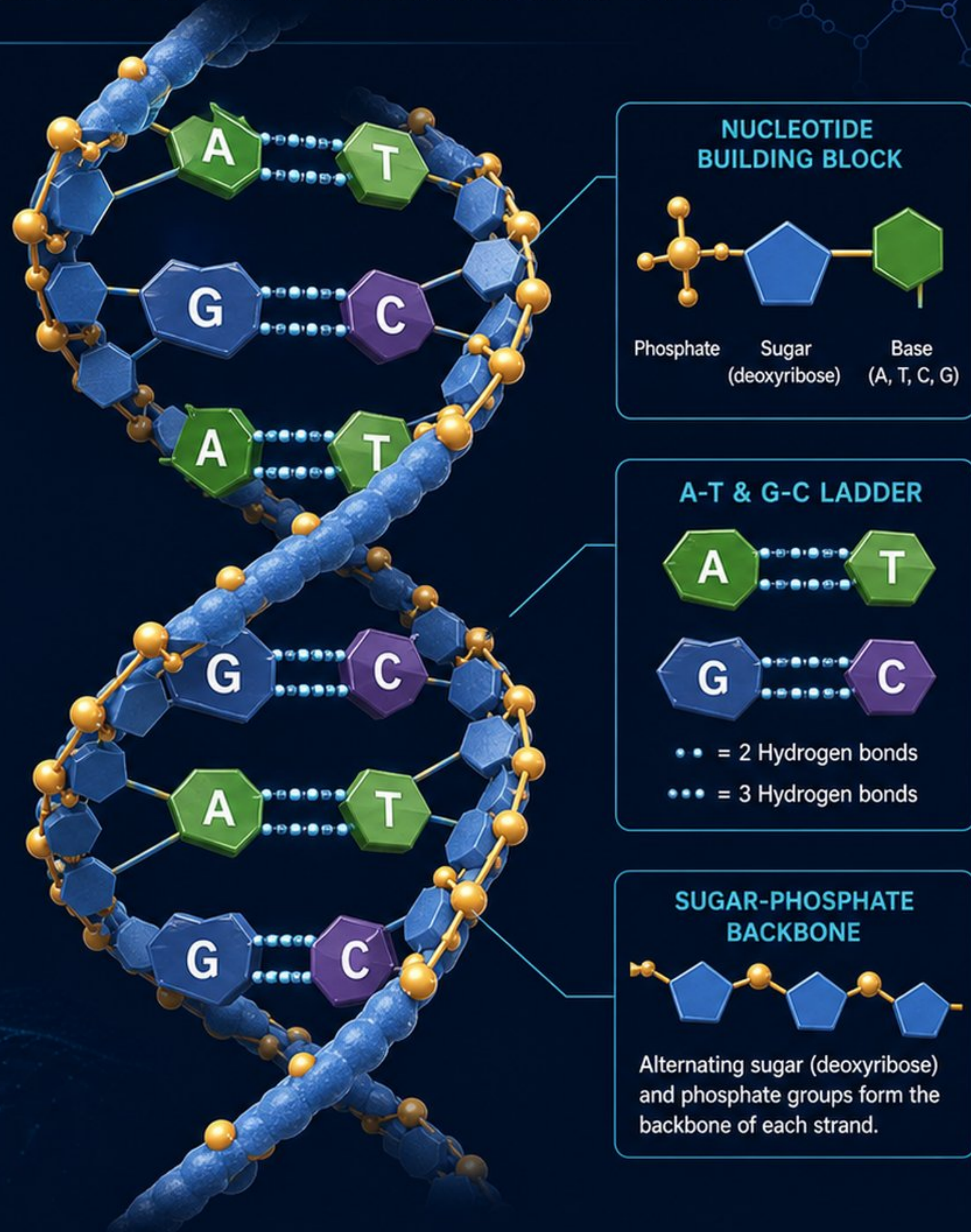
DNA: Information in Living Things

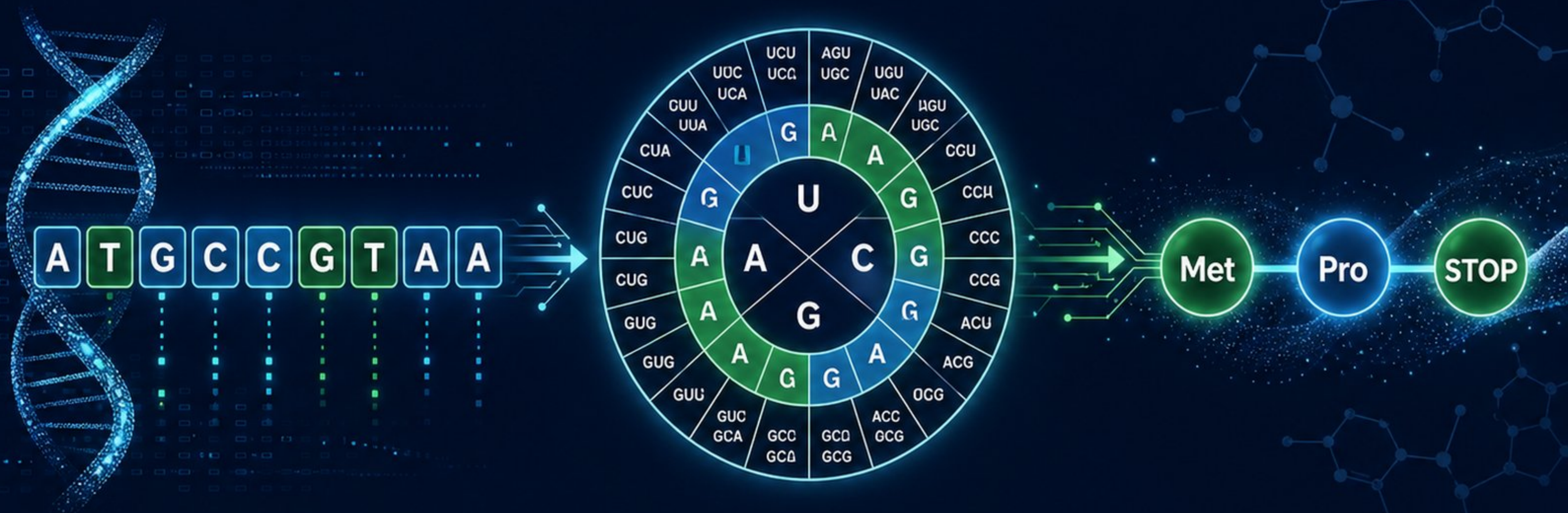
- DNA stores biological instructions using a four-letter base code (A, T, C, G).
- The double helix enables accurate copying and information transfer.
- Watson, Crick, and Franklin revealed DNA as the molecule of heredity.
- Course roadmap: structure, code, expression, replication, and modern editing.

What Is DNA? Structural Foundations

- DNA is a polymer made of repeating nucleotide building blocks
- Each nucleotide has a sugar, a phosphate, and one base (A, T, C, G)
- The double helix: two anti-parallel strands held by base pairing
- A always pairs with T; C always pairs with G
- Base sequences encode biological instructions, like letters forming words

