Control of gene expression and enzyme differentiation

1] Control of Gene Expression:

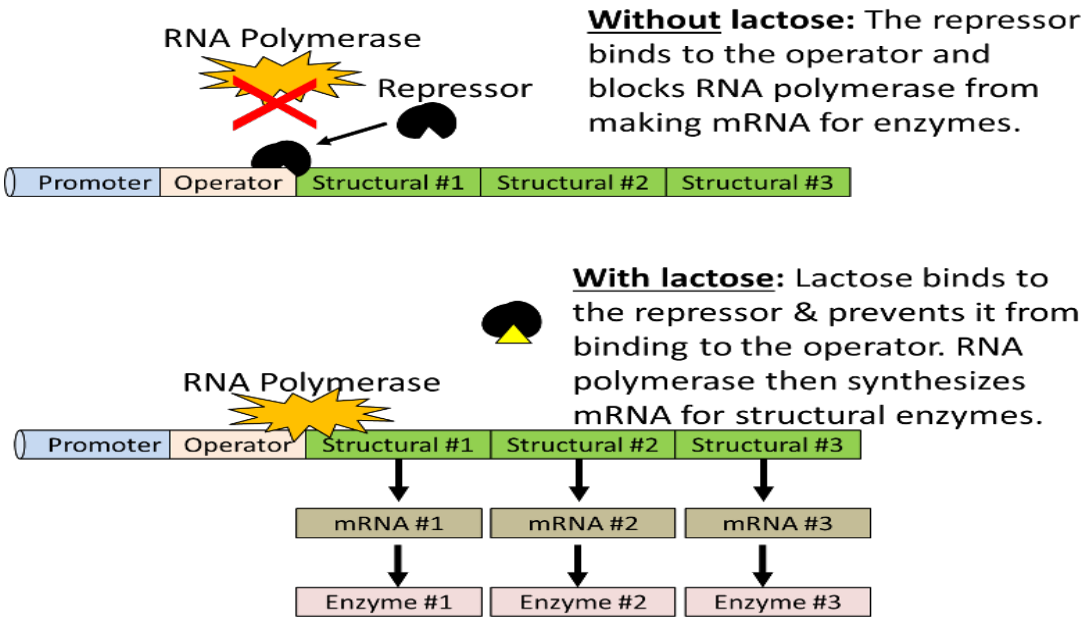
By gene expression we mean the transcription of a gene into mRNA and its subsequent translation into protein. Gene expression is primarily controlled at the level of transcription, largely as a result of binding of proteins to specific sites on DNA. In 1965 Francois Jacob, Jacques Monod, and Andre Lwoff shared the Nobel prize in medicine for their work supporting the idea that control of enzyme levels in cells is regulated by transcription of DNA. occurs through regulation of transcription, which can be either induced or repressed. These researchers proposed that production of the enzyme is controlled by an "operon," which consists a series of related genes on the chromosome consisting of an operator, a promoter, a regulator gene, and structural genes.

- The ***structural genes*** contain the code for the proteins products that are to be produced. Regulation of protein production is largely achieved by modulating access of RNA polymerase to the structural gene being transcribed.
- The ***promoter gene*** doesn't encode anything; it is simply a DNA sequence that is initial binding site for RNA polymerase.
- The ***operator gene*** is also non-coding; it is just a DNA sequence that is the binding site for the repressor.
- The ***regulator gene*** codes for synthesis of a *repressor* molecule that binds to the operator and blocks RNA polymerase from transcribing the structural genes.

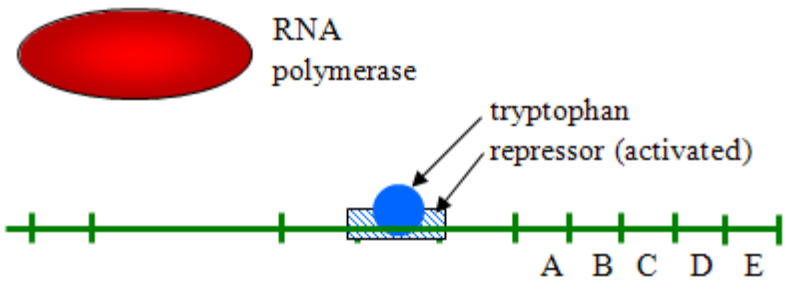
The operator gene is the sequence of non-transcribable DNA that is the repressor binding site. There is also a regulator gene, which codes for the synthesis of a repressor molecule hat binds to the operato.

- **Example of Inducible Transcription:** The bacterium E. coli has three genes that encode for enzymes that enable it to split and metabolize lactose (a sugar in milk). The promoter is the site on DNA where RNA polymerase binds in order to initiate transcription. However, the enzymes are usually present in very low concentrations, because their transcription is inhibited by a repressor protein produced by a regulator gene (see the top portion of the figure below). The repressor protein binds to the operator site and inhibits transcription. However, if lactose is present in the environment, it can bind to the repressor protein and *inactivate* it, effectively removing the blockade and enabling

transcription of the messenger RNA needed for synthesis of these genes (lower portion of the figure below).



- **Example of Repressible Transcription**: E. coli need the amino acid tryptophan, and the DNA in E. coli also has genes for synthesizing it. These genes generally transcribe continuously since the bacterium needs tryptophan. However, if tryptophan concentrations are high, transcription is repressed (turned off) by binding to a repressor protein and *activating* it as illustrated below.



