Brief content of lectures

Discipline "Genome regulation"

Lecturer Kenzhebaeva S.S.Doctor b.sc., professor

Lecture 1. Introduction. The structure of the bacterial chromosome. Features of the prokaryote genome structure (the polycistronic gene structure). Plasmids as prokaryote genome elements.

L.2 *Levels of the gene expression regulation*

L. 3 *Regulation of the gene expression at the transcriptional level in prokaryotes*

1. RNA Polymerase and Transcription
2. Repressors and Negative Control of Transcription
3. Positive Control of Transcription

Transcriptional Attenuation

The transcriptional machinery /or transcriptional apparatus/ - set of proteins required to carry out and regulate transcription.

Components are divided into:

- basal machinery — directly responsible for promoter recognition and transcription
- regulatory machinery — controls the rate at which the basal machinery carries out its job in a gene-specific manner

Cis-elements The DNA segments with which the transcriptional machinery interact are termed *cis-elements*.

Cis-elements include promoters and binding sites for activators and repressors.

Clusters of cis-elements that act over long distances — enhancers (functioned in activation), silencers (in repression).

RNA polymerase functions

1. - catalyzes the polymerization of ribonucleoside 5′-triphosphates (NTPs) as directed by a DNA template.
2. - catalyzes the growth of RNA chains always in the 5′ to 3′ direction.
3. - does not require a preformed primer to initiate the synthesis of RNA. Instead, transcription initiates *de novo* at specific sites at the beginning of genes.

L. 4 *The stages of transcriptional cycle. Initiation, formation of "open complex", elongation and termination of transcription.*

Characterization of RNA polymerase

1. RNA polymerase does not need a primer to initiate transcription.
2. The RNA product does not remain base-paired to the template DNA strand.
3. Transcription is less accurate than replication.

Transcription in Prokaryotic cells

Transcription involves three stages

- *The first step* in RNA synthesis is the binding of RNA polymerase to a DNA promoter- to form what is called a closed complex.
- *In the second step*, the closed complex undergoes a transition to the open complex in which the DNA strands separate over a distance of some 14 bp around the start site.
- *Once an RNA polymerase molecule* has bound to a promoter site and locally unwound the DNA double helix, initiation of RNA synthesis can take place.

- Once the enzyme gets further than 10 bp, it is said to have escaped the promoter.
- A stable ternary complex contains enzyme, DNA and RNA. This is a transition to elongation phase.

The sigma factors mediates binding of polymerase to the promoter.
The key functions of sigma (σ) factors involve:
I) binding core RNAP to form the RNAP holoenzyme (Eσ),
II) recruiting Eσ to target promoters,
III) promoter melting at the transcription start site, and
IV) regulating transcription by clearing and releasing the promoter from RNAP
Anti-sigma factors

L. 5 *Prokaryotic RNA polymerase, its subunit and three-dimensional structure. A variety of sigma factors. The promoter of prokaryotic genes, its structural elements*

The elongation polymerase is a processive machine that synthesizes and proofreading RNA RNA polymerase carries out two proofreading functions Pyrophosphorolytic editing: the enzyme uses its active site to catalyze the removal of an incorrectly inserted NTP.
Hydrolytic editing: the polymerase backtracks by one or more nucleotides and cleaves the RNA product, removing the error-containing sequence.
Hydrolytic editing is stimulated by Gre factors, which also serves as elongation stimulating factors.
Transcription is terminated by signals within the RNA sequences
In bacteria, terminators come in two types: **rho-independent** and **rho-dependent**

L. 6 *The features of transcription in eukaryotes. RNA processing in eukaryotes. Caping, splicing and polyadenylation of the transcripts*

Eukaryotic cells have three different polymerase (Pol I, II and III)
Whereas bacteria require only one initiation factor, several initiation factors are required for efficient and promoter-specific initiation in eukaryotes. These are called the **general transcription factors (GTFs)**.

Three different eukaryotic RNA polymerases

RNA polymerase I resides in the nucleolus and is responsible for synthesizing three of the four types of rRNA found in eukaryotic ribosomes (28S, 18S,and 5.8 S rRNA).
RNA polymerase II is found in the nucleoplasm and synthesizes precursors to mRNA, the class of RNA molecules that code for proteins.
RNA polymerase III is also a nucleoplasmic enzyme, but it synthesizes a variety of small RNAs, including tRNA precursors and the smallest type of ribosomal RNA, 5S rRNA.

Regulatory sequences

Beyond the core promoter, there are other sequence elements required for efficient transcription in vivo.
- Together these elements constitute the **regulatory sequences:**

1. promoter proximal elements,
2. upstream activator sequences,
3. enhancers, silencers,
4. boundary elements and insulators.
5. All these elements bind regulatory proteins.

Mediator consists of many subunits, some conserved from yeast to human.
Polyadenylation and termination.
Pol III transcription.

L. 7 *Mechanisms of the splicing. The role of small nuclear RNAs and protein factors. The alternative splicing, its examples.*

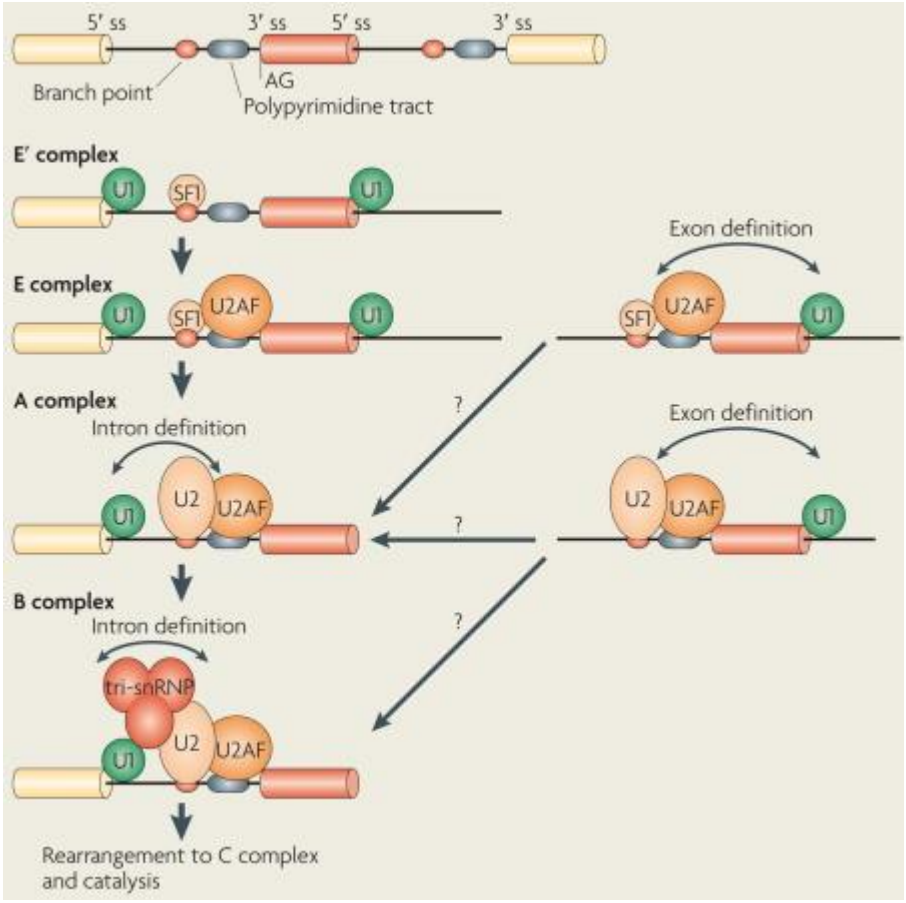
Alternative splicing of mRNA precursors provides an important means of genetic control and is a crucial step in the expression of most genes. Alternative splicing markedly affects human development, and its misregulation underlies many human diseases. Although the mechanisms of alternative splicing have been studied extensively, until the past few years we had not begun to realize fully the diversity and complexity of alternative splicing regulation by an intricate protein–RNA network. Great progress has been made by studying individual transcripts and through genome-wide approaches, which together provide a better picture of the mechanistic regulation of alternative pre-mRNA splicing.

