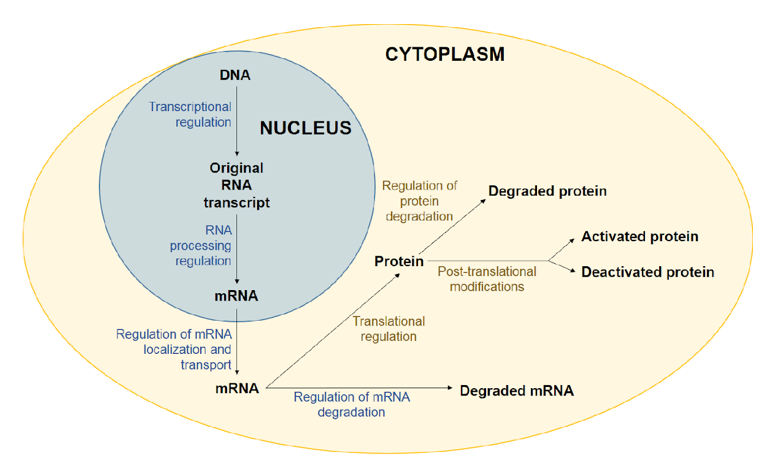
Chapter 5: Transcriptional Regulation

Converting the information stored in DNA into the functional materials of the cell is central to life. It becomes critical then to understand that not all genes are expressed in every cell all of the time. How then does each cell ‘know’ what genes it needs to express and when?

If we consider all of the different steps it takes to go from DNA to an active protein we can begin to understand all the different mechanisms a cell has to regulate gene production. Figure 5-1 illustrates all the different places along the pathway from DNA to active protein that can be regulated.

The first and most important step in regulation is that of transcription. It stands to reason, if the gene product is not needed, either at that time or ever in the cell, the best mode to regulate its production is to never express it in the first place. Therefore, regulating gene expression ensures that if a gene is not needed, it is never transcribed. This is pivotal for ensuring proper gene expression and for saving much needed sources of energy within the cell.



**Figure 5-1: Regulation of gene expression, protein production, protein activity.** There are several steps that can be regulated to determine gene expression levels or whether an active protein is produced.

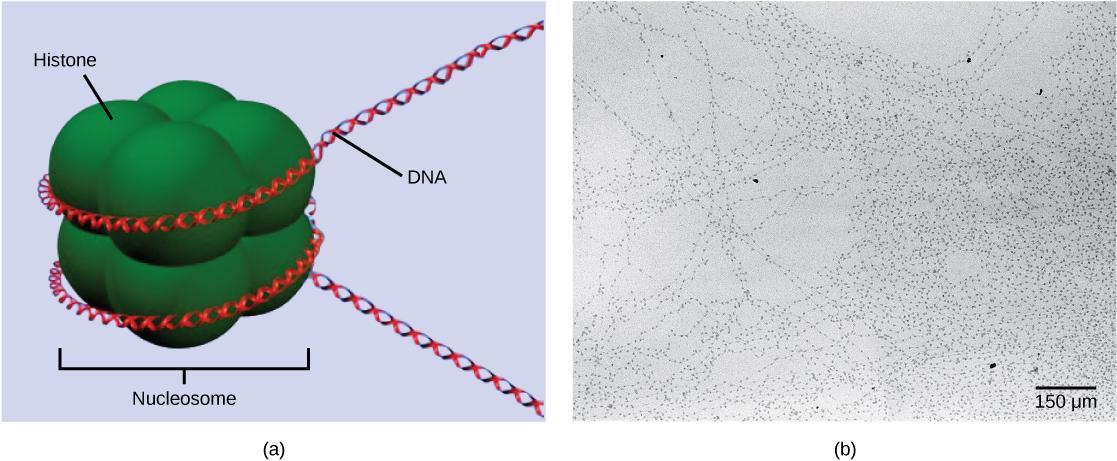
# Epigenetic Control

In genetics, there can be mutations of the DNA, which, if passed to daughter cells, is a permanent change that will be inherited by each new daughter cell until another mutation occurs at that same spot. These mutations occur in the nitrogenous bases of DNA. DNA also can be modified at a level “above” the nucleotides, this is termed epigenetics. The following section describes some of the modifications alter the three dimensional shape of DNA, but don’t necessarily affect the DNA sequence itself. These changes are often transient and can help to dynamically affect the expression of genes in cells to ensure appropriate levels of protein production. When this level of regulation is perturbed there may be misregulation of genes and disease phenotypes.

