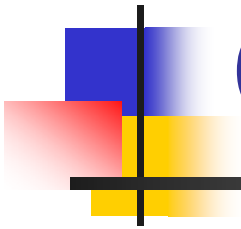


# Genetic Theory - Overview

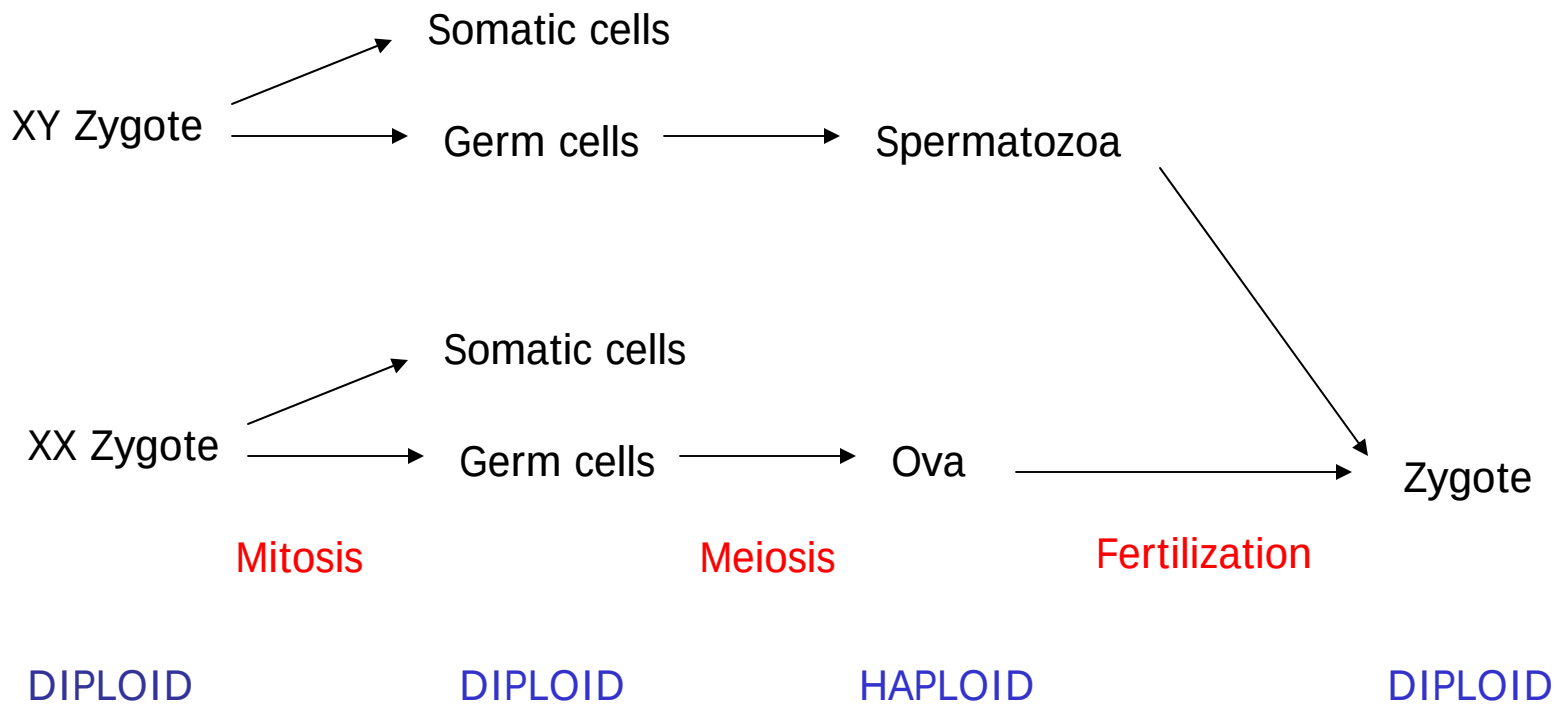
Pak Sham

International Twin Workshop
Boulder, 2005

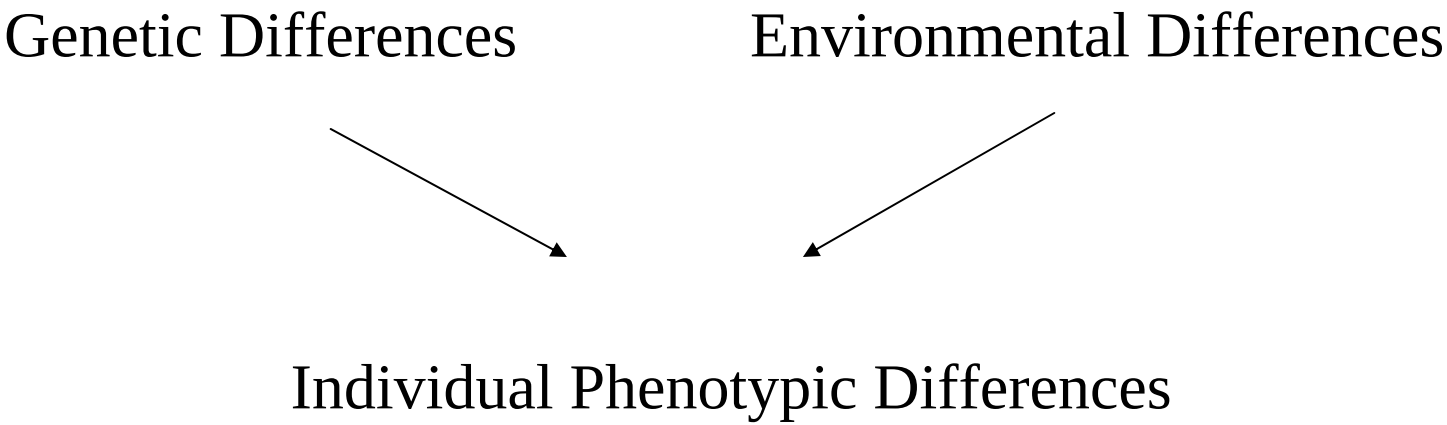
# The Human Genome

- 23 Chromosomes, each containing a DNA molecule (Watson and Crick, 1953)
- $3 \times 10^9$ base pairs, completely sequenced (Human Genome Project, 2003)
- Approximately 24,000 genes, each coding for a polypeptide chain
- Approximately $10^7$ common polymorphisms (variable sites, documented in dbSNP database)

# Genetic transmission



# Sources of Natural Variation



# Genetic Variation

- Chromosomal anomalies
- Insertions / Deletions / Translocations
- Variable sequence repeats
  - microsatellites (e.g. CACACA….)
- Single nucleotide polymorphisms (SNPs)

# Types of Genetic Disease

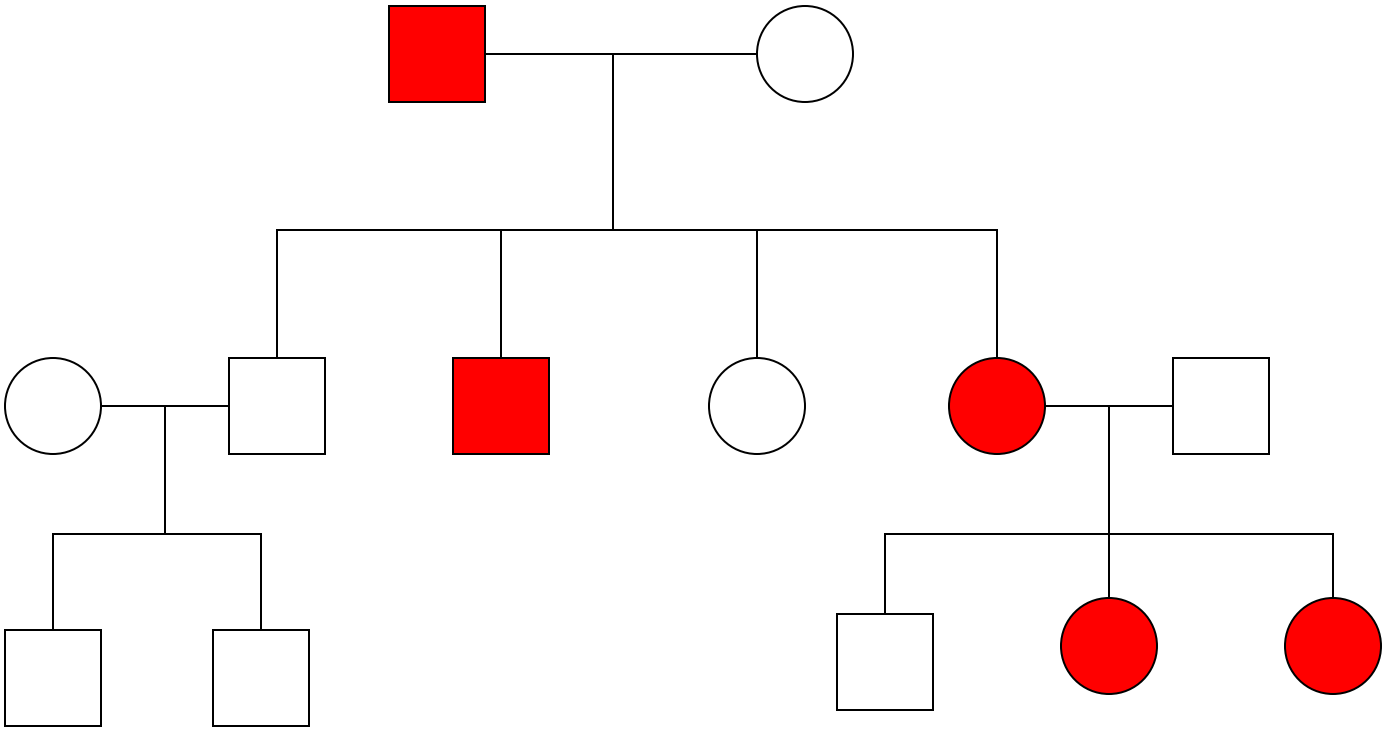
- Mendelian diseases
  - e.g. Huntington's disease, cystic fibrosis
  - A genetic mutation causes the disease
  - Environmental variation usually irrelevant
  - Usually rare
  - Occurs in isolated pedigrees

- Multifactorial diseases
  - e.g. Coronary heart disease, hypertension, schizophrenia
  - A genetic variant increases the risk of disease
  - Environmental variation usually important
  - Often common
  - Occurs in general population

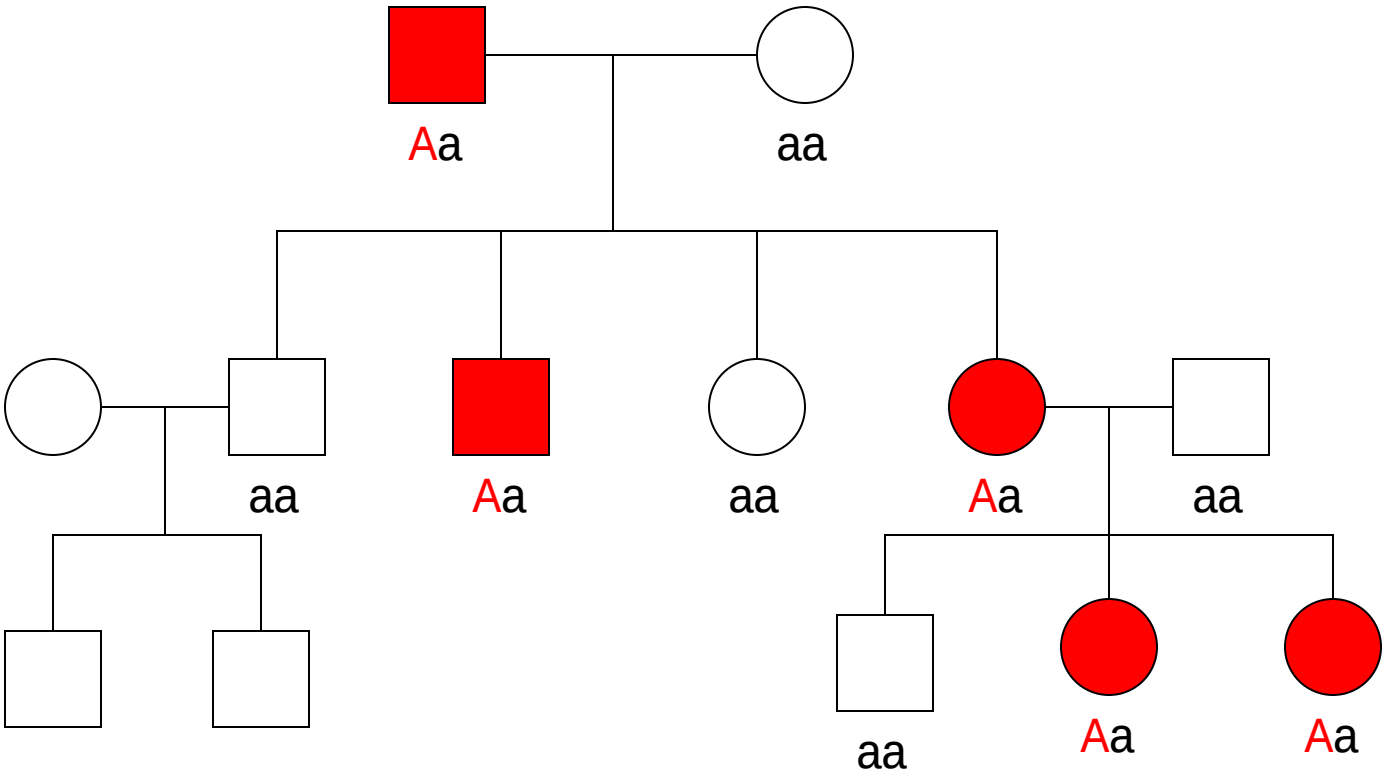
# Single-Gene Disorders

- Human Genome Project completed in 2003
- Human Gene Mutation Database contains 44,090 mutations in 1,714 genes
- Gene Test web site lists genetic tests for 1,093 diseases
- dbSNP Database Build 123 contains 10,079,771 single nucleotide polymorphisms

