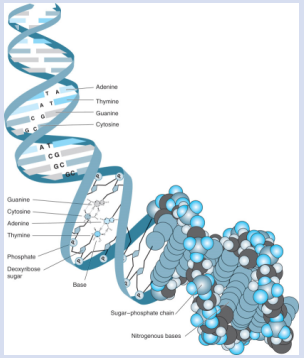
Bioinformatics



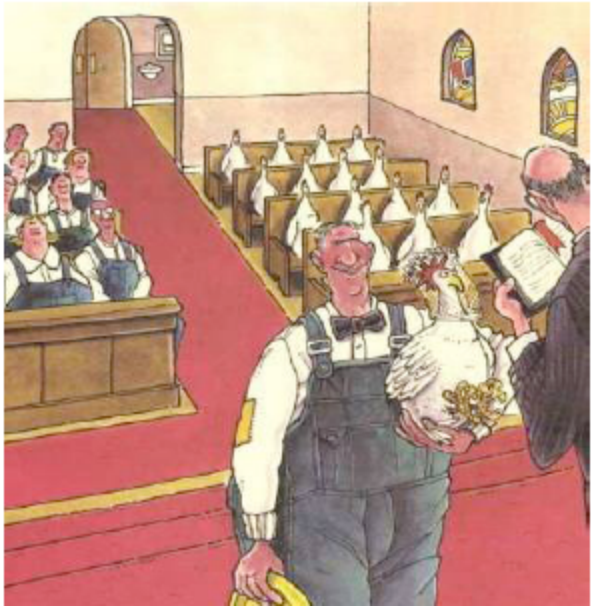
Introduction

- What is bioinformatics?

  Interdisciplinary subject between:

  – Computer science
  – Biology
  – Medicine
  – Other life sciences

- A marriage between biology and computers

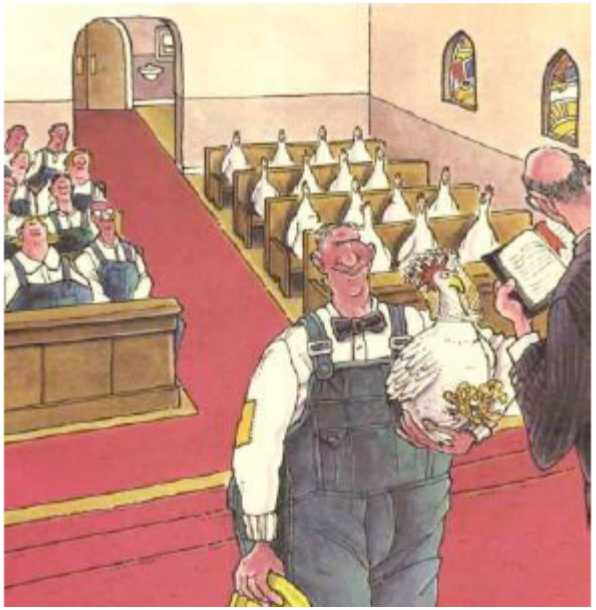


Introduction

- What is bioinformatics?

  biological and medical data on:

  – Nucliec acids:
    - RNA
    - DNA
  – Genes
  – Proteins
  – Bioinformatics is generally defined as the analysis, prediction, and modeling of biological and medical data with the help of computers

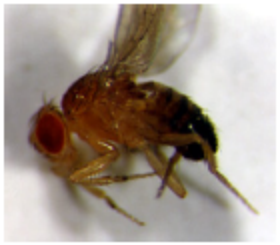


Bioinformatics introduction: continued

- Bioinformatics is the combination of biology and computing.
- DNA sequencing technologies have created massive amounts of information that can only be efficiently analyzed with computers.
- So far 70 species sequenced
  - Human, rat chimpanzee, chicken, and many others.
- As the information becomes ever so larger and more complex, more computational tools are needed to sort through the data.
  - Bioinformatics to the rescue!!!

Biological and medical data:

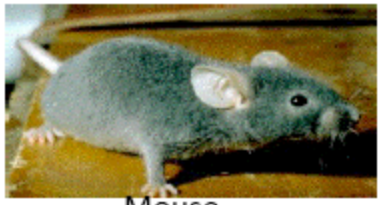
- Data storage and retrieval
- analysis of data
- Information extraction from data
- Gene functions
- Pathways
- Protein function
- Sequence and structure analysis


Drosophila


C. elegans

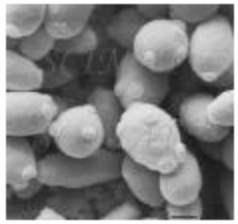

Rat

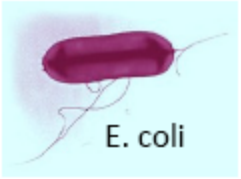

Human

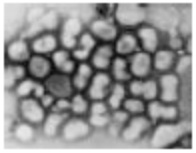

Mouse


Rice


Mosquito


Yeast


E. coli


H. influenza


Arabidopsis

Species and Genomes

History

- Molecular biology began in 1600s
- Robert Hooke (1635-1703) discovered organisms are made up of **cells**

- Robert Hooke

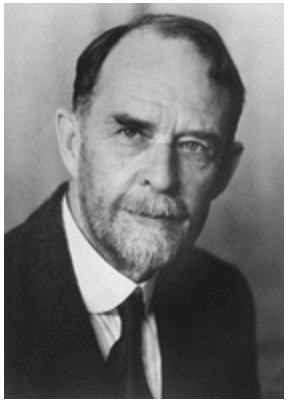
Genetics

- **1865** Gregor Mendel discover the basic rules of heredity of garden pea.
  - An individual organism has two alternative heredity units for a given trait (dominant trait v.s. recessive trait)

