

Family-Based Studies



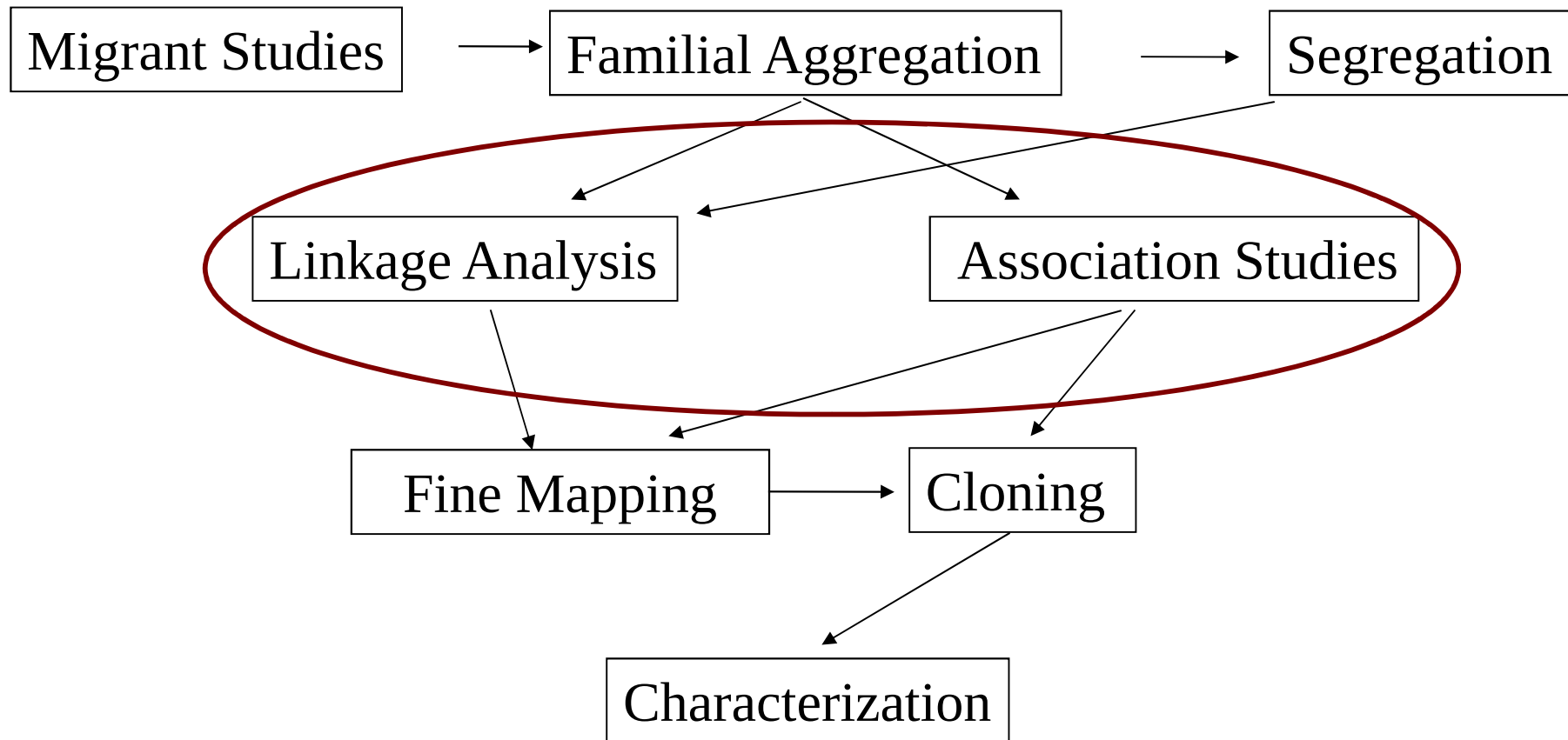
<https://belkazan.files.wordpress.com/2012/04/stripe-shirts-mother-father-child-at-the-beach-boardwalk.png?w=620>

Outline

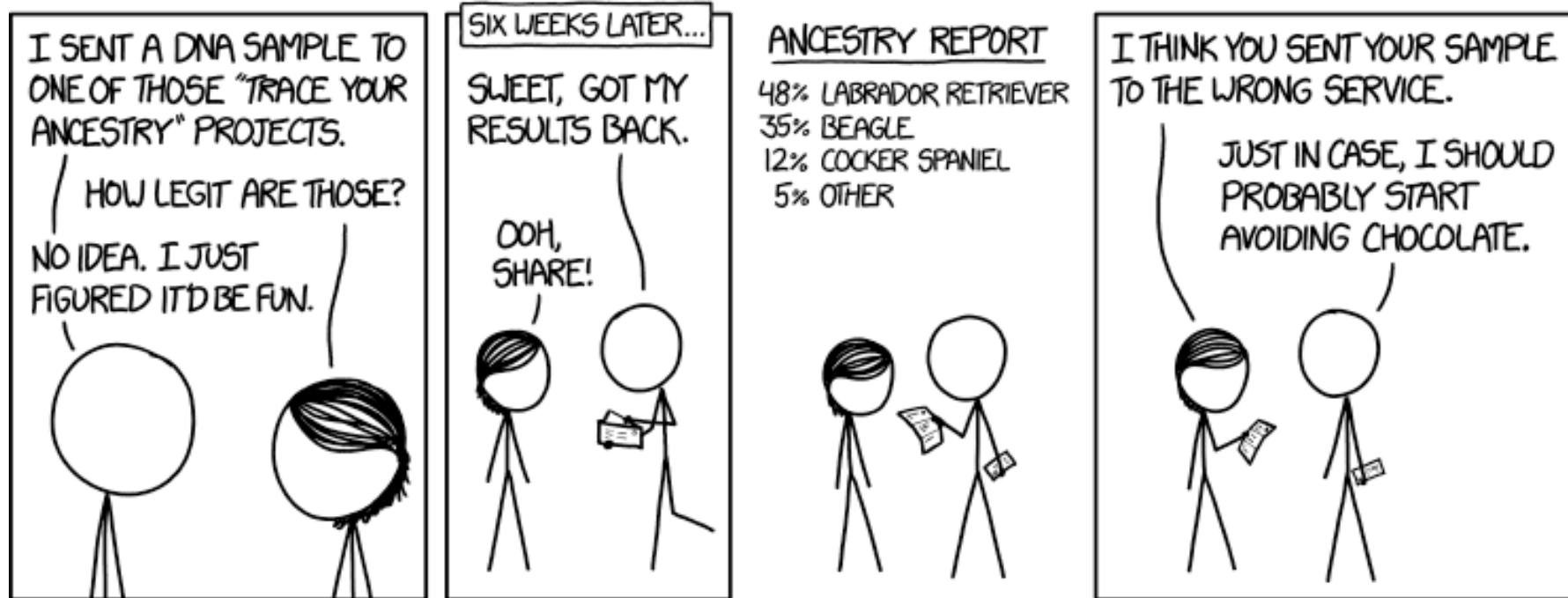
- **Background/Review**
- Family studies [PTC example]
- Linkage analysis [PTC example]
- Family-based association tests
 - Trios/TDT
 - Discordant sibships

Review: Process of Genetic Epidemiology

Defining the Phenotype



Review: Population stratification



<https://xkcd.com/1706/>

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)
 - Not based on allele frequency directly...

Outline

- Background/Review
- Family studies [PTC example]**
- Linkage analysis [PTC example]
- Family-based association tests
 - Trios/TDT
 - Discordant sibships

PTC: Bitter Taste Receptor

PTC The Genetics of Bitter Taste

In 1931, a chemist named Arthur Fox was pouring some powdered PTC into a bottle. When some of the powder accidentally blew into the air, a colleague standing nearby complained that the dust tasted bitter. Fox tasted nothing at all. Curious how they could be tasting the chemical differently, they tasted it again. The results were the same. Fox had his friends and family try the chemical then describe how it tasted. Some people tasted nothing. Some found it intensely bitter, and still others thought it tasted only slightly bitter.



The ratio of tasters to non-tasters varies between populations, but every group has some tasters and some non-tasters. On average, 75% of people can taste PTC, while 25% cannot.

<http://learn.genetics.utah.edu/content/basics/ptc/>

Geographic Distribution of PTC Phenotype



Wooding Genetics 2006, adapted from Cavalli-Sforza 1994

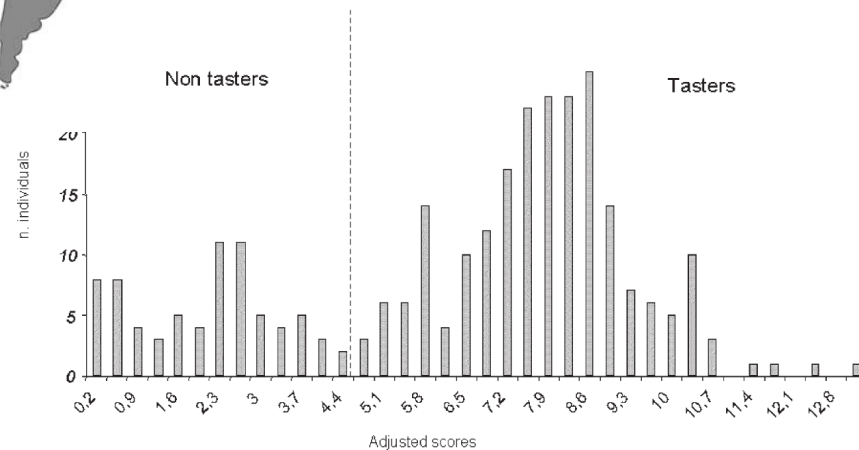


Figure 1 Distribution of PTC sensitivity in Talana. The x-axis indicates the PTC scores, adjusted for age and sex using the following formula: $n + [(a - a_M) / 20] - 0.73$ female, where n is the PTC tasted solution, a is the age of the individual and a_M is the average age of participants. The value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

Types of Genetic Studies

Family Studies *(next page for PTC...)*

Compare trait values across family members

Linkage Analysis

Compare trait values with inheritance patterns

Association

Compare trait values against genetic variants

First PTC Family Study

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

L. H. Snyder Science 1931

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)