Alternative splicing is a crucial mechanism for gene regulation and for generating proteomic diversity. Recent estimates indicate that the expression of nearly 95% of human multi-exon genes involves alternative splicing[1,2]. In metazoans, alternative splicing plays an important part in generating different protein products that function in diverse cellular processes, including cell growth, differentiation and death.

Splicing is carried out by the spliceosome, a massive structure in which five small nuclear ribonucleoprotein particles (snRNPs) and a large number of auxiliary proteins cooperate to accurately recognize the splice sites and catalyse the two steps of the splicing reaction[1,2] (Fig. 1). Spliceosome assembly (Fig. 1) begins with the recognition of the 5′ splice site by the snRNP U1 and the binding of splicing factor 1 (SF1) to the branch point[3] and of the U2 auxiliary factor (U2AF) heterodimer to the polypyrimidine tract and 3′ terminal AG[4,5]. This assembly is ATP independent and results in the formation of the E complex, which is converted into the ATP-dependent, pre-spliceosomal A complex after the replacement of SF1 by the U2 snRNP at the branch point. Further recruitment of the U4/U6–U5 tri-snRNP complex leads to the formation of the B complex, which is converted into to the catalytically active C complex after extensive conformational changes and remodelling.

Pre-mRNA splicing is a process in which intervening sequences (introns) are removed from an mRNA precursor. Splicing consists of two transesterification steps, each involving a nucleophilic attack on terminal phosphodiester bonds of the intron. In the first step this is carried out by the 2′ hydroxyl of the branch point (usually adenosine) and in the second step by the 3′ hydroxyl of the upstream (5′) exon[1,2]. This process is carried out in the spliceosome, a dynamic molecular machine the assembly of which involves sequential binding and release of small nuclear ribonucleoprotein particles (snRNPs) and numerous protein factors as well as the formation and disruption of RNA–RNA, protein–RNA and protein–protein interactions.

The basic mechanics of spliceosome assembly are well known. Briefly, the process begins with the base pairing of U1 snRNA to the 5′ splice site (ss) and the binding of splicing factor 1 (SF1) to the branch point[3] in an ATP-independent manner to form the E′ complex (see the figure; double-headed arrows indicate an interaction). The E′ complex can be converted into the E complex by the recruitment of U2 auxiliary factor (U2AF) heterodimer (comprising U2AF65 and U2AF35) to the polypyrimidine tract and 3′ terminal AG. The ATP-independent E complex is converted into the ATP-dependent pre-spliceosome A complex by the replacement of SF1 by U2 snRNP at the branch point. Further recruitment of the U4/U6–U5 tri-snRNP leads to the formation of the B complex, which contains all spliceosomal subunits that carry out pre-mRNA splicing. This is followed by extensive conformational changes and remodelling, including the loss of U1 and U4 snRNPs, ultimately resulting in the formation of the C complex, which is the catalytically active spliceosome.



Splicing and spliceosome assembly

L. 8 The post-transcriptional regulation in procariots

Post-transcriptional regulation is a very important mechanism to control gene expression in changing environments. In the past decade, a lot of interest has been directed toward the role of small RNAs (sRNAs) in bacterial post-transcriptional regulation. However, sRNAs are not the only molecules controlling gene expression at this level, RNA-binding proteins (RBPs) play an important role as well. CsrA and Hfq are the two best studied bacterial proteins of this type, but recently, additional proteins involved in post-transcriptional control have been identified. This review focuses on the general working mechanisms of post-transcriptionally active RBPs, which include (i) adaptation of the susceptibility of mRNAs and sRNAs to RNases, (ii) modulating the accessibility of the ribosome binding site of mRNAs, (iii) recruiting and assisting in the interaction of mRNAs with other molecules and (iv) regulating transcription terminator/antiterminator formation, and gives an overview of both the well-studied and the newly identified proteins that are involved in post-transcriptional regulatory processes. Additionally, the post-transcriptional mechanisms by which the expression or the activity of these proteins is regulated, are described. For many of the newly identified proteins, however, mechanistic questions remain. Most likely, more post-transcriptionally active proteins will be identified in the future

Bacteria need to survive in constantly changing environments. Therefore, they must be able to alter their gene expression in response to environmental signals, causing protein levels to be adjusted according to the needs of the cell. This can be achieved by adjusting transcription initiation with sigma factors and proteins that activate or repress transcription. However, gene expression regulation also occurs after transcription is initiated. The importance of these post-transcriptional

regulatory processes is highlighted by the weak correlation that has been observed between RNA and protein abundance

Prokaryotic post-transcriptional regulators typically modulate RNA decay, translation initiation efficiency or transcript elongation. Different types of prokaryotic post-transcriptional regulators have been identified, including small RNAs (sRNAs) and RNA-binding proteins (RBPs). sRNAs are typically defined as non-coding RNA molecules that bind with limited complementarity near the ribosome binding site (RBS) of their target mRNA, causing competition with the ribosome for binding to this region. However, the number of sRNAs that deviate from this general definition is increasing. The new insights into the post-transcriptional mechanisms of sRNAs and their role in gene expression regulation were reviewed recently. Here, RBPs involved in post-transcriptional regulation are discussed. For some of these proteins, the mechanism of action and the targets are well described, as for CsrA and Hfq. Their post-transcriptional function in *Escherichia coli* was already reported almost 20 years ago. Lately, more insight was gained into the diverse mechanisms these two well-studied proteins use to regulate the expression of their target genes and how they regulate their own expression or activity in *E. coli* and in other bacteria. Additional RBPs involved in post-transcriptional regulation have been identified only recently and not much is known about their post-transcriptional function. In this review, the general working mechanisms of RBPs are discussed first. Afterward, examples of well-known and recently identified proteins, from *E. coli* and from other bacteria, are described.