Mendel:
The Father of Genetics

Genes and RNA

- **1911 –** Thomas Hunt Morgan discovers genes on chromosomes are the discrete units of heredity
- **1911** Pheobus Aaron Theodore Lerene discovers RNA
- **1941** – George Beadle and Edward Tatum identify that genes make proteins

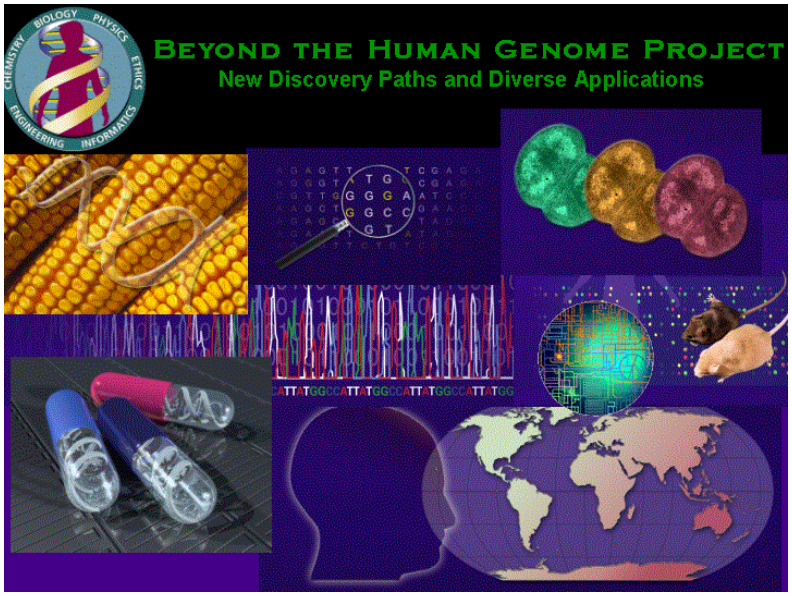


Thomas Morgan

Genome

- **1986** Human Genome Initiative announced

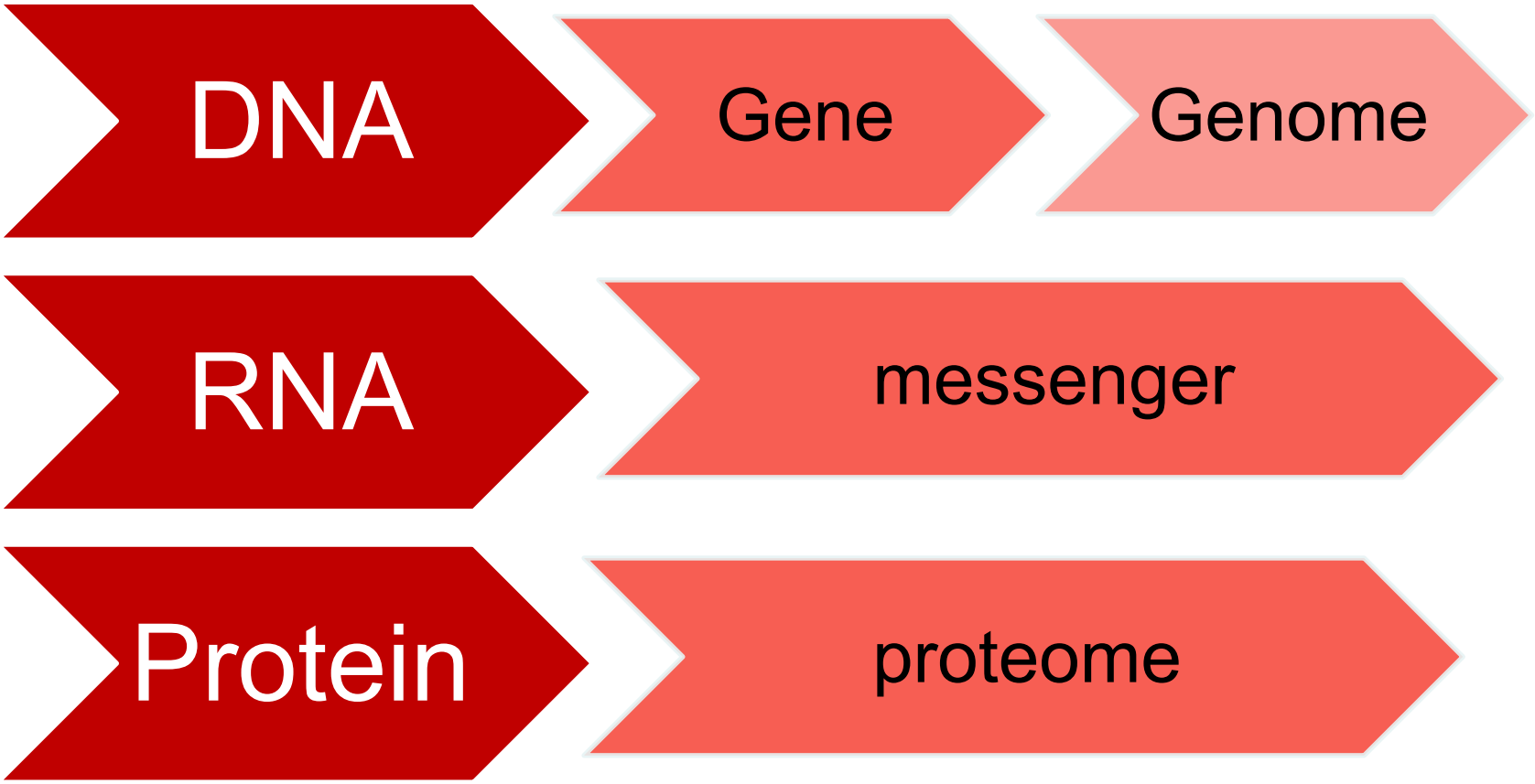
- **1990** The 15 year Human Genome project is launched by congress