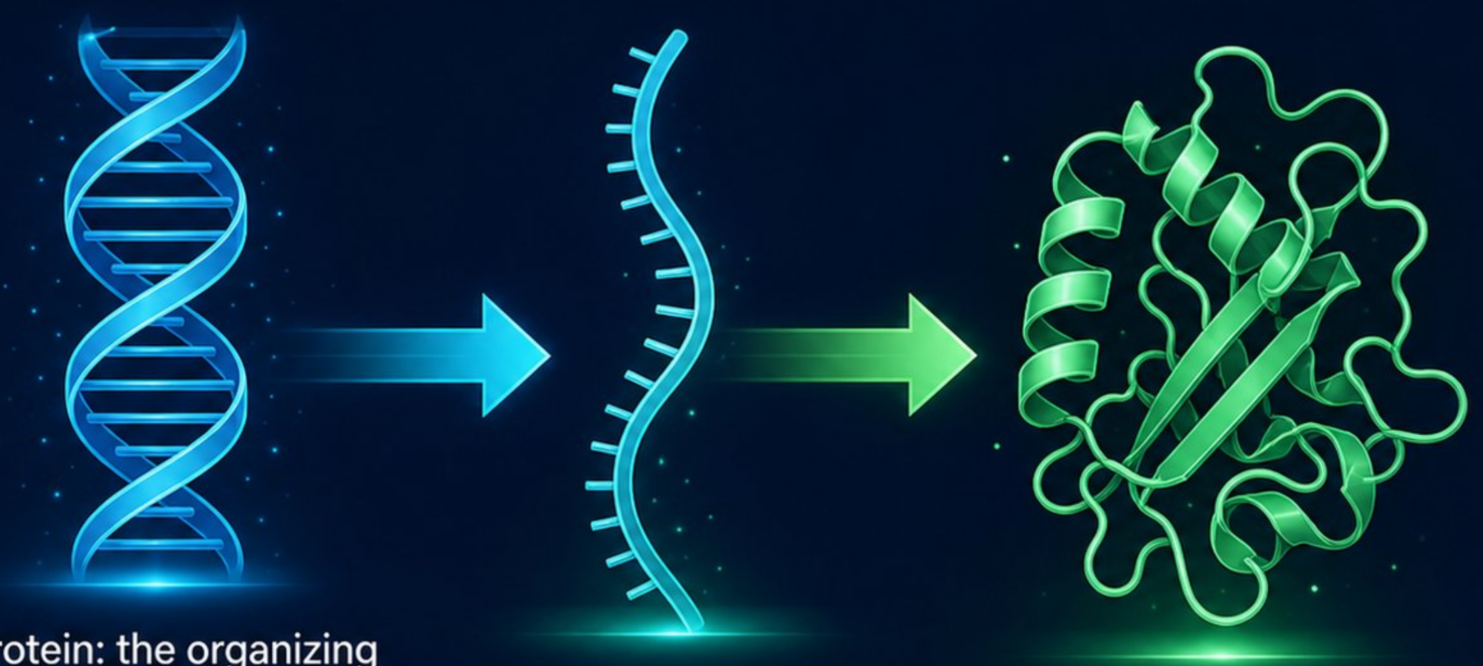


DNA as a Digital Code



- “DNA uses discrete symbols (A, T, C, G), not analog signals, akin to binary data”
- “64 possible nucleotide triplets (codons) encode only 20 amino acids”
- “The genetic code is nearly universal across species with built-in redundancy”
- “Example: DNA sequence ATG-CCG-TAA translates to Met-Pro-STOP”

