

# Chapter 4

---

## Key Concepts:

- **Number of genes:** defined at four levels, which correspond to successive stages in gene expression.
  - **genome:** the complete set of DNA sequence of an organism
    - hard to identify every gene unequivocally solely on the basis of sequence.
  - **transcriptome** (转录组): the complete set of genes expressed under particular conditions
    - the transcriptome is likely to be larger than the actual number of genes in the genome
    - including noncoding RNAs such as tRNAs, rRNAs, microRNAs (miRNAs), as well as mRNAs
  - **proteome** (蛋白质组): the complete set of polypeptides encoded by the whole genome
    - correspond to the mRNAs in the transcriptome
  - **interactome** (相互作用组): protein-protein interactions
    - the total number of independent complexes of proteins can be identified by all protein-protein interactions
- **genetic polymorphism** (遗传多态性): the coexistence of multiple alleles at a locus in a population
  - Example: Human eye color is a example of phenotypic polymorphism (表型多态性) resulting from underlying genetic polymorphism.
  - **single nucleotide polymorphism** (SNP, 单核苷酸多态性): difference between alleles of the same locus in a single nucleotide
    - On average, one SNP occurs for approximately every 1,330 bases in the human genome
    - Defined by SNPs, every human being is unique.
    - SNPs can be detected by direct comparisons of sequences from different individuals.
    - SNPs may be neutral or linked to diseases or disease susceptibilities.
- **Genome-wide association study** (GWAS, 全基因组关联分析): entire genomes of both patients and nonpatients are scanned for SNPs.
  - SNPs that are associated with the disorder (疾病) are identified
  - The disorder may be determined by a one or more genes
- **Reassociation kinetics** (复性动力学): identifies two general types of genomic sequences
  - Nonrepetitive DNA: there is only one copy in a haploid (单倍体) genome
    - mRNA is transcribed from nonrepetitive DNA
  - Repetitive DNA: more than one copy in each haploid genome
    - Moderately repetitive DNA: relatively short sequences, repeated typically 10~1,000 times. Example: genes for tRNAs and rRNAs

- Highly repetitive DNA: very short sequences (typically fewer than 100 bp), repeated many thousands of times, often organized as long regions of tandem repeats (串联重复)
- **transposons** (转座子): short sequences of DNA (up to about 5 kb) that have the ability to move to new locations in the genome and/or to make additional copies of themselves
  - selfish DNA: Transposons propagate themselves within a genome without contributing to the development and functioning of the organism.
  - junk DNA: genomic sequences without any apparent function