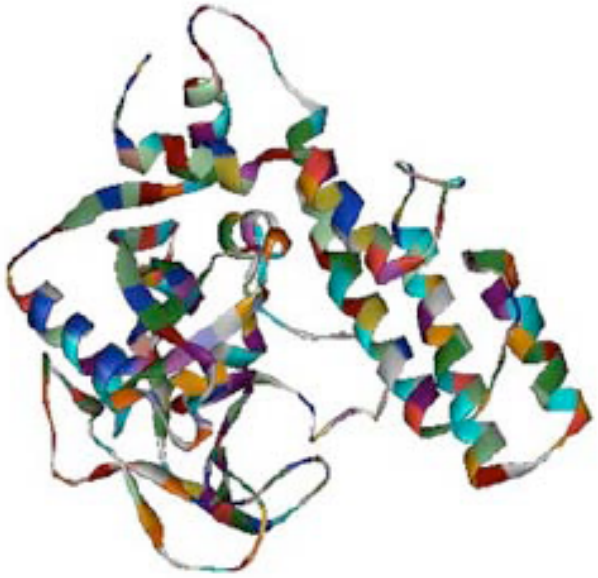


- **April 2003** Human Genome Project Completed. Mouse genome is sequenced.

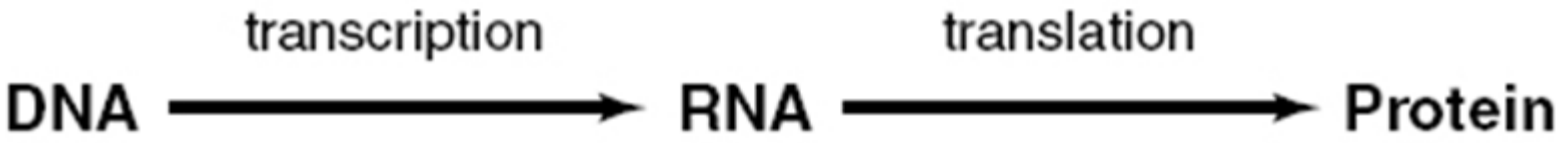
- **April 2004** Rat genome sequenced.

The three main objects





Protein 3D structure



- DNA (Deoxyribonucleic acid) is arranged into 23 chromosome pairs in the nucleus of each cell.
- DNA: 4 bases:
  - A, C, G, T
- Just like human text, genomic sequences represent a language written in A, T, C, G