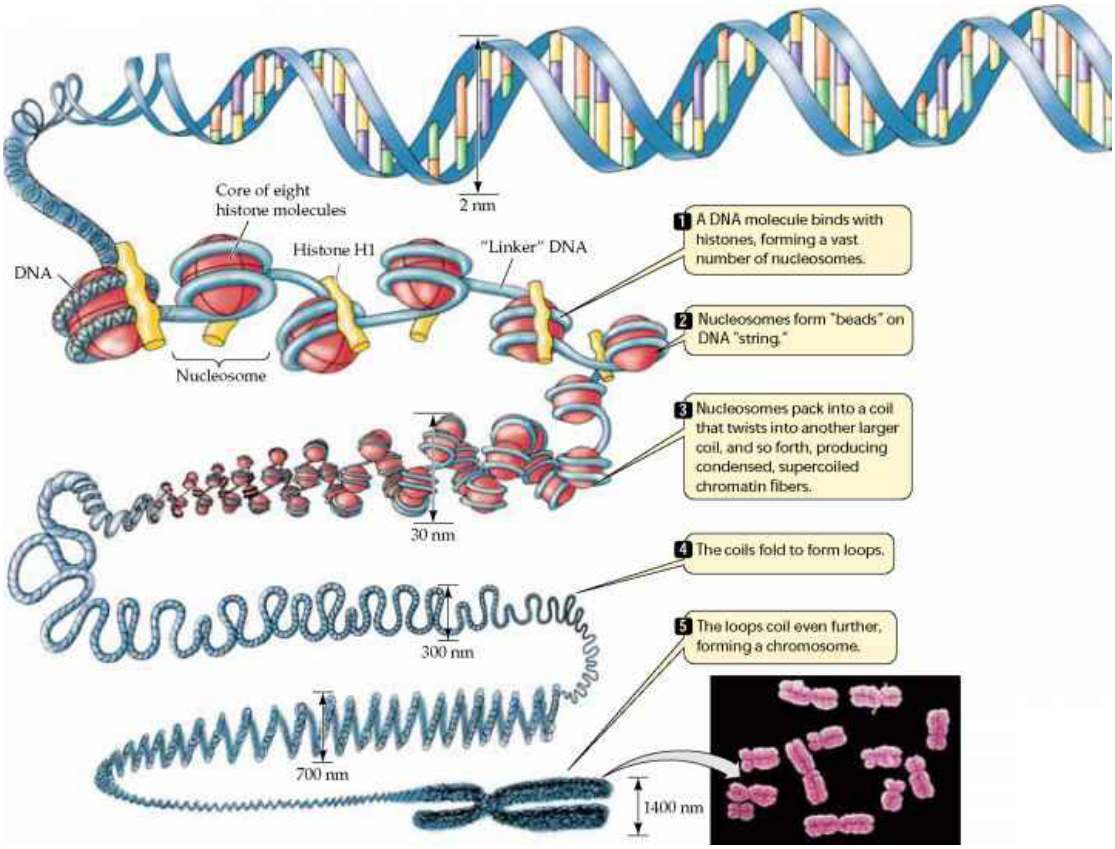
Source:
http://biowiki.ucdavis.edu/Under_Construction/BioStuff/BIO_101/Reading_and_Lecture_Notes/Control_of_Gene_Expression_in_Prokaryotes

2] Control of Gene Expression in Eukaryotes:

Eukaryotic cells have similar mechanisms for control of gene expression, but they are more complex. Consider, for example, that prokaryotic cells of a given species are all the same, but most eukaryotes are multicellular organisms with many cell types, so control of gene expression is much more complicated. Not surprisingly, gene

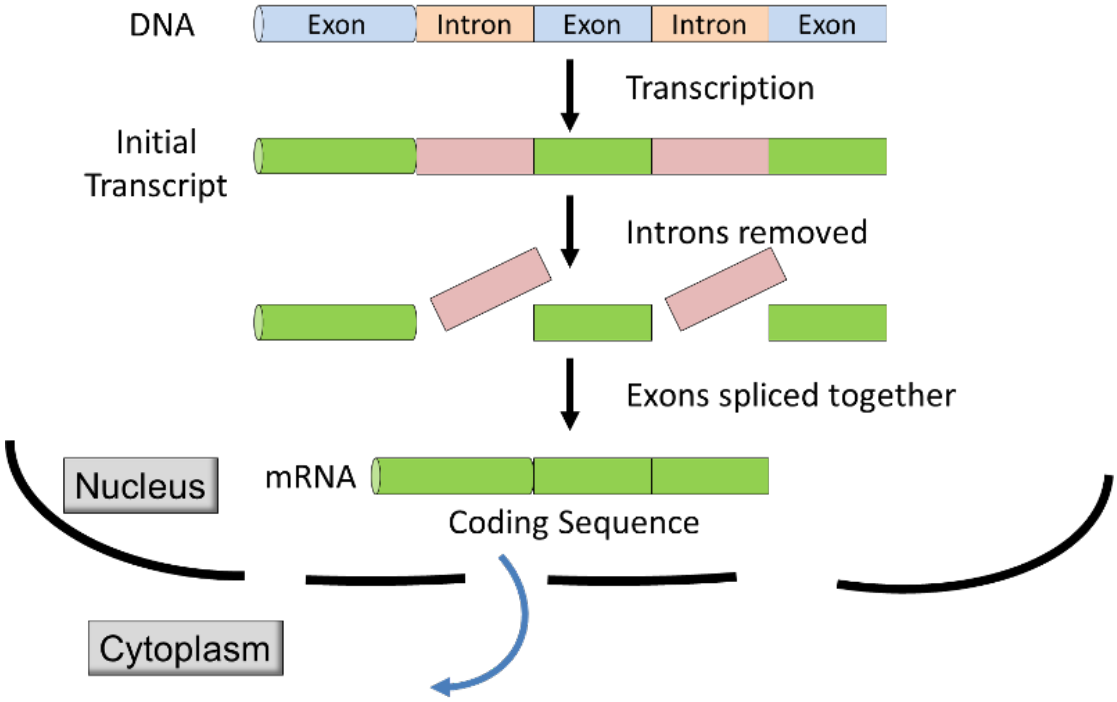
expression in eukaryotic cells is controlled by a number of complex processes which are summarized by the following list.

- After fertilization, the cells in the developing embryo become increasingly specialized, largely by turning on some genes and turning off many others. Some cells in the pancreas, for example, are specialized to synthesize and secrete digestive enzymes, while other pancreatic cells (β-cells in the islets of Langerhans) are specialized to synthesis and secrete insulin. Each type of cell has a particular pattern of expressed genes. This **differentiation** into specialized cells occurs largely as a result of turning off the expression of most genes in the cell; mature cells may only use 3-5% of the genes present in the cell's nucleus.
- Gene expression in eukaryotes may also be regulated through by alterations in the packing of DNA, which modulates the access of the cell's transcription enzymes (e.g., RNA polymerase) to DNA. The illustration below shows that chromosomes have a complex structure. The DNA helix is wrapped around special proteins called histones, and this are wrapped into tight helical fibers. These fibers are then looped and folded into increasingly compact structures, which, when fully coiled and condensed, give the chromosomes their characteristic appearance in **metaphase**.



