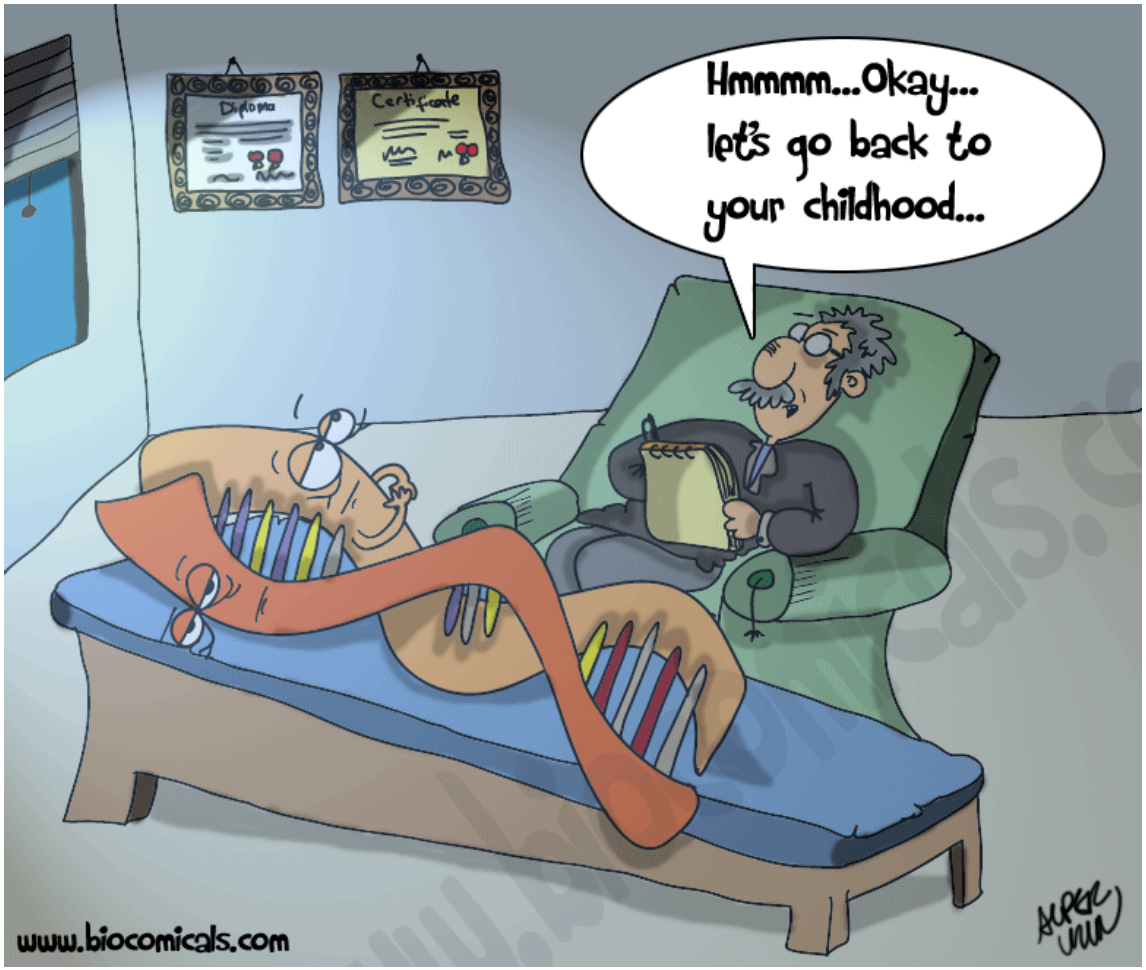
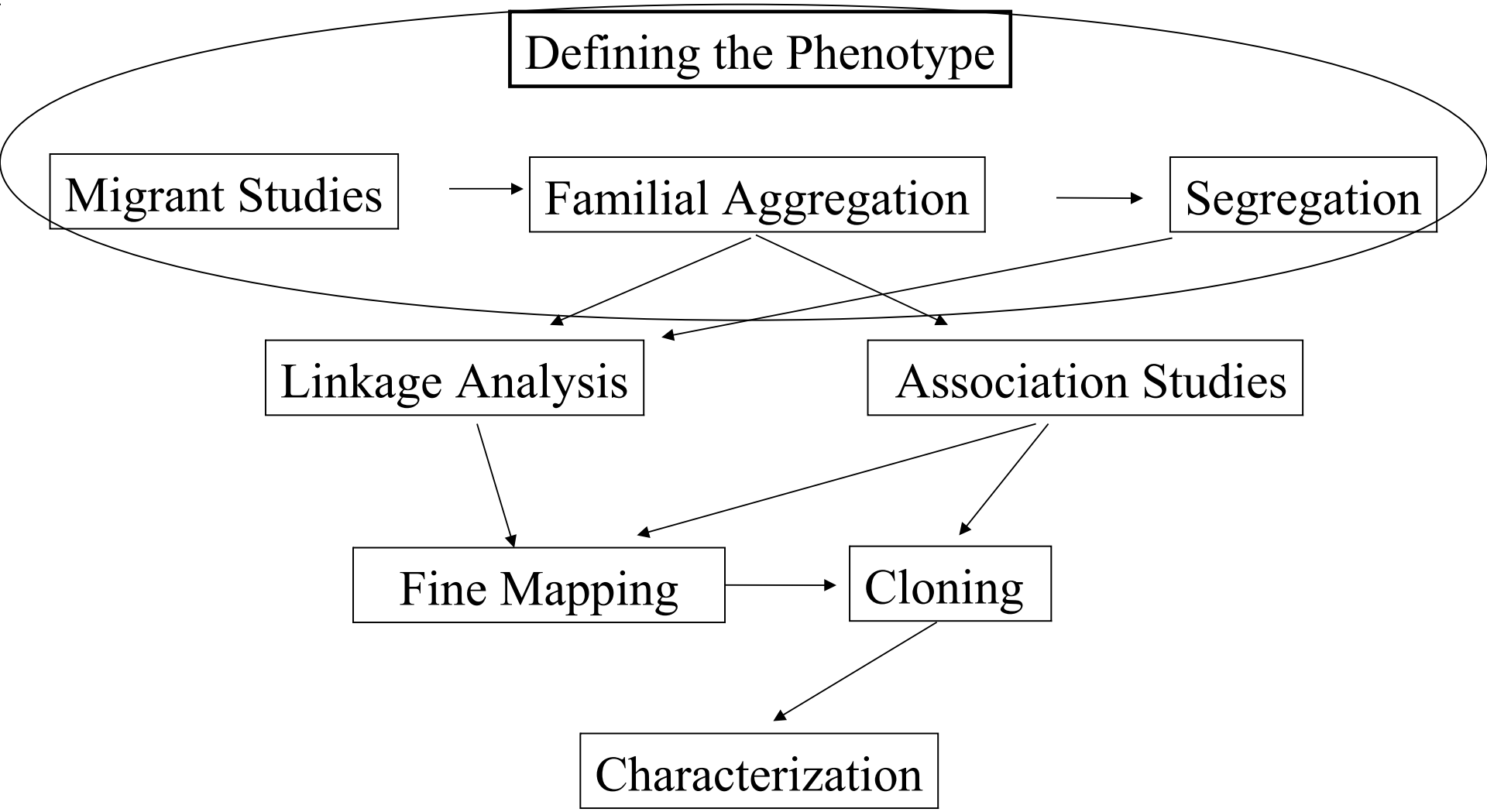
Population Genetics



Outline

- **Several methods to answer the question: Is the trait genetic?** [Austin Ch. 2]
  - Familial Aggregation and Recurrence risks
  - Heritability
  - Segregation
- Genetic concepts [Austin Ch. 3]
  - Allele Frequency Estimation
  - Hardy-Weinberg equilibrium (HWE)
  - HWE Game
  - Population Substructure/Stratification

Process of Genetic Epidemiology



Overview of is a trait genetic?

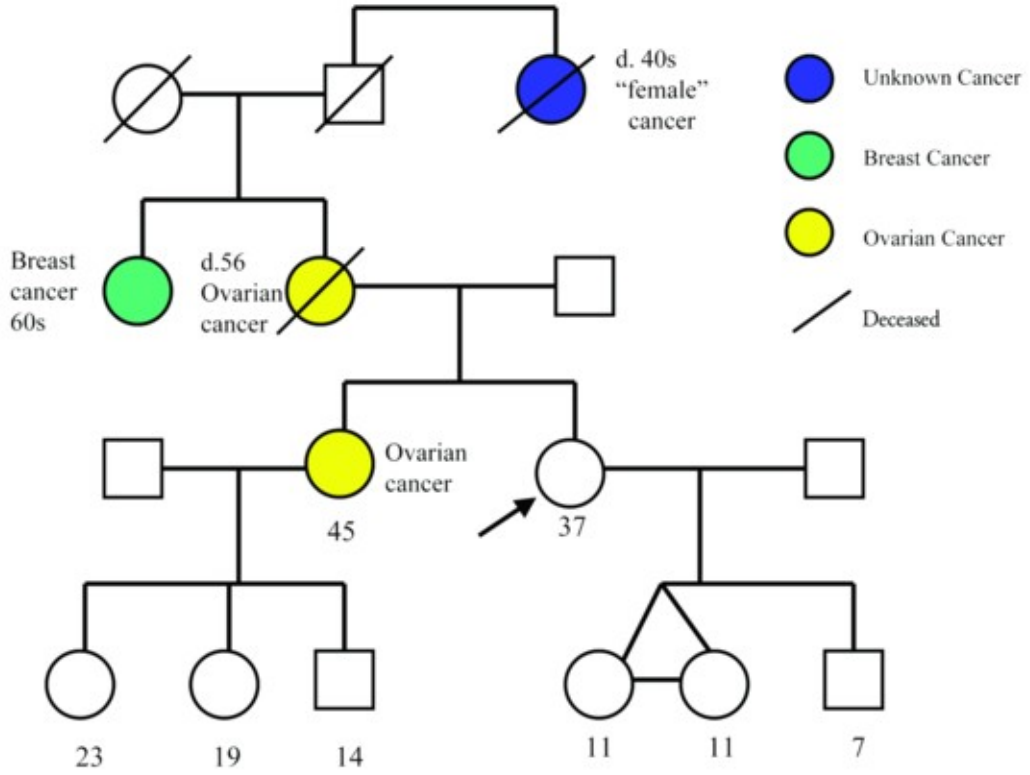
- Before we apply genetic technology to find genetic determinants of disease, first show the trait is genetic!
- Several ways to look at this
  - Familial aggregation – Does a trait occur more frequently in relatives
  - Heritability – proportion of variance explained
  - Segregation analysis – formal statistical model to the inheritance pattern

