Basic Concepts of Human Genetics

• The genetic information of an individual is contained in 23 pairs of chromosomes.
  Every human cell contains the 23 pair of chromosomes.

• One pair is called sex chromosomes
  Male: XY
  Female: XX

• Other 22 pairs of homologous chromosomes are called autosomes.

• The autosome chromosome pairs are called homologous pair. Two chromosomes in the
  same pair are called homologous chromosomes.
Figure 1.1. Human karyotype, showing a standard numbering system based on chemical staining of a chromosome preparation. Different staining methods yield different levels of detail that are used in refining the system (not shown). (From Thompson and Thompson, 1986.)
• One member of each chromosome pair is from mother; the other is from father. Father or mother transmits each of the two chromosomes with equal probability.

• The location of large scale at chromosome usually use the symbols like 10q5.6 (means at the long arm of chromosome 10, band 5.6) or 5p8.7 etc.
There are two DNA chains in one chromosome

- DNA has four bases A, G, T and C. A combined with T and G combined with C

- bp is also used as length unit of chromosome or DNA sequence
- DNA sequence has direction. There are two sides (ends) called 5’ side and 3’ side.
- The homologous chromosomes (chromosome 1, for example) have exactly same length for every individual.
Normal cell division—Mitosis

Old cells died and new cells grow. The new cells grow through normal cell division.

● The normal cell division—Mitosis process:

(1). The double DNA strands in each of the chromosomes split into two single-strands.

(2). DNA replication. After step (2), each chromosome produces another identical one.
• All the 23 pairs of the chromosomes undergo this process of replication, producing two identical sets of 23 chromosome pairs. The two sets of chromosomes are separated and distributed into two daughter cells.
23 pairs of chromosomes

replication

23 pairs of chromosomes

23 pairs of chromosomes

23 pairs of chromosomes

split to two daughter cells
The inheritance of chromosomes—Meiosis

1. The 23 pairs of chromosomes in the cell are duplicated every time a cell division occurs.

2. The only exceptions to this rule are gametes (ovum and sperm), which are produced by sex organ.

3. Gametes are produced by a special cell division called **Meiosis**.

4. **Meiosis** gives rise to daughter cells (ovum or sperm) which contain only a haploid (single chromosome, not pair) set of 22 autosomes and a sex chromosome.
The procedure of inheritance--Meiosis:

Normal cell of Male

- 23 pairs chr

Meiosis

- 23 chromosomes (sperm)

Mitosis

- 23 pairs chr.
  (zygote)

Normal cell of Female

- 23 pairs chr

Meiosis

- 23 chromosomes (ovum)

Mitosis

- All cells in human body
Genetic Terminology

• Some genetic concepts are potentially confusing, such as gene. The reason is that some concepts were introduced prior to the discovery of DNA.

—Gene
  ♦ A segment of DNA within a chromosome which has a specific genetic function.
  ♦ Length from several bps to several kbs.
  ♦ Gene is not a smallest unit of genetic material
  ♦ Before the discovery of DNA, people believe that gene is the smallest genetic unit.
**Marker (locus):** A specific position in chromosome. It may be 1 *bp* or several hundred *bps* in length.

---

**Individual 1**

![Diagram of Individual 1 with alleles AGCT and marker 7D143]

**Allele:** "AGCT"

**Individual 2**

![Diagram of Individual 2 with alleles GAGC and marker 7D143]

**Allele:** "GAGC"

---

**Alleles:** DNA sequences within a marker or locus (*gene* and *allele* sometimes have same meaning)
• Now, there are mainly two kinds of makers used today.

(1) SNPs (Single nucleotide polymorphism): 1 bp in length; there are two alleles in human population. The two alleles may be any two from \{A,G,C,T\}. Most of the SNPs were found from DNA sequences of a few individuals (Human Genome project). Here is an example of ten sequences from 5 individuals. The positions of second, 8th, 12th columns are SNPs.

GTCGAATTGGAATTGG
GTCGAATTGGAATTGG
GTCGAATTGGAATTGG
GCCGAATAGGAATTGG
GTCGAATTGGAATTGG
GTCGAATAGGAATTCG
GCCGAATTTGATTTGG GTCGAATTGGAATTGG GTCGAATTGGAATTGG

(2) Macrosatellite markers: Length from several bps to several hundred bps; many possible alleles; usually denoted by 1, 2, 3, …. 
Human Genome

- The totality of DNA characteristic of all the 23 pairs of chromosomes.
  - The human genome has about \(3 \times 10^9\) bps in length.
  - 97% of the human genome is non-coding regions called introns. 3% is responsible for controlling the human genetic behavior. The coding region is called exon.
  - There are totally about 40,000 genes, over 5000 have been identified. There are much more left
  - Human Genome Project is to identified the DNA sequence (every bp) of human genome (only a few individuals)
  - For human being, most of the place in human genome are the same. Only a very small part is different among different individuals.
Gene to protein

Genes or DNA sequences themselves are not control the phenotypes. Genes or DNA sequences control the phenotypes through protein.