## Histone Modification

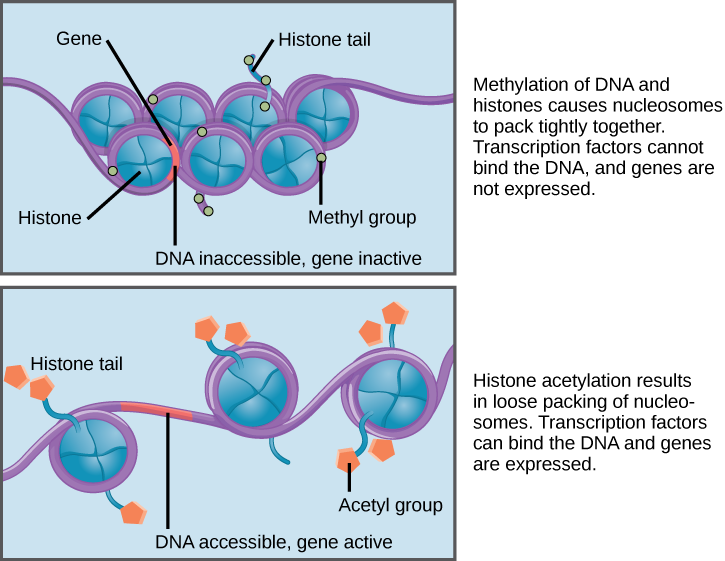
The human genome encodes over 20,000 genes, with hundreds to thousands of genes on each of the 23 human chromosomes. As discussed in an earlier unit, the DNA in the nucleus is precisely wound, folded, and compacted into chromatin so that it will fit into the nucleus. It is also organized so that specific segments can be accessed as needed for specific cell types to function.

The first level of organization, or packing, is the winding of DNA strands around histone proteins. Histones package and order DNA into structural units called nucleosome complexes, which can control the access of proteins to the DNA regions ([Figure](https://openstax.org/books/biology-2e/pages/16-3-eukaryotic-epigenetic-gene-regulation#fig-ch16_03_01) 5-2a). Under the electron microscope, this winding of DNA around histone proteins to form nucleosomes looks like small beads on a string ([Figure](https://openstax.org/books/biology-2e/pages/16-3-eukaryotic-epigenetic-gene-regulation#fig-ch16_03_01) 5-2b).



**Figure 5-2: Nucleosomes of DNA**. DNA is folded around histone proteins to create (a) nucleosome complexes. These nucleosomes control the access of proteins to the underlying DNA. When viewed through an electron microscope (b), the nucleosomes look like beads on a string. (credit “micrograph”: modification of work by Chris Woodcock)

These histone proteins can move along the string (DNA) to expose different sections of the molecule. If DNA encoding a specific gene is to be transcribed into RNA, the nucleosomes surrounding that region of DNA can slide down the DNA to open that specific chromosomal region and allow for the transcriptional machinery, including transcription factors and RNA polymerase to initiate transcription. The movement of nucleosome complexes is achieved by ATP-dependent chromatin remodeling complexes that use the energy from ATP hydrolysis to push nucleosomes along the DNA.



**Figure 5-3: Nucleosomes can slide along DNA**. When nucleosomes are spaced closely together (top), transcription factors cannot bind, and gene expression is turned off. When the nucleosomes are spaced far apart (bottom), the DNA is exposed. Transcription factors can bind, allowing gene expression to occur. Modifications to the histones and DNA affect nucleosome spacing.

How closely the histone proteins associate with the DNA is regulated by chemical signals found on both the histone proteins and on the DNA. These signals are functional groups (aka tags) added to histone proteins or to DNA and determine whether a chromosomal region should be open (euchromatic) or closed (heterochromatic) ([Figure](https://openstax.org/books/biology-2e/pages/16-3-eukaryotic-epigenetic-gene-regulation#fig-ch16_03_03) 5-3 depicts modifications to histone proteins and DNA). These tags are not permanent but may be added or removed as needed. The most common chemical groups added directly to the nucleic acids are methyl groups, while those added to histone proteins include phosphate, methyl, or acetyl groups. These functional groups are attached to specific amino acids in histone "tails" at the N-terminus of the protein and, importantly, do not alter the DNA base sequence, but they do alter how tightly wound the DNA is around the histone proteins.

