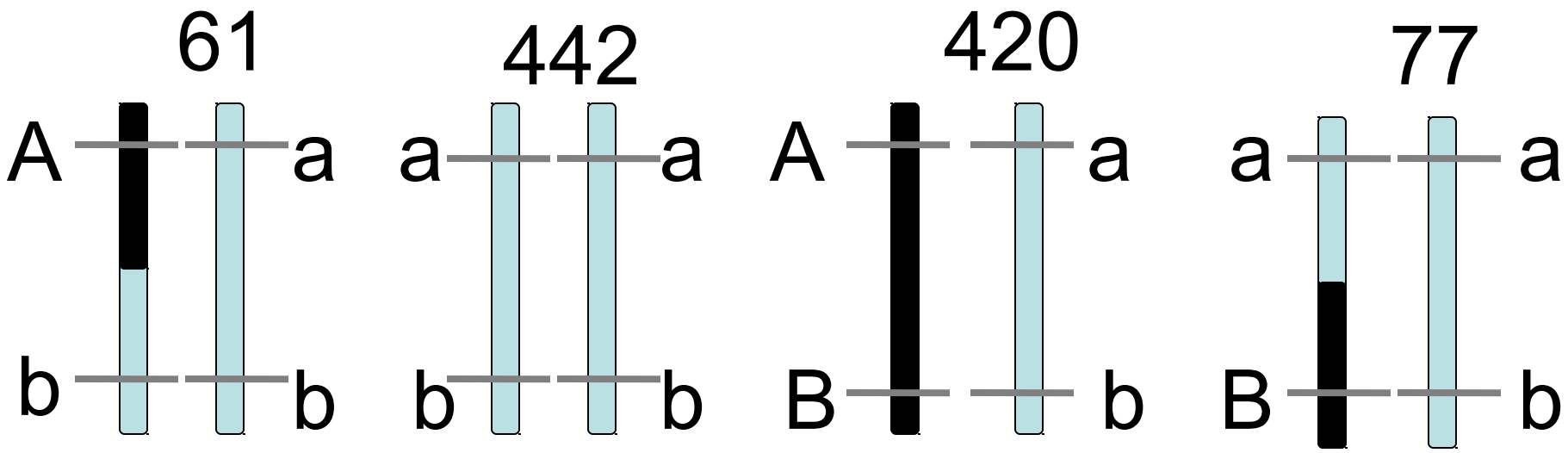
Recombination Fraction



Recombination Fraction

$\theta =$ Pr(observing a recombinant gene) =

{# Recombinants} / Total =



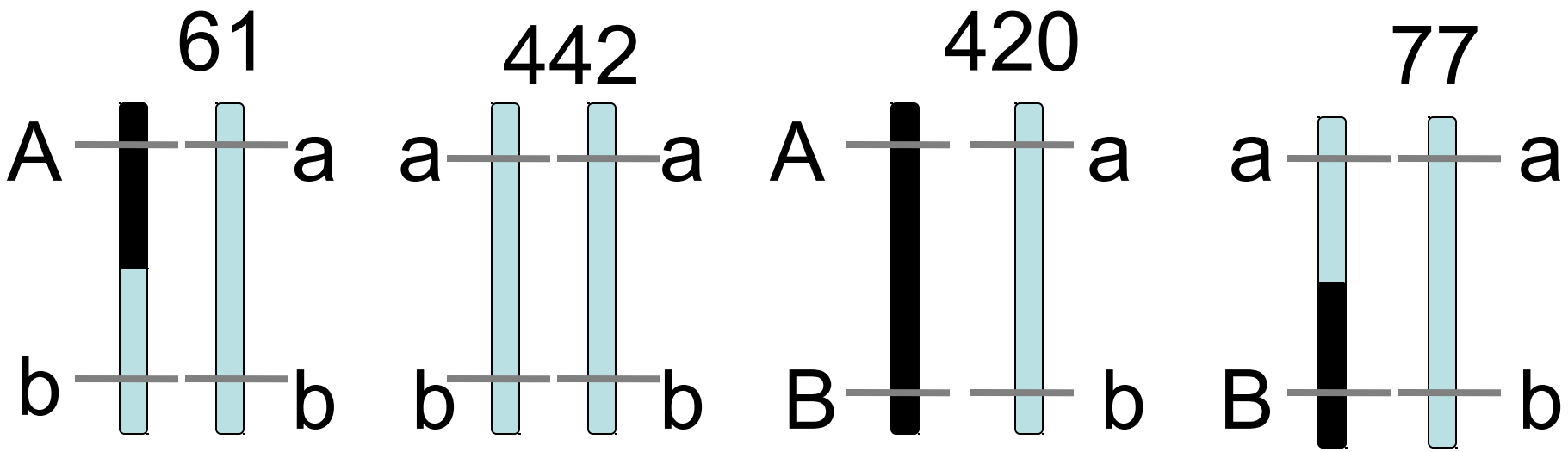
Recombination Fraction

$\theta =$ Pr(observing a recombinant gene) =

{# Recombinants} / Total =

(61+77)/(61+77+442+420)=138/1000

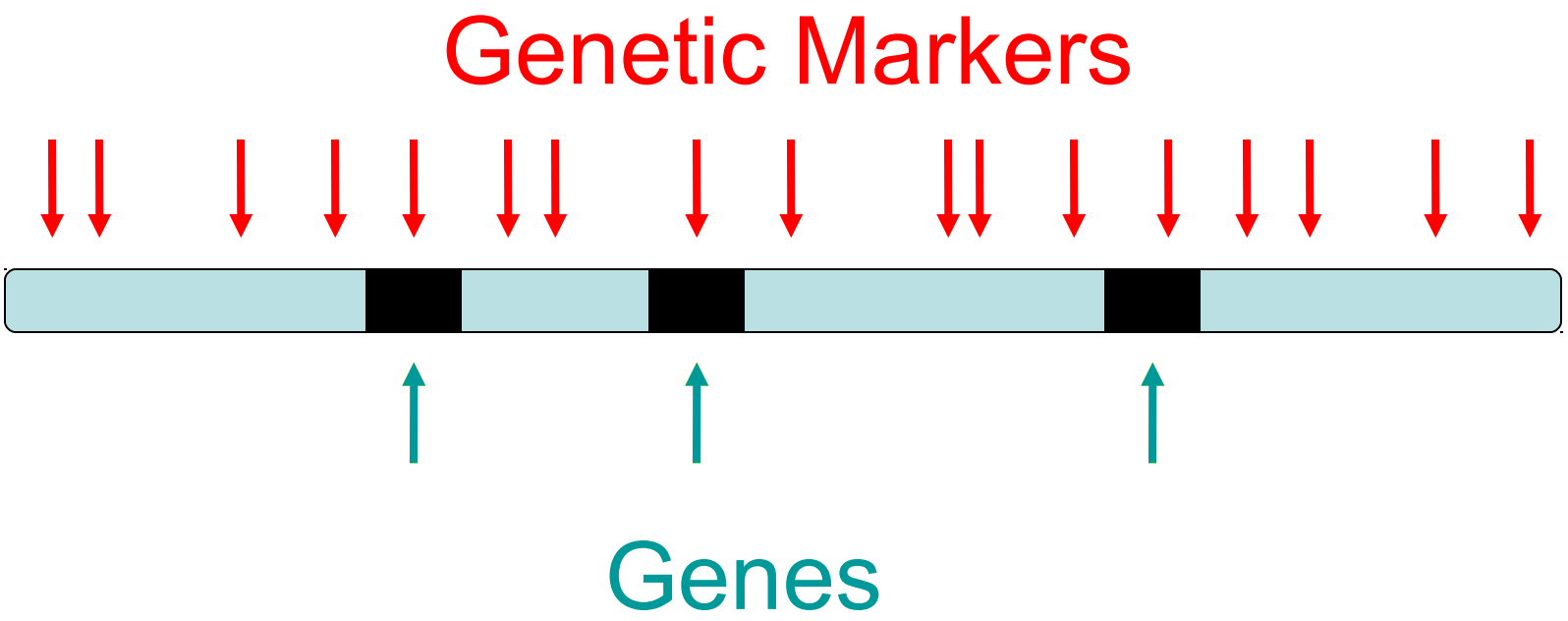
= 13.8% < 50%, indicating some linkage



Human Linkage Analysis

- RFLP Markers for Linkage (1980)
- Huntington's Disease Linkage (1983)
- Cystic Fibrosis Linkage (1985)
- Cystic Fibrosis Gene (1989)
- Huntington's Disease Gene (1993)