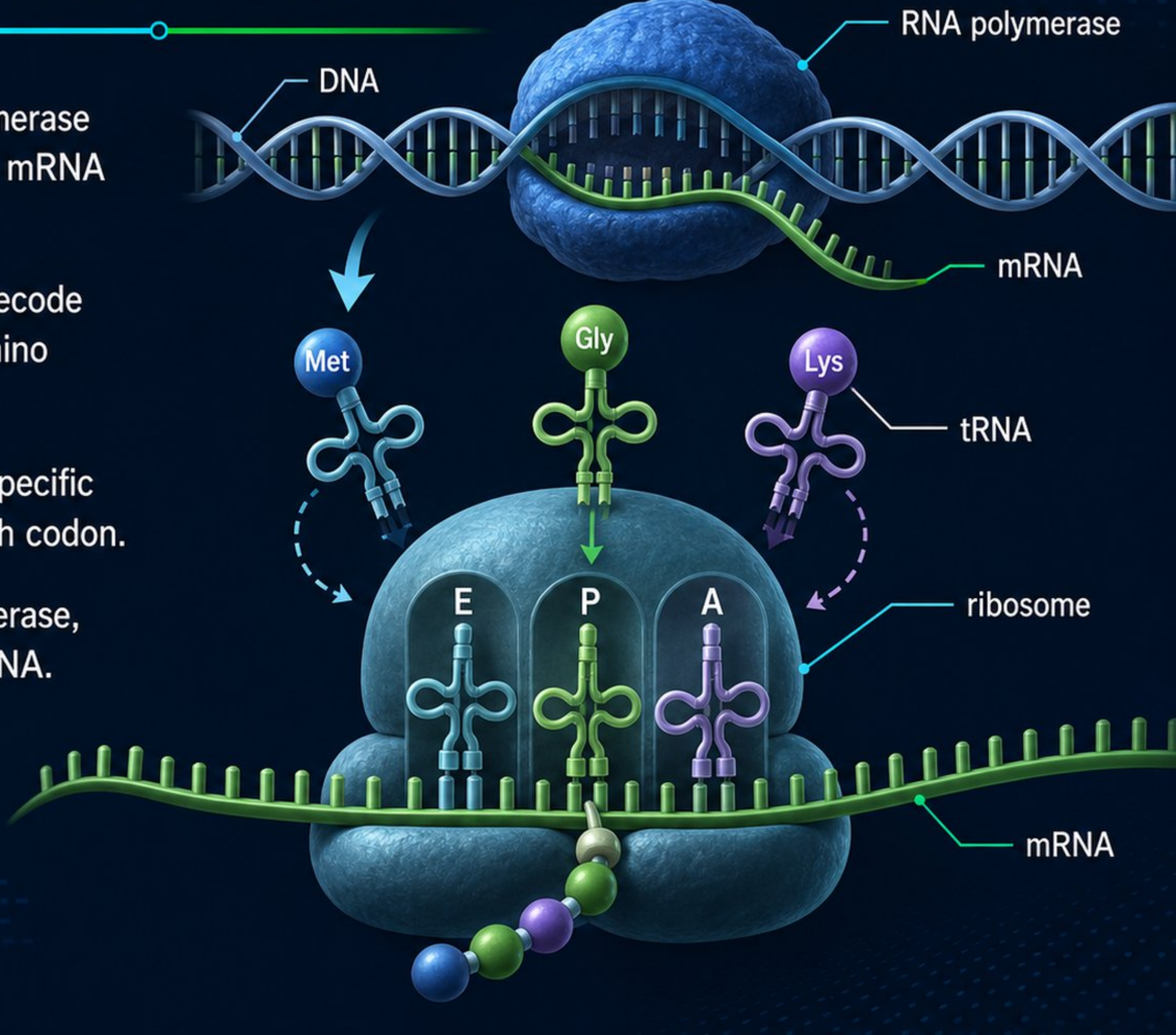
The Central Dogma:
From Information to Function



- DNA → RNA → Protein: the organizing principle of molecular biology
- Transcription: RNA polymerase copies gene information into mRNA
- Translation: ribosomes read mRNA codons and assemble amino acid chains
- Proteins are the functional expression of genetic information