General Mechanisms of Regulatory Proteins that Act Post-Transcriptionally

Bacterial post-transcriptionally active regulatory proteins typically bind RNA molecules and regulate translation initiation, stability, and transcript elongation of their RNA targets, using different regulatory mechanisms. These mechanisms include (i) adaptation of the susceptibility of the target RNAs to RNases, (ii) modulation of the accessibility of the RBS of mRNA targets for ribosome binding, (iii) acting as a chaperone for the interaction of the RNA target with other effector molecules, and (iv) modulation of transcription terminator/antiterminator structure formation, and will be described hereafter.

The post-transcriptional regulation

Post-transcriptional regulation is the control of gene expression at the RNA level. It occurs once the RNA polymerase has been attached to the gene's promoter and is synthesizing the nucleotide sequence. Therefore, as the name indicates, it occurs between the transcription phase and the translation phase of gene expression. These controls are critical for the regulation of many genes across human tissues.

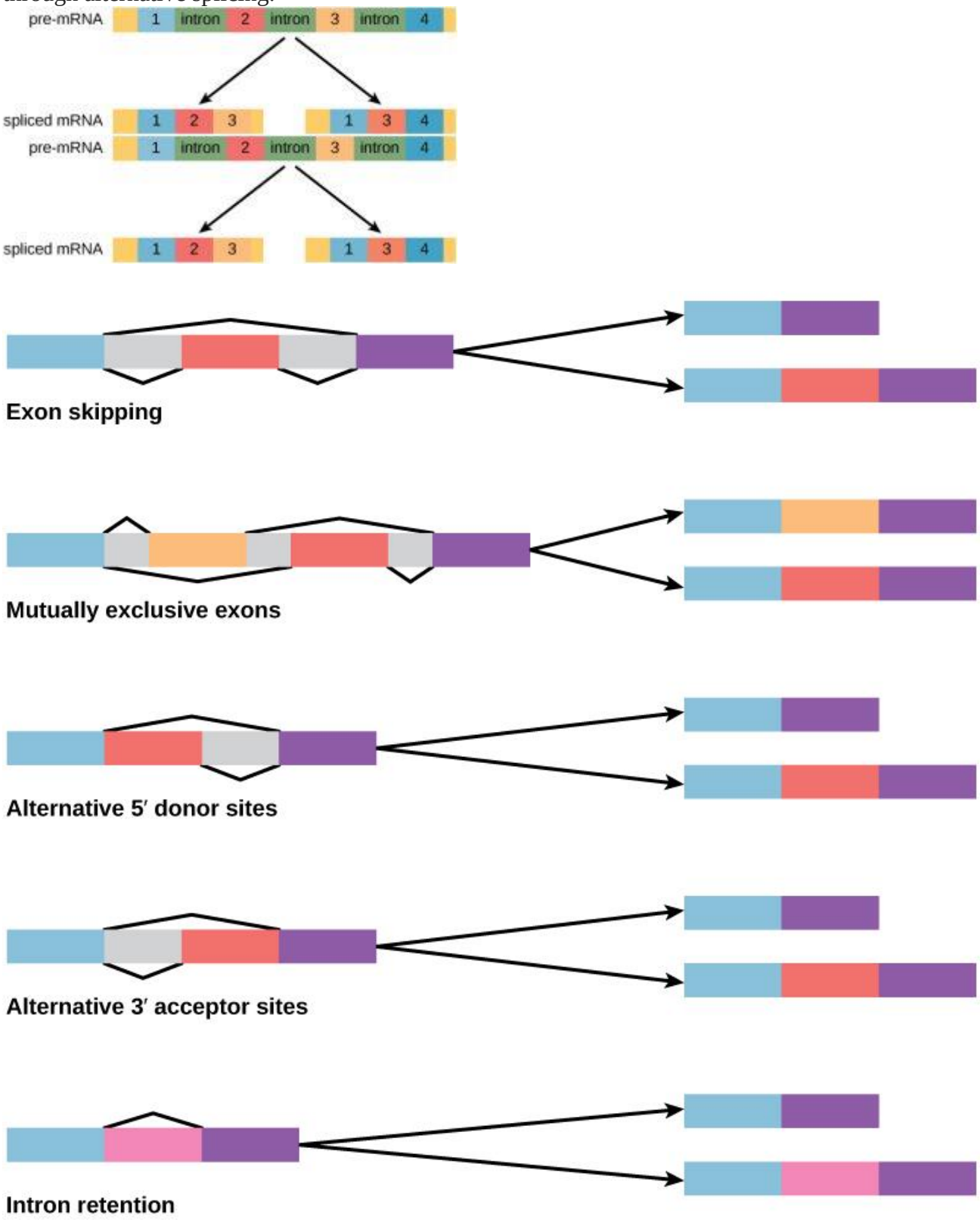
It also plays a big role in cell physiology, being implicated in pathologies such as cancer and neurodegenerative diseases

After RNA is transcribed, it must be processed into a mature form before translation can begin. This processing after an RNA molecule has been transcribed, but before it is translated into a protein, is called post-transcriptional modification. As with the epigenetic and transcriptional stages of processing, this post-transcriptional step can also be regulated to control gene expression in the cell. If the RNA is not processed, shuttled, or translated, then no protein will be synthesized.

ALTERNATIVE RNA SPLICING

In the 1970s, genes were first observed that exhibited **alternative RNA splicing**. Alternative RNA splicing is a mechanism that allows different protein products to be produced from one gene when different combinations of introns (and sometimes exons) are removed from the transcript (**Figure 1**). This alternative splicing can be haphazard, but more often it is controlled and acts as a mechanism of gene regulation, with the frequency of different splicing alternatives controlled by the cell as a way to control the production of different protein products in different cells, or at different stages of

development. Alternative splicing is now understood to be a common mechanism of gene regulation in eukaryotes; according to one estimate, 70% of genes in humans are expressed as multiple proteins through alternative splicing.



There are five basic modes of alternative splicing.

L. 9 *The gene expression regulation at the translational level*

It is estimated that the human genome encodes approximately 25,000 genes, about the same number as that for corn and nearly twice as many as that for the common fruit fly. Even more interesting is the fact that those 25,000 genes are encoded in about 1.5% of the genome. So, what exactly does the other 98.5% of our DNA do? While many mysteries remain about what all of that extra sequence is for, we know that it does contain complex instructions that direct the intricate turning on and off of gene transcription.

