Complex biological processes often require **coordinated control of the expression of many genes**. The maturation of a tadpole into a frog is largely controlled by thyroid hormone. This hormone regulates gene expression by binding to a protein, the thyroid-hormone receptor, as shown at the right. In response to the hormone’s binding, this protein **binds to specific DNA sites** in the genome and modulates the expression of nearby genes.
Many of the most important and intriguing features in modern biology and medicine, such as the pathways crucial for the development of **multicellular organisms**, the changes that distinguish **normal cells** and **cancer cells**, and the evolutionary changes **leading to new species**, entail **networks of gene-regulatory pathways**.

**Gene regulation in eukaryotes** is **significantly more complex** than in prokaryotes in several ways.

**First**, the **genomes** being regulated are **significantly larger**. The **E. coli genome** consists of a **single, circular chromosome** containing **4.6 Mb**, encoding approximately **2000 proteins**.

In comparison, one of the simplest eukaryotes, **Saccharomyces cerevisiae (baker’s yeast)**, contains **16 chromosomes** ranging in size from **0.2 to 2.2 Mb** (Figure 32.1). The yeast genome totals **12 Mb** and encodes approximately **6000 proteins**.

**The genome within a human cell** contains **23 pairs** of chromosomes ranging in size from **50 to 250 Mb**. Approximately **23,000 genes** are present within the **3000 Mb** of human DNA.
Second, whereas prokaryotic genomic DNA is relatively accessible, eukaryotic DNA is packaged into **chromatin**, a complex between the DNA and a special set of proteins (Figure 32.2).

Although the principles for the construction of chromatin are relatively simple, the **chromatin structure for a complete genome is quite complex.**

Importantly, in a given eukaryotic cell, some genes and their associated regulatory regions are **relatively accessible** for transcription and regulation, whereas **other genes are tightly packaged** and are thus rendered **inactive**.

**Eukaryotic gene regulation** frequently requires the **manipulation of chromatin structure**.
A manifestation of this complexity is the presence of many different cell types in most eukaryotes.

A liver cell, a pancreatic cell, and an embryonic stem cell contain the same DNA sequences, yet the subset of genes highly expressed in cells from the pancreas, which secretes digestive enzymes, differs markedly from the subset highly expressed in the liver, the site of lipid transport and energy transduction.

Embryonic stem cells do not express any subset of genes at high levels; the most highly expressed genes are “housekeeping” genes involved in the cytoskeleton and processes such as translation.

Table 32.1 Highly expressed protein-encoding genes of the pancreas, liver, and embryonic stem cells (as percentage of total mRNA pool)

<table>
<thead>
<tr>
<th>Rank</th>
<th>Proteins expressed in pancreas</th>
<th>%</th>
<th>Proteins expressed in liver</th>
<th>%</th>
<th>Proteins expressed in stem cells</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Procarboxypeptidase A1</td>
<td>7.6</td>
<td>Albumin</td>
<td>3.5</td>
<td>Glyceraldehyde-3-phosphate dehydrogenase</td>
<td>0.7</td>
</tr>
<tr>
<td>2</td>
<td>Pancreatic trypsinogen 2</td>
<td>5.5</td>
<td>Apolipoprotein A-I</td>
<td>2.8</td>
<td>Translation elongation factor 1 α1</td>
<td>0.6</td>
</tr>
<tr>
<td>3</td>
<td>Chymotrypsinogen</td>
<td>4.4</td>
<td>Apolipoprotein C-I</td>
<td>2.5</td>
<td>Tubulin α</td>
<td>0.5</td>
</tr>
<tr>
<td>4</td>
<td>Pancreatic trypsin 1</td>
<td>3.7</td>
<td>Apolipoprotein C-III</td>
<td>2.1</td>
<td>Translationally controlled tumor protein</td>
<td>0.5</td>
</tr>
<tr>
<td>5</td>
<td>Elastase IIIB</td>
<td>2.4</td>
<td>ATPase 6/8</td>
<td>1.5</td>
<td>Cyclophilin A</td>
<td>0.4</td>
</tr>
<tr>
<td>6</td>
<td>Protease E</td>
<td>1.9</td>
<td>Cytochrome oxidase 3</td>
<td>1.1</td>
<td>Coflin</td>
<td>0.4</td>
</tr>
<tr>
<td>7</td>
<td>Pancreatic lipase</td>
<td>1.9</td>
<td>Cytochrome oxidase 2</td>
<td>1.1</td>
<td>Nucleophosmin</td>
<td>0.3</td>
</tr>
<tr>
<td>8</td>
<td>Procarboxypeptidase B B</td>
<td>1.7</td>
<td>α1-Antitrypsin</td>
<td>1.0</td>
<td>Connexin 43</td>
<td>0.3</td>
</tr>
<tr>
<td>9</td>
<td>Pancreatic amylase</td>
<td>1.7</td>
<td>Cytochrome oxidase 1</td>
<td>0.9</td>
<td>Phosphoglycerate mutase</td>
<td>0.2</td>
</tr>
<tr>
<td>10</td>
<td>Bile-salt-stimulated lipase</td>
<td>1.4</td>
<td>Apolipoprotein E</td>
<td>0.9</td>
<td>Translation elongation factor 1 β2</td>
<td>0.2</td>
</tr>
</tbody>
</table>
The existence of stable cell types is due to differences in the epigenome, differences in chromatin structure, and covalent modifications of the DNA, not in the DNA sequence itself.
In addition, eukaryotic genes are **not** generally organized into **operons**.

