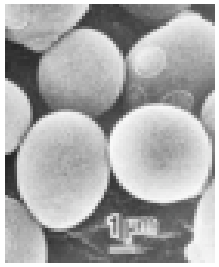
Genome 371, 1 March 2010, Lecture 13

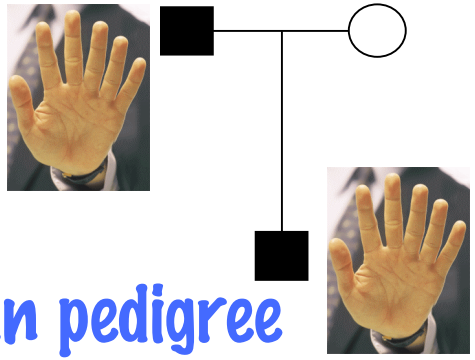
Association Mapping

- Mendelian versus Complex Phenotypes
- How to Perform an Association Study
- Why Association Studies (Can) Work
- Introduction to LOD score analysis

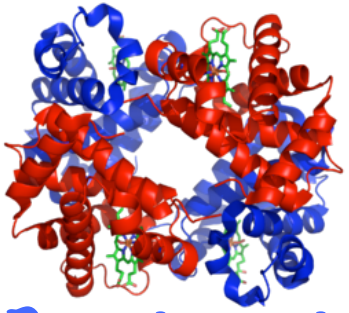
Common theme: linking genotype & phenotype



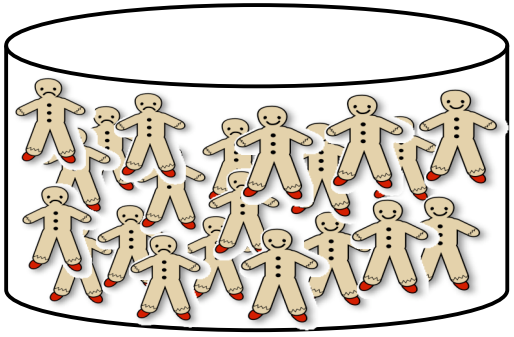
Mutant identified
in a model organism



Human pedigree
segregating a trait



Protein acting in
a biological process

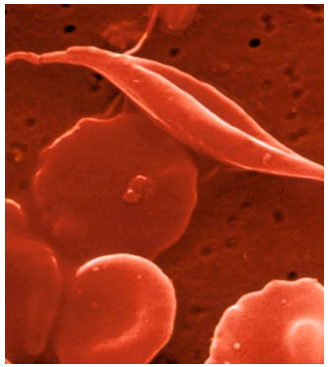


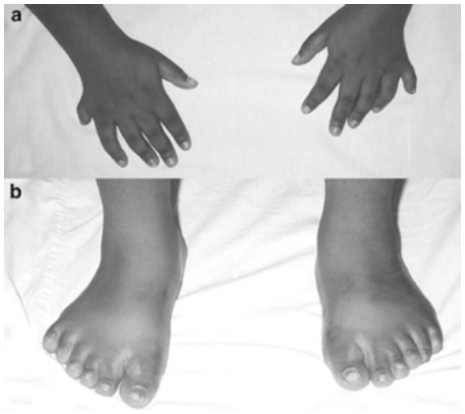
Association study

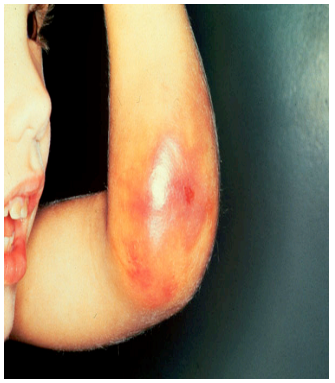
```
 946  ATT GTC TGT AGC CGA TTG GAG GAG TAC AAC AGC CAT
1009  GGA CCT TTA CGG CGT AAT CCT GGA AAC CAT GAC AAA
1072  GCT GAT GTA GAA TTT TGC CTG AGT TTG ACC CAA TAT
1135  AAT TTC AGC TTT AGA AAT ACA CTG GAA GGA TTT GCT
1198  TCT CAA AGC AGC ATG CAC AAT GCC TTG CAC ATC TAT
1261  GGA TCT GCC AAC GAT CCT ATC TTC CTT CTT CAC CAT
1324  TGG CTC CGA AGG CAC CGT CCT CTT CAA GAA GTT TAT
```

Sequence analysis

Mendelian Phenotypes

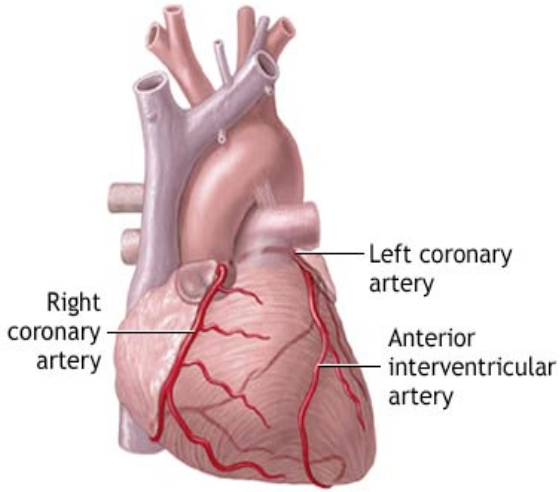


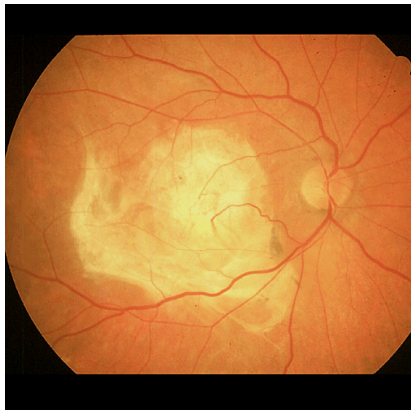




- Single gene
- Follows clear patterns of Mendelian Inheritance
- Strong relationship (correlation) between genotype and phenotype

Complex (aka “multifactorial”) Phenotypes





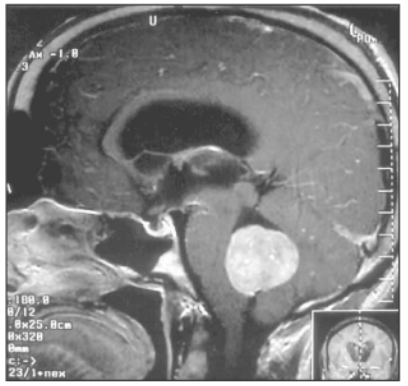


Fig 5. Sagittal T1 weighted MRI after contrast injection showing a midline cerebellar mass with posterior compression of the brain stem.

- Multiple genes
- Familial aggregation, but no clear pattern of Mendelian Inheritance
- Weak relationship (correlation) between genotype and phenotype

Complications of Genotype-Phenotype Relationship

Expressivity and penetrance can vary

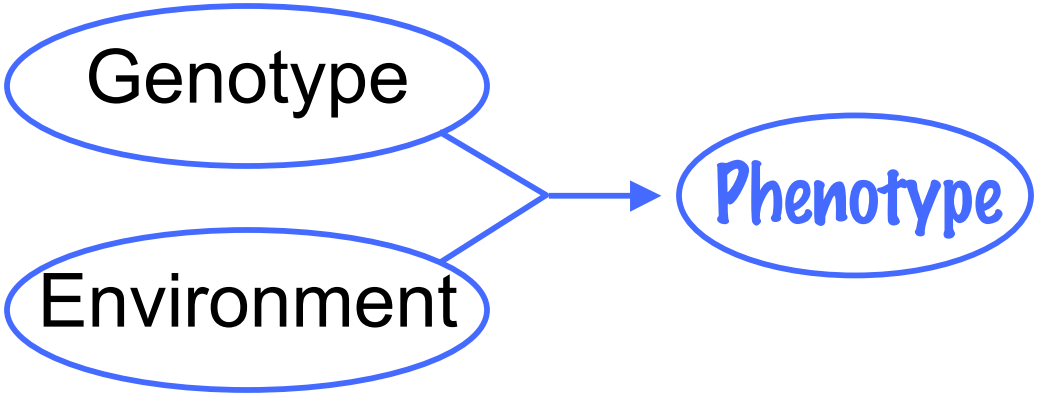
Fraction of individuals with a genotype exhibit the phenotype expected of that genotype