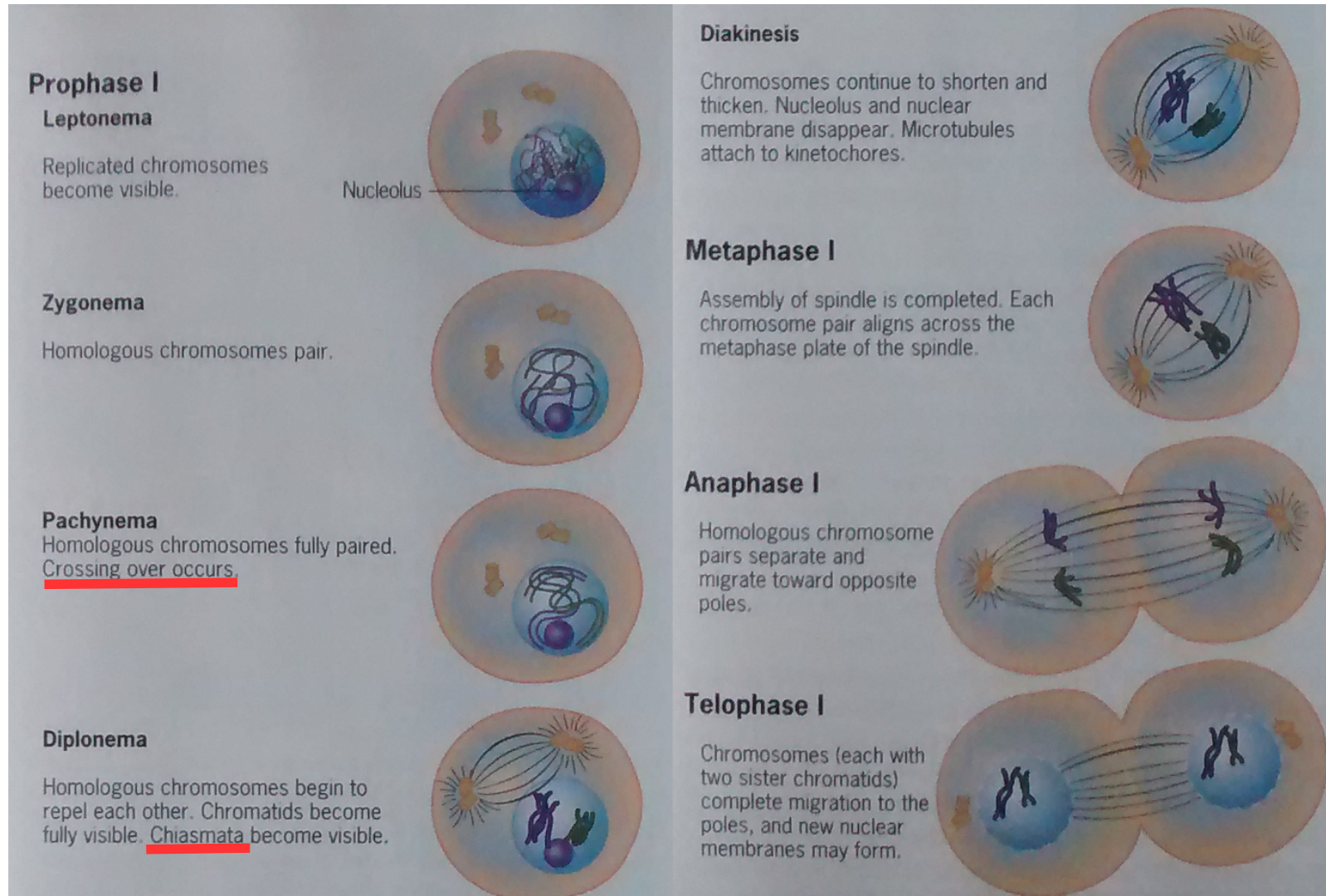


<https://belkazan.files.wordpress.com/2012/04/stripe-shirts-mother-father-child-at-the-beach-boardwalk.png?w=620>

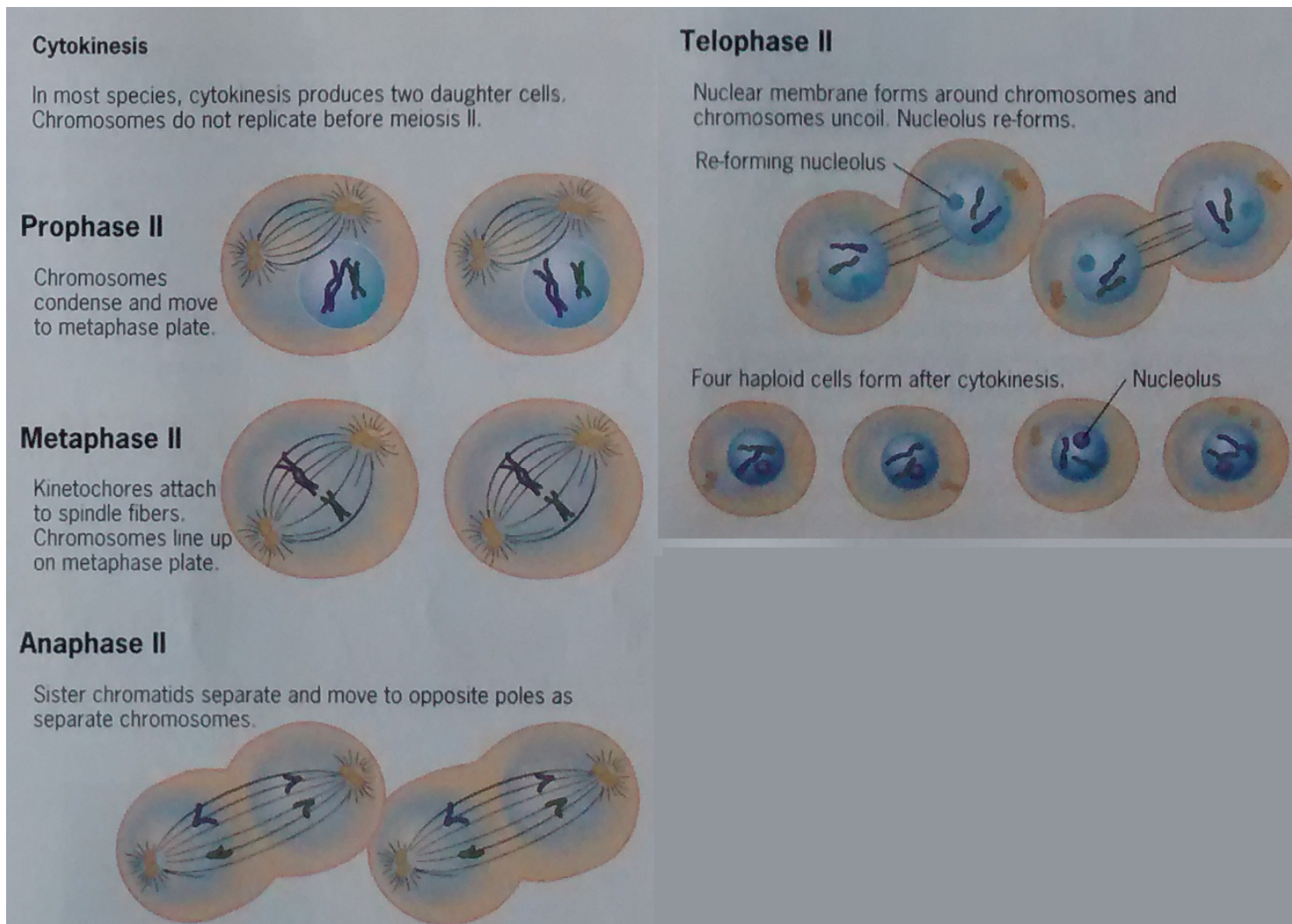
Outline

- Background/Review
- Family studies [PTC example]
- Linkage analysis [PTC example]**
- Family-based association tests
 - Trios/TDT
 - Discordant sibships

