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

XY Zygote → Somatic cells → Spermatozoa

XX Zygote → Somatic cells → Ova → Zygote

Mitosis → Meiosis → Fertilization

DI PLOID → DI PLOID → HAPLOID → DI PLOID
Sources of Natural Variation

Genetic Differences

Environmental Differences

Individual Phenotypic Differences
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

Pure Lines

F1

Intercross

3:1 Segregation Ratio
Mendel’s Experiments

Back cross

F1
Aa

Pure line
aa

1:1 Segregation ratio
<table>
<thead>
<tr>
<th>Mode of inheritance</th>
<th>Mating type</th>
<th>Segregation ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autosomal dominant</td>
<td>Affected x Normal</td>
<td></td>
</tr>
<tr>
<td>Autosomal recessive</td>
<td>Carrier x Carrier</td>
<td></td>
</tr>
<tr>
<td>X-linked dominant</td>
<td>Normal father x Affected mother</td>
<td></td>
</tr>
<tr>
<td>X-linked recessive</td>
<td>Normal father x Carrier mother</td>
<td></td>
</tr>
</tbody>
</table>
## Segregation Ratios

<table>
<thead>
<tr>
<th>Mode of inheritance</th>
<th>Mating type</th>
<th>Segregation ratio Affected:Normal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autosomal dominant</td>
<td>Affected x Normal</td>
<td>1:1</td>
</tr>
<tr>
<td>Autosomal recessive</td>
<td>Carrier x Carrier</td>
<td></td>
</tr>
<tr>
<td>X-linked dominant</td>
<td>Normal father x Affected mother</td>
<td></td>
</tr>
<tr>
<td>X-linked recessive</td>
<td>Normal father x Carrier mother</td>
<td></td>
</tr>
</tbody>
</table>
# Segregation Ratios

<table>
<thead>
<tr>
<th>Mode of inheritance</th>
<th>Mating type</th>
<th>Segregation ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autosomal dominant</td>
<td>Affected x Normal</td>
<td>1:1</td>
</tr>
<tr>
<td>Autosomal recessive</td>
<td>Carrier x Carrier</td>
<td>1:3</td>
</tr>
<tr>
<td>X-linked dominant</td>
<td>Normal father x Affected mother</td>
<td></td>
</tr>
<tr>
<td>X-linked recessive</td>
<td>Normal father x Carrier mother</td>
<td></td>
</tr>
</tbody>
</table>
## Segregation Ratios

<table>
<thead>
<tr>
<th>Mode of inheritance</th>
<th>Mating type</th>
<th>Segregation ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autosomal dominant</td>
<td>Affected x Normal</td>
<td>1:1</td>
</tr>
<tr>
<td>Autosomal recessive</td>
<td>Carrier x Carrier</td>
<td>1:3</td>
</tr>
<tr>
<td>X-linked dominant</td>
<td>Normal father x Affected mother</td>
<td>1:1</td>
</tr>
<tr>
<td>X-linked recessive</td>
<td>Normal father x Carrier mother</td>
<td></td>
</tr>
<tr>
<td>Mode of inheritance</td>
<td>Mating type</td>
<td>Segregation ratio</td>
</tr>
<tr>
<td>-------------------------</td>
<td>----------------------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Autosomal dominant</td>
<td>Affected x Normal</td>
<td>1:1</td>
</tr>
<tr>
<td>Autosomal recessive</td>
<td>Carrier x Carrier</td>
<td>1:3</td>
</tr>
<tr>
<td>X-linked dominant</td>
<td>Normal father x Affected mother</td>
<td>1:1</td>
</tr>
<tr>
<td>X-linked recessive</td>
<td>Normal father x Carrier mother</td>
<td>1:1 in sons</td>
</tr>
</tbody>
</table>
Hardy-Weinberg Law
## Parental Frequencies

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>P</td>
</tr>
<tr>
<td>Aa</td>
<td>Q</td>
</tr>
<tr>
<td>aa</td>
<td>R</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Allele</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>P+Q/2</td>
</tr>
<tr>
<td>a</td>
<td>R+Q/2</td>
</tr>
</tbody>
</table>
### Mating Type Frequencies (Random Mating)

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th>Aa</th>
<th>aa</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>$P^2$</td>
<td>PQ</td>
<td>PR</td>
</tr>
<tr>
<td>Aa</td>
<td>PQ</td>
<td>$Q^2$</td>
<td>QR</td>
</tr>
<tr>
<td>aa</td>
<td>PR</td>
<td>QR</td>
<td>$R^2$</td>
</tr>
</tbody>
</table>
# Offspring Segregation Ratios

