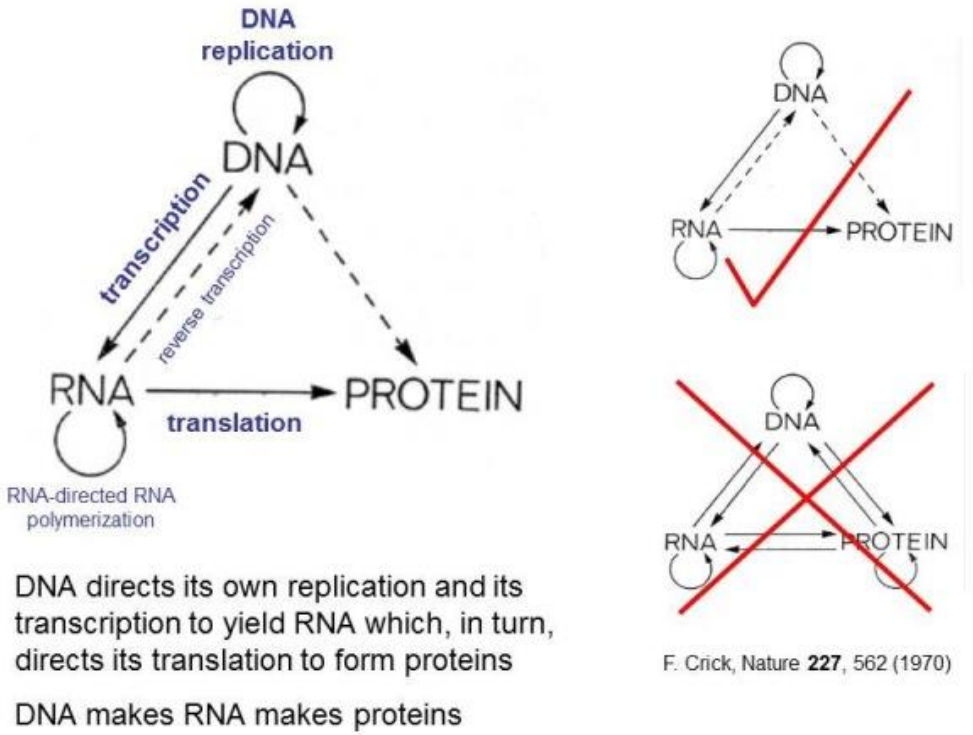
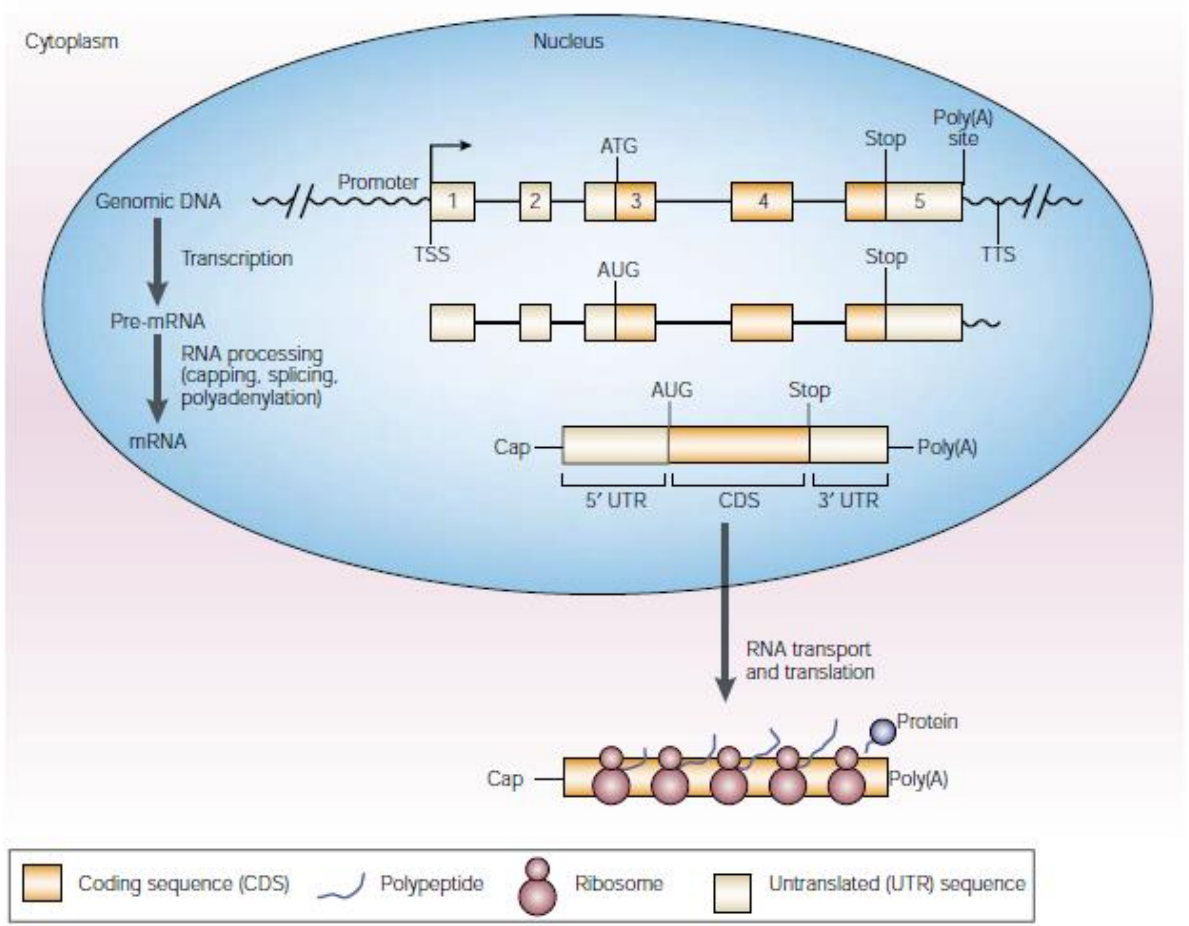
PR2_Geni: struktura genov