The degree to which a genotype is expressed — e.g., different individuals inheriting the same disease gene may be affected to different extents

Important Concept: disease causing -vs- susceptibility alleles

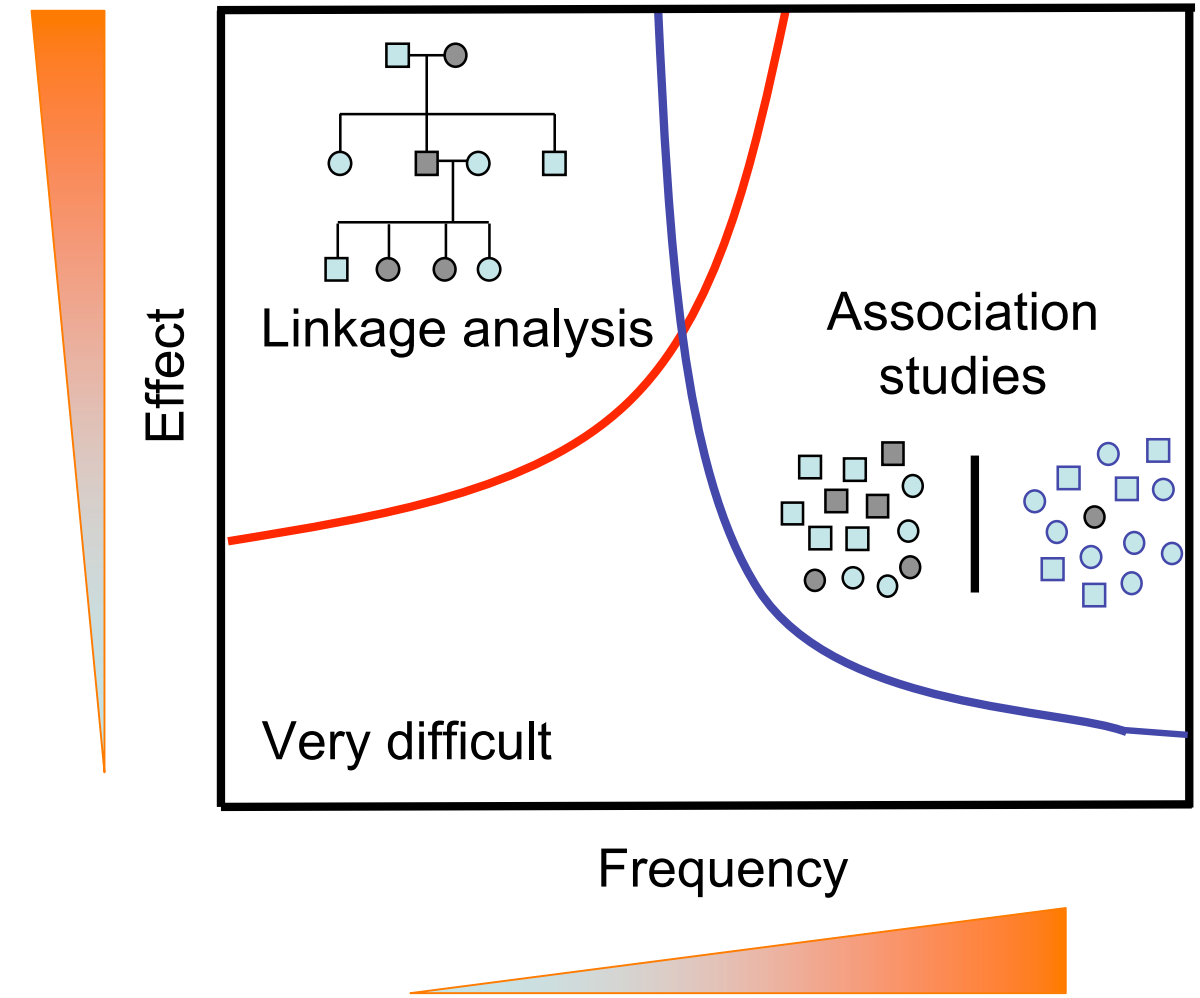
Complications of Genotype-Phenotype Relationship

Hydrangea



The genotype is an indication of the potential phenotype.
The actual phenotype depends on the environment also.

Why Are We Talking About Association Studies?



modified from D. Altschuler

How Do We Measure Effect Size?

Effect size - measure of how much familial aggregation is seen for a trait (disease) of interest

Familial relative risk measures the risk of disease in someone with an affected relative compared to the risk of the general population

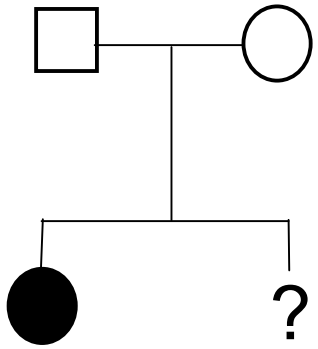
$$\lambda_S = \frac{\text{P(sib 2 has a disease | sib 1 has the disease)}}{\text{P(diseased individual in the general population)}}$$

Example: λ_S (Sibling relative risk) of Cystic Fibrosis

CF is a rare autosomal recessive mendelian disease

Population frequency ~ 8/100,000

What is the λ_S of CF?



$$\lambda_S = \frac{\text{P(sib 2 has a disease | sib 1 has the disease)}}{\text{P(diseased individual in the general population)}}$$