Source: http://www.78stepshealth.us/plasma-membrane/eukaryotic-chromosomes.html

- Similar to the operons described above for prokaryotes, eukaryotes also use regulatory proteins to control transcription, but each eukaryotic gene has its own set of controls. In addition, there are many more *regulatory proteins* in eukaryotes and the interactions are much more complex.
- In eukaryotes transcription takes place within the membrane-bound nucleus, and the initial transcript is modified before it is transported from the nucleus to the cytoplasm for translation at the ribosome s. The initial transcript in eukaryotes has coding segments (**exons**) alternating with non-coding segments (introns). Before the mRNA leaves the nucleus, the introns are removed from the transcript by a process called RNA splicing (see graphic & video below), and extra nucleotides are added to the ends of the transcript; these non-coding

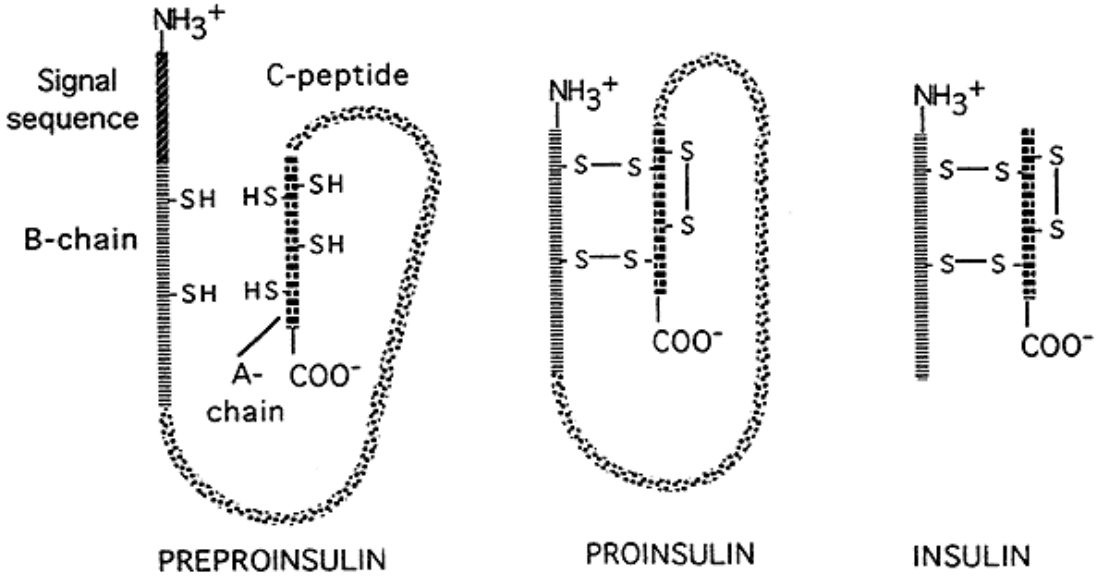


"caps" and "tails" protect the mRNA from attack by cellular enzymes and aid in recognition by the ribosomes.

Source: http://unmug.com/category/biology/organisation-control-of-genome/

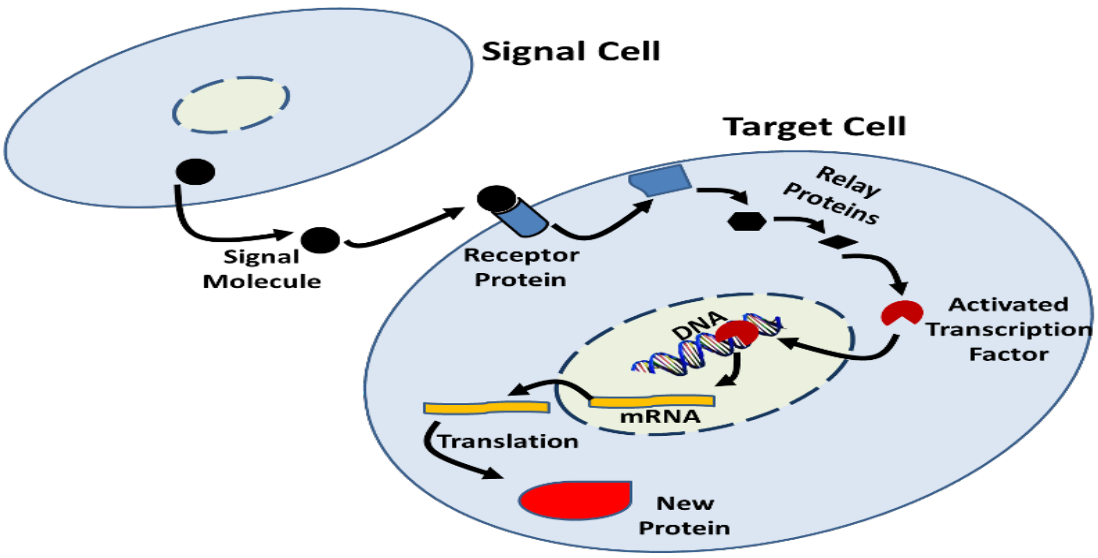
- Variation in the longevity of mRNA provides yet another opportunity for control of gene expression. Prokaryotic mRNA is very short-lived, but eukaryotic transcripts can last hours, or sometimes even weeks (e.g., mRNA for hemoglobin in the red blood cells of birds).
- The process of translation offers additional opportunities for regulation by many proteins. For example, the translation of hemoglobin mRNA is inhibited unless iron-containing heme is present in the cell.
- There are also opportunities for "post-translational" controls of gene expression in eukaryotes. Some translated polypeptides (proteins) are cut by enzymes into smaller, active final products. as illustrated in the figure below which depicts post-translational processing of the hormone insulin. Insulin is initially translated as a large, inactive precursor; a signal sequence is removed from the head of the precursor, and a large central portion (the C-chain) is cut

away, leaving two smaller peptide chains which are then linked to each other by disulfide bridges.The smaller final form is the active form of insulin.



Source:
http://www.nbs.csudh.edu/chemistry/faculty/nsturm/CHE450/19_InsulinGlucagon.htm

- Gene expression can also be modified by the breakdown of the proteins that are produced. For example, some of the enzymes involved in cell metabolism are broken down shortly after they are produced; this provides a mechanism for rapidly responding to changing metabolic demands.
- Gene expression can also be influenced by signals from other cells. There are many examples in which a signal molecule (e.g., a hormone) from one cell binds to a receptor protein on a target cell and initiates a sequence of biochemical changes (a signal transduction pathway) that result in changes within the target cell. These changes can include increased or decreased transcription as illustrated in the figure below.



