

Chapter 1

Key Concepts:

- **Genome** (基因组): complete hereditary information contained in chromosomes
 - **chromosomes** (染色体): a sequence of **DNA** (deoxyribonucleic acid, 脱氧核糖核酸)
 - **gene**: a sequence of DNA in chromosomes that encodes a single type of RNA
 - each chromosome may contain a large number of genes
 - genome size:
 - mycoplasma (支原体细菌): 500 genes
 - human: 20, 000 genes
 - rice: 50, 000 genes
 - **RNA** (ribonucleic acid, 核糖核酸): another nucleic acid
 - message RNA (mRNA, 信使RNA): translated to polypeptides (多肽)
 - ribosomal RNA (rRNA, 核糖体RNA) and transfer RNA (tRNA, 转运RNA) are functional themselves
 - **Nucleic acid** (核酸, both DNA and RNA) is a polynucleotide (多核苷酸) chain, which is a long chain of nucleotides (核苷酸)
 - each nucleotide contains a nitrogenous base (碱基), a sugar (糖基), and one or more phosphates (磷酸基)
 - nucleic acids are named for the type of sugar:
 - DNA has 2'-deoxyribose (脱氧核糖)
 - RNA has ribose (核糖)
 - four types of nitrogenous bases in DNA:
 - two purines (嘌呤): adenine (A, 腺嘌呤) and guanine (G, 鸟嘌呤)
 - two pyrimidines (嘧啶): cytosine (C, 胞嘧啶) and thymine (T, 胸腺嘧啶)
 - four types of nitrogenous bases in RNA: A, G, C, U (uracil, 尿嘧啶)
 - In RNA, U is found instead of T
 - RNA is a single polynucleotide (多核苷酸链) strand
 - the sequence of mRNA is complementary to the sequence of one of the two polynucleotide strand of DNA (antisense or template strand, 反义链或模板链)
 - the sequence of mRNA is identical (apart from the replacement of T with U) to the other strand of DNA (coding or sense strand, 编码链或有义链)
- Remark:** DNA is a sequence of A-T-C-G, RNA is a sequence of A-U-C-G
- nucleic acid sequences is written in the 5' phosphate group to 3' hydroxyl group (羟基) direction

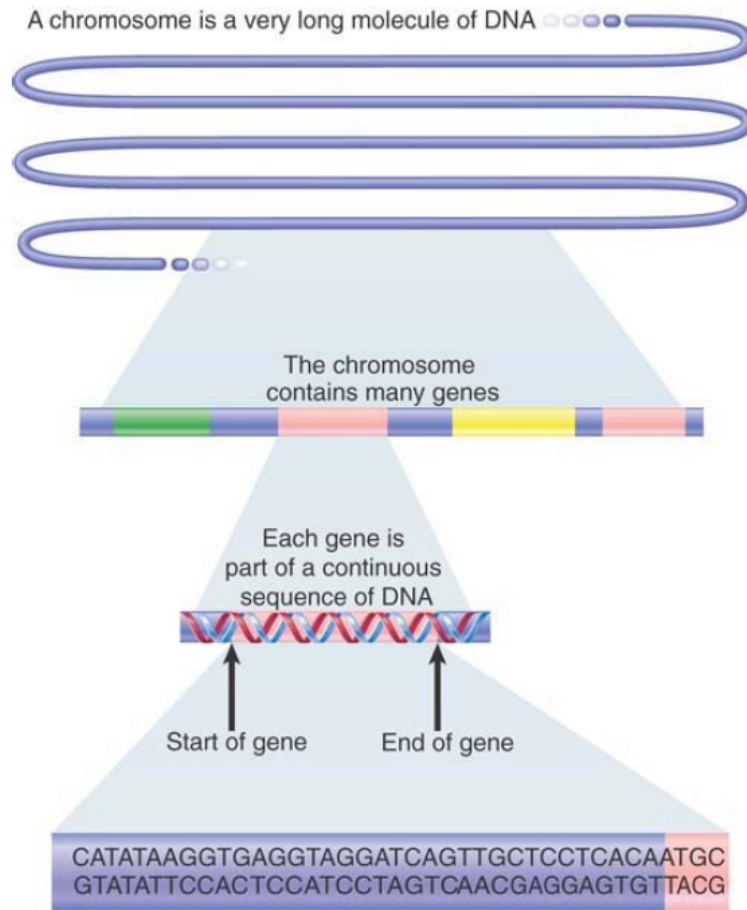


FIGURE 1.3 Each chromosome consists of a single, long molecule of DNA within which are the sequences of individual genes.