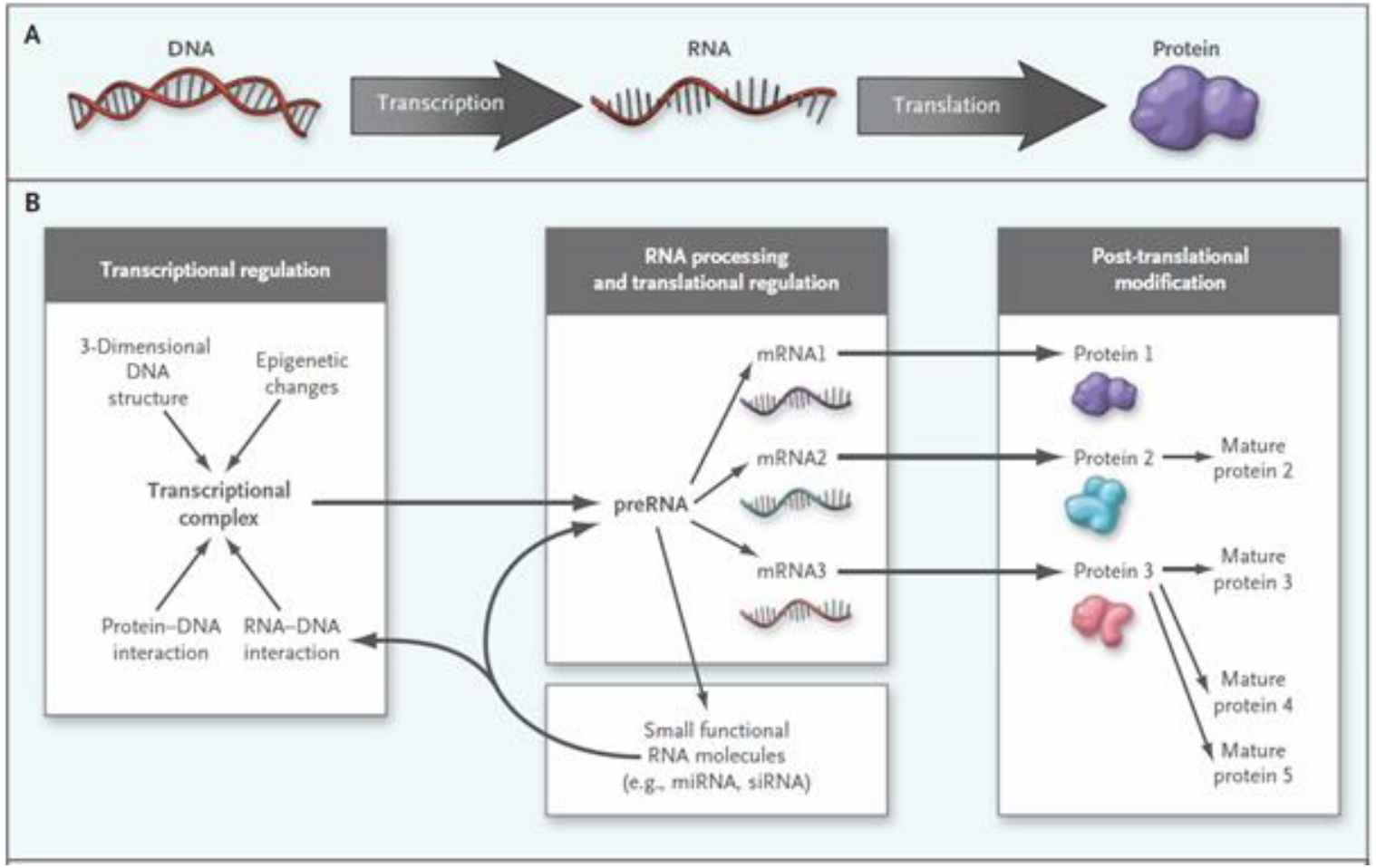
Centralna biološka dogma in koncept gena