Eukaryotes Require Complex Controls Over Gene Expression

While basic similarities in gene transcription exist between prokaryotes and eukaryotes—including the fact that RNA polymerase binds upstream of the gene on its promoter to initiate the process of transcription—multicellular eukaryotes control cell differentiation through more complex and precise temporal and spatial regulation of gene expression.

Multicellular eukaryotes have a much larger genome than prokaryotes, which is organized into multiple chromosomes with greater sequence complexity. Many eukaryotic species carry genes with the same sequences as other plants and animals. In addition, the same DNA sequences (though not the same proteins) are found within all of an organism's diploid, nucleated cells, even though these cells form tissues with drastically different appearances, properties, and functions. Why then, is there such great variation among and within such organisms? Quite simply, the way in which different genes are turned on and off in specific cells generates the variety we observe in nature. In other words, specific functions of different cell types are generated through differential gene regulation. Of course, higher eukaryotes still respond to environmental signals by regulating their genes. But there is an additional layer of regulation that results from cell-to-cell interactions within the organism that orchestrate development. Specifically, gene expression is controlled on two levels. First, transcription is controlled by limiting the amount of mRNA that is produced from a particular gene. The second level of control is through post-transcriptional events that regulate the translation of mRNA into proteins. Even after a protein is made, post-translational modifications can affect its activity.

Transcriptional Regulation in Eukaryotes

Regulation of transcription in eukaryotes is a result of the combined effects of structural properties (how DNA is "packaged") and the interactions of proteins called transcription factors. The most important structural difference between eukaryotic and prokaryotic DNA is the formation of chromatin in eukaryotes. Chromatin results in the different transcriptional "ground states" of prokaryotes and eukaryotes (Table 1).

Table 1: Overview of Differences Between Prokaryotic and Eukaryotic Gene Expression and Regulation

	Prokaryotes	**Eukaryotes**
Structure of genome	Single, generally circular genome sometimes accompanied by smaller pieces of accessory DNA, like plasmids	Genome found in chromosomes; nucleosome structure limits DNA accessibility
Size of genome	Relatively small	Relatively large
Location of gene transcription and translation	Coupled; no nucleoid envelope barrier because of prokaryotic cell structure	Nuclear transcription and cytoplasmic translation

Gene clustering	Operons where genes with similar function are grouped together	Operons generally not found in eukaryotes; each gene has its own promoter element and enhancer element(s)
Default state of transcription	On	Off
DNA structure	Highly supercoiled DNA with some associated proteins	Highly supercoiled chromatin associated with histones in nucleosomes

Transcription Factors and Combinatorial Control

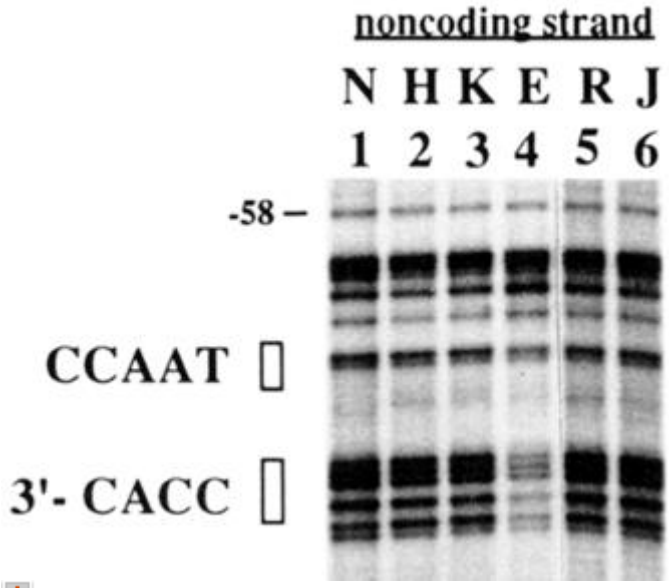


Figure 1: DNA footprinting reveals transcription factor specificity in different cell types
In vivo footprinting analysis of the human beta globin promoter shows that adult erythroblasts (E, lane 4) have footprints on important regulatory motifs (note lighter regions, especially at CACC) as compared to the other samples. Here, lane N is control DNA, lane H is HeLa cells, lane K is K562 cells, lane R is Raji cells, and lane J is Jurkat cells. Of these cells lines, none is part of the lineage leading to red blood cells.

Transcription factors (TFs) are regulatory proteins whose function is to activate (or more rarely, to inhibit) transcription of DNA by binding to specific DNA sequences. TFs have defined DNA-binding domains with up to 10^6-fold higher affinity for their target sequences than for the remainder of the DNA strand. These highly conserved sequences have been used to categorize the known TFs into various "families," such as the MADS box-containing proteins, SOX proteins, and POU factors (Remenyi *et al.*, 2004). Transcription factors can also be classified by their three-dimensional protein structure, including basic helix-turn-helix, helix-loop-helix, and zinc finger proteins. These different structural motifs result in transcription factor specificity for the consensus sequences to which they bind.

Sequence-specific transcription factors are considered the most important and diverse mechanisms of gene regulation in both prokaryotic and eukaryotic cells (Pulverer, 2005). In eukaryotes, regulation of gene expression by transcription factors is said to be combinatorial, in that it requires the coordinated interactions of multiple proteins (in contrast to prokaryotes, in which a single protein is usually all that is required).

Many genes, known as housekeeping genes, are needed by almost every type of cell and appear to be unregulated or constitutive. But at the core of cellular differentiation, manifested in the variety of cell types observed in different organisms, is the regulation of gene expression in a tissue-specific manner. The same genome is responsible for making the entire cadre of cell types, each of which has its own function—for example, red blood cells exchange oxygen, muscle cells expand and contract, and cells in the immune system recognize pathogens. Genes that regulate cell identity are turned on under very specific temporal, spatial, and environmental conditions to ensure that a cell is able to perform its designated function.

Take the example of the gene for beta globin, a protein used in red blood cells for oxygen exchange. Every cell in the human body contains the beta globin gene and the corresponding upstream regulatory sequences that regulate expression, but no cell type other than red blood cells expresses beta globin. Scientists can use a technique called DNA footprinting to map where transcription factors bind to specific regulatory sequences. When Reddy *et al*. examined the beta globin promoter in different cell types, they found that the transcription factors that could bind to the promoter sequences required for beta globin expression were expressed only in erythroblasts (immature adult red blood cells). (See Figure 1). The two consensus sequences in the beta globin promoter known for binding transcription factors, CCAAT and CACC, were protected in the erythroid cells (E), but not the other cell type

L. 10 *The post translational modulation (regulation)*

Post-translational modification (**PTM**) refers to the covalent and generally enzymatic modification of proteins following protein biosynthesis. Proteins are synthesized by ribosomes translating mRNA into polypeptide chains, which may then undergo PTM to form the mature protein product. PTMs are important components in cell signaling, as for example when prohormones are converted to hormones.