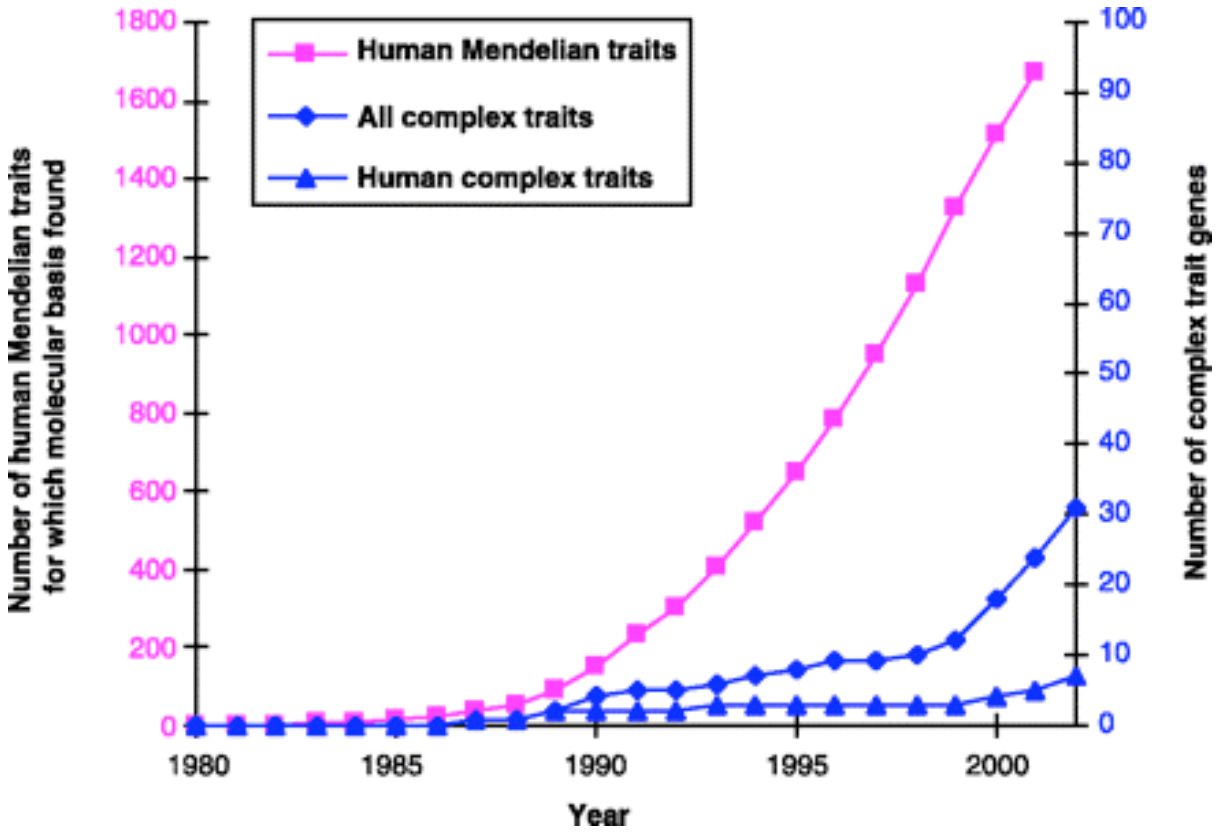
$$= \frac{1/4}{8/100{,}000}$$

$$= 3{,}125$$

λ_S of Some Complex Diseases

Disease	λ_S
Celiac disease	60
Multiple sclerosis	20
Type I Diabetes	15
Testis cancer	8
Breast cancer	2

Complex Diseases Have Proven Refractory to Linkage Analysis



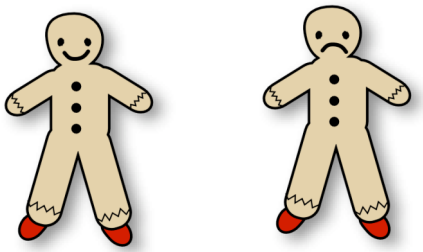
Association Studies

- Population based gene mapping method

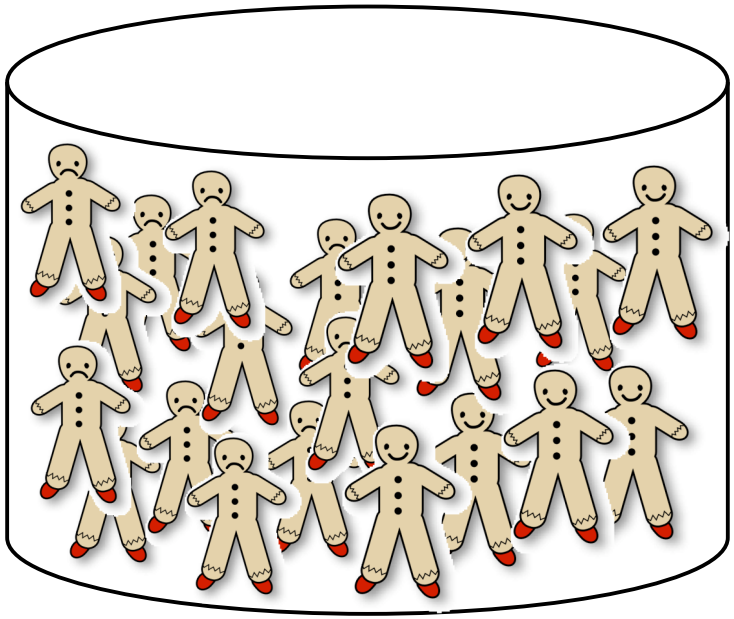
- Collect unrelated individuals with a disease (cases) and individuals without a disease (controls)

- Goal is to find an association between marker alleles and disease

A Simple Association Study

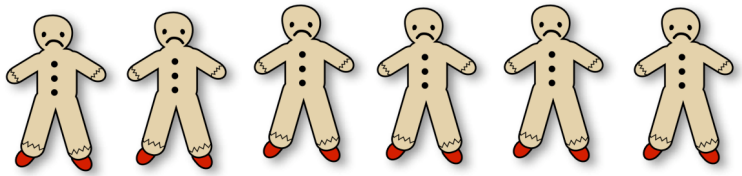


What is the genetic basis of the frowny faced gingerbread man (FFGM) disease?