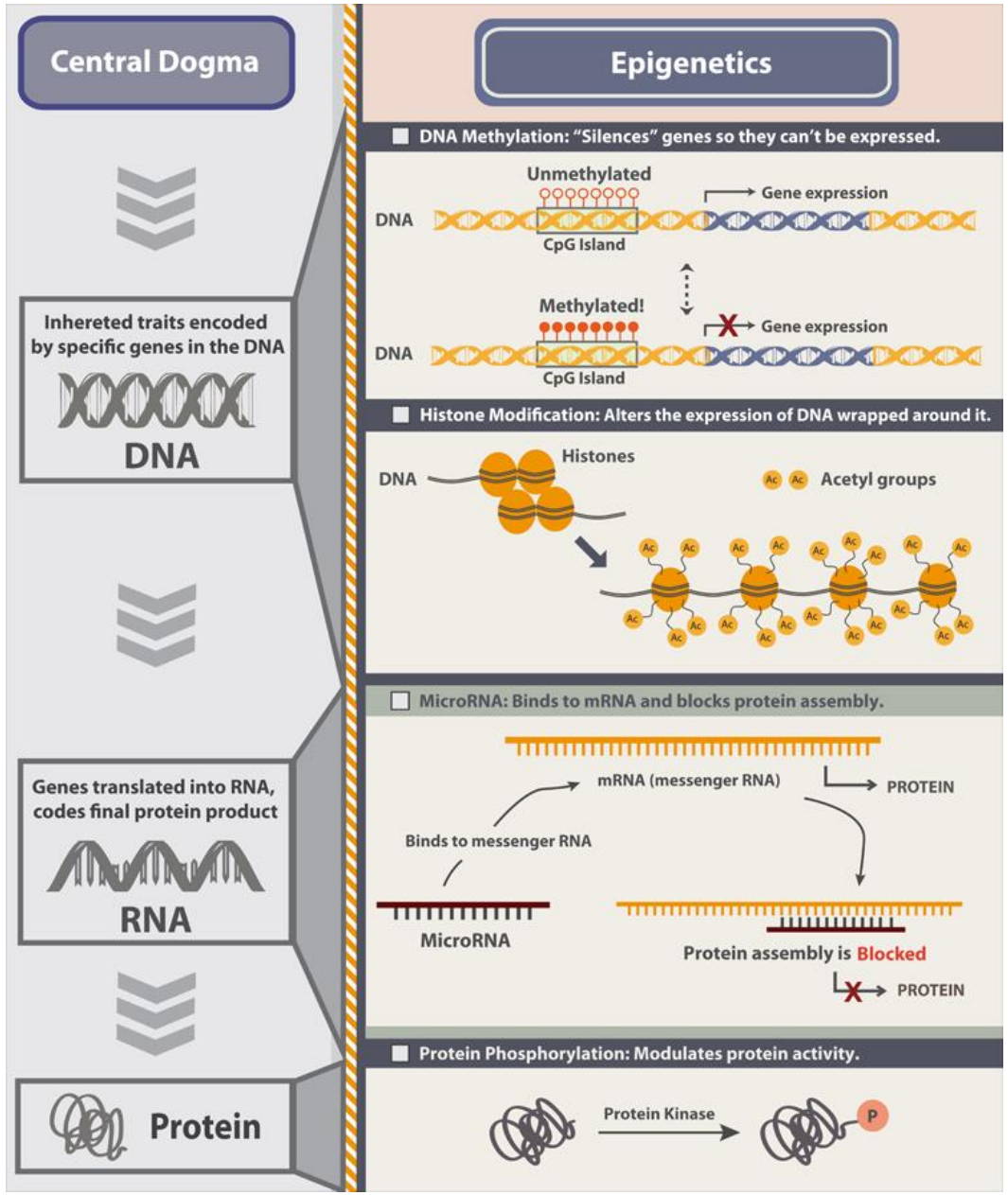
The Central Dogma. Physics has laws; biology has a Central Dogma, put forward by Francis Crick (Nature, 1970). The **Central Dogma of molecular biology** is that DNA directs its own replication and its transcription to yield RNA which, in turn, directs its translation to form proteins. ***The Central Dogma sets out a hierarchy of information. DNA is the information store***.