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th>Aa</th>
<th>aa</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AA</strong></td>
<td>AA</td>
<td>AA:Aa</td>
<td>Aa</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5:0.5</td>
<td></td>
</tr>
<tr>
<td><strong>Aa</strong></td>
<td>AA:Aa</td>
<td>AA:Aa:aa</td>
<td>Aa:aa</td>
</tr>
<tr>
<td></td>
<td>0.5:0.5</td>
<td>0.25:0.5:0.25</td>
<td>0.5:0.5</td>
</tr>
<tr>
<td><strong>aa</strong></td>
<td>Aa</td>
<td>Aa:aa</td>
<td>aa</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5:0.5</td>
<td></td>
</tr>
</tbody>
</table>
## Offspring Genotype Frequencies

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>$P^2 + PQ + Q^2/4 = (P+Q/2)^2$</td>
</tr>
<tr>
<td>Aa</td>
<td>$2PR + PQ + QR + Q^2/2 = 2(P+Q/2)(R+Q/2)$</td>
</tr>
<tr>
<td>aa</td>
<td>$R^2 + QR + Q^2/4 = (R+Q/2)^2$</td>
</tr>
</tbody>
</table>
### Offspring Allele Frequencies

<table>
<thead>
<tr>
<th>Allele</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>((P+Q/2)^2 + (P+Q/2)(R+Q/2) = P+Q/2)</td>
</tr>
<tr>
<td>a</td>
<td>((R+Q/2)^2 + (P+Q/2)(R+Q/2) = R+Q/2)</td>
</tr>
</tbody>
</table>
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 \( p^2:2pq:q^2 \), regardless of the genotype frequencies in the parents.
Hardy-Weinberg Equilibrium

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>a</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>p²</td>
<td>pq</td>
</tr>
<tr>
<td>a</td>
<td>pq</td>
<td>q²</td>
</tr>
<tr>
<td></td>
<td>p</td>
<td>q</td>
</tr>
</tbody>
</table>
# Hardy-Weinberg Disequilibrium

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>a</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>$p^2 + d$</td>
<td>$pq - d$</td>
</tr>
<tr>
<td>a</td>
<td>$pq - d$</td>
<td>$q^2 + d$</td>
</tr>
</tbody>
</table>

- $p$: Frequency of allele A
- $q$: Frequency of allele a
- $d$: Deviation from equilibrium
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

(a)  (b)  (c)  (d)
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

AABB → aabb

AaBb → aabb

AaBb aabb Aabb aaBb

Non-recombinant Recombinant
Haplotypes
Haplotypes

Paternal haplotype

Maternal haplotype

Genotype
Recombination

Parental haplotypes

Possible transmitted haplotypes

Non-recombinants

Single recombinants

Double recombinants
Linkage Equilibrium

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>pr</td>
<td>ps</td>
</tr>
<tr>
<td>a</td>
<td>qr</td>
<td>qs</td>
</tr>
<tr>
<td></td>
<td>r</td>
<td>s</td>
</tr>
</tbody>
</table>
**Linkage Disequilibrium (LD)**

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>pr+d</td>
<td>ps-d</td>
</tr>
<tr>
<td>a</td>
<td>qr-d</td>
<td>qs+d</td>
</tr>
<tr>
<td></td>
<td>r</td>
<td>S</td>
</tr>
</tbody>
</table>
Decay of LD

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

<table>
<thead>
<tr>
<th>Relationship to Proband</th>
<th>Lifetime Risk of Schizophrenia (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unrelated</td>
<td>1</td>
</tr>
<tr>
<td>First cousins</td>
<td>2</td>
</tr>
<tr>
<td>Uncles/Aunts</td>
<td>2</td>
</tr>
<tr>
<td>Nephews/Nieces</td>
<td>4</td>
</tr>
<tr>
<td>Grandchildren</td>
<td>5</td>
</tr>
<tr>
<td>Half siblings</td>
<td>6</td>
</tr>
<tr>
<td>Parents</td>
<td>6</td>
</tr>
<tr>
<td>Siblings</td>
<td>9</td>
</tr>
<tr>
<td>Children</td>
<td>13</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Zygosity</th>
<th>Concordance (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dizygotic (DZ)</td>
<td>17</td>
</tr>
<tr>
<td>Monozygotic (MZ)</td>
<td>48</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Adoptees of</th>
<th>Risk of Schizophrenia (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schizophrenic parents</td>
<td>8</td>
</tr>
<tr>
<td>Control parents</td>
<td>2</td>
</tr>
</tbody>
</table>

From: Finnish Adoption Study, as summarised in Psychiatric Genetics and Genomics. MuGuffin, Owen & Gottesman, 2002
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
→ 3 Genotypes
→ 3 Phenotypes