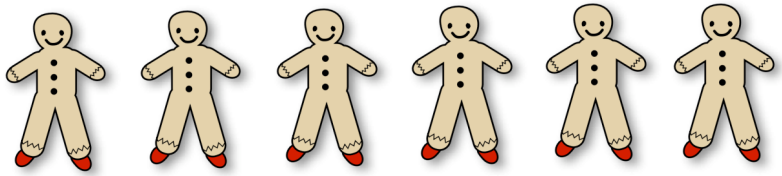


Population

1. Collect FFGM cases from the population

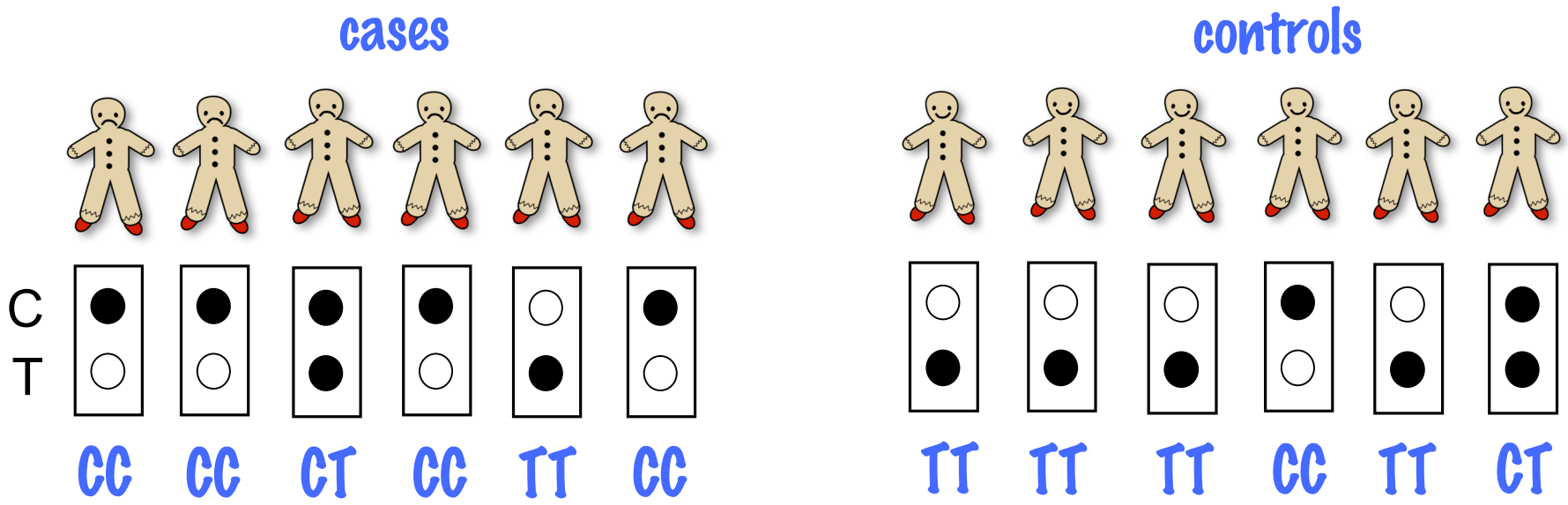


2. Collect controls from the population



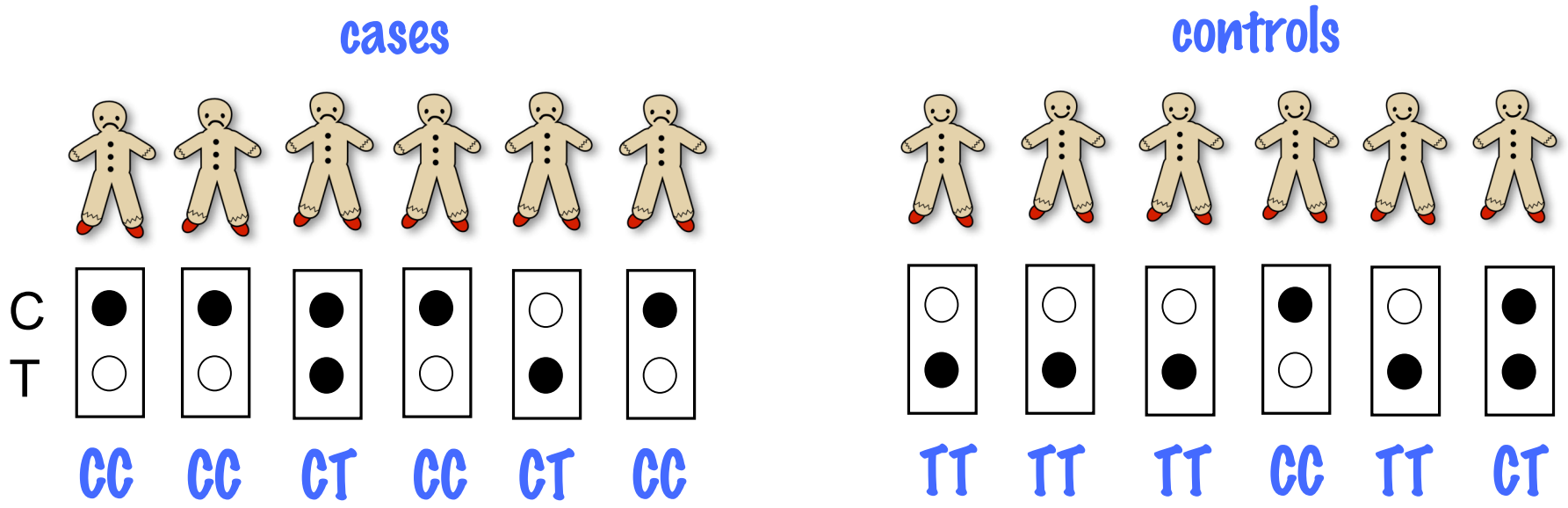
A Simple Association Study

3. Genotype a genetic marker of interest (SNP1) in:



A Simple Association Study

4. Compare frequency in cases and controls



C = 9/12 = 0.75

T = 3/12 = 0.25

C = 3/12 = 0.25

T = 9/12 = 0.75

Different… Yes!
Significant.. Maybe!

How Can We Determine Significance?

	Cases	Controls	
C	9	3	12
T	3	9	12
	12	12	24

2 x 2 "Contingency Table"

A case for our old friend $\chi^2 = \sum \frac{(O - E)^2}{E}$

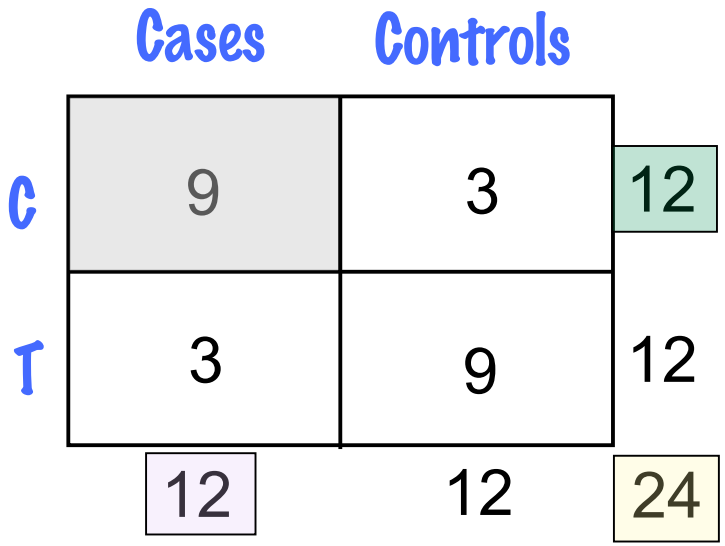
What is the Expectation? What's the Null Hypothesis?

6 Cases 6 Controls

12 C alleles 12 T alleles

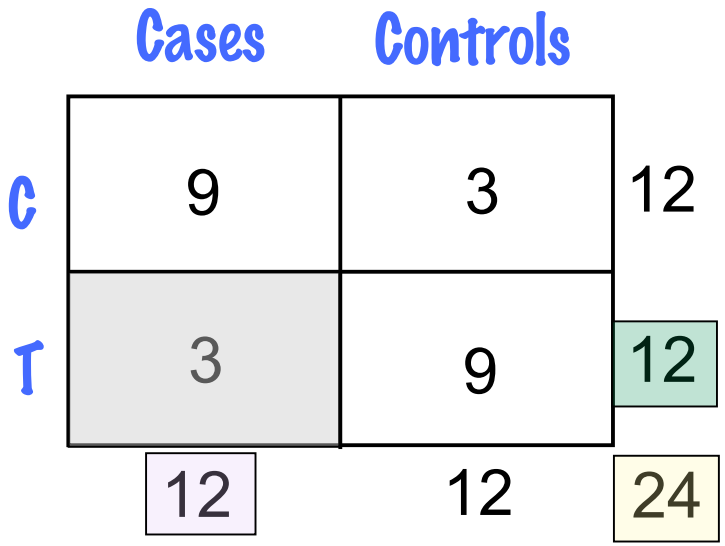
	Cases	Controls	
C	6	6	12
T	6	6	12
	12	12	24

Calculating χ^2 for an Association Study



	Obs	Exp	$(O-E)^2$	$(O-E)^2/E$
Cases C	9	(12 x 12)/24 = 6		
Cases T	3			
Controls C	3			
Controls T	9			

Calculating χ^2 for an Association Study



	Obs	Exp	$(O-E)^2$	$(O-E)^2/E$
Cases C	9	(12 x 12)/24 = 6		
Cases T	3	(12 x 12)/24 = 6		
Controls C	3			
Controls T	9			

Calculating χ^2 for an Association Study

	Cases	Controls	
C	9	3	12
T	3	9	12
	12	12	24

	Obs	Exp	$(O-E)^2$	$(O-E)^2/E$
Cases C	9	(12 x 12)/24 = 6	9	1.5
Cases T	3	(12 x 12)/24 = 6	9	1.5
Controls C	3	(12 x 12)/24 = 6	9	1.5
Controls T	9	(12 x 12)/24 = 6	9	1.5

$\chi^2 = 6.0$

χ^2 table

P	0.995	0.975	0.900	0.500	0.100	0.050	0.025	0.010	0.005
df									
1	0.000	0.000	0.016	0.455	2.706	3.841	5.024	6.635	7.879
2	0.010	0.051	0.211	1.386	4.605	5.991	7.378	9.210	10.597
3	0.072	0.216	0.584	2.366	6.251	7.815	9.348	11.345	12.838
4	0.207	0.484	1.064	3.357	7.779	9.488	11.143	13.277	14.860
5	0.412	0.831	1.610	4.351	9.236	11.070	12.832	15.086	16.750
6	0.676	1.237	2.204	5.348	10.645	12.592	14.449	16.912	18.548

Find appropriate df row

P value

Find closest χ^2 value

What does this P value mean?

Would expect a deviation from the hypothesis of this magnitude (from chance alone) about 1% of the time

Therefore, reject the null hypothesis (boo-ya)

Why Do Genetic Association Studies Work?

