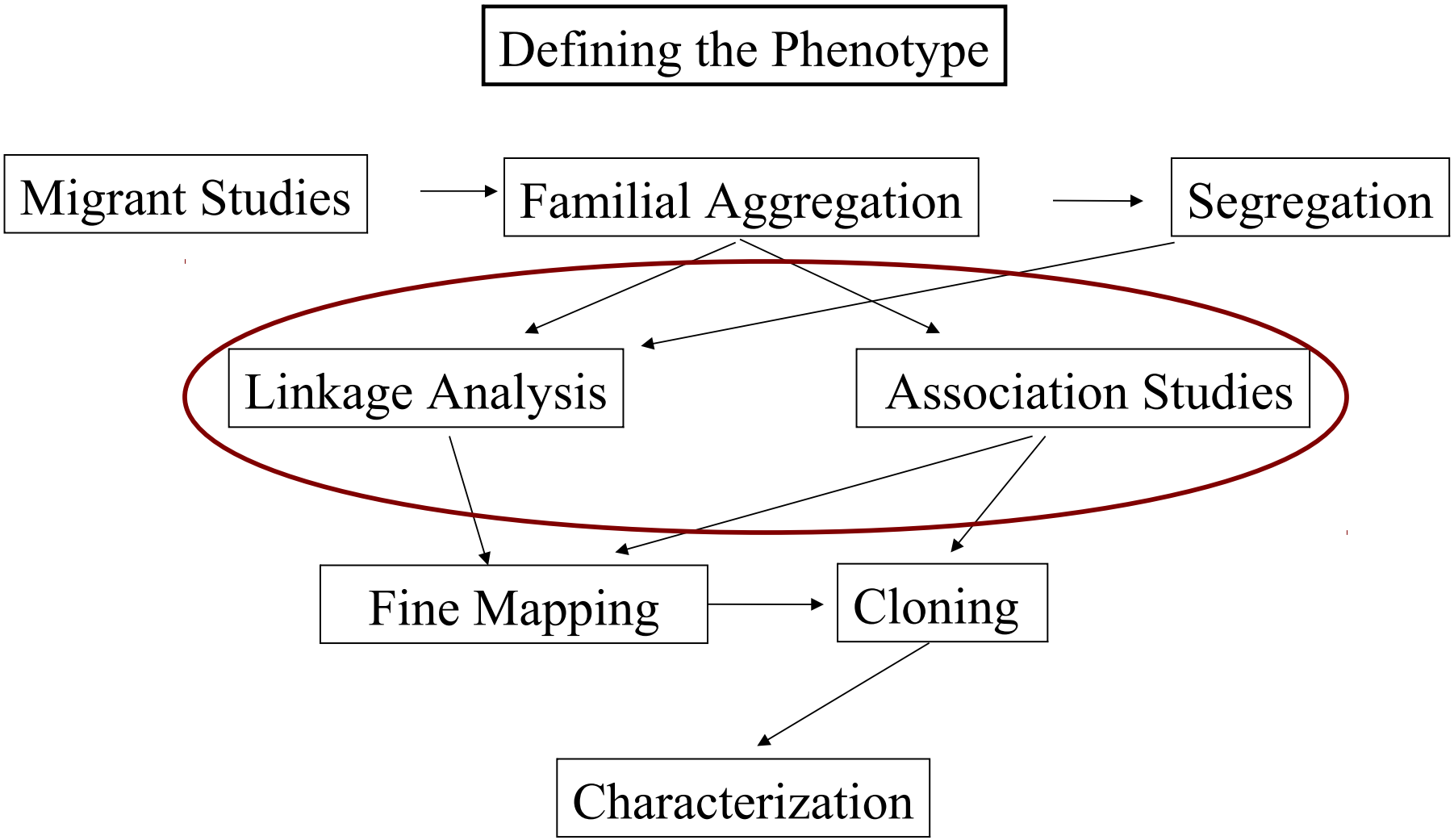
Family-Based Studies

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

Outline

- Review last weeks homeworks
- **Last few topics from previous pop gen lecture**
  - **(Had just finished HWE game)**
  - **Population stratification**
- Linkage analysis, with PTC testing example
- Family-based association tests
  - Overview
  - Trios/TDT
  - Discordant sibships

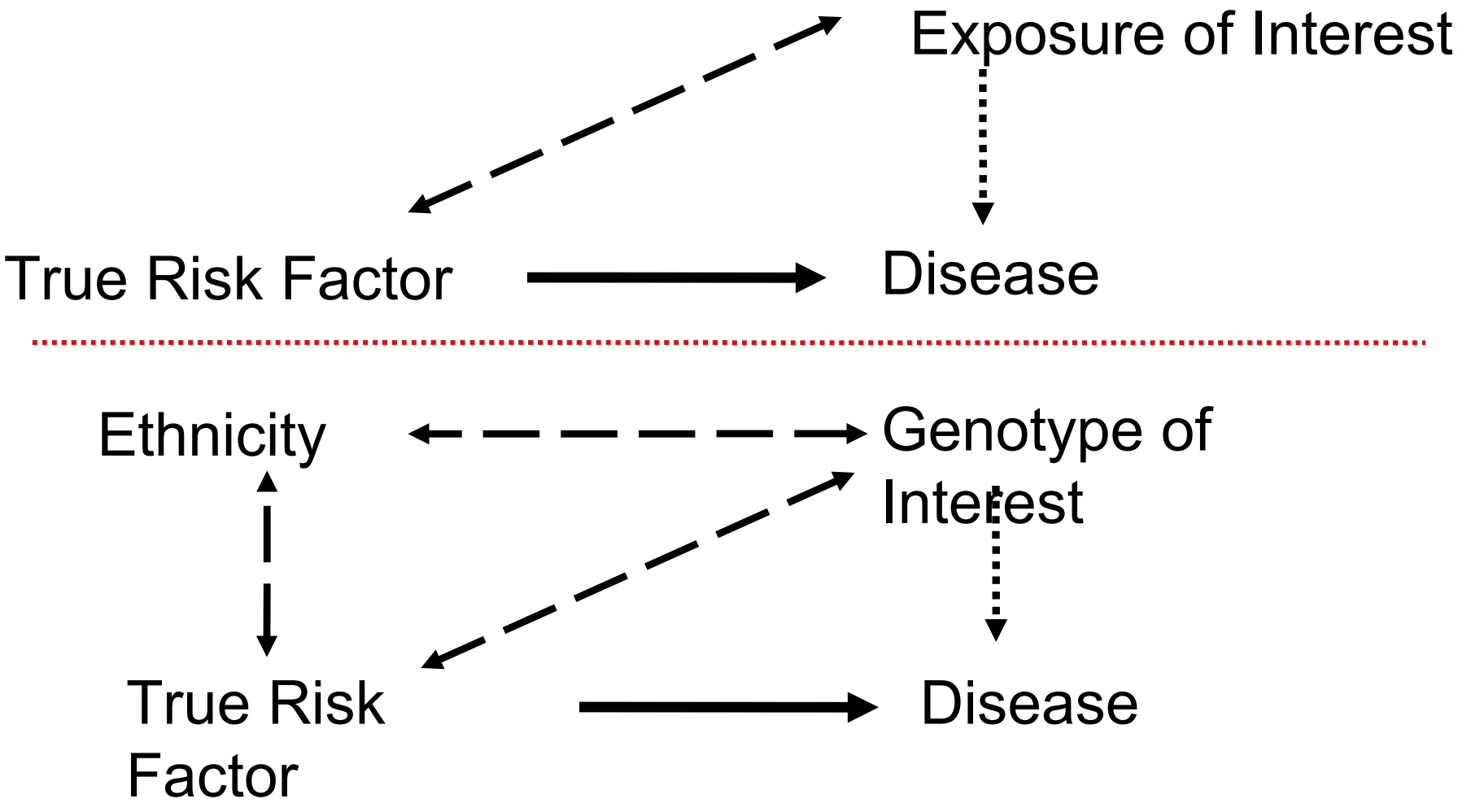
Process of Genetic Epidemiology



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

Population Stratification: Confounding



Wacholder, *JNCI*, 2000

Population Stratification: Gm3;5,13,14 in admixed sample of Native Americans of the Pima and Papago tribes

Study Population: 4,290 Pima and Papago Indians

Genetic Variant: Gm 3;5,13, 15 haplotype (Gm system of human immunoglobulin G)

Outcome: Type 2 diabetes

Question: Is the Gm 3; 5,13, 15 haplotype associated with Type 2 diabetes?

Knowler, *AJHG*, 1998

Population Stratification: Gm3;5,13,14 in admixed sample of Native Americans of the Pima and Papago tribes & Non-insulin dependent Diabetes Mellitus (NIDDM)

Full heritage American Indian population:

	+	-
Gm3;4,13,14	~1%	~99%

NIDDM prevalence: 40%

Caucasian population:

	+	-
Gm3;4,13,14	~66%	~34%

NIDDM prevalence: 15%

Combined:

Gm3,5,13,14	Case	Control
+	7.8%	29.0%
-	92.2%	71.0%

Different genotype frequency, different phenotype frequency
Unadjusted for ethnic background OR=0.27 (95% CI 0.18-0.40)

Population Stratification: Gm3;5,13,14 in admixed sample of Native Americans of the Pima and Papago tribes & Non-insulin dependent Diabetes Mellitus (NIDDM)

Index of Indian Heritage	Gm3;5,13,14 haplotype	% Diabetes
0	65.8%	18.5%
4	21.1%	28.5%
8	1.6%	39.2%

Combined:

Gm3,5,13,14	Case	Control
+	7.8%	29.0%
-	92.2%	71.0%

Different genotype frequency, different phenotype frequency
Unadjusted for ethnic background, OR=0.27 (95% CI 0.18-0.40)
Adjusted for ethnic background, OR=0.83 (95%CI 0.58-1.18)

--> Previous result just picked out race/ethnicity!