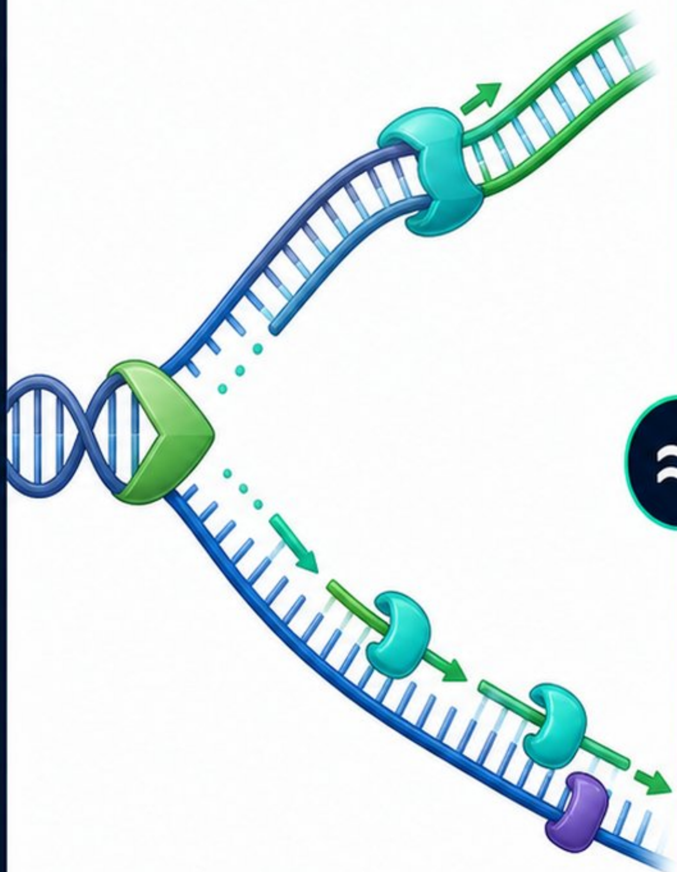
Gene Expression: A Closer Look at Transcription and Translation

- **Transcription:** RNA polymerase copies a gene's DNA into mRNA in three phases.
- **Translation:** ribosomes decode mRNA codons into an amino acid chain.
- **tRNA** molecules deliver specific amino acids matching each codon.
- **Key players:** RNA polymerase, ribosome, mRNA, and tRNA.