DNA is a negatively charged molecule, and unmodified histones are positively charged; therefore, changes in the charge of the histone will change how tightly wound the DNA molecule will be. For example, by adding chemical modifications like acetyl groups, the charge of the histone becomes less positive, and the binding of DNA to the histones is often relaxed. This allows an increase in the accessibility of the gene. Phosphate groups are negatively charged, therefore, phosphorylation of histone tails decreases the binding of positively charged histone proteins to DNA, also resulting in accessibility of the gene. Methylation of histone tails has been shown to both repress and activate gene transcription, depending on which amino acids of the histone tails are modified and the number of methyl groups added. For simplicity, we will discuss methylation as a functional group that results in gene silencing or heterochromatin. One hypothesis for how methylation causes an increase in DNA coiling is that methyl groups are nonpolar and nonpolar molecules tend to aggregate (or bind) together. Therefore, the more methyl groups there are on histones the more likely the histones will package closer together.

The commonality between all modifications to the histone tails is that either the modification is relaxing the coiling of DNA around the histones allowing the DNA to become more ‘open’ and accessible to the transcription machinery, or the modification tightens the coiling of DNA causing it to become more ‘closed’ and transcriptionally silent (refer to figure 5-3). Therefore, altering the tightness of the histone-DNA interaction opens some regions of chromatin to transcription and closes others.

## DNA Modification: Methylation

The DNA molecule itself can also be modified by methylation. DNA methylation occurs within very specific regions called CpG islands. These are regions of the DNA that contain a high frequency of cytosine and guanine dinucleotide DNA pairs (CG). The ‘p’ stands for the phosphate group that is part of the phosphodiester bond between the cytosine and guanine (CG). Figure 5-4 illustrates the difference between a gene containing CpG island and a normal gene promoter sequence. These are found in the promoter regions of genes. The cytosine base of the CG pair can be methylated, which often causes the gene to be silenced because transcription factors can no longer bind to the promoter, as the methyl group blocks that interaction. In some cases, genes that are silenced during the development of the gametes of one parent are transmitted in their silenced condition to the offspring. Such genes are said to be imprinted. Parental diet or other environmental conditions can affect the methylation patterns of genes, which in turn modifies gene expression. Additionally, external environmental changes after the offspring are born can dynamically alter the methylation status of genes, which can impact human health (Figure 5-4). Some such stimuli are trauma, physical activity, smoking, alcohol consumption, environmental pollutants, and obesity.

The methylation of histones can influence the methylation of DNA. DNA methyltransferases (enzymes important in methylating DNA) appear to be attracted to chromatin regions with specific histone modifications. Highly methylated (hypermethylated) DNA regions with deacetylated (lacking acetyl groups) histones are tightly coiled and transcriptionally inactive.



**Figure 5-4: Example of DNA sequence that is a known CpG island.** The CG sequence is in yellow. Notice the extensive amount of CG dinucleotides. The ATG in red, this denotes the start of translation for this gene. This signifies that this region is likely to be highly methylated, silencing transcription of this gene. Alterations in the sequence of CpG islands can result in misregulation of genes and inappropriate production of proteins that ultimately can result in disease phenotypes. The image on the right represents an area of DNA that is not regulated by CpG islands.

Epigenetic changes are not permanent in the same way a mutation in the DNA sequence can be, although they often persist through multiple rounds of cell division and may be heritable.

Chromatin remodeling alters the chromosomal structure (open or closed) as needed. If a gene is to be transcribed, the histone proteins and DNA in the chromosomal region encoding that gene are modified in a way that opens the promoter region to allow RNA polymerase and other proteins, called transcription factors, to bind and initiate transcription. If a gene is to remain turned off, or silenced, the histone proteins and DNA have different modifications that signal a closed chromosomal configuration. In this closed configuration, the RNA polymerase and transcription factors do not have access to the DNA, and transcription cannot occur ([Figure](https://openstax.org/books/biology-2e/pages/16-3-eukaryotic-epigenetic-gene-regulation#fig-ch16_03_03) 5-5).

