



# Overview of Basic Genetic Concepts and Terminology

# Advancements in Human Genetics



- Some of the objectives for genetic studies include:
  - ▶ Identify the genetic causes of phenotypic variation
  - ▶ Have better understanding of human evolution
  - ▶ Drug development: finding genes responsible for a disease provides valuable insight into how pathways could be targeted
- Recent decades have produced major advances in the science of genetics
- In the past few years we have seen the release of the first drafts of the entire human genome and the genomes of model organisms.
- The amount of data available for use in genetic studies has increased astronomically
  - ▶ Illumina array with more than 2.5 million SNPs
  - ▶ Next-generation Sequencing: 40 million or more SNPs and SNVs

# Challenges of Human Genetics



- The most notable experiments have unequivocal interpretation:
  - ▶ Unequivocal interpretation is rare in human genetics
  - ▶ Generally can not design the perfect experiment: have to work with data we have at our disposal
  - ▶ Interpretation is of the greatest importance
- How do our data and results inform us with respect to the fundamental questions we are trying to address?
- What are the alternative interpretations of our data?
- Is it possible to distinguish among these alternatives?
- With so much data and so many options, there is a pressing need for well-designed studies and accurate and efficient statistical methods.

# The Need for Experimental Design and Statistics

- Relative to experimental methods, analysis is fast and inexpensive
- Considering the cost of collecting family information and conducting molecular genetic experiments, we are obligated to get everything we can out of all of the data that we have at our disposal.
- Our goal for the quarter will be
  - ▶ study potential designs that incorporate genetic data
  - ▶ learn the corresponding methods for analyzing data from these designs
- Our goals in these tasks will be to:
  - ▶ understand the basic idea of each type of study
  - ▶ know the assumptions each type of analysis depends on for validity
  - ▶ understand the limitations of different types of studies
  - ▶ learn how to correctly interpret study results

# The basic structure of a gene

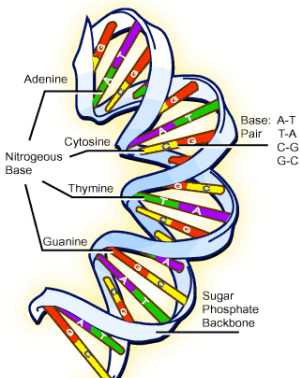


- It is well established that human characteristics are inherited from parents to offspring in discrete units called genes.
- Vast amount of info regarding the precise molecular mechanisms of genetic transmission from parent to offspring.
- A **gene** is the most fundamental unit of heredity that controls the transmission and expression of one or more traits.
- The chemical structure of a gene is deoxyribonucleic acid (DNA).

# The basic structure of a gene



- A gene can be viewed as a two long strands of DNA which are normally bound to each other lengthwise by hydrogen bonds and are twisted around each other as a double helix.
- The subunits are called nucleotides which contain the nitrogenous bases
- There are four different nitrogenous bases, called adenine (A), guanine (G), cytosine (C), and thymine (T).

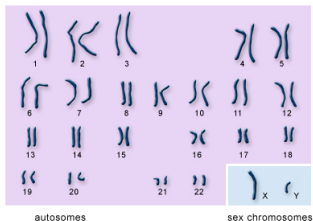


# Chromosomes



- We can think of DNA as a sequence of the four letters A, G, C, and T.
- An important feature of DNA is that A is always paired with T, and G is always paired with C.
- Genes are found on chromosomes in the nucleus of cells.
- **Chromosomes** are very long strands of DNA.
- Every species has its own characteristic number of different chromosomes.
- Humans have 23 pairs of chromosomes, 22 autosomes and 1 pair of sex chromosomes.

# Chromosomes



U.S. National Library of Medicine

- The 22 autosomal chromosomes are numbered in order of decreasing length from 1 to 22.
- For every pair of chromosomes, one is inherited from the mother of an individual and one is inherited from the father of an individual.



# Chromosomes



- Chromosomes that are of the same pair and carry the same set of genes and are called **homologous**. (e.g. both chromosome 21)
- **Mitosis** is cell division that yields two identical diploid cells, both of which have two pairs of each chromosome.
- **Meiosis** is a special type of cell division that happens in reproductive tissue yielding haploid cells (which have one of each chromosome) called **gametes**. In females, the gametes are the egg cells and in males the gametes are the sperm cells.