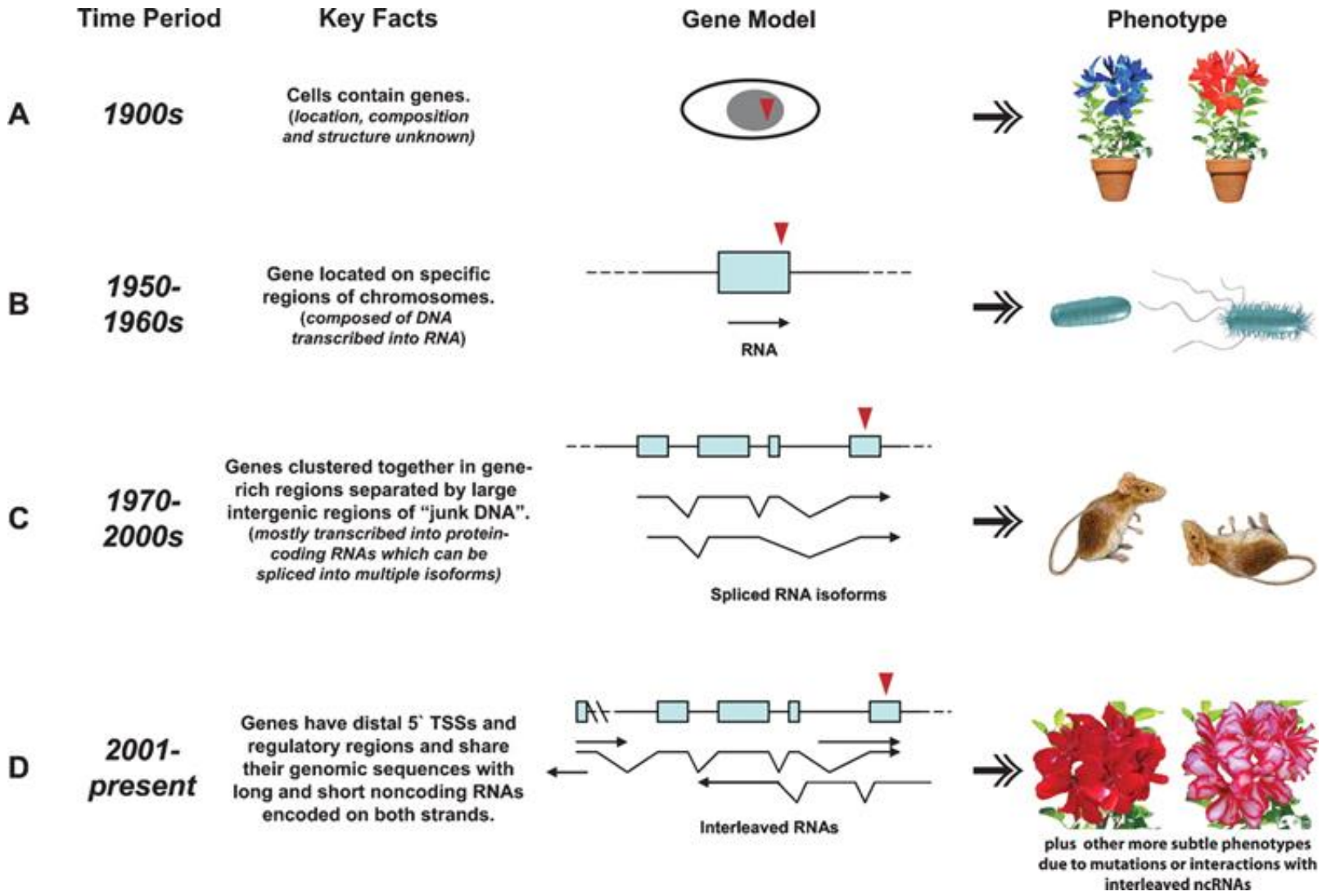


The central dogma of gene expression. In the typical process of eukaryotic gene expression, a gene is transcribed from DNA to pre-mRNA. mRNA is then produced from pre-mRNA by RNA processing, which includes the capping, splicing and polyadenylation of the transcript. It is then transported from the nucleus to the cytoplasm for translation. TSS, transcription start site; TTS, transcription termination site.