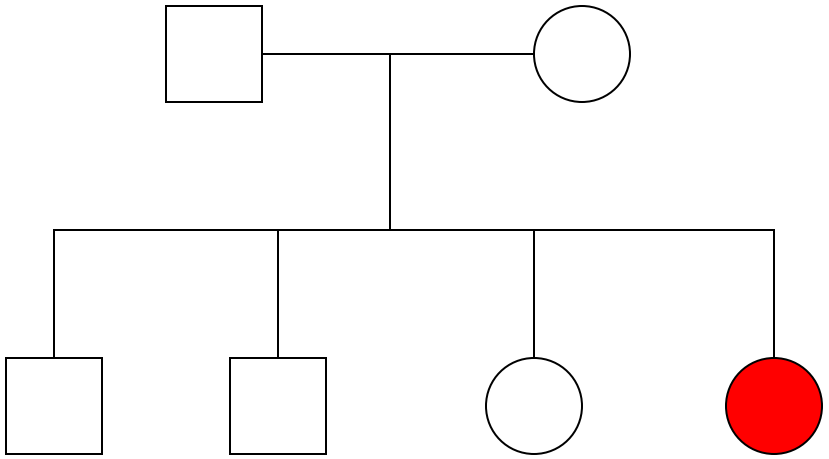
# Autosomal Dominant Disorders



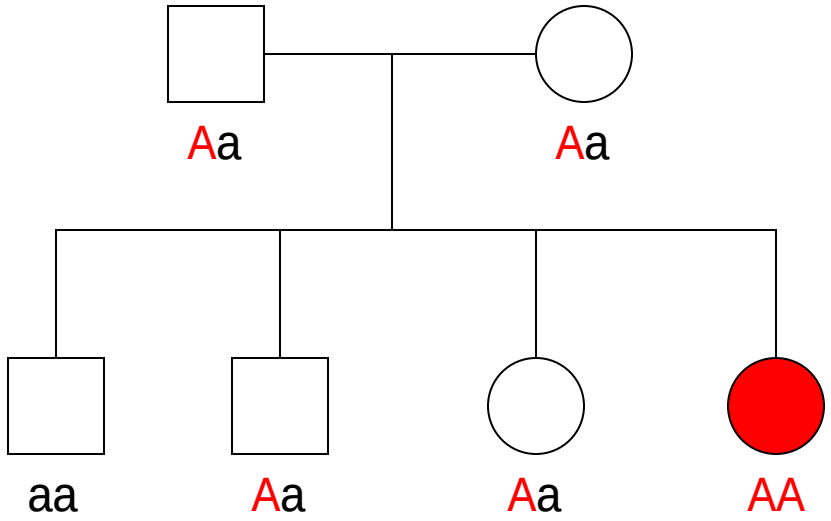
# Autosomal Dominant Disorders



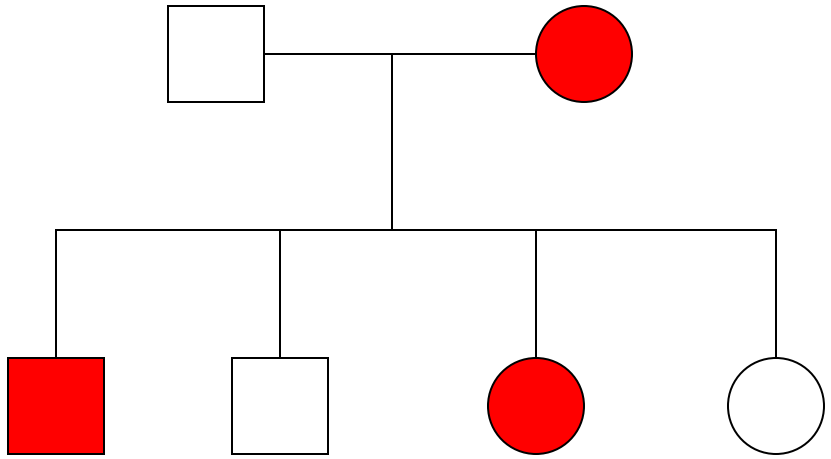
# Autosomal Recessive Disorders



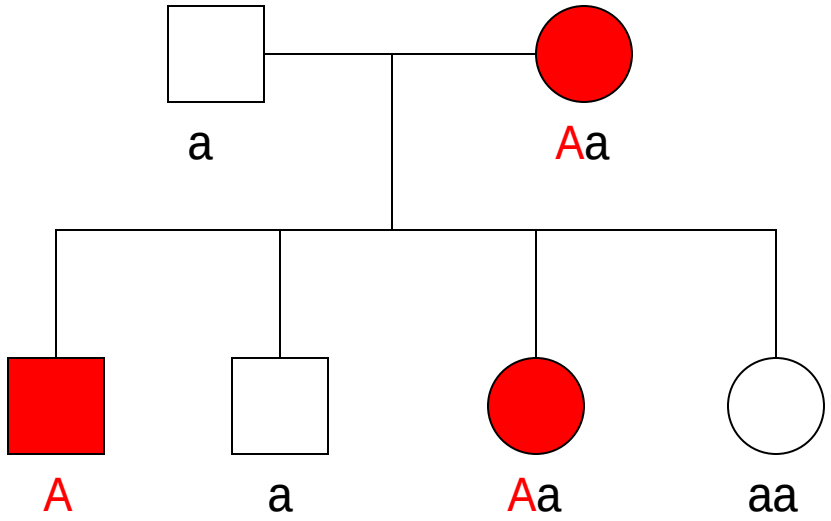
# Autosomal Recessive Disorders



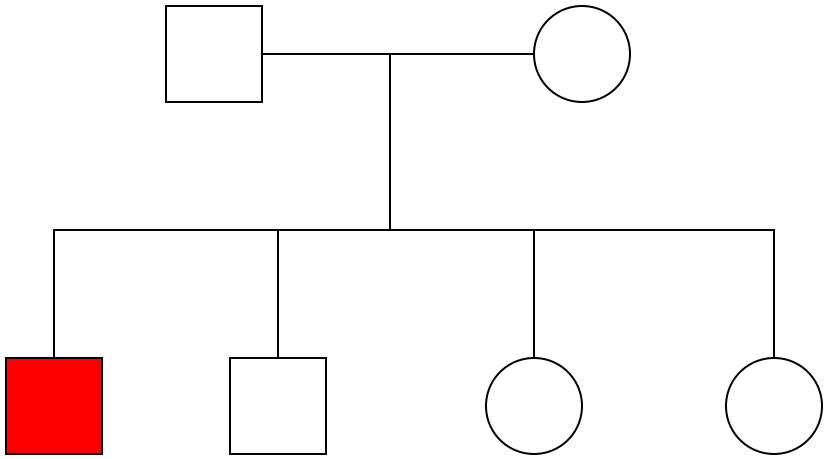
# X-linked Dominant Disorders



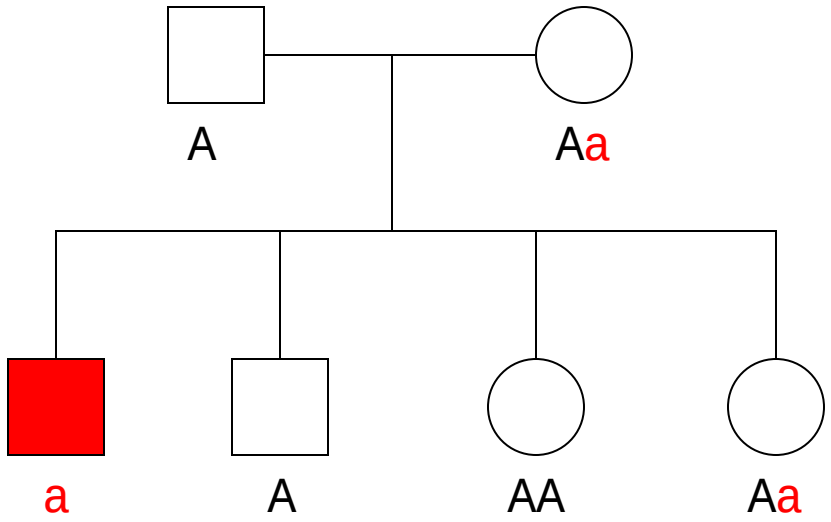
# X-linked Dominant Disorders



# X-linked Recessive Disorders



# X-linked Recessive Disorders



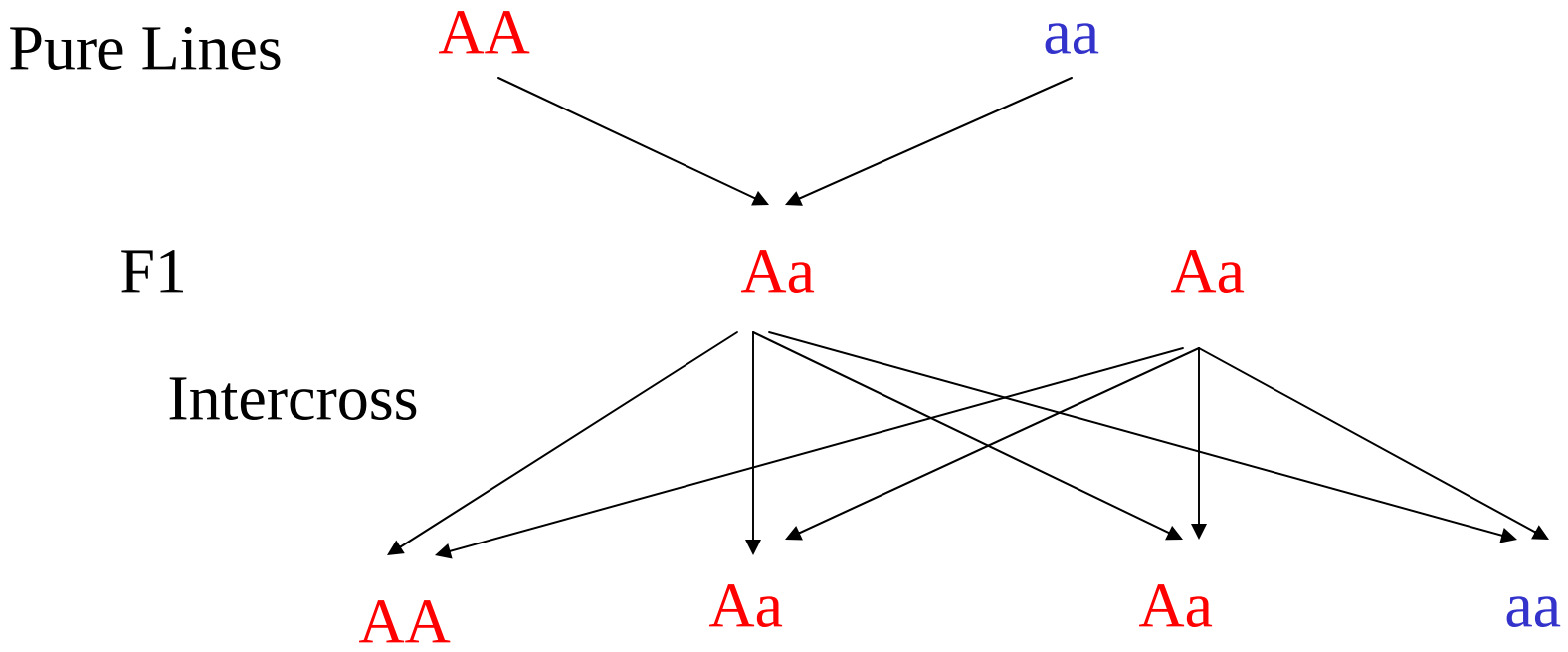
# Mendelian Segregation

# Segregation Ratios

- First discovered by Gregor Mendel in his experiments on the garden pea (published in 1866 and rediscovered in 1900)
- Form the basis of Mendel's first law:

  "law of segregation"
- Defined as the ratio of affected to normal individuals among the offspring of a particular type of mating.

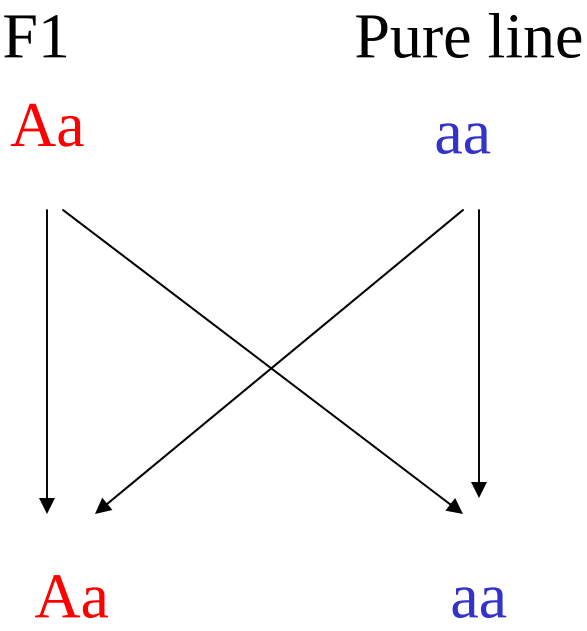
# Mendel's Experiments




3:1 Segregation Ratio

# Mendel's Experiments

Back cross




1:1 Segregation ratio

# Segregation Ratios

| Mode of inheritance | Mating type | Segregation ratio Affected:Normal |
|---|---|---|
| Autosomal dominant | Affected x Normal | |
| Autosomal recessive | Carrier x Carrier | |
| X-linked dominant | Normal father x Affected mother | |
| X-linked recessive | Normal father x Carrier mother | |

# Segregation Ratios

| Mode of inheritance | Mating type | Segregation ratio Affected:Normal |
|---|---|---|
| Autosomal dominant | Affected x Normal | 1:1 |
| Autosomal recessive | Carrier x Carrier | |
| X-linked dominant | Normal father x Affected mother | |
| X-linked recessive | Normal father x Carrier mother | |

# Segregation Ratios

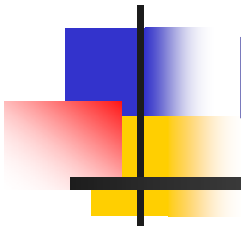
| Mode of inheritance | Mating type | Segregation ratio Affected:Normal |
|---|---|---|
| Autosomal dominant | Affected x Normal | 1:1 |
| Autosomal recessive | Carrier x Carrier | 1:3 |
| X-linked dominant | Normal father x Affected mother | |
| X-linked recessive | Normal father x Carrier mother | |

# Segregation Ratios

| Mode of inheritance | Mating type | Segregation ratio Affected:Normal |
|---|---|---|
| Autosomal dominant | Affected x Normal | 1:1 |
| Autosomal recessive | Carrier x Carrier | 1:3 |
| X-linked dominant | Normal father x Affected mother | 1:1 |
| X-linked recessive | Normal father x Carrier mother | |

# Segregation Ratios

| Mode of inheritance | Mating type | Segregation ratio Affected:Normal |
|---|---|---|
| Autosomal dominant | Affected x Normal | 1:1 |
| Autosomal recessive | Carrier x Carrier | 1:3 |
| X-linked dominant | Normal father x Affected mother | 1:1 |
| X-linked recessive | Normal father x Carrier mother | 1:1 in sons |



# Hardy-Weinberg Law

# Parental Frequencies

| Genotype | Frequency |
|---|---|
| AA | P |
| Aa | Q |
| aa | R |

| Allele | Frequency |
|---|---|
| A | P+Q/2 |
| a | R+Q/2 |

# Mating Type Frequencies (Random Mating)

| | **AA** | **Aa** | **aa** |
|---|---|---|---|
| **AA** | $P^2$ | PQ | PR |
| **Aa** | PQ | $Q^2$ | QR |
| **aa** | PR | QR | $R^2$ |

# Offspring Segregation Ratios

| | **AA** | **Aa** | **aa** |
|---|---|---|---|
| **AA** | AA | AA:Aa<br>0.5:0.5 | Aa |
| **Aa** | AA:Aa<br>0.5:0.5 | AA:Aa:aa<br>0.25:0.5:0.25 | Aa:aa<br>0.5:0.5 |
| **aa** | Aa | Aa:aa<br>0.5:0.5 | aa |

# Offspring Genotype Frequencies

| Genotype | Frequency |
|---|---|
| AA | $P^2+PQ+Q^2/4 = (P+Q/2)^2$ |
| Aa | $2PR+PQ+QR+Q^2/2 = 2(P+Q/2)(R+Q/2)$ |
| aa | $R2+QR+Q^2/4 = (R+Q/2)^2$ |

# Offspring Allele Frequencies

| Allele | Frequency |
|---|---|
| A | $(P+Q/2)^2 + (P+Q/2)(R+Q/2) = P+Q/2$ |
| a | $(R+Q/2)^2 + (P+Q/2)(R+Q/2) = R+Q/2$ |

# Hardy-Weinberg Equilibrium

In a large population under random mating:

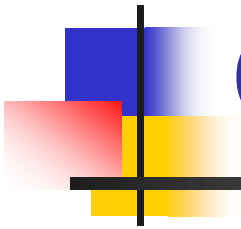
- Allele frequencies in the offspring, denoted as **p** and **q**, are the same as those in the parental generation.
- Genotype frequencies in the offspring will follow the ratios $\mathbf{p^2}$:**2pq**:$\mathbf{q^2}$, regardless of the genotype frequencies in the parents.

