

Eukaryotic Genomes: Organization, Regulation, and Evolution

1. Eukaryotic DNA is precisely combined with large amounts of protein.
 - a. Eukaryotic chromosomes contain an enormous amount of DNA relative to their condensed length. Each human chromosome averages about 1.5×10^8 nucleotide pairs. If extended, each DNA molecule would be about 4 cm long, thousands of times longer than the cell diameter.
 - b. The chromosomes fit into the nucleus through an elaborate, multilevel system of DNA packing.
 - c. **Histone** proteins are responsible for the first level of DNA packaging.
 - i. The mass of histone in chromatin is approximately equal to the mass of DNA.
 - ii. Their positively charged amino acids bind tightly to negatively charged DNA.
 - iii. The five types of histones are very similar from one eukaryote to another, and similar proteins are found in prokaryotes. The conservation of histone genes during evolution reflects their pivotal role in organizing DNA within cells.
 - iv. Histones leave the DNA only transiently during DNA replication and they stay with the DNA during transcription.
 - d. Interphase chromosomes have highly condensed areas, **heterochromatin**, and less compacted areas, **euchromatin**.
 - a. Heterochromatin DNA is largely inaccessible to transcription enzymes presumably because they cannot reach the DNA.
 - b. Looser packing of euchromatin makes its DNA accessible to enzymes and available for transcription.
- II. In addition to its role in packing DNA inside the nucleus, chromatin organization regulates gene expression.
 - A. **Histone acetylation** (addition of an acetyl group —COCH₃) and deacetylation appear to play a direct role in the regulation of gene transcription.
 1. Acetylated histones grip DNA less tightly, providing easier access for transcription proteins in this region.
 2. Some of the enzymes responsible for acetylation or deacetylation are associated with or are components of transcription factors that bind to promoters.
 3. Thus histone acetylation enzymes may promote the initiation of transcription not only by modifying chromatin structure, but also by binding to and recruiting components of the transcription machinery.
 - B. **DNA methylation** is the attachment by specific enzymes of methyl groups (—CH₃) to DNA bases after DNA synthesis.
 1. Inactive DNA is generally highly methylated compared to DNA that is actively transcribed.
 2. DNA methylation proteins recruit histone deacetylation enzymes, providing a mechanism by which DNA methylation and histone deacetylation cooperate to repress transcription.
 3. Once methylated, genes usually stay that way through successive cell divisions.
 4. Methylation enzymes recognize sites on one strand that are already methylated and correctly methylate the daughter strand after each round of DNA replication.
 5. This methylation patterns accounts for **genomic imprinting** in which methylation turns off either the maternal or paternal alleles of certain genes at the start of development.
 - a. The chromatin modifications just discussed do not alter DNA sequence, and yet they may be passed along to future generations of cells.
 - b. Inheritance of traits by mechanisms not directly involving the nucleotide

sequence is called **epigenetic inheritance**.

- III. Transcription initiation is controlled by proteins that interact with DNA and with each other.
- A. Chromatin-modifying enzymes provide initial control of gene expression by making a region of DNA either more available or less available for transcription.
 - B. A cluster of proteins called a *transcription initiation complex* assembles on the promoter sequence at the “upstream” end of the gene.
 - C. One component, RNA polymerase II, transcribes the gene, synthesizing a primary RNA transcript or pre-mRNA.
 - D. Several **control elements** are associated with most eukaryotic genes.
 1. Control elements are noncoding DNA segments that regulate transcription by binding certain proteins.
 2. These control elements and the proteins they bind are critical to the precise regulation of gene expression in different cell types.
 - E. To initiate transcription, eukaryotic RNA polymerase requires the assistance of proteins called **transcription factors**.
 1. Only a few general transcription factors independently bind a DNA sequence such as the TATA box within the promoter.
 2. Others in the initiation complex are involved in protein-protein interactions, binding each other and RNA polymerase II.
 3. The interaction of general transcription factors and RNA polymerase II with a promoter usually leads to only a low rate of initiation and production of few RNA transcripts.
 4. In eukaryotes, high levels of transcription of particular genes depend on the interaction of control elements with *specific* transcription factors.
 5. Some control elements, named *proximal control elements*, are located close to the promoter.
 6. *Distant control elements*, **enhancers**, may be thousands of nucleotides away from the promoter or even downstream of the gene or within an intron.
 - a. A given gene may have several enhancers, each active at a different time or in a different cell type or location in the organism.
 - b. An **activator** is a protein that binds to an enhancer to stimulate transcription of a gene.
 7. Protein-mediated bending of DNA brings bound activators in contact with a group of *mediator proteins* that interact with proteins at the promoter. This helps assemble and position the initiation complex on the promoter.
 8. Eukaryotic genes also have **repressor** proteins to inhibit expression of a gene.
 - a. Eukaryotic repressors can cause inhibition of gene expression by blocking the binding of activators to their control elements or to components of the transcription machinery or by turning off transcription even in the presence of activators.
 9. Some activators and repressors act indirectly to influence chromatin structure.
 - a. Some activators recruit proteins that acetylate histones near the promoters of specific genes, promoting transcription.
 - b. Some repressors recruit proteins that deacetylate histones, reducing transcription or *silencing* the gene.
 - c. Recruitment of chromatin-modifying proteins seems to be the most common mechanism of repression in eukaryotes.
 10. For many genes, the particular *combination* of control elements associated with the gene may be more important than the presence of a single unique control element in regulating transcription of the gene. Even with only a dozen control element

