

CHAPTER 19

THE ORGANIZATION AND CONTROL OF EUKARYOTIC GENOMES

I. □ Introduction

- A. Gene expression in eukaryotes has two main differences from the same process in prokaryotes.
 - 1. First, the typical multicellular eukaryotic genome is much larger than that of a bacterium.
 - 2. Second, cell specialization limits the expression of many genes to specific cells.
- B. The estimated 35,000 genes in the human genome includes an enormous amount of DNA that does not program the synthesis of RNA or protein.
- C. This DNA is elaborately organized.
 - 1. Not only is the DNA associated with protein to form chromatin, but the chromatin is organized into higher organizational levels.
- D. Level of packing is one way that gene expression is regulated.
 - 1. Densely packed areas are inactivated.
 - 2. Loosely packed areas are being actively transcribed.

II. Eukaryotic Chromatin Structure

- A. Chromatin structure is based on successive levels of DNA packing
 - 1. While the single circular chromosome of bacteria is coiled and looped in a complex, but orderly manner, eukaryotic chromatin is far more complex.
 - 2. Eukaryotic DNA is precisely combined with large amounts of protein.
 - 3. During interphase of the cell cycle, chromatin fibers are usually highly extended within the nucleus.
 - 4. Eukaryotic chromosomes contain an enormous amount of DNA relative to their condensed length.
 - a) *Each human chromosome averages about 2×10^8 nucleotide pairs.*
 - b) *If extended, each DNA molecule would be about 6 cm long, thousands of times longer than the cell diameter.*
 - c) *This chromosome and 45 other human chromosomes fit into the nucleus.*
 - d) *This occurs through an elaborate, multilevel system of DNA packing.*
 - 5. **Histone** proteins are responsible for the first level of DNA packaging.
 - a) *Their positively charged amino acids bind tightly to negatively charged DNA.*
 - b) *The five types of histones are very similar from one eukaryote to another and are even present in bacteria.*
 - 6. Unfolded chromatin has the appearance of beads on a string, a **nucleosome**, in which DNA winds around a core of histone proteins.
 - 7. The beaded string seems to remain essentially intact throughout the cell cycle.
 - 8. Histones leave the DNA only transiently during DNA replication.
 - 9. They stay with the DNA during transcription.
 - a) *By changing shape and position, nucleosomes allow RNA-synthesizing polymerases to move along the DNA.*
 - 10. As chromosomes enter mitosis the beaded string undergoes higher-order packing.
 - 11. The beaded string coils to form the 30-nm chromatin fiber.
 - 12. This fiber forms *looped domains* attached to a scaffold of nonhistone proteins.
 - 13. In a mitotic chromosome, the looped domains coil and fold to produce the characteristic metaphase chromosome.
 - 14. These packing steps are highly specific and precise with particular genes located in the same places.

- a) *Interphase chromatin is generally much less condensed than the chromatin of mitosis.*
- b) *While the 30-nm fibers and looped domains remain, the discrete scaffold is not present.*
- c) *The looped domains appear to be attached to the nuclear lamina and perhaps the nuclear matrix.*
- d) *The chromatin of each chromosome occupies a restricted area within the interphase nucleus.*
- e) *Interphase chromosomes have areas that remain highly condensed, **heterochromatin**, and less compacted areas, **euchromatin**.*

III. Genome Organization at the DNA Level

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

1. In prokaryotes, most of the DNA in a genome codes for protein (or tRNA and rRNA), with a small amount of noncoding DNA, primarily regulators.
2. In eukaryotes, most of the DNA (about 97% in humans) does *not* code for protein or RNA.
 - a) *Some noncoding regions are regulatory sequences.*
 - b) *Other are introns.*
 - c) *Finally, even more of it consists of **repetitive DNA**, present in many copies in the genome.*
3. In mammals about 10 -15% of the genome is *tandemly repetitive DNA*, or **satellite DNA**.
 - a) *These differ in intrinsic density from other regions, leading them to form a separate band after differential ultracentrifugation.*
4. These sequences (up to 10 base pairs) are repeated up to several hundred thousand times in series.
5. There are three types of satellite DNA, differentiated by the total length of DNA at each site.
 - a) *A number of genetic disorders are caused by abnormally long stretches of tandemly repeated nucleotide triplets within the affected gene.*
 - b) *Fragile X syndrome is caused by hundreds to thousands of repeats of CCG in the leader sequence of the fragile X gene.*
Problems at this site lead to mental retardation.
 - c) *Huntington's disease, another neurological syndrome, occurs due to repeats of CAG that are translated into a proteins with a long string of glutamines.*
 - d) *The severity of the disease and the age of onset of these diseases are correlated with the number of repeats.*
Much of the satellite DNA appears to play a structural role at telomeres and centromeres.
 - e) *The DNA at the centromeres is essential for the separation of sister chromatids during cell division and may help organize the chromatin within the nucleus.*
 - f) *The telomeres protect genes from being lost as the DNA shortens with each round of replication and they bind to proteins that protect the ends of chromosomes from degradation and fusion with other chromosomes.*
Artificial chromosomes, each consisting of an origin of replication site, a centromere, and two telomeres, will be replicated and distributed to daughter cells if inserted into a cell.
6. About 25-40% of most mammalian genomes consists of *interspersed repetitive DNA*.
 - a) *Sequences hundreds to thousands of base pairs long appear at multiple sites in the genome.*
 - b) *The "dispersed" copies are similar but usually not identical to each other.*
 - c) *One common family of interspersed repetitive sequences, **Alu elements**, is transcribed into RNA molecules with unknown roles in the cell role.*