Outline

- Several methods to answer the question: Is the trait genetic? [Austin Ch. 2]
  - **Familial Aggregation and Recurrence risks**
  - Heritability
  - Segregation
- Genetic concepts [Austin Ch. 3]
  - Allele Frequency Estimation
  - Hardy-Weinberg equilibrium (HWE)
  - HWE Game
  - Population Substructure/Stratification

Familial Aggregation

- Does the phenotype tend to run in families?
- Ascertain family members by “probands”, i.e., affected individuals



Familial aggregation

"Family case-control" approach

Family history of disease	Yes	No
# Cases	a	b
# Controls	c	d

Odds Ratio = OR
= (a/b)/(c/d) = ad/bc

"Family cohort" approach

	Relative has disease? Yes	No
# Relatives of Cases	a	b
# Relatives of Controls	c	d

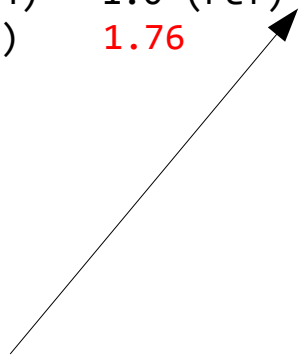
Relative Risk = RR
= (a/(a+b))/(c/(c+d))

- Be careful with "family history"
  - E.g., breast cancer, only females
  - E.g., late onset diseases, relative may not generally have disease
- Generally done by interview (cost reasons), though medical records would be more accurate

Khoury et al., 1993, and in Austin

Familial aggregation

	No of Cases (%)	No of Controls (%)	Adjusted OR	95% CI
Family history of pancreatic cancer in 1st-degree relative				
No	1,107(93.6)	1,152(96.4)	1.0 (ref)	(1.19,2.61)
Yes	76(6.4)	43(3.6)	1.76	



Unadjusted OR=(a/b)/(c/d)=(76/43)/(1107/1152)=1.84

ln(OR) = 0.61

SE(ln(OR)) = sqrt(1/a + 1/b + 1/c + 1/d)

= sqrt(1/76 + 1/43 + 1/1107 + 1/1152) = 0.195

ln(OR) +/- 1.96*SE(ln(OR)) = 0.61 +/- 1.96*0.195

= (0.228, 0.992)

Exp((0.228, 0.992)) = (1.25, 2.70) *Unadjusted* 95% CI

Jacobs et al., 2010; Austin

Familial aggregation

	No of Cases (%)	No of Controls (%)	Adjusted OR	95% CI
Family history of pancreatic cancer in 1st-degree relative				
No	1,107(93.6)	1,152(96.4)	1.0 (ref)	(1.19,2.61)
Yes	76(6.4)	43(3.6)	1.76	
Number of affected 1st-degree Relatives with pancreatic cancer				
One	70(92.1)	42(97.7)	1.70	(1.14,2.53)
Two or more	6(7.9)	1(2.3)	4.26	(0.48,37.7)
Age at diagnosis of youngest Affected 1st-degree relative				
>= 60 yrs	38(65.5)	22(66.7)	1.82	(1.05,3.13)
< 60 yrs	20(34.5)	11(33.3)	1.64	(0.71,3.34)

Jacobs et al., 2010; Austin

Recurrence ('Familial') Risk Ratios

- Compares the probability a subject is affected given they have an affected family member to the population risk:

  $\lambda_R = K_R/K$,

  where K_R is the risk to relatives of type R

  K is the population risk

λ_S = recurrence risk to siblings of probands versus the general population risk.

(Recurrence Risks are not covered in Austin)

Estimating Recurrence Risk Ratios (RRR)

- With case-control data, calculate RRR as:

  Proportion of affected relatives of the cases (observed) /

  Proportion of affected relatives of controls (expected) (assumed to estimate K, the population prevalence of disease)

$$\text{Using } \lambda_R = P(Y_2 = 1 \mid Y_1 = 1) / K$$

- The higher the value of λ_R the stronger the genetic effect

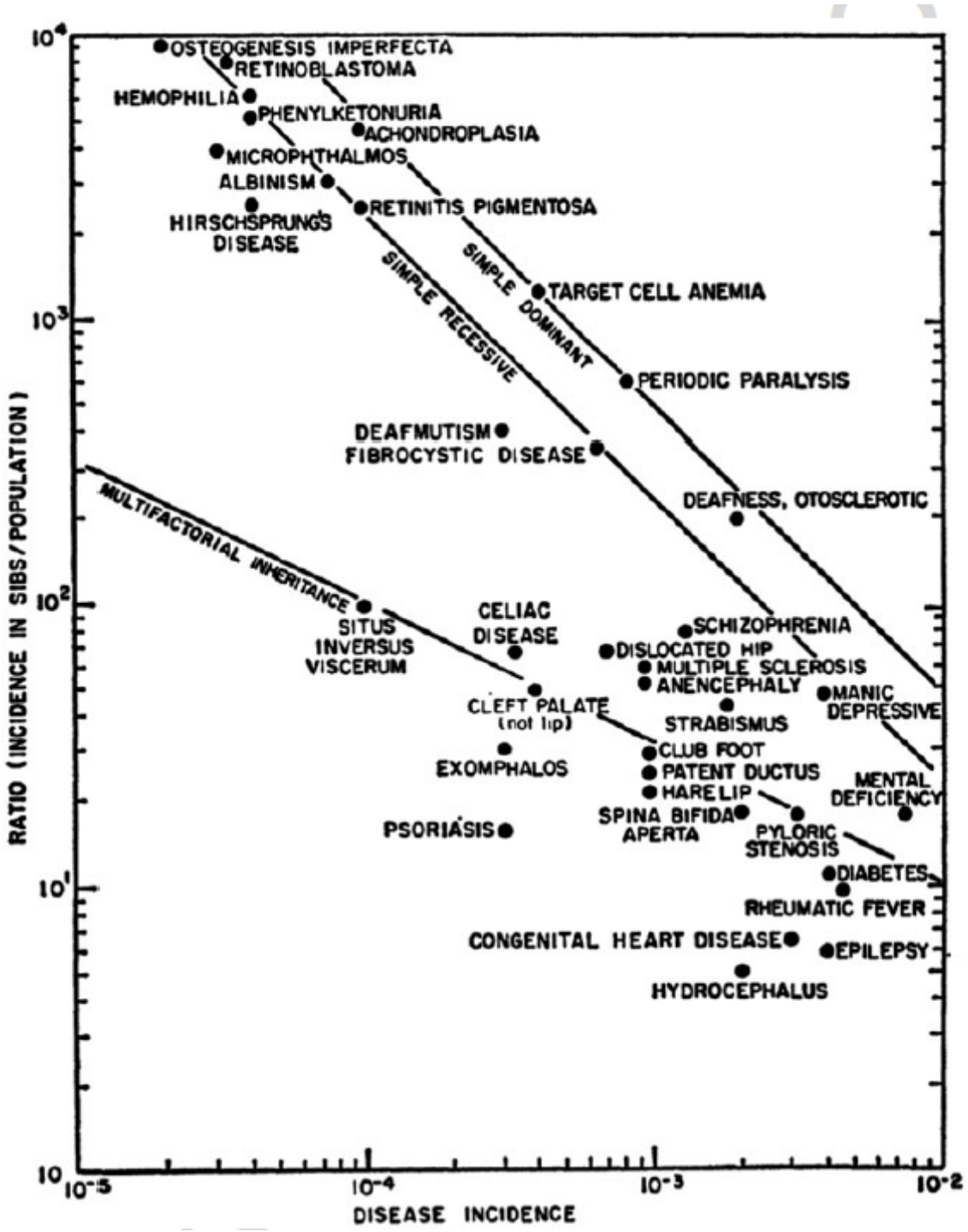
Examples of λ_s

• Alzheimer Disease	3-4
• Rheumatoid Arthritis	12
• Schizophrenia	13
• Type I Diabetes	15
• Multiple Sclerosis	20-30
• Neural Tube Defects	25-50
• Autism	75-150

Defining the phenotype...!