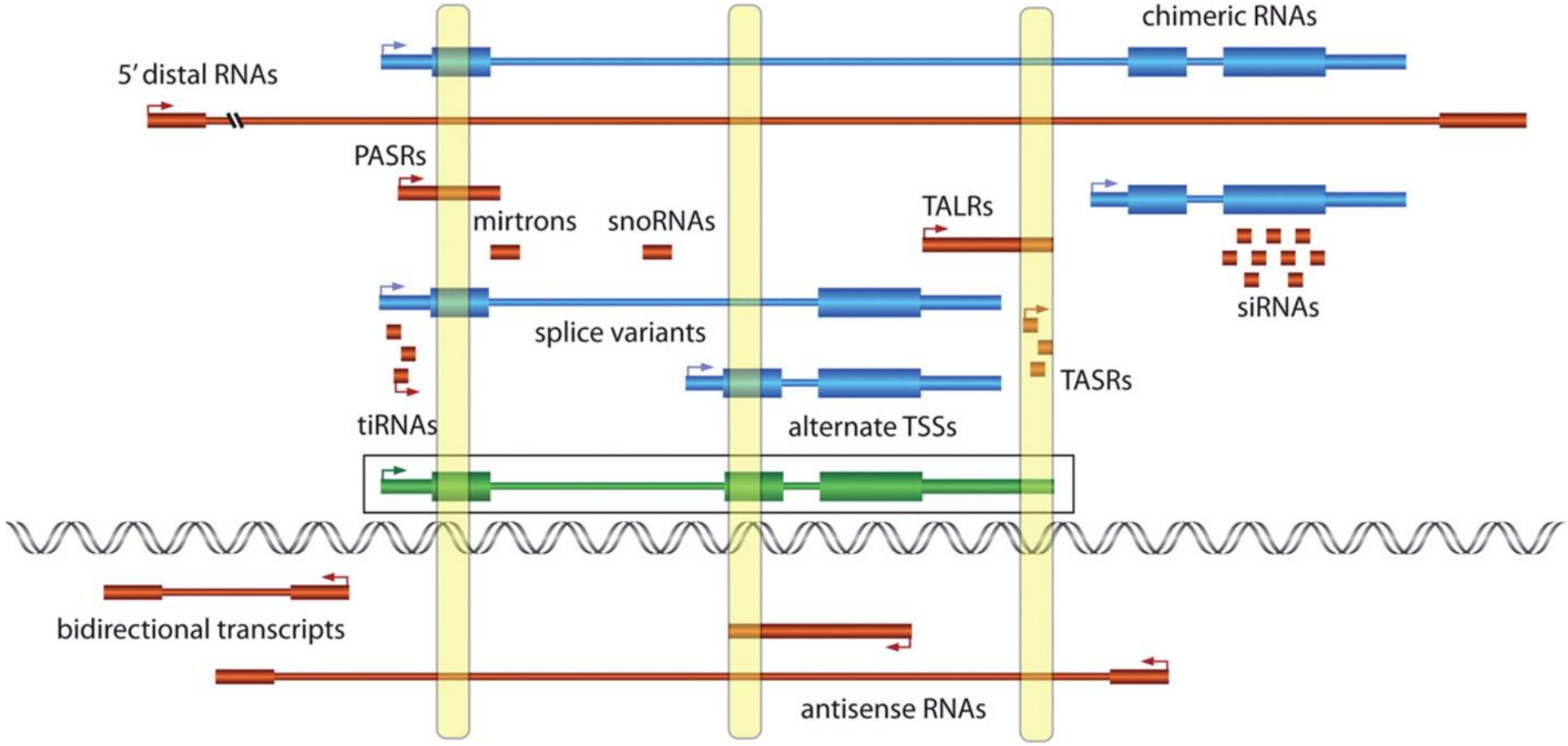


The Increasing Complexity of the Central Dogma of Molecular Biology. The ***flow of genomic information from DNA to RNA to protein remains the basis for understanding genomic function*** (Panel A). A single gene can yield an extensive array of gene products, depending on the environment in which it is expressed, thereby expanding the repertoire of the 20,000 or so genes in the human genome (Panel B). The initial event of gene expression, transcription, is regulated by means of a complex choreography of events involving the three-dimensional DNA structure, covalent chemical, or epigenetic, modifications of the DNA backbone, and interactions between protein and DNA and between RNA and DNA. Translation is similarly complex and tightly regulated by interactions between messenger RNA (mRNA) and proteins. Processing of single-precursor RNA (preRNA) molecules can yield multiple RNA products, including microRNA (miRNA) and small interfering RNA (siRNA) molecules. Post-translational modification of proteins also contributes greatly to the diversity of the output of the human genome through modifications of individual immature proteins (e.g., folding, cleavage, and chemical modifications), which yield an array of related protein products.

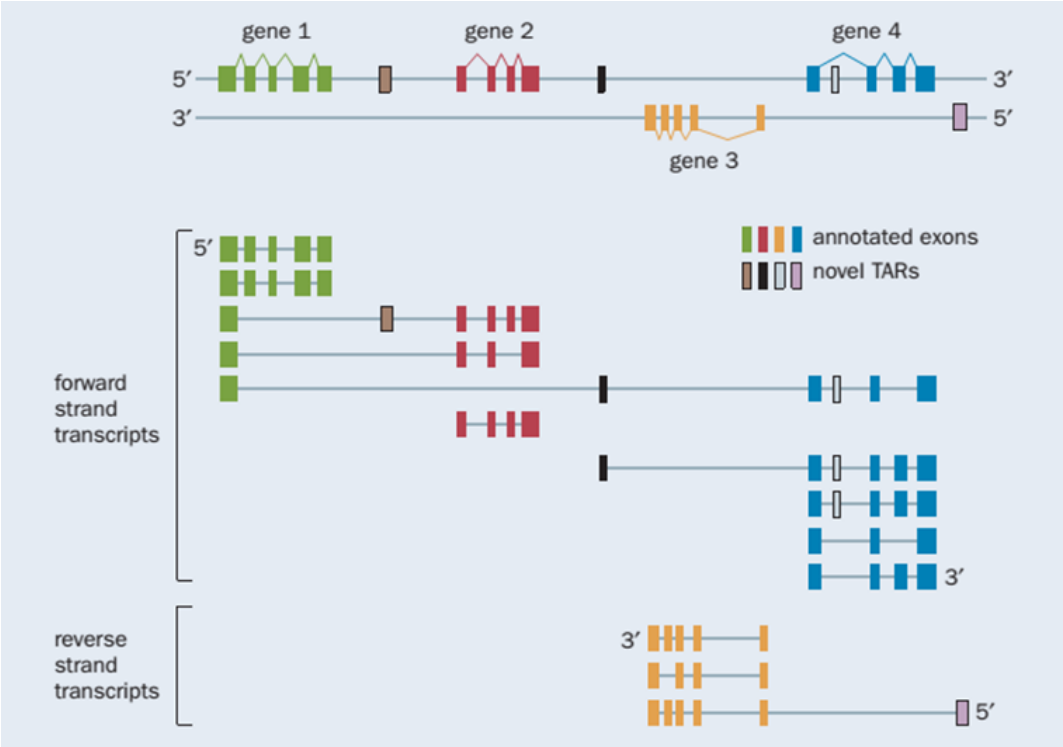

Central Dogma
Inhereted traits encoded by specific genes in the DNA
DNA
Genes translated into RNA, codes final protein product
RNA
Protein
Epigenetics
DNA Methylation: "Silences" genes so they can't be expressed.
Unmethylated
Gene expression
DNA
CpG Island
Methylated!
Gene expression
DNA
CpG Island
Histone Modification: Alters the expression of DNA wrapped around it.
DNA
Histones
Acetyl groups
MicroRNA: Binds to mRNA and blocks protein assembly.
mRNA (messenger RNA)
PROTEIN
Binds to messenger RNA
MicroRNA
Protein assembly is Blocked
PROTEIN
Protein Phosphorylation: Modulates protein activity.
Protein Kinase
P