# Hardy-Weinberg Equilibrium

| | A | a | |
|---|---|---|---|
| A | $p^2$ | pq | p |
| a | pq | $q^2$ | q |
| | p | q | |

# Hardy-Weinberg Disequilibrium

| | A | a | |
|---|---|---|---|
| A | $p^2+d$ | pq-d | p |
| a | pq-d | $q^2+d$ | q |
| | p | q | |



# Genetic Linkage

# Genetic Markers

- Classical
  - Mendelian Disorders
  - Blood groups
  - HLA Antigens
- Molecular genetic
  - Microsatellites (e.g. CACACA... )
  - Single-nucleotide polymorphisms (e.g. C/T)

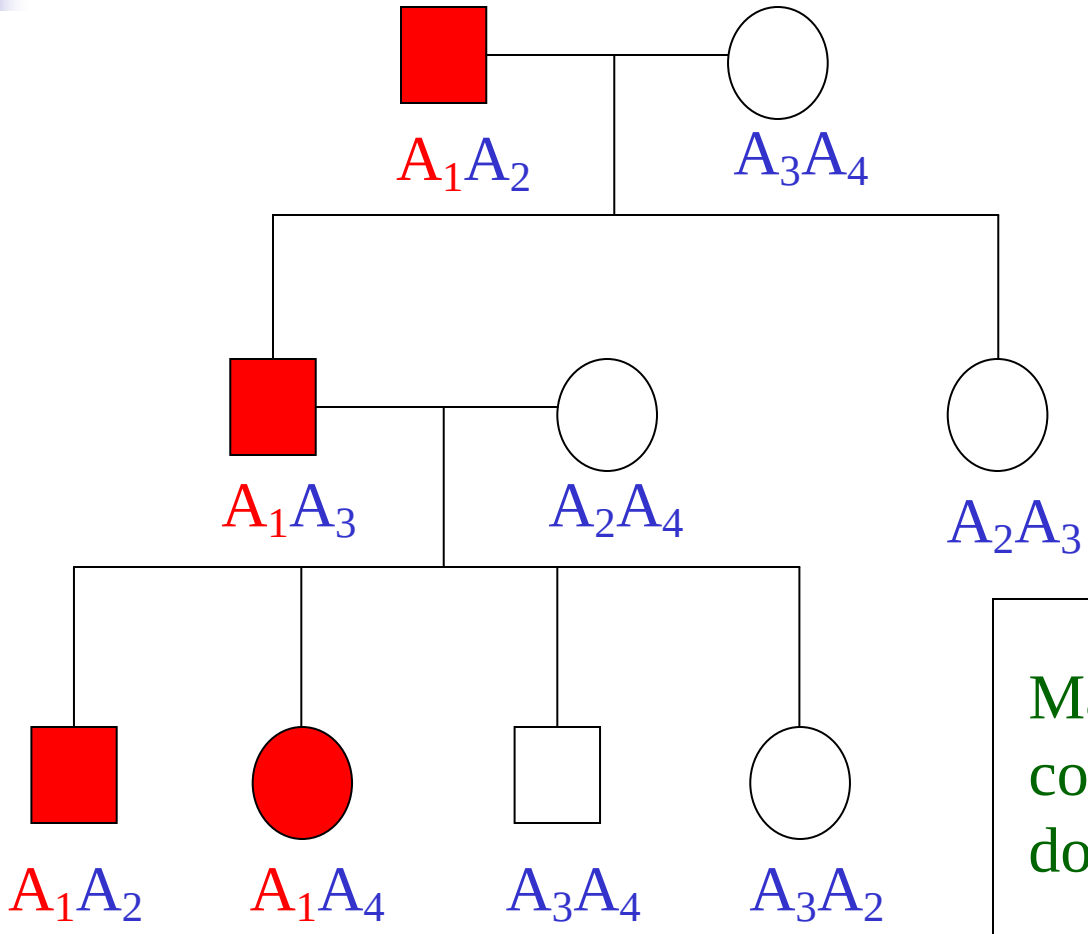
# High-Throughput Genotyping

- Extreme multiplexing (multiple markers)
- DNA Pooling (multiple samples)

Maximum throughput of SEQUENOM system at the HKU Genome Research Centre is 100,000 genotypes / day, at a cost of US$ 0.2 per genotype

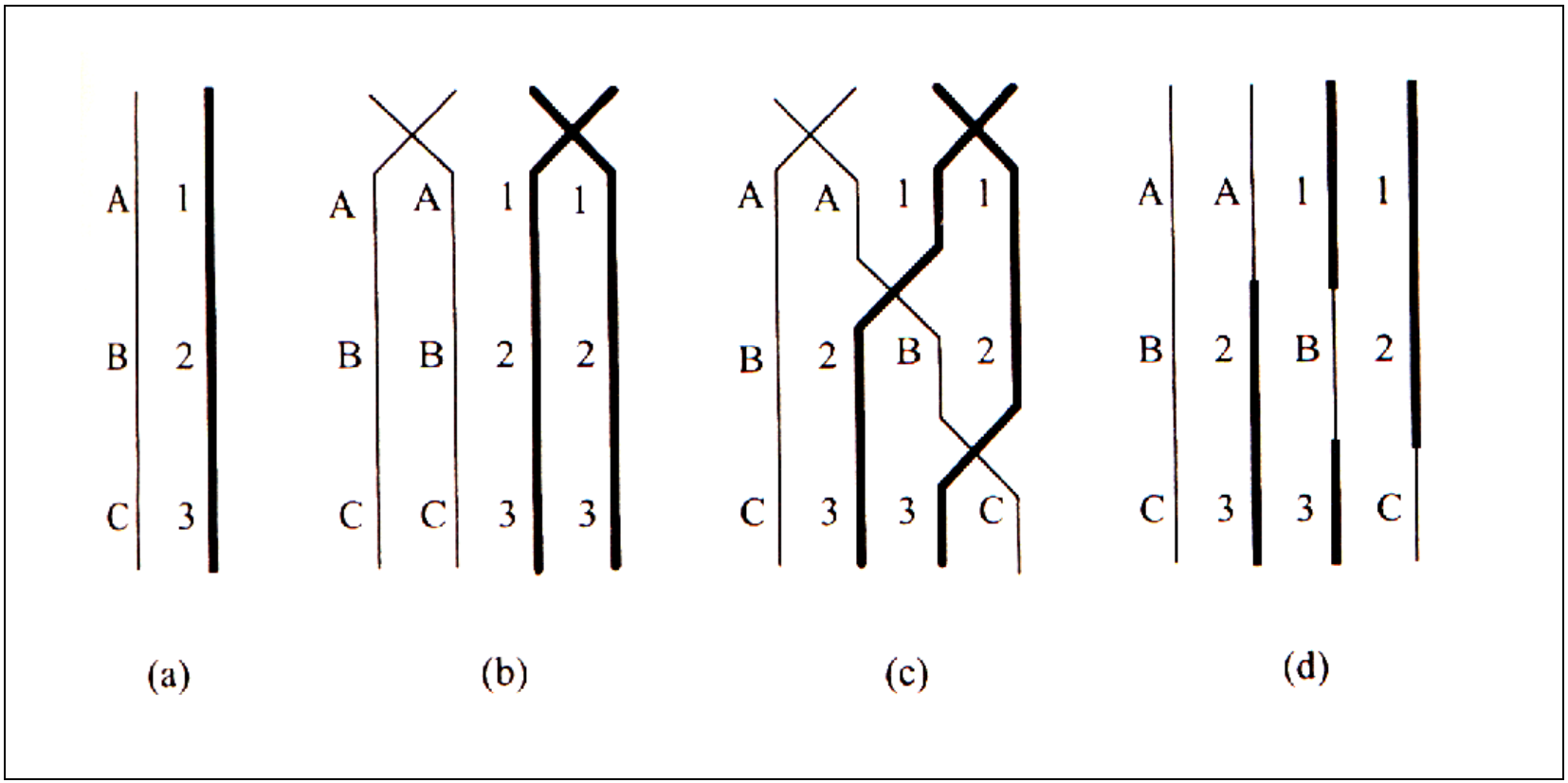
Cost of genotyping set to decrease further – eventually enabling whole-genome association studies to be done.

# Linkage = Co-segregation



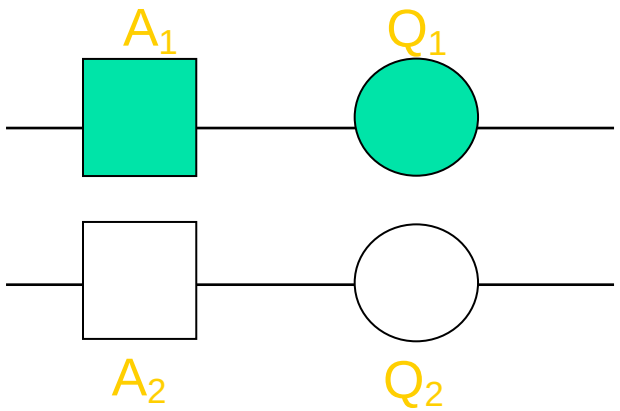

Marker allele $A_1$ cosegregates with dominant disease

# Crossing-over in meiosis

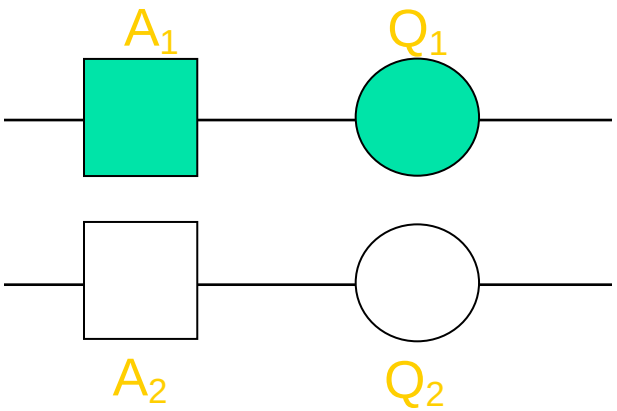


# Recombination

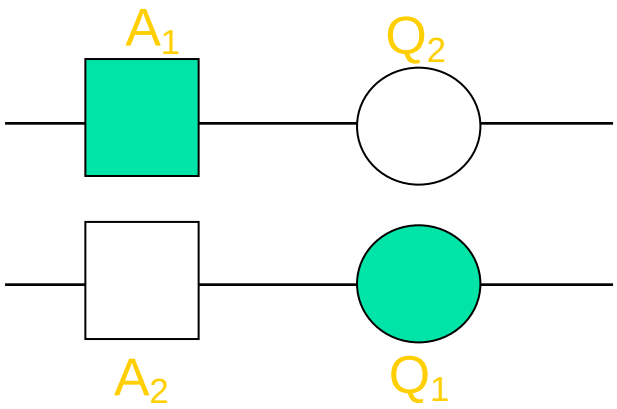
Parental genotypes




Likely gametes
(Non-recombinants)




Unlikely gametes
(Recombinants)