Instead, genes that encode proteins for steps within a given pathway are **often spread widely across the genome**.

This characteristic requires that **other mechanisms** function to regulate genes in a **coordinated** way.
Despite these differences, some aspects of gene regulation in eukaryotes are quite similar to those in prokaryotes.

In particular, activator and repressor proteins that recognize specific DNA sequences are central to many gene-regulatory processes.

In this chapter, we will focus first on chromatin structure.

We will then turn to transcription factors—DNA-binding proteins similar in many ways to the prokaryotic proteins that we encountered in the preceding chapter.

Eukaryotic transcription factors can act directly by interacting with the transcriptional machinery or indirectly by influencing chromatin structure.

Finally, we examine selected posttranscriptional gene-regulatory mechanisms, including those based on microRNAs, an important class of gene-regulatory molecules discovered in recent years.
32.1 Eukaryotic DNA Is Organized into Chromatin

Eukaryotic DNA is tightly bound to a group of small basic proteins called histones.

In fact, histones constitute half the mass of a eukaryotic chromosome.

The entire complex of a cell’s DNA and associated protein is called chromatin.

Chromatin serves to compact and organize eukaryotic DNA and its presence has dramatic consequences for gene regulation.
Nucleosomes are complexes of DNA and histones

Chromatin is made up of **repeating units**, each containing **200 bp of DNA** and two copies each of **four histone proteins H2A, H2B, H3, and H4**. Histones have **strikingly basic properties** because a quarter of the residues in each histone are either **arginine or lysine**, **positively charged amino acids** that strongly interact with the negatively charged DNA.

The protein complex is called the **histone octamer**. The repeating units of the histone octamer and the associated DNA are known as **nucleosomes**. Chromatin viewed with the electron microscope has the appearance of **beads on a string**.
Each bead has a diameter of approximately 100 Å. Partial digestion of chromatin with DNase yields the isolated beads. These particles consist of fragments of DNA about **200 bp** in length bound to the histone octamer.

More-extensive digestion yields a shorter DNA fragment of **145 bp bound to the octamer**. The smaller complex formed by the histone octamer and the 145-bp DNA fragment is the nucleosome core particle.

The DNA connecting core particles in undigested chromatin is called **linker DNA**. **Histone H1 binds**, in part, to the linker DNA.
DNA wraps around histone octamers to form nucleosomes

The overall structure of the nucleosome was revealed through **electron microscopic** and **x-ray crystallographic** studies pioneered by Aaron Klug (1982 Nobel).

More recently, the three-dimensional structure of a reconstituted nucleosome core particle (Figure 32.3) was determined to higher resolution by x-ray diffraction methods.