- SNP1 is associated with FFGM

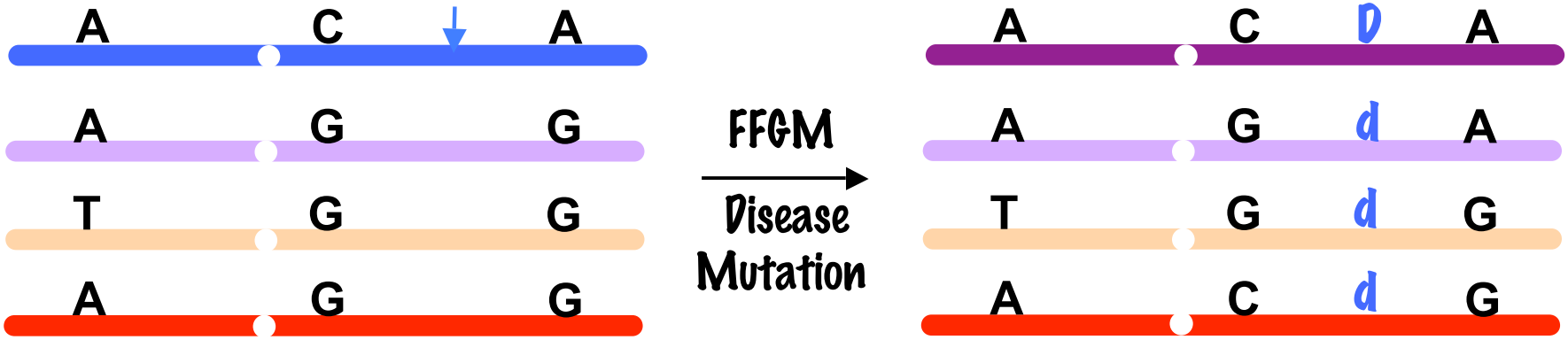
  C allele more frequent in FFGM cases compared to controls

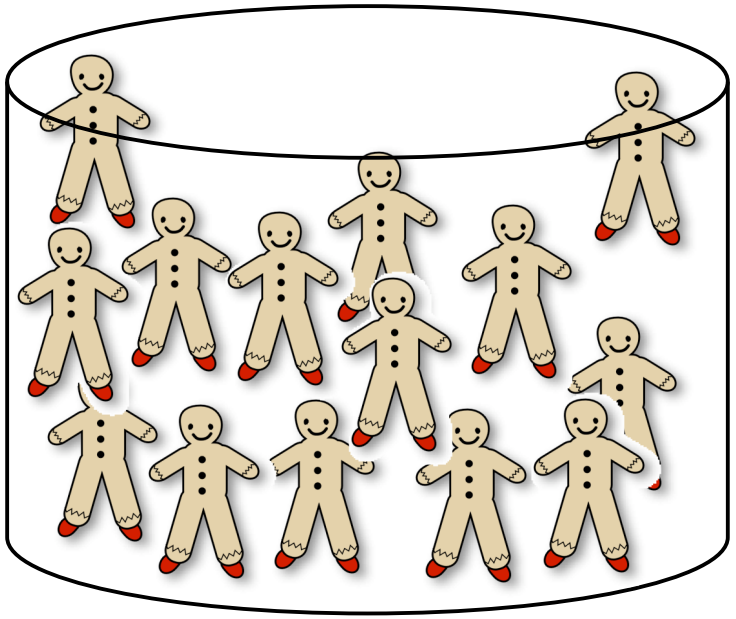
- Does this mean the C allele causes FFGM?

  No - it could just be on the same haplotype as the FFGM disease allele

  Alleles between two or more loci can be correllated

Why Genetic Association Studies Work: Part I





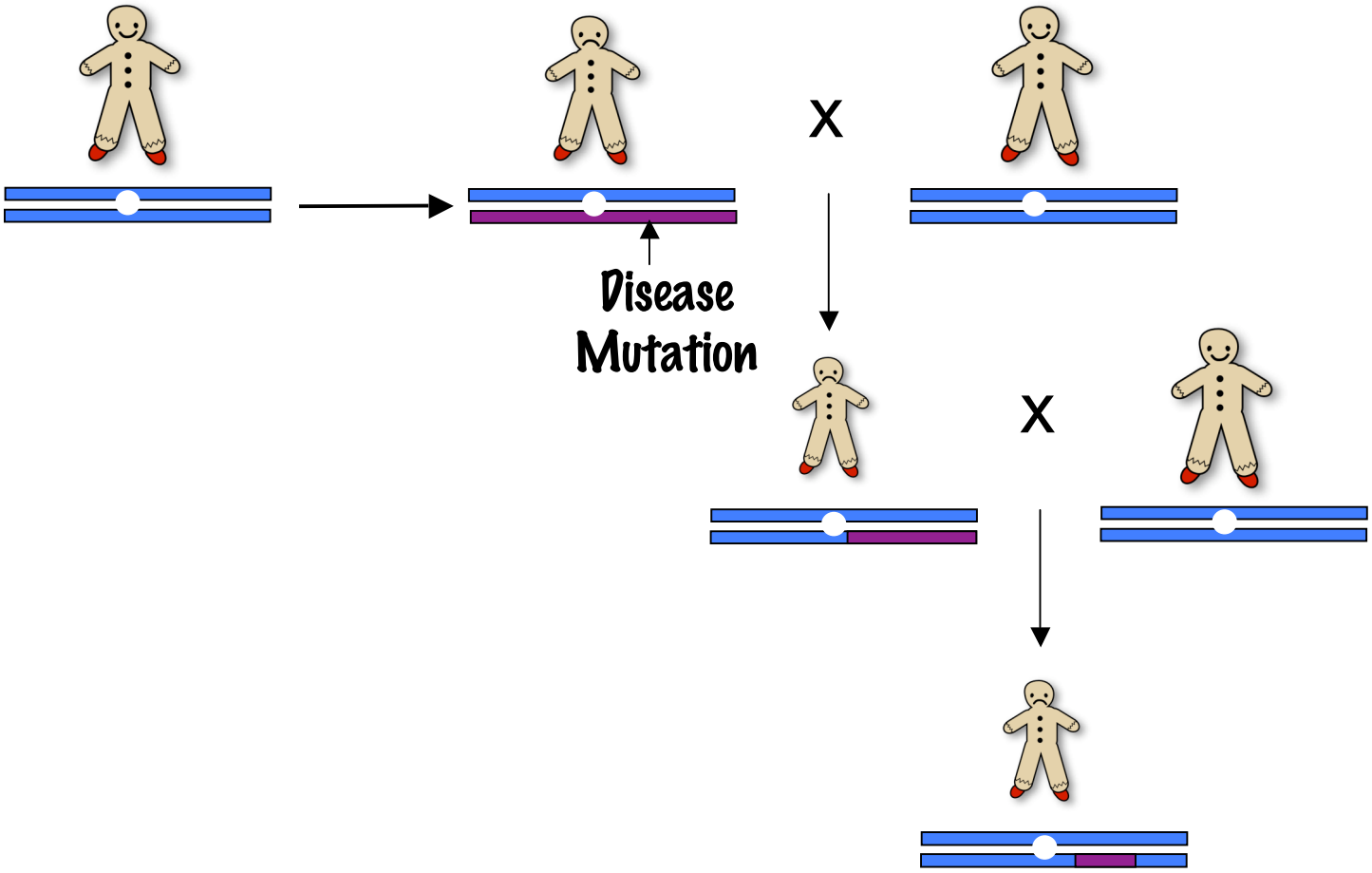
Population

Mutations arise on particular halotype backgrounds

Creates association among alleles

This is why association studies "work"

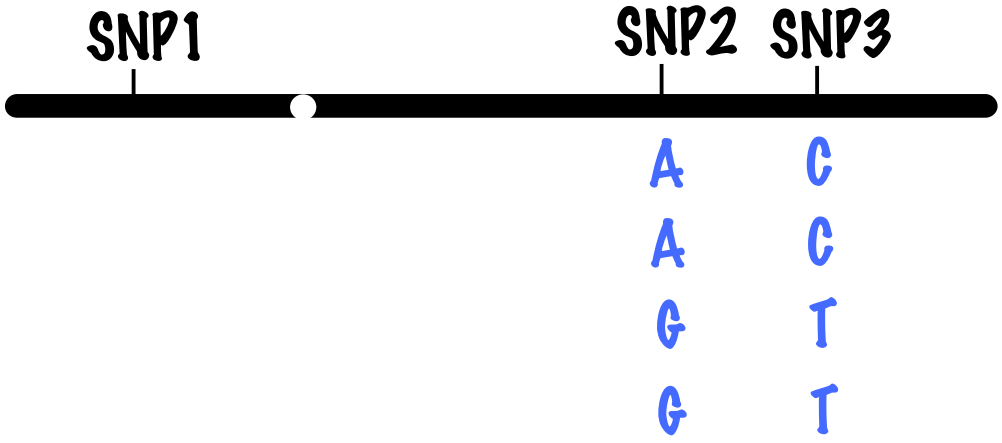
Why Genetic Association Studies Work: Part II



How could we ever map anything if allelic associations extend over a whole chromosome?