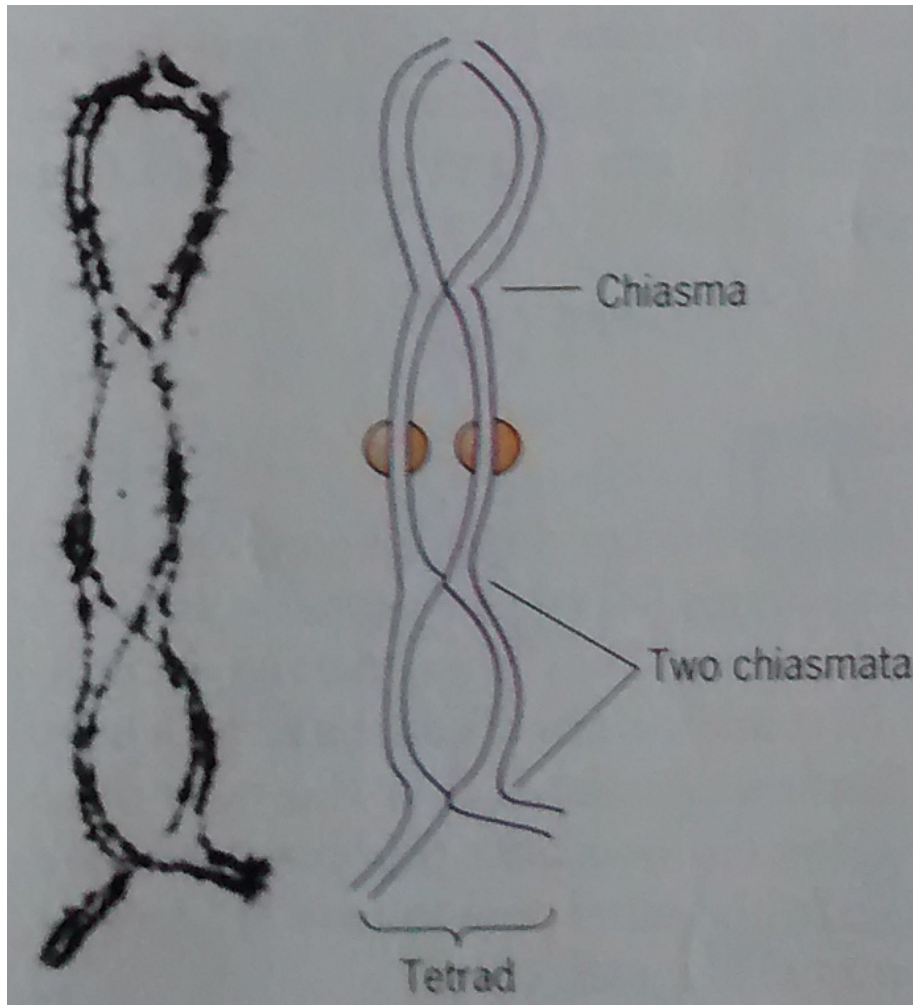
Meiosis – Crossing over



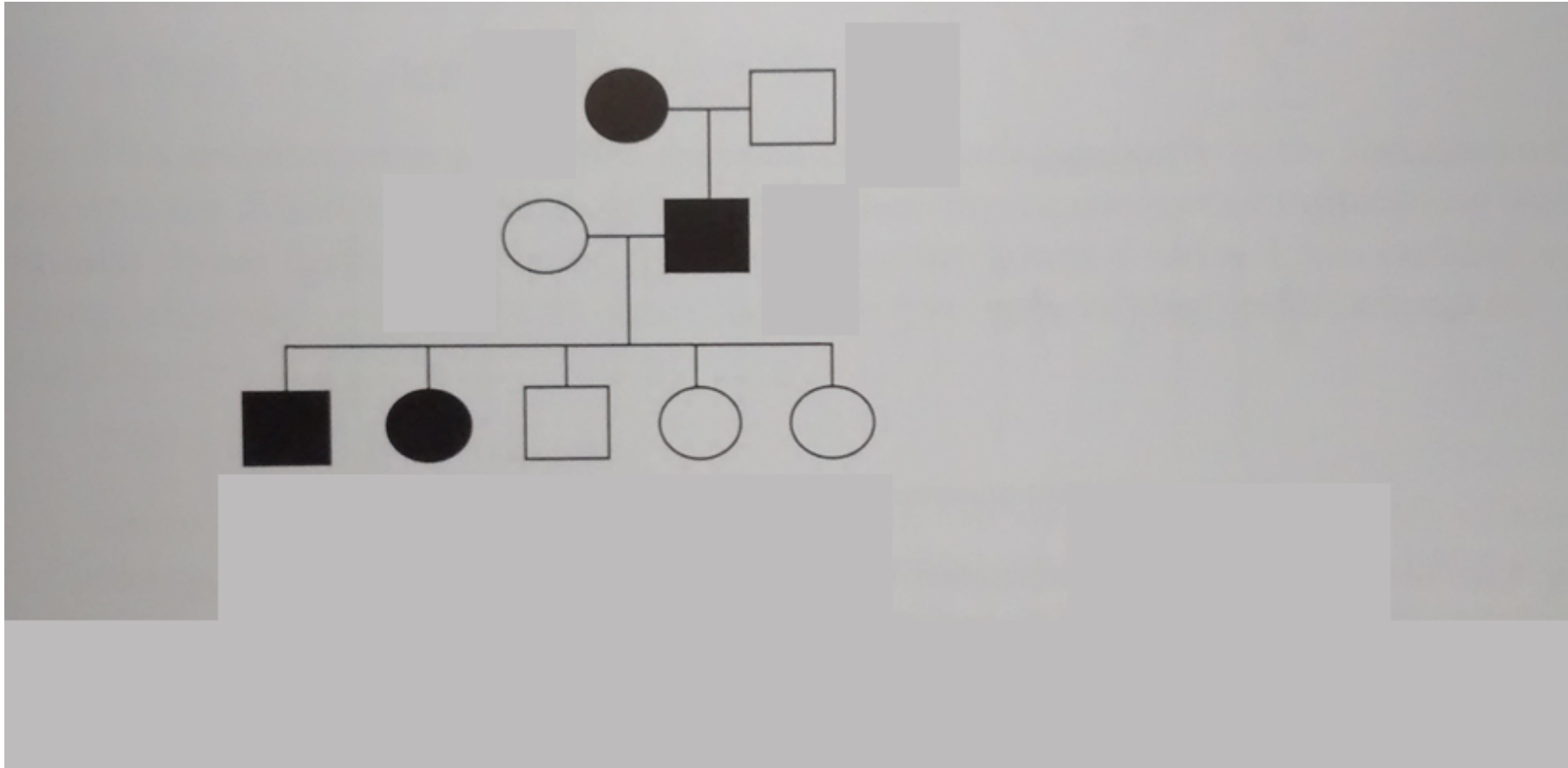
Meiosis – Crossing over



Crossing over – Chiasmata



Idea behind linkage – crossing over



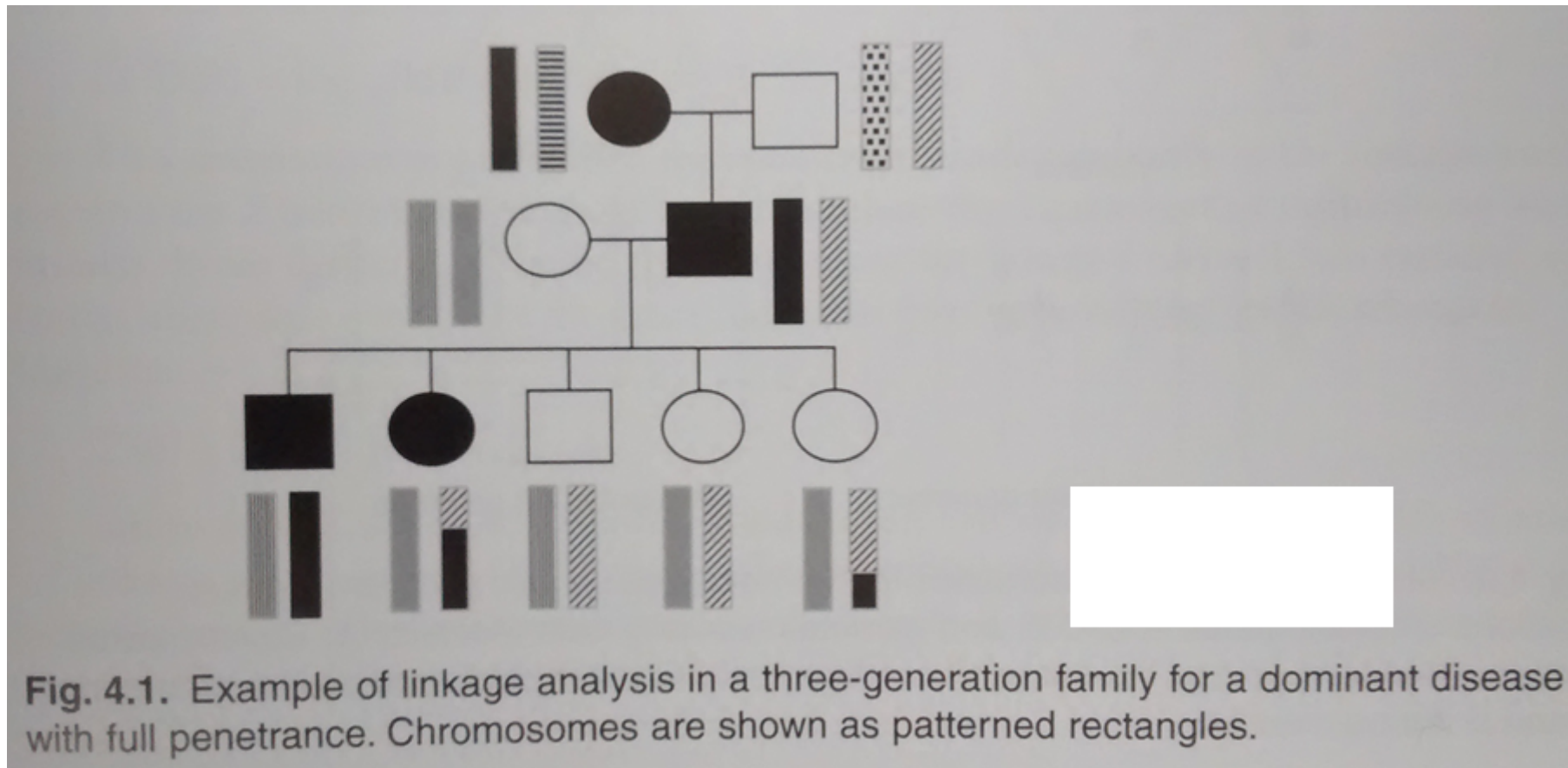
PAUSE: Think then take a poll!

What kind of disease is this, assuming full penetrance?

Dominant? Recessive? Neither? Both?

Austin, Genetic Epidemiology

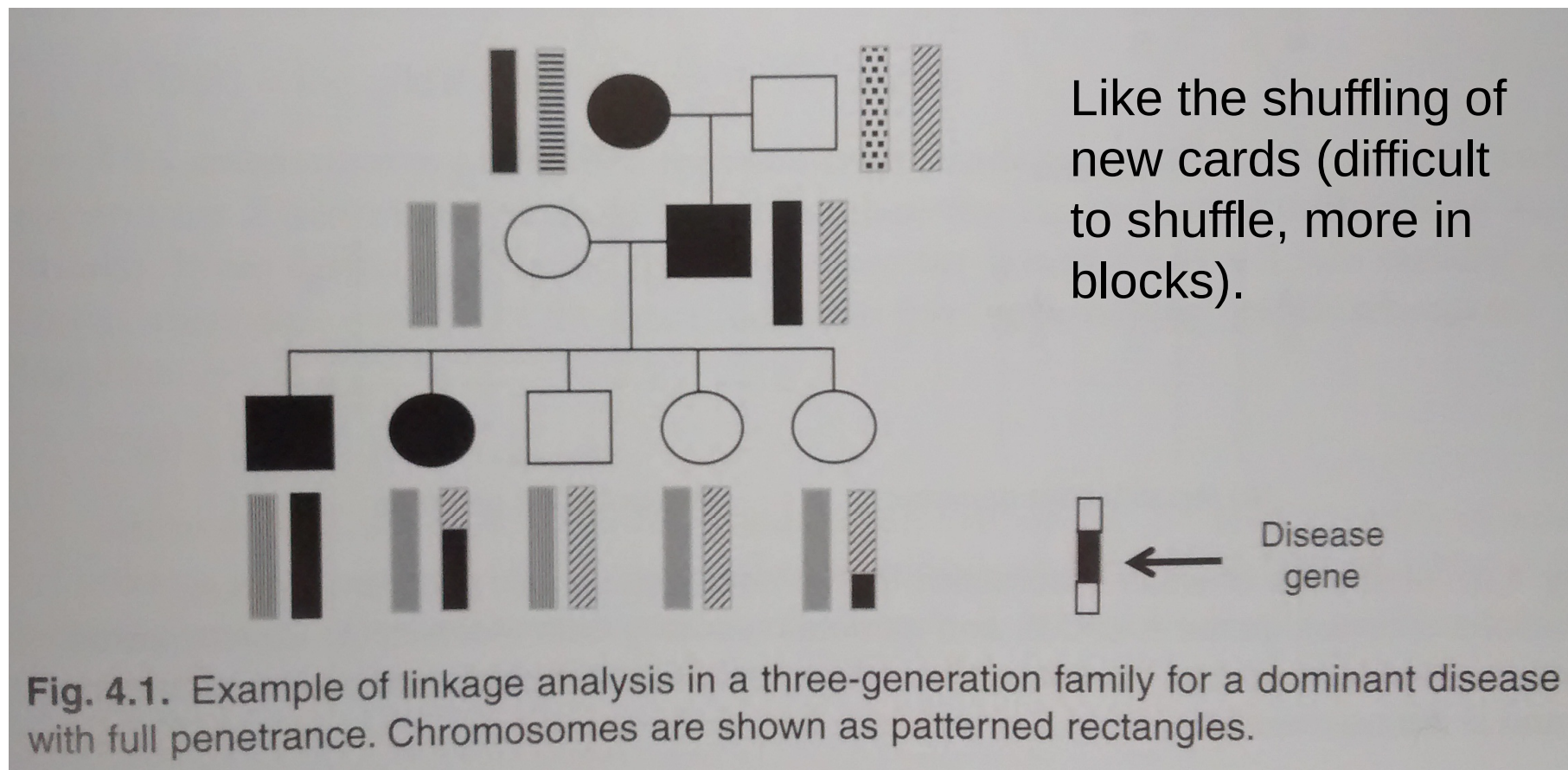
Idea behind linkage



PAUSE:

Can you narrow the genetic region responsible?