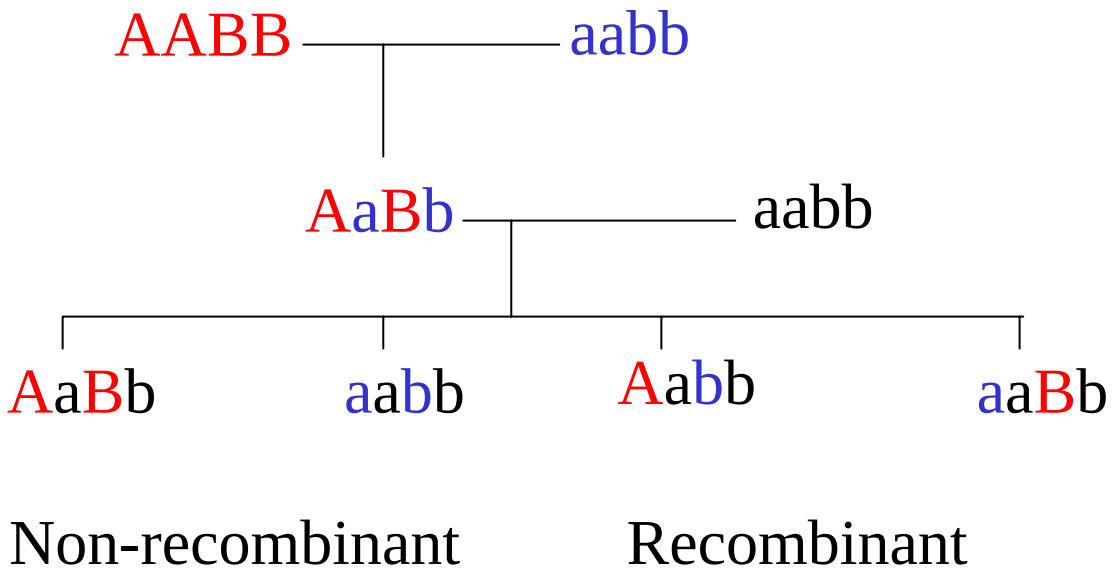


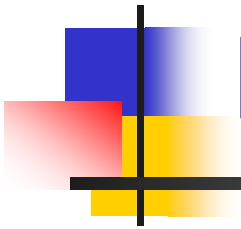
# Recombination fraction

Recombination fraction between two loci

= Proportion of gametes that are recombinant with respect to the two loci

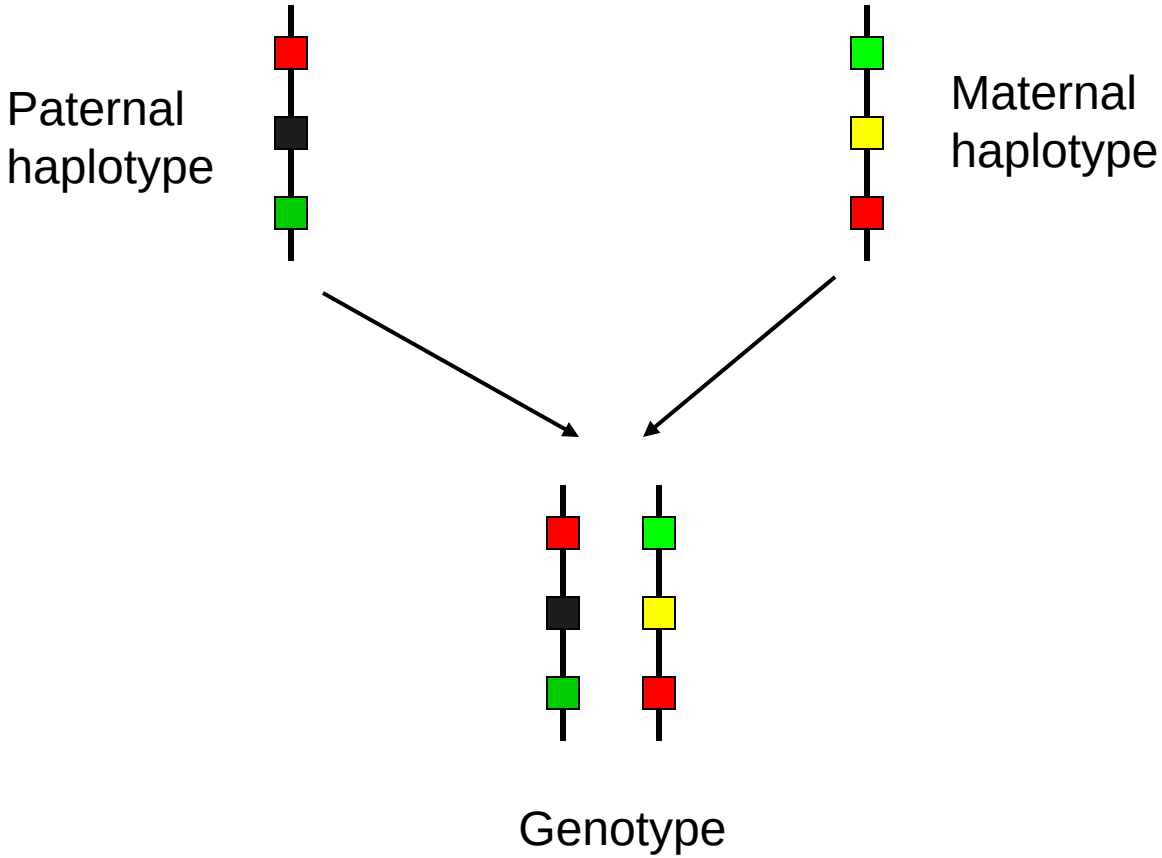
# Double Backcross : Fully Informative Gametes





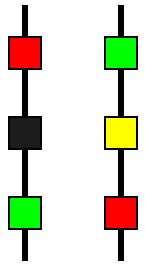
# Haplotypes

# Haplotypes

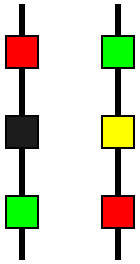


# Recombination

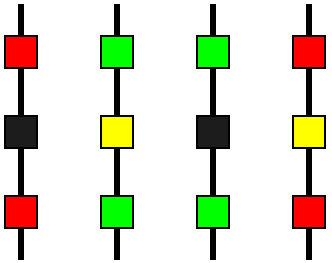
**Parental haplotypes**



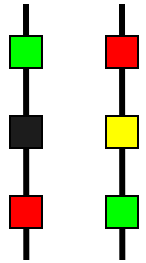
**Possible transmitted haplotypes**



Non-recombinants



Single recombinants



Double recombinants

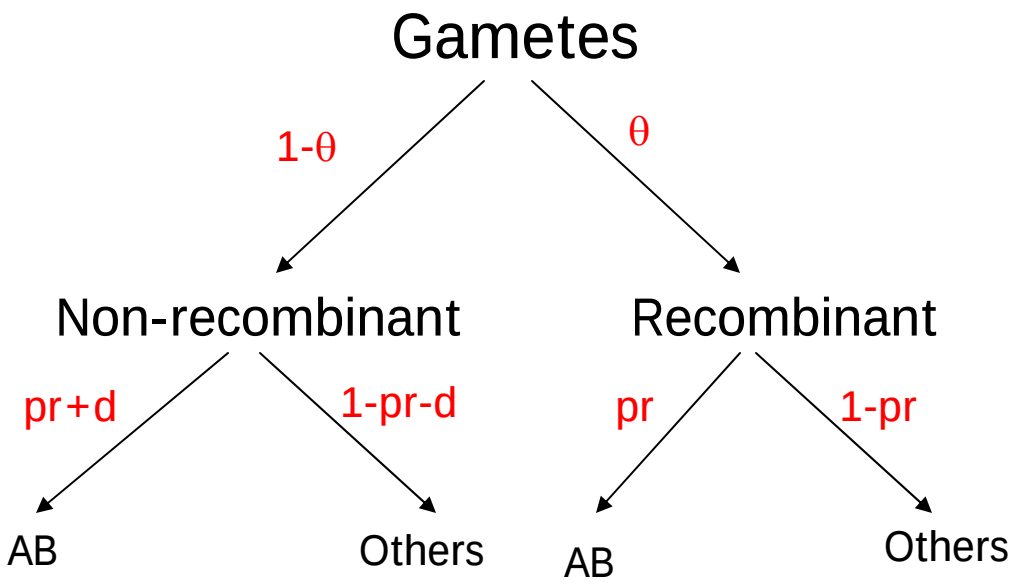
# Linkage Equilibrium

| | B | b | |
|---|---|---|---|
| A | pr | ps | p |
| a | qr | qs | q |
| | r | s | |

# Linkage Disequilibrium (LD)

| | B | b | |
|---|---|---|---|
| A | pr+d | ps-d | p |
| a | qr-d | qs+d | q |
| | r | S | |

# Decay of LD




Frequency of AB gametes = $(1-\theta)(pr+d)+\theta pr = pr+(1-\theta)d$

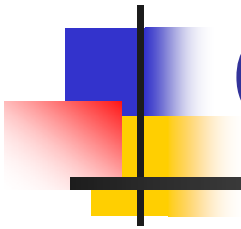
# Single-Gene Disorders: Some Historical Landmarks

- 1902: First identified single-gene disorder - alkaptonuria
- 1956: First identified disease-causing amino acid change: sickle-cell anaemia
- 1961: First screening program: phenylketonuria
- 1983: First mapped to chromosomal location: Huntington's disease
- 1986: First positionally cloned - chronic granulomatous disease, Duchenne muscular dystrophy
- 1987: First autosomal recessive disease cloned – cystic fibrosis

# Types of Genetic Disease

- Mendelian diseases
  - e.g. Huntington's disease, cystic fibrosis
  - A genetic mutation causes the disease
  - Environmental variation usually irrelevant
  - Usually rare
  - Occurs in isolated pedigrees

- Multifactorial diseases
  - e.g. Coronary heart disease, hypertension, schizophrenia
  - A genetic variant increases the risk of disease
  - Environmental variation usually important
  - Often common
  - Occurs in general population



# Genetic Study Designs

# Family Studies

Case – Control Family Design

Compares risk in relatives of case and controls

Some terminology

- Proband
- Secondary case
- Lifetime risk / expectancy (morbid risk)

Problem: Familial aggregation can be due to shared family environment as well as shared genes

# Family Studies: Schizophrenia

| Relationship to Proband | Lifetime Risk of Schizophrenia (%) |
|---|---|
| Unrelated | 1 |
| First cousins | 2 |
| Uncles/Aunts | 2 |
| Nephews/Nieces | 4 |
| Grandchildren | 5 |
| Half siblings | 6 |
| Parents | 6 |
| Siblings | 9 |
| Children | 13 |

From: Psychiatric Genetics and Genomics. MuGuffin, Owen & Gottesman, 2002

# Twin Studies

Studies risk of disease (concordance rates) in cotwins of affected MZ and DZ Twin

Under the equal environment assumption, higher MZ than DZ concordance rate implies genetic factors

Problems:

- Validity of equal environment assumption
- Generalizability of twins to singletons

# Twin Studies: Schizophrenia

| Zygosity | Concordance (%) |
|---|---|
| Dizygotic (DZ) | 17 |
| Monozygotic (MZ) | 48 |

From: Psychiatric Genetics and Genomics. MuGuffin, Owen & Gottesman, 2002

# Adoption Studies

Adoptees' method compares

- Adoptees with an affected parent
- Adoptees with normal parents

Adoptee's family method compares

- Biological relatives of adoptees
- Adoptive relatives of adoptees

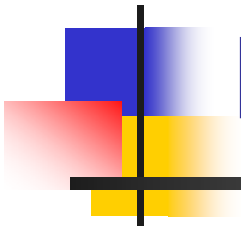
Problems:

- Adoption correlated with ill-health/psychopathology in parents
- Adoptive parents often rigorously screened

# Adoption Studies: Schizophrenia

| Adoptees of | Risk of Schizophrenia (%) |
|---|---|
| Schizophrenic parents | 8 |
| Control parents | 2 |

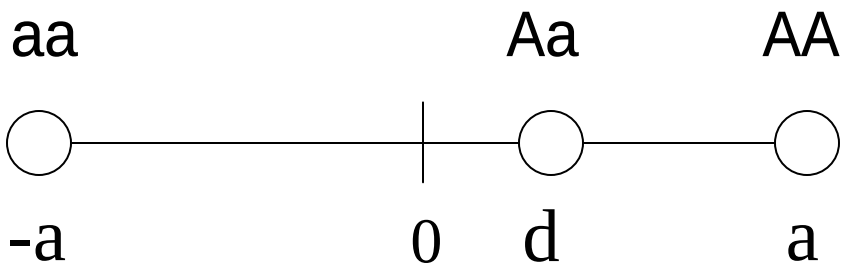
From: Finnish Adoption Study, as summarised in Psychiatric Genetics and Genomics. MuGuffin, Owen & Gottesman, 2002



# Biometrical Genetic Model

# Biometrical Genetic Model

| | Trait Means |
|---|---|
| AA | a |
| Aa | d |
| aa | -a |




Fisher's convention: the midpoint between the trait means of the two homozygous genotypes is designated as 0

# Biometrical Model: Mean

| **Genotype** | AA | Aa | aa |
|---|---|---|---|
| **Frequency** | $p^2$ | 2pq | $q^2$ |
| **Trait (x)** | a | d | -a |

**Mean** $m = p^2(a) + 2pq(d) + q^2(-a)$

$= (p-q)a + 2pqd$

# Biometrical Model: Variance

| **Genotype** | AA | Aa | aa |
|---|---|---|---|
| **Frequency** | $p^2$ | $2pq$ | $q^2$ |
| **$(x-m)^2$** | $(a-m)^2$ | $(d-m)^2$ | $(-a-m)^2$ |

**Variance** $= (a-m)^2p^2 + (d-m)^2 2pq + (-a-m)^2q^2$

$= 2pq[a+(q-p)d]^2 + (2pqd)^2$

$= V_A + V_D$

# Average Allelic Effect

Effect of gene substitution: $a \rightarrow A$

If background allele is a, then effect is (a+d)
If background allele is A, then effect is (a-d)

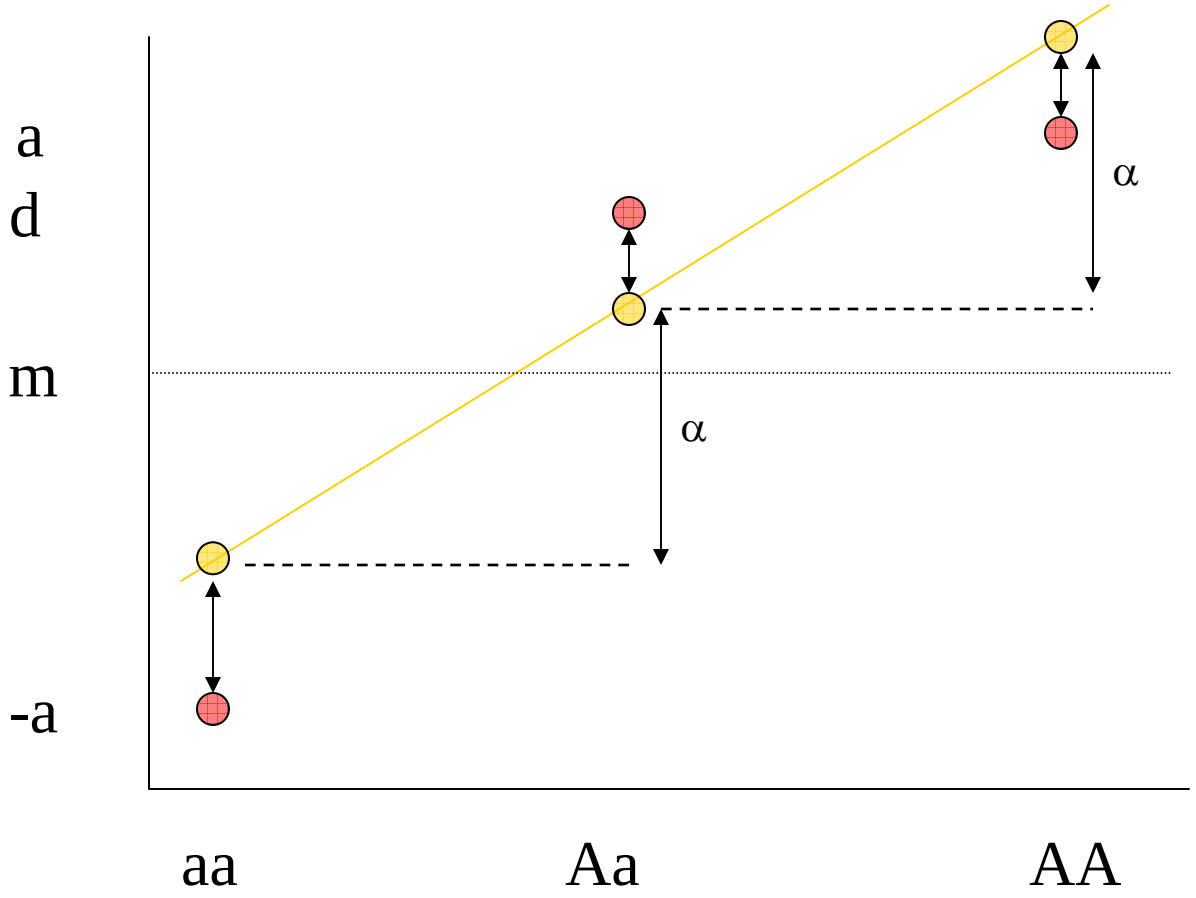
Average effect of gene substitution is therefore

$$\alpha = q(a+d) + p(a-d) = a + (q-p)d$$

Additive genetic variance is therefore

$$V_A = 2pq\alpha^2$$

# Additive and Dominance Variance



# Cross-Products of Deviations for Pairs of Relatives

| | **AA** | **Aa** | **aa** |
|---|---|---|---|
| **AA** | $(a-m)^2$ | | |
| **Aa** | $(a-m)(d-m)$ | $(d-m)^2$ | |
| **aa** | $(a-m)(-a-m)$ | $(-a-m)(d-m)$ | $(-a-m)^2$ |

The covariance between relatives of a certain class is the weighted average of these cross-products, where each cross-product is weighted by its frequency in that class.