The Case of Linkage -v- Association

What is the primary difference between the concepts of linkage and association?



Linkage: relationship between two or more loci

Association: relationship between the *alleles* of two or more loci

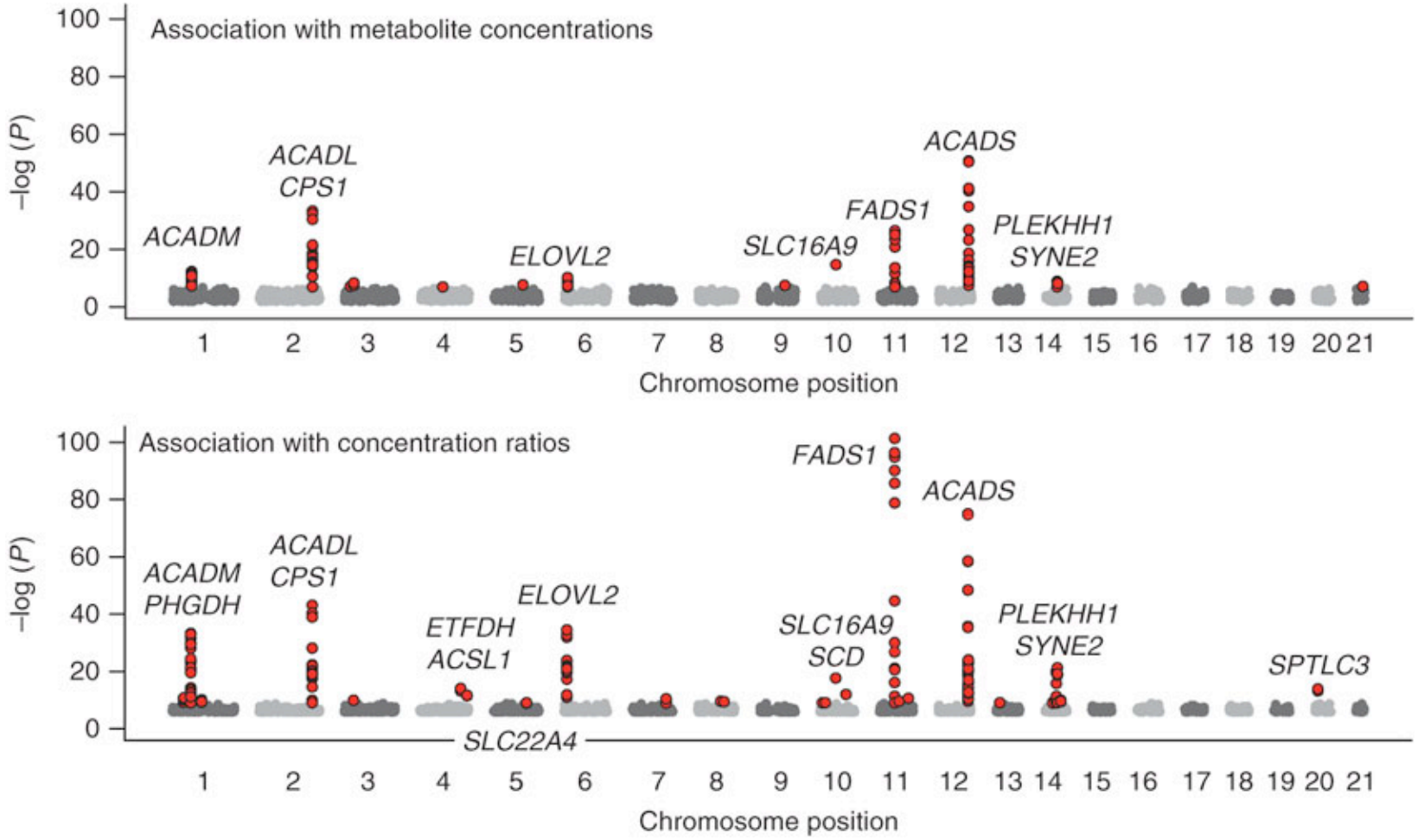
Beyond the Basics

Genome-Wide Association Studies (GWAS)

"A genome-wide perspective of genetic variation in human metabolism"

Nat. Genet. 2009, **42:137**

Measured 163 metabolites in 1,809 individuals and performed a GWAS:

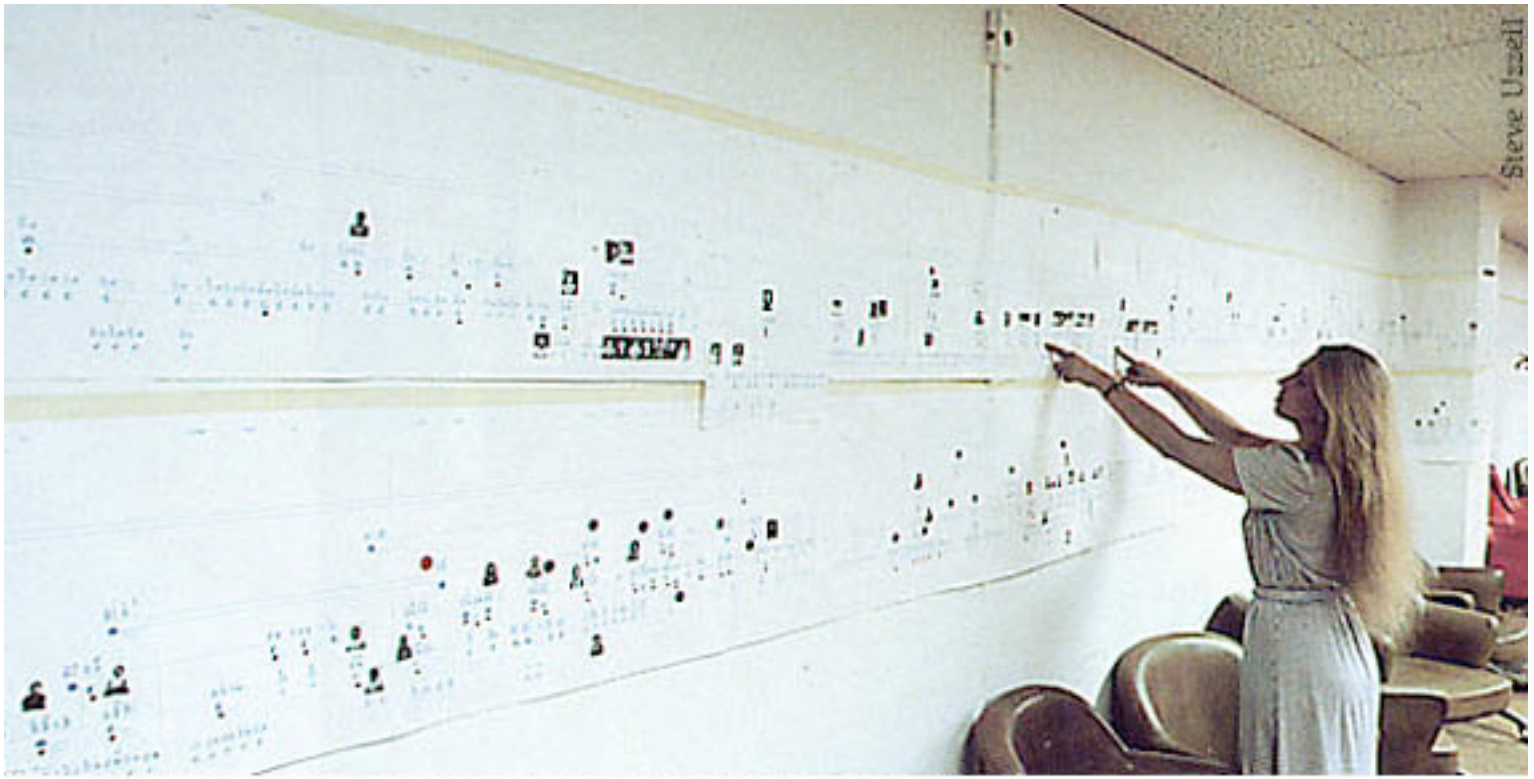


LOD Score Analysis in Humans

Gene mapping in humans

»Mapping using pedigrees—LOD score analysis, haplotype analysis

»Mapping using populations—association studies

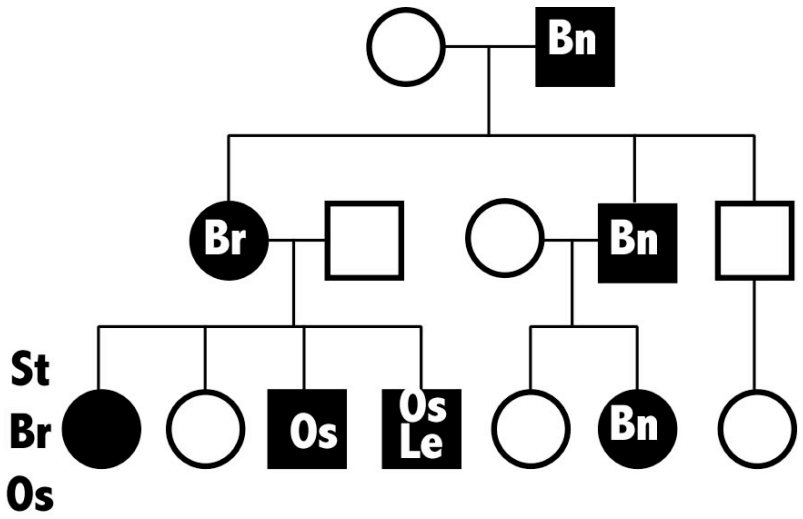


Mapping and cloning human genes

Based on position "positional cloning"

- Find linkage (or association) of disease gene to a chromosomal landmark (e.g., a polymorphic site)
- Clone the sequences from that region
- Identify **candidate genes** in that region (does it look like a gene, act like a gene, etc.)
- **Tests**: are the candidates mutated in disease patients?
- **Model organism or tissue culture**: If we knock out that gene, do we get the predicted mutant phenotype?

Linkage mapping of “familial” traits



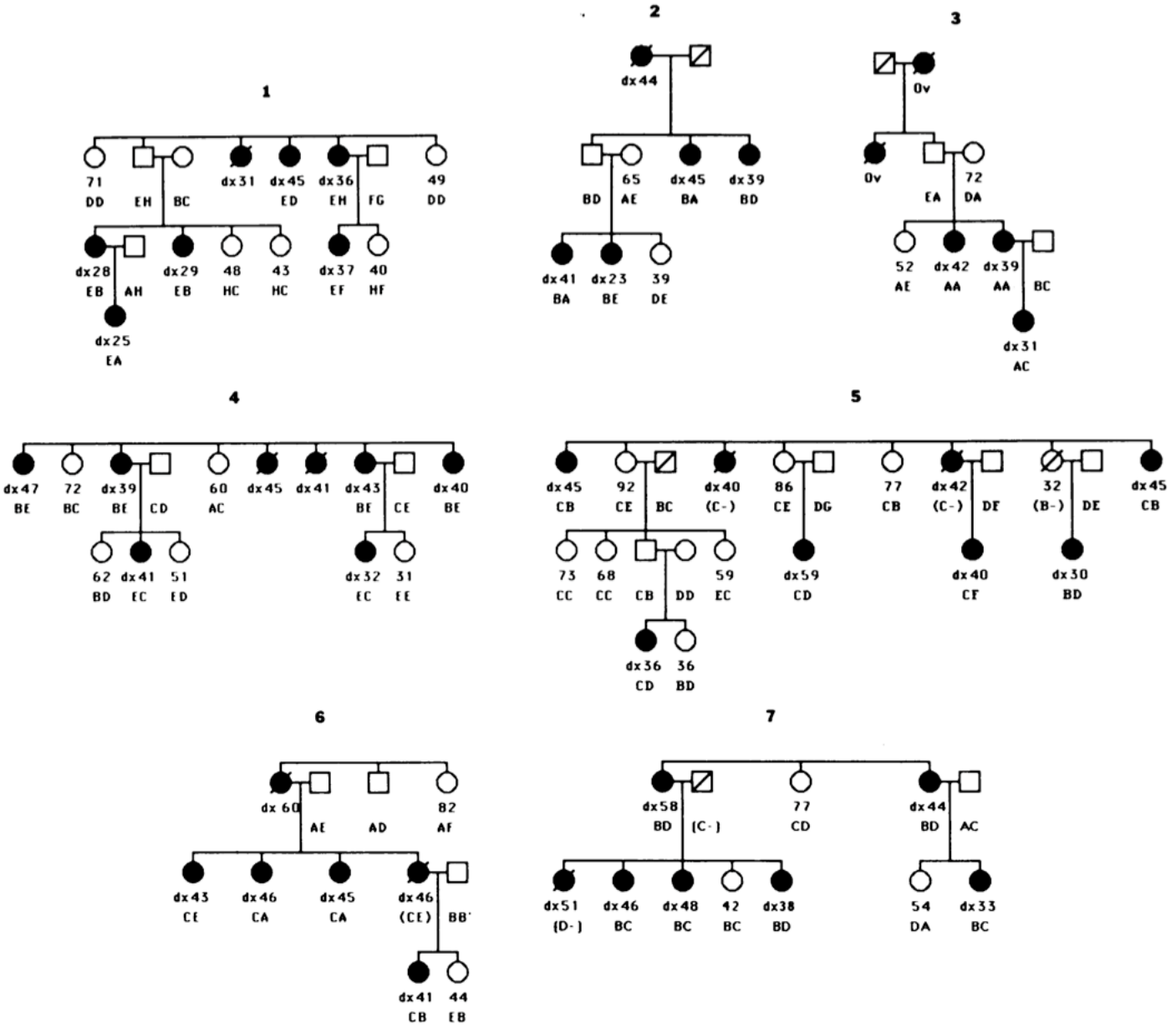
Bn = brain
Br = breast
Le = leukemia
Os = osteosarcoma
St = soft tissue

Li-Fraumeni syndrome: p53 checkpoint defect

Positional cloning of the BRCA1 gene

Initial mapping by LOD score analysis

(M-C King lab, 1990)



The problem with humans

Stumbling blocks…

- Which polymorphic loci to test?
- Not all meioses are informative
- Pedigrees may be too small to detect linkage with confidence

Which polymorphic loci to test for linkage?

Don't know ahead of time… so do trial and error!