Limitations of recurrence risks

- Recurrence risks depend on mode of inheritance and disease frequency.
- Single gene diseases have high recurrence risks
- More common complex diseases have lower values (e.g., CHD).
- Here, hard to distinguish genetic versus environmental effects.



Laird and Lange, Fundamentals of Modern Statistical Genetics 2011, (Figure 4.1, p. 53)

Outline

- Several methods to answer the question: Is the trait genetic? [Austin Ch. 2]
  - Familial Aggregation and Recurrence risks
  - **Heritability**
  - Segregation
- Genetic concepts [Austin Ch. 3]
  - Allele Frequency Estimation
  - Hardy-Weinberg equilibrium (HWE)
  - HWE Game
  - Population Substructure/Stratification

Heritability Analysis

- Evaluates the genetic contribution to a trait P in terms of variance explained, based on an additive model of each additional genetic loci (additive polygenic model).
- Phenotype = Mean + Genetics + Environment

$$P = \mu + G + E$$

- Overall variation in Phenotype P = Var(P) =

$$\sigma^2_P = \sigma^2_G + \sigma^2_E$$

Broad Sense Heritability

- “Broad sense” heritability, $H^2=\sigma^2_G/\sigma^2_P$

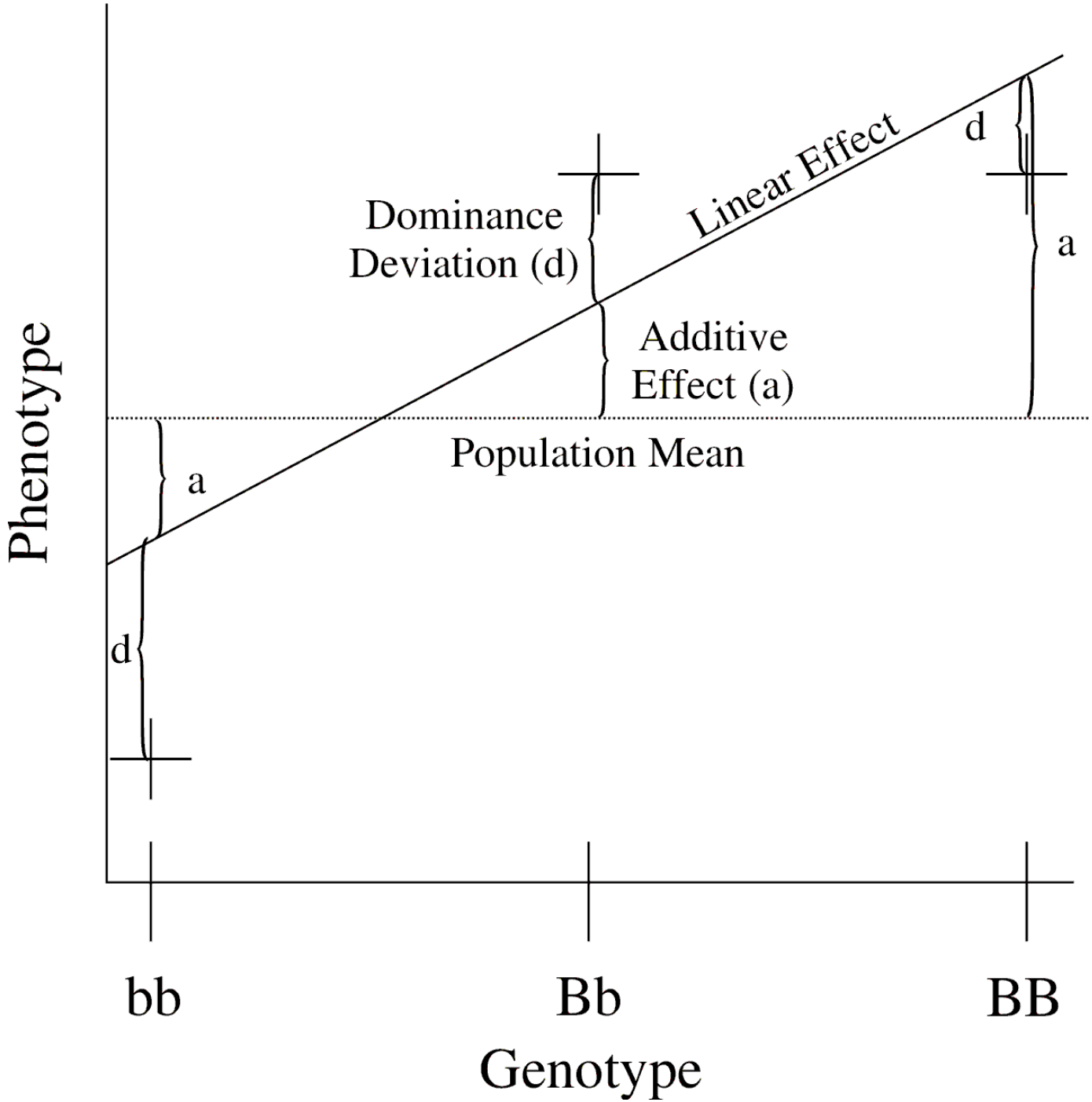
- Proportion of the phenotypic variance attributable to genetic influences

Narrow Sense Heritability

- (Recall previous slide: $\sigma^2_P = \sigma^2_G + \sigma^2_E$)
- $\sigma^2_G = \sigma^2_A + \sigma^2_D$
  - A: Additive effect
  - D: Dominance effect
- "Broad sense": $H^2=\sigma^2_G/\sigma^2_P=(\sigma^2_A + \sigma^2_D)/\sigma^2_P$
- "Narrow sense": $h^2 = \sigma^2_A/\sigma^2_P$
  - Proportion of variance explained only by additive genetic effects
  - This is what people usually are referring to, if don't say

Narrow Sense Heritability

- (Proportion of phenotypic variance that is explained only by additive genetic effects)
- Additive is largest part, and commonly focus only on this.
- Can estimate from regression:
  - –Additive: 0, 1, 2 for alleles
  - –Dominance: 0, 1, 0 for departure
- *Or estimate from twin studies...*


Phenotype
Linear Effect
d
Dominance
Deviation (d)
a
Additive
Effect (a)
Population Mean
a
d
bb
Bb
BB
Genotype

Liability model

Two copies of capitol letter (risk allele)

- Two unlinked loci, with allele a and risk allele A; and allele b and risk allele B.
- Increase the risk by 1 unit on liability scale

Table 2.3. Genotypes, liability scores, and genotype frequencies under a polygenic model with two loci.

Genotype(s)	Liability score	Frequency
aabb	0	1
Aabb, aAbb, aaBb, aabB	+1	4
AAbb, AaBb, AabB, aABb, aAbB, aaBB	+2	6
AABb, AAbB, AaBB, aABB	+3	4
AABB	+4	1

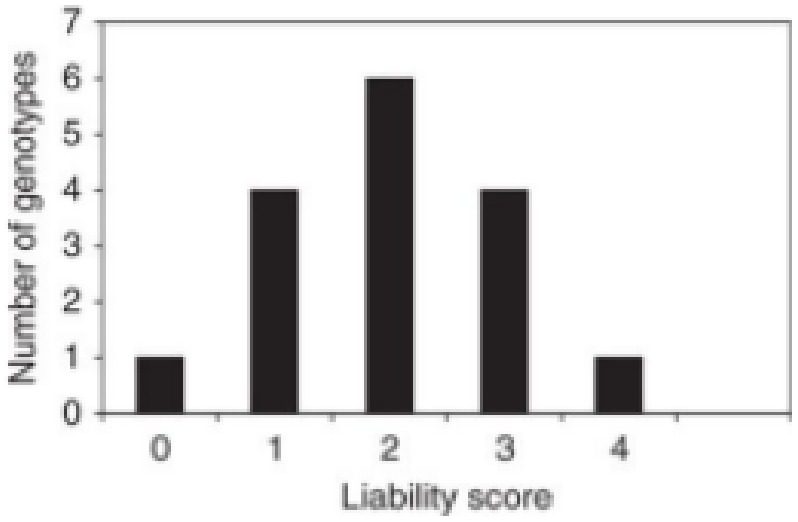


Fig. 2.1. Frequency distribution of genotypes by liability score under a polygenic model with two loci.

Austin 2013

- *Three* unlinked loci Aa, Bb, Cc (capitol is risk allele)
- Increase the risk by 1 unit on liability scale
- If dichotomous, threshold determines disease status

Table 2.4. Genotypes, liability scores, and genotype frequencies under a polygenic model with three loci.