Genome-wide linkage analysis



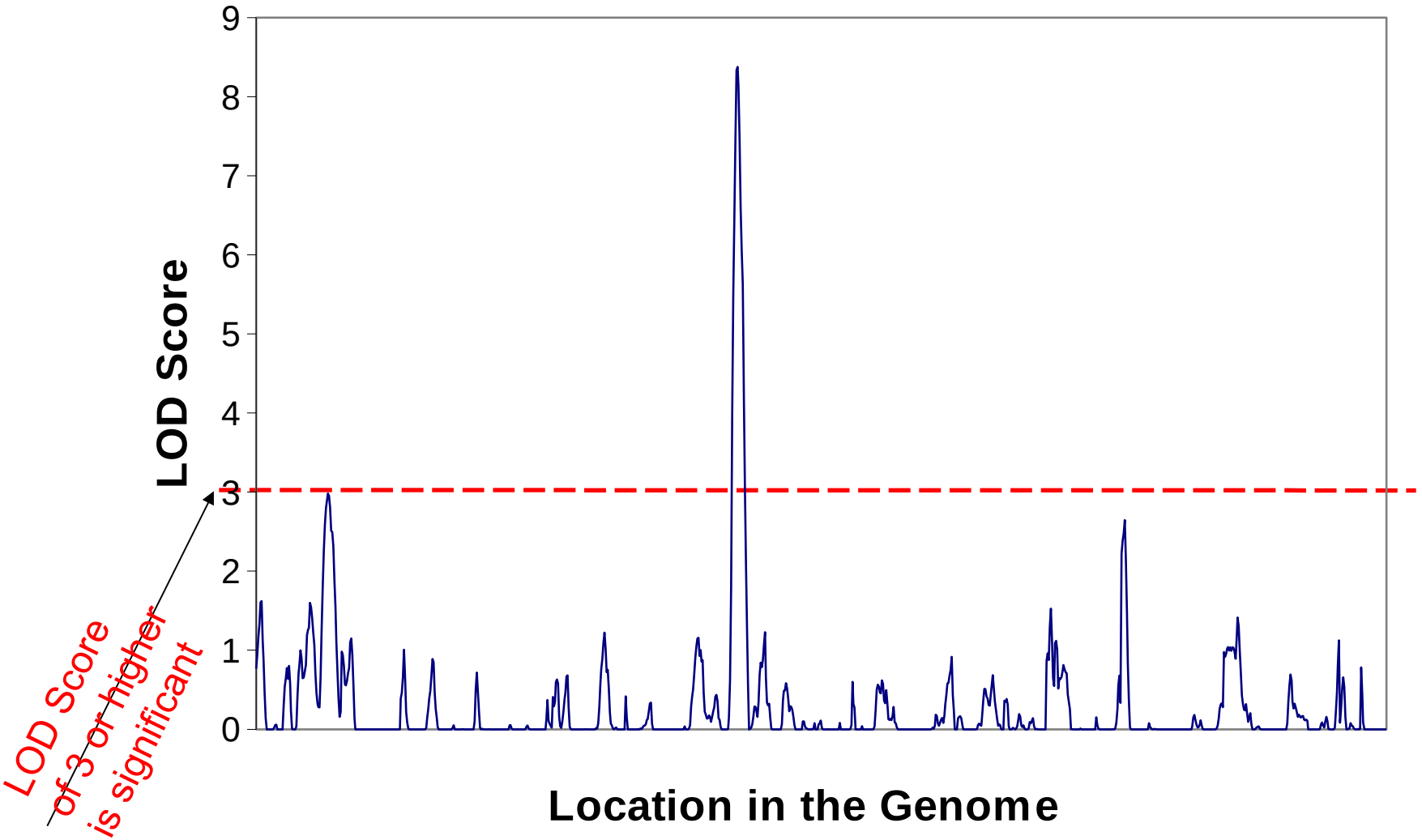
Linkage Analysis

LOD score based on recombination

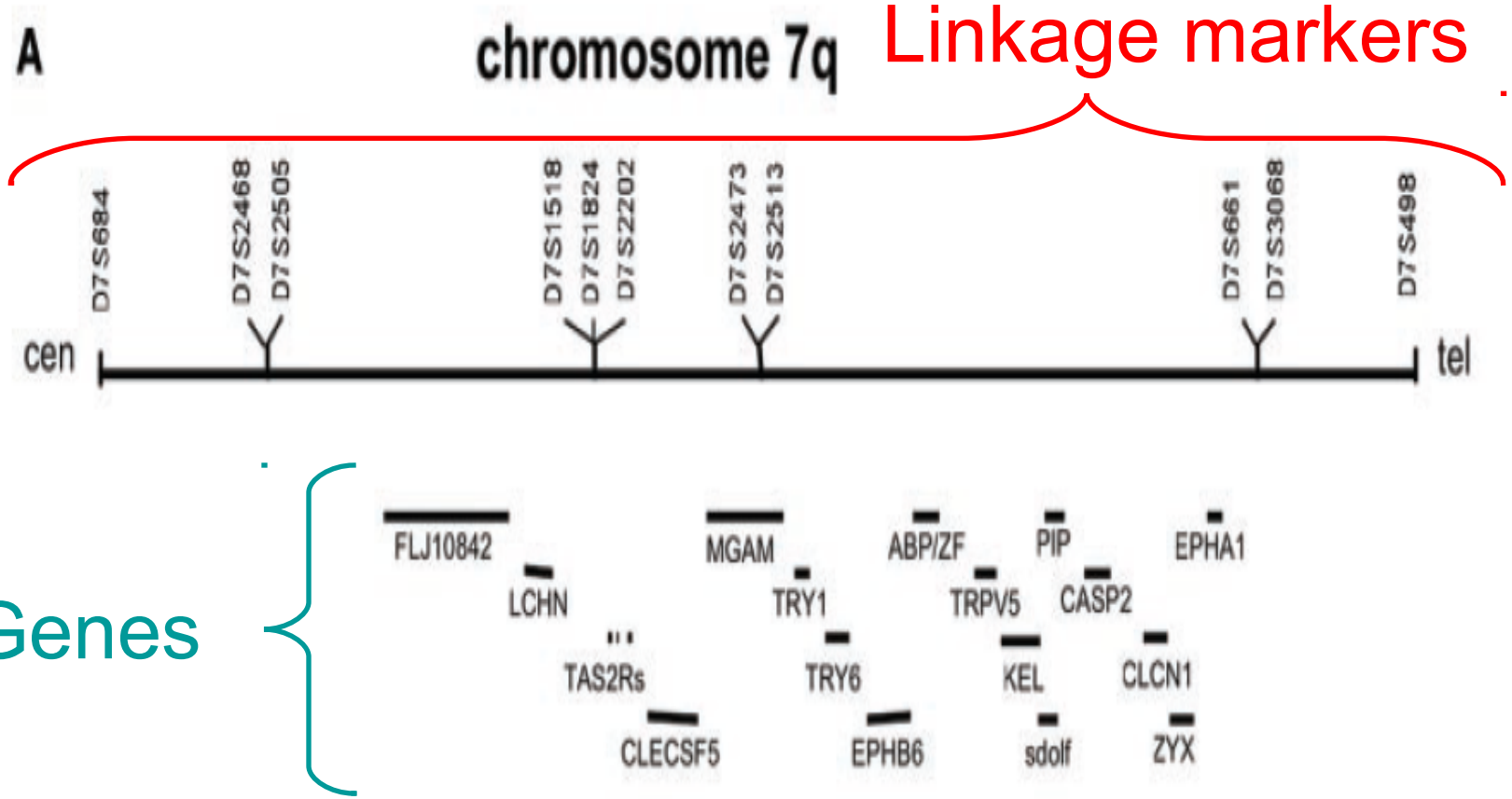
$$\text{LOD}(\theta) = \log \left(\frac{(\theta)^{R} \; (1 - \theta)^{NR}}{(\theta_{null} = 1/2)^{R + NR}} \right)$$