# Covariance of MZ Twins

| | **AA** | **Aa** | **aa** |
|---|---|---|---|
| **AA** | $p^2$ | | |
| **Aa** | 0 | 2pq | |
| **aa** | 0 | 0 | $q^2$ |

$$\text{Covariance} = (a-m)^2p^2 + (d-m)^2 2pq + (-a-m)^2q^2$$
$$= 2pq[a+(q-p)d]^2 + (2pqd)^2$$
$$= V_A + V_D$$

# Covariance for Parent-offspring (P-O)

| | **AA** | **Aa** | **aa** |
|---|---|---|---|
| **AA** | $p^3$ | | |
| **Aa** | $p^2q$ | $pq$ | |
| **aa** | 0 | $pq^2$ | $q^3$ |

$$\text{Covariance} = (a-m)^2p^3 + (d-m)^2pq + (-a-m)^2q^3 + (a-m)(d-m)2p^2q + (-a-m)(d-m)2pq^2$$
$$= pq[a+(q-p)d]^2$$
$$= V_A / 2$$

# Covariance for Unrelated Pairs (U)

| | **AA** | **Aa** | **aa** |
|---|---|---|---|
| **AA** | $p^4$ | | |
| **Aa** | $2p^3q$ | $4p^2q^2$ | |
| **aa** | $p^2q^2$ | $2pq^3$ | $q^4$ |

$$\text{Covariance} = (a-m)^2p^4 + (d-m)^2 4p^2q^2 + (-a-m)^2q^4 + (a-m)(d-m)4p^3q + (-a-m)(d-m)4pq^3 + (a-m)(-a-m)2p^2q^2 = 0$$

# Covariance for DZ twins

Genotype frequencies are weighted averages:

¼ MZ twins (when $\pi$=1)

½ Parent-offspring (when $\pi$ =1)

¼ Unrelated (when $\pi$ =0)

Covariance = $\frac{1}{4}(V_A+V_D) + \frac{1}{2}(V_A/2) + \frac{1}{4}(0)$

= $\frac{1}{2}V_A + \frac{1}{4}V_D$

# Covariance: General Relative Pair

$$
\begin{aligned}
\text{Covariance} &= \text{Prob}(\pi=1)(V_A+V_D) \\
&\quad + \text{Prob}(\pi=1/2)(V_A/2) \\
&\quad + \text{Prob}(\pi=0)(0) \\
&= (\text{Prob}(\pi=1)+\text{Prob}(\text{IBD}=1/2)/2)V_A \\
&\quad + \text{Prob}(\pi=1)\,V_D \\
&= E(\pi)V_A + \text{Prob}(\pi=1)V_D \\
&= 2\Phi V_A + \Delta V_D
\end{aligned}
$$

# Total Genetic Variance

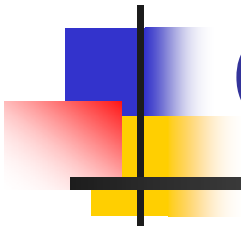
Heritability is the combined effect of all loci

total component = sum of individual loci components

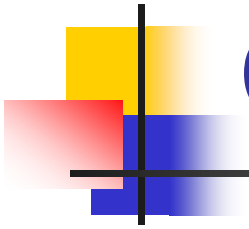
$$\mathbf{V}_A = V_{A1} + V_{A2} + \ldots + V_{AN}$$

$$\mathbf{V}_D = V_{D1} + V_{D2} + \ldots + V_{DN}$$

| Correlations | | MZ | DZ | P-O | U |
|---|---|---|---|---|---|
| $\mathbf{V}_A$ $(2\Phi)$ | 1 | 0.5 | 0.5 | 0 | |
| $\mathbf{V}_D$ $(\Delta)$ | | 1 | 0.25 | 0 | 0 |



# Quantitative Genetics

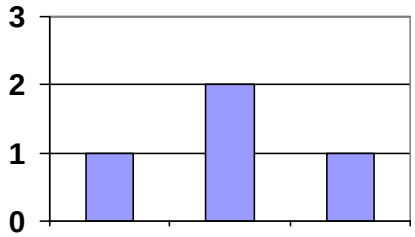


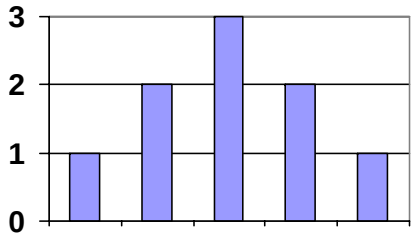
# Quantitative Genetics

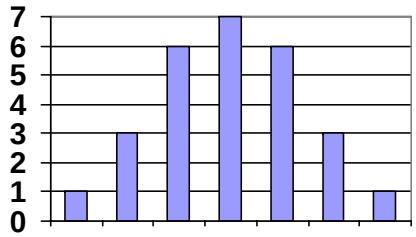
- Examples of quantitative traits
  - Blood Pressure (BP)
  - Body Mass Index (BMI)
  - Blood Cholesterol Level
  - General Intelligence (G)
- Many quantitative traits are relevant to health and disease

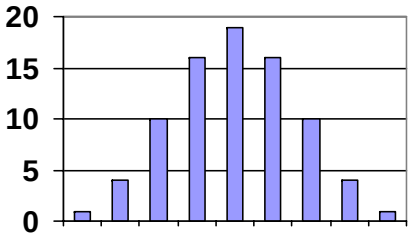
# Quantitative Traits

| 1 Gene | 2 Genes | 3 Genes | 4 Genes |
|---|---|---|---|
| → 3 Genotypes | → 9 Genotypes | → 27 Genotypes | → 81 Genotypes |
| → 3 Phenotypes | → 5 Phenotypes | → 7 Phenotypes | → 9 Phenotypes |



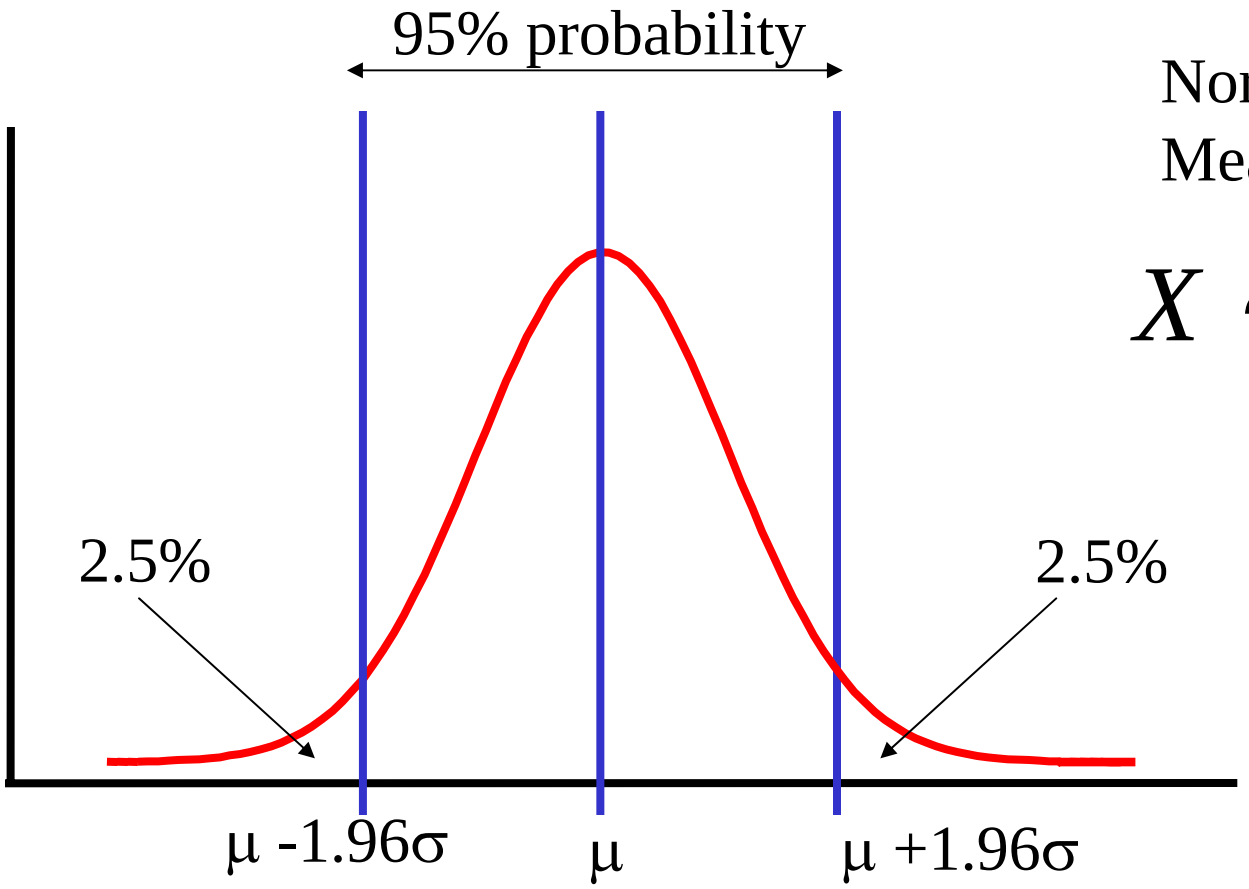




Central Limit Theorem → Normal Distribution

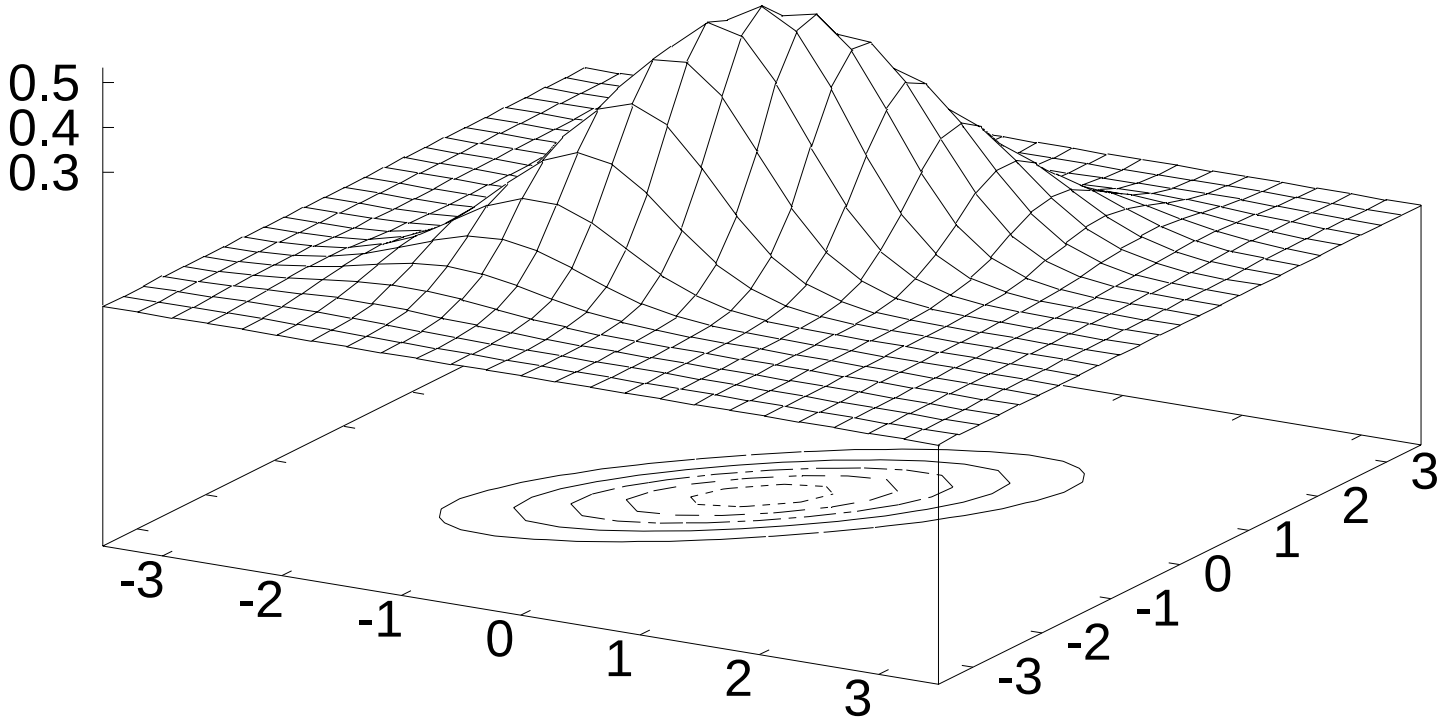
# Continuous Variation



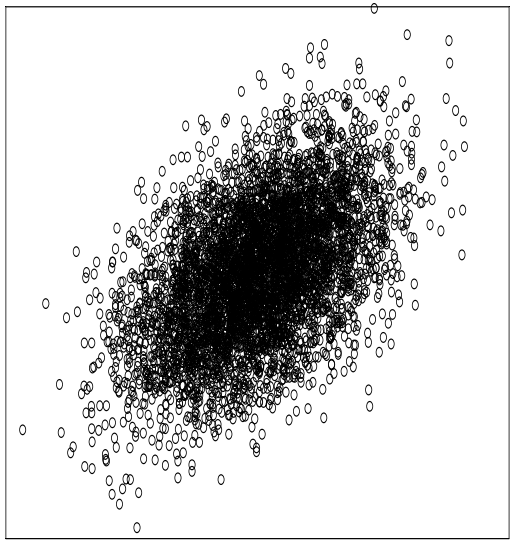

Normal distribution
Mean μ, variance $\sigma^2$

$$X \sim N(\mu, \sigma^2)$$

# Bivariate normal



# Familial Covariation




*Bivariate normal disttribution*

$$X \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}$$

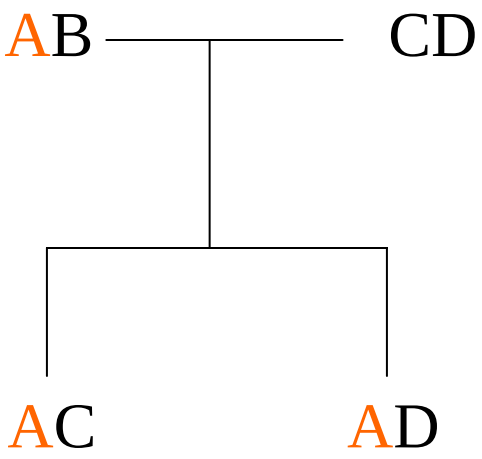
# Correlation due to Shared Factors

Francis Galton: Two Journeys starting at same time


Journey Times: A+B and A+C

Shared A ⟶ Covariance ⟶ Correlation

# Shared Genes




Gene A is shared:

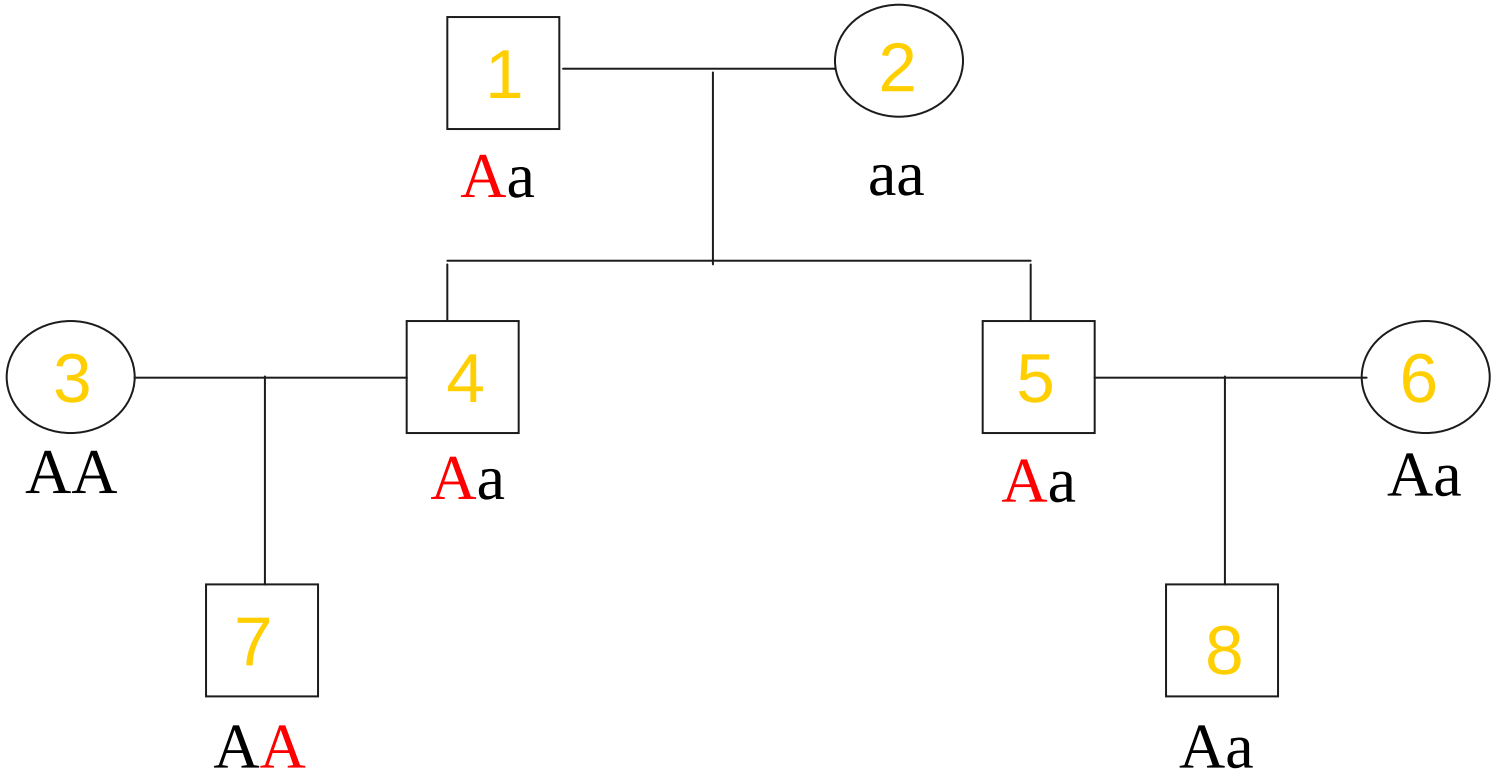
= Identity-By-Descent (IBD)

$\Rightarrow$ Shared Phenotypic Effects

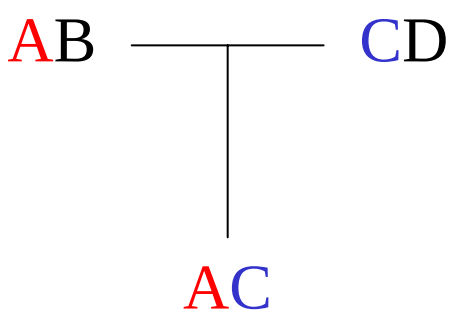
At any chromosomal location, two individuals can share 0, 1 or 2 alleles.

# Identity by Descent (IBD)

- Two alleles are IBD if they are descended from and replicates of the same ancestral allele

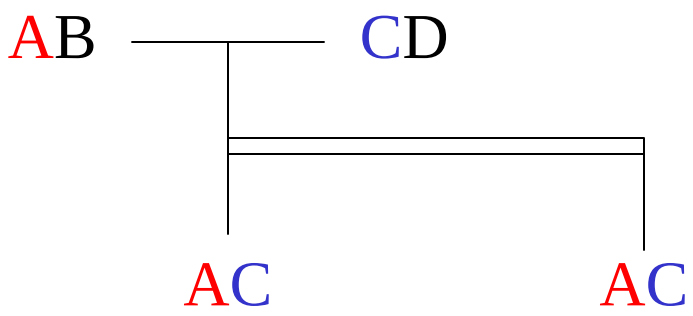


# IBD: Parent-Offspring



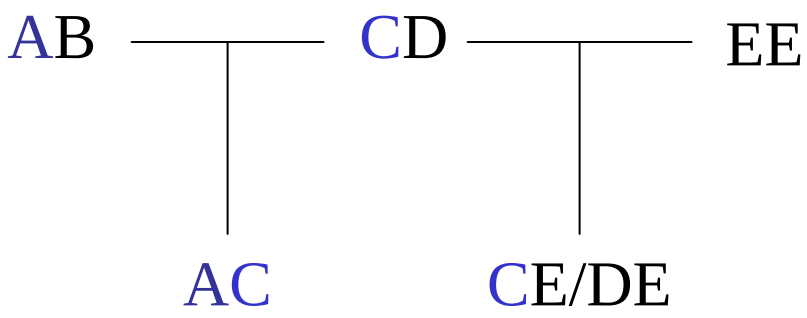

If the parents are unrelated,
then parent-offspring pairs always share 1 allele IBD

# IBD: MZ Twins



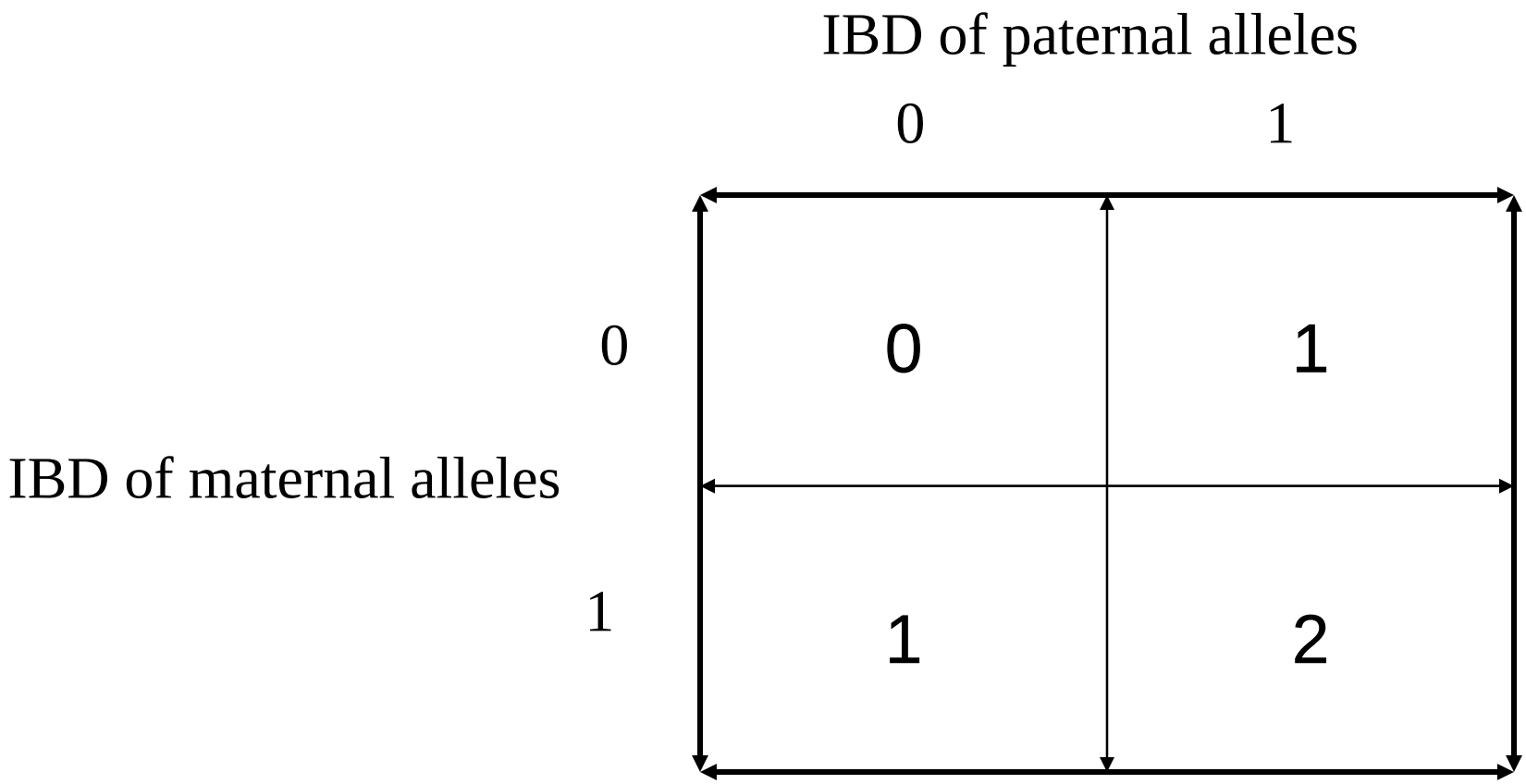

MZ twins always share 2 alleles IBD

# IBD: Half Sibs




| IBD Sharing | Probability |
|---|---|
| 0 | ½ |
| 1 | ½ |

# IBD: Full Sibs



# IBD: Full Sibs

| IBD Sharing | Probability |
|---|---|
| 0 | 1/4 |
| 1 | 1/2 |
| 2 | 1/4 |

Average IBD sharing = 1

# Genetic Relationships

Φ (kinship coefficient): Probability of IBD between two alleles drawn at random, one from each individual, at the same locus.

Δ: Probability that both alleles at the same locus are IBD

| Relationship | Φ | Δ |
|---|---|---|
| MZ twins | 0.5 | 1 |
| Parent-offspring | 0.25 | 0 |
| Full sibs | 0.25 | 0.25 |
| Half sibs | 0.125 | 0 |

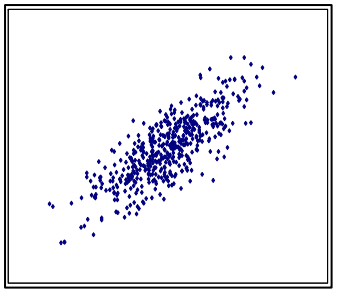
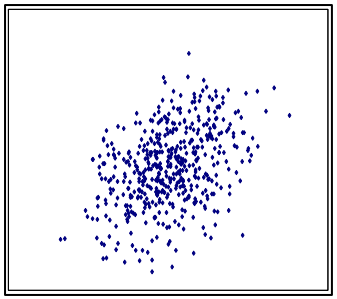
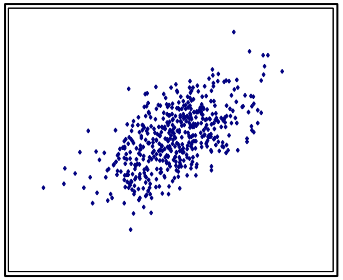
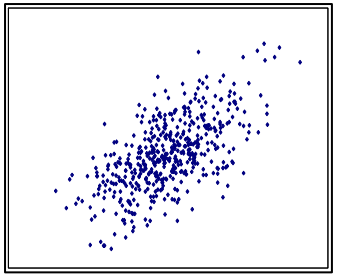
# Proportion of Alleles IBD ($\pi$)

Proportion of alleles IBD = Number of alleles IBD / 2

| Relatiobship | $\Phi$ | E($\pi$) | Var($\pi$) |
|---|---|---|---|
| MZ | 0.5 | 1 | 0 |
| Parent-Offspring | 0.25 | 0.5 | 0 |
| Full sibs | 0.25 | 0.5 | 0.125 |
| Half sibs | 0.125 | 0.25 | 0.0625 |

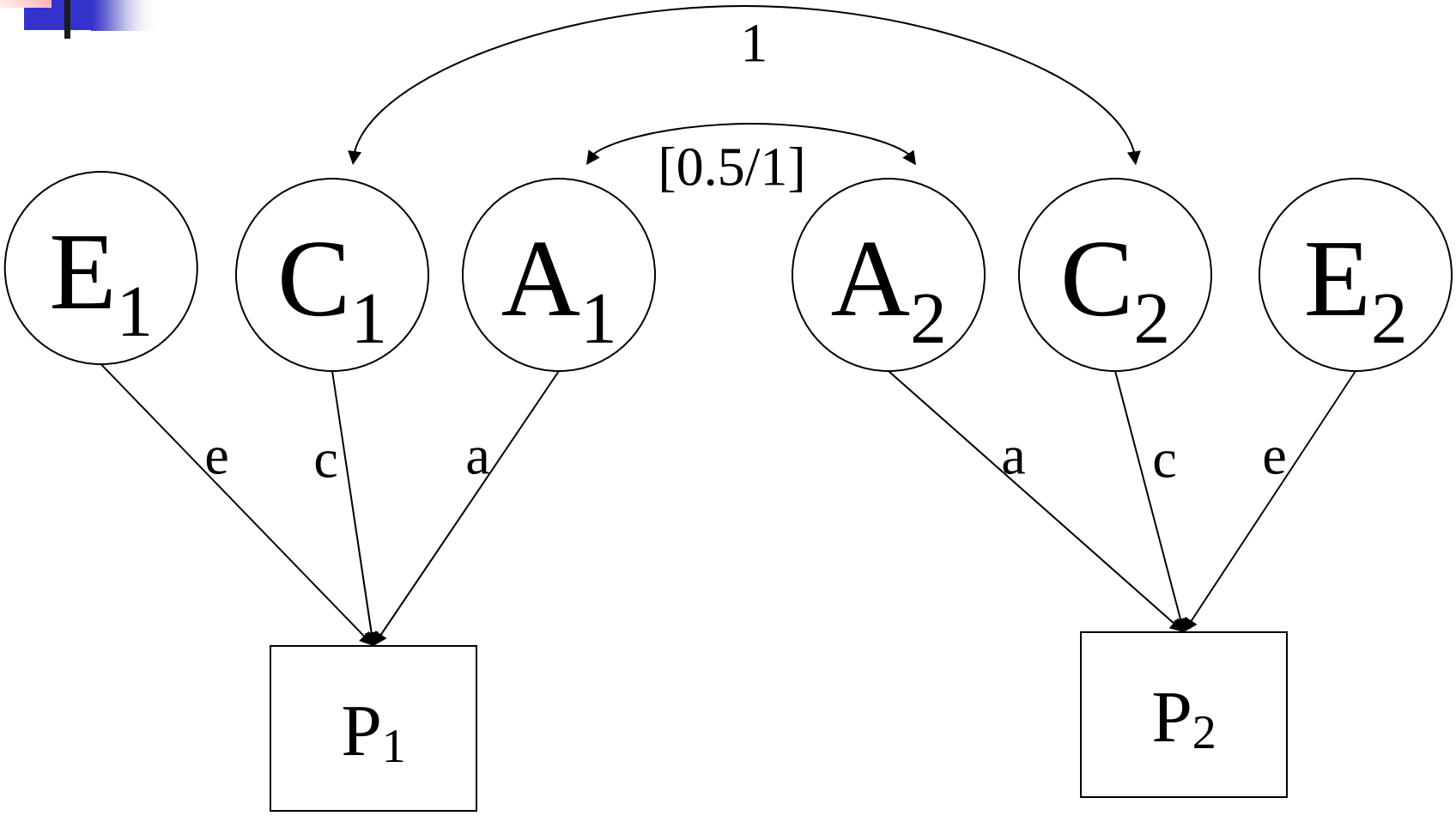
Most relationships demonstrate variation in $\pi$ across the chromosomes