Genotype(s)	Liability score	Frequency
aabbcc	0	1
Aabbcc, aAbbcc, aaBbcc, aabBcc, aabbCc, aabbcC	+1	6
AAbbcc, AaBbcc, AabBcc, ...	+2	15
AABbcc, AAbBcc, AaBBcc, ...	+3	20
AABBcc, AABbCc, AABbcC, ...	+4	15
AABBCc, AABBcC, AABbCC, ...	+5	6
AABBCC	+6	1

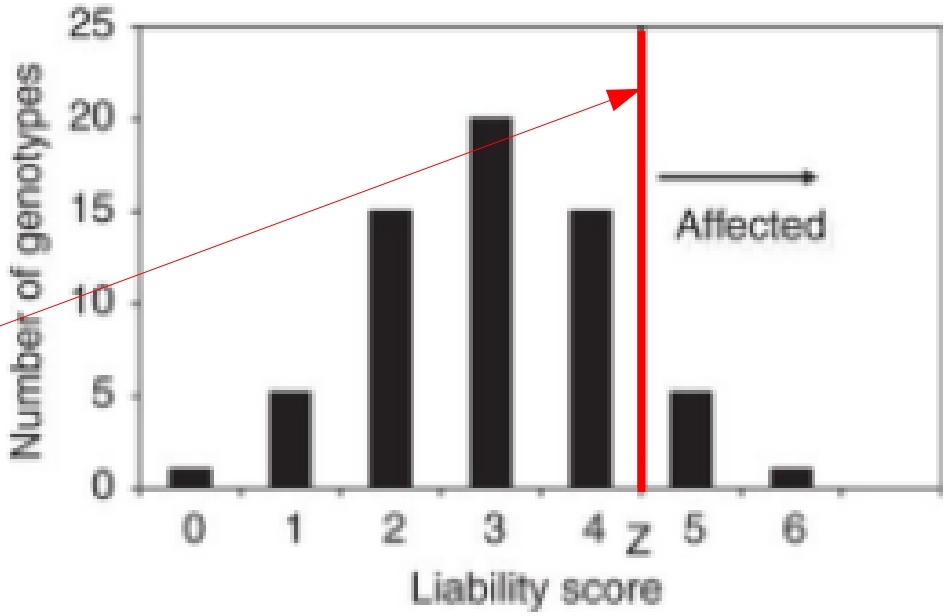
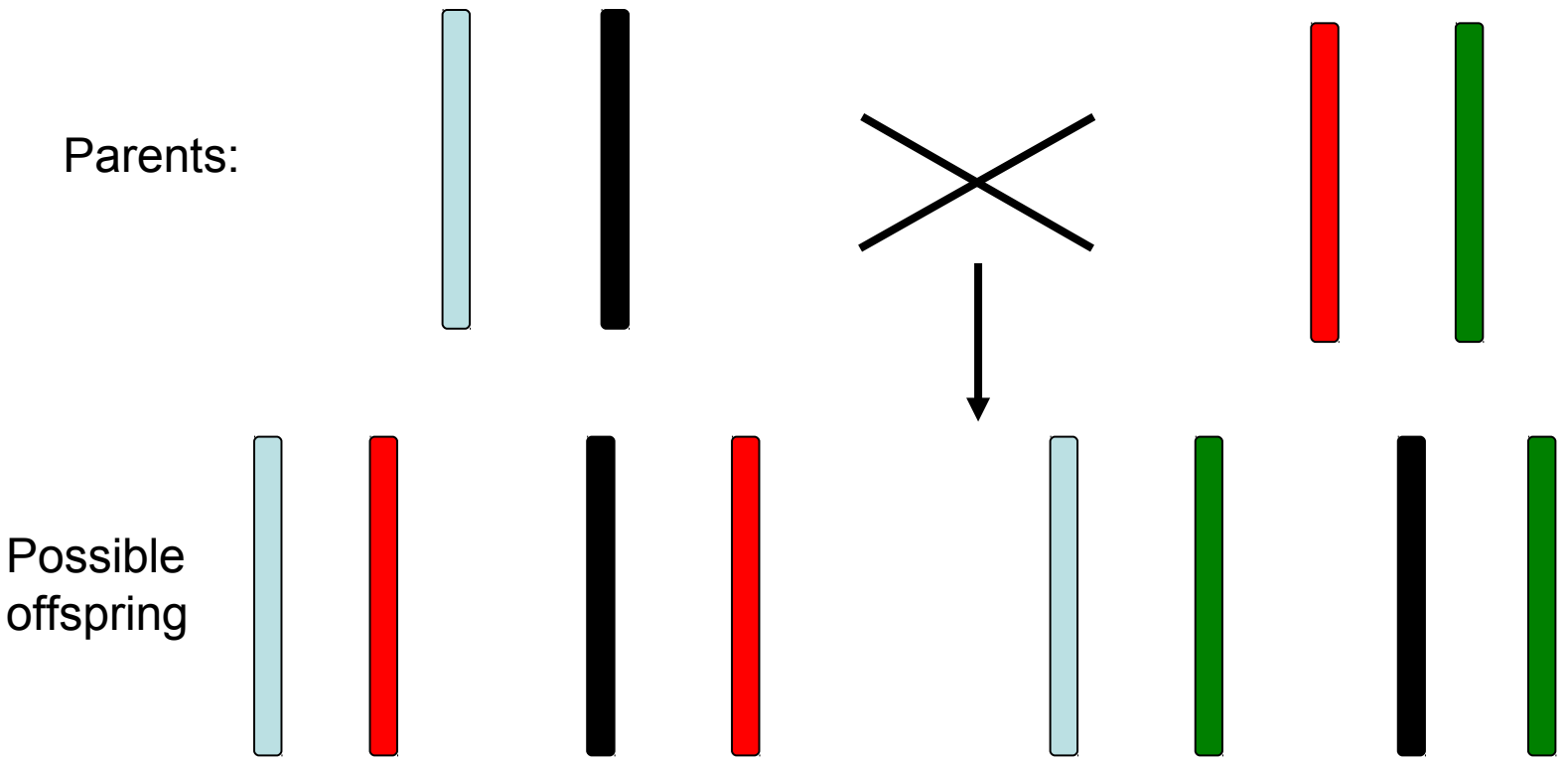


Fig. 2.2. Frequency distribution of genotypes by liability score under a polygenic model with three loci and threshold value Z for dichotomous traits.

Twin Studies

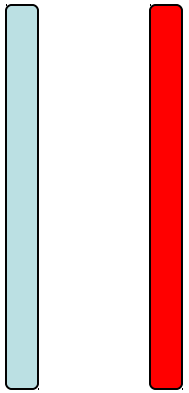
(More background than is in the text...)

- Compare the phenotype correlation or disease concordance rates of MZ (identical) and DZ (fraternal) twins.



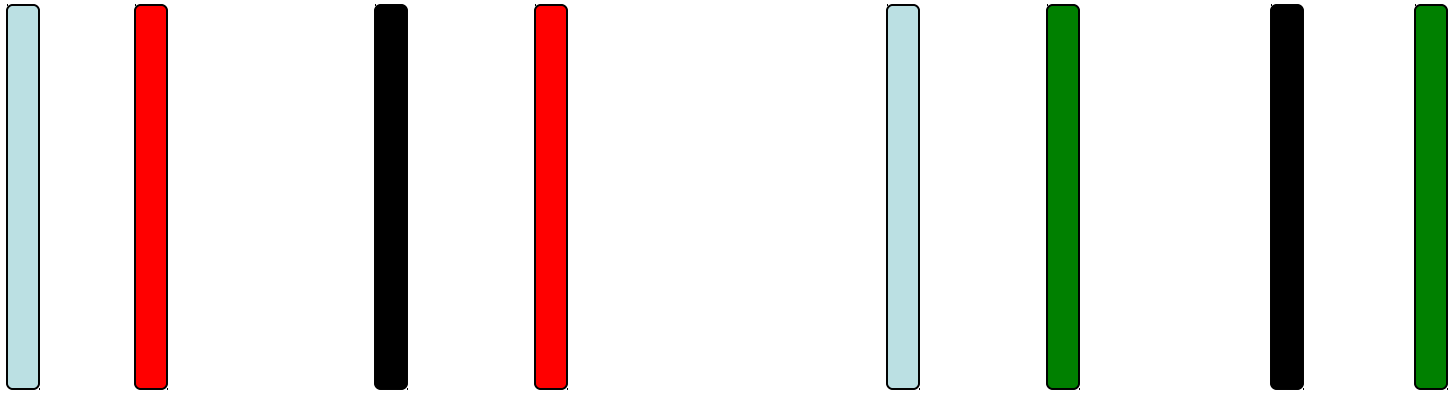
MZ Twins (Identical)

If twin 1 is



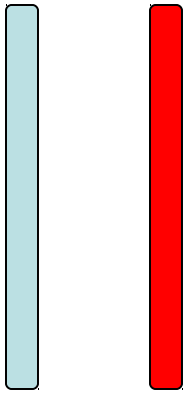
Both alleles are shared identical by descent (IBD)

Then twin 2 must be:



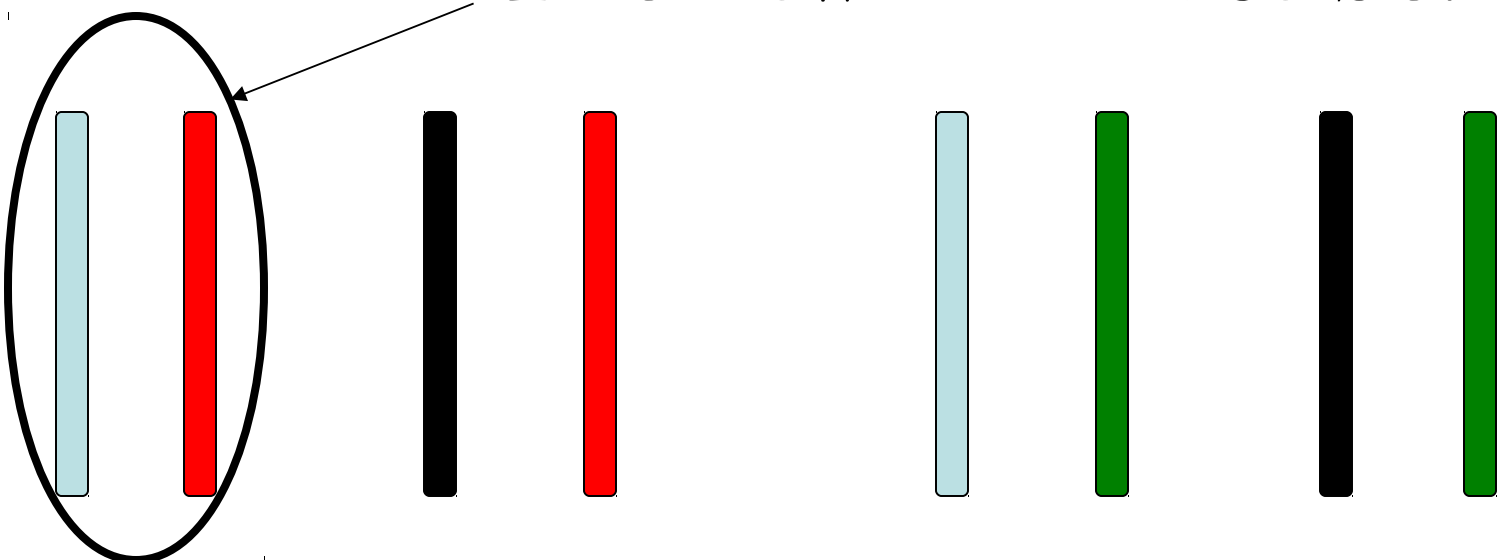
MZ Twins (Identical)

If twin 1 is

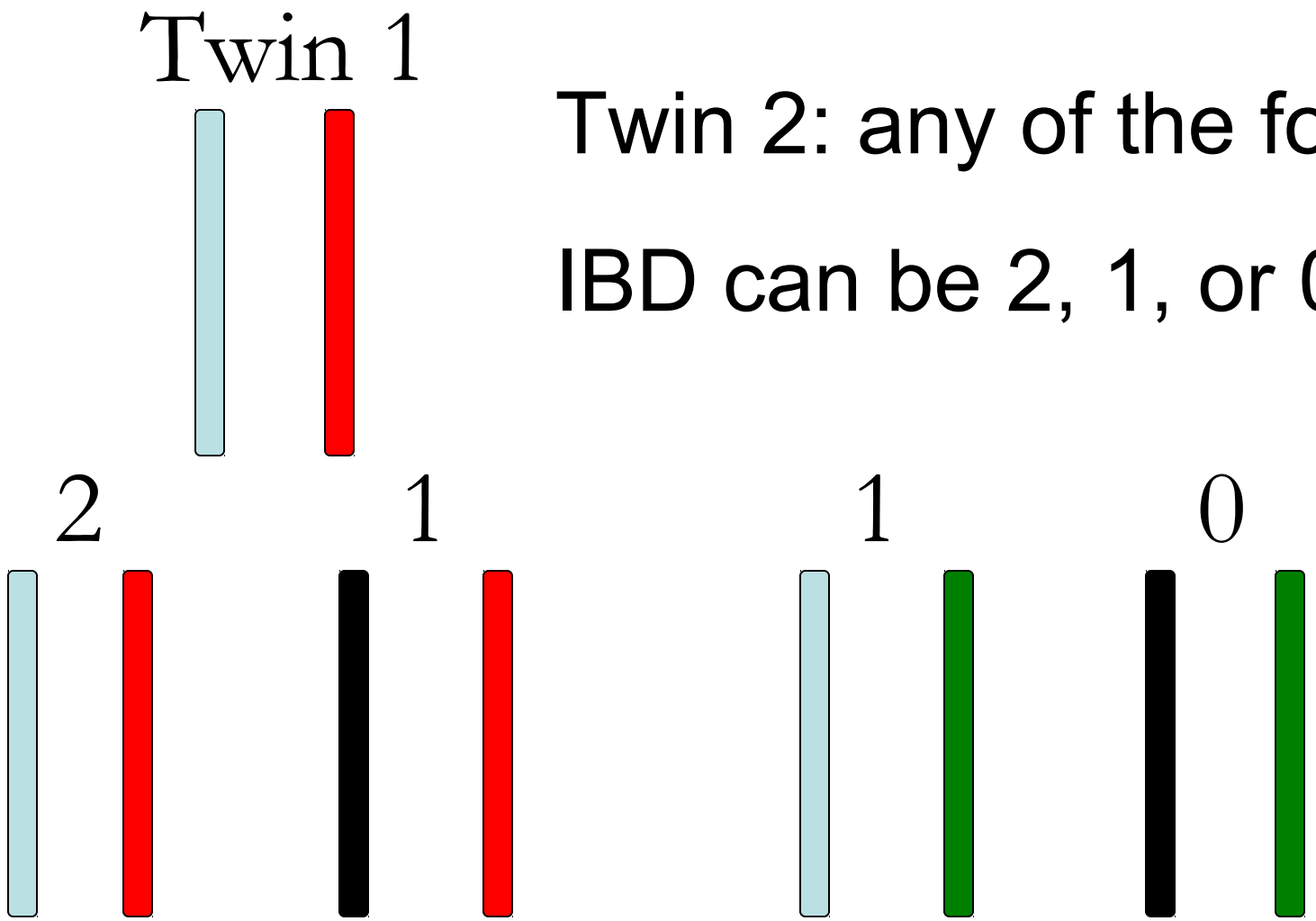


Both alleles are shared identical by descent (IBD)

Then twin 2 must be:



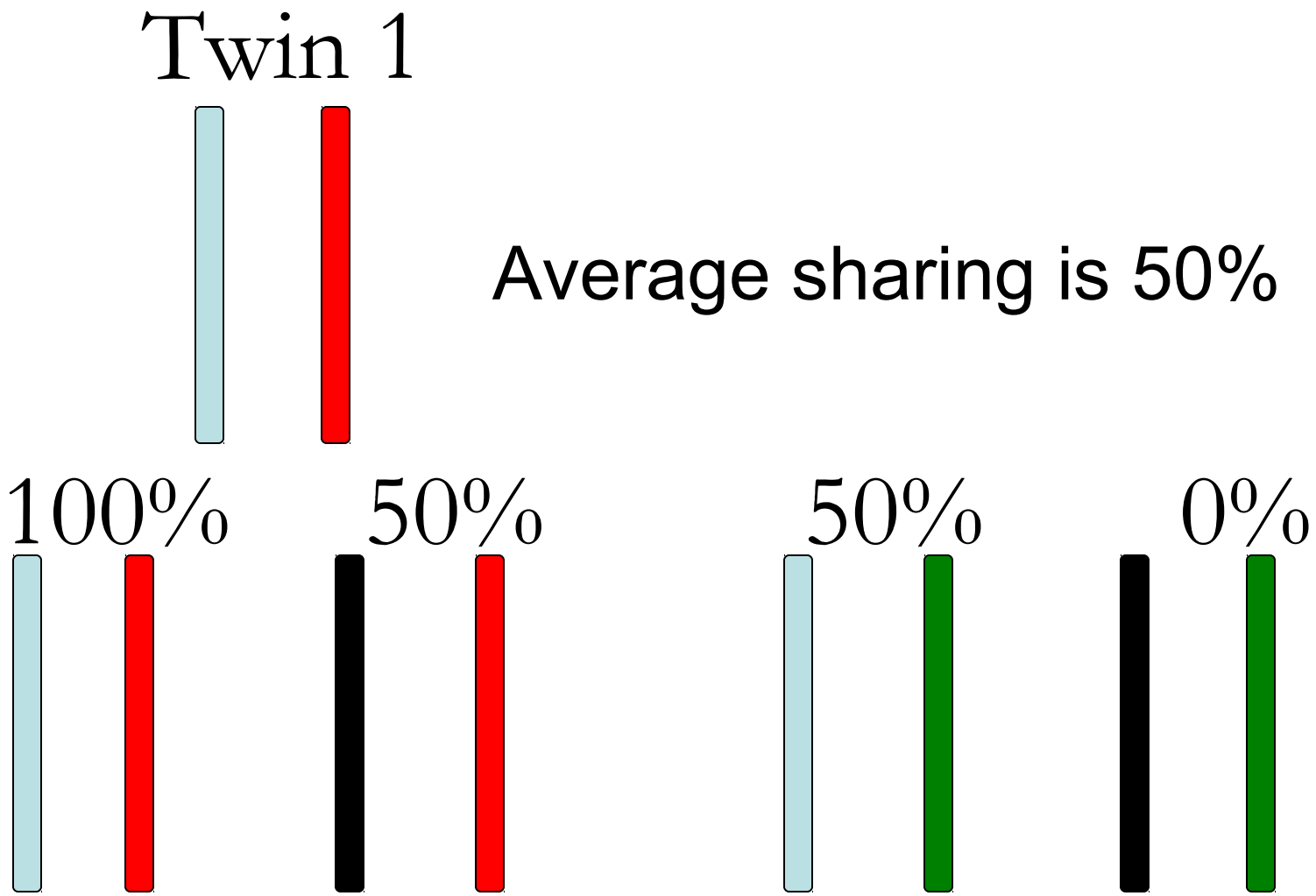
DZ Twins (Fraternal)



Twin 2: any of the four

IBD can be 2, 1, or 0

DZ Twins (Fraternal)