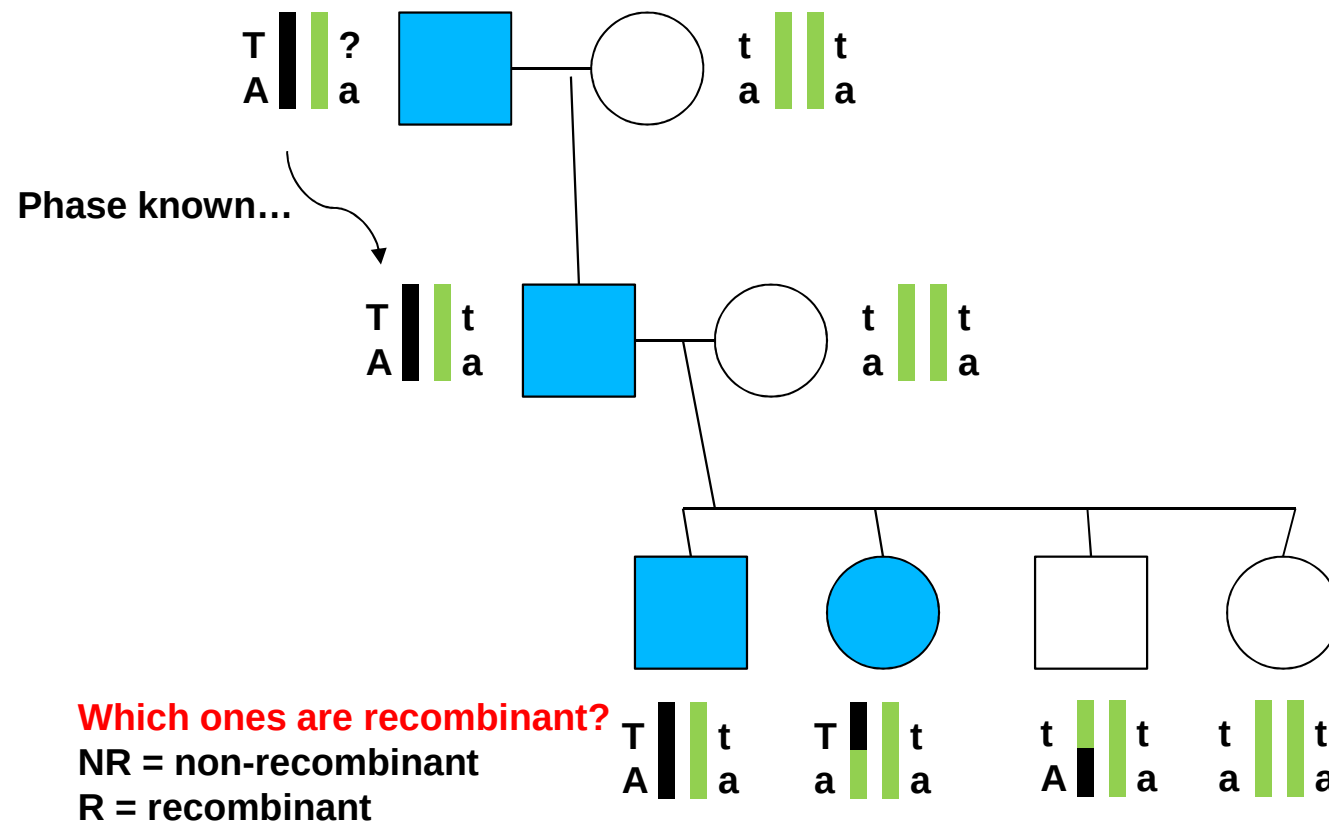
Idea behind linkage



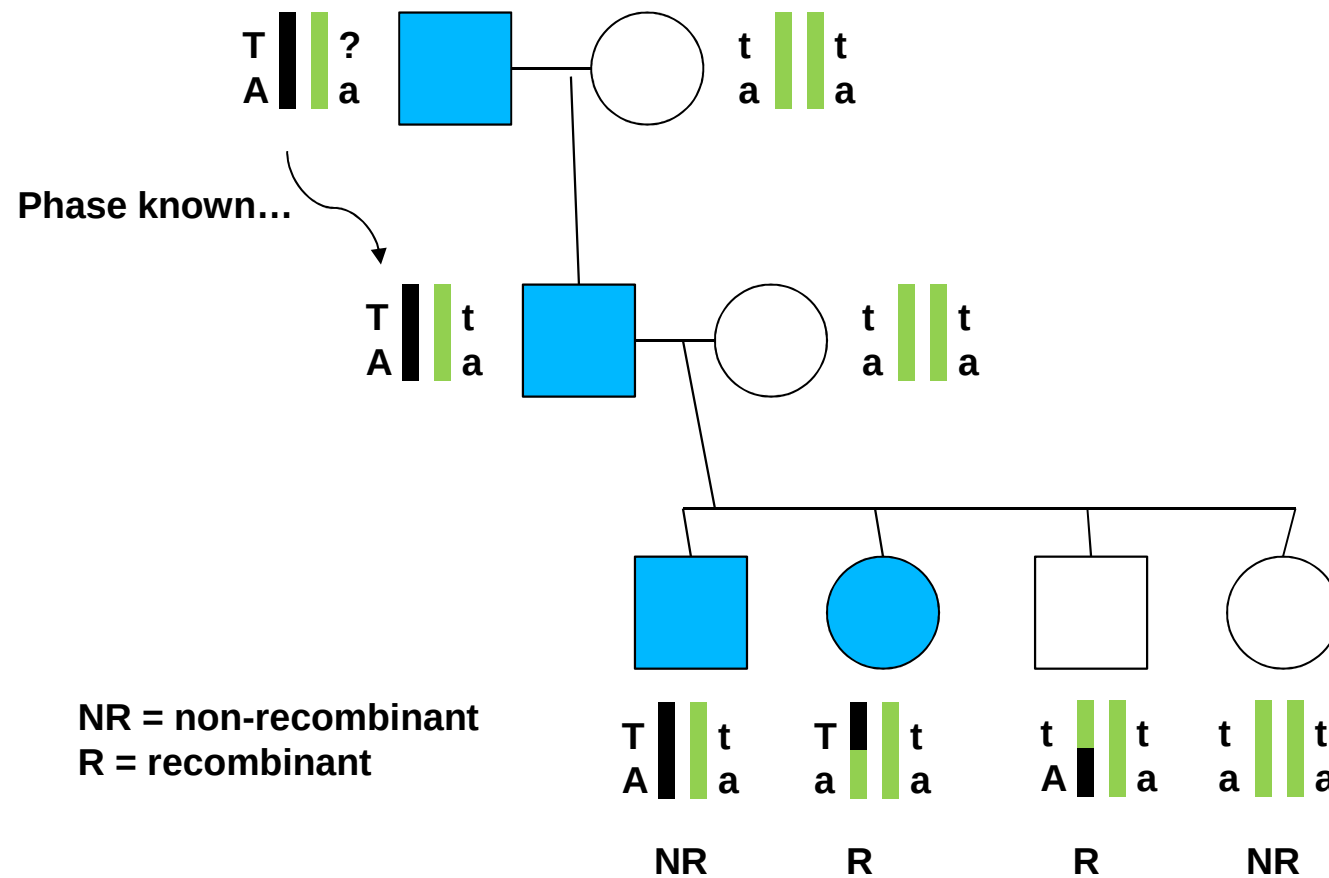
Linkage Analysis

- Narrow down position of disease gene
- No biological knowledge needed
- Genetic markers (not disease gene)
- It's all about the ***recombination***

PTC, dominant(-ish..., more later)



PTC, dominant(-ish..., more later)



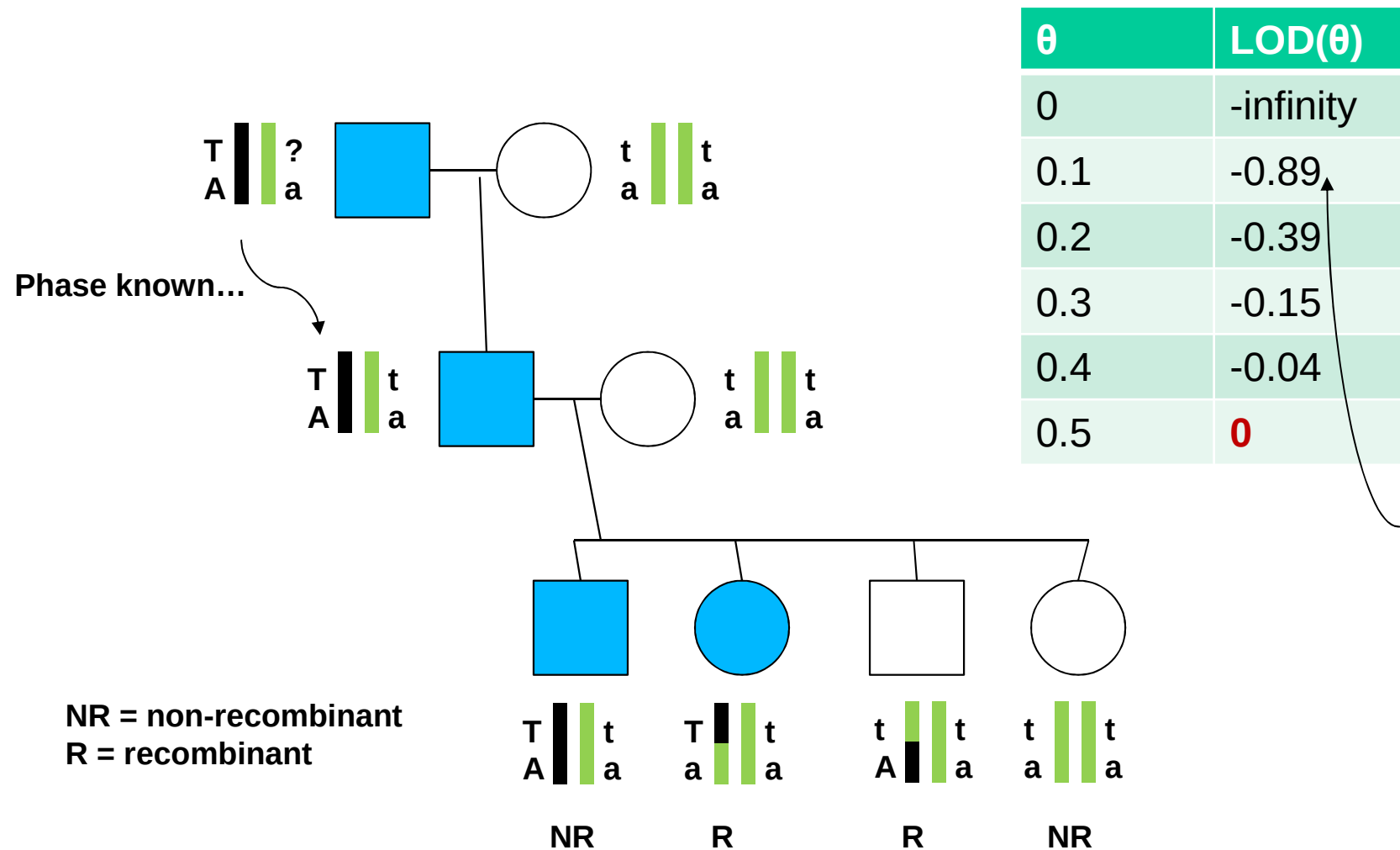
Linkage Analysis

LOD score based on recombination

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

- $\theta_{\text{null}} = 1/2$, completely unlinked have 50% chance of recombining
- The LOD score is the maximum of this function. You could maximize it easily here (take derivative, etc.), but this is difficult in more complicated examples (e.g., phase unknown, multi-generational pedigrees). So do a grid search over some possible values.
- Effectively a likelihood ratio test.
- < -2 evidence of lack of linkage
- > 3 evidence of linkage
- Not definitive in -2 to 3

PTC, dominant(-ish..., more later)



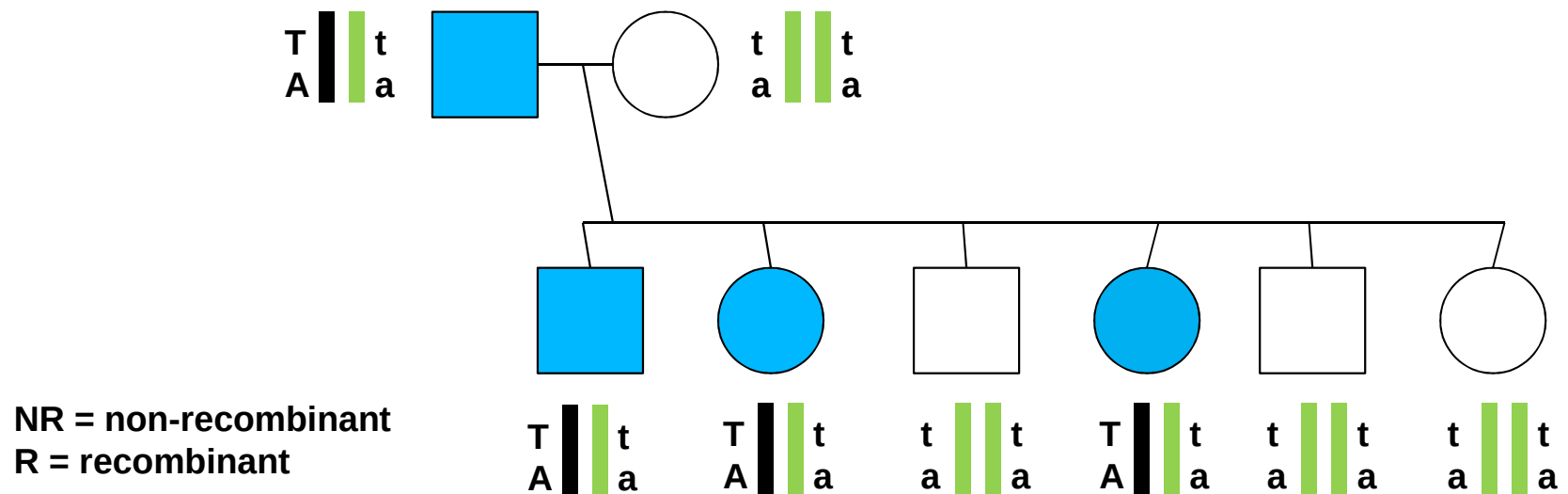
$$\text{LOD}(0.1) = \log_{10}[(0.1)^2(1-0.1)^2/0.5^4] = -0.89$$

PTC, dominant(-ish..., more later)

- LOD score=0
- Not definitive
- What now???