Source: http://sites.saschina.org/emily01px2016/2014/11/23/a-variety-of-intercellular-and-intracellular-signal-transmissions-mediate-gene-expression/

- The RNA Interference system (RNAi) is yet another mechanism by which cells control gene expression by shutting off translation of mRNA. RNAi can also be used to shut down translation of viral proteins when a cell is infected by a virus. The RNAi system also has the potential to be exploited therapeutically.

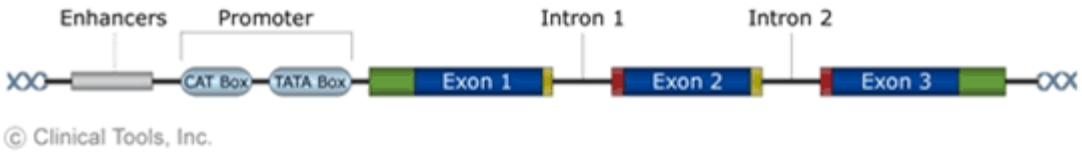
3] The process of gene expression involves two main stages:

Transcription: the production of messenger RNA (mRNA) by the enzyme RNA polymerase, and the processing of the resulting mRNA molecule.

Translation: the use of mRNA to direct protein synthesis, and the subsequent post-translational processing of the protein molecule.

Some genes are responsible for the production of other forms of RNA that play a role in translation, including transfer RNA (tRNA) and ribosomal RNA (rRNA).

4] A structural gene involves a number of different components:



- Exons. Exons code for amino acids and collectively determine the amino acid sequence of the protein product. It is these portions of the gene that are represented in final mature mRNA molecule.
- Introns. Introns are portions of the gene that do not code for amino acids, and are removed (spliced) from the mRNA molecule before translation.

5] Gene control regions

- **Start site**. A start site for transcription.
- **A promoter**. A region a few hundred nucleotides 'upstream' of the gene (toward the 5' end). It is not transcribed into mRNA, but plays a role in controlling the transcription of the gene. Transcription factors bind to specific nucleotide sequences in the promoter region and assist in the binding of RNA polymerases.
- **Enhancers**. Some transcription factors (called activators) bind to regions called 'enhancers' that increase the rate of transcription. These sites may be thousands of nucleotides from the coding sequences or within an intron. Some enhancers are

conditional and only work in the presence of other factors as well as transcription factors.

- **Silencers**. Some transcription factors (called repressors) bind to regions called 'silencers' that depress the rate of transcription.

Transcription

> Transcription is the process of RNA synthesis, controlled by the interaction of promoters and enhancers. Several different types of RNA are produced, including **messenger RNA**(**mRNA**), which specifies the sequence of amino acids in the protein product, plus **transfer RNA** (**tRNA**) and **ribosomal RNA** (**rRNA**), which play a role in the translation process.

Transcription involves four steps:

1. **Initiation**. The DNA molecule unwinds and separates to form a small **open complex**. RNA polymerase binds to the promoter of the **template strand**.
2. **Elongation**. RNA polymerase moves along the template strand, synthesising an mRNA molecule. In prokaryotes RNA polymerase is a holoenzyme consisting of a number of subunits, including a **sigma factor** (transcription factor) that recognises the promoter. In eukaryotes there are three RNA polymerases: I, II and III. The process includes a proofreading mechanism.
3. **Termination**. In prokaryotes there are two ways in which transcription is terminated. In **Rho-dependent termination**, a protein factor called "Rho" is responsible for disrupting the complex involving the template strand, RNA polymerase and RNA molecule. In **Rho-independent termination**, a loop forms at the end of the RNA molecule, causing it to detach itself. Termination in eukaryotes is more complicated, involving the addition of additional adenine nucleotides at the 3' of the RNA transcript (a process referred to as **polyadenylation**).
4. **Processing**. After transcription the RNA molecule is processed in a number of ways: introns are removed and the exons are spliced together to form a mature mRNA molecule consisting of a single protein-coding sequence. RNA synthesis involves the normal base pairing rules, but the base thymine is replaced with the base **uracil**.

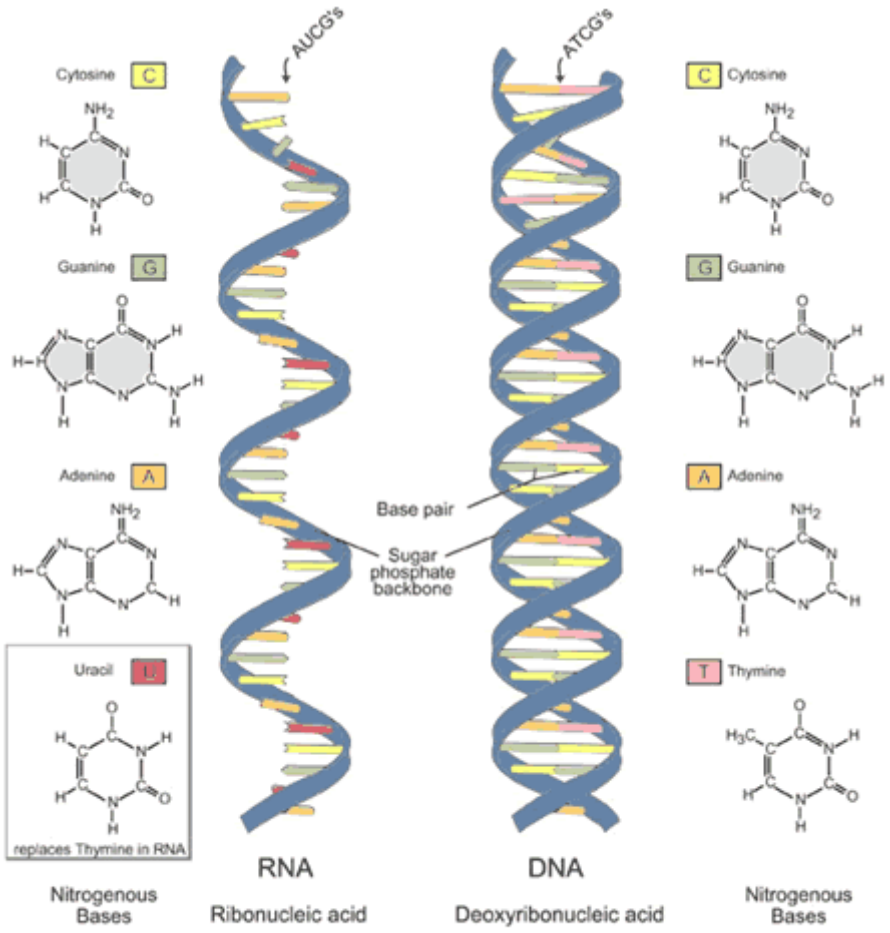


Image adapted from: National Human Genome Research Institute. Talking Glossary of Genetic Terms. Available at: www.genome.gov/Pages/Hyperion//DIR/VIP/Glossary/Illustration/rna.shtml.