sequences, a large number of combinations are possible.

- a. A particular combination of control elements will be able to activate transcription only when the appropriate activator proteins are present, such as at a precise time during development or in a particular cell type.
 - b. The use of different combinations of control elements allows fine regulation of transcription with a small set of control elements.
11. Remember that in prokaryotes, coordinately controlled genes are often clustered into an operon with a single promoter and other control elements upstream. The genes of the operon are transcribed into a single mRNA and translated together.
12. In contrast, very few eukaryotic genes are organized this way. Some coexpressed genes are clustered near each other on the same chromosome.
- a. Each eukaryotic gene in these clusters has its own promoter and is individually transcribed.
 - b. The coordinate regulation of clustered genes in eukaryotic cells is thought to involve changes in the chromatin structure that makes the entire group of genes either available or unavailable for transcription.
 - c. More commonly, genes coding for the enzymes of a metabolic pathway are scattered over different chromosomes.
 - d. Coordinate gene expression in eukaryotes depends on the association of a specific control element or combination of control elements with every gene of a dispersed group.
 - e. A common group of transcription factors binds to all the genes in the group, promoting simultaneous gene transcription.

IV. Post-transcriptional mechanisms play supporting roles in the control of gene expression.

- A. By using regulatory mechanisms that operate after transcription, a cell can rapidly fine-tune gene expression in response to environmental changes without altering its transcriptional patterns.
- B. RNA processing in the nucleus and the export of mRNA to the cytoplasm provide opportunities for gene regulation that are not available in bacteria.
- C. In **alternative RNA splicing**, different mRNA molecules are produced from the same primary transcript, depending on which RNA segments are treated as exons and which as introns.
 1. Regulatory proteins specific to a cell type control intron-exon choices by binding to regulatory sequences within the primary transcript.
- D. The life span of an mRNA molecule is an important factor in determining the pattern of protein synthesis.
 1. Prokaryotic mRNA molecules may be degraded after only a few minutes.
 2. Eukaryotic mRNAs typically last for hours, days, or weeks.
 3. A common pathway of mRNA breakdown begins with enzymatic shortening of the poly-A tail.
 - a. This triggers the enzymatic removal of the 5' cap.
 - b. This is followed by rapid degradation of the mRNA by nucleases.
 4. During the past few years, researchers have found small single-stranded RNA molecules called **microRNAs**, or **miRNAs**, that bind to complementary sequences in mRNA molecules.
 - a. miRNAs are formed from longer RNA precursors that fold back on themselves, forming a long hairpin structure stabilized by hydrogen bonding.
 - b. An enzyme called Dicer cuts the double-stranded RNA into short fragments.