# Classical Twin Analysis

| | MZ Twins | DZ Twins | |
|---|---|---|---|
| Average genetic sharing | 100% | 50% | |
| Phenotypic correlation |  |  | > ⇒ Genetic influences |
| |  |  | = ⇒ No genetic influences |

Note: Equal Environment Assumption

# ACE Model for twin data



# Implied covariance matrices

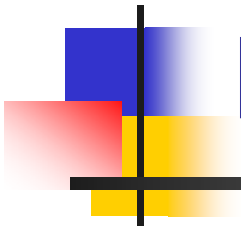
$$\Sigma_{MZ} = \begin{bmatrix} a^2 + c^2 + e^2 & \\ a^2 + c^2 & a^2 + c^2 + e^2 \end{bmatrix}$$

$$\Sigma_{DZ} = \begin{bmatrix} a^2 + c^2 + e^2 & \\ \frac{1}{2}a^2 + c^2 & a^2 + c^2 + e^2 \end{bmatrix}$$

$\Rightarrow$ Difference between MZ and DZ covariance ~ Genetic Variance / 2

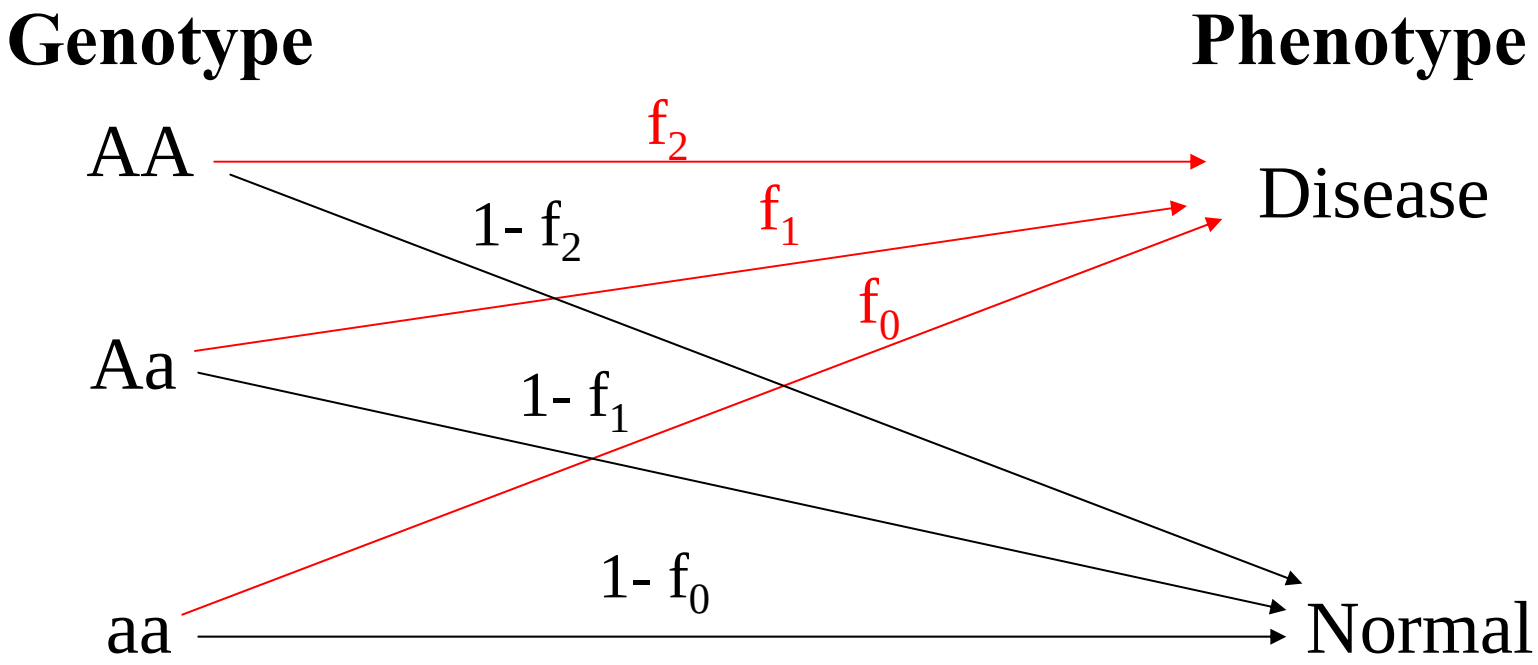
# Heritability

- Is proportion of phenotypic variance due to genetic factors
- Is population-specific
- May change with changes in the environment
- A high heritability does not preclude effective prevention or intervention
- Most human traits have heritability of 30% – 90%



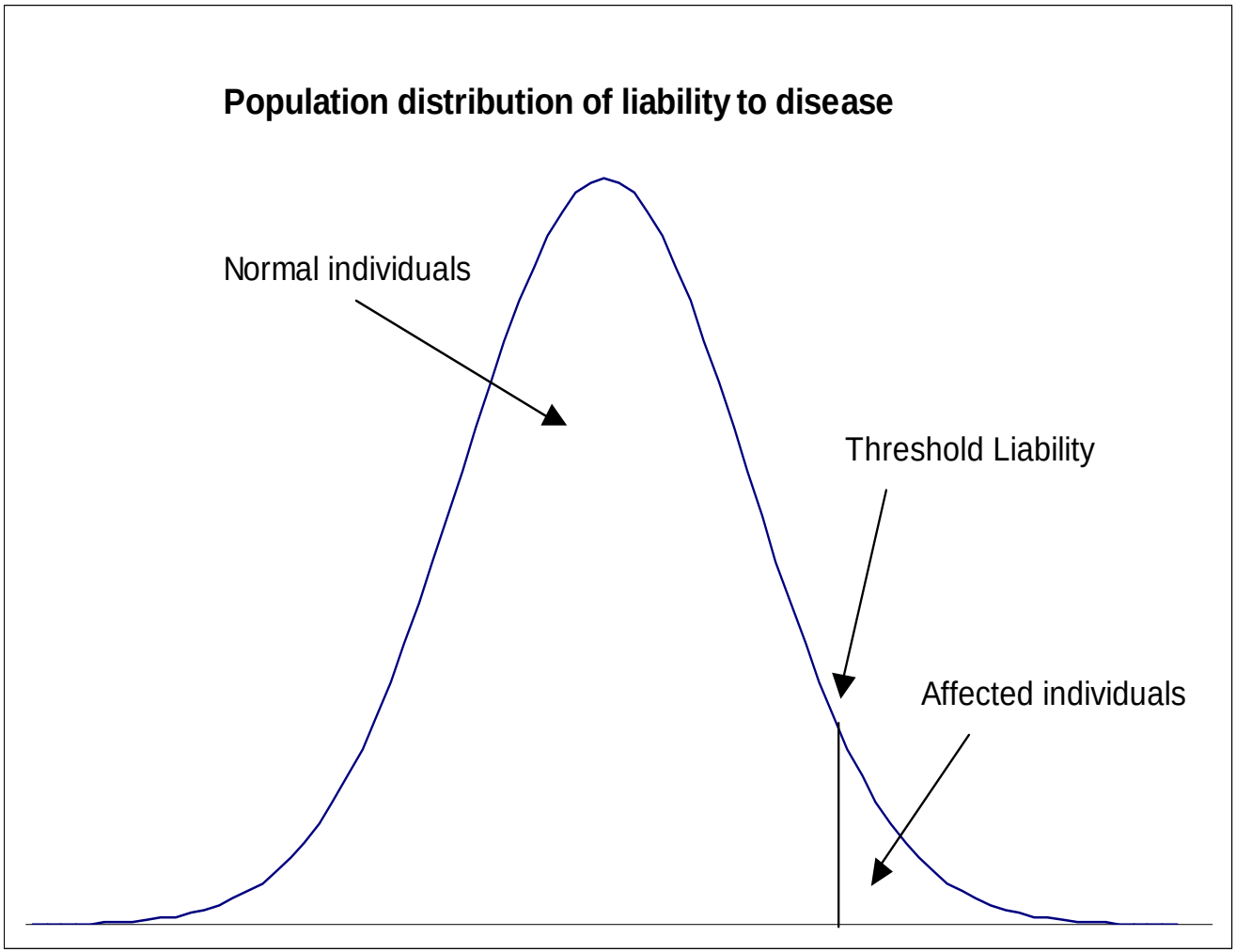
# Liability-Threshold Models

# Single Major Locus (SML) Model

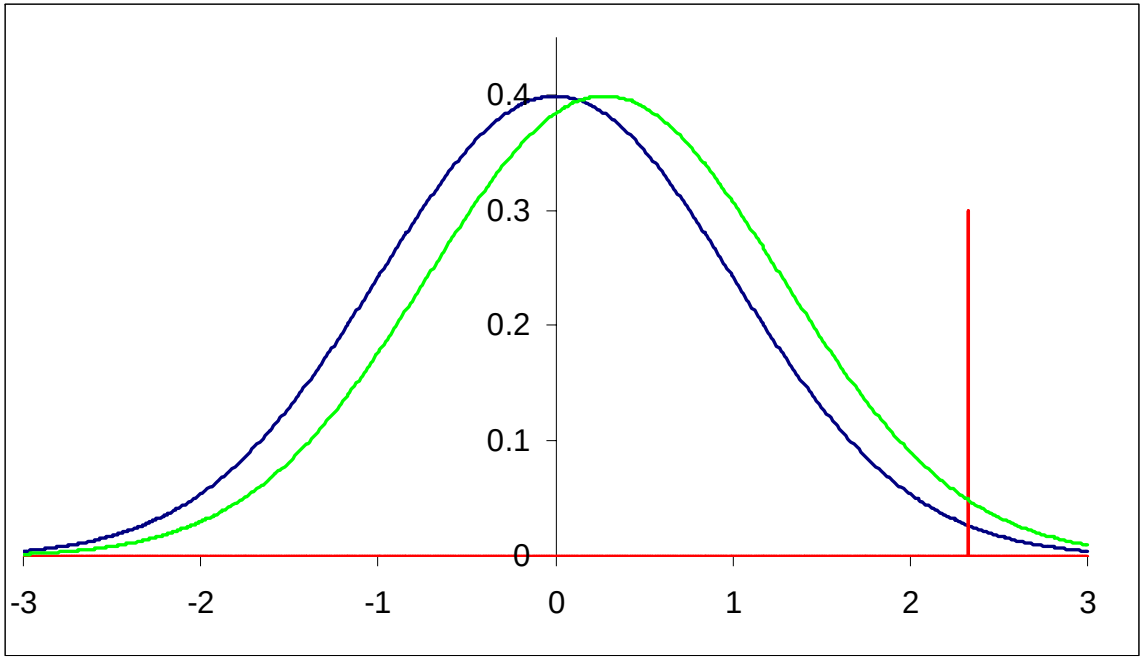



"Penetrance parameters"

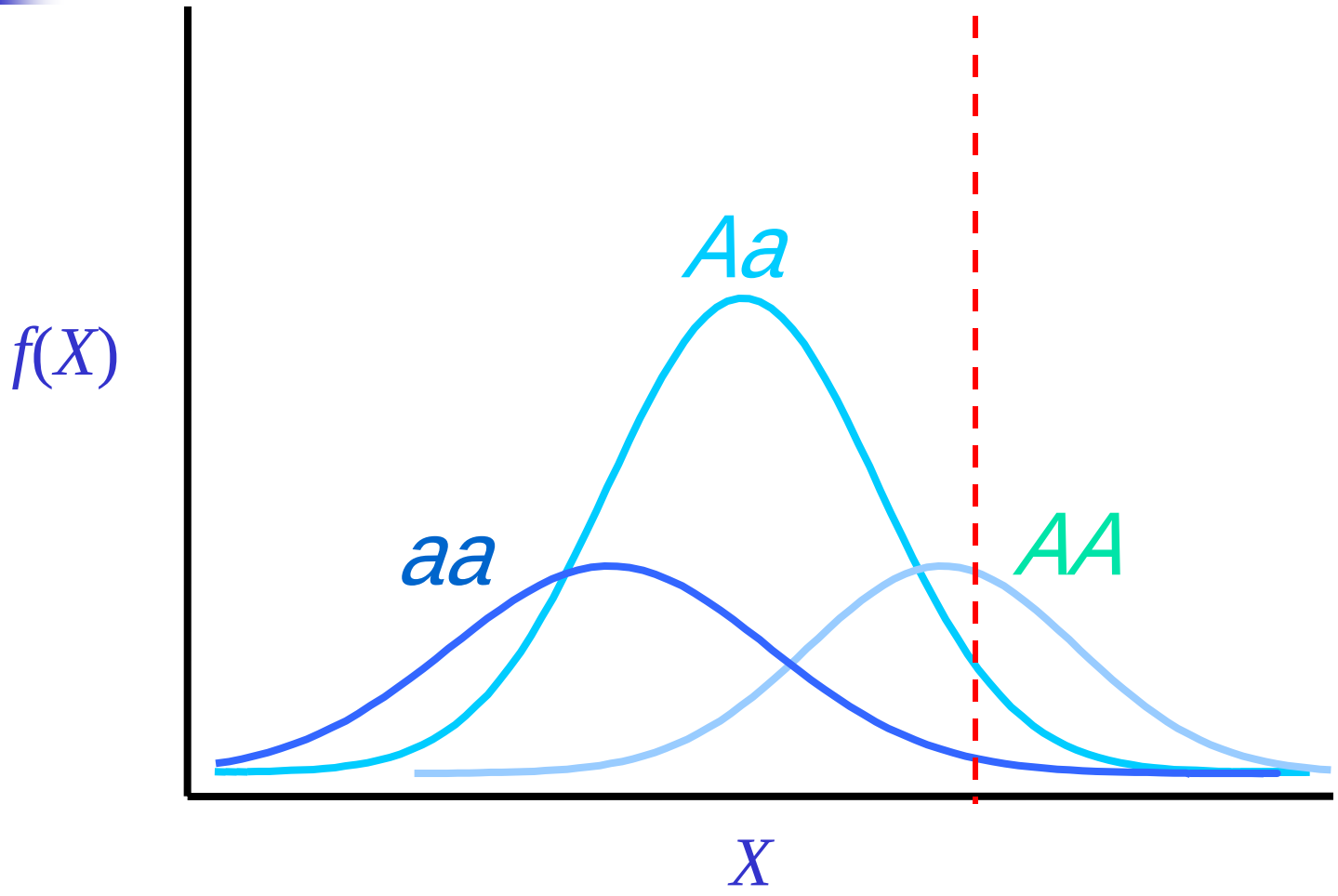
# Liability-Threshold Model

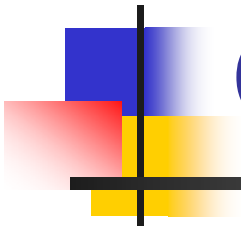


# Liability-threshold model



# Threshold Model with SML

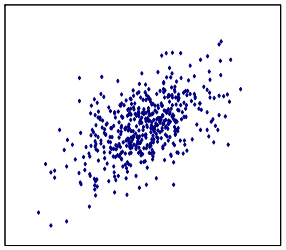
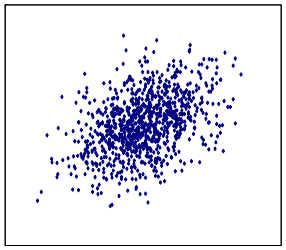
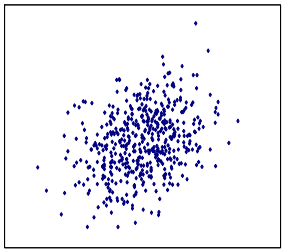
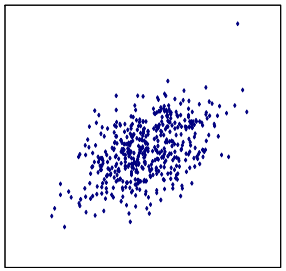
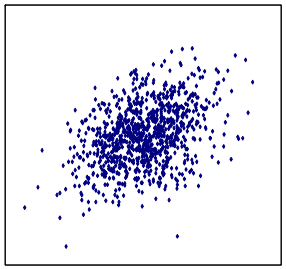
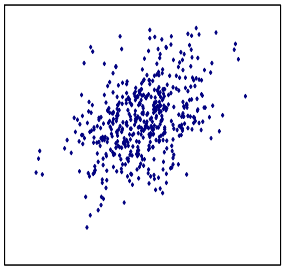