Translation

In translation the mature mRNA molecule is used as a template to assemble a series of amino acids to produce a polypeptide with a specific amino acid sequence. The complex in the cytoplasm at which this occurs is called a ribosome. Ribosomes are a mixture of ribosomal proteins and ribosomal RNA (rRNA), and consist of a large subunit and a small subunit.

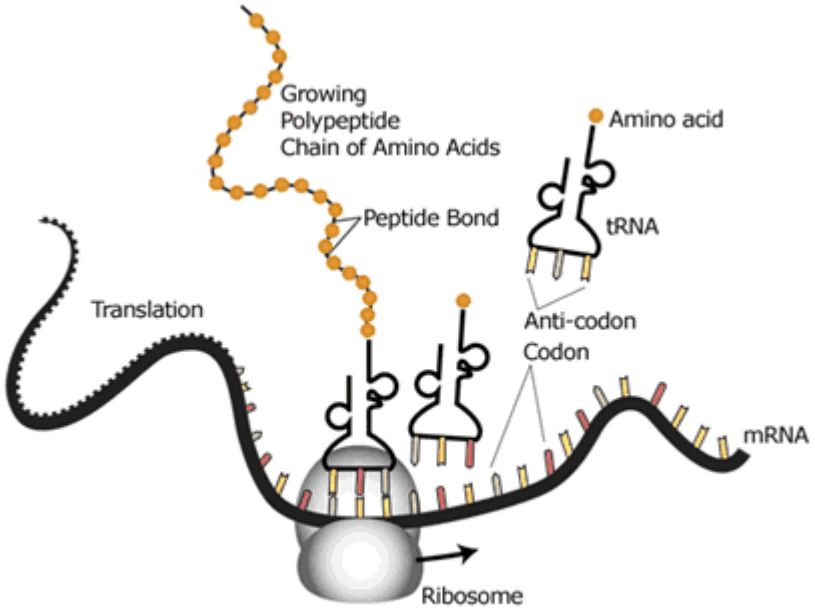


Image adapted from: National Human Genome Research Institute.

Translation involves four steps:

1. Initiation. The small subunit of the ribosome binds at the 5' end of the mRNA molecule and moves in a 3' direction until it meets a start codon (AUG). It then forms a complex with the large unit of the ribosome complex and an initiation tRNA molecule.
2. Elongation. Subsequent codons on the mRNA molecule determine which tRNA molecule linked to an amino acid binds to the mRNA. An enzyme peptidyl transferase links the amino acids together using peptide bonds. The process continues, producing a chain of amino acids as the ribosome moves along the mRNA molecule.
3. Termination. Translation in terminated when the ribosomal complex reached one or more stop codons (UAA, UAG, UGA). The ribosomal complex in eukaryotes is larger and more complicated than in prokaryotes. In addition, the processes of transcription and translation are divided in eukaryotes between the nucleus (transcription) and the cytoplasm (translation), which provides more opportunities for the regulation of gene expression.
4. Post-translation processing of the protein

Gene regulation

Gene regulation is a label for the cellular processes that control the rate and manner of gene expression. A complex set of interactions between genes, RNA molecules, proteins (including transcription factors) and other components of the expression

system determine when and where specific genes are activated and the amount of protein or RNA product produced.

Some genes are expressed continuously, as they produce proteins involved in basic metabolic functions; some genes are expressed as part of the process of cell differentiation; and some genes are expressed as a result of cell differentiation.

Mechanisms of gene regulation include:

- Regulating the rate of transcription. This is the most economical method of regulation.
- Regulating the processing of RNA molecules, including alternative splicing to produce more than one protein product from a single gene.
- Regulating the stability of mRNA molecules.
- Regulating the rate of translation.

Transcription factors are proteins that play a role in regulating the transcription of genes by binding to specific regulatory nucleotide sequences.

5] Cell-Extrinsic Regulation of Gene Expression

Gene expression is regulated by factors both extrinsic and intrinsic to the cell. Cell-extrinsic factors that regulate expression include environmental cues, such as small molecules, secreted proteins, temperature, and oxygen. These cues can originate from other cells within the organism, or they can come from the organism's environment. Within the organism, cells communicate with each other by sending and receiving secreted proteins, also known as growth factors, morphogens, cytokines, or signaling molecules. Receipt of these signaling molecules triggers intercellular signaling cascades that ultimately cause semipermanent changes in transcription or expression of genes. Such changes in gene expression can include turning genes completely on or off, or just slightly tweaking the level of transcript produced. This process is thought to regulate a vast number of cell behaviors, including cell fate decisions during embryogenesis, cell function, and chemotaxis.

In addition, gene expression changes can lead to changes in an entire organism, such as molting in insects. In *Drosophila*, for example, the molting process is regulated by levels of a hormone called ecdysone. This hormone acts as a signal, triggering a cascade of events and leading to changes in gene expression. Not surprisingly, the genes that are expressed in response to ecdysone are also the genes that are involved in the molting process. Thus, ecdysone acts on the organism level as a cell-extrinsic factor to bring about physiologically meaningful changes in gene expression.

What is also interesting is that scientists can learn more about a physiological process like metamorphosis by studying how gene expression patterns change over time. For example, although researchers were aware that ecdysone results in a

decrease of transcription from some loci, such as those involved in the glycolytic pathway, microarray data suggest that ecdysone-induced metamorphosis also downregulates genes involved with fatty acid oxidation, amino acid metabolism, oxidative phosphorylation, and other pathways. This suggests that there is a more global repression of metabolic activity during molting (Figure 2). Specifically, during metamorphosis, the larval muscle cells are degraded, and muscle-specific genes are downregulated (Figure 2B). Simultaneously, the development of the nervous system begins, and the genes involved in neuronal differentiation are induced (Figure 2C).