IBD Sharing

	# of alleles shared IBD			
	2	1	0	
	Pr(2)	Pr(1)	Pr(0)	Prop IBD
Relationship				
Self, MZ twins	1	0	0	1
Parent, Offspring	0	1	0	1/2
Full siblings	1/4	1/2	1/4	1/2
Gr-child, Gr-prt	0	1/4	3/4	1/4
First cousins	0	1/4	3/4	1/8

Proportion of alleles shared IBD

= # alleles x Pr(# alleles) / 2

Twin Studies

- ACE Model: $\sigma^2_P = \sigma^2_A + \sigma^2_C + \sigma^2_E$
  - A: Additive genetics
  - C: Common Environment
  - E: Unique Environment
- Correlation in phenotype (P_1, P_2) among twins:
  - $Corr_{mz}(P_1, P_2) = r_{mz} = A + C$ [100% genes + Env]
  - $Corr_{mz}(P_1, P_2) = r_{dz} = ½A + C$ [50% genes + Env]
- Heritability: $h^2 = 2(r_{mz} - r_{dz})$

Example of Twin Study: PCa

- Twin registry (Sweden, Denmark, and Finland)
  7,231 MZ and 13,769 DZ Twins (male)

Twin	Concordant pairs (A)	Discordant pairs (B+C)	Concordance 2A / (2A+B+C)
MZ	40	299	0.21 (r_{mz})
DZ	20	584	0.06 (r_{dz})

Heritability = $2(r_{mz} - r_{dz}) = 2(0.21-0.06)$
$= 0.30$

Lichtenstein et al NEJM 2000 13;343:78-85.
(actual paper estimate uses Structural Equation Modeling, so differs)

Heritability extras

- Can also model interactions (covered in text)
  - ACE model will overestimate the heritability in the presence of interactions
- Other violations also... Main point is that the ACE model is based on some assumptions that might not be true, and bias our heritability estimates

Heritability ***misconceptions*** (from text) – why?

- h^2 is the proportion of phenotype passed on to the next generation
- High h^2 implies genetic determination
- Low h^2 implies no additive genetic variance
- h^2 is informative about the nature of between group differences
- Large h^2 implies genes of large effect

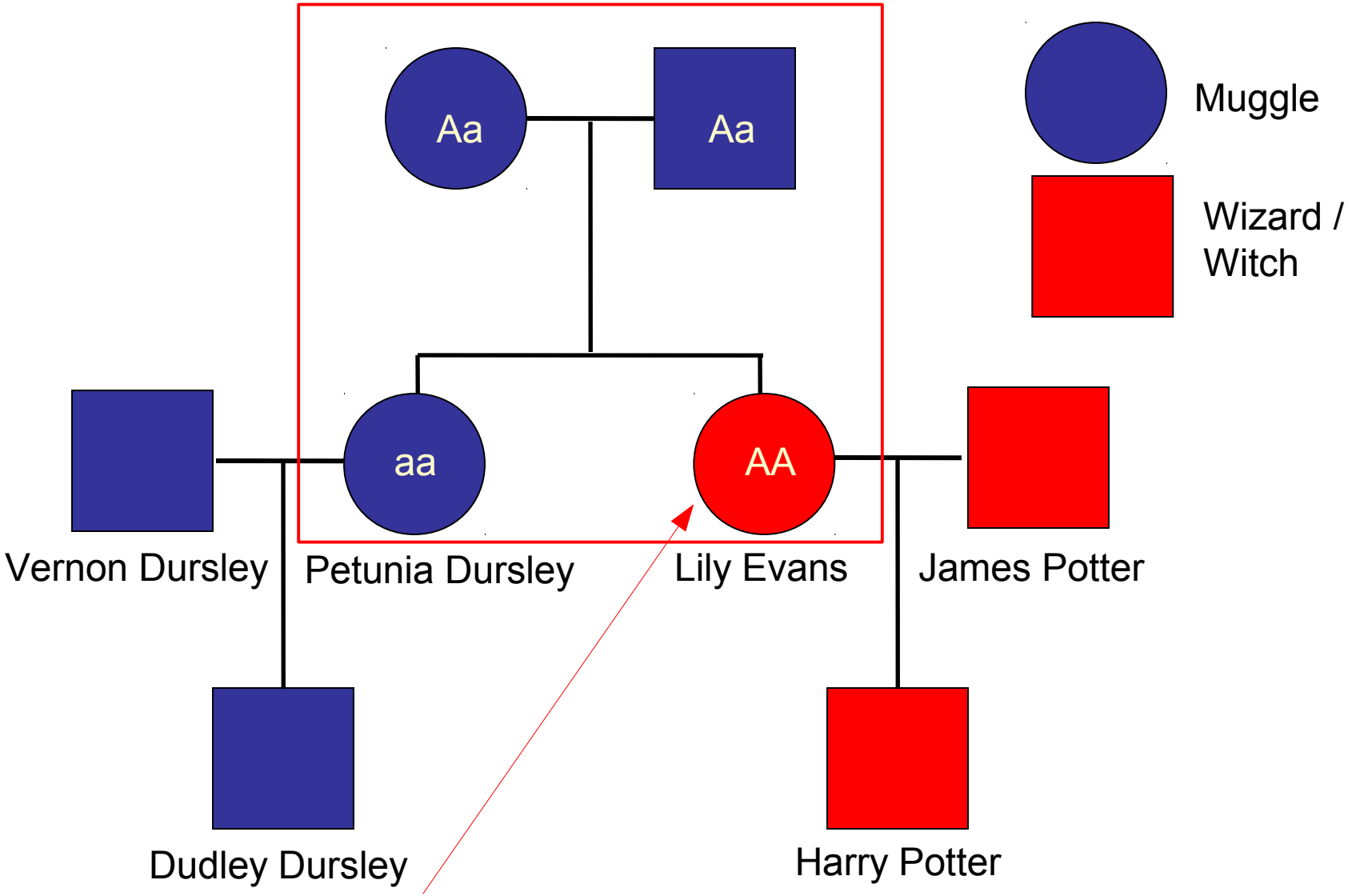
Outline

- Several methods to answer the question: Is the trait genetic? [Austin Ch. 2]
  - Familial Aggregation and Recurrence risks
  - Heritability
  - **Segregation**
- Genetic concepts [Austin Ch. 3]
  - Allele Frequency Estimation
  - Hardy-Weinberg equilibrium (HWE)
  - HWE Game
  - Population Substructure/Stratification

Segregation Analysis

- Study families.
- Estimate ‘mode of inheritance’ & what type of genetic variant might be causal.
- Determine whether the disease appears to follow particular patterns across generations.
- Estimate whether variants are rare or common, etc.

Segregation: Harry Potter's Pedigree



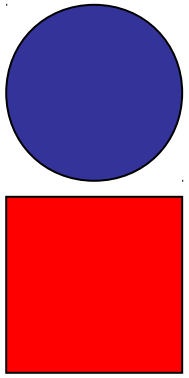
We thought this might be recessive because Lily is affected, but her parents are not (so maybe both are carriers). There are other reasons, e.g., not fully penetrant, that it only increases her risk/probability of being a wizard/witch.

Segregation Analysis

- What is the best model of inheritance for observed families?
    - Dominant
    - Recessive
    - Additive
  - Disease allele frequency?
  - Magnitude of risk? (might not be fully penetrant)

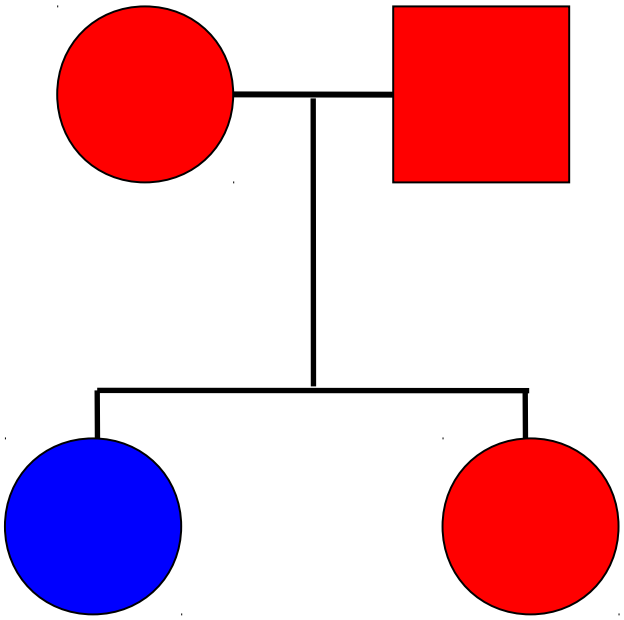
- Fit formal genetic models to data on disease phenotypes of family members.
- The parameters of the model are generally fitted finding the values that maximize the probability (likelihood) of the observed data.
- This information is useful in parametric linkage analysis, which assumes a defined model of inheritance.

What is the mode of inheritance here?



