

Chapter 19

Eukaryotic Genomes: Organization, Regulation, and Evolution

Lecture Outline

Overview

- Two features of eukaryotic genomes present a major information-processing challenge.
 - First, the typical multicellular eukaryotic genome is much larger than that of a prokaryotic cell.
 - Second, cell specialization limits the expression of many genes to specific cells.
- The estimated 25,000 genes in the human genome include an enormous amount of DNA that does not code for RNA or protein.
- This DNA is elaborately organized.
 - Not only is the DNA associated with protein, but also this DNA-protein complex called **chromatin** is organized into higher structural levels than the DNA-protein complex in prokaryotes.

A. The Structure of Eukaryotic Chromatin

1. Chromatin structure is based on successive levels of DNA packing.

- While the single circular chromosome of bacteria is coiled and looped in a complex but orderly manner, eukaryotic chromatin is far more complex.
- Eukaryotic DNA is precisely combined with large amounts of protein.
 - The resulting chromatin undergoes striking changes in the course of the cell cycle.
- During interphase of the cell cycle, chromatin fibers are usually highly extended within the nucleus.
- As a cell prepares for meiosis, its chromatin condenses, forming a characteristic number of short, thick chromosomes that can be distinguished with a light microscope.
- 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.
 - This chromosome and 45 other human chromosomes fit into the nucleus.
 - This occurs through an elaborate, multilevel system of DNA packing.
- **Histone** proteins are responsible for the first level of DNA packaging.
 - The mass of histone in chromatin is approximately equal to the mass of DNA.

- Their positively charged amino acids bind tightly to negatively charged DNA.
- 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.
- Unfolded chromatin has the appearance of beads on a string.
 - In this configuration, a chromatin fiber is 10 nm in diameter (*the 10-nm fiber*).
- Each bead of chromatin is a **nucleosome**, the basic unit of DNA packing.
 - The “string” between the beads is called *linker DNA*.
- A nucleosome consists of DNA wound around a protein core composed of two molecules each of four types of histone: H2A, H2B, H3, and H4.
 - The amino acid (N-terminus) of each histone protein (the *histone tail*) extends outward from the nucleosome.
 - A molecule of a fifth histone, H1, attaches to the DNA near the nucleosome.
- The beaded string seems to remain essentially intact throughout the cell cycle.
- Histones leave the DNA only transiently during DNA replication.
- They stay with the DNA during transcription.
 - By changing shape and position, nucleosomes allow RNA-synthesizing polymerases to move along the DNA.
- The next level of packing is due to the interactions between the histone tails of one nucleosome and the linker DNA and nucleosomes to either side.
 - With the aid of histone H1, these interactions cause the 10-nm to coil to form the *30-nm chromatin fiber*.
- This fiber forms *looped domains* attached to a scaffold of nonhistone proteins to make up a *300-nm fiber*.
- In a mitotic chromosome, the looped domains coil and fold to produce the characteristic metaphase chromosome.
- These packing steps are highly specific and precise, with particular genes located in the same places on metaphase chromosomes.
- Interphase chromatin is generally much less condensed than the chromatin of mitotic chromosomes, but it shows several of the same levels of higher-order packing.
 - Much of the chromatin is present as a 10-nm fiber, and some is compacted into a 30-nm fiber, which in some regions is folded into looped domains.
 - An interphase chromosome lacks an obvious scaffold, but its looped domains seem to be attached to the nuclear lamina on the inside of the nuclear envelope, and perhaps also to fibers of the nuclear matrix.
- The chromatin of each chromosome occupies a specific restricted area within the interphase nucleus.
- Interphase chromosomes have highly condensed areas, **heterochromatin**, and less compacted areas, **euchromatin**.
- Heterochromatin DNA is largely inaccessible to transcription enzymes.

- Looser packing of euchromatin makes its DNA accessible to enzymes and available for transcription.