Modifications affect nucleosome spacing and gene expression. (credit: modification of work by NIH)



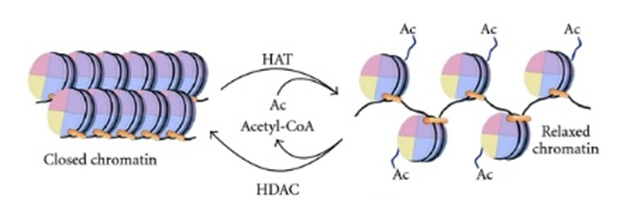
**Figure 5-5: Histone proteins and DNA nucleotides can be modified chemically.** Modifications affect nucleosome spacing and gene expression. (credit: modification of work by NIH)

## Chromatin Remodeling

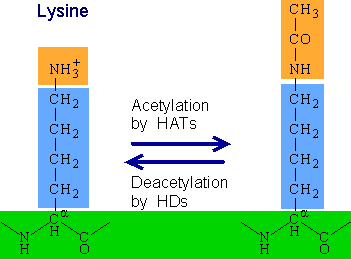
Chromatin remodeling is the dynamic modification of the structure of chromatin to allow transcription factors access to a gene’s promoter, and thereby control gene expression. Such remodeling is typically carried out by covalent histone modifications by specific enzymes, e.g., histone acetyltransferases (HATs), deacetylases (HDACs), methyltransferases, and kinases. Besides actively regulating gene expression, dynamic remodeling of chromatin is important in several key biological processes, DNA replication and repair; apoptosis; chromosome segregation as well as development and pluripotency. Mistakes in chromatin remodeling proteins are associated with human diseases, including cancer. Targeting chromatin remodeling pathways is currently becoming a primary therapeutic strategy in the treatment of several cancers.

**Covalent histone modification enzymes**

There are several enzymes that cells express that can modify the chemical tags on histone tails. The two opposing enzymes responsible for acetylation are called HATs and HDACs. HATs are enzymes that acetylate conserved lysine amino acids on histone proteins by transferring an acetyl group from acetyl-CoA to lysine. As mentioned before, this modification is often associated with increased gene expression and is often found in areas of euchromatin. HDACs are enzymes that remove acetyl groups from lysine and cause the chromatin to become more tightly packed (aka heterochromatin). Figures 5-6 and 5-7 below depicts the relationship between these enzymes and their impact on chromatin structure and what the chemical modification looks like on a lysine amino acid, respectively.

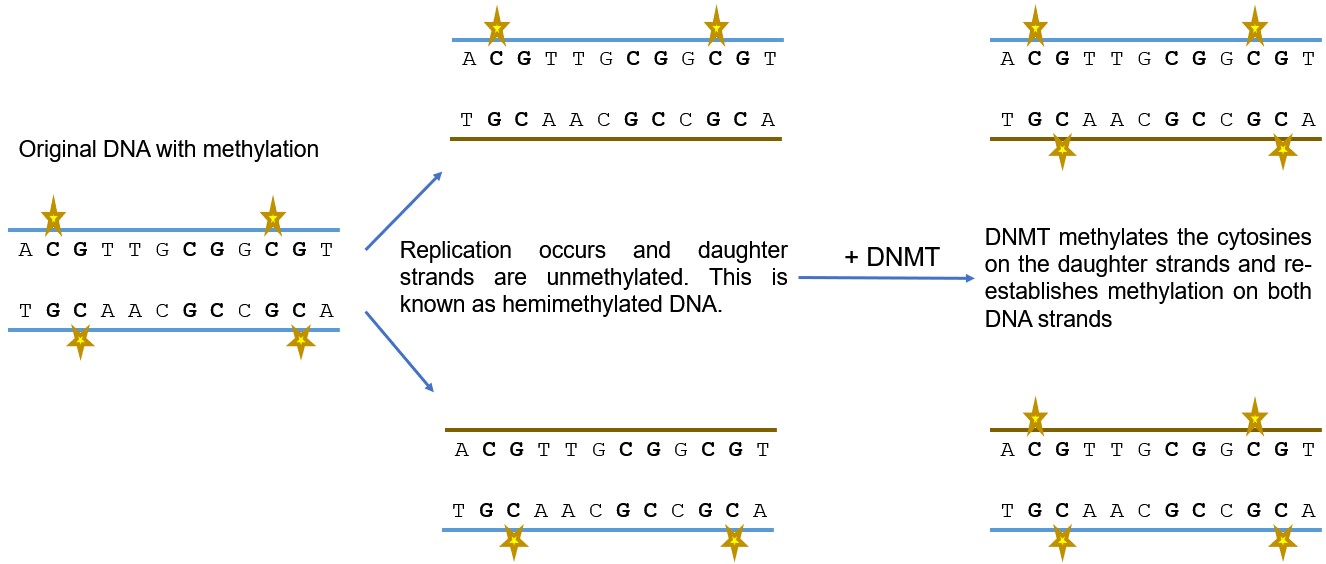


**Figure 5-6: Acetylation/deacetylation cycle and its typical effect on chromatin structure**. Histone acetyl transferases (HATs) add acetyl groups to the unstructured ‘tails’ of histone proteins, usually opening the chromatin structure. Histone deacetylases (HDACs) remove acetyl groups from histone tails, typically resulting in a closed chromatin structure.