-DNA
-RNA
-Protein

DNA

transcription

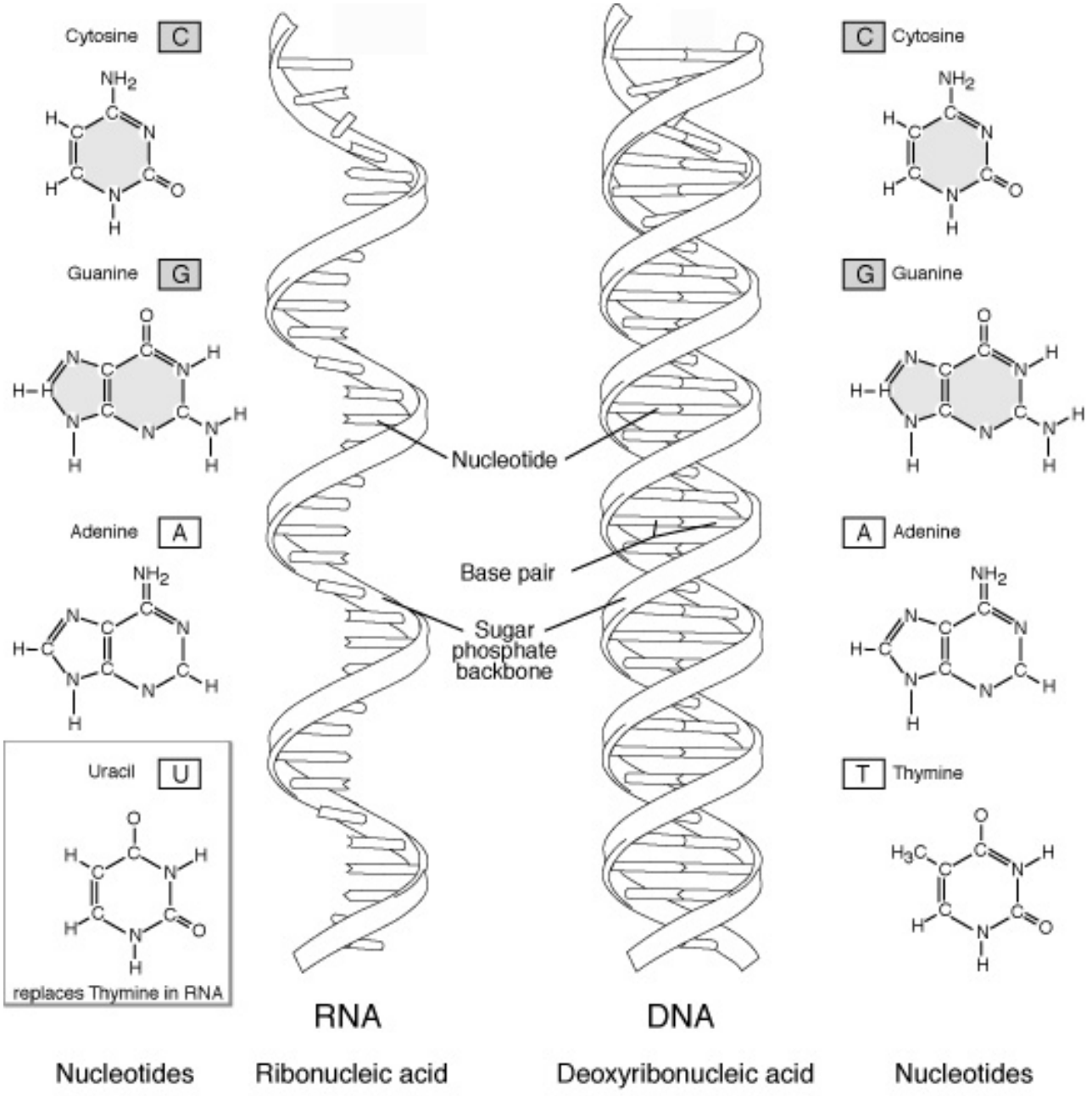
↓

RNA

translation

↓

Protein



-DNA
-RNA
-Protein

DNA

transcription

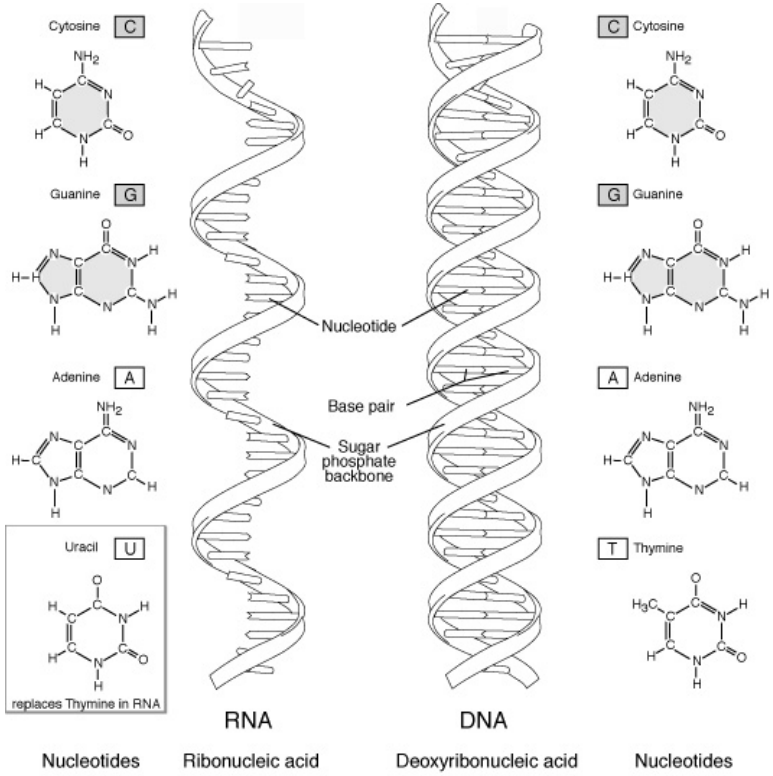
↓

RNA

translation

↓

Protein



Protein: is a chain of amino acid

- there are around 20 amino acids known in human body

A protein is usually composed of 50 to 400+ amino acids.