B. The Control of Gene Expression

1. *Gene expression is regulated primarily at the transcription step.*

- Like unicellular organisms, the tens of thousands of genes in the cells of multicellular eukaryotes are continually turned on and off in response to signals from their internal and external environments.
- Gene expression must be controlled on a long-term basis during **cellular differentiation**, the divergence in form and function as cells in a multicellular organism specialize.
 - A typical human cell probably expresses about 20% of its genes at any given time.
 - Highly specialized cells, such as nerves or muscles, express only a tiny fraction of their genes.
 - Although all the cells in an organism contain an identical genome, the subset of genes expressed in the cells of each type is unique.
 - The differences between cell types are due to differential gene expression, the expression of different genes by cells with the same genome.
- The genomes of eukaryotes may contain tens of thousands of genes.
 - For quite a few species, only a small amount of the DNA—1.5% in humans—codes for protein.
 - Of the remaining DNA, a very small fraction consists of genes for rRNA and tRNA.
 - Most of the rest of the DNA seems to be largely noncoding, although researchers have found that a significant amount of it is transcribed into RNAs of unknown function.
- Problems with gene expression and control can lead to imbalance and diseases, including cancers.
- Our understanding of the mechanisms controlling gene expression in eukaryotes has been enhanced by new research methods, including advances in DNA technology.
- In all organisms, the expression of specific genes is most commonly regulated at transcription, often in response to signals coming from outside the cell.
 - The term *gene expression* is often equated with transcription.
 - With their greater complexity, eukaryotes have opportunities for controlling gene expression at additional stages.
- Each stage in the entire process of gene expression provides a potential control point where gene expression can be turned on or off, sped up or slowed down.
 - A web of control connects *different* genes and their products.
- These levels of control include chromatin packing, transcription, RNA processing, translation, and various alterations to the protein product.

2. *Chromatin modifications affect the availability of genes for transcription.*

- In addition to its role in packing DNA inside the nucleus, chromatin organization regulates gene expression.
 - Genes of densely condensed heterochromatin are usually not expressed, presumably because transcription proteins cannot reach the DNA.

- A gene's location relative to nucleosomes and to attachment sites to the chromosome scaffold or nuclear lamina can affect transcription.
- Chemical modifications of chromatin play a key role in chromatin structure and gene expression.
- Chemical modifications of histones play a direct role in the regulation of gene transcription.
- The N-terminus of each histone molecule in a nucleosome protrudes outward from the nucleosome.
 - These histone tails are accessible to various modifying enzymes, which catalyze the addition or removal of specific chemical groups.
- **Histone acetylation** (addition of an acetyl group —COCH₃) and deacetylation appear to play a direct role in the regulation of gene transcription.
 - Acetylated histones grip DNA less tightly, providing easier access for transcription proteins in this region.
 - Some of the enzymes responsible for acetylation or deacetylation are associated with or are components of transcription factors that bind to promoters.
 - 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.
- **DNA methylation** is the attachment by specific enzymes of methyl groups (—CH₃) to DNA bases after DNA synthesis.
- Inactive DNA is generally highly methylated compared to DNA that is actively transcribed.
 - For example, the inactivated mammalian X chromosome in females is heavily methylated.
- Genes are usually more heavily methylated in cells where they are not expressed.
 - Demethylating certain inactive genes turns them on.
 - However, there are exceptions to this pattern.
- DNA methylation proteins recruit histone deacetylation enzymes, providing a mechanism by which DNA methylation and histone deacetylation cooperate to repress transcription.
- In some species, DNA methylation is responsible for long-term inactivation of genes during cellular differentiation.
 - Once methylated, genes usually stay that way through successive cell divisions.
 - Methylation enzymes recognize sites on one strand that are already methylated and correctly methylate the daughter strand after each round of DNA replication.
- 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.
- The chromatin modifications just discussed do not alter DNA sequence, and yet they may be passed along to future generations of cells.
 - Inheritance of traits by mechanisms not directly involving the nucleotide sequence is called **epigenetic inheritance**.
- Researchers are amassing more and more evidence for the importance of epigenetic information in the regulation of gene expression.
 - Enzymes that modify chromatin structure are integral parts of the cell's machinery for regulating transcription.