Ancestry Informative Markers

- Polymorphisms with known allele frequency differences across ancestral groups
- Useful in estimating ancestry in admixed individuals
- Example: Duffy locus (codes for blood group)
  - 100% sub-Saharan Africans vs. other groups
  - protects *P. vivax* (malaria)

Example AIM: Duffy locus (rs2814778)

Malaria resistance

http://www.ncbi.nlm.nih.gov/projects/SNP

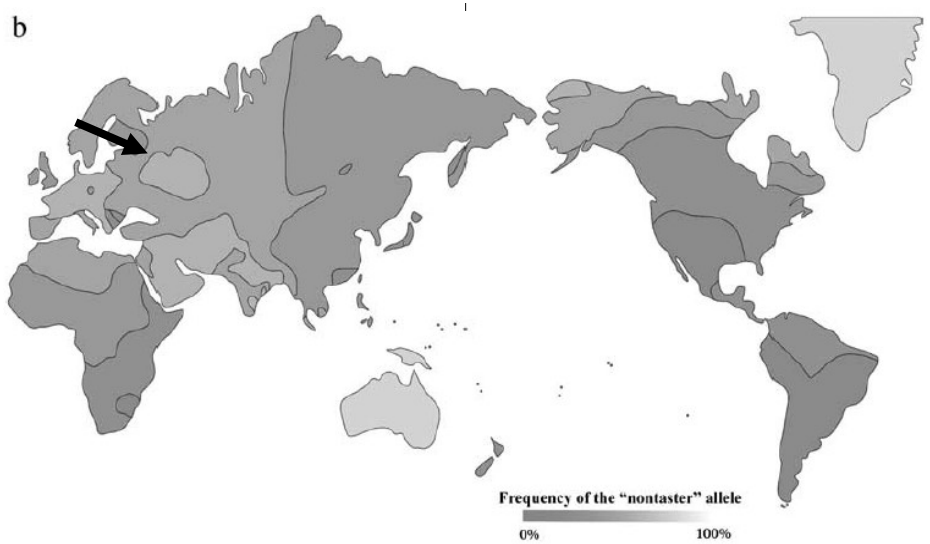
	Sample Ascertainment				Genotype Detail					Alleles	
ss#	Population	Individual Group	Chrom. Sample Cnt.	Source	A/A	A/G	G	G/G	HWP	A	G
ss105436776	P1	24 African/African American, 31 Caucasian, 23 Hispanic, 24 Pacific Rim	204	GF	0.706	0.118		0.176	0.001	0.765	0.235
	CAUC1		62	GF	1.000					1.000	
	AFR1		48	GF		0.292		0.708	0.752	0.146	0.854
	HISP1		46	GF	0.783	0.174		0.043	0.273	0.870	0.130
	PAC1		48	GF	0.958	0.042			1.000	0.979	0.021
	HAPMAP-ASW		98	IG	0.041	0.327		0.633	1.000	0.204	0.796
	HAPMAP-LWK		180	IG		0.011		0.989	1.000	0.006	0.994
	HAPMAP-MEX		100	IG	0.920	0.080			1.000	0.960	0.040
	HAPMAP-MKK		286	IG	0.007	0.126		0.867	0.752	0.070	0.930
	HAPMAP-TSI		176	IG	0.977	0.023			1.000	0.989	0.011
	P3	76 African/African American, 66 Caucasian, 49 Native American/Hispanic, 32 Pacific Rim	532	GF	0.737	0.019		0.244	0.001	0.746	0.254
	CAUC3		122	GF	0.984	0.016			1.000	0.992	0.008
	AFR3		148	GF	0.095	0.027		0.878	0.001	0.108	0.892
	HISP3		92	GF	0.957	0.043			1.000	0.978	0.022
	PAC3		170	GF	1.000					1.000	
	ENSEMBL_Watson		2	IG	1.000					1.000	
	ENSEMBL_Venter		2	IG	1.000					1.000	
ss119047871	YRI		2	IG				1.000			1.000
ss164197181	YRI	Sub-Saharan African	2	IG				1.000			1.000

Outline

- Review last weeks homeworks
- Last few topics from previous pop gen lecture
  - (Had just finished HWE game)
  - Population stratification
- **Linkage analysis, with PTC testing example**
- Family-based association tests
  - Overview
  - Trios/TDT
  - Discordant sibships

Family-based analysis shown here are completely robust to population stratification

Geographic Distribution of PTC Phenotype



Wooding Genetics 2006, adapted from Cavalli-Sforza 1994

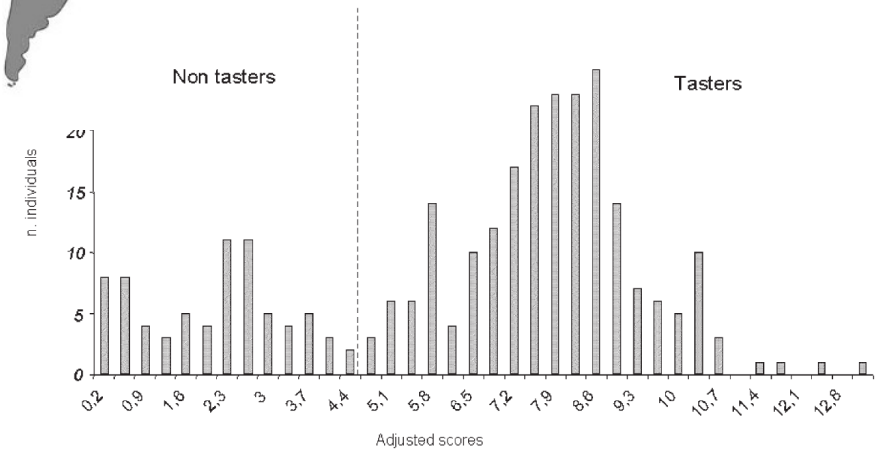


Figure 1 Distribution of PTC sensitivity in Talana. The *x*-axis indicates the PTC scores, adjusted for age and sex using the following formula: $n + [(a - a_M)/20] - 0.73$female, where n is the PTC tasted solution, a is the age of the individual and a_M is the average age of participants. The value 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

Why do Family Studies?

Is the trait genetic?

What is the mode of transmission?

Dominant

Recessive

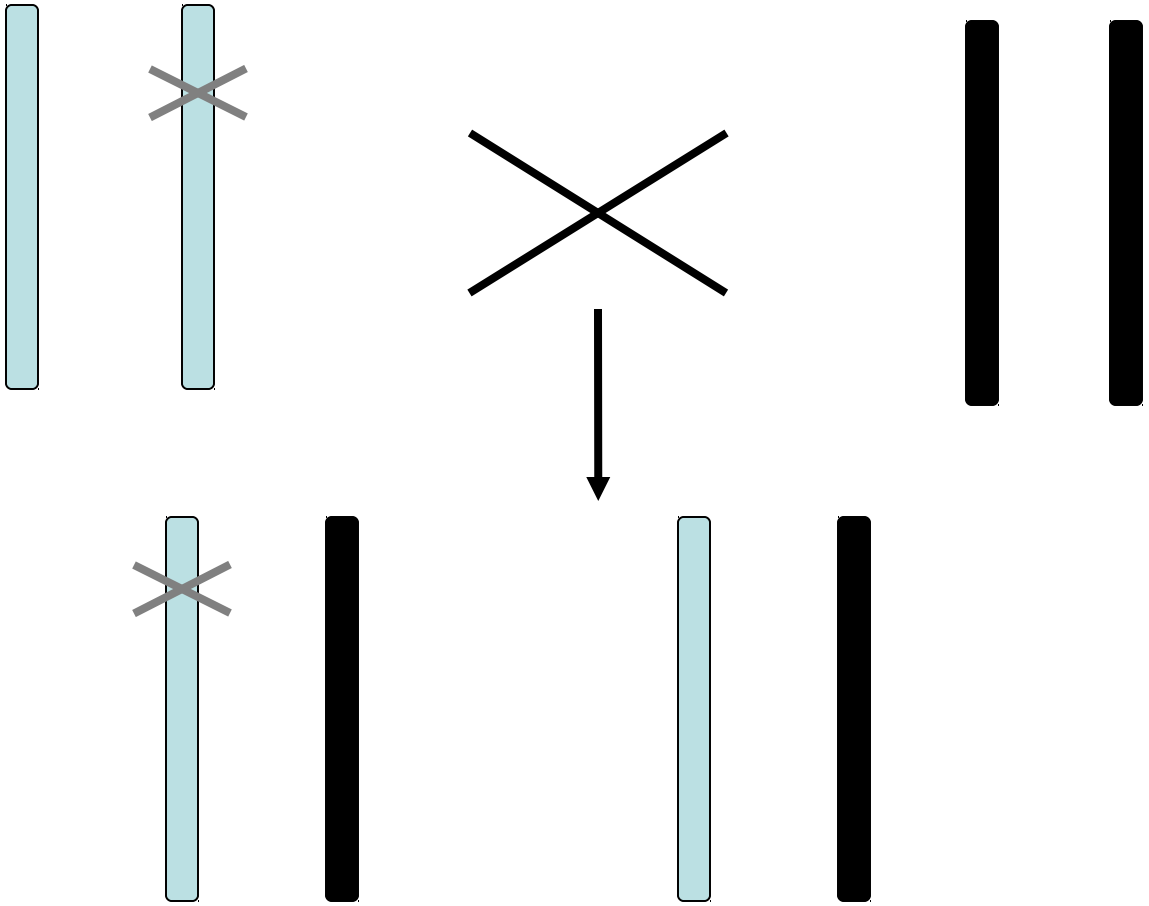
Additive

Polygenic (Multiple genes involved)