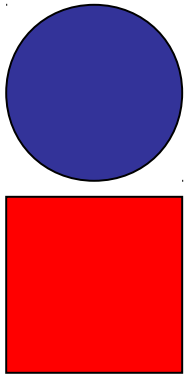
Unaffected

Affected



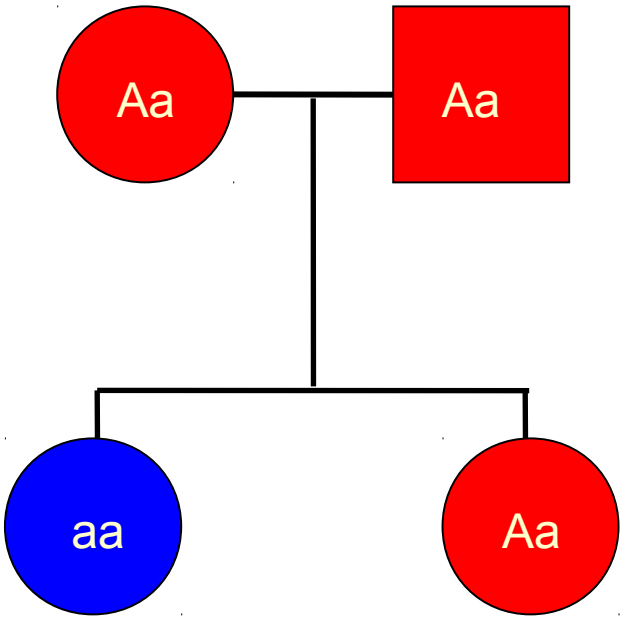
Assume *complete penetrance* here; i.e., if you have the allele, you will have the disease.

What is the mode of inheritance here?



Unaffected

Affected



Dominant: Two affected individuals have an unaffected offspring. [If recessive, then both parents hom for disease allele, and so must offspring.]

Assume *complete penetrance* here; i.e., if you have the allele, you will have the disease.

Outline

- Several methods to answer the question: Is the trait genetic? [Austin Ch. 2]
  - Familial Aggregation and Recurrence risks
  - Heritability
  - Segregation
- Genetic concepts [Austin Ch. 3]
  - **Allele Frequency Estimation**
  - Hardy-Weinberg equilibrium (HWE)
  - HWE Game
  - Population Substructure/Stratification

Allele Frequency

- Diploid, autosomal locus with 2 alleles: **A** and **a**
- Allele frequency is the fraction:

$$\frac{\text{No. of particular allele}}{\text{No. of all alleles in population}}$$

Allele (Gamete) Frequency

- Let p = Freq(A) frequency of the dominant allele
- Let q = Freq(a) frequency of the recessive allele

Then, $p + q = 1$

Genotype Frequency

- p^2 = frequency of homozygous dominant genotype
- q^2 = frequency of homozygous recessive genotype
- 2pq = frequency of heterozygous genotype

Then, $p^2 + 2pq + q^2 = 1$

Estimating Allele Frequencies from Genotype Frequencies

Genotypes:	AA	Aa	aa
Frequency:	p^2	2pq	q^2

- Frequency of **A** allele = $p^2 + ½ (2pq)$
- Frequency of **a** allele = $q^2 + ½ (2pq)$

Ex. Calculation: Allele Frequencies

Assume N=200 in each of two populations

- Pop 1: 90 AA 40 Aa 70 aa (N=200)
- Pop 2: 45 AA 130Aa 25 aa (N=200)

In Pop 1:

- p = 90/200 + ½ (40/200) = 0.45 + 0.10 = 0.55
- q = 70/200 + ½ (40/200) = 0.35 + 0.10 = 0.45

In Pop 2:

- p = 45/200 + ½ (130/200) = 0.225 + 0.325 = 0.55
- q = 25/200 + ½ (130/200) = 0.125 + 0.325 = 0.45

Take home points

- $p + q = 1$ (sum of the allele frequencies = 1)
- $p^2 + 2pq + q^2 = 1$ (sum of the genotype frequencies = 1)
- Two populations with markedly different genotype frequencies can have the same allele frequencies

Outline

- Several methods to answer the question: Is the trait genetic? [Austin Ch. 2]
  - Familial Aggregation and Recurrence risks
  - Heritability
  - Segregation
- Genetic concepts [Austin Ch. 3]
  - Allele Frequency Estimation
  - **Hardy-Weinberg equilibrium (HWE)**
  - HWE Game
  - Population Substructure/Stratification

Hardy-Weinberg

The **Hardy–Weinberg principle** states that both allele and genotype frequencies in a population remain constant—that is, they are in equilibrium—from generation to generation unless specific disturbing influences are introduced

$$p^2 + 2pq + q^2 = 1$$

Hardy-Weinberg Assumptions

Allele frequencies do not vary **IF**:

- Large population
- Random mating
- No in or out migration
- No isolated groups within the population
- No mutation
- No selection (no allele is advantageous)

Example: Test of Hardy-Weinberg Equilibrium (HWE)

- We have the following observed genotype frequencies: 100 GG, 30 AG, 20AA

- Frequency of the G allele:

  $p = 100/150 + 0.5(30/150) = 0.767$

- Frequency of the A allele

  $q = 20/150 + 0.5(30/150) = 0.233 \quad = 1-p$

Test of Hardy-Weinberg Equilibrium

2. Calculate expected genotype frequencies based on HW: $p^2 + 2pq + q^2 = 1$

GG

p^2 = 0.767*0.767 = 0.588

$n(p^2)$ = 150(0.588) = 88.2

AG

2pq = 2*0.767*0.233 = 0.357

n(2pq) = 150(0.357) = 53.6

AA

q^2 = 0.233*0.233 = 0.054

$n(q^2)$ = 150(0.054) = 8.1

Test of Hardy-Weinberg Equilibrium

3. Compare expected genotype counts (E) to observed genotype counts (O)

	E	O	(O-E)^2/E
GG	88.2	100	1.58
AG	53.6	30	10.39
AA	8.1	20	17.48

			29.45

Chi-square test

$= \Sigma_i (O_i - E_i)^2/E_i = 29.4$ (chi square distribution with 1 degree of freedom)

$p = 6.6 \times 10^{-8}$ > Out of H-W

HWE can be easily expanded to account for any number of alleles at a locus

- 3 allele case (p_1, p_2, p_3)

  Allele frequencies: $p_1 + p_2 + p_3 = 1$

  Genotype frequencies:

  $p_1^2 + p_2^2 + p_3^2 + 2p_1p_2 + 2p_1p_3 + 2p_2p_3 = 1$

- 4 allele case (p_1, p_2, p_3, p_4)

  Allele frequencies: $p_1 + p_2 + p_3 + p_4 = 1$

  Genotype frequencies:

  $p_1 2 + p_2 2 + p_3 2 + p_4^2 + 2p_1p_2 + 2p_1p_3 + 2p_2p_3 + 2p_3p_4 = 1$

Application of Hardy-Weinberg Equilibrium

For genetic association studies:

- Used as QC measure to assess the accuracy of the genotyping method
- Expect SNPs to be in HWE among control populations (ethnic-specific)
- Violations of HWE could indicate genotyping errors or bias in data

Outline

- Several methods to answer the question: Is the trait genetic? [Austin Ch. 2]
  - Familial Aggregation and Recurrence risks
  - Heritability
  - Segregation
- Genetic concepts [Austin Ch. 3]
  - Allele Frequency Estimation
  - Hardy-Weinberg equilibrium (HWE)
  - **HWE Game**
  - Population Substructure/Stratification

HWE Game

1. Everyone receives ~5 pairs of cards
2. Two allele model: Red (R allele) & Black (B allele)
3. Random Mating: Exchange one card from each pair with another person (keep cards face down)
4. Determine genotype frequency: RR, RB, BB
5. Determine allele frequency: R, B