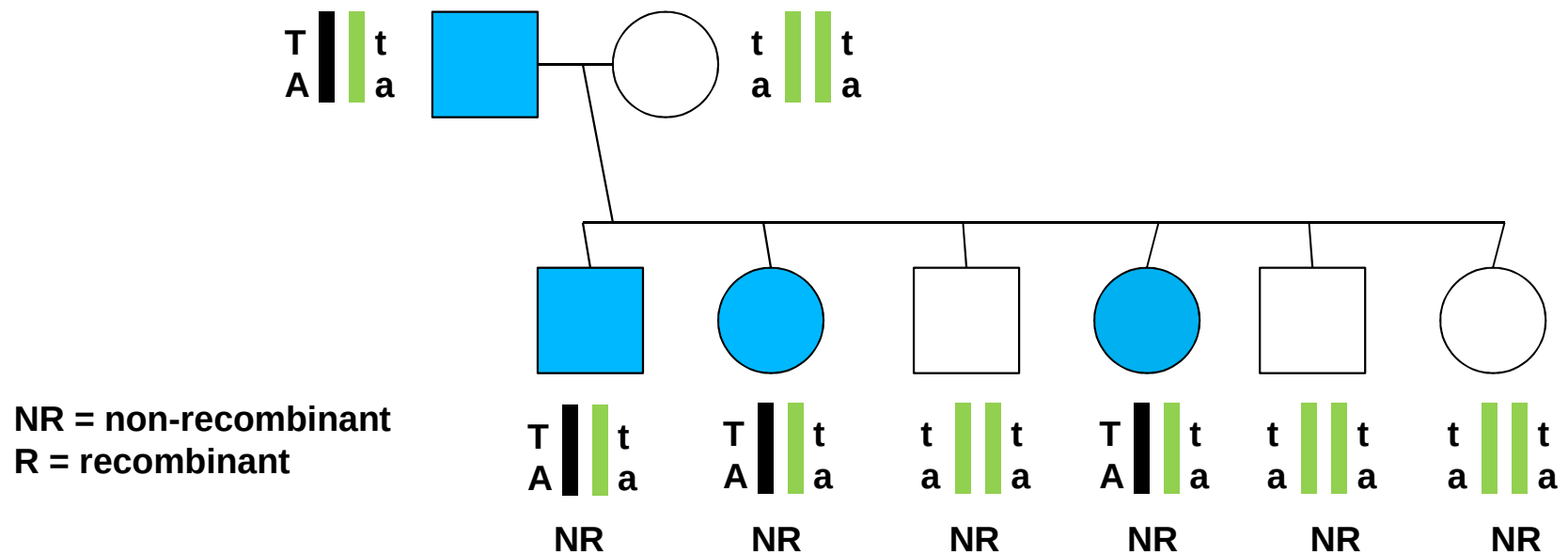
Get more families!!!

Assume phase known still...



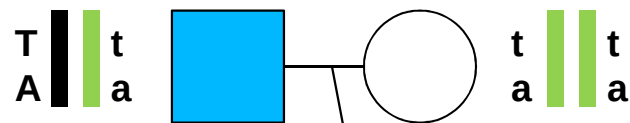
Get more families!!!

Assume phase known still...



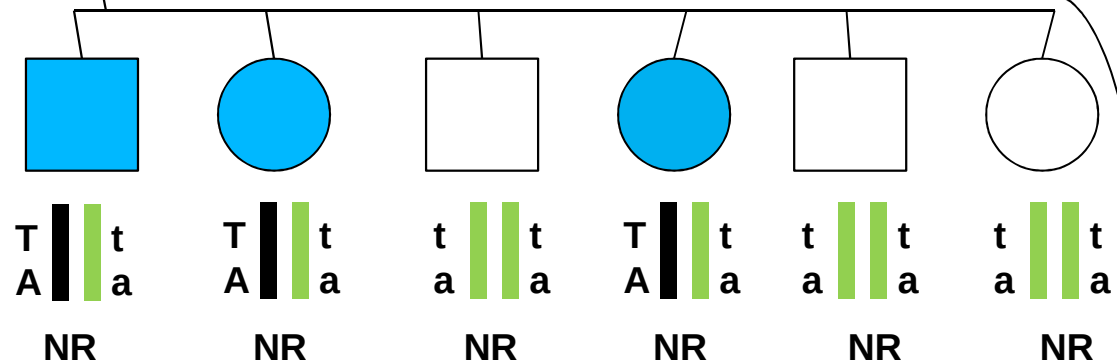
Get more families!!!

Assume phase known still...



θ	LOD(θ)
0	1.81
0.1	1.53
0.2	1.22
0.3	0.88
0.4	0.48
0.5	0

NR = non-recombinant
R = recombinant



$$\text{LOD}(0.1) = \log_{10}[(0.1)^0(1-0.1)^6/0.5^4] = 1.53$$

Add lod scores of families together...

θ	LOD(θ)
0	-infinity
0.1	-0.89
0.2	-0.39
0.3	-0.15
0.4	-0.04
0.5	0

+

θ	LOD(θ)
0	1.81
0.1	1.53
0.2	1.22
0.3	0.88
0.4	0.48
0.5	0

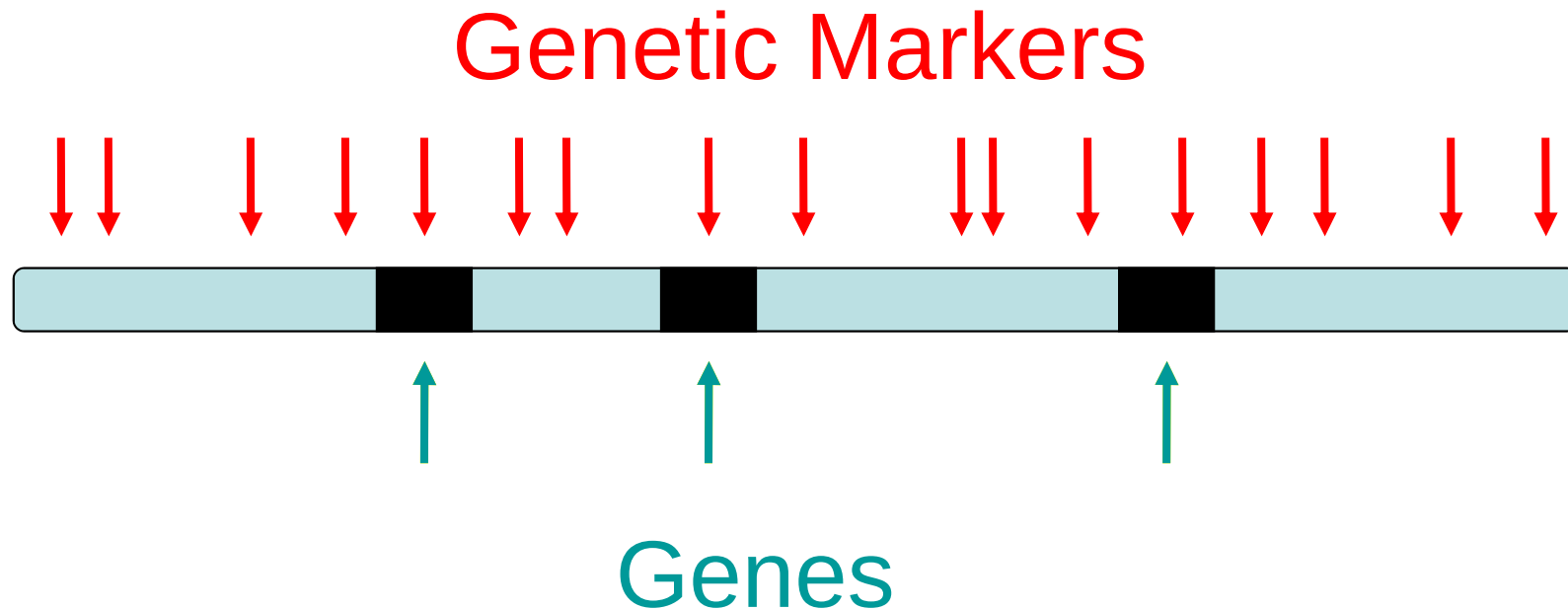
=

θ	LOD(θ)
0	-infinity
0.1	0.64
0.2	0.83
0.3	0.73
0.4	0.44
0.5	0

Now what?

- LOD score of 0.83, not definitive
- Suppose we had more families, LOD score of -4 (i.e., not that locus)?
- Test more loci?