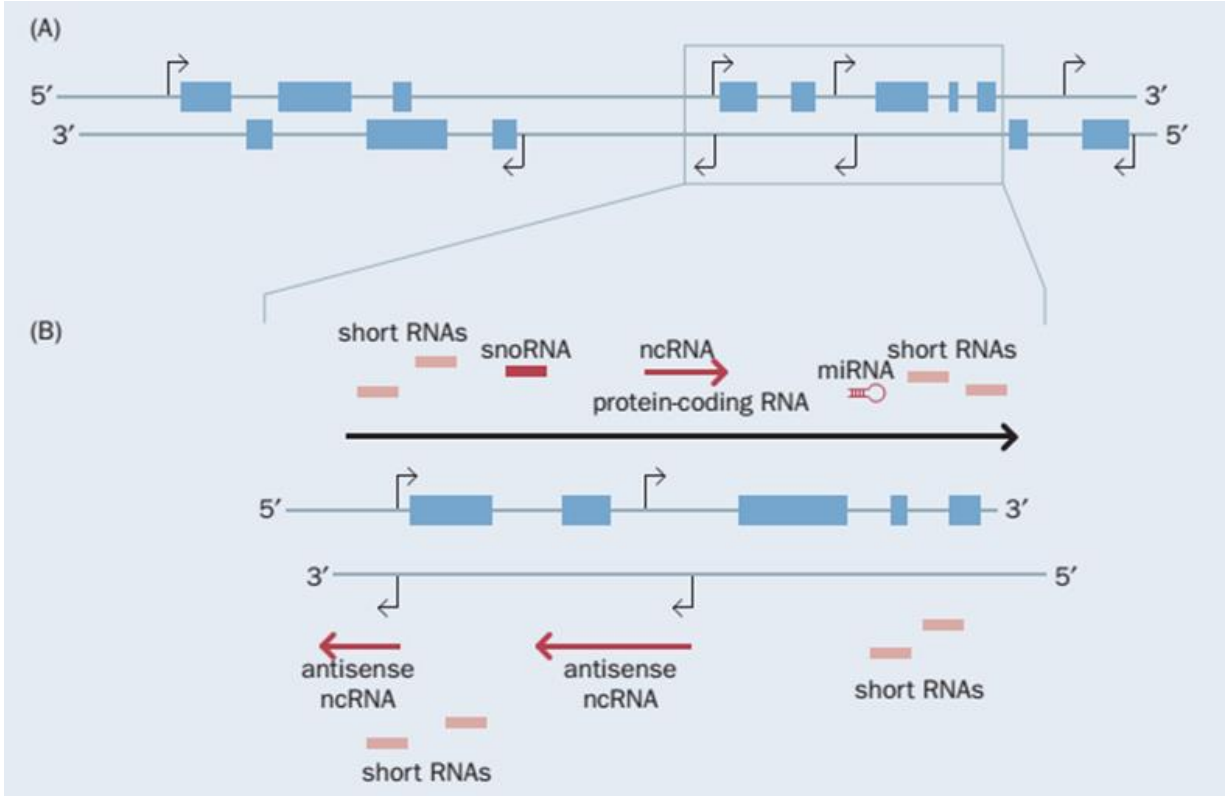
Evolution of the gene model and its relationship to wild-type and mutant phenotypes. Over the past century, the definition of a gene has been improved and refined from its conceptual origin in the early 1900's (A) with ***the discovery of RNA and DNA structures (B), splicing (C), and lastly, widespread unannotated transcription (D)***. Exonic regions are depicted as blue boxes with transcripts shown as arrows below (spliced and unspliced). A hypothetical mutation is shown as a red triangle. Note that as the definition of a gene grows to include multiple transcripts, a single mutation can now affect many different transcripts and thus potentially could have multiple and more subtle phenotypes.



Overview of pervasive transcription and its implications on the gene concept. A representation of a traditional gene (boxed) shown in context with associated coding and non-coding transcripts identified in various transcriptomic analyses. The highlighted sections exemplify how experimental approaches targeting regions within one ‘gene’ may inadvertently target other intersecting transcripts and thereby yield confounding results. Abbreviations: PASRs, promoter-associated small RNAs; TALRs, terminal-associated long RNAs; tiRNAs, tiny RNAs.