Protein: like the DNA molecule that is a chain of base pair, each protein molecule is a linear chain of subunits called **amino acids**.

A protein Molecule
Amino acids have 20 different forms which usually denoted by (The first three letters of different amino acids except in the case of Asn (asparagines); Gln (glutamine); Ile (isoleucine); and Trp (tryptophane)

<table>
<thead>
<tr>
<th>Ala</th>
<th>Arg</th>
<th>Asn</th>
<th>Asp</th>
<th>Cys</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gln</td>
<td>Glu</td>
<td>Gly</td>
<td>His</td>
<td>Leu</td>
</tr>
<tr>
<td>Ile</td>
<td>Lys</td>
<td>Met</td>
<td>Phe</td>
<td>Pro</td>
</tr>
<tr>
<td>Ser</td>
<td>Stop</td>
<td>Thr</td>
<td>Trp</td>
<td>Val</td>
</tr>
</tbody>
</table>

The physical and chemical properties of a protein molecule are largely determined by its sequence of the amino acids and its shape structure.
The DNA sequence specifies *amino acids* sequence (protein), and therefore the structure and function of protein.

The decoding of the information in the DNA into proteins involve two steps called transcription and translation.

**Transcription:**
- A single strand of DNA synthesis a single-stranded RNA
- RNA: similar to DNA but have letter U instead of T. So, RNA sequence contains four letters A, U, G and C.
- Before translation, RNA transcripts are processed by the deletion of certain non-coding sequence. The processed RNA chain is called messenger RNA (mRNA).
**Translation:**

• This step translates mRNA to protein.

• There is a direct relationship between the base sequence of mRNA and the amino acid sequence of its protein product. This relation is called the genetic code.

**Coding rule**

• Each “codeword” (or codon) is a triplet of nucleotides (each word is consist of three letters in length).

• So, there are $4^3 = 64$ possible “code words”. Since there are only 20 different amino acids, the genetic coding is degenerate, that is, some mRNA code words translate to the sample amino acid. The relation is given below:
### Table 1. Genetic coding

<table>
<thead>
<tr>
<th>First position 5’ end</th>
<th>Second position</th>
<th>Third Position 3’ end</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>U</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>Ser</td>
<td>Trp</td>
</tr>
<tr>
<td>Phe</td>
<td>Ser</td>
<td>Trp</td>
</tr>
<tr>
<td>Leu</td>
<td>Ser</td>
<td>Stop</td>
</tr>
<tr>
<td>Leu</td>
<td>Ser</td>
<td>Stop</td>
</tr>
<tr>
<td><strong>C</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Pro</td>
<td>His</td>
</tr>
<tr>
<td>Leu</td>
<td>Pro</td>
<td>His</td>
</tr>
<tr>
<td>Leu</td>
<td>Pro</td>
<td>Gln</td>
</tr>
<tr>
<td>Leu</td>
<td>Pro</td>
<td>Gln</td>
</tr>
<tr>
<td><strong>A</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile</td>
<td>Thr</td>
<td>Asn</td>
</tr>
<tr>
<td>Ile</td>
<td>Thr</td>
<td>Asn</td>
</tr>
<tr>
<td>Ile</td>
<td>Thr</td>
<td>Lys</td>
</tr>
<tr>
<td>Met</td>
<td>Thr</td>
<td>Lys</td>
</tr>
<tr>
<td><strong>G</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td>Ala</td>
<td>Asp</td>
</tr>
<tr>
<td>Val</td>
<td>Ala</td>
<td>Asp</td>
</tr>
<tr>
<td>Val</td>
<td>Ala</td>
<td>Glu</td>
</tr>
<tr>
<td>Val</td>
<td>Ala</td>
<td>Glu</td>
</tr>
</tbody>
</table>
Genotype, phenotype and haplotype

- **Genotype**: At a specific locus there is an allele in each of the two homologous chromosomes. The two alleles together are called genotype.

- **Phenotype**: Observable, such as height, color of eye, etc.
**Example:** Blood type (ABO locus, three allele A, B and O)

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>A/A, A/O</td>
</tr>
<tr>
<td>B</td>
<td>B/B, B/O</td>
</tr>
<tr>
<td>AB</td>
<td>A/B</td>
</tr>
<tr>
<td>O</td>
<td>O/O</td>
</tr>
</tbody>
</table>

Here, A and B both mask the presence of the O allele.
— A and B are said to be **dominant** to O;
— O is **recessive** to A and B.
— A and B are **codominant**.