B. Gene families have evolved by duplication of ancestral genes

1. While most genes are present as a single copy per haploid set of chromosomes, **multigene families** exist as a collection of identical or very similar genes.
2. These likely evolved from a single ancestral gene.
3. The members of multigene families may be clustered or dispersed in the genome.
4. *Identical genes* are multigene families that are clustered tandemly.
5. These usually consist of the genes for RNA products or those for histone proteins.
6. For example, the three largest rRNA molecules are encoded in a single transcription unit that is repeated tandemly hundreds to thousands of times.

7. This transcript is cleaved to yield three rRNA molecules that combine with proteins and one other kind of rRNA to form ribosomal subunits.
8. Nonidentical genes have diverged since their initial duplication event.
9. The classic example traces the duplication and diversification of the two related families of globin genes, α (alpha) and β (beta), of hemoglobin.
10. The α subunit family is on human chromosome 16 and the β subunit family is on chromosome 11.
11. The different versions of each globin subunit are expressed at different times in development, fine-tuning function to changes in environment.
12. Within both the α and β families are sequences that are expressed during the embryonic, fetal, and/or adult stage of development.
13. The embryonic and fetal hemoglobins have higher affinity for oxygen than do adult forms, ensuring transfer of oxygen from mother to developing fetus.
14. These gene families probably arise by repeated gene duplications that occur as errors during DNA replication and recombination.
15. The differences in genes arise from mutations that accumulate in the gene copies over generations.
16. These mutations may even lead to enough changes to form **pseudogenes**, DNA segments that have sequences similar to real genes but that do not yield functional proteins.
17. The locations of the two globin families on different chromosomes indicate that they (and certain other gene families) probably arose by transposition.

C. Gene amplification, loss, or rearrangement can alter a cell's genome during an organism's lifetime

1. In addition to rare mutations, the nucleotide sequence of an organism's genome may be altered in a systematic way during its lifetime.
2. Because these changes do not affect gametes, they are not passed on to offspring and their effects are confined to particular cells and tissues.
 - a) In **gene amplification**, certain genes are replicated as a way to increase expression of these genes.
 - b) In amphibians, the genes for rRNA not only have a normal complement of multiple copies but millions of additional copies are synthesized in a developing ovum.
 - c) This assists the cell in producing enormous numbers of ribosomes for protein synthesis after fertilization.
 - d) These extra copies exist as separate circles of DNA in nucleoli and are degraded when no longer needed.
 - e) Amplification of genes for drug resistance is often seen in surviving cancer cells after chemotherapy.
 - f) In some insect cells, whole or parts of chromosomes are lost early in development.
3. Rearrangement of the loci of genes in somatic cells may have a powerful effect on gene expression.
4. Transposons are genes that can move from one location to another within the genome.
 - a) Up to 50% of the corn genome and 10% of the human genome are transposons.
 - b) If one "jumps" into a coding sequence of another gene, it can prevent normal gene function as seen in the pigment of this morning glory leaf.
 - c) If the transposon is inserted in a regulatory area, it may increase or decrease transcription.
 - d) 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.
 - e) Reverse transcriptase uses the RNA molecule originally transcribed from the retrotransposon as a template to synthesize a double stranded DNA copy.
5. Retrotransposons facilitate replicative transposition, populating the eukaryotic genome with multiple copies of its sequence.
 - a) Retrovirus may have evolved from escaped and packaged retrotransposons.
 - b) Even retrotransposons that lack the gene for reverse transcriptase (like Alu elements) can be moved using enzymes encoded for by other retrotransposons in the genome.
6. Major rearrangements of at least one set of genes occur during immune system differentiation.
7. B lymphocytes produce **immunoglobins**, or antibodies, that specifically recognize and combat viruses, bacteria, and other invaders.