The **four types of histone** that make up the protein core are homologous and similar in structure (Figure 32.4). The eight histones in the core are arranged into a \((\text{H3})_2(\text{H4})_2\) tetramer and a pair of H2A–H2B dimers.
The tetramer and dimers come together to form a left-handed superhelical ramp around which the DNA wraps. In addition, each histone has an **amino-terminal tail** that extends out from the core structure. These tails are flexible and contain a number of lysine and arginine residues. As we shall see, **covalent modifications of these tails play an essential role in regulating gene expression.**
The winding of DNA around the nucleosome core contributes to the **packing of DNA** by decreasing its linear extent.

Clearly, the nucleosome is **just the first step in DNA compaction**.

**What is the next step?**

The nucleosomes themselves are arranged in a helical array approximately 360 Å across, forming a **series of stacked layers** approximately 110 Å apart (Figure 32.5).

The folding of these **fibers** of nucleosomes into loops further compacts DNA.
32.2 Transcription Factors Bind DNA and Regulate Transcription Initiation

The gene control region for a typical eucaryotic gene
The roles of eukaryotic transcription factors are different in several ways.

First, whereas the DNA-binding sites crucial for the control of gene expression in prokaryotes are usually quite close to promoters, those in eukaryotes can be farther away from promoters and can exert their action at a distance.

Third, in prokaryotes, transcription factors usually interact directly with RNA polymerase. In eukaryotes, some transcription factors interact directly with RNA polymerase, whereas others interact with other proteins associated with RNA polymerase and still others act by modifying the chromatin structure.

Second, most prokaryotic genes are regulated by single transcription factors, and multiple genes in a pathway are expressed in a coordinated fashion because such genes are often transcribed as part of a polycistronic mRNA.

In eukaryotes, the expression of each gene is typically controlled by multiple transcription factors, and the coordinated expression of different genes depends on having similar transcription-factor-binding sites in each gene in the set.
A range of DNA-binding structures are employed by eukaryotic DNA-binding proteins
A range of DNA-binding structures are employed by eukaryotic DNA-binding proteins

The structures of many eukaryotic DNA-binding proteins have been determined and a range of structural motifs have been observed, but we will focus on three that reveal the common features and the diversity of these motifs.

The first class of eukaryotic DNA-binding unit that we will consider is the homeodomain (Figure 32.6). The structure of this domain and its mode of recognition of DNA are very similar to those of the prokaryotic helix-turn-helix proteins.

In eukaryotes, homeodomain proteins often form heterodimeric structures, sometimes with other homeodomain proteins, that recognize asymmetric DNA sequences.
The second class of eukaryotic DNA-binding unit comprises the basic-leucine zipper (bZip) proteins (Figure 32.7).

This DNA-binding unit consists of a pair of long α helices.

The first part of each α helix is a basic region that lies in the major groove of the DNA and makes contacts responsible for DNA-site recognition.

The second part of each α helix forms a coiled-coil structure with its partner.

Because these units are often stabilized by appropriately spaced leucine residues, these structures are often referred to as leucine zippers.
The final class of eukaryotic DNA-binding units that we will consider here are the Cys$_2$His$_2$ zinc-finger domains (Figure 32.8).

A DNA-binding unit of this class comprises tandem sets of small domains, each of which binds a zinc ion through conserved sets of two cysteine and two histidine residues.

These domains, often called zinc-finger domains, form a string that follows the major groove of DNA. An a helix from each domain makes specific contact with the edges of base pairs within the groove.

The human genome encodes several hundred proteins that contain zinc-finger domains of this class.

We will encounter another class of zinc-based DNA-binding domain when we consider nuclear hormone receptors in Section 32.3.
Activation domains interact with other proteins

The activation domains of transcription factors generally recruit other proteins that promote transcription.

In some cases, these activation domains interact directly with RNA polymerase II or closely associated proteins.

The activation domains act through intermediary proteins that bridge between the transcription factors and the polymerase. An important target of activators is mediator, a complex of 25 to 30 subunits conserved from yeast to human beings, that acts as a bridge between transcription factors and promoter-bound RNA polymerase II (Figure 32.9).
Activation domains are less conserved than DNA-binding domains. In fact, very little sequence similarity has been found. For example, they may be acidic, hydrophobic, glutamine rich, or proline rich. However, certain features are common to activation domains.