**Figure 5-7: Example of the chemical change in a lysine residue upon acetylation by a HAT.** The structure highlighted in green is the backbone structure of the amino acid. The blue and orange represent the R-group of lysine. The blue structure is unchanged by the HAT. The terminal amino group is replaced by an acetyl group, highlighted in orange.

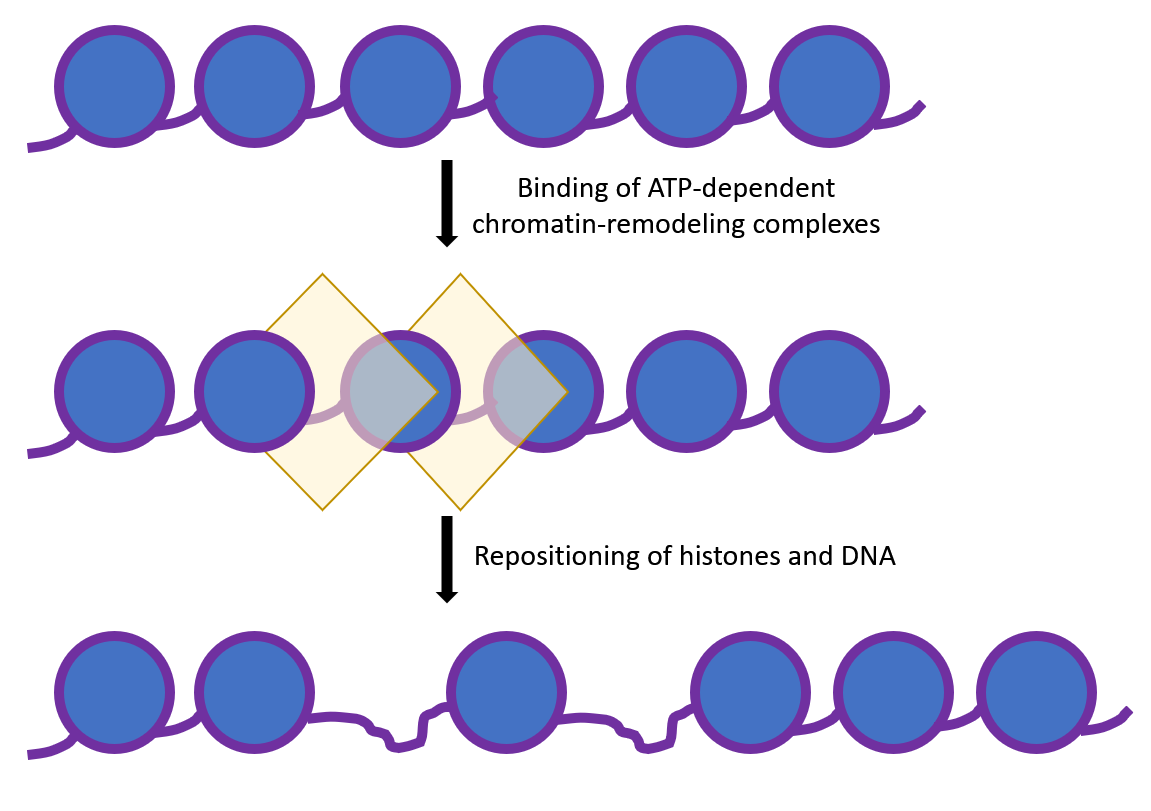
Histone methyltransferases (HMTs) are enzymes that can add one or more methyl groups to either lysine or arginine amino acids found in the tail region of histones (figure 5-7). Research suggests that depending on the symmetry and location of the methylation, it can be either repressive or activating. Alternatively, DNA methyltransferases are enzymes that add a methyl group directly to cytosine in DNA (rather than a histone tail modification), to produce 5-methylcytosine, which is typically repressive or causes transcriptional silencing. There are two major times when DNA methyltransferases (DNMTs) work (1) is during early embryonic development to begin silencing genes in specific cells to begin helping the cell know what type of cell it will become, and (2) to maintain methylation on the daughter strands of DNA after replication. The first type is called *de novo* methylation. It is thought that the DNMT recognizes a structure in the DNA and potentially patterns of modifications in histone tails and this helps recruit the enzyme to sequences of DNA that should be methylated. The maintenance version of methylation occurs after replication. When DNA is replicated a parental strand is used as a template to make a daughter strand. The parental strand will retain its methyl groups on specific cytosine bases, but the newly replicated strands do not contain these marks, therefore, DNMT recognizes the pattern on the parental strand and methylates the daughter strand, thereby ensuring that the epigenetic patterns are maintained from parent cell to daughter cell (see figure 5-8).