- a) *Each differentiated cell and its descendants produce one specific type of antibody that attacks a specific invader.*
 - b) *As an unspecialized cell differentiates into a B lymphocyte, functional antibody genes are pieced together from physically separated DNA regions.*
8. Each immunoglobulin consists of four polypeptide chains, each with a constant region and a variable region, giving each antibody its unique function.
 - a) *As a B lymphocyte differentiates, one of several hundred possible variable segments is connected to the constant section by deleting the intervening DNA.*
 9. The random combinations of different variable and constant regions create an enormous variety of different polypeptides.
 10. Further, these individual polypeptides combine with others to form complete antibody molecules.
 11. As a result, the mature immune system can make millions of different kinds of antibodies from millions of subpopulations of B lymphocytes.

IV. The Control of Gene Expression

A. Each cell of a multicellular eukarote expresses only a small fraction of its genes

1. 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.
2. Gene expression must be controlled on a long-term basis during **cellular differentiation**, the divergence in form and function as cells specialize.
 - a) *Highly specialized cells, like nerves or muscles, express only a tiny fraction of their genes.*
 - b) *Problems with gene expression and control can lead to imbalance and diseases, including cancers.*
 - c) *Our understanding of the mechanisms controlling gene expression in eukaryotes has been enhanced by new research methods and technologies.*
 - d) *Control of gene activity in eukaryotes involves some of the principles described for prokaryotes.*
 - e) *The expression of specific genes is commonly regulated at the transcription level by DNA-binding proteins that interact with other proteins and with external signals.*
 - f) *With their greater complexity, eukaryotes have opportunities for controlling gene expression at additional stages.*

B. The control of gene expression can occur at any step in the pathway from gene to functional protein: an overview

1. Each stage in the entire process of gene expression provides a potential control point where gene expression can be turned on or off, speeded up or slowed down.
2. A web of control connects *different* genes and their products.
3. These levels of control include chromatin packing, transcription, RNA processing, translation, and various alterations to the protein product.

C. Chromatin modifications affect the availability of genes for transcription

- a) *In addition to its role in packing DNA inside the nucleus, chromatin organization impacts regulation.*
 - b) *Genes of densely condensed heterochromatin are usually not expressed, presumably because transcription proteins cannot reach the DNA.*
 - c) *A gene's location relative to nucleosomes and to attachment sites to the chromosome scaffold or nuclear lamina can affect transcription.*
 - d) *Chemical modifications of chromatin play a key role in chromatin structure and transcription regulation.*
2. **DNA methylation** is the attachment by specific enzymes of methyl groups (-CH₃) to DNA bases after DNA synthesis.
 3. Inactive DNA is generally highly methylated compared to DNA that is actively transcribed.

- a) *For example, the inactivated mammalian X chromosome in females is heavily methylated.*
 - b) *Genes are usually more heavily methylated in cells where they are not expressed.*
 - c) *Demethylating certain inactive genes turns them on.*
 - d) *However, there are exceptions to this pattern.*
4. In some species DNA methylation is responsible for long-term inactivation of genes during cellular differentiation.
 - a) *Once methylated, genes usually stay that way through successive cell divisions.*
 - b) *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.
 6. **Histone acetylation** (addition of an acetyl group $-\text{COCH}_3$) and deacetylation appear to play a direct role in the regulation of gene transcription.
 - a) *Acetylated histones grip DNA less tightly, providing easier access for transcription proteins in this region.*
 - b) *Some of the enzymes responsible for acetylation or deacetylation are associated with or are components of transcription factors that bind to promoters.*
 - c) *In addition, some DNA methylation proteins recruit histone deacetylation enzymes, providing a mechanism by which DNA methylation and histone deacetylation cooperate to repress transcription.*