(Likelihood ratio test)

PTC Linkage Analysis



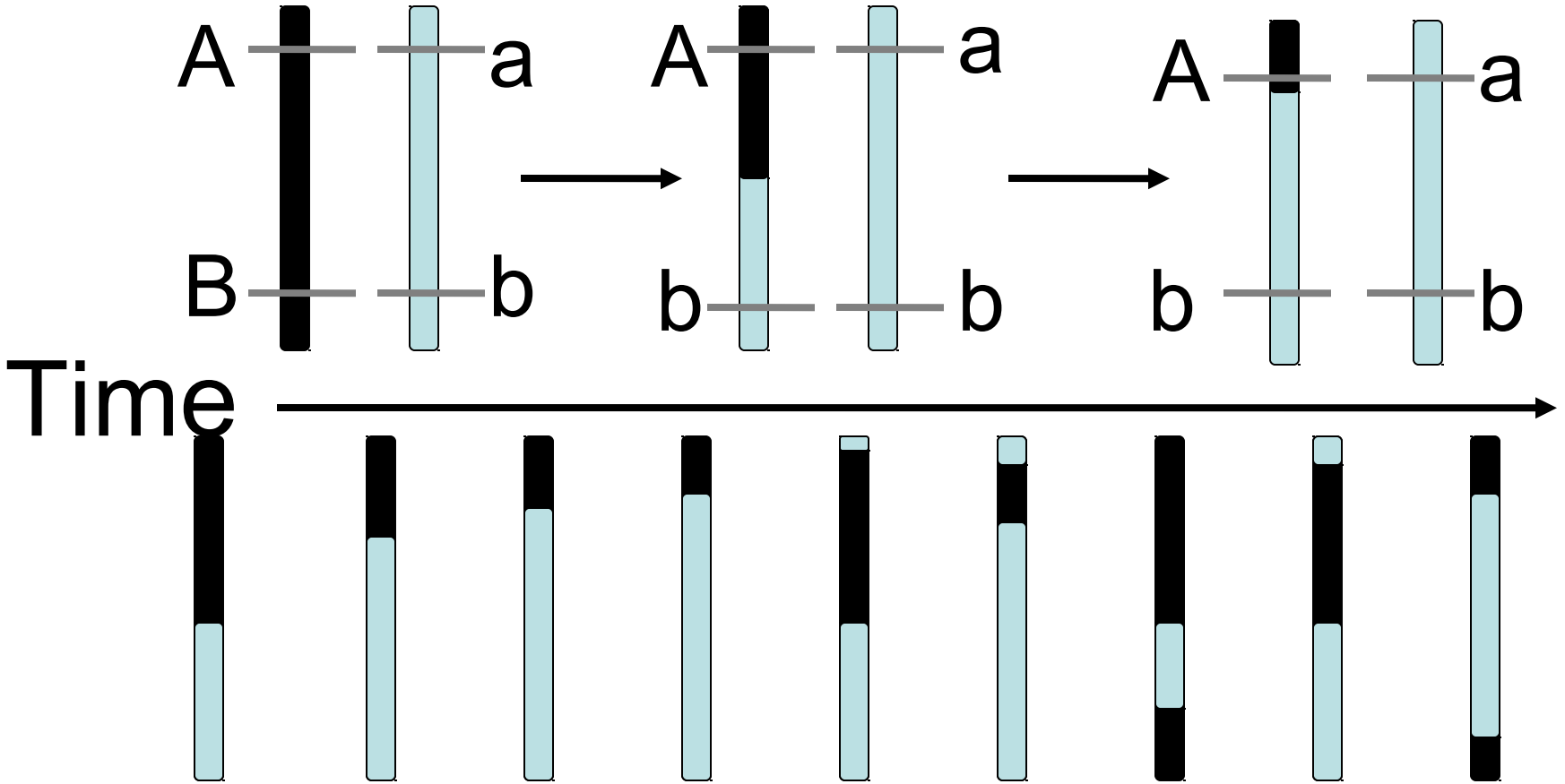
After you have found a linkage peak: Fine Mapping (more details later)