- look for linkage between the disease and polymorphic site #1
- then repeat with polymorphic site # 2

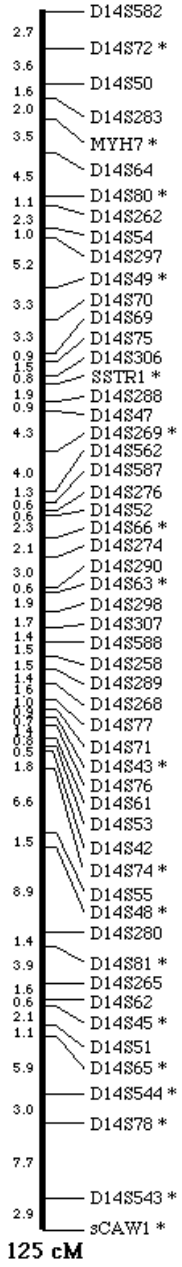
  ...etc.

A molecular marker map—

Variation in DNA sequence at specific chromosomal locations is mapped

These sites are used as landmarks in mapping genes

e.g., chr 14 map



The problem with humans

Stumbling blocks…

- Which polymorphic loci to test?
- Not all meioses are informative
- Pedigrees may be too small to detect linkage with confidence

Not all matings are informative…

A test for informative vs. non-informative meiosis: can we tell if the gamete was recombinant?

If we can tell: the meiosis is informative

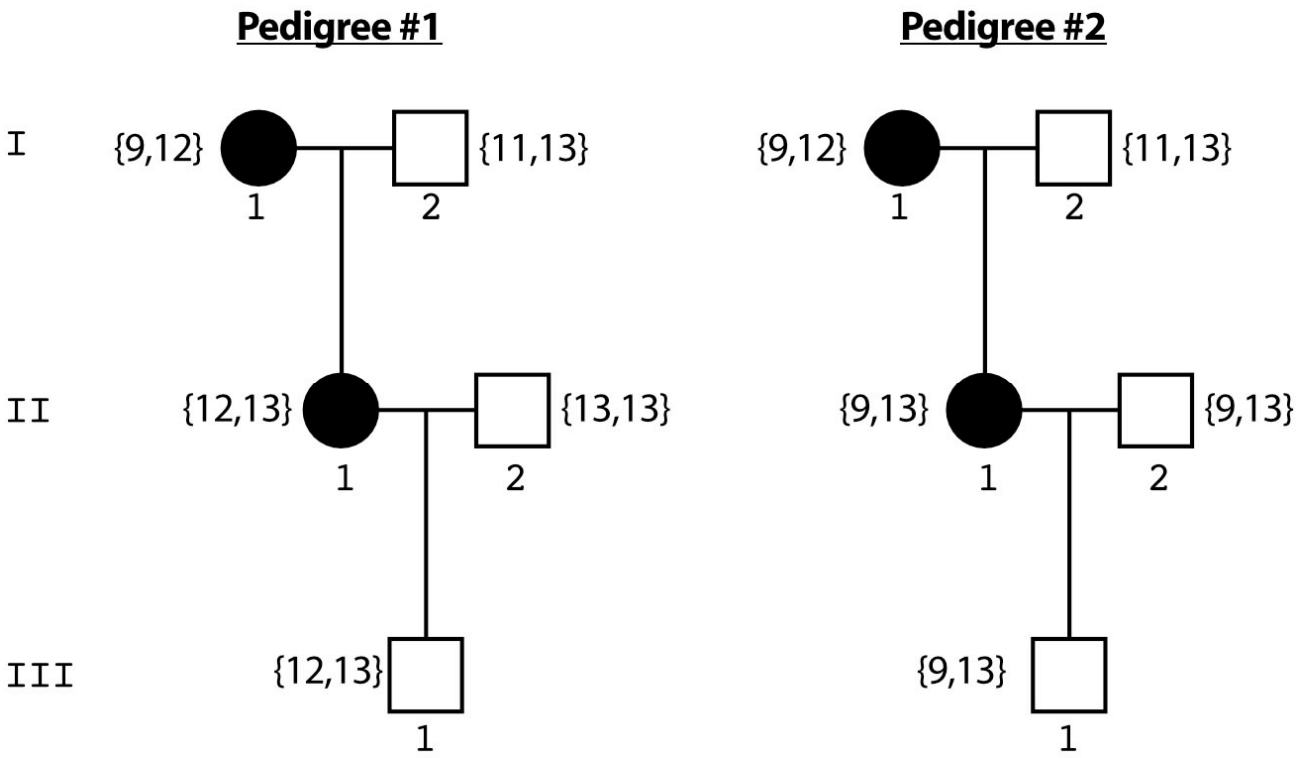
- yes, gamete was recombinant
  or
  no, not recombinant

} both are informative

cannot tell if the gamete was recombinant } → non-informative

Practice question

The two pedigrees show inheritance of an **autosomal dominant trait** (**D** = disease, dominant; **d** = normal, recessive). Numbers in {curly brackets} indicate alleles of a microsatellite repeat polymorphic locus. For each pedigree, state whether the meiosis **in II-1** is <u>informative</u> or <u>uninformative</u>, giving the parental types for II-1 in each case.



The problem with humans

Stumbling blocks…

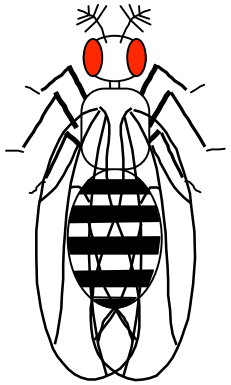
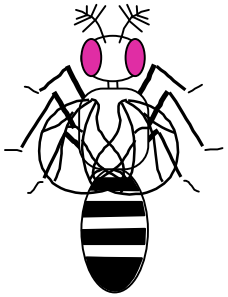
» Which polymorphic loci to test?

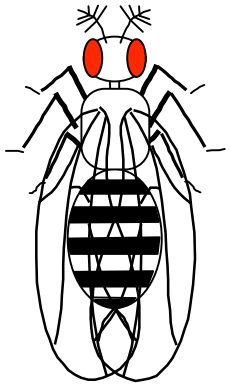
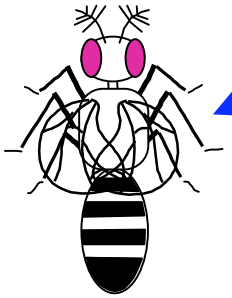
» Not all meioses are informative

» Pedigrees may be too small to detect linkage with confidence

Purebreeding--every fly a known genotype

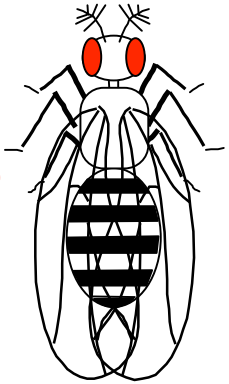
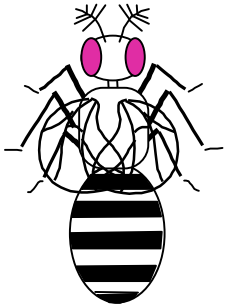
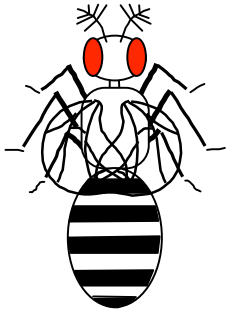
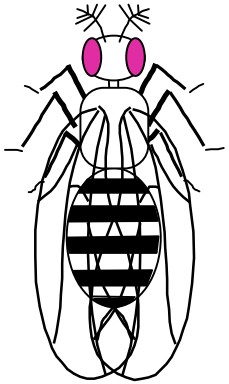
P_1 cross:

$\frac{pr^+ \quad vg^+}{pr^+ \quad vg^+}$  X  $\frac{pr \quad vg}{pr \quad vg}$

F_1 : $\frac{pr^+ \quad vg^+}{pr \quad vg}$  X 

Known "Phase"

Testcross Parent

1339  1195  151  154 

Mapping a gene using molecular markers (cont'd)

Stumbling blocks…

» Which polymorphic loci to test?

» Not all meioses are informative

» Pedigrees may be too small to detect linkage with confidence

A solution—

Play the odds: What is more likely to give this pedigree outcome, linkage or non-linkage?