Mendel's laws

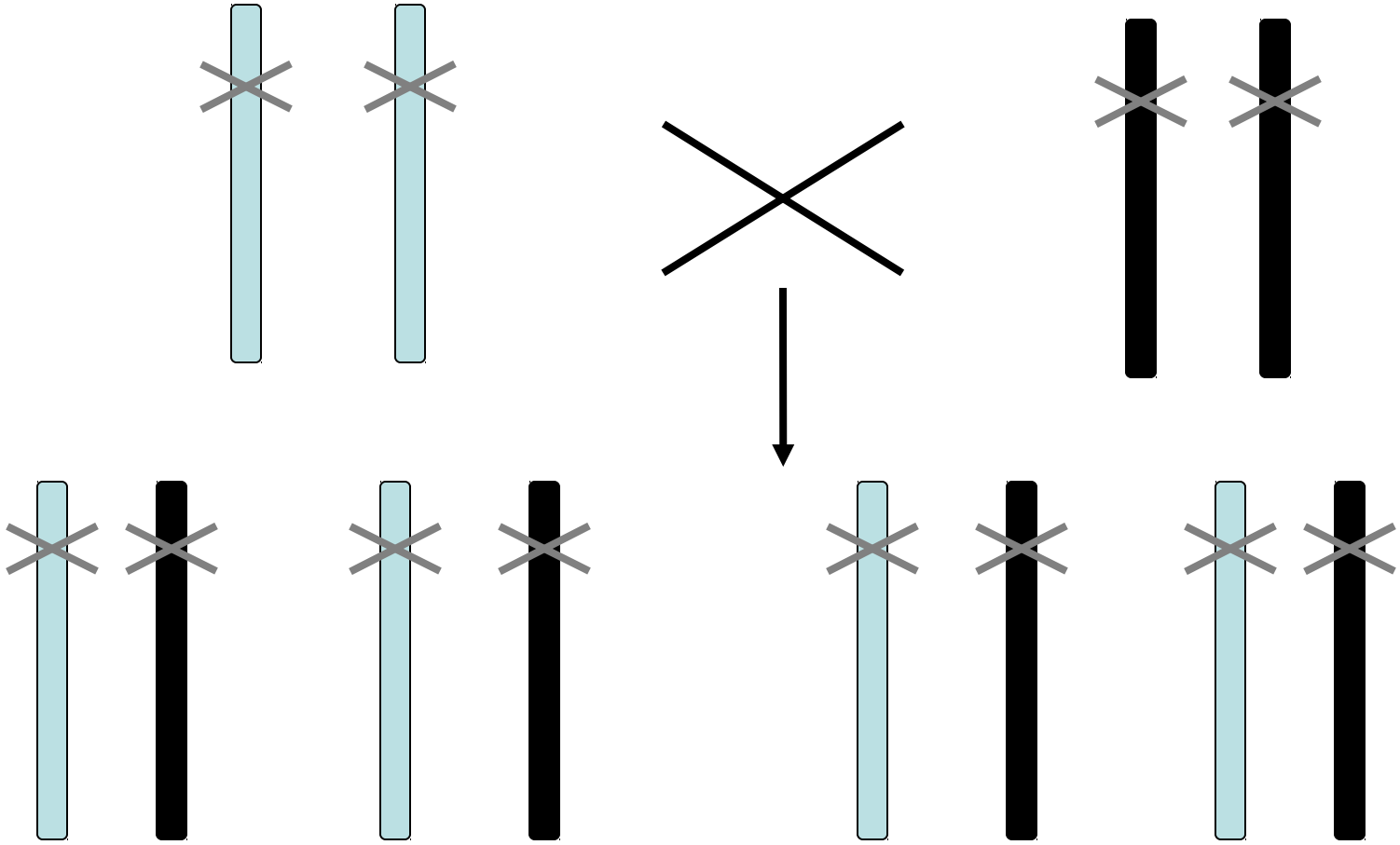
- Recall the playing cards example...
- One allele from each parent for each gene
  - Many family based tests based on this, rather than estimating allele frequencies (case-control)

Recall: Mutation and Meiosis



Recall: E.g., Recessive trait

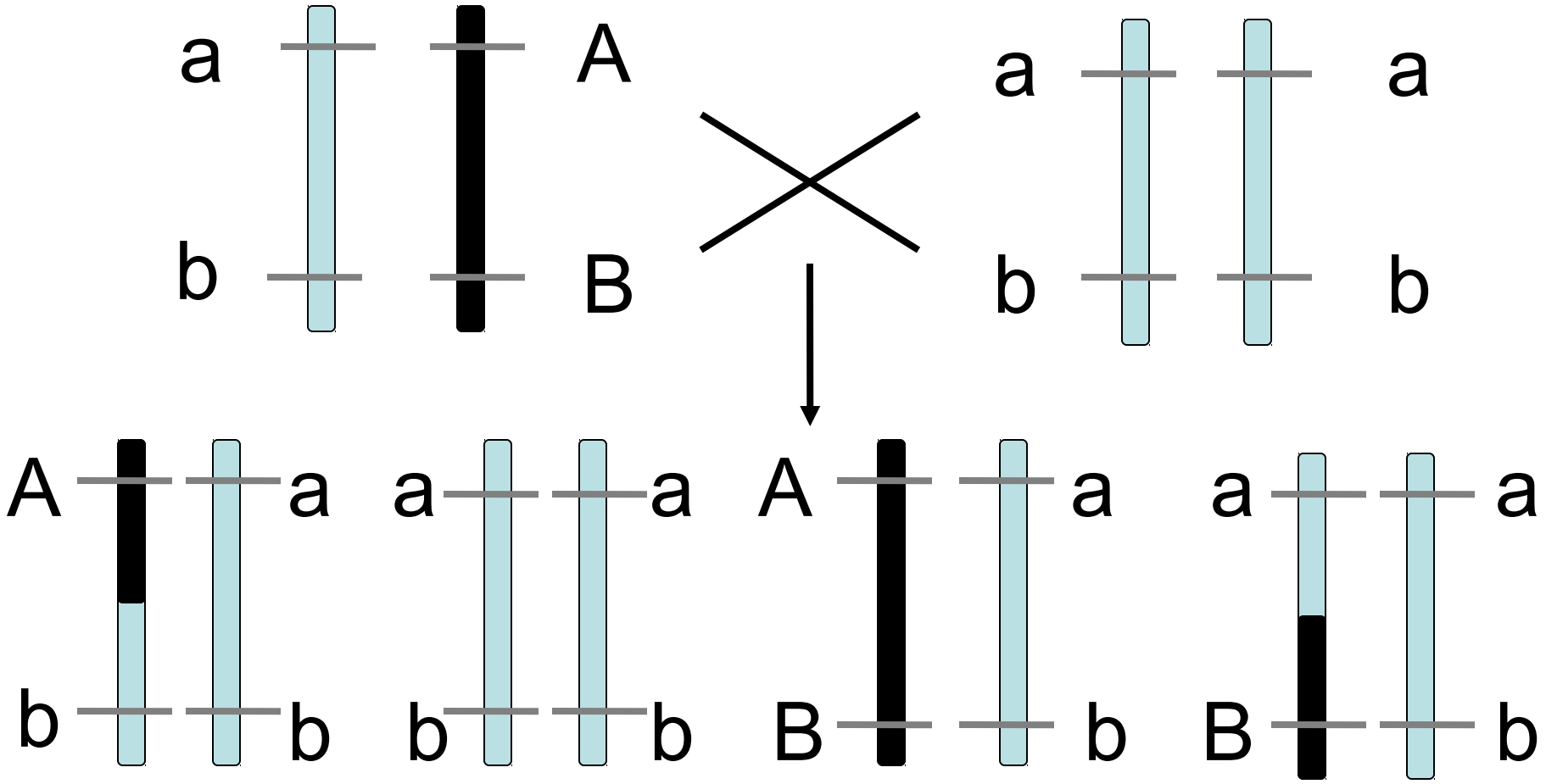
(Know the trait, know the genotype at the disease loci)