3. Transcription initiation is controlled by proteins that interact with DNA and with each other.

- Chromatin-modifying enzymes provide initial control of gene expression by making a region of DNA either more available or less available for transcription.
- A cluster of proteins called a *transcription initiation complex* assembles on the promoter sequence at the “upstream” end of the gene.
 - One component, RNA polymerase II, transcribes the gene, synthesizing a primary RNA transcript or pre-mRNA.
 - RNA processing includes enzymatic addition of a 5' cap and a poly-A tail, as well as splicing out of introns to yield a mature mRNA.
- Multiple **control elements** are associated with most eukaryotic genes.
 - Control elements are noncoding DNA segments that regulate transcription by binding certain proteins.
 - These control elements and the proteins they bind are critical to the precise regulation of gene expression in different cell types.
- To initiate transcription, eukaryotic RNA polymerase requires the assistance of proteins called **transcription factors**.
 - *General* transcription factors are essential for the transcription of *all* protein-coding genes.
 - Only a few general transcription factors independently bind a DNA sequence such as the TATA box within the promoter.
 - Others in the initiation complex are involved in protein-protein interactions, binding each other and RNA polymerase II.
- 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.
- In eukaryotes, high levels of transcription of particular genes depend on the interaction of control elements with *specific* transcription factors.
- Some control elements, named *proximal control elements*, are located close to the promoter.
- *Distant control elements*, **enhancers**, may be thousands of nucleotides away from the promoter or even downstream of the gene or within an intron.
- A given gene may have multiple enhancers, each active at a different time or in a different cell type or location in the organism.
- An **activator** is a protein that binds to an enhancer to stimulate transcription of a gene.
 - 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.
- Eukaryotic genes also have **repressor** proteins to inhibit expression of a gene.
 - 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.
- Some activators and repressors act indirectly to influence chromatin structure.
 - Some activators recruit proteins that acetylate histones near the promoters of specific genes, promoting transcription.
 - Some repressors recruit proteins that deacetylate histones, reducing transcription or *silencing* the gene.

- Recruitment of chromatin-modifying proteins seems to be the most common mechanism of repression in eukaryotes.
- The number of nucleotide sequences found in control elements is surprisingly small.
- 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 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.
 - The use of different combinations of control elements allows fine regulation of transcription with a small set of control elements.
- 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.
- In contrast, very few eukaryotic genes are organized this way.
- Recent studies of the genomes of several eukaryotic species have found that some coexpressed genes are clustered near each other on the same chromosome.
 - Each eukaryotic gene in these clusters has its own promoter and is individually transcribed.
 - 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.
- More commonly, genes coding for the enzymes of a metabolic pathway are scattered over different chromosomes.
- 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.
- A common group of transcription factors binds to all the genes in the group, promoting simultaneous gene transcription.
 - For example, a steroid hormone enters a cell and binds to a specific receptor protein in the cytoplasm or nucleus, forming a hormone-receptor complex that serves as a transcription activator.
 - Every gene whose transcription is stimulated by that steroid hormone has a control element recognized by that hormone-receptor complex.
 - Other signal molecules control gene expression indirectly by triggering signal-transduction pathways that lead to activation of transcription.
- Systems for coordinating gene regulation probably arose early in evolutionary history and evolved by the duplication and distribution of control elements within the genome.

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

- Gene expression may be blocked or stimulated by any posttranscriptional step.