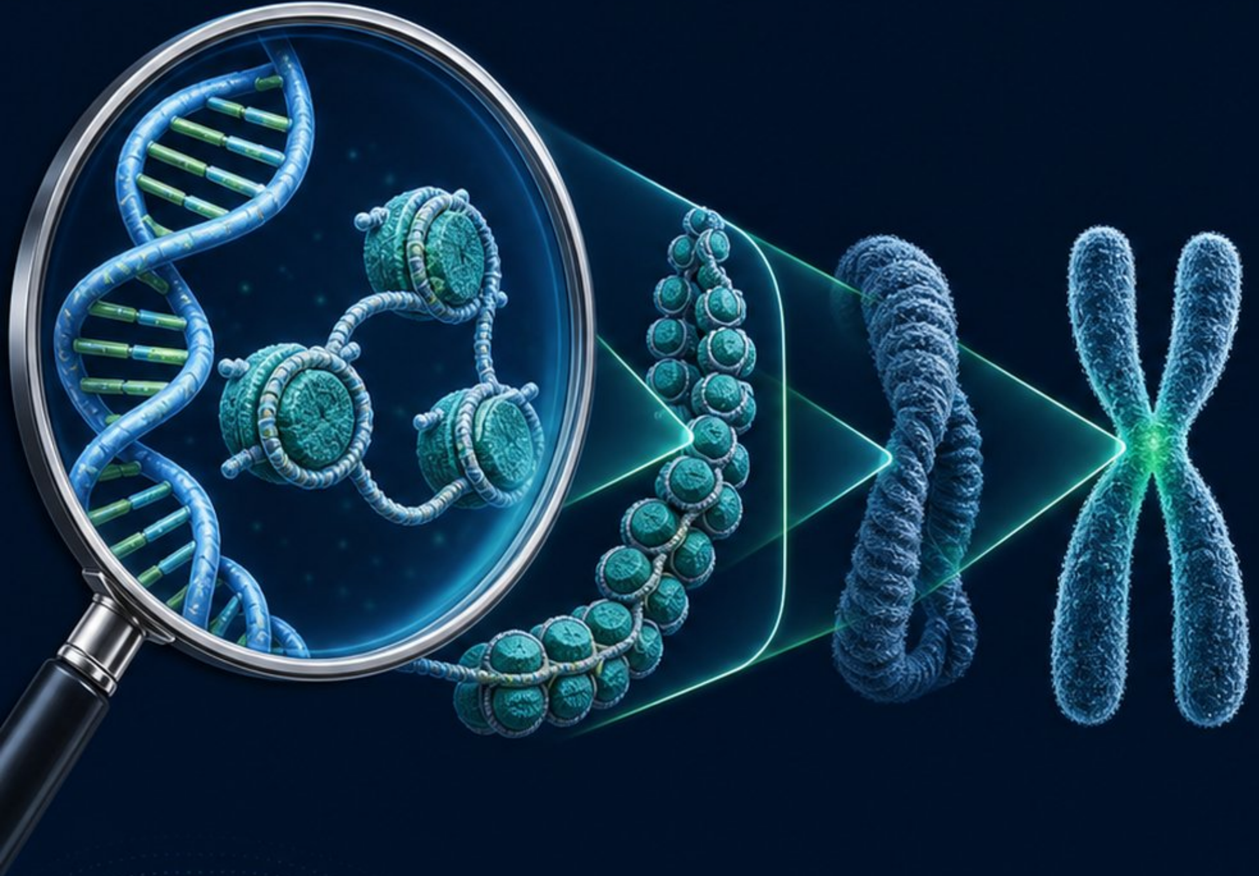


Copying the Information: DNA Replication

- Each strand serves as a template for a new complementary strand
- Semi-conservative: one old and one new strand in each daughter helix
- Leading strand synthesized continuously; lagging strand formed as Okazaki fragments