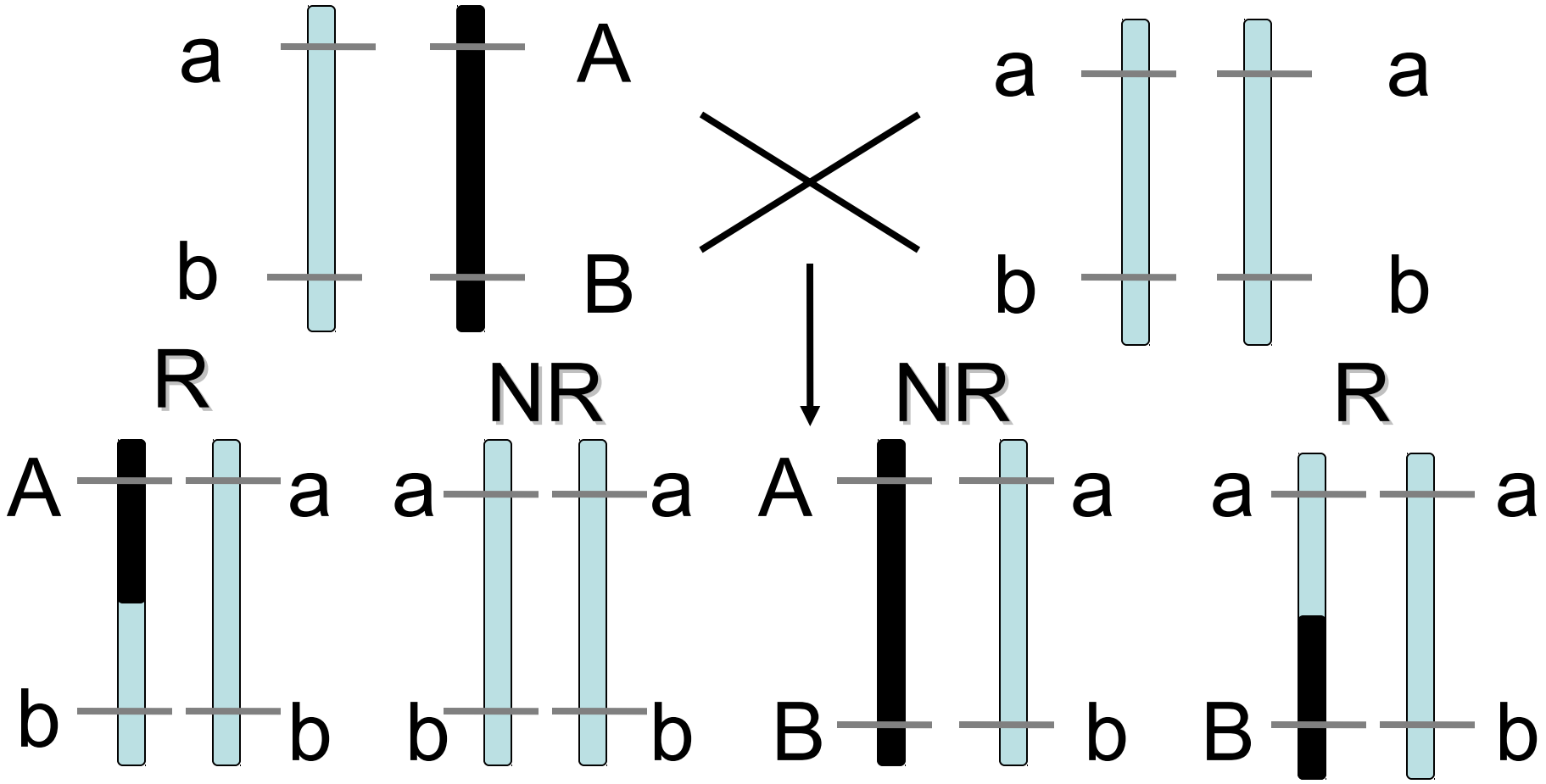
Linkage Analysis

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

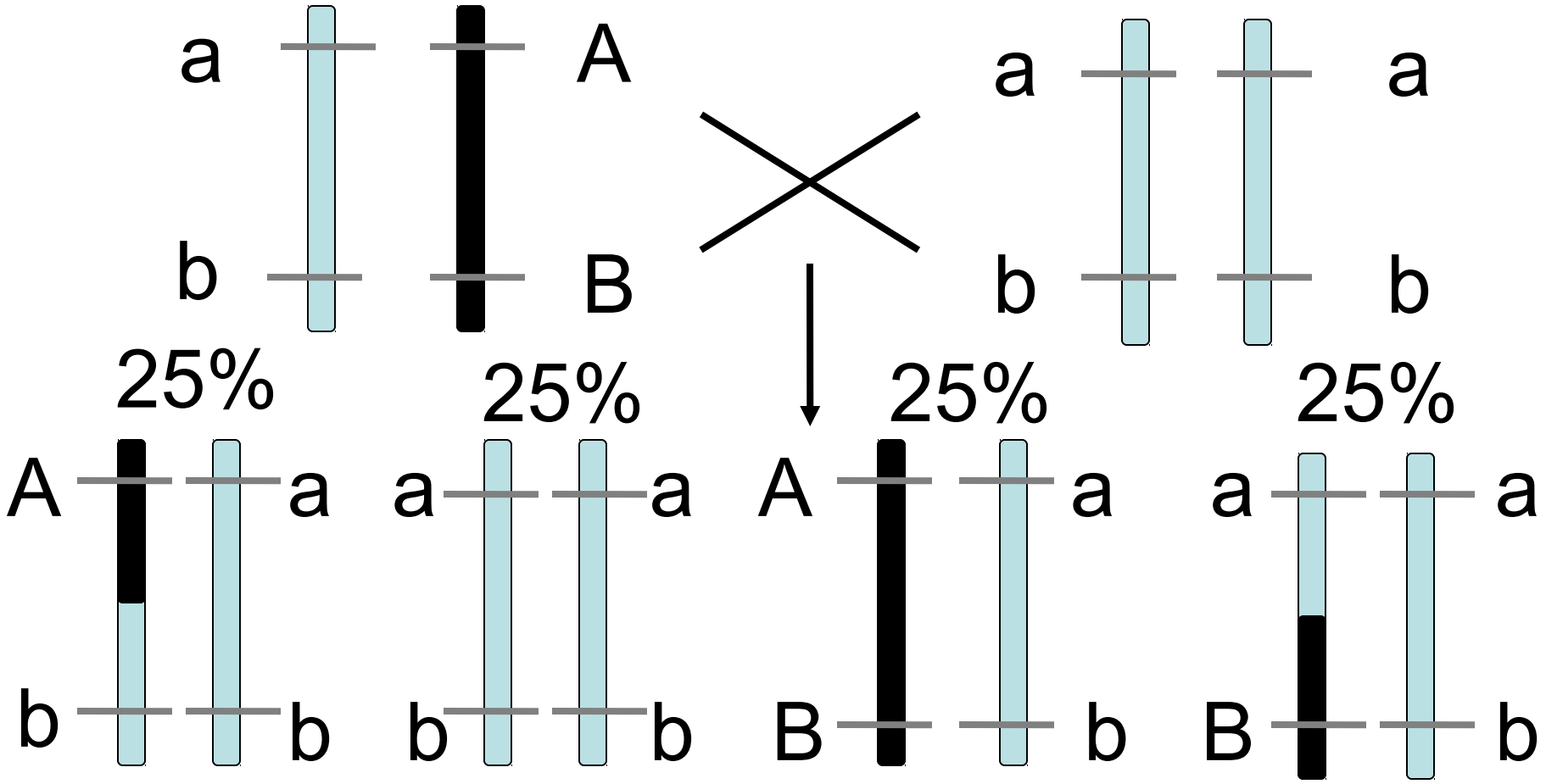
Recombination – was there recombination?



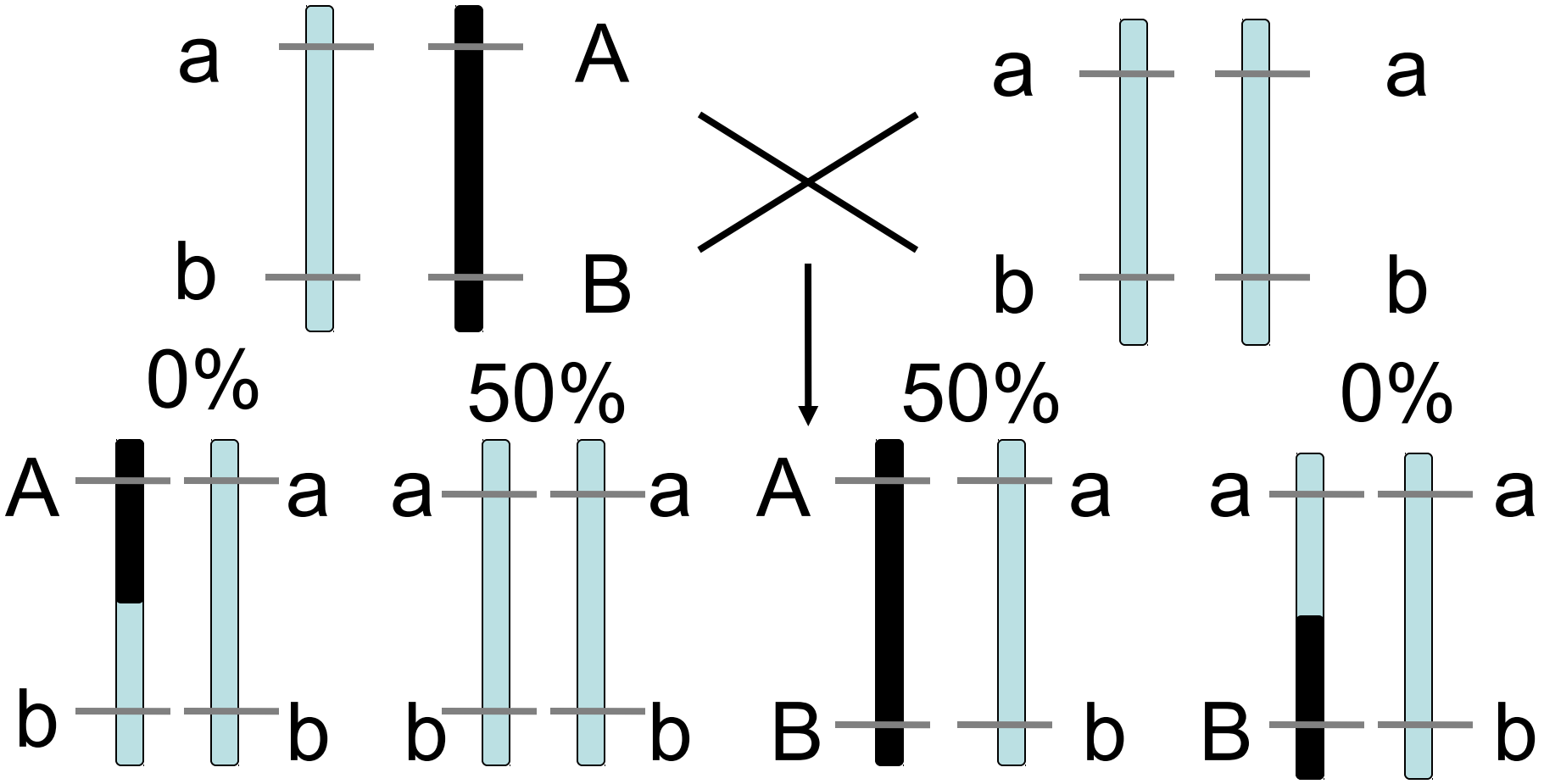
Recombination– was there recombination?