**Figure 5-8: Action of DNA methyltransferases (DNMTs).** The original DNA is methylated at four cytosine bases as denoted by the stars. After replication, the daughter strands (in gold) are not methylated. The methylation marks on the parental strand (blue) help orient DNMTs to know where another cytosine base needs to be methylated. Once the daughter strands are methylated the correct methylation pattern has been re-established. (Illustration by Jennell Talley)

Chromatin Remodeling Complex

Another mechanism the cell has to alter the structure of the chromatin is an **ATP-dependent chromatin-remodeling complex**. These complexes consist of a number of proteins that work together to reposition the histones in relationship to the DNA. Since energy is required to move the histones around, ATP is hydrolyzed (hence the term “ATP-dependent”). It is thought that the complex interacts with both the histone proteins and the DNA to either make the chromatin more or less packed (see figure 5-9). Interestingly, ATP-dependent chromatin-remodeling complexes often work in tandem with enzymes that add functional groups to histones, coordinating both of these processes to either ‘open’ or ‘close’ regions of DNA.



**Figure 5-9. ATP-dependent chromatin-remodeling complex.** **The DNA is represented in purple and the histones are in blue. The DNA is wound around histone proteins forming a nucleosome. ATP-dependent chromatin-remodeling complexes bind to the DNA/histone complex and help to reposition the DNA. In this example, the remodeling complexes are “relaxing” or “opening” the chromatin structure to make DNA more accessible. (Illustration by Jennell Talley)**

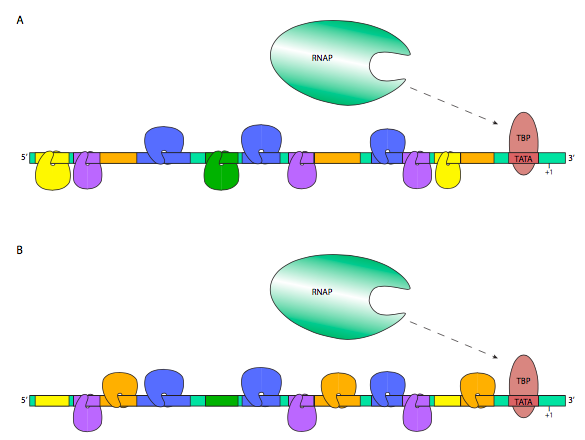
# Turning gene expression on and off

Each cell, whether prokaryotic or eukaryotic, cannot possibly express every gene at the same time. It would be incredibly wasteful of resources and counterproductive to the cell’s specific function. The right genes must be “turned on” or “turned off” based on the needs of the cell at specific times and also depending on the type of cell it is, as different cells express different genes. There is a complex network of control mechanisms that regulate the differential expression of genes.

Transcription regulation usually starts with extracellular environmental signaling. The signals may be chemicals in the air, in the water, or in the case of multicellular organisms, in blood, lymph, or other extracellular fluids. Signal molecules in higher organisms include hormones released at the appropriate time in a sequential developmental program of gene expression, or in response to nutrient levels in body fluids. Some signal molecules get into cells by binding to specific intracellular receptors to convey their instructions. Others bind to cell surface receptors that transduce their ‘information’ into intracellular molecular signals. (See Chapter 11) When signaling leads to gene regulation, the responding cells ultimately produce or activate existing transcription factors. These, in turn, recognize and bind to specific regulatory DNA sequences associated with the genes that they control. DNA sequences that bind transcription factors are relatively short. They can lie proximal (close) to the transcription start site of a gene, and/or distal (far from) to it. We will see that some regulatory DNA sequences are enhancers, turning on or increasing gene transcription when bound by transcription factors or other regulatory proteins. Others are silencers, down-regulating, or suppressing transcription of a gene. Finally, DNA regulatory sequences are hidden behind a thicket of chromatin proteins in eukaryotes. When patterns of gene expression in cells change during development, chromatin is re-organized, cells differentiate, and new tissues and organs form. To this end, new patterns of gene expression and chromatin configuration in a cell must be remembered in its descendants.