- Helicase unwinds DNA; DNA polymerase builds and proofreads; ligase joins fragments
- High fidelity replication acts like a data backup with built-in checksums



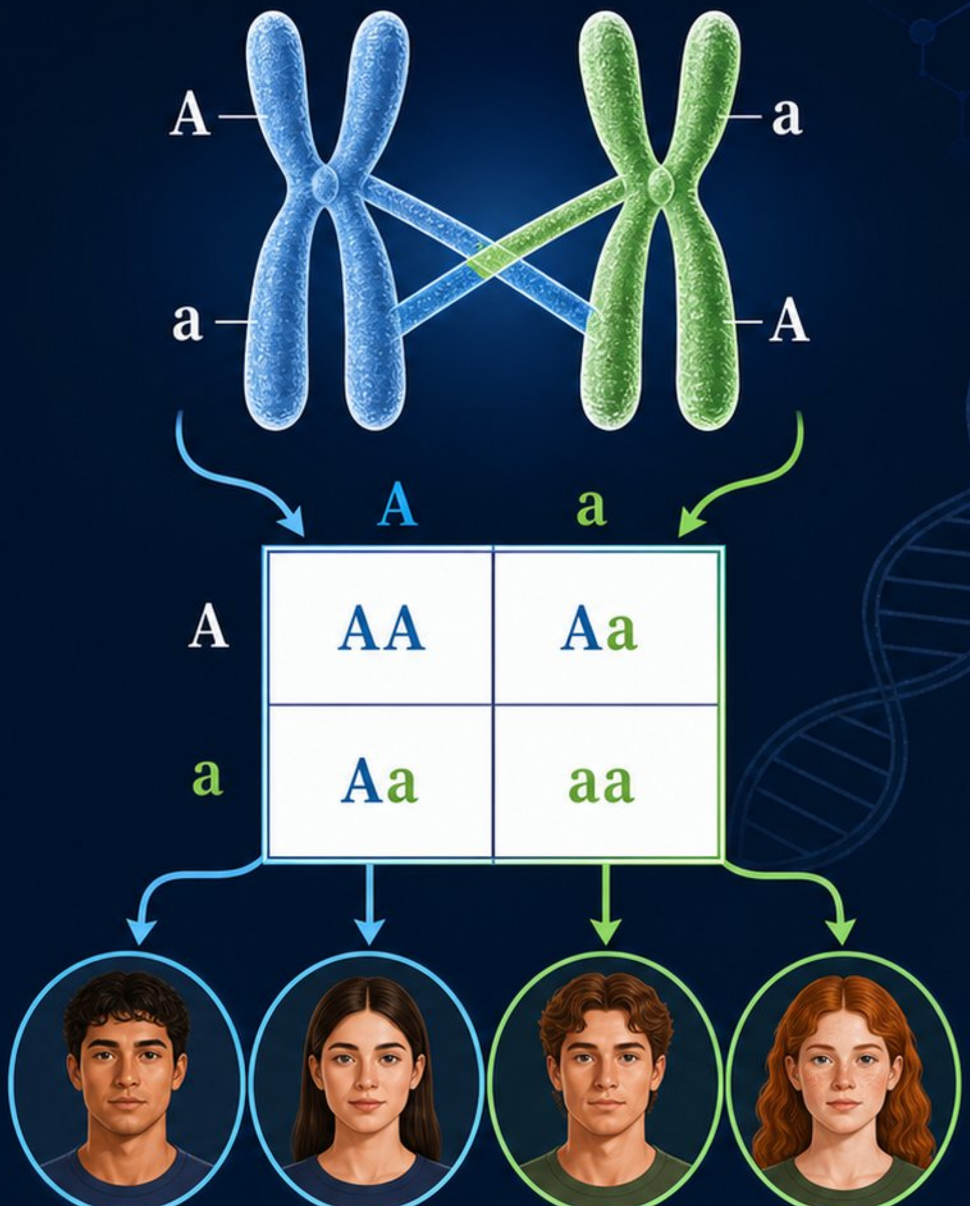
Organizing the Library: Chromosomes and Genomes