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

1. Chromatin-modifying enzymes provide a coarse adjustment to gene expression by making a region of DNA either more available or less available for transcription.
2. Fine-tuning begins with the interaction of transcription factors with DNA sequences that control specific genes.
3. Initiation of transcription is the most important and universally used control point in gene expression.
4. A eukaryotic gene and the DNA segments that control transcription include introns and exons, a promoter sequence upstream of the gene, and a large number of other **control elements**.
 - a) *Control elements are noncoding DNA segments that regulate transcription by binding transcription factors.*
 - b) *The transcription initiation complex assembles on the promoter sequence and one component, RNA polymerase, then transcribes the gene.*
 - c) *Eukaryotic RNA polymerase is dependent on transcription factors before transcription begins.*
 - d) *One transcription factor recognizes the TATA box.*
 - e) *Others in the initiation complex are involved in protein-protein interactions.*
 - f) *High transcription levels require additional transcription factors binding to other control elements.*
 - g) *Distant control elements, **enhancers**, may be thousands of nucleotides away from the promoter or even downstream of the gene or within an intron.*
5. Bending of DNA enables transcription factors, **activators**, bound to enhancers to contact the protein initiation complex at the promoter.
 - a) *This helps position the initiation complex on the promoter.*
 - b) *Eukaryotic genes also have repressor proteins that bind to DNA control elements called silencers.*
 - c) *At the transcription level, activators are probably more important than repressors, because the main regulatory mode of eukaryotic cells seems to be activation of otherwise silent genes.*
 - d) *Repression may operate mostly at the level of chromatin modification.*
6. The hundreds of eukaryotic transcription factors follow only a few basic structural principles.
 - a) *Each protein generally has a **DNA-binding domain** that binds to DNA and a protein-binding domain that recognizes other transcription factors.*
7. A surprisingly small number of completely different nucleotide sequences are found in DNA control elements.
8. Members of a dozen or so sequences about four to ten base pairs long appear again and again in the control elements for different genes.

9. For many genes, the particular combination of control elements associated with the gene may be more important than the presence of a control element unique to the gene.
 - a) *In prokaryotes, coordinately controlled genes are often clustered into an operon with a single promoter and other control elements upstream.*
 - b) *The genes of the operon are transcribed into a single mRNA and translated together.*
 - c) *In contrast, only rarely are eukaryotic genes organized this way.*
 - d) *Genes coding for the enzymes of a metabolic pathway may be scattered over different chromosomes.*
 - e) *Even if genes are on the same chromosome, each gene has its own promoter and is individually transcribed.*
10. Coordinate gene expression in eukaryotes probably depends on the association of a specific control element or collection of control elements with every gene of a dispersed group.
11. A common group of transcription factors bind to them, promoting simultaneous gene transcription.
 - a) *For example, steroid hormones enter a cell and bind to a specific receptor protein in the cytoplasm or nucleus.*
 - b) *After allosteric activation of these proteins, they function as transcription activators.*
 - c) *Other signal molecules can control gene expression indirectly by triggering signal-transduction pathways that lead to transcription activators.*

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

- a) *Gene expression may be blocked or stimulated by any post-transcriptional step.*
- b) *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.*
- c) *RNA processing in the nucleus and the export of mRNA to the cytoplasm provide opportunities for gene regulation that are not available in bacteria.*
- d) *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.*
- e) *The life span of a mRNA molecule is an important factor in determining the pattern of protein synthesis.*
- f) *Prokaryotic mRNA molecules may be degraded after only a few minutes.*
- g) *Eukaryotic mRNAs typically endure for hours and can even last days or weeks.*
- h) *For example, in red blood cells the mRNAs for the hemoglobin polypeptides are unusually stable and are translated repeatedly in these cells.*
- i) *A common pathway of mRNA breakdown begins with enzymatic shortening of the poly(A) tail.*
- j) *This triggers the enzymatic removal of the 5' cap.*
- k) *This is followed by rapid degradation of the mRNA by nucleases.*
- l) *Nucleotide sequences in the untranslated trailer region at the 3' end affect mRNA stability.*
- m) *Transferring such a sequence from a short-lived mRNA to a stable mRNA results in quick mRNA degradation.*
2. Translation of specific mRNAs can be blocked by regulatory proteins that bind to specific sequences or structures within the 5' leader region of mRNA.
 - a) *This prevents attachment to ribosomes.*
3. Protein factors required to initiate translation in eukaryotes offer targets for simultaneously controlling translation of *all* the mRNA in a cell.
 - a) *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, until after fertilization or during daylight in plants).*
 - b) *Finally, eukaryotic polypeptides must often be processed to yield functional proteins.*
 - c) *This may include cleavage, chemical modifications, and transport to the appropriate destination.*
 - d) *Regulation may occur at any of these steps.*
 - e) *For example, cystic fibrosis results from mutations in the genes for a chloride ion channel protein that prevents it from reaching the plasma membrane.*
 - f) *The defective protein is rapidly degraded.*
4. The cell limits the lifetimes of normal proteins by selective degradation.

- a) Many proteins, like the cyclins in the cell cycle, must be short-lived to function appropriately.
- 5. Proteins intended for degradation are marked by the attachment of ubiquitin proteins.
- 6. Giant **proteosomes** recognize the ubiquitin and degrade the tagged protein.