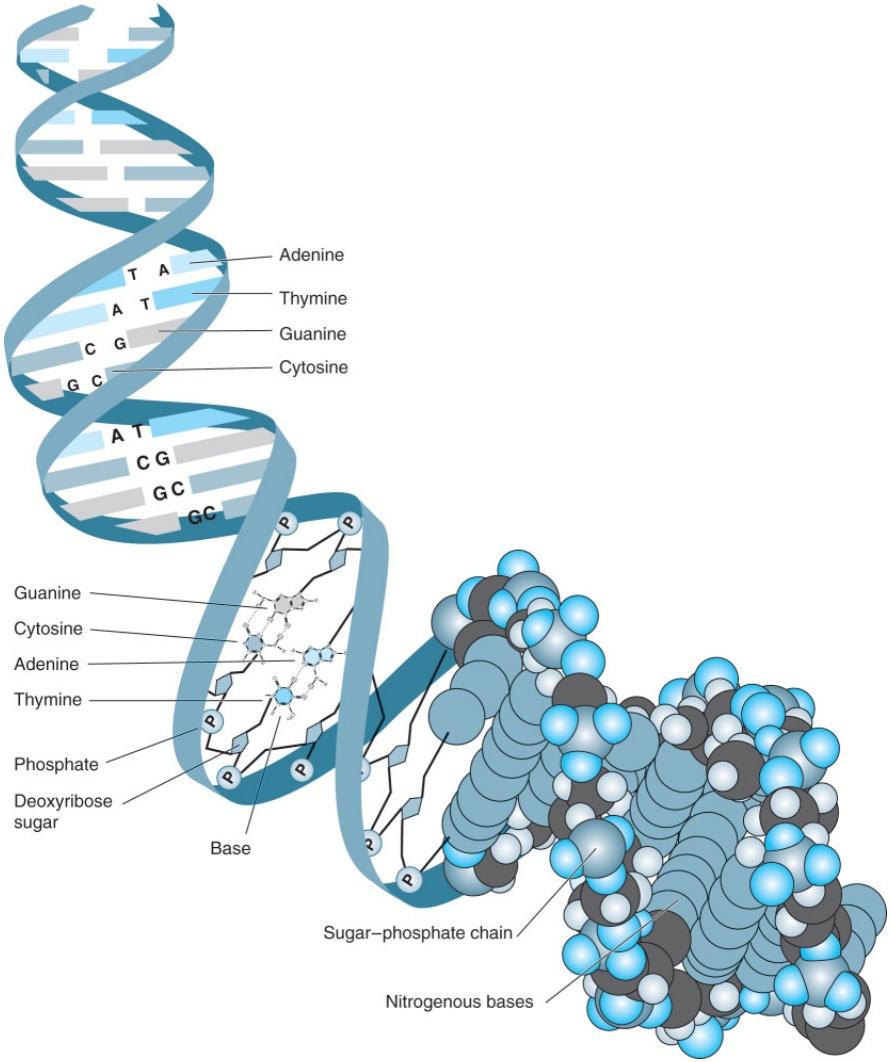
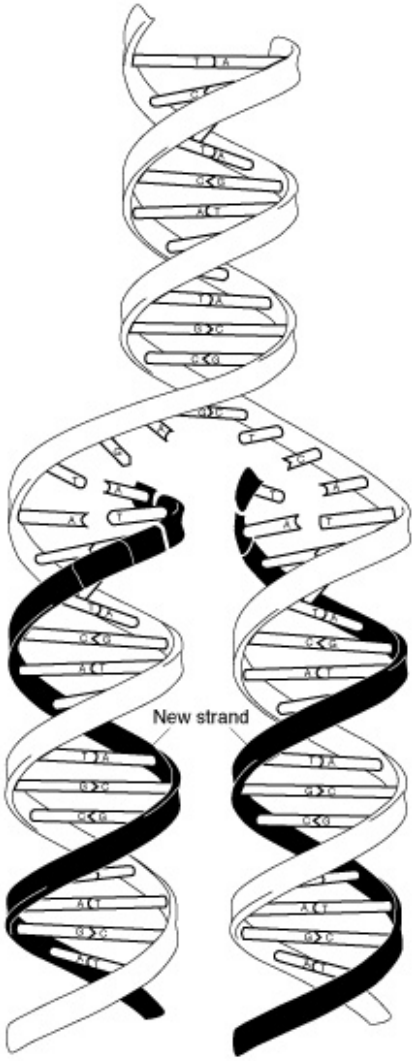
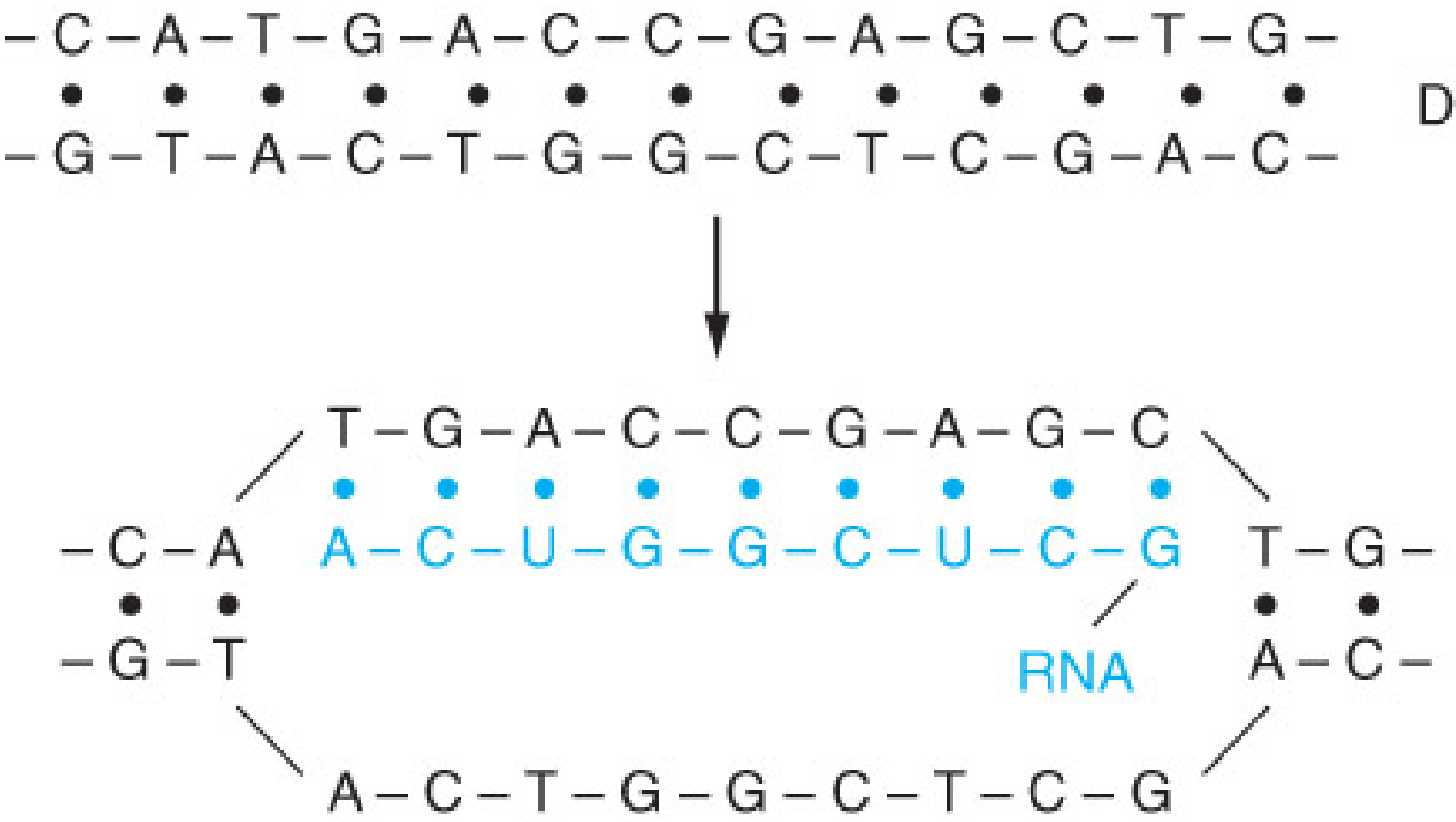