2 Genes
→ 9 Genotypes
→ 5 Phenotypes

3 Genes
→ 27 Genotypes
→ 7 Phenotypes

4 Genes
→ 81 Genotypes
→ 9 Phenotypes

Central Limit Theorem → Normal Distribution
Continuous Variation

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

$X \sim N(\mu, \sigma^2)$
Bivariate normal
Familial Covariation

Bivariate normal distribution

\[ X \sim N(\mu, \Sigma) \]

\[ \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \]

\[ \Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} \]
Correlation due to Shared Factors

Francis Galton: Two Journeys starting at same time

A

Denmark Hill → Victoria

B

Paddington

C

Brixton

Journey Times: A+B and A+C

Shared A → Covariance → Correlation
Shared Genes

Gene A is shared:

\[ = \text{Identity-By-Descent (IBD)} \]

⇒ 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.
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

<table>
<thead>
<tr>
<th>IBD Sharing</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>½</td>
</tr>
<tr>
<td>1</td>
<td>½</td>
</tr>
</tbody>
</table>
IBD: Full Sibs

IBD of paternal alleles

0 1

0 1 2

IBD of maternal alleles
### IBD: Full Sibs

<table>
<thead>
<tr>
<th>IBD Sharing</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1/4</td>
</tr>
<tr>
<td>1</td>
<td>1/2</td>
</tr>
<tr>
<td>2</td>
<td>1/4</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Relationship</th>
<th>Φ</th>
<th>Δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>MZ twins</td>
<td>0.5</td>
<td>1</td>
</tr>
<tr>
<td>Parent-offspring</td>
<td>0.25</td>
<td>0</td>
</tr>
<tr>
<td>Full sibs</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>Half sibs</td>
<td>0.125</td>
<td>0</td>
</tr>
</tbody>
</table>
Proportion of Alleles IBD ($\pi$)

Proportion of alleles IBD = Number of alleles IBD / 2

<table>
<thead>
<tr>
<th>Relationship</th>
<th>$\Phi$</th>
<th>$E(\pi)$</th>
<th>$Var(\pi)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MZ</td>
<td>0.5</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Parent-Offspring</td>
<td>0.25</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>Full sibs</td>
<td>0.25</td>
<td>0.5</td>
<td>0.125</td>
</tr>
<tr>
<td>Half sibs</td>
<td>0.125</td>
<td>0.25</td>
<td>0.0625</td>
</tr>
</tbody>
</table>

Most relationships demonstrate variation in $\pi$ across the chromosomes.
<table>
<thead>
<tr>
<th>Type of Relationship</th>
<th>Average Genetic Sharing</th>
</tr>
</thead>
<tbody>
<tr>
<td>MZ Twins</td>
<td>1</td>
</tr>
<tr>
<td>Parent - offspring</td>
<td>0.5</td>
</tr>
<tr>
<td>Full sibs (including DZ Twins)</td>
<td>0.5</td>
</tr>
<tr>
<td>Half Sibs</td>
<td>0.25</td>
</tr>
<tr>
<td>Aunt/Uncle - Nephew/Niece</td>
<td>0.25</td>
</tr>
<tr>
<td>First Cousins</td>
<td>0.125</td>
</tr>
</tbody>
</table>

If genetic factors are involved in a disease, then the closer the relationship, the greater the similarity in disease status.
Classical Twin Analysis

**MZ Twins**
- Average genetic sharing: 100%
- Phenotypic correlation: >

**DZ Twins**
- Average genetic sharing: 50%
- Phenotypic correlation: =

⇒ Genetic influences
⇒ No genetic influences

Note: Equal Environment Assumption
ACE Model for twin data
Implied covariance matrices

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

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

⇒ 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

Genotype

- AA
- Aa
- aa

 Phenotype

- Disease
- Normal

"Penetrance parameters"
Liability-Threshold Model

Population distribution of liability to disease

Normal individuals

Threshold Liability

Affected individuals
Liability-threshold model
Threshold Model with SML

\[ f(X) \]

\[ x \]

\[ AA \]

\[ Aa \]

\[ aa \]
Quantitative Trait Linkgage
QTL Linkage Analysis

DZ Twins / Sibling Pairs

Local genetic sharing

<table>
<thead>
<tr>
<th>2</th>
<th>1</th>
<th>0</th>
</tr>
</thead>
</table>

Linkage

Phenotypic correlation

No linkage
QTL linkage model for sib pairs

\[ \hat{\pi} \]

\[ 0.5 \]
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

Example: Apolipoprotein E ε4 allele increases susceptibility to Alzheimer’s disease
Analysis of Means

Genotype

- AA
- Aa
- aa

Phenotype

- No association
- Association
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

Multiple generations

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