- 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.
- RNA processing in the nucleus and the export of mRNA to the cytoplasm provide opportunities for gene regulation that are not available in bacteria.
- 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.
 - Regulatory proteins specific to a cell type control intron-exon choices by binding to regulatory sequences within the primary transcript.
- The life span of an mRNA molecule is an important factor in determining the pattern of protein synthesis.
- Prokaryotic mRNA molecules may be degraded after only a few minutes.
- Eukaryotic mRNAs typically last for hours, days, or weeks.
 - In red blood cells, mRNAs for hemoglobin polypeptides are unusually stable and are translated repeatedly.
- A common pathway of mRNA breakdown begins with enzymatic shortening of the poly-A tail.
 - This triggers the enzymatic removal of the 5' cap.
 - This is followed by rapid degradation of the mRNA by nucleases.
- Nucleotide sequences in the untranslated trailer region at the 3' end affect mRNA stability.
 - Transferring such a sequence from a short-lived mRNA to a normally stable mRNA results in quick mRNA degradation.
- 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.
 - miRNAs are formed from longer RNA precursors that fold back on themselves, forming a long hairpin structure stabilized by hydrogen bonding.
 - An enzyme called Dicer cuts the double-stranded RNA into short fragments.
 - 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.
 - The miRNA-protein complex then degrades the target mRNA or blocks its translation.
- The phenomenon of inhibition of gene expression by RNA molecules is called **RNA interference (RNAi)**.
 - **Small interfering RNAs (siRNAs)** are similar in size and function to miRNAs and are generated by similar mechanisms in eukaryotic cells.
- Cellular RNAi pathways lead to the destruction of RNAs and may have originated as a natural defense against infection by RNA viruses.
 - Whatever their origin, RNAi plays an important role in regulating gene expression in the cell.
- 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.

- mRNAs may be stored in egg cells without poly-A tails of sufficient size to allow translation initiation.
 - At the appropriate time during development, a cytoplasmic enzyme adds more A residues, allowing translation to begin.
- Protein factors required to initiate translation in eukaryotes offer targets for simultaneously controlling translation of *all* mRNAs in a cell.
 - 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).
- Finally, eukaryotic polypeptides must often be processed to yield functional proteins.
 - This may include cleavage, chemical modifications, and transport to the appropriate destination.
- The cell limits the lifetimes of normal proteins by selective degradation.
 - Many proteins, like the cyclins in the cell cycle, must be short-lived to function appropriately.
- Proteins intended for degradation are marked by the attachment of ubiquitin proteins.
- Giant protein complexes called **proteasomes** recognize the ubiquitin and degrade the tagged protein.
 - Mutations making cell cycle proteins impervious to proteasome degradation can lead to cancer.

C. The Molecular Biology of Cancer

1. Cancer results from genetic changes that affect the cell cycle.

- Cancer is a disease in which cells escape the control methods that normally regulate cell growth and division.
 - 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.
- 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.
- The agent of such changes can be random spontaneous mutations or environmental influences such as chemical carcinogens, X-rays, or certain viruses.
- In 1911, Peyton Rous discovered a virus that causes cancer in chickens.
 - Since then, scientists have recognized a number of *tumor viruses* that cause cancer in various animals, including humans.
 - All tumor viruses transform cells into cancer cells through the integration of viral nucleic acid into host cell DNA.
- Cancer-causing genes, **oncogenes**, were initially discovered in retroviruses, but close counterparts, **proto-oncogenes**, have been found in other organisms.
- The products of proto-oncogenes are proteins that stimulate normal cell growth and division and play essential functions in normal cells.

- 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.
 - 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.
- Cancer cells frequently have chromosomes that have been broken and rejoined incorrectly.
 - This may translocate a fragment to a location near an active promoter or other control element.
 - Movement of transposable elements may also place a more active promoter near a proto-oncogene, increasing its expression.
- Amplification increases the number of copies of the proto-oncogene in the cell.
- 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.
- 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.
 - Some tumor-suppressor proteins normally repair damaged DNA, preventing the accumulation of cancer-causing mutations.
 - Others control the adhesion of cells to each other or to an extracellular matrix, crucial for normal tissues and often absent in cancers.
 - Still others are components of cell-signaling pathways that inhibit the cell cycle.