Post-translational modifications can occur on the amino acid side chains or at the protein's C- or N- termini. They can extend the chemical repertoire of the 20 standard amino acids by modifying an existing functional group or introducing a new one such as phosphate.

Phosphorylation is a very common mechanism for regulating the activity of enzymes and is the most common post-translational modification.

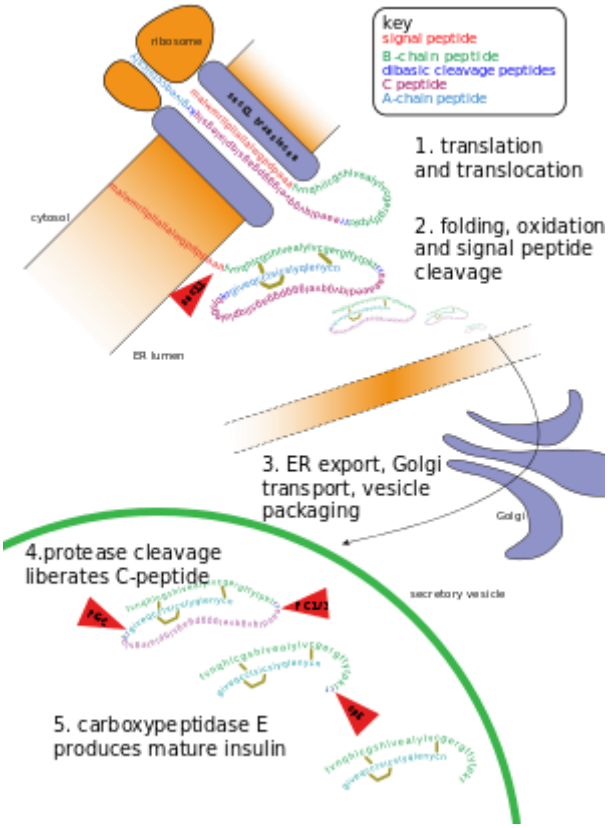
Many eukaryotic and prokaryotic proteins also have carbohydrate molecules attached to them in a process called glycosylation, which can promote protein folding and improve stability as well as serving regulatory functions. Attachment of lipid molecules, known as lipidation, often targets a protein or part of a protein attached to the cell membrane.

Other forms of post-translational modification consist of cleaving peptide bonds, as in processing a propeptide to a mature form or removing the initiator methionine residue. The formation of disulfide bonds from cysteine residues may also be referred to as a post-translational modification.

For instance, the peptide hormone insulin is cut twice after disulfide bonds are formed, and a propeptide is removed from the middle of the chain; the resulting protein consists of two polypeptide chains connected by disulfide bonds.

Some types of post-translational modification are consequences of oxidative stress. Carbonylation is one example that targets the modified protein for degradation and can result in the formation of protein aggregates.[

Specific amino acid modifications can be used as biomarkers indicating oxidative damage.
Sites that often undergo post-translational modification are those that have a functional group that can serve as a nucleophile in the reaction: the hydroxyl groups of serine, threonine, and tyrosine; the amine forms of lysine, arginine, and histidine; the thiolate anion of cysteine; the carboxylates of aspartate and glutamate; and the N- and C-termini. In addition, although the amide of asparagine is a weak nucleophile, it can serve as an attachment point for glycans. Rarer modifications can occur at oxidized methionines and at some methylenes in side chains.
Post-translational modification of proteins can be experimentally detected by a variety of techniques, including mass spectrometry, Eastern blotting, and Western blotting. Additional methods are provided in the external links sections



Post-translational modification of insulin. At the top, the ribosome translates a mRNA sequence into a protein, insulin, and passes the protein through the endoplasmic reticulum, where it is cut, folded, and held in shape by disulfide (-S-S-) bonds. Then the protein passes through the golgi apparatus, where it is packaged into a vesicle. In the vesicle, more parts are cut off, and it turns into mature insulin.

L. 11. *Features of genome regulation at organells*

Protist mitochondrial genomes show a very wide range of gene content, ranging from three genes for respiratory chain components in Apicomplexa and dinoflagellates to nearly 100 genes in *Reclinomonas americana*. In many organisms the rRNA genes are fragmented, although still functional. Some protist mitochondria encode a full set of tRNAs, while others rely on imported molecules. There is similarly a wide variation in mitochondrial genome organization, even among closely related groups. Mitochondrial gene expression and control are generally poorly characterized. Transcription probably relies on a ‘viral-type’ RNA polymerase, although a ‘bacterial-type’ enzyme may be involved in some cases. Transcripts are heavily edited in many lineages. The chloroplast genome

generally shows less variation in gene content and organization, although greatly reduced genomes are found in dinoflagellate algae and non-photosynthetic organisms. Genes in the former are located on small plasmids in contrast to the larger molecules found elsewhere. Control of gene expression in chloroplasts involves transcriptional and post-transcriptional regulation. Redox poise and the ATP/ADP ratio are likely to be important determinants. Some protists have an additional extranuclear genome, the nucleomorph, which is a remnant nucleus. Nucleomorphs of two separate lineages have a number of features in common.

The mitochondrion originated from an α-proteobacterial endosymbiont acquired by a host that was unable to carry out aerobic respiration, but whose nature is otherwise controversial . The majority of the symbiont's genes were either transferred to the nucleus or lost completely. A number of eukaryotic lineages have secondarily adopted an anaerobic lifestyle, with consequent modification of the mitochondrion to form hydrogenosomes or mitosomes, as discussed elsewhere in this issue. This modification has been accompanied by partial or complete loss of the mitochondrial genome. The chloroplast originated from the acquisition of an oxygenic photosynthetic eubacterial endosymbiont by a non-photosynthetic host, with subsequent reduction of the symbiont genome. Although it is frequently supposed that a single primary endosymbiosis gave rise to all chloroplasts, the evidence for this has been questioned and there is increasing evidence that the photosynthetic chromatophore of the amoeba *Paulinella* may represent a clearly independent chloroplast origin.

It is recognized that multiple 'secondary' endosymbiotic acquisitions of photosynthetic eukaryotes have occurred. In some secondary endosymbiotic lineages, the nucleus of the intermediate photosynthetic eukaryote persists as a 'nucleomorph' between two of the four membranes surrounding the chloroplast.