- c. One of the two strands is degraded. The other miRNA strand associates with a protein complex and directs the complex to any mRNA molecules with a complementary sequence.
 - d. The miRNA-protein complex then degrades the target mRNA or blocks its translation.
 - e. The phenomenon of inhibition of gene expression by RNA molecules is called **RNA interference (RNAi)**.
 - f. **Small interfering RNAs (siRNAs)** are similar in size and function to miRNAs and are generated by similar mechanisms in eukaryotic cells.
 - g. Cellular RNAi pathways lead to the destruction of RNAs and may have originated as a natural defense against infection by RNA viruses.
- E. Translation of specific mRNAs can be blocked by regulatory proteins that bind to specific sequences or structures within the 5' leader region of mRNA. This prevents attachment of ribosomes.
- F. mRNAs may be stored in egg cells without poly-A tails of sufficient size to allow translation initiation.
- 1. At the appropriate time during development, a cytoplasmic enzyme adds more A residues, allowing translation to begin.
- G. Protein factors required to initiate translation in eukaryotes offer targets for simultaneously controlling translation of *all* mRNAs in a cell.
- 1. This allows the cell to shut down translation if environmental conditions are poor (for example, shortage of a key constituent) or until the appropriate conditions exist (for example, after fertilization in an egg or during daylight in plants).
- V. Cancer results from genetic changes that affect the cell cycle.
- A. Cancer is a disease in which cells escape the control methods that normally regulate cell growth and division.
- B. The gene regulation systems that go wrong during cancer are the very same systems that play important roles in embryonic development, the immune response, and other biological processes.
- C. The genes that normally regulate cell growth and division during the cell cycle include genes for growth factors, their receptors, and the intracellular molecules of signaling pathways. Mutations altering any of these genes in somatic cells can lead to cancer.
- D. The agent of such changes can be random spontaneous mutations or environmental influences such as chemical carcinogens, X-rays, or certain viruses.
- E. All tumor viruses transform cells into cancer cells through the integration of viral nucleic acid into host cell DNA.
- F. Cancer-causing genes, **oncogenes**, were initially discovered in retroviruses, but close counterparts, **proto-oncogenes**, have been found in other organisms.
- 1. The products of proto-oncogenes are proteins that stimulate normal cell growth and division and play essential functions in normal cells.
 - 2. A proto-oncogene becomes an oncogene following genetic changes that lead to an increase in the proto-oncogene's protein production or the activity of each protein molecule.
 - 3. These genetic changes include movements of DNA within the genome, amplification of the proto-oncogene, and point mutations in the control element of the proto-oncogene.
- G. Cancer cells frequently have chromosomes that have been broken and rejoined incorrectly.
- 1. This may translocate a fragment to a location near an active promoter or other control element.
 - 2. Movement of transposable elements may also place a more active promoter near a

proto-oncogene, increasing its expression.

- H. Amplification increases the number of copies of the proto-oncogene in the cell.
 - I. A point mutation in the promoter or enhancer of a proto-oncogene may increase its expression. A point mutation in the coding sequence may lead to translation of a protein that is more active or longer-lived.
 - J. Mutations to **tumor-suppressor genes**, whose normal products *inhibit* cell division, also contribute to cancer. Any decrease in the normal activity of a tumor-suppressor protein may contribute to cancer.
 - 1. Some tumor-suppressor proteins normally repair damaged DNA, preventing the accumulation of cancer-causing mutations.
 - 2. Others control the adhesion of cells to each other or to an extracellular matrix, crucial for normal tissues and often absent in cancers.
 - 3. Still others are components of cell-signaling pathways that inhibit the cell cycle.
- VI. Oncogene proteins and faulty tumor-suppressor proteins interfere with normal signaling pathways.
- A. The proteins encoded by many proto-oncogenes and tumor-suppressor genes are components of cell-signaling pathways.
 - B. Mutations in the products of two key genes, the *ras* proto-oncogene, and the *p53* tumor suppressor gene occur in 30% and 50% of human cancers, respectively.
 - C. Both the Ras protein and the p53 protein are components of signal-transduction pathways that convey external signals to the DNA in the cell's nucleus.
 - 1. Ras, the product of the ***ras* gene**, is a G protein that relays a growth signal from a growth factor receptor on the plasma membrane to a cascade of protein kinases.
 - 2. At the end of the pathway is the synthesis of a protein that stimulates the cell cycle.
 - 3. Many *ras* oncogenes have a point mutation that leads to a hyperactive version of the Ras protein that can issue signals on its own, resulting in excessive cell division.
 - D. The *p53* gene, named for its 53,000-dalton protein product, is often called the “guardian angel of the genome.”
 - 1. Damage to the cell's DNA acts as a signal that leads to expression of the *p53* gene.
 - 2. The p53 protein is a transcription factor for several genes.
 - 3. It can activate the *p21* gene, which halts the cell cycle.
 - 4. It can turn on genes involved in DNA repair.
 - 5. When DNA damage is irreparable, the p53 protein can activate “suicide genes” whose protein products cause cell death by apoptosis.
 - 6. A mutation that knocks out the *p53* gene can lead to excessive cell growth and cancer.
 - E. More than one somatic mutation is generally needed to produce the changes characteristic of a full-fledged cancer cell.
 - 1. About a half dozen DNA changes must occur for a cell to become fully cancerous.
 - 2. These usually include the appearance of at least one active oncogene and the mutation or loss of several tumor-suppressor genes.
 - a. Since mutant tumor-suppressor alleles are usually recessive, mutations must knock out *both* alleles.
 - b. Most oncogenes behave as dominant alleles and require only one mutation.
 - F. In many malignant tumors, the gene for telomerase is activated, removing a natural limit on the number of times the cell can divide.
 - G. Viruses promote cancer development by integrating their DNA into that of infected cells. By this process, a retrovirus may donate an oncogene to the cell.
 - 1. Alternatively, insertion of viral DNA may disrupt a tumor-suppressor gene or convert a proto-oncogene to an oncogene.