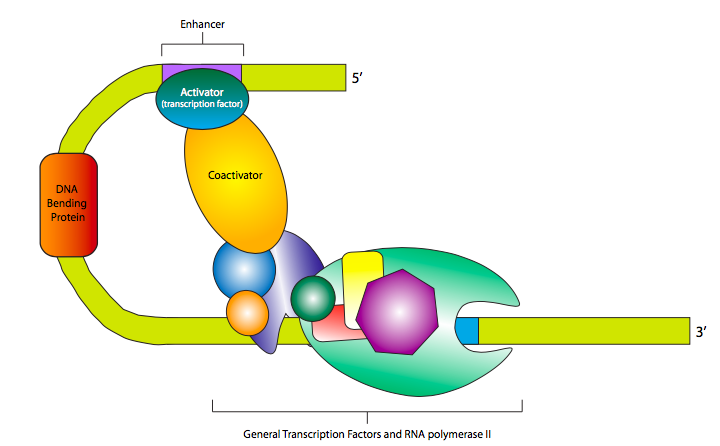
## Regulation of gene expression in eukaryotes

Eukaryotic genes are always transcribed one per mRNA, so one promoter region would be regulated for each instance of gene expression. The previous chapter described the formation of a preinitiation complex of transcription factors for RNA polymerase II (Chapter 3). These transcription factors (e.g. TFIID, TFIIH, etc.) are known as general transcription factors, and are required for transcription of any gene at any level. However, there are also specific transcription factors, usually referred to simply as transcription factors (TF) that modulate the frequency of transcription of particular genes (Fig 5-10). Some upstream elements and their associated TFs are fairly common, while others are gene or gene-family specific. An example of the former is the upstream element AACCAAT and its associated transcription factor, CP1. Another transcription factor, Sp1, is similarly common, and binds to a consensus sequence of ACGCCC. Both are used in the control of the beta-globin gene, along with more specific transcription factors, such as GATA-1, which binds a consensus AAGTATCACT and is primarily produced in blood cells. This illustrates another option found in eukaryotic control that is not found in prokaryotes: tissue-specific gene expression. Genes, DNA sequences that code for proteins, are technically available to any and every cell, but obviously the needs of a blood cell differ a great deal from the needs of a liver cell, or a neuron. Therefore, each cell may produce transcription factors that are specific to its cell or tissue type. These transcription factors can then allow or repress expression of multiple genes that help define this particular cell type, assuming they all have the recognition sequences for the TFs. These recognition sequences are also known as response elements (RE).



**Figure 5-10: Eukaryotic transcription regulation.** Eukaryotic transcription factors can work in complex combinations. In this Figure, the transcription factors hanging downward are representative of repressors (inhibitory TFs), while those riding upright on the DNA are considered activators. Thus, the RNA polymerase (RNAP) in (A) has a lower probability of transcribing this gene, while the RNAP in (B) is more likely to, perhaps because the TF nearest the promoter interacts with the RNAP to stabilize its interactions with TFIID. In this way, the same gene may be expressed in very different amounts and at different times depending on the transcription factors expressed in a particular cell type.