Some lineages have lost photosynthetic function, with consequent reduction of the chloroplast genome, analogous to the reduction of the mitochondrial genome in anaerobic eukaryotes. It is not clear whether any formerly photosynthetic eukaryotes have completely lost a chloroplast genome, or indeed a chloroplast compartment . There is tremendous variation among mitochondrial genomes, less among chloroplast genomes, which we illustrate in the following discussion. For the purposes of this discussion, we include macroalgal relatives of the microalgae. For simplicity, we use the term chloroplast to cover all photosynthetic organelles, whatever their pigment type, as well as non-photosynthetic organelles derived from them during evolution or development.

CHLOROPLAST GENE EXPRESSION AND CONTROL

Most of our understanding of gene expression and its control in chloroplasts comes from work on land plants and the unicellular green alga *Chlamydomonas*. The latter is particularly valuable for such studies as it is (i) readily amenable to genetic analysis and (ii) able to grow heterotrophically on medium containing fixed carbon, so non-photosynthetic mutants remain viable.

Land plant chloroplasts contain a nuclear-encoded, viral-type RNA polymerase (NEP, or nuclear-encoded polymerase), which is used to transcribe 'genetic system' genes in the chloroplasts, such as those for the 'plastid-encoded RNA polymerase' (PEP), which in turn transcribes genes for proteins involved in photosynthesis. However, algae appear only to have a chloroplast-encoded bacterial-type RNA polymerase (with the possible exception of dinoflagellates), so the possibility of control by modulation of the NEP does not exist). The PEP is thought to resemble that of *E. coli*, where the RNA polymerase is encoded by the *rpoA, B* and *C* genes. In algae, the *rpoC* gene is normally divided into two separate genes *rpoC1* and *rpoC2*. An exception to this is in *Chlamydomonas* where significant gene rearrangement has occurred and where the location and identity of the *rpoA* and *rpoC1* genes are still unclear. The PEP presumably requires sigma factors, and there is variation in the number of these. Land plants have several nuclear genes for them, as does the red alga *Cyanidioschyzon merolae* . Many other algae, such as *Chlamydomonas reinhardtii*, appear to

have only a single sigma factor), and in *Guillardia theta*, a sigma factor is encoded in the nucleomorph.

L. 12. *Functional sites of genome*

Functional genomics is a branch that integrates molecular biology and cell biology studies, and deals with the whole structure, function and regulation of a gene in contrast to the gene-by-gene approach of classical molecular biology technique. It aims to relate the phenotype and genotype on genome level and includes processes such as transcription, translation, protein-protein interaction and epigenetic regulation. This involves comprehensive analysis to understand genes, their functional roles and variable levels of protein expression. The Human Genome Project (HGP) (Collins *et al.*, 2003; Green *et al.*, 2015) is an integral part of functional genomics. It elucidated that the human genome contains 3164.7 million nucleotide bases with the total number of ~20,000 genes. The size of chromosomes is in between 47 and 250 Mb. Chromosome 1 has a maximum number of genes (5078) and the MT has fewest (37). Functional genomics is characterized by the following distinct research areas:

Functional genomics advances have led to the development of high-throughput techniques that enable expression profiling within discrete brain regions and specific cell types. These exciting technologies, when combined with solid experimental design and validation using alternative molecular- and cellular-based strategies, comprise a paradigm that is useful for assessing mechanisms underlying the pathophysiology of PD. Specifically, individual genes and classes of transcripts that comprise discrete signaling pathways can be evaluated in vitro as well as in the midbrain, SNPC, and/or striatum of relevant animal models of PD and human postmortem brains of subjects afflicted with PD as well as exploring potential biomarker and therapeutic strategies through the use of blood and cerebrospinal fluid as RNA/miRNA sources.

L.13. *Regulation of genes expression in eukaryotes*

Gene regulation is the process of turning genes on or off. Gene regulation can occur at any point of the transcription-translation process but most often occurs at the transcription level.

Proteins that can be activated by other cells and signals from the environment are called transcription factors. Transcription factors bind to regulatory regions of the gene and increase or decrease the level of transcription. Other mechanisms of gene regulation include regulating the processing of RNA, the stability of mRNA and the rate of translation.

Turning the correct genes on and off is an essential component to maintaining a cell's functionality.

Epigenetic modifications

Gene/genome editing/CRISPR

Many scientists have contributed to the development of genome-editing technology. Emmanuelle Charpentier, PhD, and Jennifer Douda, PhD, are often considered pioneers in the field. In 2015, they published a paper on using a bacterial system called clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) or CRISPR-Cas9 technology to edit genomes. The CRISPR technology can make precise changes in human DNA by slicing out the incorrect portion of the gene and replacing it. It is a complicated process, but simply put, "guide" RNA and a bacterial enzyme, called Cas-9, bind to and cut DNA. A repair template with the desired change is inserted where the DNA has been cut. Multiple DNA edits can be made simultaneously.

Editing DNA with CRISPR has many advantages. For example, genome editing could potentially prevent or treat genetic diseases such as cystic fibrosis, hemophilia and sickle cell anemia. Research is also being done on DNA editing in the treatment of more complex diseases, such as cancer. CRISPR technology is quick and fairly easy for trained scientists.
Although there are many benefits of using CRISPR, the technology also has some limitations. Although CRISPR technology is precise, it is not perfect. It sometimes cuts DNA that is similar to the guide RNA, but not exact.
CRISPR has been one of the biggest scientific achievements of the century. However, with progress comes considerations. There are complicating ethical issues to evaluate when considering DNA editing. For example, is it appropriate to edit the genomes of human embryos? Should we cure disease? Do edits we make today have unforeseen impacts to future generations? How does commercialization of gene editing technologies fit in? Should CRISPR technology be available to the scientific masses, or should its use be limited to selected experts? These questions remain up for debate as conversations about CRISPR technology continue. While this debate continues, leaders in genetics and bioethics have proposed a moratorium on germline gene editing.

L. 14. *MicroRNA as regulators gene expression*

MicroRNAs (miRNAs) are a class of non-coding RNAs that play important roles in regulating gene expression. The majority of miRNAs are transcribed from DNA sequences into primary miRNAs and processed into precursor miRNAs, and finally mature miRNAs. In most cases, miRNAs interact with the 3′ untranslated region (3′ UTR) of target mRNAs to induce mRNA degradation and translational repression. However, interaction of miRNAs with other regions, including the 5′ UTR, coding sequence, and gene promoters, have also been reported. Under certain conditions, miRNAs can also activate translation or regulate transcription. The interaction of miRNAs with their target genes is dynamic and dependent on many factors, such as subcellular location of miRNAs, the abundancy of miRNAs and target mRNAs, and the affinity of miRNA-mRNA interactions. miRNAs can be secreted into extracellular fluids and transported to target cells via vesicles, such as exosomes, or by binding to proteins, including Argonautes. Extracellular miRNAs function as chemical messengers to mediate cell-cell communication. In this review, we provide an update on canonical and non-canonical miRNA biogenesis pathways and various mechanisms underlying miRNA-mediated gene regulations. We also summarize the current knowledge of the dynamics of miRNA action and of the secretion, transfer, and uptake of extracellular miRNAs.