V. The Molecular Biology of Cancer

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

- a) *Cancer is a disease in which cells escape from the control methods that normally regulate cell growth and division.*
 - b) *The agent of such changes can be random spontaneous mutations or environmental influences such as chemical carcinogens or physical mutagens.*
 - c) *Cancer-causing genes, **oncogenes**, were initially discovered in retroviruses, but close counterparts, **proto-oncogenes** were found in other organisms.*
 - d) *The products of proto-oncogenes, proteins that stimulate normal cell growth and division, have essential functions in normal cells.*
 - e) *An oncogene arises from a genetic change that leads to an increase in the proto-oncogene's protein or the activity of each protein molecule.*
 - f) *These genetic changes include movements of DNA within the genome, amplification of proto-oncogenes, and point mutations in the gene.*
2. Malignant cells frequently have chromosomes that have been broken and rejoined incorrectly.
 - a) *This may translocate a fragment to a location near an active promotor or other control element.*
 3. Amplification increases the number of gene copies.
 4. A point mutation may lead to translation of a protein that is more active or longer-lived.
 - a) *Mutations to genes whose normal products inhibit cell division, **tumor-suppressor genes**, also contribute to cancer.*
 - b) *Any decrease in the normal activity of a tumor-suppressor protein may contribute to cancer.*
 - c) *Some tumor-suppressor proteins normally repair damaged DNA, preventing the accumulation of cancer-causing mutations.*
 - d) *Others control the adhesion of cells to each other or to an extracellular matrix, crucial for normal tissues.*
 - e) *Still others are components of cell-signaling pathways that inhibit the cell cycle.*

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

1. 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.
2. 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.
3. Ras, the product of the **ras gene**, is a G protein that relays a growth signal from a growth factor receptor to a cascade of protein kinases.
 - a) *At the end of the pathway is the synthesis of a protein that stimulates the cell cycle.*
 - b) *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.*
4. The tumor-suppressor protein encoded by the normal *p53* gene is a transcription factor that promotes synthesis of growth-inhibiting proteins.
5. A mutation that knocks out the *p53* gene can lead to excessive cell growth and cancer.
 - a) *Mutations that result in stimulation of growth-stimulating pathways or deficiencies in growth-inhibiting pathways lead to increased cell division.*
6. The *p53* gene, named for its 53,000-dalton protein product, is often called the "guardian angel of the genome".
7. Damage to the cell's DNA acts as a signal that leads to expression of the *p53* gene.
8. The p53 protein is a transcription factor for several genes.

- a) It can activate the p21 gene, which halts the cell cycle.
- b) It can turn on genes involved in DNA repair.
- c) When DNA damage is irreparable, the p53 protein can activate "suicide genes" whose protein products cause cell death by apoptosis.

C. Multiple mutations underlie the development of cancer

1. More than one somatic mutation is generally needed to produce the changes characteristic of a full-fledged cancer cell.
2. 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.
3. Colorectal cancer, with 135,000 new cases in the U.S. each year, illustrates a multi-step cancer path.
4. The first sign is often a polyp, a small benign growth in the colon lining with fast dividing cells.
5. Through gradual accumulation of mutations that activate oncogenes and knock out tumor-suppressor genes, the polyp can develop into a malignant tumor.
6. About a half dozen DNA changes must occur for a cell to become fully cancerous.
7. 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.
8. In many malignant tumors, the gene for telomerase is activated, removing a natural limit on the number of times the cell can divide.
 - a) Viruses, especially retroviruses, play a role in about 15% of human cancer cases worldwide.
 - b) These include some types of leukemia, liver cancer, and cancer of the cervix.
 - c) Viruses promote cancer development by integrating their DNA into that of infected cells.
 - d) By this process, a retrovirus may donate an oncogene to the cell.
 - e) Alternatively, insertion of viral DNA may disrupt a tumor-suppressor gene or convert a proto-oncogene to an oncogene.
9. The fact that multiple genetic changes are required to produce a cancer cell helps explain the predispositions to cancer that run in some families.
 - a) 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.
10. Geneticists are devoting much effort to finding inherited cancer alleles so that predisposition to certain cancers can be detected early in life.
 - a) About 15% of colorectal cancers involve inherited mutations, especially to DNA repair genes or to the tumor-suppressor gene APC.
 - Normal functions of the APC gene include regulation of cell migration and adhesion.
 - Between 5-10% of breast cancer cases, the 2nd most common U.S. cancer, show an inherited predisposition.
 - Mutations to one of two tumor-suppressor genes, BRCA1 and BRCA2, increases the risk of breast and ovarian cancer.