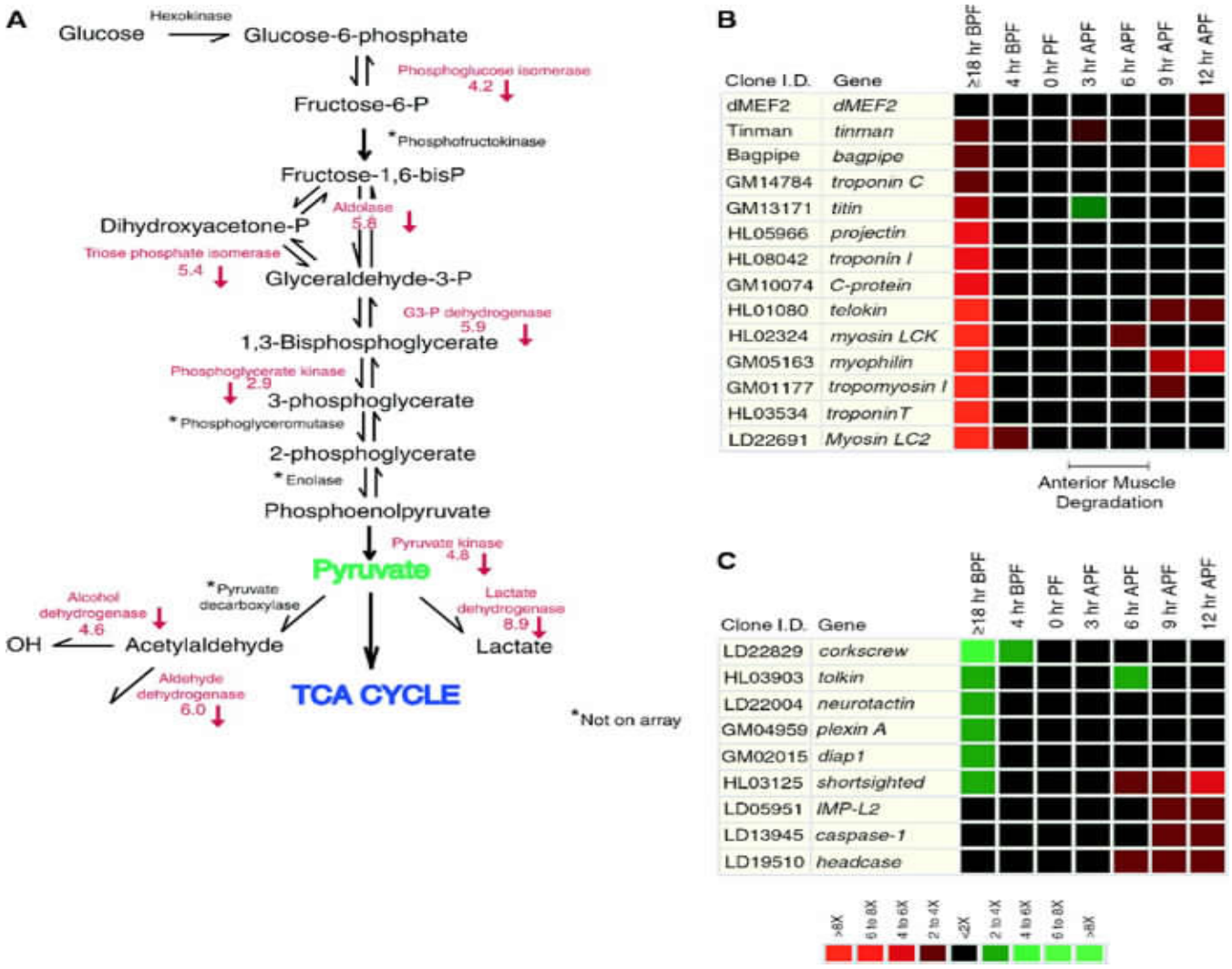


Figure 1: Microarray data collected at different times during metamorphosis reveals the effects of the ecdysone pulse on many downstream genetic pathways.
(A) Changes in ecdysone levels affect the glycolytic pathway. Levels of a number of enzymes involved in this pathway are decreased as a result of the ecdysone pulse; these enzymes are listed in red next to the reactions they catalyze. (B) This array shows expression changes in various structural and regulatory genes involved in muscle formation (myogenesis) in response to the ecdysone pulse. (C) This array shows how the ecdysone pulse alters expression of multiple genes involved in central

nervous system restructuring, apoptosis, and cellular differentiation during metamorphosis. In both of the microarrays, red means that the gene was downregulated, while green means that the gene was upregulated. Expression levels were measured at various points before and after pupal formation (PF). (BPF = before pupal formation; APF = after pupal formation).

6] Cell-Intrinsic Regulation of Gene Expression

Although differentiation is not thought to occur by permanent loss of genetic material, DNA can be modified in a way that affects gene expression. For instance, DNA and its associated histone proteins (together known as chromatin) can be chemically modified by a cell's own machinery. Chromatin modification can affect gene expression by changing the accessibility of genes to transcription factors, in either a positive or a negative manner. Two major classes of such chemical modifications include DNA methylation and histone modification (methylation and/or acetylation). These changes are often described as epigenetic because they do not act to alter the primary DNA sequence but instead act at a level just above the DNA sequence. Although DNA methylation and histone modification are not genetic, cells have mechanisms to copy this epigenetic information during their division so that their daughter cells contain the same regulatory data

Changes in chromatin modification play an important role in regulating gene expression during developmental cell-type specification as well. For example, chromatin-modifying proteins play an essential role in muscle cell differentiation via interactions with key muscle-promoting transcription factors MyoD and MEF. That is, these factors are thought to help recruit chromatin modifying factors, such as histone acetyltransferases and deacetylases. In so doing, MyoD and MEF alter access to their target sites upstream of muscle differentiation genes. For instance, MyoD binds histone acetyltransferases p300 and PCAF, and this activity is essential for muscle cell differentiation . This example provides evidence for a link among chromatin modifications, transcription factors, and, ultimately, cell-fate-specific changes in gene expression.

Chromatin modification can be stable over the life of an organism, thereby effectively permanently influencing gene expression. However, that is not to say that chromatin modification is irreversible. For instance, chromatin can become mismodified in certain cancers , suggesting that, although important, the change is not permanent. Moreover, chromatin modifications are usually erased and reset during the production of gametes, such that the adult program of intrinsic cues is replaced with a program more suited to embryonic development.