Genome-wide linkage analysis

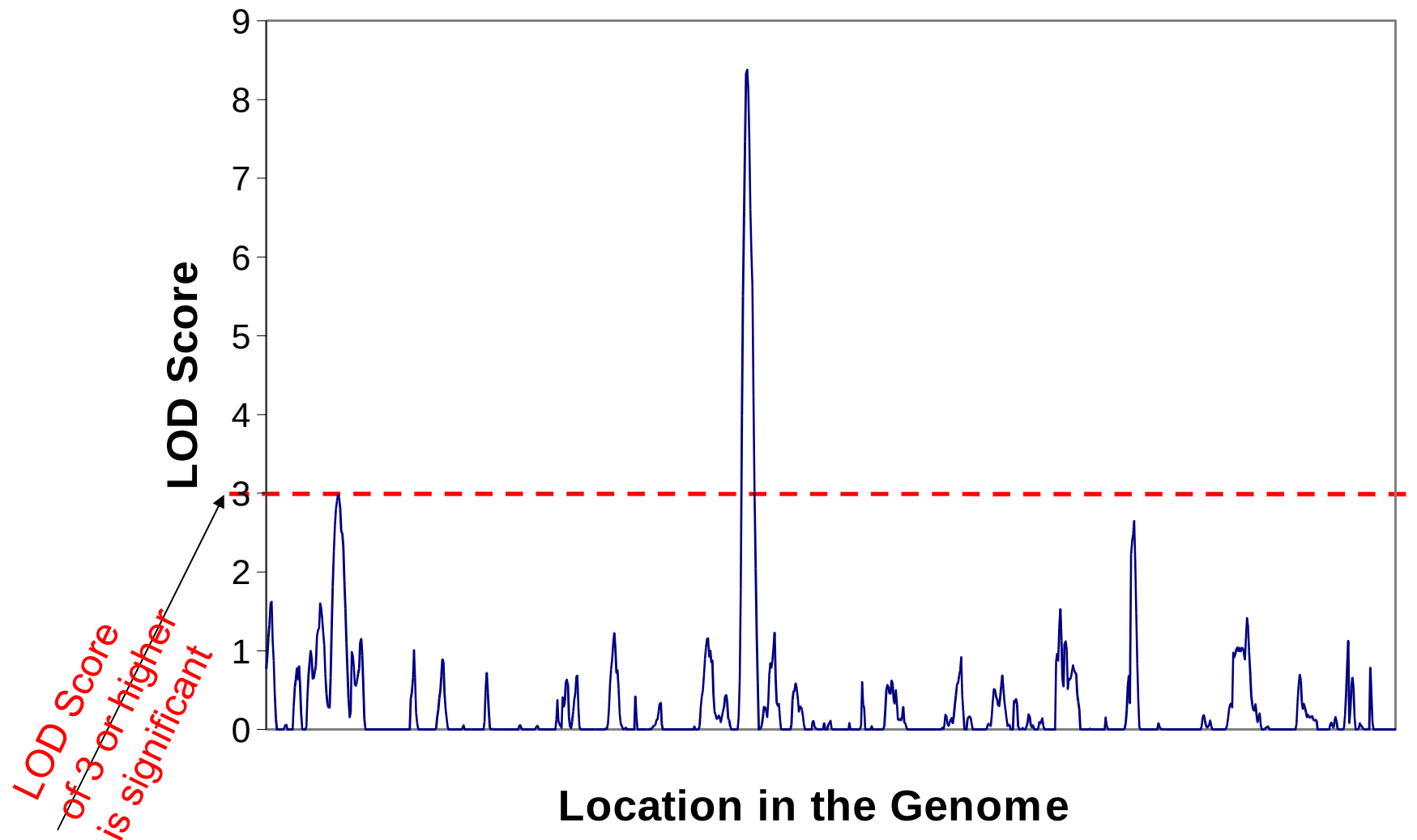


(Markers can be much less dense than genome-wide association studies later on...)

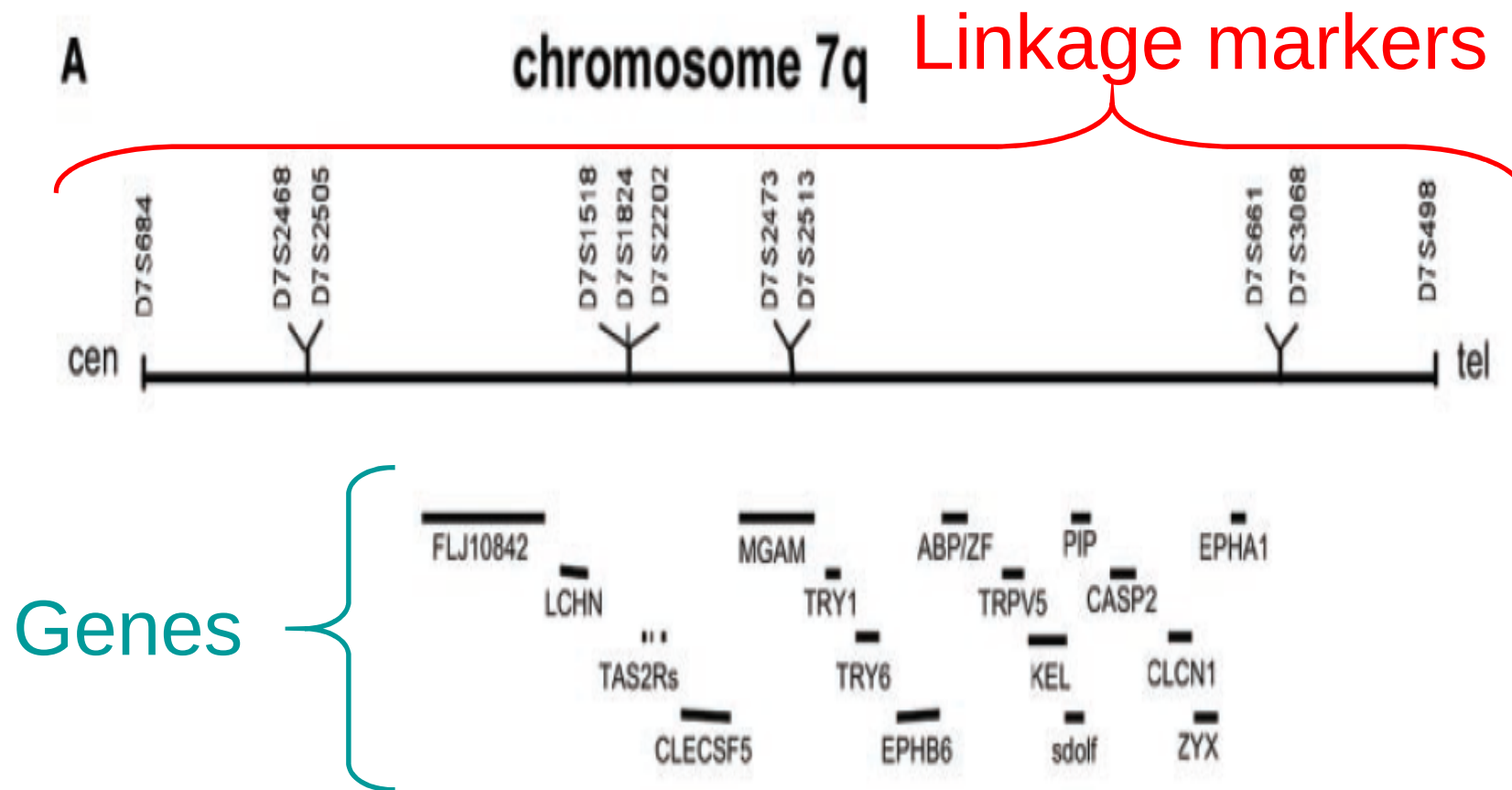
Human Linkage Analysis

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

PTC Linkage Analysis

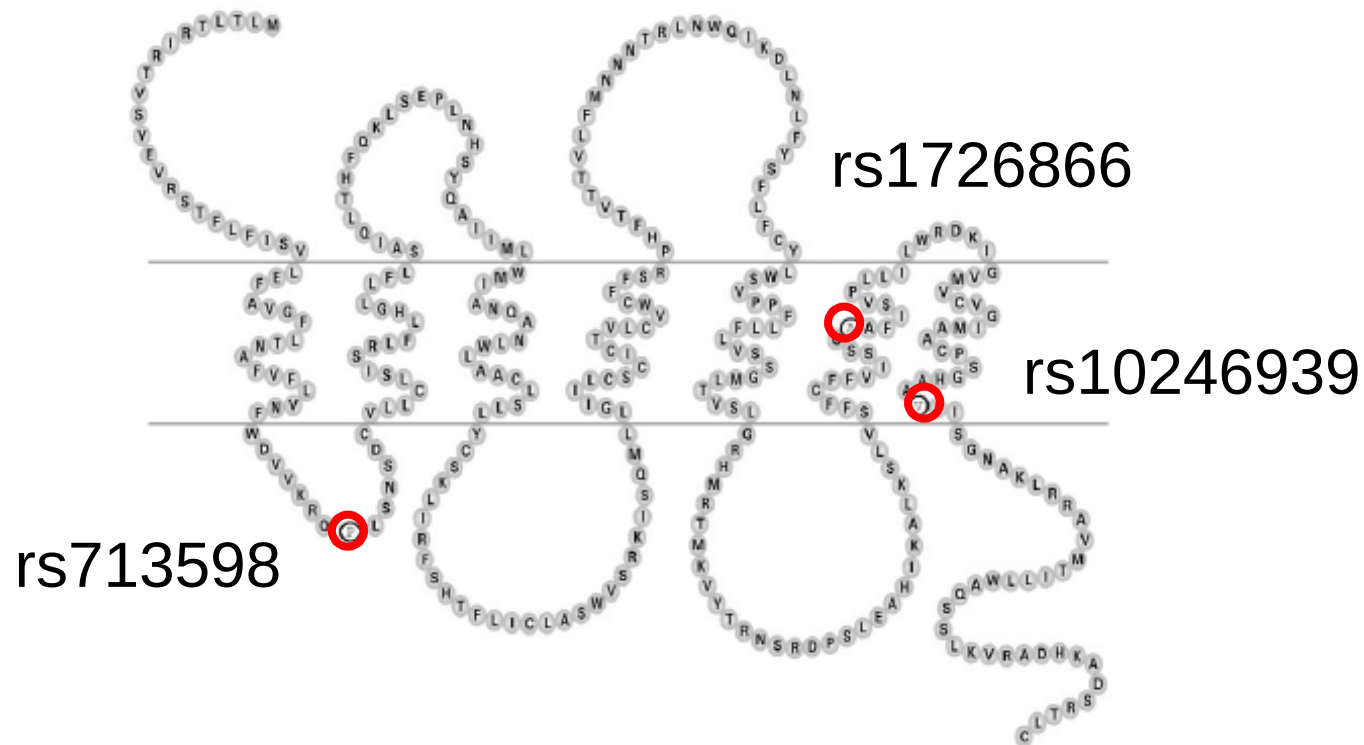


After you have found a linkage peak:
Fine Mapping (more details later)