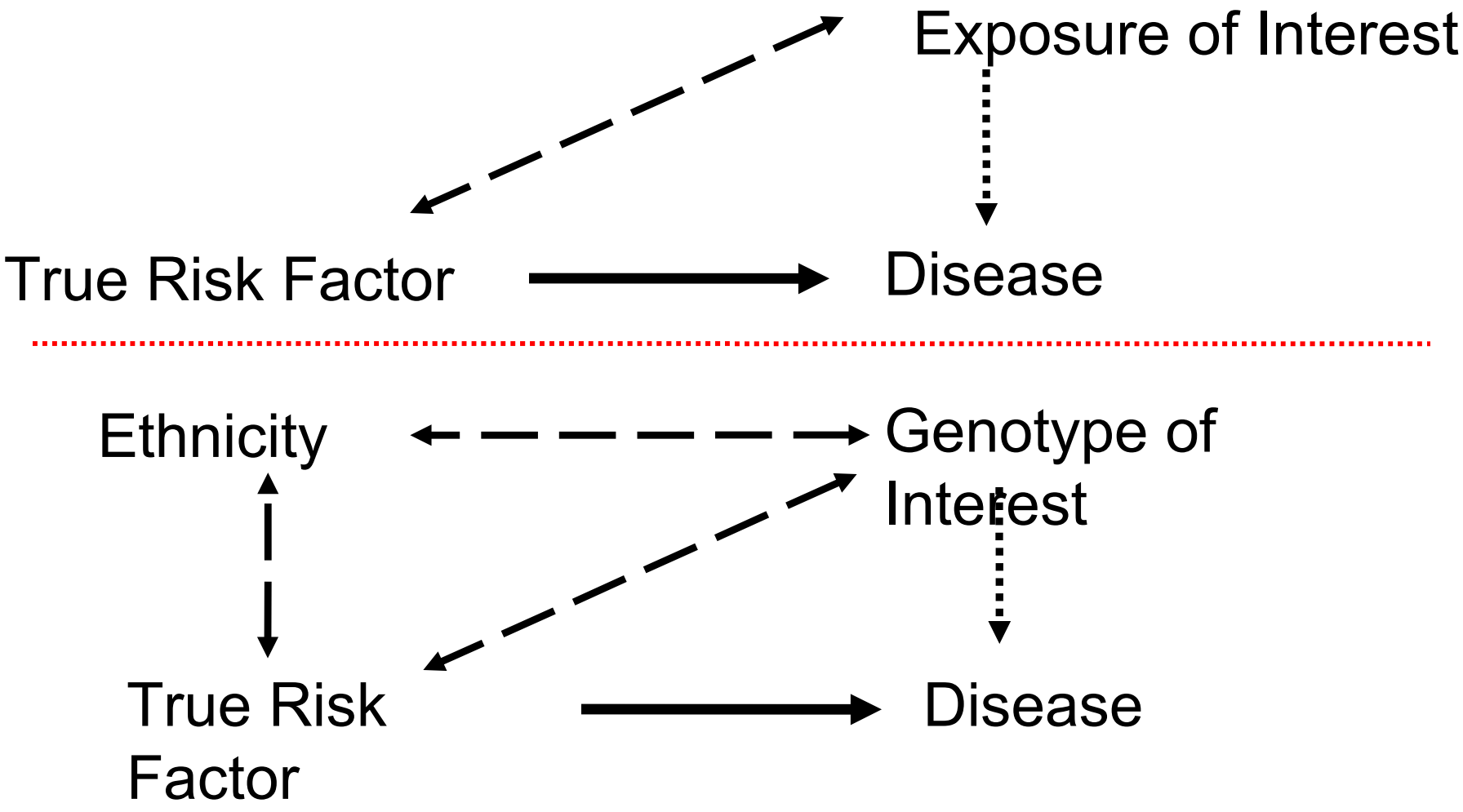
Outline

- Several methods to answer the question: Is the trait genetic? [Austin Ch. 2]
  - Familial Aggregation and Recurrence risks
  - Heritability
  - Segregation
- Genetic concepts [Austin Ch. 3]
  - Allele Frequency Estimation
  - Hardy-Weinberg equilibrium (HWE)
  - HWE Game
  - **Population Substructure/Stratification**

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

Full heritage American Indian population

	+	-
Gm3;5,13,14	~1%	~99%

NIDDM prevalence ~40%

Caucasian population

	+	-
Gm3;5,13,14	~66%	~34%

NIDDM prevalence ~15%

Gm3,5,13,14 haplotype	Cases	Controls
+	7.80%	29.00%
-	92.20%	71.00%

Unadjusted for ethnic background

OR = 0.27 (95% 0.18-0.40)

Different genotype frequency, different phenotype frequency

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

Index of Indian heritage	Gm3;5,13,14 haplotype	% Diabetes
0	65.8%	18.5%
4	42.1%	28.5%
8	1.6%	39.2%

Gm3,5,13,14 haplotype	Cases	Controls
+	7.80%	29.00%
-	92.20%	71.00%

Adjusted for ethnic background

OR = 0.83 (95% 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)

	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		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		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

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

Population inbreeding

Population inbreeding occurs when there is a preference of mating between close relatives or because of geographic isolation in a population. This will cause deviations in HWE by causing a deficit of heterozygotes.

How to quantify the amount of inbreeding in a population?

- Inbreeding coefficient, ***F***
- The probability that a random individual in the population inherits two copies of the same allele from a common ancestor
- *F* ranges 0 to 1:

  F is low in random mating populations

  F close to 1 in self-breeding population (plants)

The amount of inbreeding between two *individuals*

en.wikipedia.org/wiki/Inbreeding

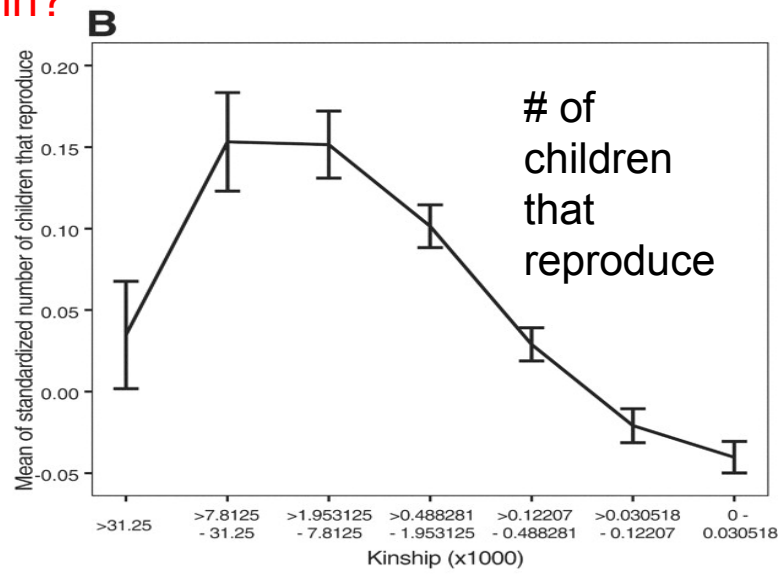
Typical inbreeding coefficient percentages are as follows, assuming no previous inbreeding between any parents:

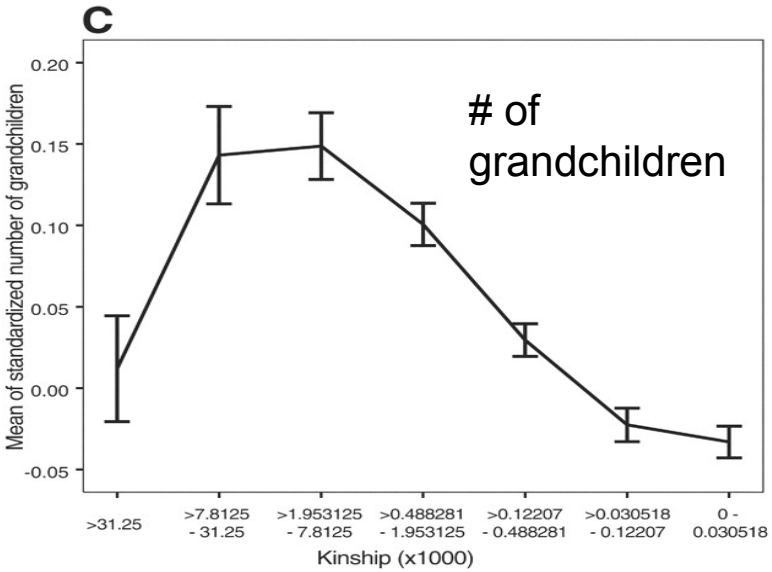
- Father/daughter, mother/son or brother/sister → 25% ($1/4$)
- Grandfather/granddaughter or grandmother/grandson → 12.5% ($1/8$)
- Half-brother/half-sister → 12.5% ($1/8$)
- Uncle/niece or aunt/nephew → 12.5% ($1/8$)
- Great-grandfather/great-granddaughter or great-grandmother/great-grandson → 6.25% ($1/16$)
- Half-uncle/niece or half-aunt/nephew → 6.25% ($1/16$)
- First cousins → 6.25% ($1/16$)
- First cousins once removed or half-first cousins → 3.125% ($1/32$)
- Second cousins or first cousins twice removed → 1.5625% ($1/64$)
- Second cousins once removed or half-second cousins → 0.78125% ($1/128$)
- Third cousins or second cousins twice removed → 0.390625% ($1/256$)
- Third cousins once removed or half-third cousins → 0.1953125% ($1/512$)

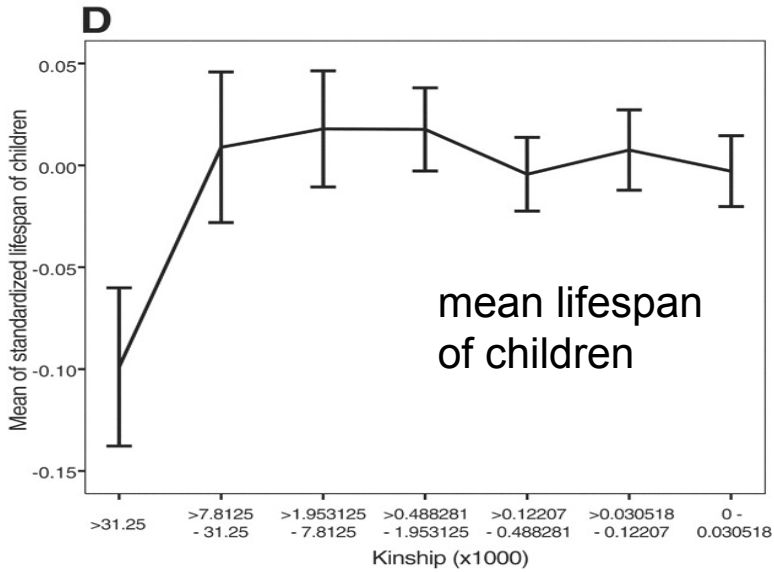
Kinship & Reproduction: Icelandic couples

Should you marry your 3rd-cousin?

"The first interval of kinship represents all couples related at the level of second cousins or closer, the second interval represents couples related at the level of third cousins and up to the level of second cousins, and so on, with each subsequent category representing steps to fourth, fifth, sixth, and seventh cousins and the final category representing couples with no known relationship and those with relationships up to the level of eighth cousins."







Helgason, *Science*, 2008