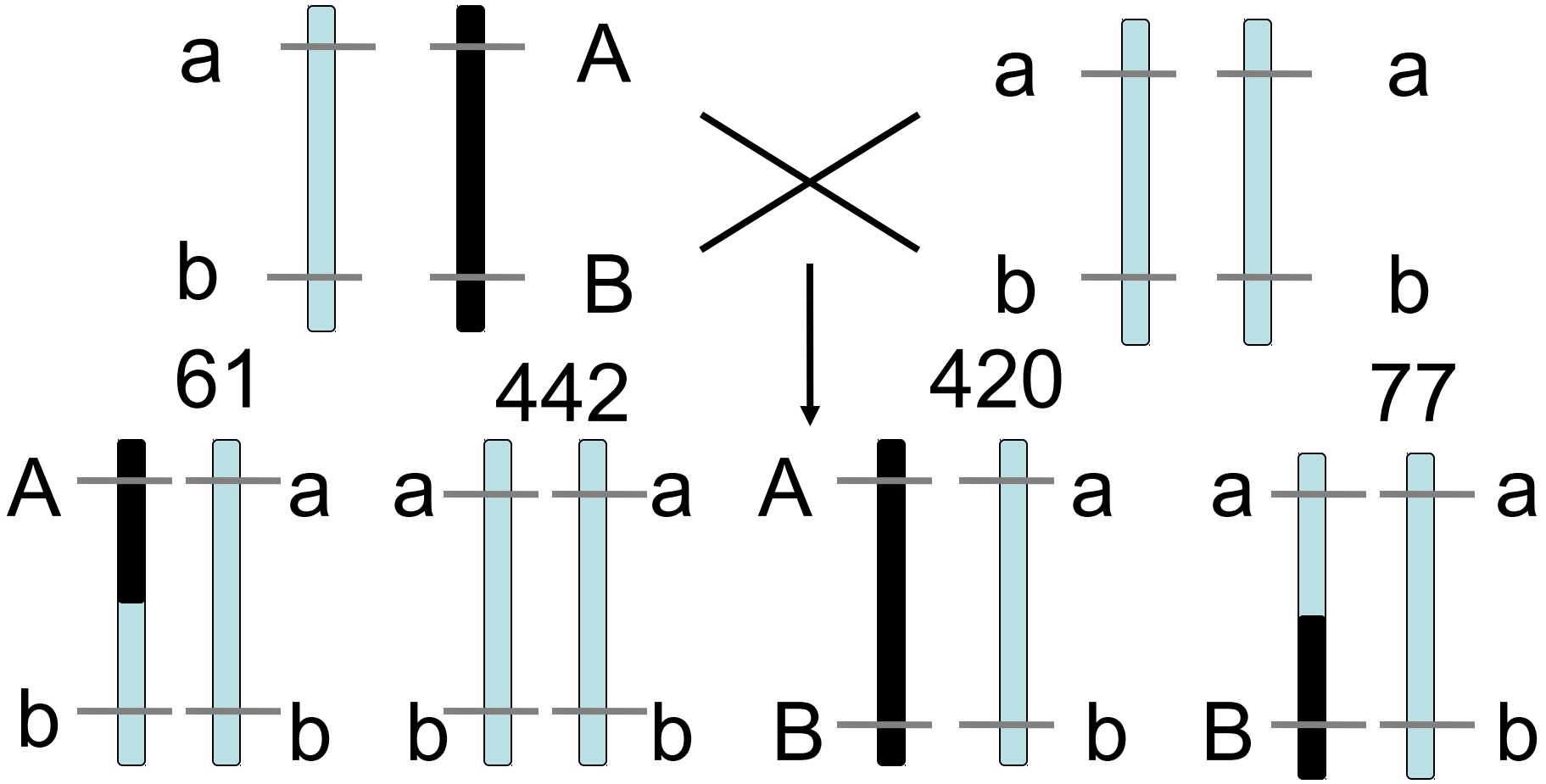
Independent Assortment



No recombination

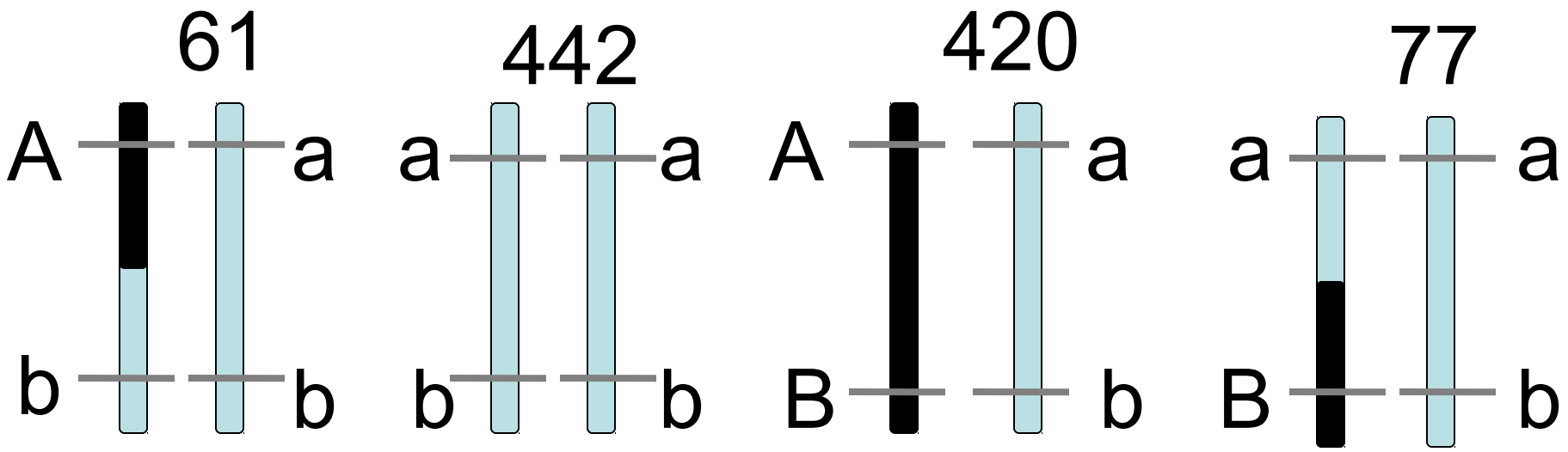


Recombination Fraction



Recombination Fraction

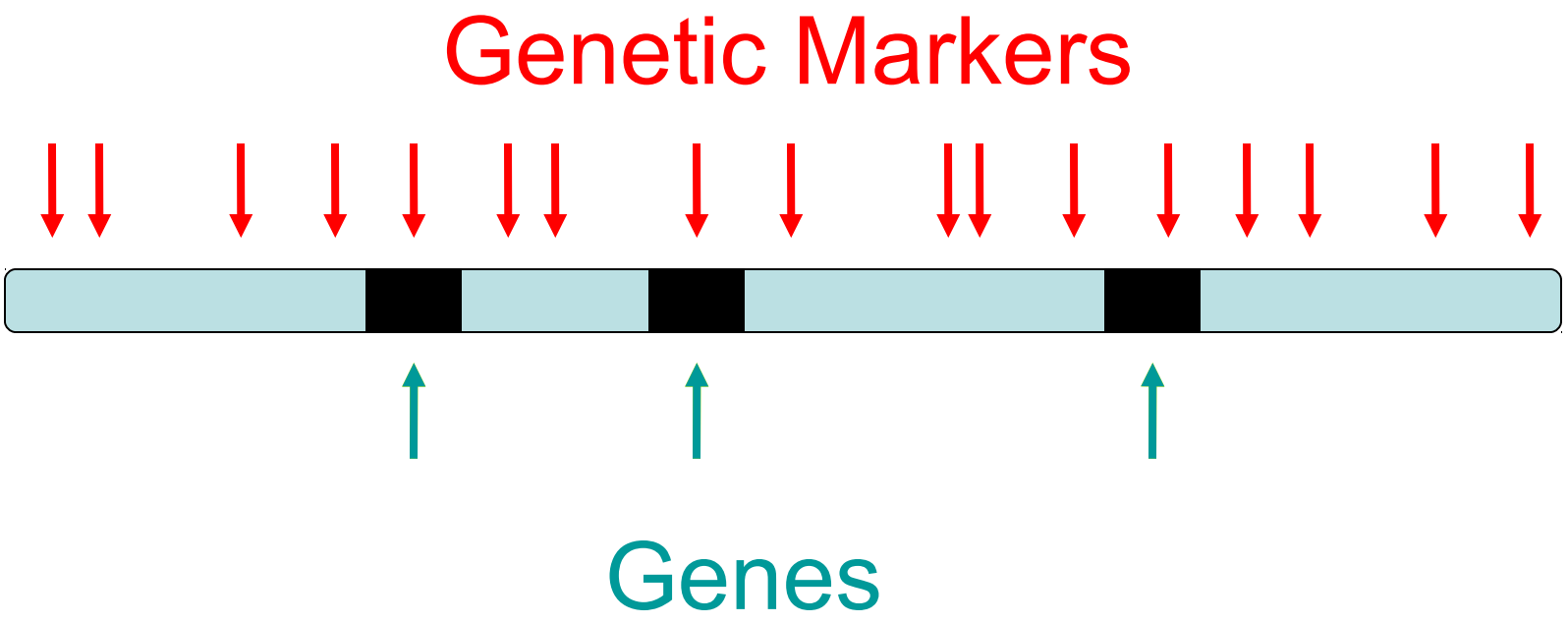
Recombination Fraction θ =
Recombinants / Total =
(61+77)/(61+77+442+420)=138/1000
= 13.8% < 50%, indicating some linkage