Kim et al. Science 2003

TAS2R38 Receptor Structure



Kim et al. J Dent Res 2004

3 SNPs form 3 Haplotypes

Taster



Non-taster



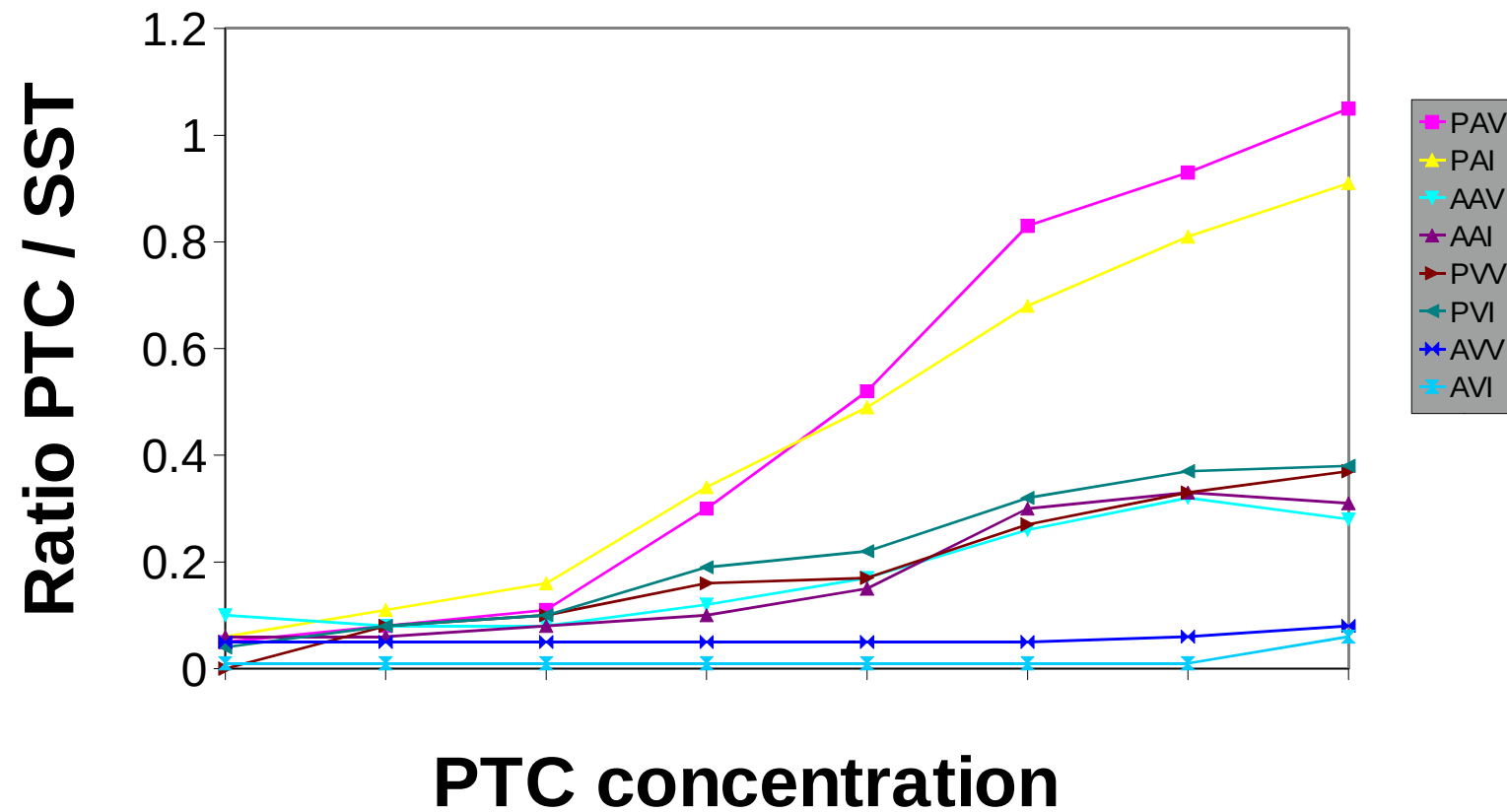
Rare



Predicted Effect of the 3 SNPs

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

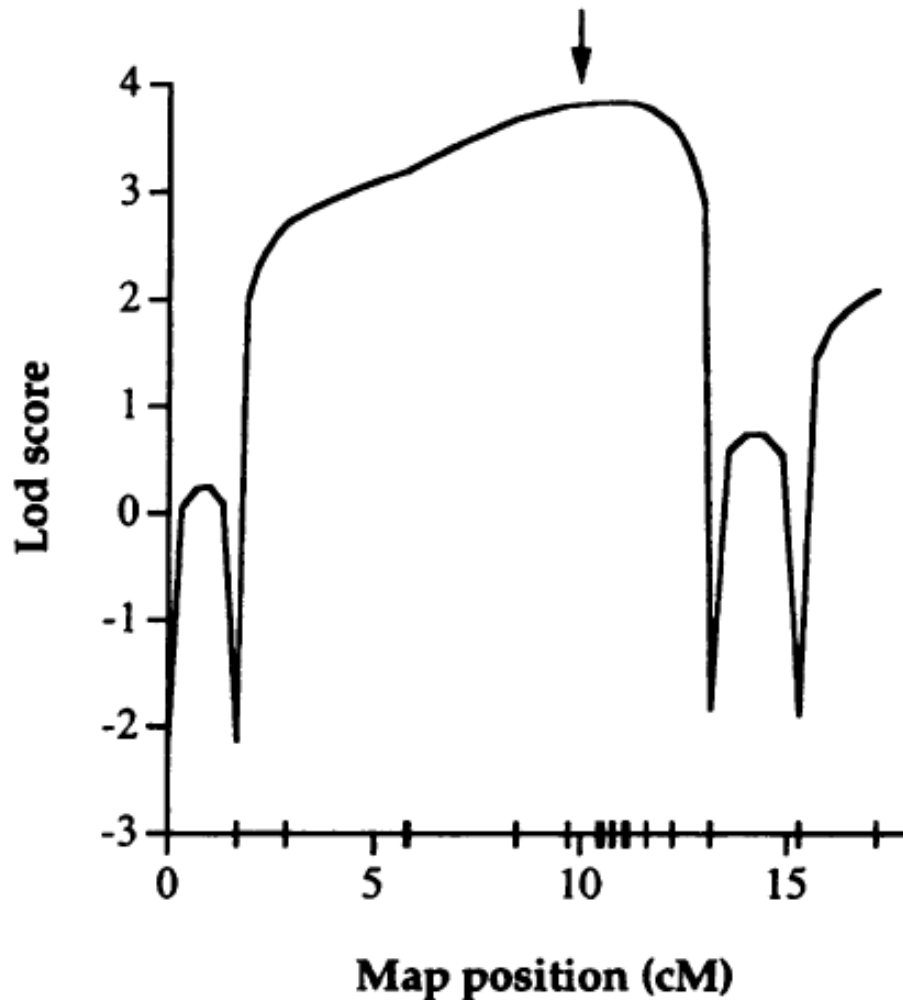
TAS2R38 Haplotype function in vitro



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

- BRCA2 Breast cancer
- HOXB13 G84E
 - Took so long to find because they were initially looking in the wrong place!

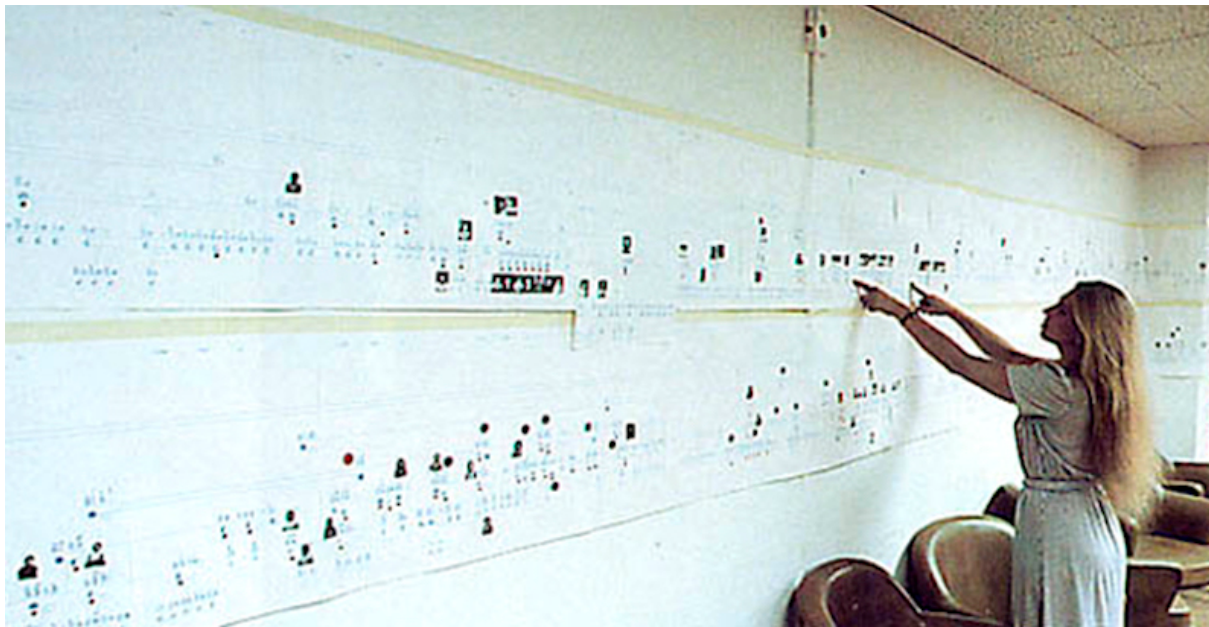
Multipoint Linkage Analysis



- 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!!!



Nancy Wexler examining a family tree with nearly 10,000 people, on a figure over 100 feet (30 meters) long!

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

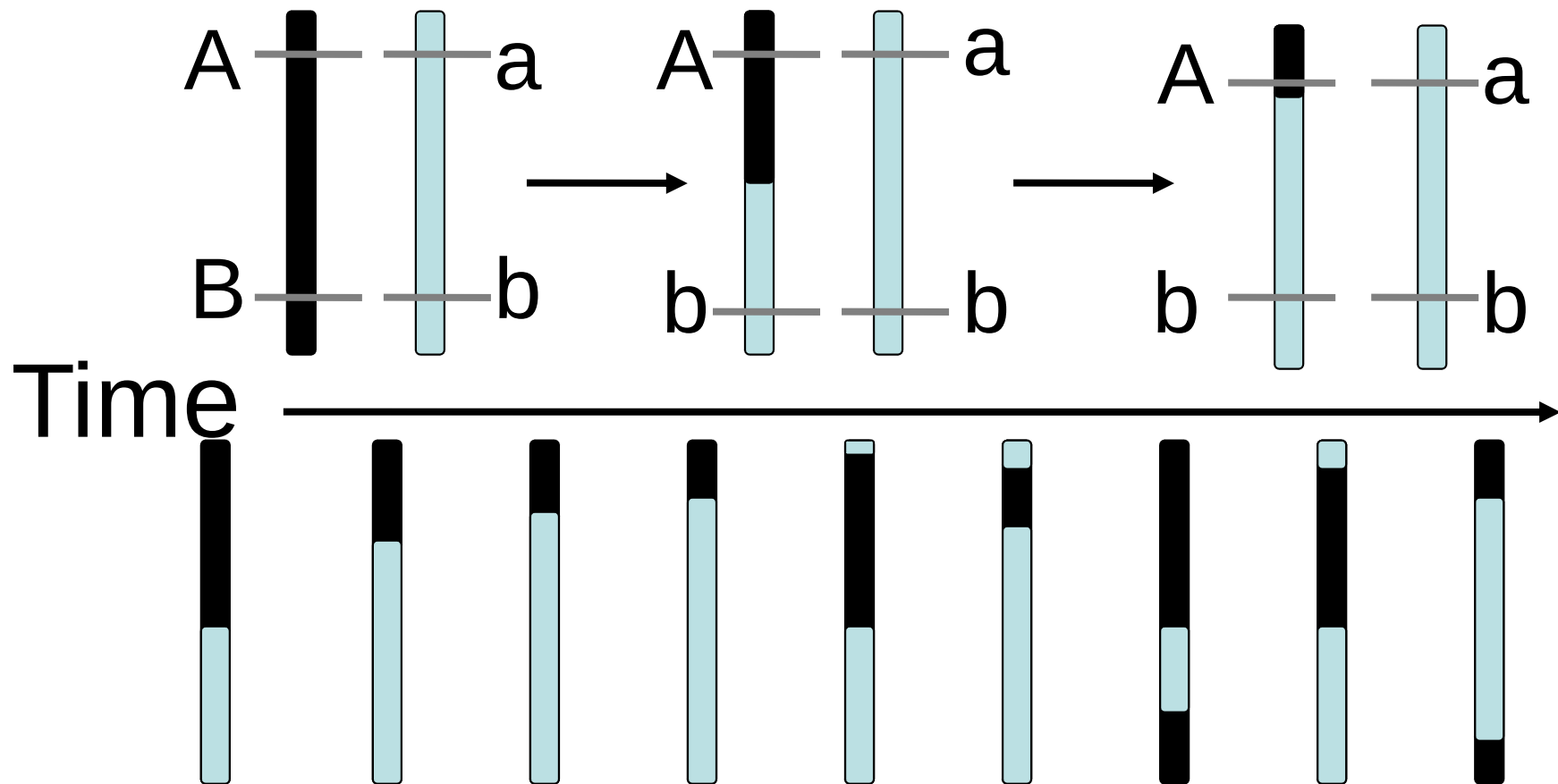
- Background/Review
- Family studies [PTC example]
- Linkage analysis [PTC example]
- Family-based association tests**
 - Trios/TDT
 - Discordant sibships

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 (next slide...)
- *Much smaller* family structures than for linkage!