Figure 2.5: A DNA molecule is made up of two strands of nucleotides

DNA Replication

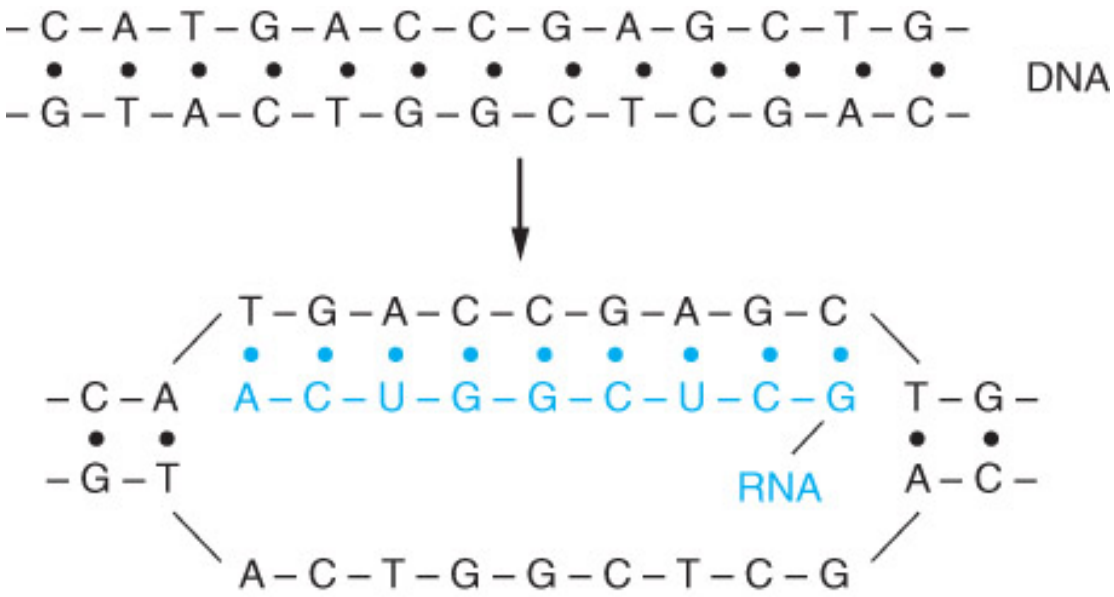
- Prior to cell division, all the genetic instructions must be "copied" so that each new cell will have a complete set
- DNA polymerase is the enzyme that copies DNA





DNA: **C G A T**

RNA: **C G A U** ***==> notice: RNA is single strand.***



- DNA: C G A T

- RNA: C G A U

- Proteins: 20 letters from the English alphabet
  (*all letters except for B, J, O, U, X, and Z*).