2. Some viruses produce proteins that inactivate p53 and other tumor-suppressor proteins, making the cell more prone to becoming cancerous.
- H. The fact that several genetic changes are required to produce a cancer cell helps explain the predispositions to cancer that run in some families.
1. An individual inheriting an oncogene or a mutant allele of a tumor-suppressor gene will be one step closer to accumulating the necessary mutations for cancer to develop.
- VII. Repetitive DNA and other noncoding sequences account for much of a eukaryotic genome.
- A. In eukaryotes, most of the DNA (98.5% in humans) does *not* code for protein or RNA.
 - B. Gene-related regulatory sequences and introns account for 24% of the human genome.
 - C. Most intergenic DNA is **repetitive DNA**, present in many copies in the genome.
 - D. Transposable elements and related sequences make up 44% of the entire human genome.
 1. Eukaryotic transposable elements are of two types: transposons, which move within a genome by means of a DNA intermediate, and retrotransposons, which move by means of an RNA intermediate, a transcript of the retrotransposon DNA.
 - a. Transposons can move by a “cut and paste” mechanism, which removes the element from its original site, or by a “copy and paste” mechanism, which leaves a copy behind.
 - b. Retrotransposons always leave a copy at the original site, since they are initially transcribed into an RNA intermediate.
 2. Most transposons are **retrotransposons**, in which the transcribed RNA includes the code for an enzyme that catalyzes the insertion of the retrotransposon and may include a gene for reverse transcriptase.
 - a. Reverse transcriptase uses the RNA molecule originally transcribed from the retrotransposon as a template to synthesize a double-stranded DNA copy.
 3. Copies of transposable elements and related sequences are scattered throughout eukaryotic genomes.
 - a. A single unit is hundreds or thousands of base pairs long, and the dispersed “copies” are similar but not identical to one another.
 - b. Some of the copies are transposable elements and some are related sequences that have lost the ability to move.
 4. In primates, a large portion of transposable element–related DNA consists of a family of similar sequences called *Alu elements*.
 - a. These sequences account for approximately 10% of the human genome.
 - b. *Alu* elements are about 300 nucleotides long, shorter than most functional transposable elements, and they do not code for protein.
 - c. Many *Alu* elements are transcribed into RNA molecules, however, their cellular function is unknown.
 5. Repetitive DNA that is not related to transposable elements probably arose by mistakes that occurred during DNA replication or recombination.
 - a. Repetitive DNA accounts for about 15% of the human genome.
 - b. Five percent of the human genome consists of large-segment duplications in which 10,000 to 300,000 nucleotide pairs seem to have been copied from one chromosomal location to another.
 - c. *Simple sequence DNA* contains many copies of tandemly repeated short sequences of 15–500 nucleotides.
 - (1) There may be as many as several hundred thousand repetitions of a nucleotide sequence.
 - (2) Simple sequence DNA makes up 3% of the human genome.