Introduction

The discovery of the first microRNA (miRNA), *lin-4*, in 1993 by the Ambros and Ruvkun groups in *Caenorhabditis elegans* () has revolutionized the field of molecular biology. Years before, *lin-4* was characterized by the Horvitz's lab as one of the genes that regulate temporal development of *C. elegans* larvae (). Later in 1987, the same group found that a mutation in *lin-4* had an opposite phenotype to a mutation in another gene, *lin-14*, yet a *lin-14* suppressor mutation in a *null-lin-4* line was wildtype (. Both Ambros and Ruvkun continued to study *lin-4* and *lin-14* after leaving the Horvitz's lab, only to discover later that *lin-4* was not a protein-coding RNA but indeed a small non-coding RNA. They also found that *lin-14* was post-transcriptionally downregulated through its 3′ untranslated region (UTR) and that *lin-4* had a complementary sequence to that of the 3′ UTR of *lin-14* (1). Therefore, they proposed that *lin-4* regulates *lin-14* at the post-transcriptional level. Since then, miRNAs have been detected in all animal model systems and some were shown to be highly conserved across species. New miRNAs are still being discovered and their roles in gene regulation are well recognized.

miRNAs are small non-coding RNAs, with an average 22 nucleotides in length. Most miRNAs are transcribed from DNA sequences into primary miRNAs (pri-miRNAs) and processed into precursor miRNAs (pre-miRNAs) and mature miRNAs. In most cases, miRNAs interact with the 3′ UTR of target mRNAs to suppress expression. However, interaction of miRNAs with other regions, including the 5′ UTR, coding sequence, and gene promoters, have also been reported. Furthermore, miRNAs have been shown to activate gene expression under certain conditions. Recent studies have suggested that miRNAs are shuttled between different subcellular compartments to control the rate of translation, and even transcription.
miRNAs are critical for normal animal development and are involved in a variety of biological processes. Aberrant expression of miRNAs is associated with many human diseases.
In addition, miRNAs are secreted into extracellular fluids. Extracellular miRNAs have been widely reported as potential biomarkers for a variety of diseases and they also serve as signaling molecules to mediate cell-cell communications.
We have provided a brief overview of the different pathways of miRNA biogenesis in animals and the expanding complexity of their regulation of gene expression. Moreover, we have discussed the dynamics of miRNA intracellular localization and function. Finally, we have summarized the secretion and circulation of miRNAs and the potential roles of extracellular miRNAs in mediating intercellular communications.

Biogenesis of miRNAs
miRNA biogenesis starts with the processing of RNA polymerase II/III transcripts post- or co-transcriptionally. About half of all currently identified miRNAs are intragenic and processed mostly from introns and relatively few exons of protein coding genes, while the remaining are intergenic, transcribed independently of a host gene and regulated by their own promoters. Sometimes miRNAs are transcribed as one long transcript called clusters, which may have similar seed regions, and in which case they are considered a family. The biogenesis of miRNA is classified into canonical and non-canonical pathways (Figure 1).

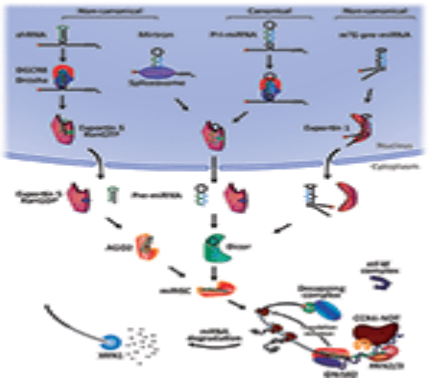


Figure 1. MicroRNA biogenesis and mechanism of action. Canonical miRNA biogenesis begins with the generation of the pri-miRNA transcript. The microprocessor complex, comprised of Drosha and DiGeorge Syndrome Critical Region 8 (DGCR8), cleaves the pri-miRNA to produce the precursor-miRNA (pre-miRNA). The pre-miRNA is exported to the cytoplasm in an Exportin5/RanGTP-dependent manner and processed to produce the mature miRNA duplex. Finally, either the 5p or 3p strands of the mature miRNA duplex is loaded into the Argonaute (AGO) family of proteins to form a miRNA-induced silencing complex (miRISC). In the non-canonical pathways, small hairpin RNA (shRNA) are initially cleaved by the microprocessor complex and exported to the cytoplasm via Exportin5/RanGTP. They are further processed via AGO2-dependent, but Dicer-independent,

cleavage. Mirtrons and 7-methylguanine capped (m^7G)-pre-miRNA are dependent on Dicer to complete their cytoplasmic maturation, but they differ in their nucleocytoplasmic shuttling. Mirtrons are exported via Exportin5/RanGTP while m^7G-pre-miRNA are exported via Exportin1. All pathways ultimately lead to a functional miRISC complex. In most cases, miRISC binds to target mRNAs to induce translational inhibition, most likely by interfering with the eIF4F complex. Next, GW182 family proteins bound to Argonaute recruit the poly(A)-deadenylases PAN2/3 and CCR4-NOT. PAN2/3 initiates deadenylation while the CCR4-NOT complex completes the process, leading to removal of the m^7G cap on target mRNA by the decapping complex. Decapped mRNA may then undergo 5′−3′ degradation via the exoribonuclease XRN1. Modified from Hayder et al.

L. 15. *Genomics of Gene Regulation*Gene regulation is the process of turning genes on and off. During early development, cells begin to take on specific functions. Gene regulation ensures that the appropriate genes are expressed at the proper times. Gene regulation can also help an organism respond to its environment. Gene regulation is accomplished by a variety of mechanisms including chemically modifying genes and using regulatory proteins to turn genes on or off.
Metazoan life is dependent on the proper temporal and spatial control of gene expression within the many cells—essentially all with the identical genome—that make up the organism. While much is understood about how individual gene regulatory elements function, many questions remain about how they interact to maintain correct regulation globally throughout the genome. In this review we summarize the basic features and functions of the crucial regulatory elements promoters, enhancers, and insulators and discuss some of the ways in which proper interactions between these elements is realized. We focus in particular on the role of core promoter sequences and propose explanations for some of the contradictory results seen in experiments aimed at understanding insulator function. We suggest that gene regulation depends on local genomic context and argue that more holistic in vivo investigations that take into account multiple local features will be necessary to understand how genome-wide gene regulation is maintained.