2. Oncogene proteins and faulty tumor-suppressor proteins interfere with normal signaling pathways.

- The proteins encoded by many proto-oncogenes and tumor-suppressor genes are components of cell-signaling pathways.
- 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.
- 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.
- 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.
 - At the end of the pathway is the synthesis of a protein that stimulates the cell cycle.
 - 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.
- The *p53* gene, named for its 53,000-dalton protein product, is often called the “guardian angel of the genome.”
- Damage to the cell's DNA acts as a signal that leads to expression of the *p53* gene.
- The p53 protein is a transcription factor for several genes.
 - It can activate the *p21* gene, which halts the cell cycle.
 - It can turn on genes involved in DNA repair.

- When DNA damage is irreparable, the p53 protein can activate “suicide genes” whose protein products cause cell death by apoptosis.
- A mutation that knocks out the *p53* gene can lead to excessive cell growth and cancer.

3. *Multiple mutations underlie the development of cancer.*

- More than one somatic mutation is generally needed to produce the changes characteristic of a full-fledged cancer cell.
- If cancer results from an accumulation of mutations, and if mutations occur throughout life, then the longer we live, the more likely we are to develop cancer.
- Colorectal cancer, with 135,000 new cases and 60,000 deaths in the United States each year, illustrates a multistep cancer path.
- The first sign is often a polyp, a small benign growth in the colon lining.
 - The cells of the polyp look normal but divide unusually frequently.
- Through gradual accumulation of mutations that activate oncogenes and knock out tumor-suppressor genes, the polyp can develop into a malignant tumor.
- About a half dozen DNA changes must occur for a cell to become fully cancerous.
- These usually include the appearance of at least one active oncogene and the mutation or loss of several tumor-suppressor genes.
 - Since mutant tumor-suppressor alleles are usually recessive, mutations must knock out *both* alleles.
 - Most oncogenes behave as dominant alleles and require only one mutation.
- In many malignant tumors, the gene for telomerase is activated, removing a natural limit on the number of times the cell can divide.
- Viruses, especially retroviruses, play a role in about 15% of human cancer cases worldwide.
 - These include some types of leukemia, liver cancer, and cancer of the cervix.
- 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.
- Alternatively, insertion of viral DNA may disrupt a tumor-suppressor gene or convert a proto-oncogene to an oncogene.
- Some viruses produce proteins that inactivate p53 and other tumor-suppressor proteins, making the cell more prone to becoming cancerous.
- The fact that multiple genetic changes are required to produce a cancer cell helps explain the predispositions to cancer that run in some families.
 - 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.
- Geneticists are devoting much effort to finding inherited cancer alleles so that predisposition to certain cancers can be detected early in life.
 - About 15% of colorectal cancers involve inherited mutations, especially to DNA repair genes or to the tumor-suppressor gene *adenomatous polyposis coli*, or *APC*.
 - Normal functions of the *APC* gene include regulation of cell migration and adhesion.
 - Even in patients with no family history of the disease, *APC* is mutated in about 60% of colorectal cancers.

- Between 5–10% of breast cancer cases show an inherited predisposition.
 - This is the second most common type of cancer in the United States, striking more than 180,000 women annually and leading to 40,000 annual deaths.
 - Mutations to one of two tumor-suppressor genes, *BRCA1* and *BRCA2*, increase the risk of breast and ovarian cancer.
- A woman who inherits one mutant *BRCA1* allele has a 60% probability of developing breast cancer before age 50 (versus a 2% probability in an individual with two normal alleles).
- *BRCA1* and *BRCA2* are considered tumor-suppressor genes because their wild-type alleles protect against breast cancer and because their mutant alleles are recessive.
- Recent evidence suggests that the BRCA2 protein is directly involved in repairing breaks that occur in both strands of DNA.