- (3) Much of the genome's simple sequence DNA is located at chromosomal telomeres and centromeres, suggesting that it plays a structural role.
 - (a) Telomeric DNA prevents gene loss as DNA shortens with each round of replication and also binds proteins that protect the ends of a chromosome from degradation or attachment to other chromosomes.

VIII. Gene families have evolved by duplication of ancestral genes.

- A. Sequences coding for proteins and structural RNAs compose a mere 1.5% of the human genome. If introns and regulatory sequences are included, gene-related DNA makes up 25% of the human genome.
- B. In humans, solitary genes present in one copy per haploid set of chromosomes make up only half of the total coding DNA.
- C. The rest occurs in **multigene families**, collections of identical or very similar genes.
- D. Some multigene families consist of *identical* DNA sequences that may be clustered tandemly. These code for RNA products or for histone proteins.
 1. For example, the three largest rRNA molecules are encoded in a single transcription unit that is repeated tandemly hundreds to thousands of times.
 2. This transcript is cleaved to yield three rRNA molecules that combine with proteins and one other kind of rRNA to form ribosomal subunits.
- E. Also found in some gene family clusters are several **pseudogenes**, DNA sequences similar to real genes that do not yield functional proteins. Random mutations accumulating over time in the pseudogenes have destroyed their function.

IX. Duplications, rearrangements, and mutations of DNA contribute to genome evolution.

- A. The earliest forms of life likely had a minimal number of genes, including only those necessary for survival and reproduction.
- B. The size of genomes has increased over evolutionary time, with the extra genetic material providing raw material for gene diversification.
- C. An accident in meiosis can result in one or more extra sets of chromosomes, a condition known as polyploidy.
 1. In a polyploid organism, one complete set of genes can provide essential functions for the organism.
 2. The genes in the extra set may diverge by accumulating mutations.
 3. These variations may persist if the organism carrying them survives and reproduces.
 4. In this way, genes with novel functions may evolve.
- D. Errors during meiosis due to unequal crossing over during Prophase I can lead to duplication of individual genes.
- E. Slippage during DNA replication can result in deletion or duplication of DNA regions. Such errors can lead to regions of repeats, such as simple sequence DNA.
 1. Duplication events can lead to the evolution of genes with related functions. The necessary function can be provided by one copy, while other copies of the gene accumulated random mutations.
 2. Some mutations may have altered the function of the protein product in ways that were beneficial to the organism without changing the function of the original gene.
- F. Rearrangement of existing DNA sequences has also contributed to genome evolution.
 1. The presence of introns in eukaryotic genes may have promoted the evolution of new and potentially useful proteins by facilitating the duplication or repositioning of exons in the genome.
 2. A particular exon within a gene could be duplicated on one chromosome and

- deleted from the homologous chromosome.
3. The gene with the duplicated exon would code for a protein with a second copy of the encoded domain.
 4. This change in the protein's structure could augment its function by increasing its stability or altering its ability to bind a particular ligand.
 5. Mixing and matching of different exons within or between genes owing to errors in meiotic recombination is called *exon shuffling* and could lead to new proteins with novel combinations of functions.
- G. The persistence of transposable elements as a large percentage of eukaryotic genomes suggests that they play an important role in shaping a genome over evolutionary time.
1. These elements can contribute to evolution of the genome by promoting recombination, disrupting cellular genes or control elements, and carrying entire genes or individual exons to new locations.
 2. The presence of homologous transposable element sequences scattered throughout the genome allows recombination to take place between different chromosomes.
 3. The movement of transposable elements around the genome can have several direct consequences.
 - a. If a transposable element "jumps" into the middle of a coding sequence of a protein-coding gene, it prevents the normal functioning of that gene.
 - b. If a transposable element inserts within a regulatory sequence, it may increase or decrease protein production.
 - c. During transposition, a transposable element may transfer genes to a new position on the genome or may insert an exon from one gene into another gene.
 - d. Transposable elements can lead to new coding sequences when an *Alu* element hops into introns to create a weak alternative splice site in the RNA transcript.
 - (1) Splicing will usually occur at the regular splice sites, producing the original protein. Occasionally, splicing will occur at the new weak site.
 - (2) In this way, alternative genetic combinations can be "tried out" while the function of the original gene product is retained.