# Chromosomes

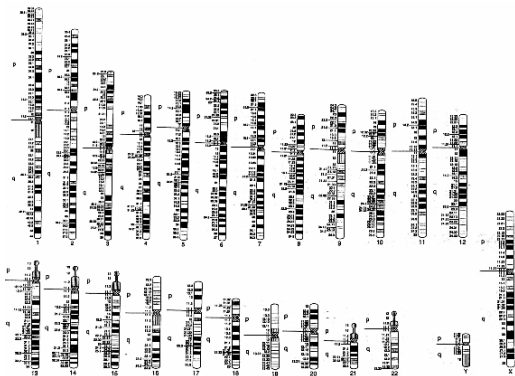


- The **centromere** is a region of the chromosome that is the attachment site for the spindle fiber that moves the chromosome during cell division.
- The centromere defines two arms of the chromosome, the short arm **p** and the long arm **q**.
- When treated with special stains, each arm appears to be divided into a number of bands, which are numbered from the centromere.

# Chromosomes



- The approximate location of a gene is often specified by the chromosome number (i.e. 1,2,...,22,X,Y), the arm (p,q), and the band (1,2,3,...).
- Example: 12q24.2



# More Genetic Terminology



- More than 99 percent of loci of the DNA sequences on the 23 chromosome pairs are identical in all humans
- A **genetic marker** is a strand of DNA that is polymorphic: has some variation in the human population.
- A genetic marker can have two or more different states and we define an **allele** to be the state at a marker.
- Single Nucleotide Polymorphism (SNP) has two allelic types: highly abundant (1 per 1000 base-pairs)
- Short Tandem Repeats (microsatellites) GTAGTAGTAGTAGTA...

# More Genetic Terminology



- For a chromosome pair, the two alleles at a single genetic marker are called an individual's **genotype**.
- **Homozygous** genotypes have alleles that are identical.
- Genotypes that have two different alleles are said to be **Heterozygous** .
- A **haplotype** is a sequence of alleles along a chromosome.

# The Human Genome

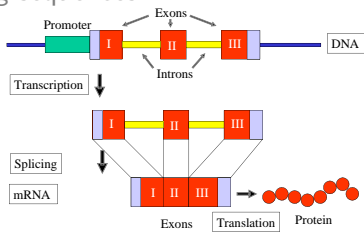


- The entire DNA characteristics of a species is called its genome.
- The human genome has about 3 billion base pairs per haploid.
- Approximately 2% of the human genome is coding and 98% of the human genome is non-coding.
- A gene is a sequence of DNA that is transcribed into mRNA (messenger RNA), which, in turn, is translated into protein.
- For RNA, uracil (U) is substituted for thymine in DNA.
- In a recent build of the human genome, annotation data are available for approximately 32,000 genes with around 18,000 confirmed genes.
- Genes vary enormously in length from less than a thousand base pairs to over a million base pairs (megabases).

# Coding Sequences



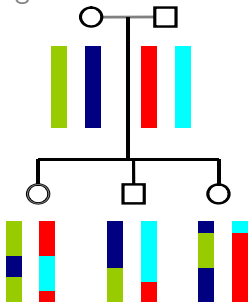
- Genes do not form a continuous sequence but consists of several coding segments called **exons** that are separated by non-coding segments called **introns**.
- Non-coding regions and introns are sometimes called "junk" DNA.
- This term can be misleading because non-coding regions may indeed have a function.
- Some non-coding regions are known to be involved in the regulation of nearby coding sequences.



# Recombination



- A chromosome inherited by an offspring from a parent is actually a mosaic of the parent's two chromosomes.
- **Genetic Recombination** → genetic material is exchanged between a chromosome of paternal origin and the corresponding chromosome of maternal origin.





# Genetic and Physical Maps



- **Crossovers** are the points of exchange
- **Recombination fraction** between two loci on a chromosome is the probability that the two loci end up on regions of different origin → occurs when the two loci are separated by an odd number of crossovers
- **Genetic Maps** → give the order and distances (recombination fraction) between genes or genetic markers.
- **Physical Maps** → sets of ordered markers and the physical distance (base pairs) between them

# Short Questions



- It is known that about 22 percent of the double-stranded DNA of an organism consists of thymine. Can the other base percentages be determined? If so, what are they?
- A certain DNA virus has a base ratio of  $(A+G)/(C+T)=0.85$ . Is this single- or double-stranded DNA? Explain.