- **Double-helix model** (James Watson and Francis Crick, 1953):
 - Complementary base pairing (碱基互补配对): A-T and C-G, connected by hydrogen bonds
 - G-C content: the sum of the proportions of G and C bases, range from 26% to 74%
 - Antiparallel: two polynucleotide chains running in opposite directions

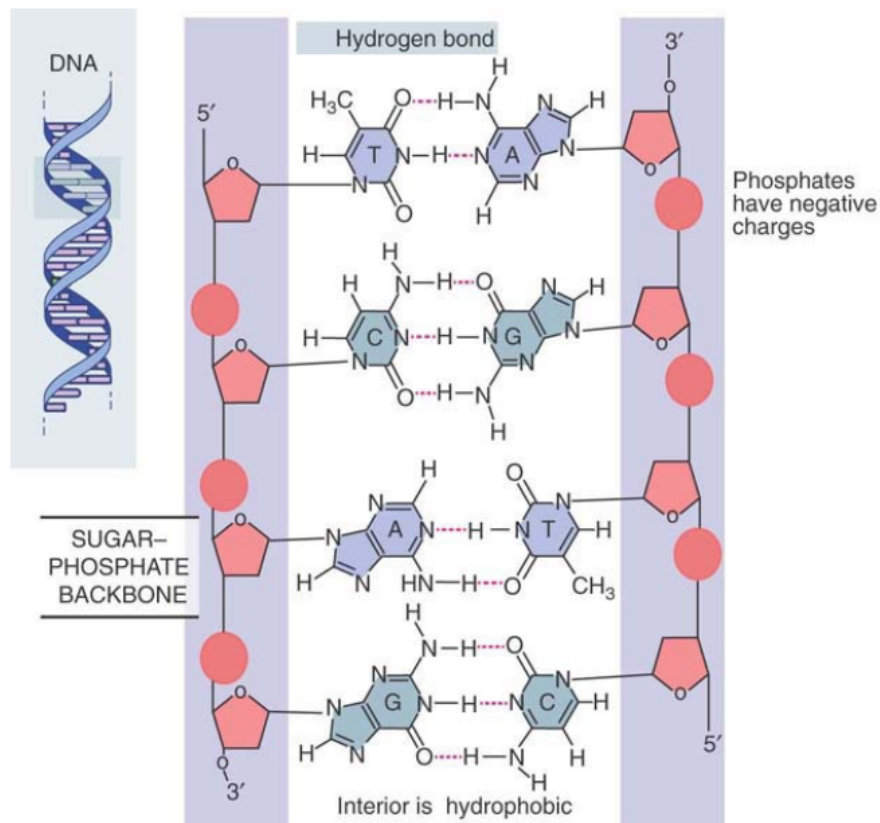


FIGURE 1.12 The double helix maintains a constant width because purines always face pyrimidines in the complementary A-T and G-C base pairs. The sequence in the figure is T-A, C-G, A-T, G-C.

- **DNA semiconservative replication** (半保留复制, Matthew Meselson and Franklin Stahl, 1958): the parental DNA carries a “heavy” density label, such as a suitable isotope ^{15}N

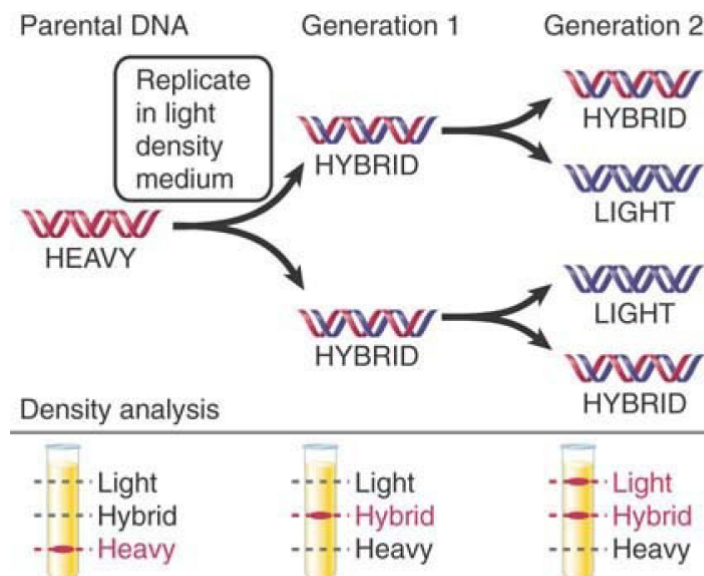


FIGURE 1.16 Replication of DNA is semiconservative.

- **Genetic Central Dogma:** Relationship of DNA - RNA - Protein
 - **Transcription** (转录): DNA produces **RNA**
 - **Translation** (翻译): RNA produce proteins, which serve diverse functions in an organism