First, they are often redundant; that is, a part of the activation domain can be deleted without loss of function.

Second, they are modular and can activate transcription when paired with a variety of DNA-binding domains.

Third, activation domains can act synergistically: two activation domains acting together create a stronger effect than either acting separately.

We have been considering the case in which gene control increases the expression level of a gene.

In many cases, the expression of a gene must be decreased by blocking transcription. The agents in such cases are transcriptional repressors.

Like activators, transcriptional repressors act in many cases by altering chromatin structure.
Multiple transcription factors interact with eukaryotic regulatory regions

In contrast with the regulators of prokaryotic transcription, few eukaryotic transcription factors have any effect on transcription on their own. Instead, each factor recruits other proteins to **build up large complexes** that interact with the transcriptional machinery to activate transcription.

A major advantage of this mode of regulation is that **a given regulatory protein can have different effects, depending on what other proteins are present** in the same cell.

This phenomenon, called **combinatorial control**, is crucial to multicellular organisms that have many different cell types. Even in unicellular eukaryotes such as yeast, combinatorial control allows the generation of distinct cell types.
Enhancers can stimulate transcription in specific cell types

Transcription factors can often act even if their binding sites lie at a considerable distance from the promoter. These distant regulatory sites are called enhancers (Chapter 29).

Enhancers function by serving as binding sites for specific transcription factors. An enhancer is effective only in the specific cell types in which appropriate regulatory proteins are expressed. In many cases, these DNA-binding proteins influence transcription initiation by perturbing the local chromatin structure to expose a gene or its regulatory sites rather than by direct interactions with RNA polymerase.

This mechanism accounts for the ability of enhancers to act at a distance. The properties of enhancers are illustrated by studies of the enhancer controlling the muscle isoform of creatine kinase (Figure 32.10).

![Diagram of enhancer controlling muscle isoform of creatine kinase](image)
The results of mutagenesis and other studies revealed the presence of an enhancer located between 1350 and 1050 base pairs upstream of the start site of the gene for this enzyme.

Experimentally inserting this enhancer near a gene not normally expressed in muscle cells is sufficient to cause the gene to be expressed at high levels in muscle cells but not in other cells (Figure 32.11).
Induced pluripotent stem cells can be generated by introducing four transcription factors into differentiated cells.

A cell that can differentiate into all cell types of the adult organism is pluripotent.
A cell that can differentiate into all cell types, including the placental tissue, is totipotent.
An important application illustrating the power of transcription factors is the development of induced pluripotent stem (iPS) cells.

Pluripotent stem cells have the ability to differentiate into many different cell types on appropriate treatment. Previously isolated cells derived from embryos show a very high degree of pluripotency.

Over time, researchers identified dozens of genes in embryonic stem cells that contributed to this pluripotency when expressed.

In a remarkable experiment reported for mouse cells in 2006 and human cells in 2007,

Shinya Yamanaka (2012 Nobel) demonstrated that just four genes out of this entire set could induce pluripotency in already-differentiated skin cells.

Yamanaka introduced genes encoding four transcription factors into skin cells called fibroblasts. The fibroblasts de-differentiated into cells that appeared to have characteristics very nearly identical with those of embryonic stem cells (Figure 32.12).
These iPS cells represent **powerful new research tools** and, potentially, a new class of therapeutic agents.

The proposed concept is that a sample of **a patient’s fibroblasts** could be readily isolated and **converted into iPS cells**.

These iPS cells could then be treated to differentiate into a desired cell type that could then be transplanted into the patient.

For example, such an approach might be used to restore **a particular class of nerve cells** that had been depleted by **a neurodegenerative disease**.

Although the field of iPS cell research is still in its very early stages, **it holds great promise** as a possible approach to treatment for many common and difficult-to-treat diseases.

---

**Fibroblast**: synthesizes extracellular matrix and collagen

*Derived from Human iPS Cells*

*Cell, 2007, 131(5), 861*
From Fibroblast to iPS