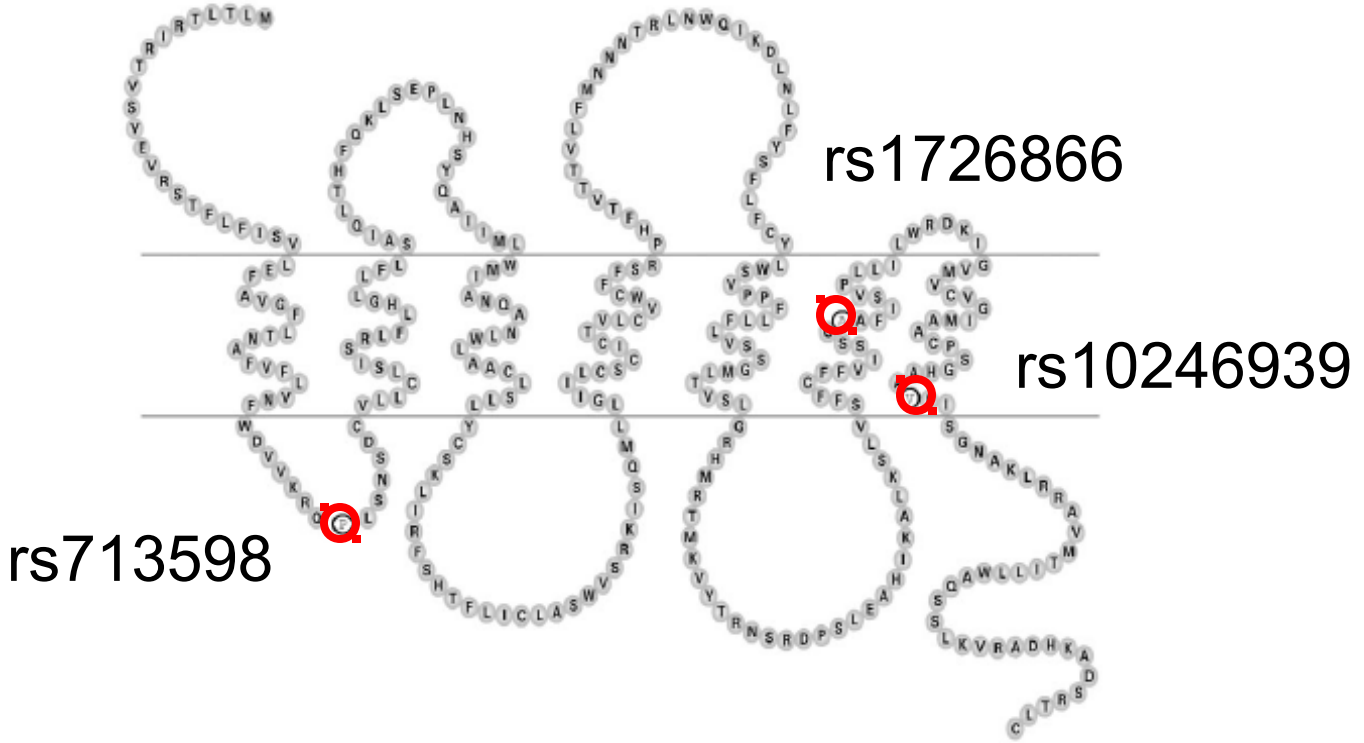
Kim et al. Science 2003

Key Idea: Linkage and Linkage Disequilibrium

Linkage Disequilibrium blocks (LD) are much finer resolution than just linkage (basically the effect of linkage over time, over many more generations) [Shuffle cards more and more...]