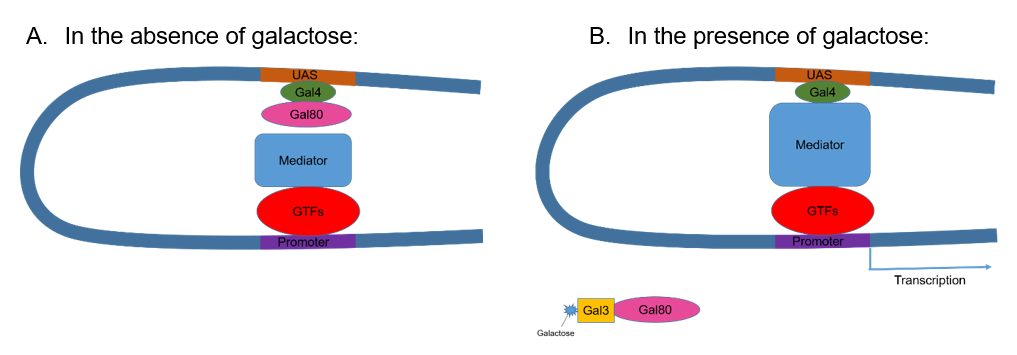
Very often, a combination of many transcription factors, both **activators** and **repressors**, are responsible for the ultimate expression rate of a given eukaryotic gene. This is referred to as **combinatorial control**. This can be done in a graded fashion, in which expression becomes stronger or weaker as more activators or repressors are bound, respectively, or it can be a binary mode of control, in which a well-defined group of TFs are required to turn on transcription, and missing just one can effectively shut down transcription entirely. In the first case, activating TFs generally bind to the GTFs or RNA Polymerase II directly to help them recognize the promoter more efficiently or stably, while repressing TFs may bind to the activating TFs, or to the GTFs or RNAP II, in preventing recognition of the promoter, or destabilizing the RNAP II preinitiation complex. In the second case, activation hinges on the building of an enhanceosome, in which transcription factors, protein scaffolding elements and **coactivators**, proteins that link upstream elements to the promoter region, come together to position and stabilize the preinitiation complex and RNAP II on the promoter. The most prominent and nearly ubiquitous coactivator is named **Mediator**, and binds to the C terminal domain of the β’ subunit of RNA polymerase II and also to a variety of transcription factors (Figure 5-11).



**Figure 5-11: Transcription initiation.** An enhancer can stabilize or recruit components of the transcription machine through a coactivator protein (one common version of a coactivator is the mediator).

## Yeast Gal4 as an example of eukaryotic transcription regulation

Yeasts are unicellular fungi that typically metabolize maltose as a primary source of energy. In the event that the preferred sugars are not available, there are necessary enzymes that must be transcribed and translated in order to enable the yeast to metabolize different sugars. When yeasts are cultured in media containing exclusively galactose, there is a greater than 1000-fold increase in the mRNA levels (i.e. increased transcription) of five enzymes required for galactose metabolism. For example, the preparatory reaction for galactose metabolism requires galactokinase, which phosphorylates galactose to produce galactose-1-phosphate. When galactose is absent, then it is not an efficient use of the cell’s resources to produce these enzymes, and thus the transcription is repressed by the repressor protein Gal80 (Fig 5-12A). In the presence of galactose, these enzymes are necessary and thus transcription must be stimulated by the Gal4 protein, which functions as a transcriptional activator. Gal4 binds to the UAS (Upstream Activating Sequence) which is located upstream of the TATA box promoter site of genes involved in galactose metabolism. The Gal4 protein has two major functional domains, the UAS*GAL* domain (DNA binding domain) at the N-terminus and the transcription activation domain at the C-terminus. The transcription activation domain is recognized and bound by Gal80 when transcription should be repressed in the absence of galactose. In the presence of galactose, galactose binds to the transcriptional regulator Gal3, which then in turn binds the repressor Gal80. This releases Gal80 from blocking the Gal4 transcriptional activator (Fig 5-12B). The transcription machinery in Figure 5-12 includes Mediator, which is a multiprotein complex that is required for transcription. Mediator functions as a coactivator by interacting with transcription factors, DNA, and RNA polymerase to facilitate transcription. In yeasts with a mutant form of Gal4 that is nonfunctional, there is no activation of galactose metabolism genes and these yeasts are unable to metabolize galactose.



**Figure 5-12: Transcriptional regulation by Gal4.** A) In the absence of galactose, the activator protein Gal4 binds the UAS (Upstream activating sequence) and interaction with the transcription machinery is inhibited by the repressor protein Gal80. B) In the presence of galactose, galactose binds Gal3 which, in turn, binds Gal80, allowing Gal4 to activate the transcriptional machinery.

## Post-transcriptional regulation of gene expression

It is important to note that the regulation of gene expression continues beyond the initial transcription of the mRNA. Expression can be upregulated or downregulated at the level of RNA processing, transport of the mRNA out of the nucleus, translation, and post-translational modification. Regulation at these various levels can allow the eukaryotic cell slower or quicker responses to external stimuli as appropriate (See Fig 5-1).