D. Genome Organization at the DNA Level

1. Repetitive DNA and other noncoding sequences account for much of a eukaryotic genome.

- Several trends are evident when we compare the genomes of prokaryotes to those of eukaryotes.
- There is a general trend from smaller to larger genomes, but with fewer genes in a given length of DNA.
 - Humans have 500 to 1,500 times as many base pairs in their genome as most prokaryotes, but only 5 to 15 times as many genes.
- Most of the DNA in a prokaryote genome codes for protein, tRNA, or rRNA.
 - The small amount of noncoding DNA consists mainly of regulatory sequences.
- In eukaryotes, most of the DNA (98.5% in humans) does *not* code for protein or RNA.
 - Gene-related regulatory sequences and introns account for 24% of the human genome.
 - Introns account for most of the difference in average length of eukaryotic (27,000 base pairs) and prokaryotic genes (1,000 base pairs).
 - Most intergenic DNA is **repetitive DNA**, present in multiple copies in the genome.
 - Transposable elements and related sequences make up 44% of the entire human genome.
- The first evidence for transposable elements came from geneticist Barbara McClintock's breeding experiments with Indian corn (maize) in the 1940s and 1950s.
- 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.
 - 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.
 - Retrotransposons always leave a copy at the original site, since they are initially transcribed into an RNA intermediate.
- 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.
 - Reverse transcriptase uses the RNA molecule originally transcribed from the retrotransposon as a template to synthesize a double-stranded DNA copy.

- Multiple copies of transposable elements and related sequences are scattered throughout eukaryotic genomes.
 - A single unit is hundreds or thousands of base pairs long, and the dispersed “copies” are similar but not identical to one another.
 - Some of the copies are transposable elements and some are related sequences that have lost the ability to move.
 - Transposable elements and related sequences make up 25–50% of most mammalian genomes, and an even higher percentage in amphibians and angiosperms.
- In primates, a large portion of transposable element–related DNA consists of a family of similar sequences called *Alu elements*.
 - These sequences account for approximately 10% of the human genome.
 - *Alu* elements are about 300 nucleotides long, shorter than most functional transposable elements, and they do not code for protein.
 - Many *Alu* elements are transcribed into RNA molecules.
 - However, their cellular function is unknown.
- Repetitive DNA that is not related to transposable elements probably arose by mistakes that occurred during DNA replication or recombination.
 - Repetitive DNA accounts for about 15% of the human genome.
 - 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.
- *Simple sequence DNA* contains many copies of tandemly repeated short sequences of 15–500 nucleotides.
 - There may be as many as several hundred thousand repetitions of a nucleotide sequence.
 - Simple sequence DNA makes up 3% of the human genome.
 - Much of the genome’s simple sequence DNA is located at chromosomal telomeres and centromeres, suggesting that it plays a structural role.
 - The DNA at centromeres is essential for the separation of chromatids in cell division and may also help to organize the chromatin within the interphase nucleus.
 - 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.

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

- 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.
- In humans, solitary genes present in one copy per haploid set of chromosomes make up only half of the total coding DNA.
- The rest occurs in **multigene families**, collections of identical or very similar genes.
- Some multigene families consist of *identical* DNA sequences that may be clustered tandemly.
 - These code for RNA products or for histone proteins.
 - For example, the three largest rRNA molecules are encoded in a single transcription unit that is repeated tandemly hundreds to thousands of times.

- This transcript is cleaved to yield three rRNA molecules that combine with proteins and one other kind of rRNA to form ribosomal subunits.
- Two related families of *nonidentical genes* encode globins, a group of proteins that include the α (alpha) and β (beta) polypeptide sequences of hemoglobin.
- The different versions of each globin subunit are expressed at different times in development, allowing hemoglobin to function effectively in the changing environment of the developing animal.
 - Within both the α and β families are sequences that are expressed during the embryonic, fetal, and/or adult stage of development.
 - In humans, the embryonic and fetal hemoglobins have higher affinity for oxygen than do adult forms, ensuring transfer of oxygen from mother to developing fetus.
 - Also found in the globin gene family clusters are several **pseudogenes**, DNA sequences similar to real genes that do not yield functional proteins.