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

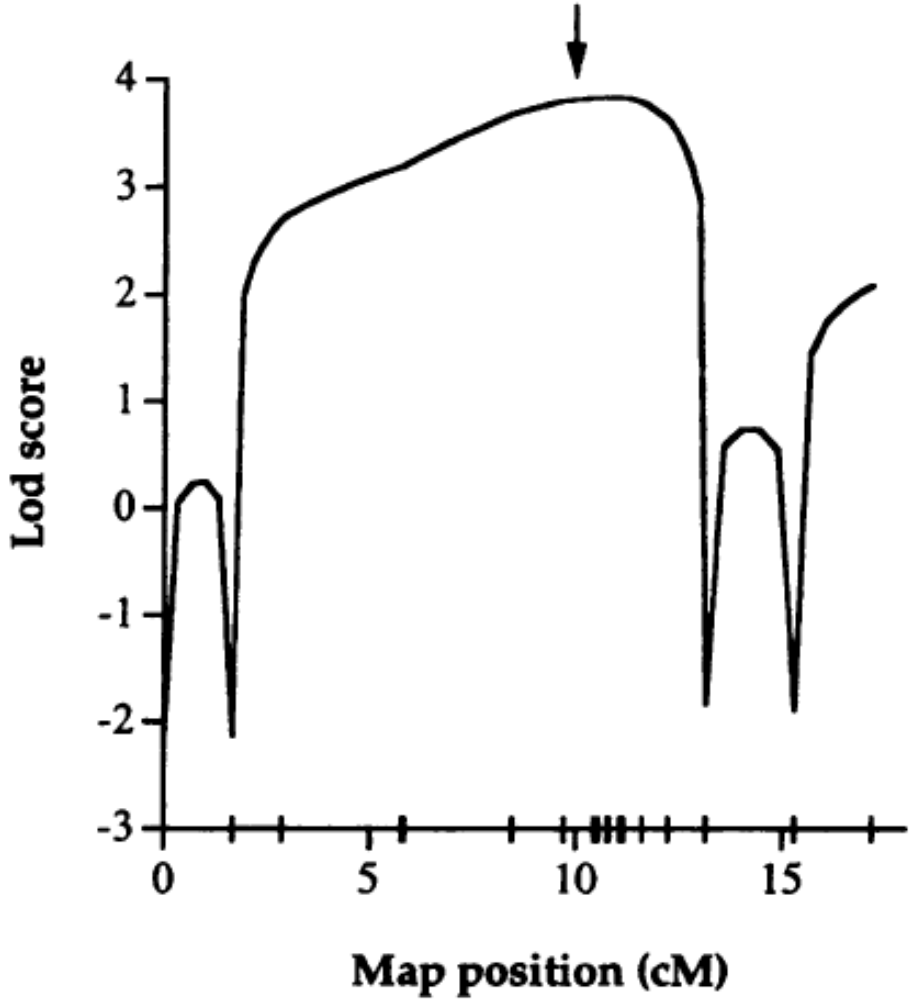


Linkage Analysis

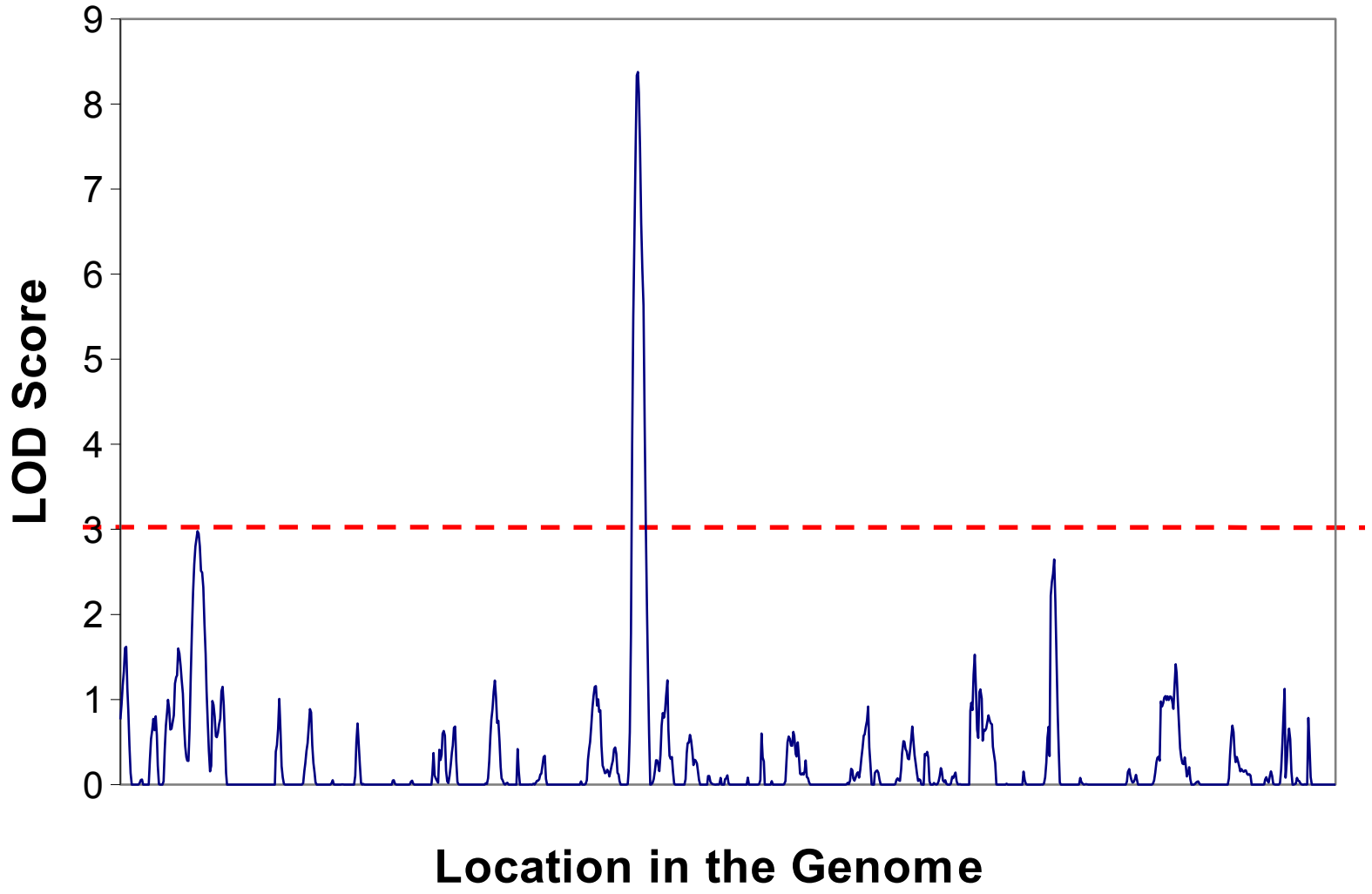
LOD score based on recombination

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

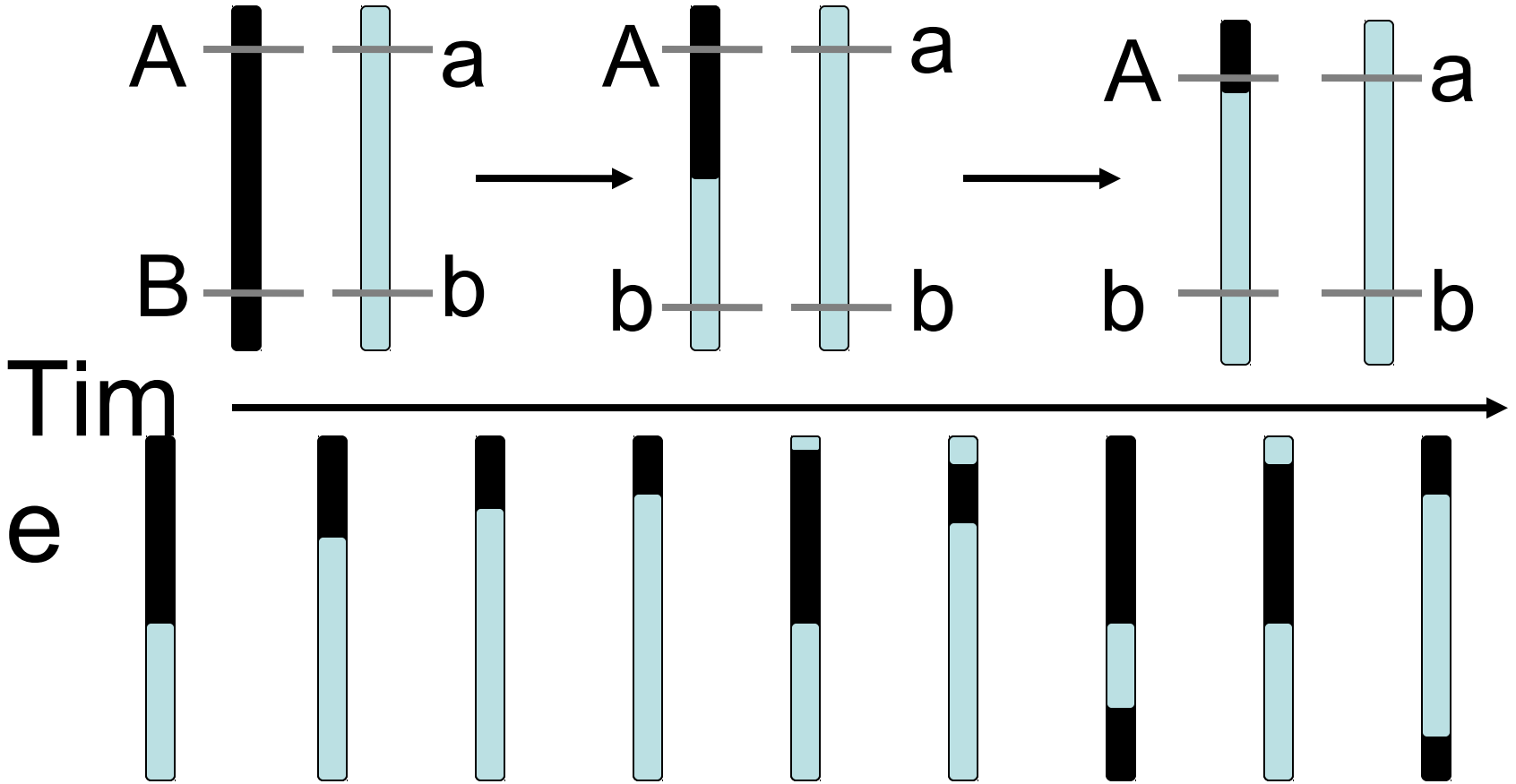
Multipoint Linkage Analysis



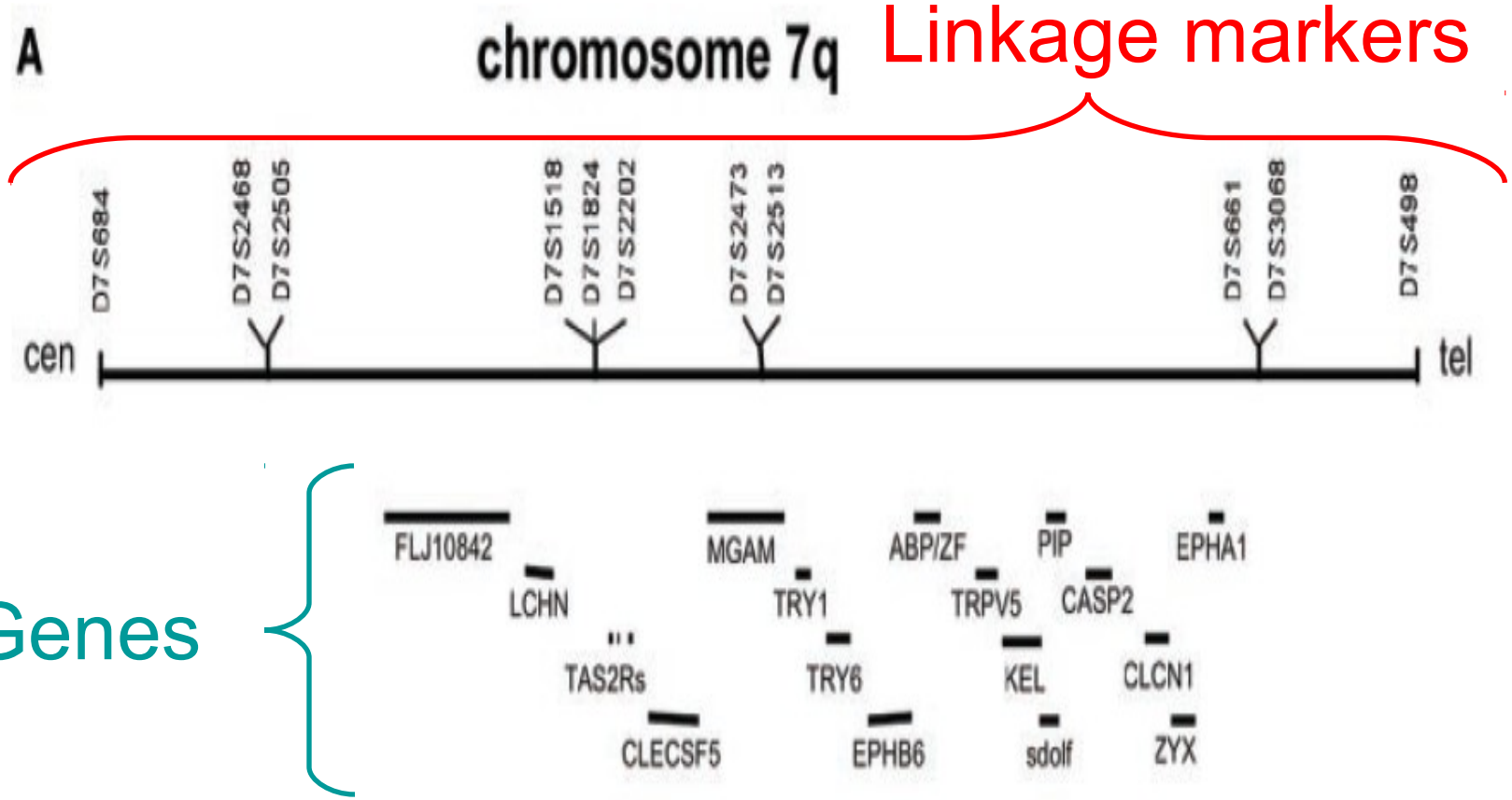
PTC Linkage Analysis



Linkage Disequilibrium

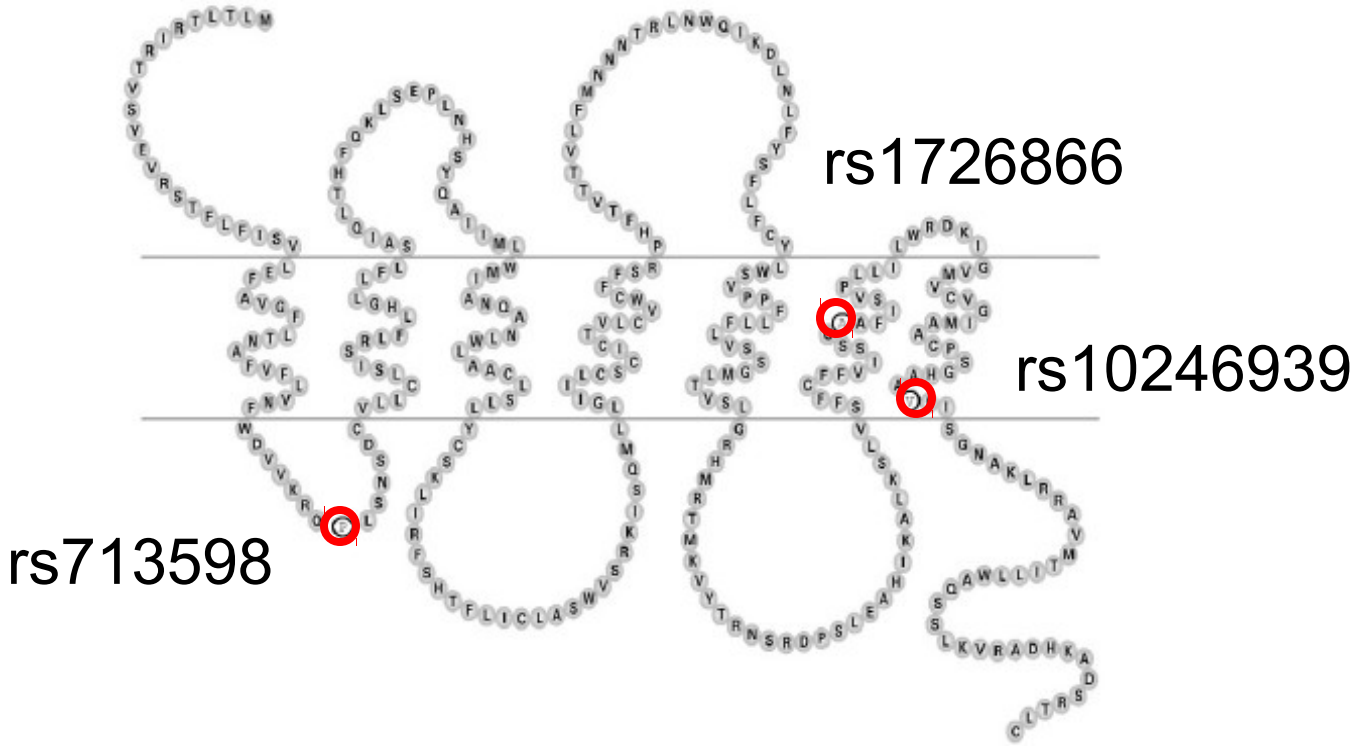


Fine Mapping



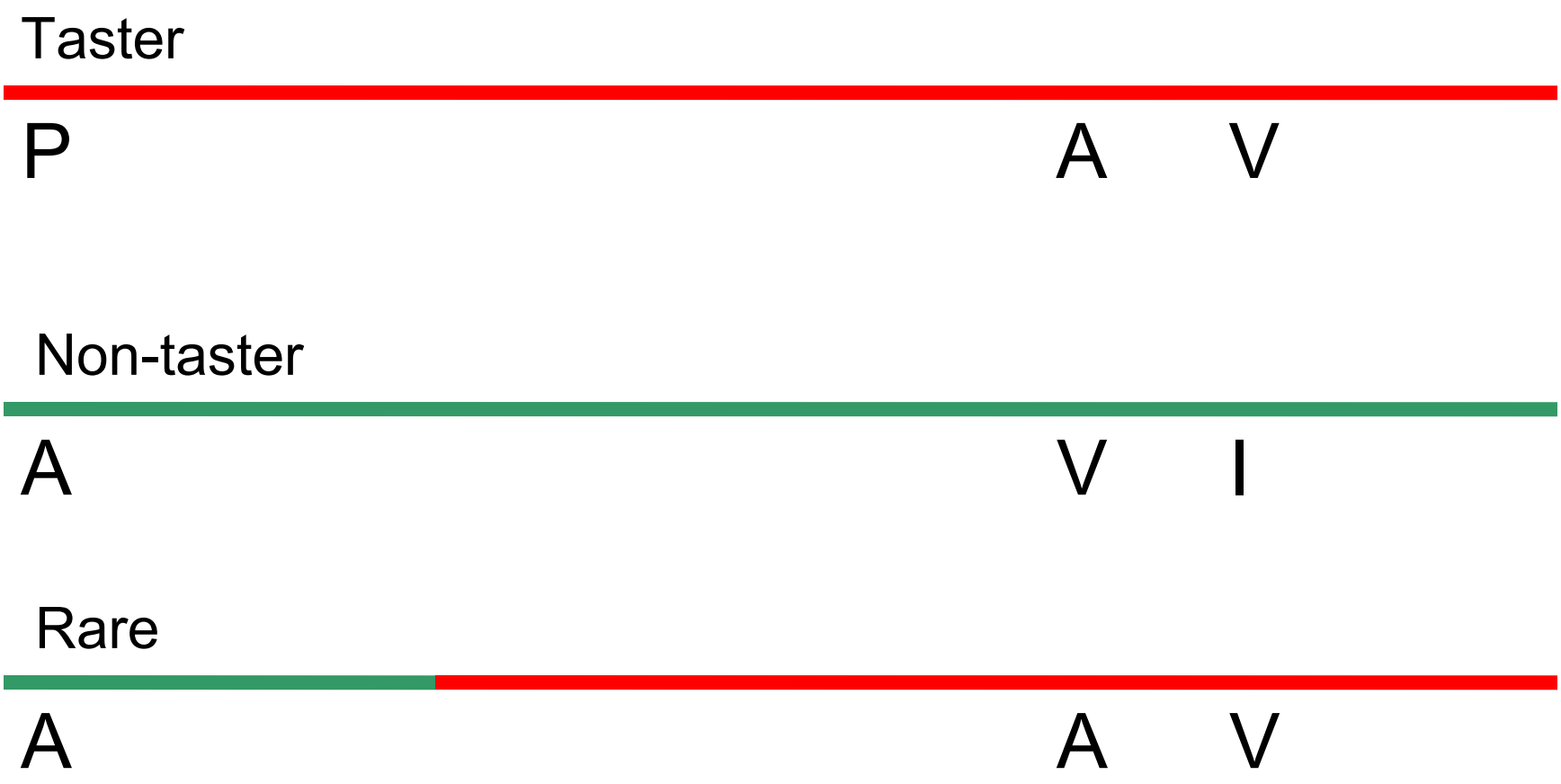
Kim et al. Science 2003

TAS2R38 Receptor Structure



Kim et al. J Dent Res 2004

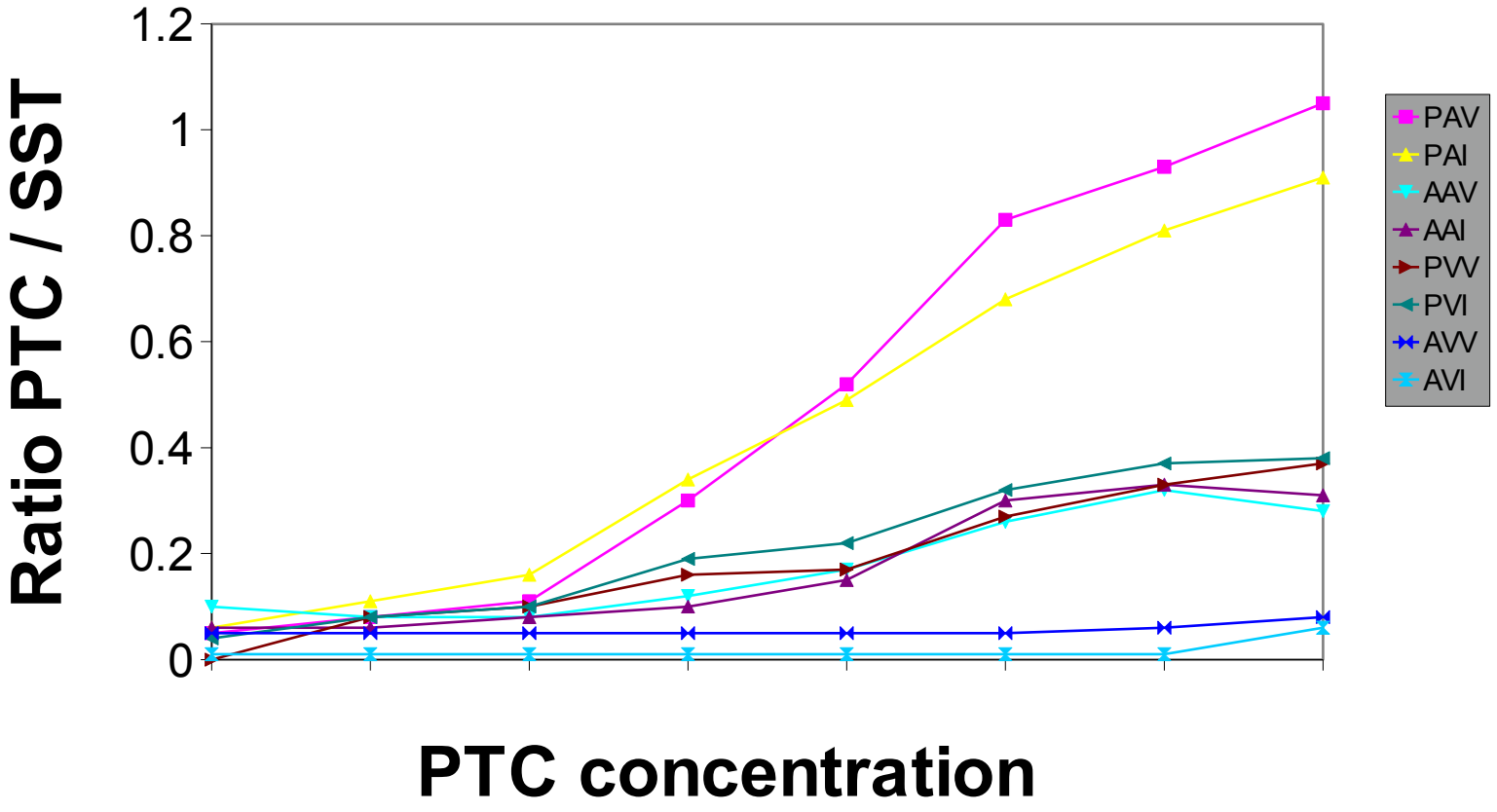
3 SNPs form 3 Haplotypes



Predicted Effect of the 3 SNPs

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

TAS2R38 Haplotype function in vitro



Outline

- Review last weeks homeworks
- Last few topics from previous pop gen lecture
  - (Had just finished HWE game)
  - Population stratification
- Linkage analysis, with PTC testing example
- **Family-based association tests**
  - **Overview**
  - **Trios/TDT**
  - **Discordant sibships**

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)=?

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 known
  - dichotomous phenotype (though Q-TDT), only affected offspring
- Count transmissions from heterozygote parents, and compare to expected transmissions
  - Mendel's laws of segregation (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)

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	
		A	a
Transmitted parental allele	A	2	0
	a	0	0

AA - AA
|
AA

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

Both parents homozygous

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

aa-aa
|
aa

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

AA-Aa

|

AA,Aa

.5 .5 ← Pr

		Non-transmitted parental allele: A	Non-transmitted parental allele: 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

.5 .5 ← 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, 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
- 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