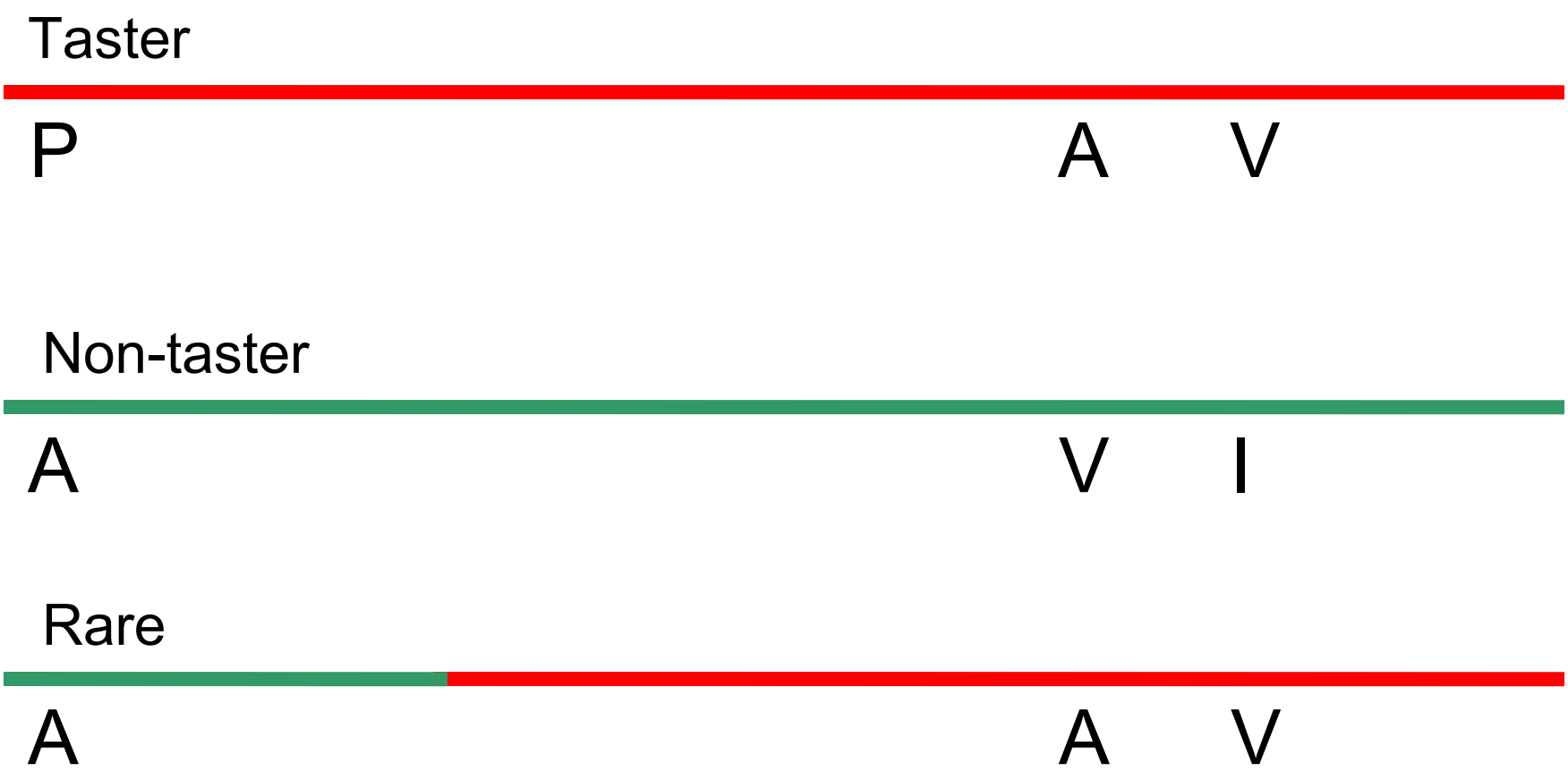


TAS2R38 Receptor Structure



Kim et al. J Dent Res 2004

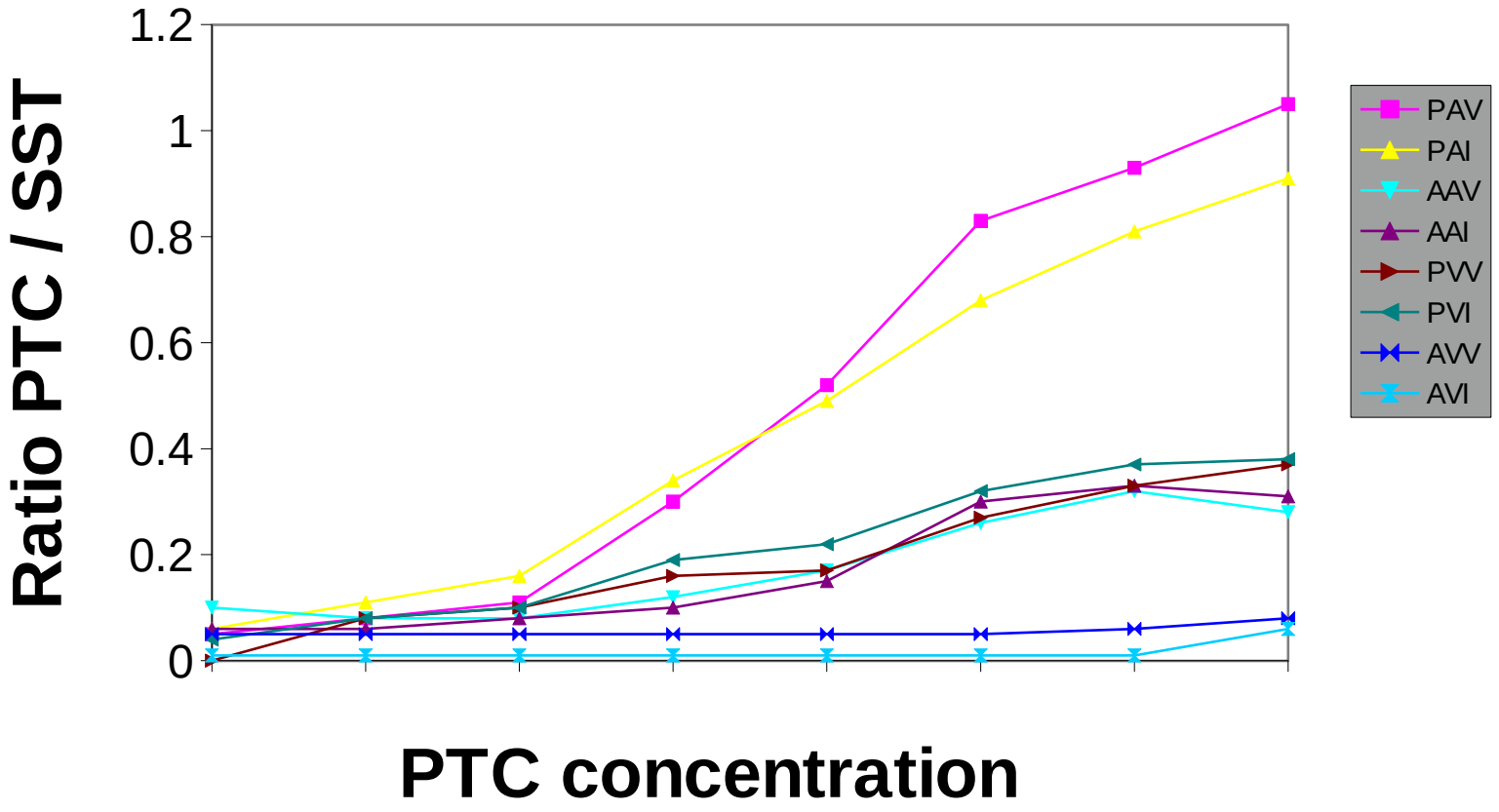
3 SNPs form 3 Haplotypes



Predicted Effect of the 3 SNPs

Amino Acid Substitution	PAM250
P --> A	1
A --> V	0
V --> I	4

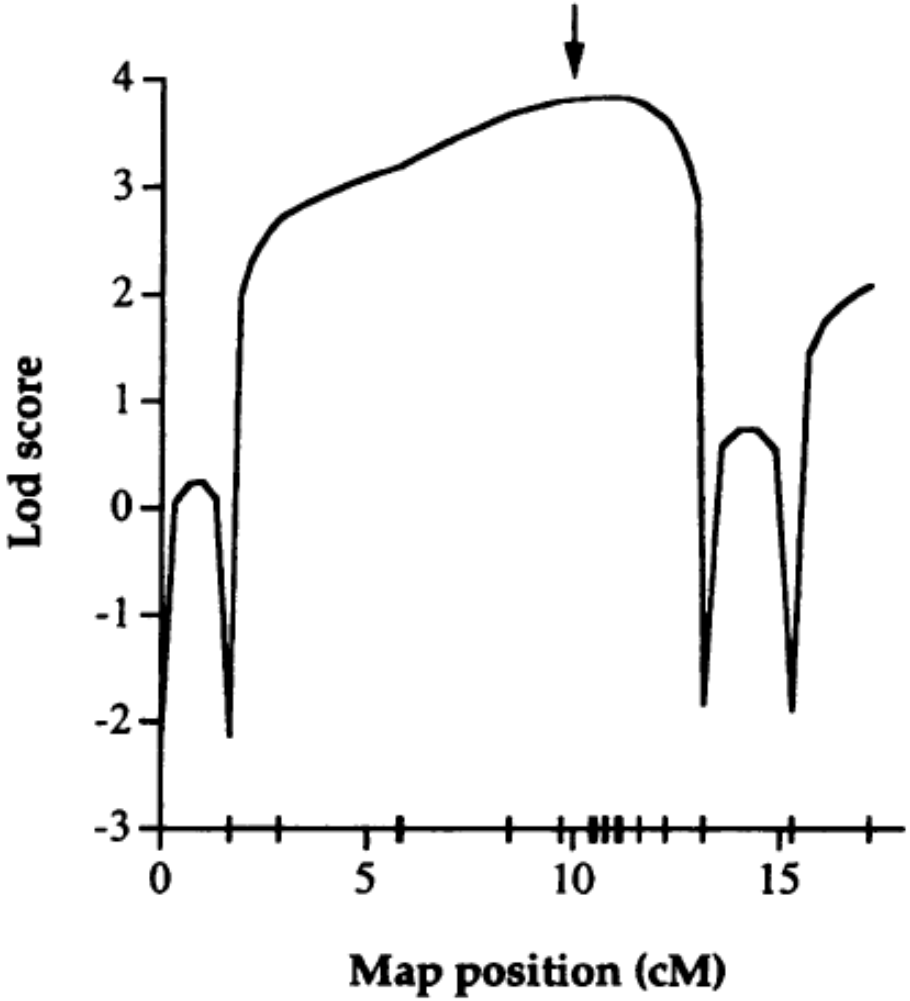
TAS2R38 Haplotype function in vitro



Exome sequencing and examples in Austin book following up on linkage peaks:

- BRCA2 Breast cancer
- HOXB14 G84E
  - Took so long to find because they were initially looking in the wrong place!

Multipoint Linkage Analysis



- Can base LOD score on multiple surrounding markers, same basic idea

- Austin refers to other articles
  - Khoury et al. (1993)
  - Ott (1999)

Non-parametric linkage analysis

- Large multi-generational families most powerful for parametric linkage analysis
  - Difficult to ascertain!!!
- Non-parametric tests available for >= 2 affected siblings
- Instead of testing θ directly, test excessive sharing identity by descent (IBD)