In fact, embryonic cell types are known to contain a unique set of chromatin modifications that are different from those found in adult cell types. This has led to the tantalizing proposal that chromatin modification helps lock in changes in gene expression that are required during development. The permanent silencing of the genes involved only in embryogenesis could then drive the development of cells toward more mature cell types. By blocking accessibility of transcription machinery, for example, chromatin modification could prevent the need for continued repression

through active binding of a repressive transcription factor. Alternatively, the genes required for an adult cell type might contain chromatin modifications (especially histone acetylation) that cause the DNA to become open and, therefore, more accessible to the transcription machinery.

Interestingly, embryonic cell types have been found to contain a signature chromatin modification in the regions that regulate the expression of genes involved in early embryonic development . Such regions were found to contain chromatin modifications with both silencing and promoting characteristics. The finding of these bivalent (two-directional) markers in association with genes important for embryonic development has led to the belief that embryonic cells exist in a special epigenetic state, wherein they can choose to remain embryonic (as in an embryonic stem cell) or to differentiate (as in normal development), and bivalent domains provide a means by which to quickly choose between the two options.

Together, these lines of evidence have led to an emerging hypothesis that cell-cell signaling and epigenetic changes converge to guide cell differentiation decisions both during development and beyond.

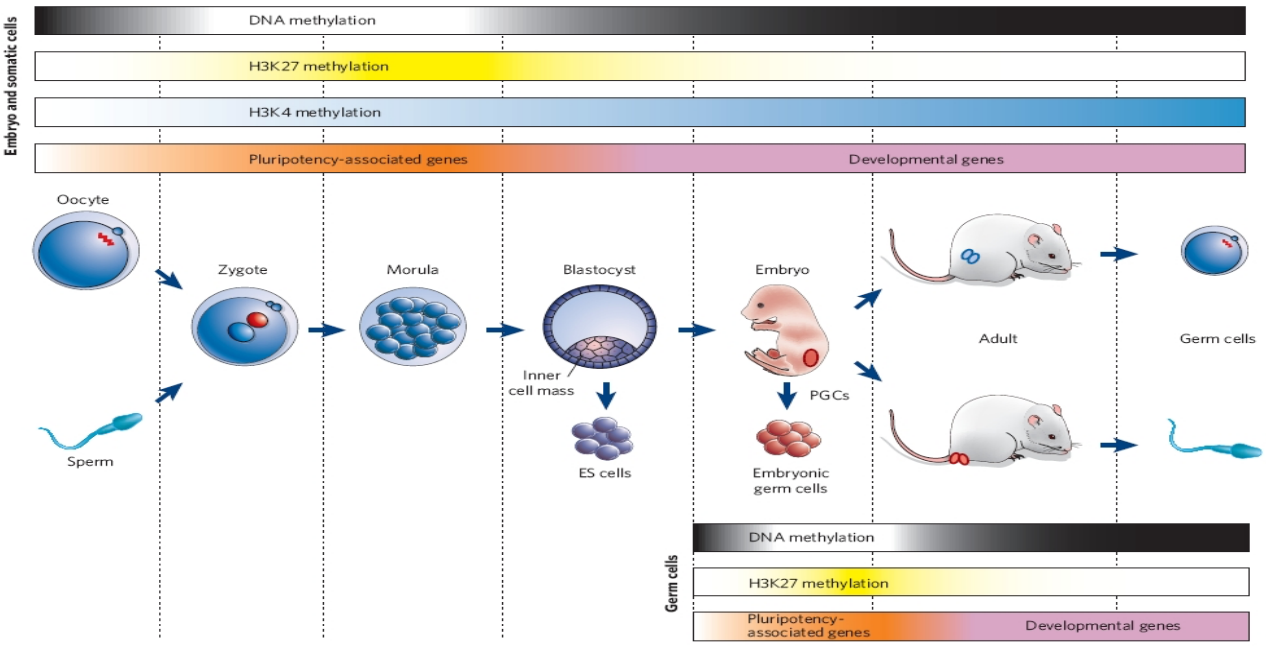


Figure 2 : Epigenetic gene regulation during mammalian development.

This figure depicts key developmental events together with global epigenetic modifications and gene expression patterns. Very early in development, DNA methylation is erased. In addition, pluripotency-associated genes begin to be expressed, and developmental genes are repressed by the PcG protein system and H3K27 methylation. During the differentiation of pluripotent cells such as embryonic stem (ES) cells, pluripotency-associated genes are repressed, potentially permanently, as a result of DNA methylation. At the same time, developmental genes begin to be expressed, and there is an increase in H3K4 methylation. During the early development of primordial germ cells (PGCs), DNA methylation and repressive

histone modifications (such as H3K9 methylation) are also erased. Pluripotency-associated genes are re-expressed during a time window that allows embryonic germ cells to be derived in culture. Imprinted genes are demethylated during this period, and developmental genes are expressed afterwards. Flexible histone marks such as H3K27 methylation enable developmental genes to be silenced for a short time in pluripotent cells. By contrast, DNA methylation enables the stable silencing of imprinted genes, transposons, and some pluripotency-associated genes.

7] gene control by metabolites and metabolic enzymes

Metabolism and gene regulation are two fundamental biological processes that are essential to all living organisms. Homeostasis, cell growth, and differentiation all require the coordination of metabolic state and gene expression programs. Nevertheless, how expression of the genome adapts to metabolic state or environmental changes is not well understood. One level of regulation involves signal transduction pathways that control key transcription factors that act as master regulators of gene expression programs. In addition, post-translational modifications (PTMs) of chromatin play a major role in the activation or repression of gene transcription. These include acetylation, methylation, and phosphorylation of the histones and DNA methylation. Some of these chromatin modifications are involved in the maintenance of stable patterns of gene expression, usually referred to as epigenetic regulation. Pertinently, the activity of enzymes that modulate chromatin is critically dependent on central metabolites as cofactors or cosubstrates. Thus, the availability of metabolites that are required for the activity of histone-modifying enzymes may connect metabolism to chromatin structure and gene expression. Finally, selective metabolic enzymes act in the nucleus to adjust gene transcription in response to changes in metabolic state. Here, we review the interface between metabolism and gene transcription. We focus on the molecular mechanisms involved and discuss unresolved issues and implications for development and disease.