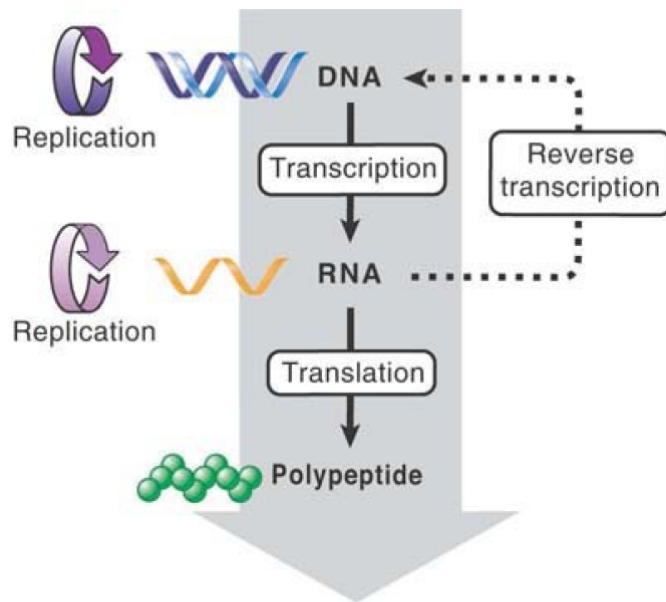


FIGURE 1.20 The central dogma states that information in nucleic acid can be perpetuated or transferred, but the transfer of information into a polypeptide is irreversible.

- DNA are genetic materials in cells. Some viruses use RNA.
- The expression of transcription and translation is usually unidirectional, except some special retroviruses infect cells (reverse transcription)
- **Mutation:** changes in the sequence of DNA
 - mutations can occur spontaneously or can be induced by mutagens (诱变剂)
 - the frequency of mutation for most base pairs is statistically equivalent
 - hotspots: the frequency of mutation is increased by at least an order of magnitude
 - The reason for hotspots is chemical modification of bases, e.g., methylation (甲基化)
- **One gene-one polypeptide hypothesis:**
 - most structural genes (most) encode mRNAs, which are translated to polypeptides
 - some genes encode rRNAs or tRNAs that are not translated and are functional themselves
- **Genetic code:** relationship between a sequence of DNA and polypeptide
 - coding sequence is read in groups of three nucleotides, which is called a codon (密码子)
 - each codon represents one amino acid (氨基酸)
 - nonoverlapping principle:
 - an individual nucleotide is part of only one codon
 - coding sequence is read sequentially. Different parts cannot be read independently
 - every coding sequence has three possible reading frames
 - A sequence that can be translated into polypeptide has a reading frame: begin with a special initiation codon (AUG) and end at one of three termination codons (UAA, UAG, UGA)
 - Usually only one of the three possible reading frames can be translated and the other two are closed by frequent termination signals
 - If a sequence is closed in all three reading frames, it cannot encode polypeptide.

- **Gene expression:** the process that uses information from a gene to synthesize an RNA or polypeptide, including transcription (DNA \Rightarrow mRNA) and translation (mRNA \Rightarrow polypeptid)
 - coding region: the part of nucleotides in mRNA that contain the codons for the amino acids
 - noncoding region: other parts of nucleotides. These regions are not translated directly, but important for mRNA stability and translation.
 - RNA processing:
 - intron (内含子): regions of noncoding sequence embedded in coding sequence. These sequence are spliced after transcription.
 - exon (外显子): regions of coding sequence are joined together in the mature mRNA

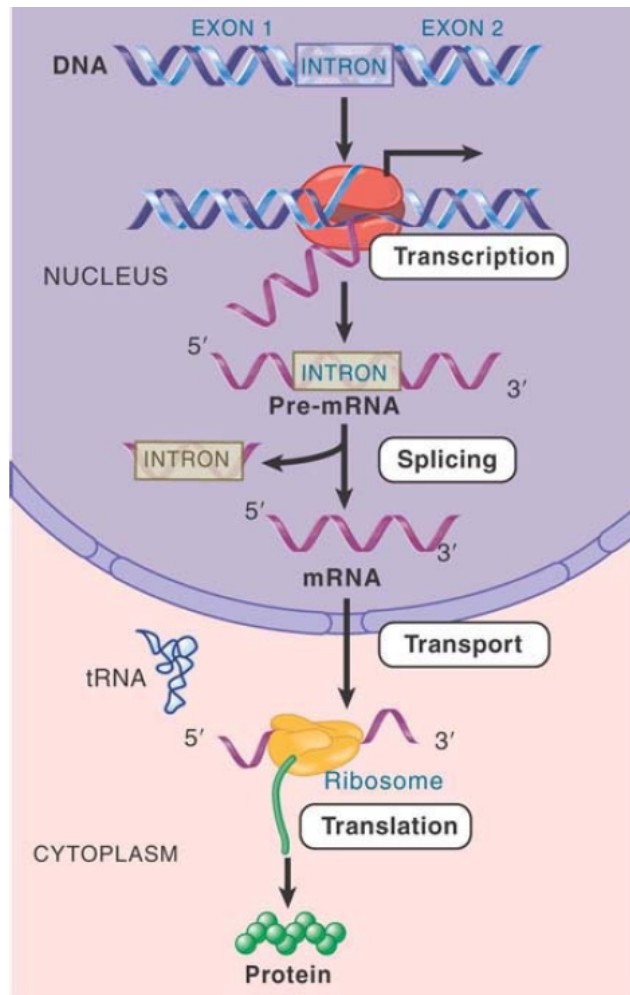


FIGURE 1.46 In eukaryotes, transcription occurs in the nucleus and translation occurs in the cytoplasm.