  - Always share exactly one allele IBD with parent, and so on

Outline

- Review last weeks homeworks
- Linkage analysis, with PTC testing example
  - Background/Review
  - Linkage analysis
- **Family-based association tests**
  - **Overview**
  - **Trios/TDT**
  - **Discordant sibships**

Moving from linkage to association...

- Linkage: Estimating & Testing recombination rates θ
- Association:
  - Difference in genotype frequencies (next lecture)
  - Difference in genotype transmission frequency (discuss next)
- Family based tests next based on composite hypothesis: No linkage, no association
- LD block much smaller than linkage, so much finer resolution (recall slide earlier over time...)
- *Much smaller* family structures than for linkage!

Mendel's laws

- Recall the playing cards example...
- One allele from each parent for each gene
  - Many family based tests based on this, rather than estimating allele frequencies (case-control)

Mendelian transmission: Ex

- E.g., parents are Aa, Aa:
  - P(offspring=AA | Mother=Aa,Father=Aa)=?
  - P(offspring=Aa | Mother=Aa,Father=Aa)=?
  - P(offpsring=aa | Mother=Aa, Father=Aa)=?

M\F	A	a
A	AA	Aa
a	Aa	aa

Mendelian transmission: Ex

- E.g., parents are Aa, Aa:
  - P(offspring=AA | Mother=Aa,Father=Aa)=1/4
  - P(offspring=Aa | Mother=Aa,Father=Aa)=1/2
  - P(offpsring=aa | Mother=Aa, Father=Aa)=1/4

M\F	A	a
A	AA	Aa
a	Aa	aa

Conditioning on parents...

Mendelian transmission: Ex

- E.g., parents are AA, Aa:
  - P(offspring=AA | Mother=AA,Father=Aa)=?
  - P(offspring=Aa | Mother=AA,Father=Aa)=?
  - P(offpsring=aa | Mother=AA, Father=Aa)=?

M\F	A	a
A	AA	Aa
A	AA	Aa

Mendelian transmission: Ex

- E.g., parents are AA, Aa:
  - P(offspring=AA | Mother=AA,Father=Aa)=1/2
  - P(offspring=Aa | Mother=AA,Father=Aa)=1/2
  - P(offpsring=aa | Mother=AA, Father=Aa)=0

M\F	A	a
A	AA	Aa
A	AA	Aa

Mendelian transmission: Ex

- E.g., parents are AA, AA:
  - P(offspring=AA | Mother=AA,Father=AA)=?
  - P(offspring=Aa | Mother=AA,Father=AA)=?
  - P(offpsring=aa | Mother=AA, Father=AA)=?