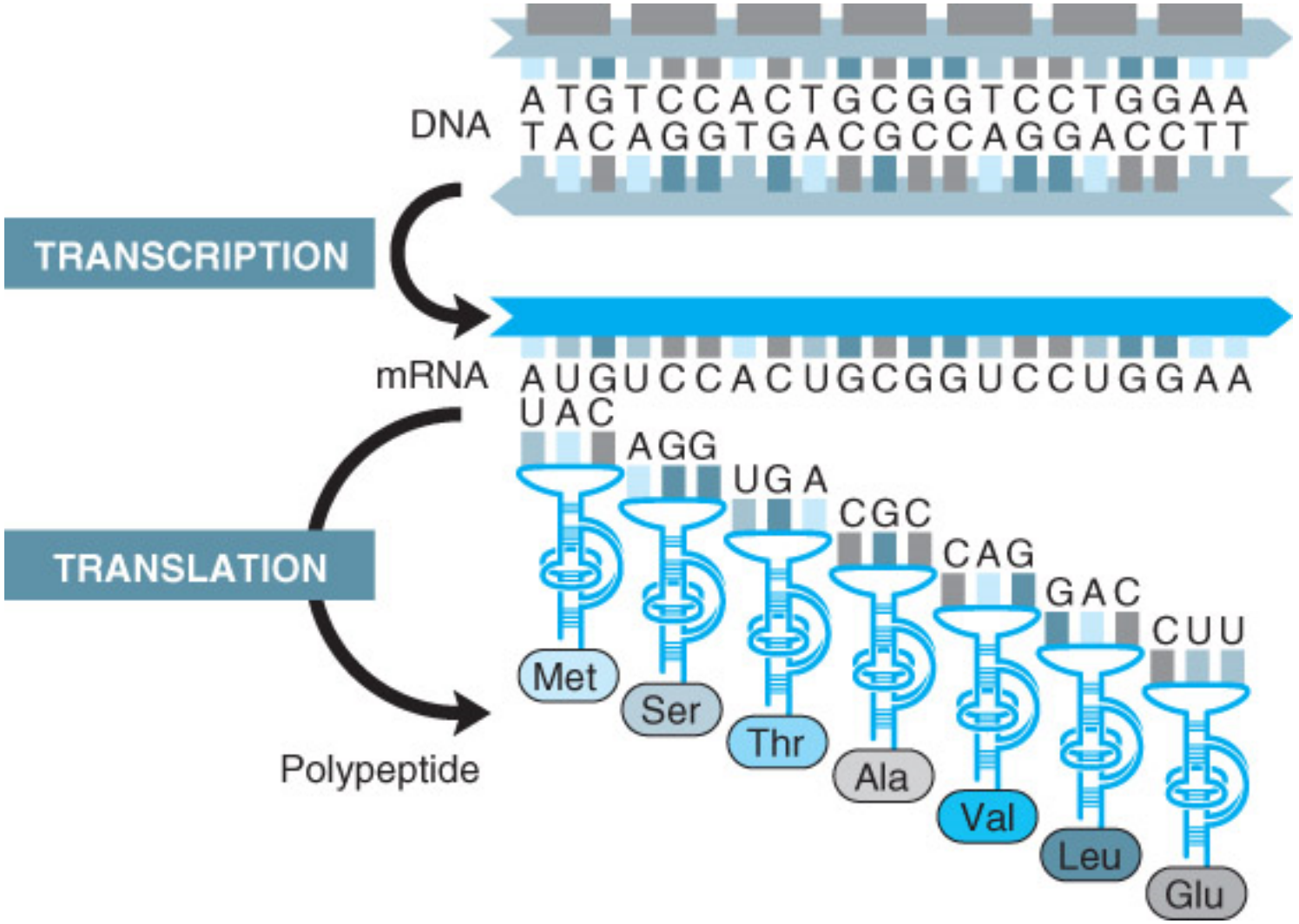
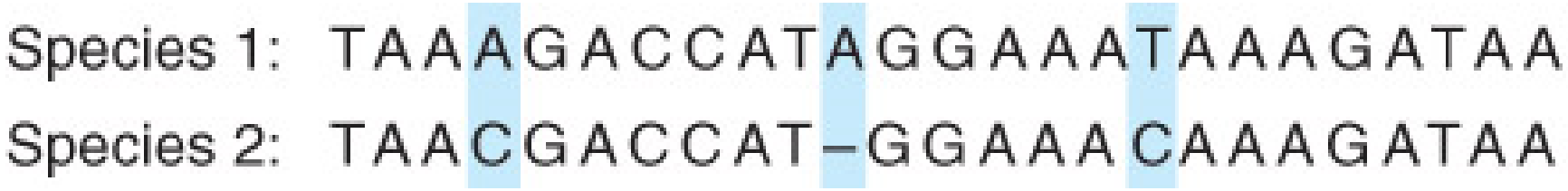


Diagram showing how a single gene encoded in DNA is transcribed and translated to produce a protein