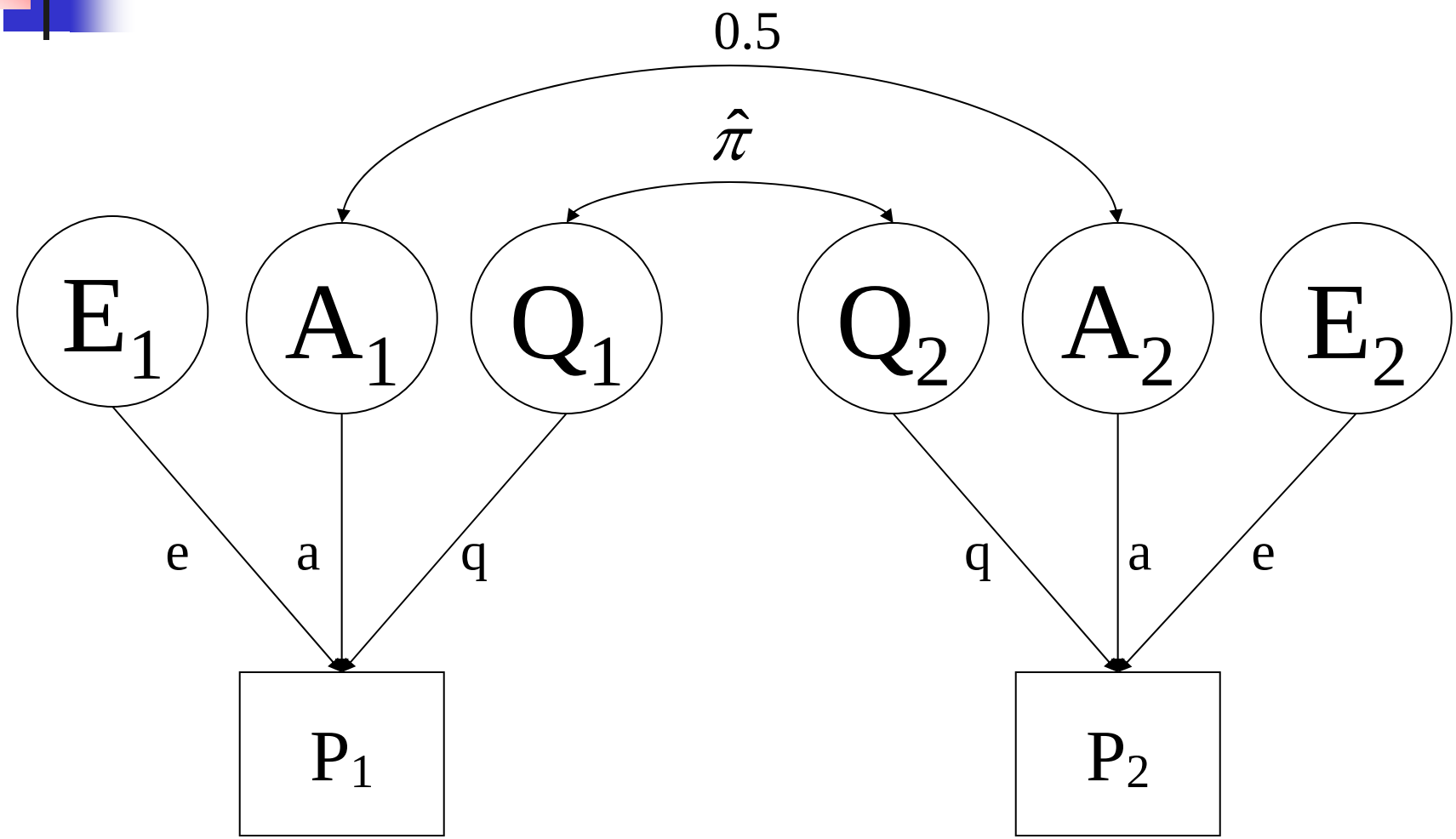
# Quantitative Trait Linkage

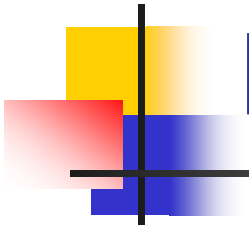
# QTL Linkage Analysis

DZ Twins / Sibling Pairs

| Local genetic sharing | 2 | 1 | 0 | |
|---|---|---|---|---|
| Phenotypic correlation |  |  |  | Linkage |
| |  |  |  | No linkage |

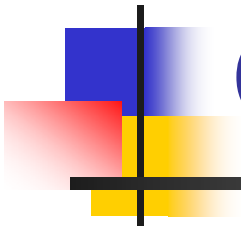
# QTL linkage model for sib pairs





# Exercise

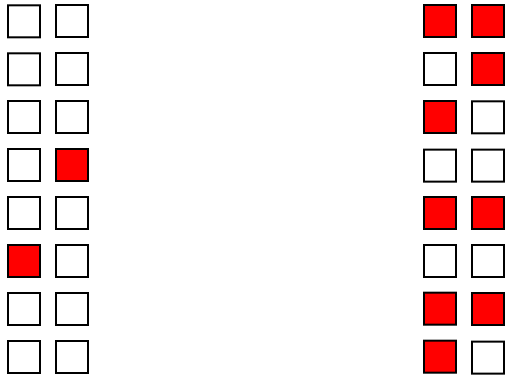
- From the path diagram write down the implied covariance matrices for sib pairs with proportion IBD sharing of 0, 0.5 and 1.



# Quantitative Association

# Allelic Association

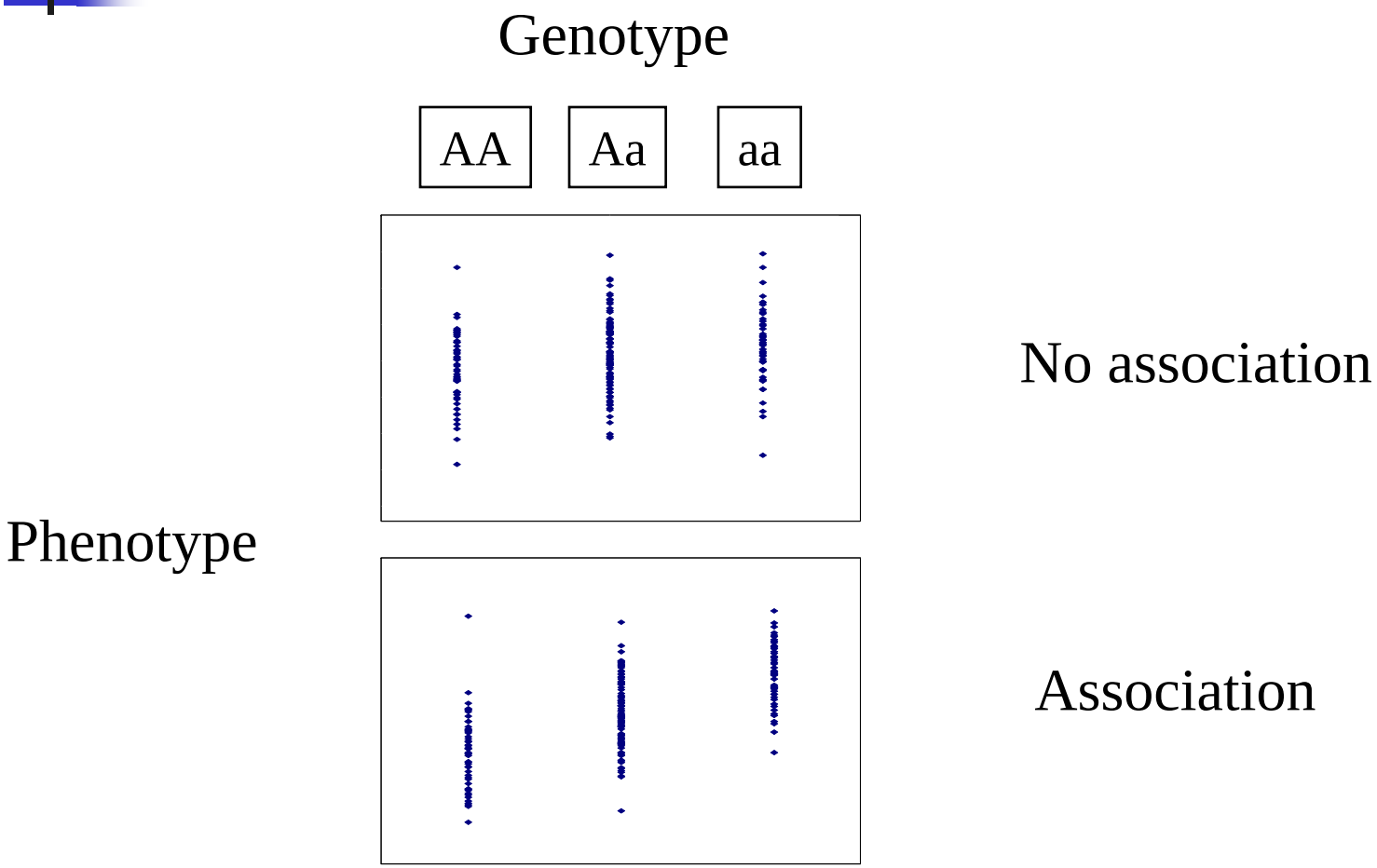
- disease susceptibility allele is more frequent in cases than in controls



Controls Cases

Example: Apolipoprotein E ε4 allele increases susceptibility to Alzheimer's disease

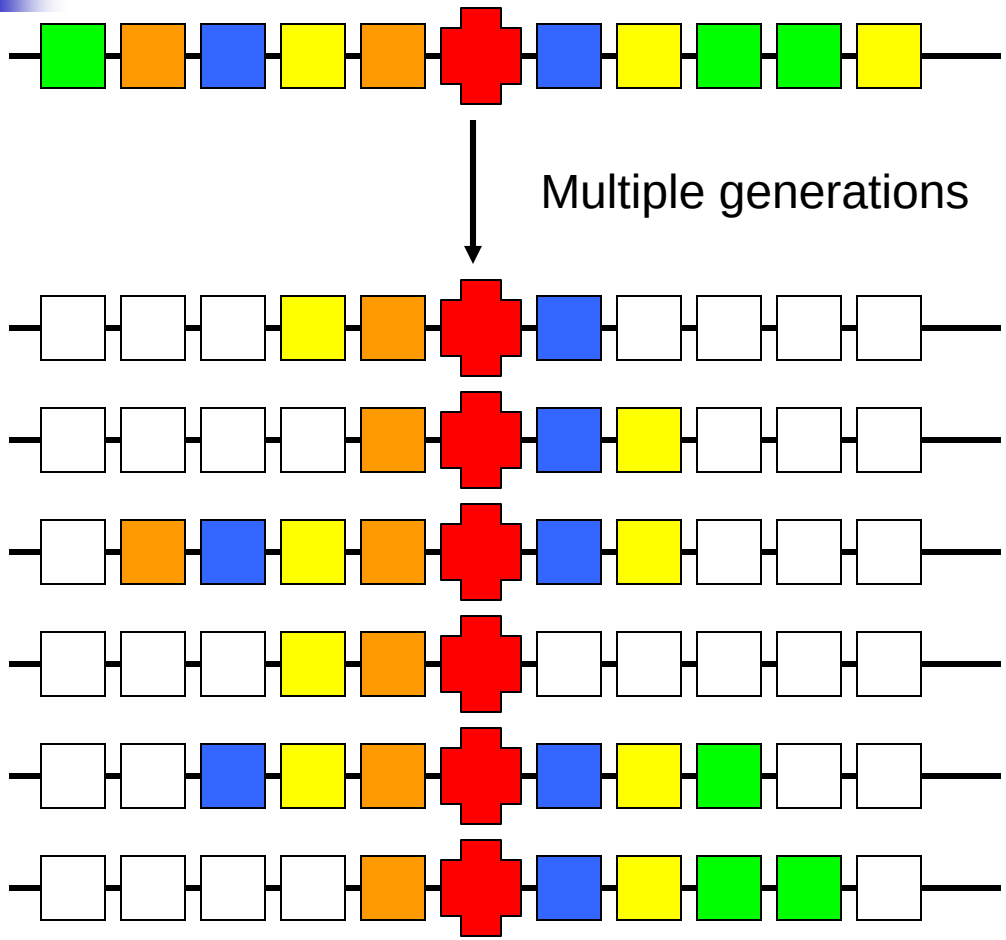
# Analysis of Means



# Causes of association

- Direct: allele increases risk of disease
- Indirect: allele associated with a risk-increasing allele through tight linkage
- “Spurious”: allele associated with disease through confounding variable (e.g. population substructure).

# Haplotype association




Mutational event on ancestral chromosome

Present mutation-bearing chromosomes with variable preserved region

# Complex Disorders: Some Historical Landmarks

- 1875: Use of twins to disentangle nature from nurture (Galton)
- 1918: Polygenic model proposed to reconcile quantitative and Mendelian genetics (Fisher)
- 1965: Liability-threshold model postulated for common congenital malformations (Carter)
- 1960's: Association between blood groups and HLA antigens with disease
- 1990's: Identification of APOE-e4 as a susceptibility allele for dementia
- 2000's: International HapMap Project