- DNA is packaged from double helix to nucleosomes, chromatin, and compact chromosomes.
- A genome is the complete set of an organism's genetic information.
- Humans are diploid with two information copies; sperm and eggs are haploid.
- Benchmark genome sizes: human ~3.2 Gb, fruit fly ~175 Mb, *E. coli* ~4.6 Mb.
- Genome is the information store; the organism is its physical expression.

Heredity: Passing Information Across Generations

- Alleles act as discrete units, transmitting dominant or recessive traits
- Meiosis shuffles genetic information via independent assortment and crossing over
- Sequence differences (genotype) produce observable traits (phenotype)
- Single-gene mutations link directly to phenotypes like sickle cell or cystic fibrosis

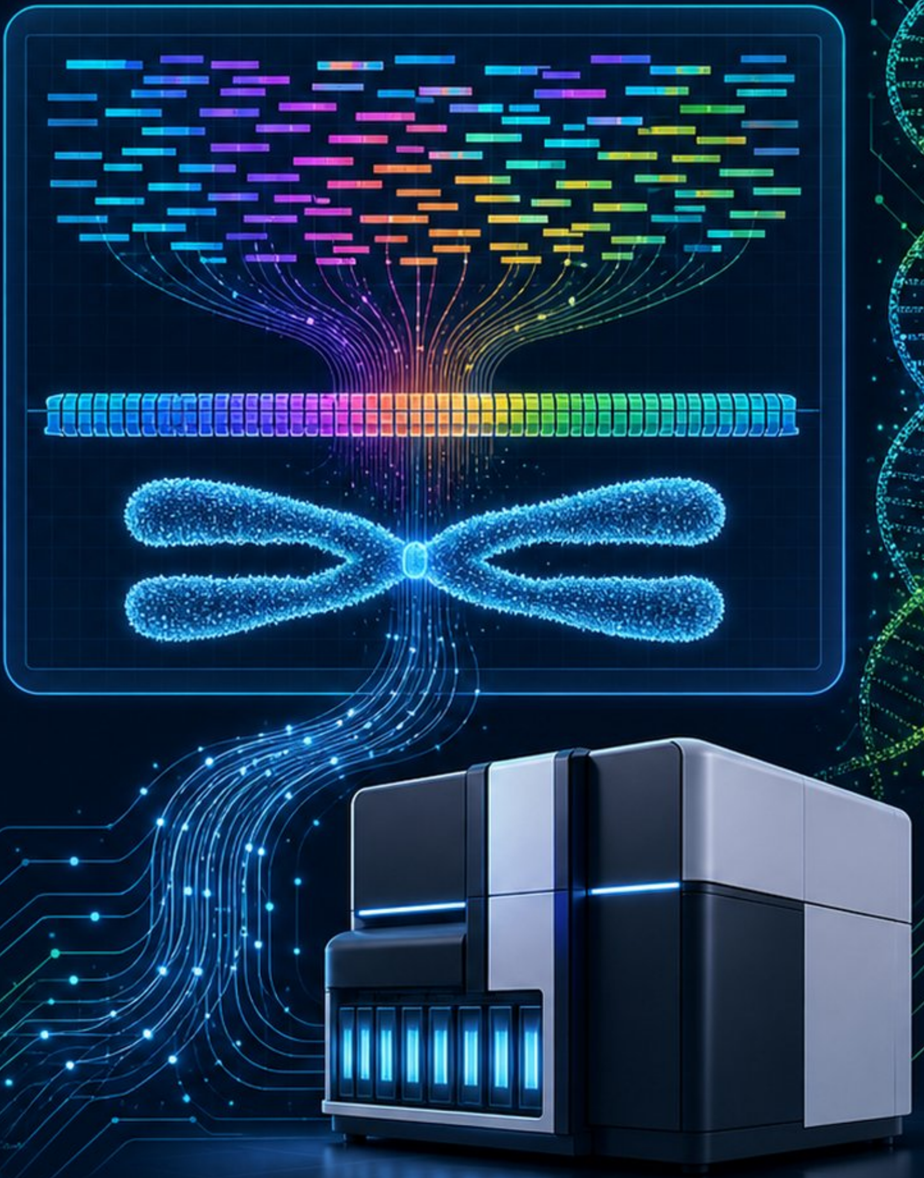


When Information Changes: Mutation and Variation

- Point mutations: substitutions, silent, missense, nonsense changes
- Insertions, deletions, duplications alter gene length and function
- Impact spectrum: neutral, beneficial, or harmful outcomes
- ~50–100 new SNVs per generation, with strong paternal age effect
- Caused by replication errors, radiation, mutagens; repair systems guard information

Reading the Code: DNA Sequencing

- Determining the precise order of bases in a DNA molecule
- Sanger sequencing evolved into massively parallel next-generation sequencing
- Billions of short reads assembled computationally into complete genomes
- Applications: medical diagnostics, ancestry, forensics, biodiversity



Editing the Code: CRISPR and Modern Genome Tools

- CRISPR-Cas9 functions as programmable ‘search and replace’ for DNA sequences
- AI-designed enzymes (OpenCRISPR-1, SynTnpBs) expand editing precision beyond natural proteins
- Small-molecule switches (PRINCE/Little Prince) enable on-demand control of therapeutic editing
- Prime editing enters first-in-human trials; compact editors now fit single AAV vectors
- CRISPR-based diagnostics rapidly detect cancer mutations from liquid biopsies

DNA as a Digital Storage Medium

- Encode digital data (text, images, video) into synthetic DNA base sequences
- Density: over 200 PB per gram; durability for millennia
- 2026 advances: Harvard's semiconductor chip, DNA Diamond framework, DNA-MGC+ codec
- Bottlenecks: synthesis cost, read/write speed, random access latency
- Currently strongest for deep archival scenarios

Ethics, Society, and Information Ownership

- Genetic privacy: your DNA can be read from a discarded coffee cup
- Heritable editing: global debate and moratoria on germline interventions
- Who owns your genome—you, researchers, or corporations?
- Equitable access: benefits must reach all, not just the privileged

DNA Information in Context:
Framework and Future Directions

- DNA: the most durable, replicable, and editable information system known
- Biological lens: evolution, heredity, and the flow of genetic information
- Technological lens: sequencing, CRISPR, and DNA-based data storage
- Societal lens: equity, ethics, and information ownership in medicine
- Build mental models to engage responsibly with biology and biotechnology