Now, almost of all the markers with codominant alleles. In this case, we can say that Genotype is observable.
**Haplotype:** Sequence of alleles along a chromosome

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>B</td>
<td>c</td>
</tr>
</tbody>
</table>

Chromosome pair

Locus 1  locus 2  locus 3

Genotype  A/a    B/B    C/c

The two haplotypes are  ABC and aBc.

- In practice (for codominant alleles), we can only observe multilocus genotype \{A/a  B/B  C/c\}. So the possible haplotype pairs are \{ ABC, aBc\} and \{ABc, aBC\}
### Typical data set as follow:

<table>
<thead>
<tr>
<th>Ind</th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0/0</td>
<td>0/1</td>
<td>4/5</td>
</tr>
<tr>
<td>2</td>
<td>1/0</td>
<td>0/2</td>
<td>8/7</td>
</tr>
<tr>
<td>3</td>
<td>1/1</td>
<td>2/1</td>
<td>9/6</td>
</tr>
<tr>
<td>4</td>
<td>0/1</td>
<td>3/2</td>
<td>8/8</td>
</tr>
<tr>
<td>5</td>
<td>0/0</td>
<td>0/3</td>
<td>6/5</td>
</tr>
<tr>
<td>6</td>
<td>1/0</td>
<td>2/2</td>
<td>7/7</td>
</tr>
</tbody>
</table>

- **2 alleles**
- **4 alleles**
- **6 alleles**

For individual 6, two haplotypes are \{1,2; 0,2,7\}.

For individual 1, we do not know the two haplotypes. Two possibilities are \{0,0,4; 0,1,5\} or \{0,1,4; 0,0,5\}

---

**Individual 1**

- **homozygote**
- **heterozygote**
Example: We sample $n$ individuals from a population. Each individual has genotypes at $m$ bi-allelic markers (a marker with two possible alleles). For $n=m=3$, the genotypes are as following

1. A/a B/b C/c
2. A/A B/B C/c
3. A/A B/b C/c

Questions:

(1) Is there any formula for the number of possible haplotype pairs.
(2) How to estimate the haplotype frequencies in the population.
Pedigree (ABO locus)

- Individual 7 is not a biological child of the parents 3 and 4 (Mendelian in consistency)
- Individuals 1, 2 and 4 are called found whose parents are not in the pedigree.

- Individual 7 is not a biological child of the parents 3 and 4 (Mendelian in consistency)
Pedigree (ABO locus and another marker)

Is 5 the biological child of 3 and 4? Yes!

Why 5’s paternal haplotype OD is different from her father’s two haplotypes AD and Od.

This is called recombination.
• The probability of recombination between two markers is called **recombination rate** between these two markers.

• The recombination rate increases with the physical distance becoming larger.

• The genetic distance between two markers means the **recombination rate**
Relationship between Physical distance and Genetic distance

- **Physical distance:** *unit* bp, kb

- **Genetic distance:** Using recombination rate (cM—centi-Morgan, Morgan)
  \[ 1 \text{ cM} = 1\% \text{ of recombination rate.} \]

- **Relationship** (When physical distance is small, that is, < 1000 kb)
  \[ 1 \text{ cM} \approx 1000 \text{ kb} \]
**Linkage**

Let \( \theta \) denote the recombination rate between two markers \( M \) and \( m \). If \( \theta < 1/2 \), marker \( M \) and \( m \) are said to have **linkage**, otherwise (\( \theta < 1/2 \)) the two markers are said to be in **linkage equilibrium**. (recombination rate either \( <1/2 \) or \( =1/2 \))

**Example:**

If \( \theta = 0 \), then \( ? = d \),

If \( \theta < 1/2 \), then the

\( ? = d \) with probability \( >1/2 \).

If \( \theta = 1/2 \), \( ? = d \) and \( ? = D \) with

Equal probability.
Reference


   (Chapter 1)


   (Chapter 1)


29
Homework
1. Suppose that segments of DNA sequences (one chain of the DNA) of Chromosome 1 for 10 individuals are as follow:
   ATCGAATTGGAATTGG
   GTCGAATTGGAATTGG
   ATCGAATTGGAATTGG
   ACCGAATTGGAATTGG
   GTCGAATTGGAATTGG
   ATCGAATTGGAATTGG
   ACCGAATTGGAATTGG
   GTCGAATTGGAATTGG
   ATCGAATTGGAATTGG
   ACCGAATTGGAATTGG
   Which positions are SNPs?
2. What are the major differences between mitosis and meiosis?
3. Given a definition of Gene, Marker (locus), allele, recombination rate, genotype, haplotype, and phenotype. If possible, give a example.