The basics of metabolism and gene expression control

Metabolism is the total of all chemical reactions in cells and organisms that maintain life. Metabolism can be divided into two classes: catabolic processes (the breakdown of molecules that usually results in the release of energy) and anabolic processes (the synthesis of components such as proteins, lipids, and nucleic acids, which costs energy). Cellular (or intermediary) metabolism is organized in separate chemical pathways formed by a chain of linked enzymatic reactions in which the product of one step is the substrate for the next. A small set of cofactors, nonprotein compounds that are required for an enzymatic activity, is used in a multitude of reactions to mediate the transfer of chemical groups for abbreviations of key metabolites and metabolic enzymes). A classic example is nicotinamide (NAM) adenine dinucleotide (NAD), which exists in an oxidized (NAD^+) or reduced (NADH) form. NAD^+ is used as an electron acceptor by dehydrogenases in a wide variety of pathways to remove electrons from their substrate, producing NADH. Conversely, NADH serves as an electron donor for reductases (yielding NAD^+). Thus, cofactors like NAD^+/NADH are used broadly and recycled continuously in metabolic reactions. In eukaryotic cells, the

functional specialization of metabolic pathways is enabled by compartmentalization. For example, oxidative phosphorylation and the tricarboxylic acid (TCA) cycle take place on the inner membrane of mitochondria, but glycolysis occurs in the cytosol. The bulk of intermediary metabolism happens in the cytoplasm, whereas the nucleus is largely dedicated to the replication, maintenance, and expression of the eukaryotic genome (Fig. 1).

Control of gene expression underlies cell differentiation and development and allows a cell to respond to signals and environmental changes. Gene transcription by DNA-dependent RNA polymerases is the primary level at which genes are regulated. Eukaryotic nuclear genes are transcribed by three different RNA polymerases (polymerase I [Pol I], Pol II, and Pol III). RNA Pol I transcribes most ribosomal RNA genes, whereas RNA Pol III transcribes mainly tRNAs, 5S RNA, and some additional small RNAs. Thus, Pol I and Pol III transcription is responsible for major components of the protein synthesis machinery and is closely linked to nutrient availability and cell growth. RNA Pol II and a set of auxiliary general (or basal) transcription factors are responsible for the expression of all protein-encoding genes and a diverse group of regulatory noncoding RNA genes. The differential regulation of RNA Pol II initiation and elongation is fundamental to gene expression during development, homeostasis, and disease. Sequence-specific DNA-binding transcription factors that bind promoters and enhancer elements impose gene selectivity on RNA Pol II transcription. These transcription factors function through the recruitment of transcriptional corepressors or coactivators to regulatory DNA elements. Many of these transcriptional coregulators target the structure of chromatin to either enable or block gene transcription.

The packaging of eukaryotic genomic DNA into chromatin constitutes a major level of gene transcription. The nucleosome is the basic repeat unit of eukaryotic chromatin, comprising 147 base pairs (bp) of DNA wrapped tightly around a protein core formed by two copies each of the core histones: H2A, H2B, H3, and H4. Put simply, nucleosomes create a barrier to the accessibility of genes and regulatory DNA sequences, thereby controlling transcription. Consequently, by mediating the assembly, sliding, restructuring, or ejection of nucleosomes, ATP-dependent chromatin remodeling factors play a central role in the regulation of gene expression). A distinct biochemical mechanism involves a plethora of PTMs of specific residues within the histone N-terminal tails, which protrude from the nucleosome. These modifications, including acetylation, methylation, and phosphorylation, can modulate the folding of the chromatin fiber or direct recruitment of regulatory proteins, such as transcription factors or other chromatin regulators.

Transcriptional activity is intimately associated with specific histone modifications. For example, acetylation of histone H3 Lys27 (H3K27ac) marks active genes, whereas trimethylation of the same residue (H3K27me3) by the Polycomb-repressive complex 2 (PRC2) leads to gene silencing. Histone acetylation is generally associated with active chromatin irrespective of which residue is modified. In contrast, the consequences of histone methylation are determined by the specific residue that is targeted or even by the number of methyl groups added. Methylation of H3K4, H3K36, or H3K79 is usually associated with active transcription, whereas

methylation of H3K9, H3K20, or H3K27 marks transcriptionally repressed chromatin. The recognition of histone PTMs by specific protein domains is important for the formation of repressive or active chromatin structures. For example, the bromodomain acts as an acetyl-lysine-binding module, enhancing the recruitment of transcription factors to chromatin. In contrast, the binding of the chromodomain of HP1 to H3K9me3 mediates heterochromatin formation. It is important to note that chromatin PTMs are dynamic; they are continuously placed and removed by antagonizing sets of enzymes. Moreover, enzymes that modify histones also target and modulate the activity of transcription factors (and other proteins). Thus, caution should be used in equating the regulatory role of a chromatin-modifying enzyme to a particular histone PTM. Reflecting their key role in gene control, cancer genome sequencing studies revealed that chromatin-modulating enzymes are frequently mutated in human cancers .
The DNA itself is also subject to modifications. In particular, methylation of cytosines (5-methylcytosine [5mC]) by DNA methyltransferases (DNMTs) within CpG islands in vertebrate genomes is linked tightly to gene transcription. CpG islands, the most common promoter motifs in vertebrate genomes, are GC-rich DNA sequences of several hundreds of base pairs that are highly enriched for CpG dinucleotides. CpG promoters are stably silenced by a high level of CpG methylation. Many tumors are characterized by aberrant CpG island methylation, which is referred to as a CpG island methylator phenotype. DNA methylation can be reversed passively (i.e., lost during several rounds of DNA replication) or removed actively by the ten-eleven translocation (TET) enzymes). TET enzymes oxidize 5mC to 5-hydroxymethylcytosine (5hmC), which is the first step in active DNA demethylation. Moreover, 5hmC may function as a chromatin mark, but more research on this potential function is needed.

In summary, sequence-specific transcription factors provide gene specificity but act through the recruitment of coregulators. Many of these coregulators function by modulating chromatin structure; e.g., by opening it up to allow DNA access to other transcriptional regulators. Furthermore, histone PTMs can be recognized by selective regulatory proteins, including basal transcription factors or other histone-modulating enzymes. This way, chromatin modifications can lead to a cascade of events that either promote or block gene expression.