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

- The earliest forms of life likely had a minimal number of genes, including only those necessary for survival and reproduction.
- The size of genomes has increased over evolutionary time, with the extra genetic material providing raw material for gene diversification.
- An accident in meiosis can result in one or more extra sets of chromosomes, a condition known as polyploidy.
 - In a polyploid organism, one complete set of genes can provide essential functions for the organism.
 - The genes in the extra set may diverge by accumulating mutations.
 - These variations may persist if the organism carrying them survives and reproduces.
 - In this way, genes with novel functions may evolve.
- Errors during meiosis due to unequal crossing over during Prophase I can lead to duplication of individual genes.
- 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.
- Major rearrangements of at least one set of genes occur during immune system differentiation.
- Duplication events can lead to the evolution of genes with related functions, such as the α -globin and β -globin gene families.
 - A comparison of gene sequences within a multigene family indicates that they all evolved from one common ancestral globin gene, which was duplicated and diverged about 450–500 million years ago.
- After the duplication events, the differences between the genes in the globin family arose from mutations that accumulated in the gene copies over many generations.
 - The necessary function provided by an α -globin protein was fulfilled by one gene, while other copies of the α -globin gene accumulated random mutations.
 - Some mutations may have altered the function of the protein product in ways that were beneficial to the organism without changing its oxygen-carrying function.
- The similarity in the amino acid sequences of the various α -globin and β -globin proteins supports this model of gene duplication and mutation.

- Random mutations accumulating over time in the pseudogenes have destroyed their function.
- In other gene families, one copy of a duplicated gene can undergo alterations that lead to a completely new function for the protein product.
- The genes for lysozyme and α -lactalbumin are good examples.
 - Lysozyme is an enzyme that helps prevent infection by hydrolyzing bacterial cell walls.
 - α -lactalbumin is a nonenzymatic protein that plays a role in mammalian milk production.
- Both genes are found in mammals, while only lysozyme is found in birds.
 - The two proteins are similar in their amino acids sequences and 3-D structures.
- These findings suggest that at some time after the bird and mammalian lineage had separated, the lysozyme gene underwent a duplication event in the mammalian lineage but not in the avian lineage.
 - Subsequently, one copy of the duplicated lysozyme gene evolved into a gene encoding α -lactalbumin, a protein with a completely different function.
- Rearrangement of existing DNA sequences has also contributed to genome evolution.
 - 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.
 - A particular exon within a gene could be duplicated on one chromosome and deleted from the homologous chromosome.
 - The gene with the duplicated exon would code for a protein with a second copy of the encoded domain.
- This change in the protein's structure could augment its function by increasing its stability or altering its ability to bind a particular ligand.
- 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.
- 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.
- 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.
- The presence of homologous transposable element sequences scattered throughout the genome allows recombination to take place between different chromosomes.
 - Most of these alterations are likely detrimental, causing chromosomal translocations and other changes in the genome that may be lethal to the organism.
 - Over the course of evolutionary time, an occasional recombination may be advantageous.
 - The movement of transposable elements around the genome can have several direct consequences.
 - * 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.
 - * If a transposable element inserts within a regulatory sequence, it may increase or decrease protein production.

- 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.
- 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.
 - * Splicing will usually occur at the regular splice sites, producing the original protein.
 - * Occasionally, splicing will occur at the new weak site.
- In this way, alternative genetic combinations can be “tried out” while the function of the original gene product is retained.
- These processes produce no effect or harmful effects in most individual cases.
- However, over long periods of time, the generation of genetic diversity provides more raw material for natural selection to work on during evolution.