Key Idea: Linkage and Linkage Disequilibrium

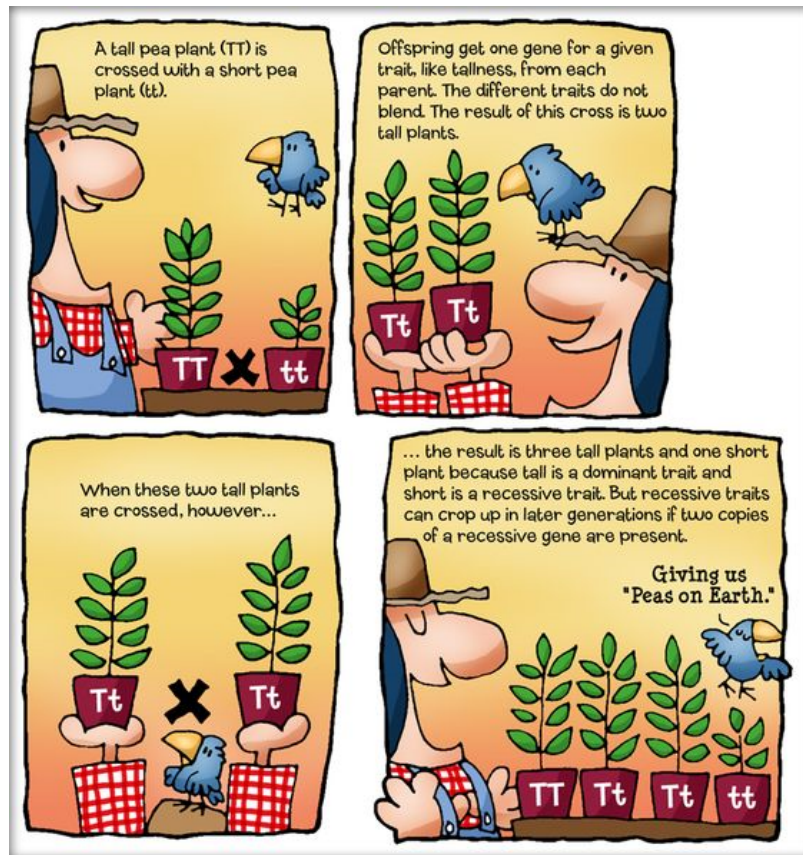
Association: Linkage Disequilibrium blocks (LD) are much finer resolution than just linkage (basically the effect of linkage over time, over many more generations) [Shuffle cards more and more...]



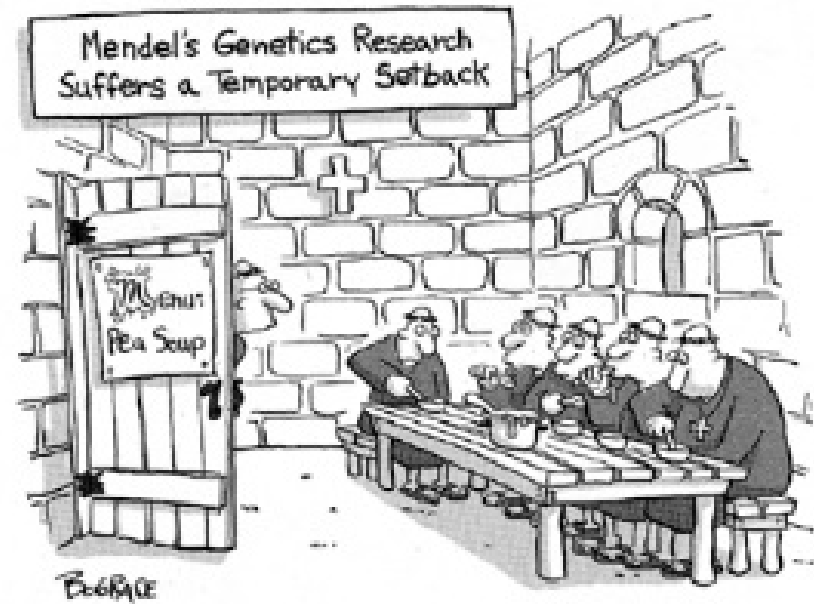
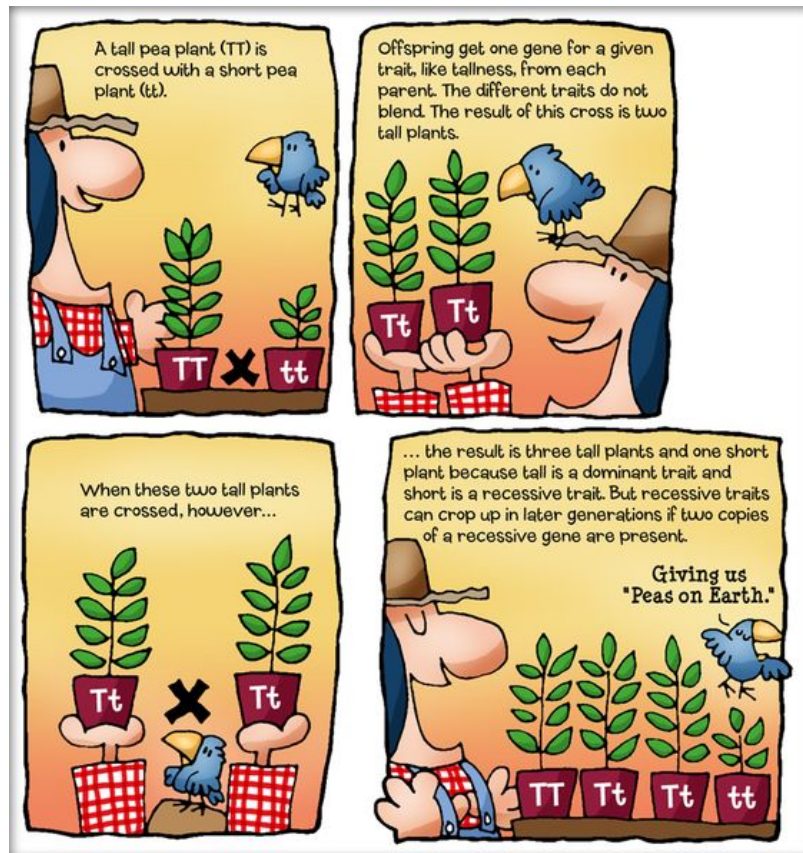
Mendel's laws

- 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)

Mendel's laws



Mendel's laws



Mendelian transmission: Ex

- E.g., parents are Aa, Aa:
 - $P(\text{offspring}=AA \mid \text{Mother}=Aa, \text{Father}=Aa)=?$
 - $P(\text{offspring}=Aa \mid \text{Mother}=Aa, \text{Father}=Aa)=?$
 - $P(\text{offspring}=aa \mid \text{Mother}=Aa, \text{Father}=Aa)=?$

M\F	A	a
A	AA	Aa
a	Aa	aa

Mendelian transmission: Ex

- E.g., parents are Aa, Aa:
 - $P(\text{offspring}=AA \mid \text{Mother}=Aa, \text{Father}=Aa)=1/4$
 - $P(\text{offspring}=Aa \mid \text{Mother}=Aa, \text{Father}=Aa)=1/2$
 - $P(\text{offspring}=aa \mid \text{Mother}=Aa, \text{Father}=Aa)=1/4$

M\F	A	a
A	AA	Aa
a	Aa	aa

Conditioning on
parents...

Mendelian transmission: Ex

- E.g., parents are AA, Aa:
 - $P(\text{offspring}=AA \mid \text{Mother}=AA, \text{Father}=Aa)=?$
 - $P(\text{offspring}=Aa \mid \text{Mother}=AA, \text{Father}=Aa)=?$
 - $P(\text{offspring}=aa \mid \text{Mother}=AA, \text{Father}=Aa)=?$

M\F	A	a
A	AA	Aa
A	AA	Aa

Mendelian transmission: Ex

- E.g., parents are AA, Aa:
 - $P(\text{offspring}=AA \mid \text{Mother}=AA, \text{Father}=Aa)=1/2$
 - $P(\text{offspring}=Aa \mid \text{Mother}=AA, \text{Father}=Aa)=1/2$
 - $P(\text{offspring}=aa \mid \text{Mother}=AA, \text{Father}=Aa)=0$

M\F	A	a
A	AA	Aa
A	AA	Aa

Mendelian transmission: Ex

- E.g., parents are AA, AA:
 - $P(\text{offspring}=AA \mid \text{Mother}=AA, \text{Father}=AA)=?$
 - $P(\text{offspring}=Aa \mid \text{Mother}=AA, \text{Father}=AA)=?$
 - $P(\text{offspring}=aa \mid \text{Mother}=AA, \text{Father}=AA)=?$

M\F	A	A
A	AA	AA
A	AA	AA

Mendelian transmission: Ex

- E.g., parents are AA, AA:
 - $P(\text{offspring}=AA \mid \text{Mother}=AA, \text{Father}=AA)=1$
 - $P(\text{offspring}=Aa \mid \text{Mother}=AA, \text{Father}=AA)=0$
 - $P(\text{offspring}=aa \mid \text{Mother}=AA, \text{Father}=AA)=0$

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 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)
- Composite null hypothesis: No linkage, no association

Trios: Transmission Disequilibrium Test (TDT)

		Non-transmitted parental allele	
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

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

PAUSE: Talk to your neighbor and construct a rough test! Try starting with (1) what cells are important? (hint, not all of them) (2) Think about whether you might want to add or subtract the two remaining values. (3) Is this like a test with matched pairs of subjects? EDUCATIONAL RESEARCH: Making you think/struggle about it will make you remember more!

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
Transmitted parental allele	A	A	a	 AA
	a	2	0	
		0	0	

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

Both parents homozygous

		Non-transmitted parental allele		aa-aa
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	AA-aa
Transmitted parental allele	A	1	0	 Aa
	a	0	1	

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

One parent heterozygous

		Non-transmitted parental allele		AA-Aa		
		A	a			
Transmitted parental allele	A	1	1	AA, Aa		
	a	0	0	.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		Aa - aa		
		A	a			
Transmitted parental allele	A	0	1	Aa, aa		
	a	0	1	.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		Aa - Aa
		A	a	
Transmitted parental allele	A	0	2	AA, Aa, aa
	a	0	0	.25, .5, .25 ← Pr

		Non-transmitted parental allele				Non-transmitted parental allele	
		A	a			A	a
Transmitted parental allele	A	0	1	Transmitted parental allele	A	0	0
	a	1	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		Insulin Dependent Diabetes Mellitus (IDDM)
		A	a	
Transmitted parental allele	A	?	78	
	a	46	?	

- 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

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)*...
 - *"If you cannot get rid of the family skeleton, you may as well make it dance" (G.B. Shaw) → Nan Laird's favorite quote on this... (big proponent of FBAT)*

Discordant sibships

- Conditional logistic regression
 - $P(Y_1=1|Y_1+Y_2=1, g_1, g_2, \dots)$
 - 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 (i.e., as odds ratios)
- You will go through an example in the homework

Summary

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