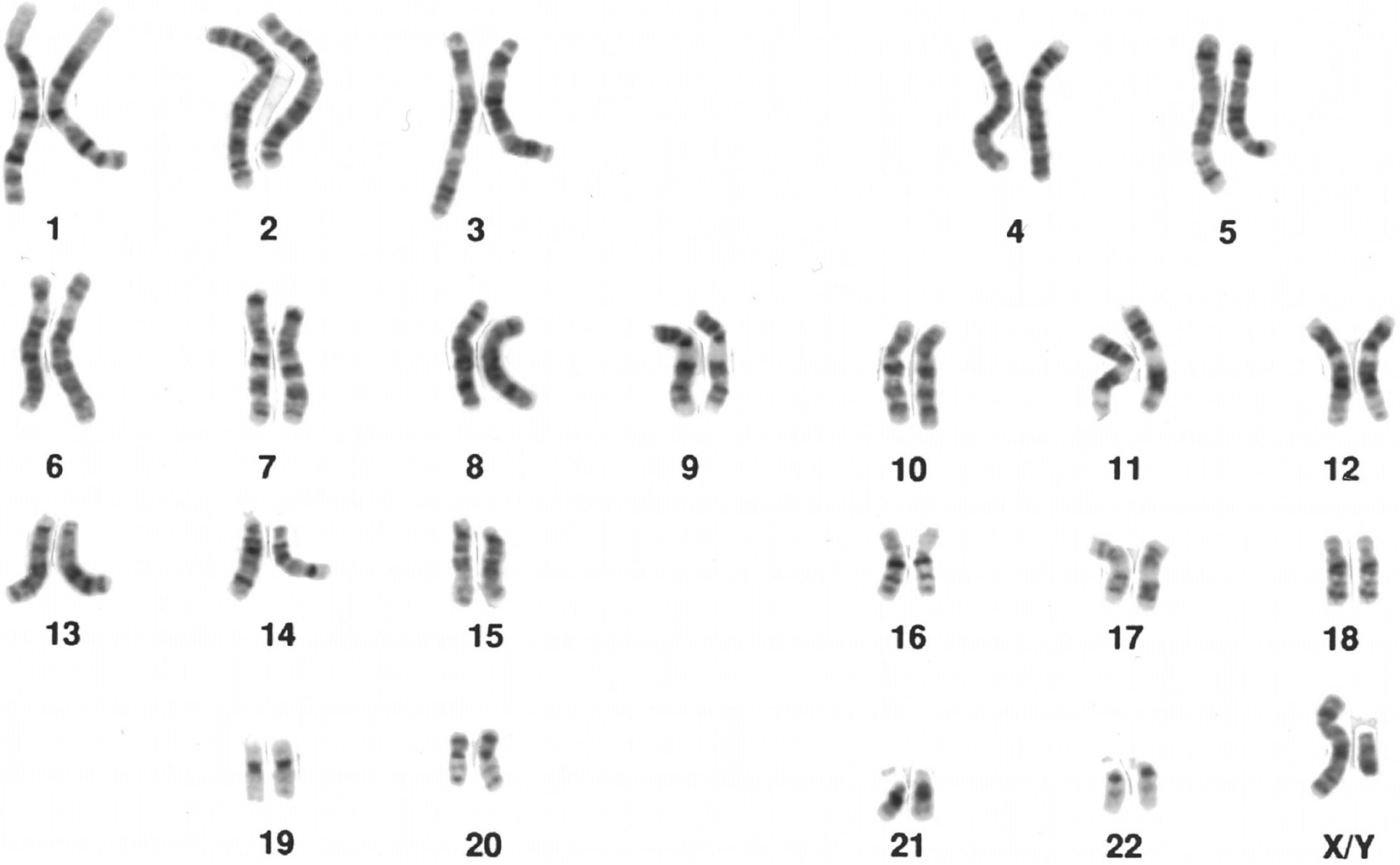
>gi | 28302128 | ref | NM_000518.4 | Homo sapiens hemoglobin, beta (HBB), mRNA

ACATTTGCTTCTGACACAACTGTGTTCACTAGCAACCTCAAACAGACACCATGGTGCATCTGACTCCTGA
GGAGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCTGGGC
AGGCTGCTGGTGGTCTACCCTTGGACCCAGAGGTTCTTTGAGTCCTTTGGGGATCTGTCCACTCCTGATG
CTGTTATGGGCAACCCTAAGGTGAAGGCTCATGGCAAGAAAGGTCTCGGTGCCTTTAGTGATGGCCTGGC
TCACCTGGACAACCTCAAGGGCACCTTTGCCACACTGAGTGAGCTGCACTGTGACAAGCTGCACGTGGAT
CCTGAGAACTTCAGGCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCCCATCACTTTGGCAAAGAATTCA
CCCCACCAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTAATGCCCTGGCCCACAAGTATCA
CTAAGCTCGCTTTCTTGCTGTCCAATTTCTATTAAAGGTTCCTTTGTTCCCTAAGTCCAACTACTAAACT
GGGGGATATTATGAAGGGCCTTGAGCATCTGGATTCTGCCTAATAAAAAACATTTATTTTCATTGC

The human HBB gene in FASTA format

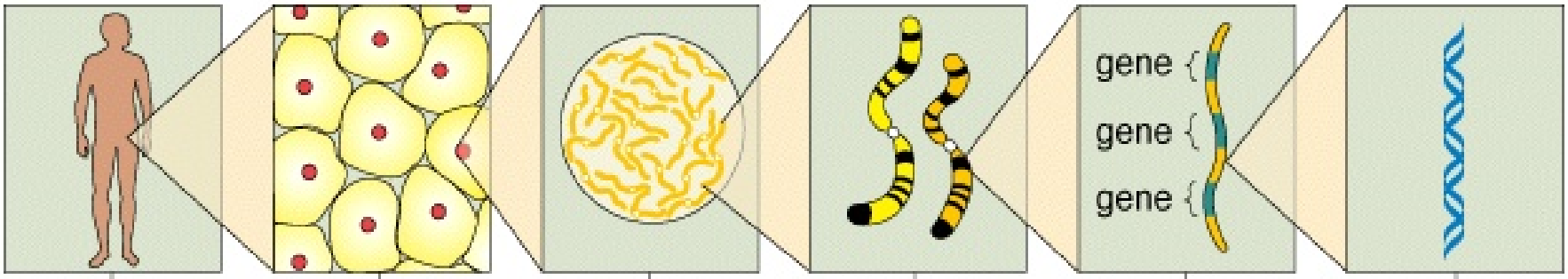


Determining the similarity of two or more genes: by aligning their nucleotide sequences as well as possible; the differences due to mutation are shown in boxes



The 23 sets of chromosomes in the human genome

In Summary



Organism (human)

A human body is made up of trillions of cells.

Each cell nucleus contains an identical complement of chromosomes.

One specific chromosome pair

Each chromosome is one long DNA molecule, and genes are functional regions of this DNA.

DNA is a double helix.