M\F	A	A
A	AA	AA
A	AA	AA

Mendelian transmission: Ex

- E.g., parents are AA, AA:
  - P(offspring=AA | Mother=AA,Father=AA)=1
  - P(offspring=Aa | Mother=AA,Father=AA)=0
  - P(offpsring=aa | Mother=AA, Father=AA)=0

M\F	A	A
A	AA	AA
A	AA	AA

Homozygote parents are "non-informative" (no variation in offspring's conditional genotype distribution)

Trios: Transmission Disequilibrium Test (TDT)

- Test based on transmissions from parents to offspring
- Assumptions:
  - Parents' and offspring genotypes are known
  - Dichotomous phenotype (thought Q-TDT), only affected offspring
- Count transmissions from heterozygote parents, and compare to expected transmissions
  - Mendel's laws of seqregation (previous slides) not control group
  - Test for over/under-transmission of alleles in cases (intuition)
- Conditional test
  - Offspring affection status
  - Parental genotypes (conditions out allele frequencies, which is what case-control is based on testing)
- Composite null hypothesis: No linkage, no association

Spielmen et al., AJHG 1993

Trios: Transmission Disequilibrium Test (TDT)

		Non-transmitted parental allele	
		A	a
Transmitted parental allele	A	w	x
	a	y	z

- w AA parents (transmit one A, do not transmit other A)
- z aa parents (transmit one a, do not transmit other a)
- x Aa parents that transmit A, do not transmit a
- y Aa parents that transmit a, do not transmit A

Possible Parental Configurations

- **AA-AA**, **AA-Aa**, **AA-aa**, Aa-AA, **Aa-Aa**, **Aa-aa**, aa-AA, aa-Aa, **aa-aa**
  - (Ones not bolded are symmetric for what we will do next, e.g., AA-Aa == Aa-AA
  - Six possible configurations

Both parents homozygous

		Non-transmitted parental allele		AA-AA
		A	a	|
Transmitted parental allele	A	2	0	AA
	a	0	0	

- Offspring genotype is deterministic, no variation, not informative!

Both parents homozygous

		Non-transmitted parental allele		aa-aa
		A	a	|
Transmitted parental allele	A	0	0	aa
	a	0	2	

- Offspring genotype is deterministic, no variation, not informative!

Both parents homozygous

		Non-transmitted parental allele	
		A	a
Transmitted parental allele	A	1	0
	a	0	1

AA-aa
|
Aa

- Offspring genotype is deterministic, no variation, not informative!

One parent heterozygous

		Non-transmitted parental allele	
		A	a
Transmitted parental allele	A	1	1
	a	0	0

AA-Aa

|

AA,Aa

.5 .5 ← Pr

		Non-transmitted parental allele	
		A	a
Transmitted parental allele	A	1	0
	a	1	0

- Variation from one parent

One parent heterozygous

		Non-transmitted parental allele	
		A	a
Transmitted parental allele	A	0	1
	a	0	1

Aa-aa
|
Aa,aa
.5 .5 ← Pr

		Non-transmitted parental allele	
		A	a
Transmitted parental allele	A	0	0
	a	1	1

- Variation from one parent

Both parents heterozygous

		Non-transmitted parental allele	
		A	a
Transmitted parental allele	A	0	2
	a	0	0

Aa-Aa

|

AA,Aa,aa

.25,.5,.25 ← Pr

		Non-transmitted parental allele	
		A	a
Transmitted parental allele	A	0	1
	a	1	0

		Non-transmitted parental allele	
		A	a
Transmitted parental allele	A	0	0
	a	2	1

- Variation from both parents

Trios: Transmission Disequilibrium Test (TDT)

		Non-transmitted parental allele	
		A	a
Transmitted parental allele	A	w	x
	a	y	z

- w AA parents (transmit one A, do not transmit other A)
- z aa parents (transmit one a, do not transmit other a)
- x Aa parents that transmit A, do not transmit a
- y Aa parents that transmit a, do not transmit A

Transmission Disequilibrium Test (TDT)

		Non-transmitted parental allele	
		A	a
Transmitted parental allele	A	w	x
	a	y	z

- No variation in w or z (recall homozygous parents non informative)
- $(x-y)^2/(x+y) \sim \chi_1^2$; it's just special case of McNemar's test
- Think of it as testing are there an excess of the A allele in the affected offspring than would happen by Mendel's laws?

Transmission Disequilibrium Test (TDT)

		Non-transmitted parental allele	
		A	a
Transmitted parental allele	A	?	78
	a	46	?

Insulin Dependent Diabetes Mellitus (IDDM)

- Example from the text: 94 families, 78 parents transmit allele A, 46 transmit allele a
- $(78-46)^2/(78+46)=8.26$, p-value=0.004

Spielman et al., 1993

Limitations of TDT

- Limitations
  - Only affected offspring
  - Only dichotomous phenotypes
  - Bi-allelic markers
  - Additive genetic model
  - No missing parents
  - Can’t do multiple markers, multiple phenotypes
- Extensions exist for most everything you'd want to do, e.g., FBAT, *...(next slide)...*

Discordant sibships

- Conditional logistic regression
  - $P(Y_1=1|Y_1+Y_2=1,g_1,g_2,\ldots)$
  - Matching each sib together, conditions on the fact that *they have discordant phenotypes*
  - Standard model for disease as in logistic regression, just matching based on family strata
    - Interpret coefficients, etc., in the same way
- You will go through an example in the homework

Summary

- Family-based studies can be robust to population substructure (based on Mendelian transmissions, not population allele frequencies)
- Robust to HWE failure
- **Powerful for very rare, highly penetrant traits**