Maintaining proper control of gene expression is fundamental for all organisms. Although much is known about how individual metazoan genes are regulated, how correct patterns of gene activation are maintained genome-wide is not well understood. Every gene lies adjacent to another gene, and many genes have multiple differentially regulated transcripts. Genes can be nested inside other genes or overlap one another on opposite strands of the DNA. Within the nucleus, chromatin is arranged in a three-dimensional fashion such that genes that are far apart on the chromosome, or are on different chromosomes, become closely juxtaposed. Given such complexity in genomic organization, it is a wonder that gene expression can be correctly sorted out: when regulatory elements are able to act over large distances and ignore intervening elements, how is one regulatory element able to target a specific gene while at the same time bypassing other nearby promoters? We consider here some answers to this question. Our focus is not on broad epigenetic mechanisms such as heterochromatic silencing and Polycomb-mediated repression of large chromatin domains (reviewed by [1]) but rather on local-scale events such as the differential expression of several genes lying in an apparently similar chromatin state or physical region (**Fig. 1**). We begin with a brief review of the main regulatory elements that influence gene expression and genomic organization-promoters, enhancers, and insulators-followed by a discussion of possible mechanisms for ensuring faithful gene regulation. We highlight the often overlooked role of core promoter sequences in mediating specific enhancer-promoter interactions and describe some of the challenges of trying to understand genome-wide events using approaches centered on single genes or regulatory elements. We suggest that a more holistic view of regulation, taking into account the full set of local genomic features, will be needed to fully understand how gene expression throughout the genome is properly maintained.

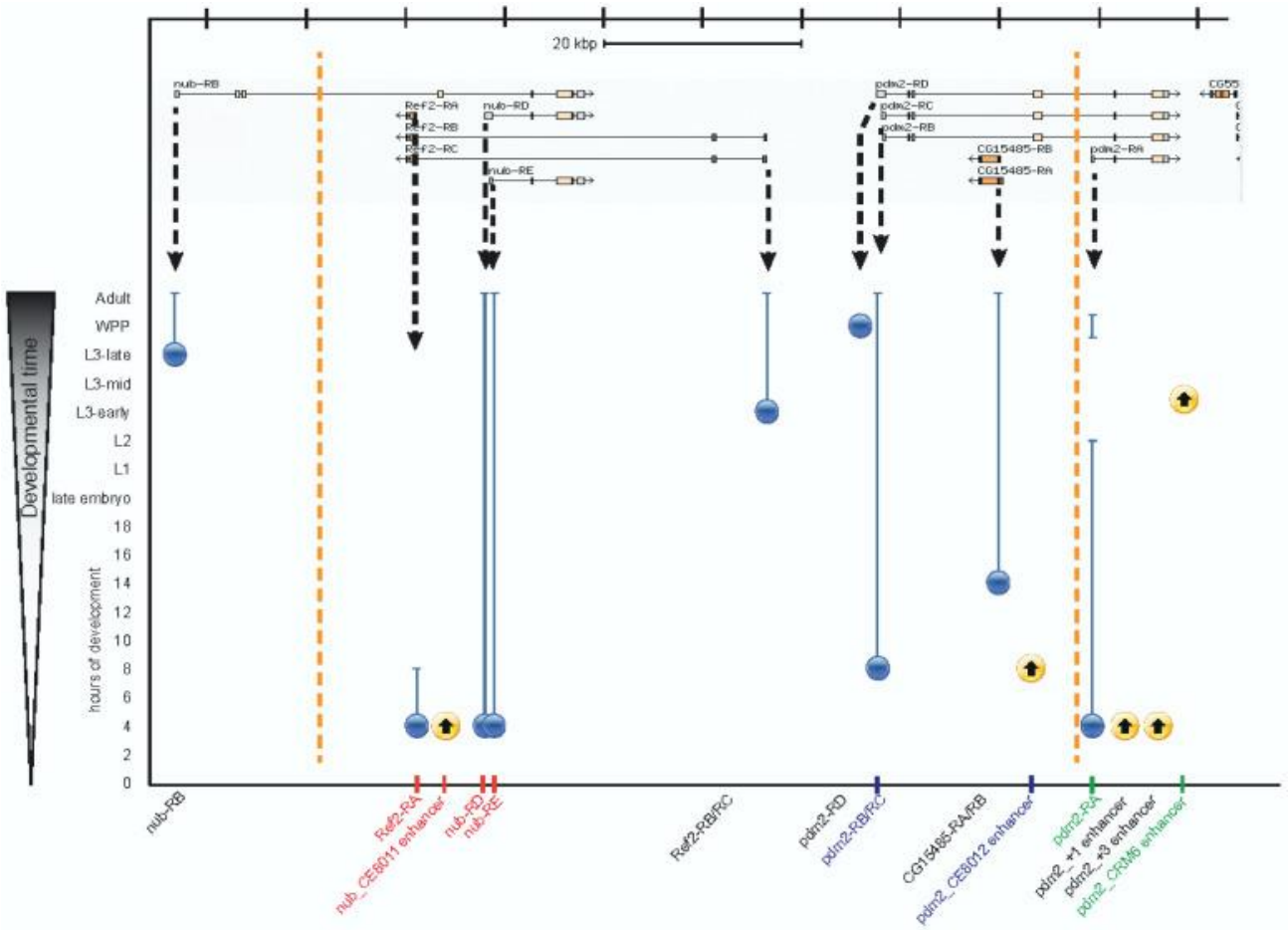


Figure 1. Genomic region showing promoters, enhancers, and insulators. Pictured is a 100 kb fragment of the *Drosophila* genome (chr2L:12,593,026..12,693,025), based on the FlyBase v.FB2013_04 genome annotation. Transcripts for the genes *nub*, *Ref2*, *pdm2*, and *CG15485* are shown along the top of the figure. Each promoter is highlighted with a vertical black dashed arrow. Insulators are depicted by orange dashed lines and enhancers by yellow circles containing upward-pointing arrows; the arrows connote that while the time when enhancers become active is known, how long they remain active generally has not been described. Enhancer names are drawn from the REDfly database. Developmental time is portrayed vertically on the y-axis; not all stages are shown and the axis is not to scale. Blue circles and lines depict gene expression from each promoter based on RNA-seq data provided as part of the genome annotation. Circles represent the onset of expression and lines continued expression, which sometimes must be inferred as it is not always possible to determine from which promoter the later expression originates. Note that of the seven promoters located between the two insulators, only three appear to be co-regulated, potentially by the *nub_CE8011* enhancer (red text). Promoter *pdm2-RB/RC* may be regulated by the *pdm2_CE8012* enhancer, but the other nearby transcripts are not expressed at the time when this enhancer becomes active (blue text). Interestingly, enhancer *pdm2_CRM6* is active exactly when promoter *pdm2-RA* is inactive, raising the possibility that it engages in insulator bypass to activate one of the other *pdm2* promoters or that its native role is as a negative regulatory element and its classification as an enhancer is due to experimental artifact (green text).