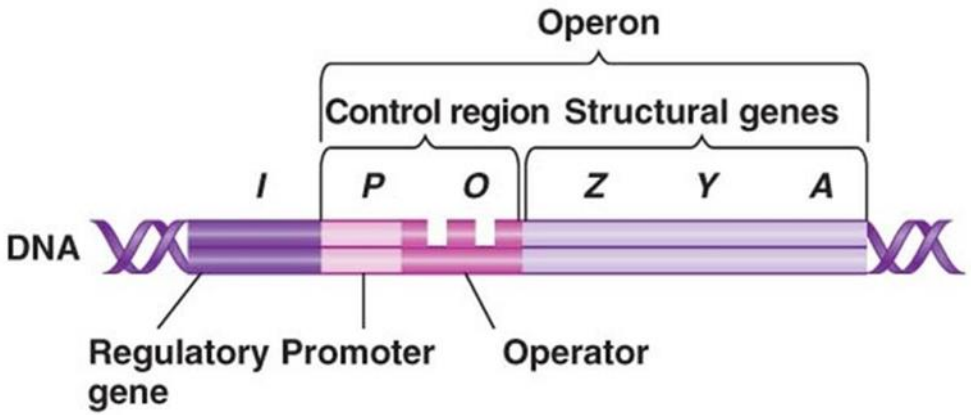


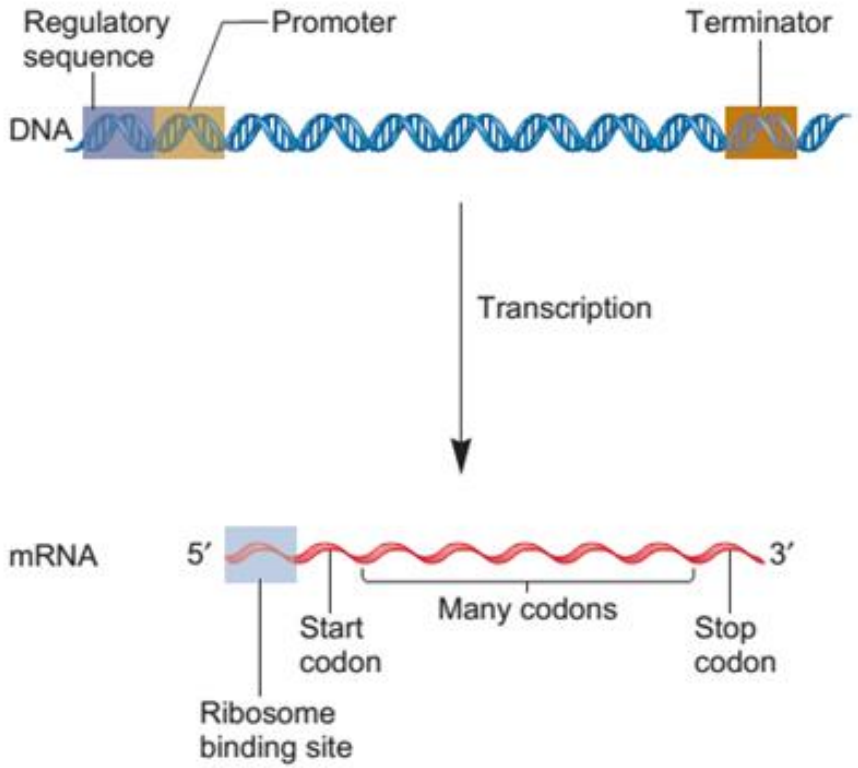
Blurring of gene boundaries at the transcript level. In the past, the four genes at the top would be expected to behave as discrete non-overlapping transcription units. As shown by recent analyses, the reality is more complicated. ***A variety of transcripts often links exons in neighboring genes. The transcripts frequently include sequences from previously unsuspected transcriptionally active regions (TARs)***.



Extensive transcriptional complexity of human genes. (A) **Human genes are frequently transcribed on both strands**, as shown in this hypothetical gene cluster. (B) **A single gene can have multiple transcriptional start sites** (right-angled arrows) as well as **many interleaved coding and noncoding transcripts**. Exons are shown as blue boxes. Known short RNAs such as small nucleolar RNAs (snoRNAs) and microRNAs (miRNAs) can be processed from intronic sequences, and novel species of short RNAs that cluster around the beginning and end of genes have recently been discovered.

Struktura genov pri prokariontih: operoni





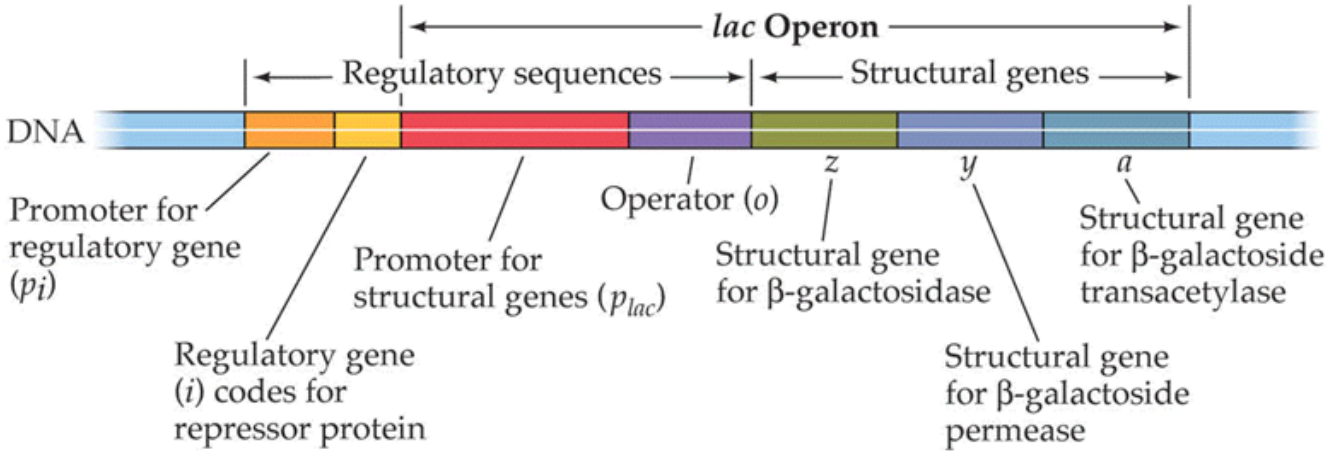
DNA:

- **Regulatory sequences:** site for the binding of regulatory proteins; the role of regulatory proteins is to influence the rate of transcription. Regulatory sequences can be found in a variety of locations.
- **Promoter:** site for RNA polymerase binding; signals the beginning of transcription.
- **Terminator:** signals the end of transcription.

mRNA:

- **Ribosomal binding site:** site for ribosome binding; translation begins near this site in the mRNA. In eukaryotes, the ribosome scans the mRNA for a start codon.
- **Start codon:** specifies the first amino acid in a polypeptide sequence, usually a formylmethionine (in bacteria) or a methionine (in eukaryotes).
- **Codons:** 3-nucleotide sequences within the mRNA that specify particular amino acids. The sequence of codons within mRNA determines the sequence of amino acids within a polypeptide.
- **Stop codon:** specifies the end of polypeptide synthesis.
- Bacterial mRNA may be polycistronic, which means it encodes two or more polypeptides.

Organization of sequences of a bacterial gene and its mRNA transcript. This figure depicts the general organization of sequences that are needed to create a functional gene that encodes an mRNA.



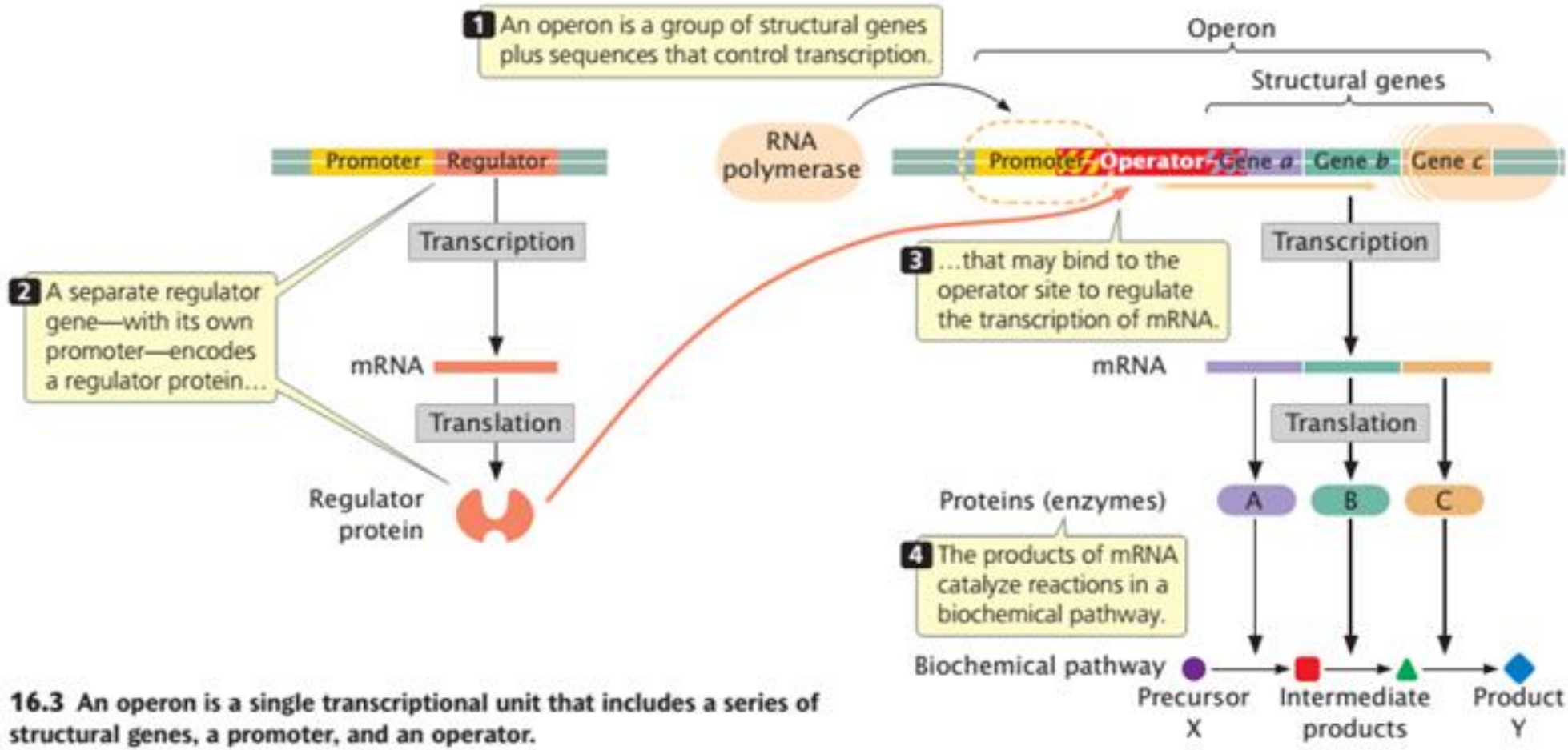
LIFE: THE SCIENCE OF BIOLOGY, Seventh Edition, Figure 13.16 The *lac* Operon of *E. coli*

Bacterial Operons Are Coregulated Gene Clusters

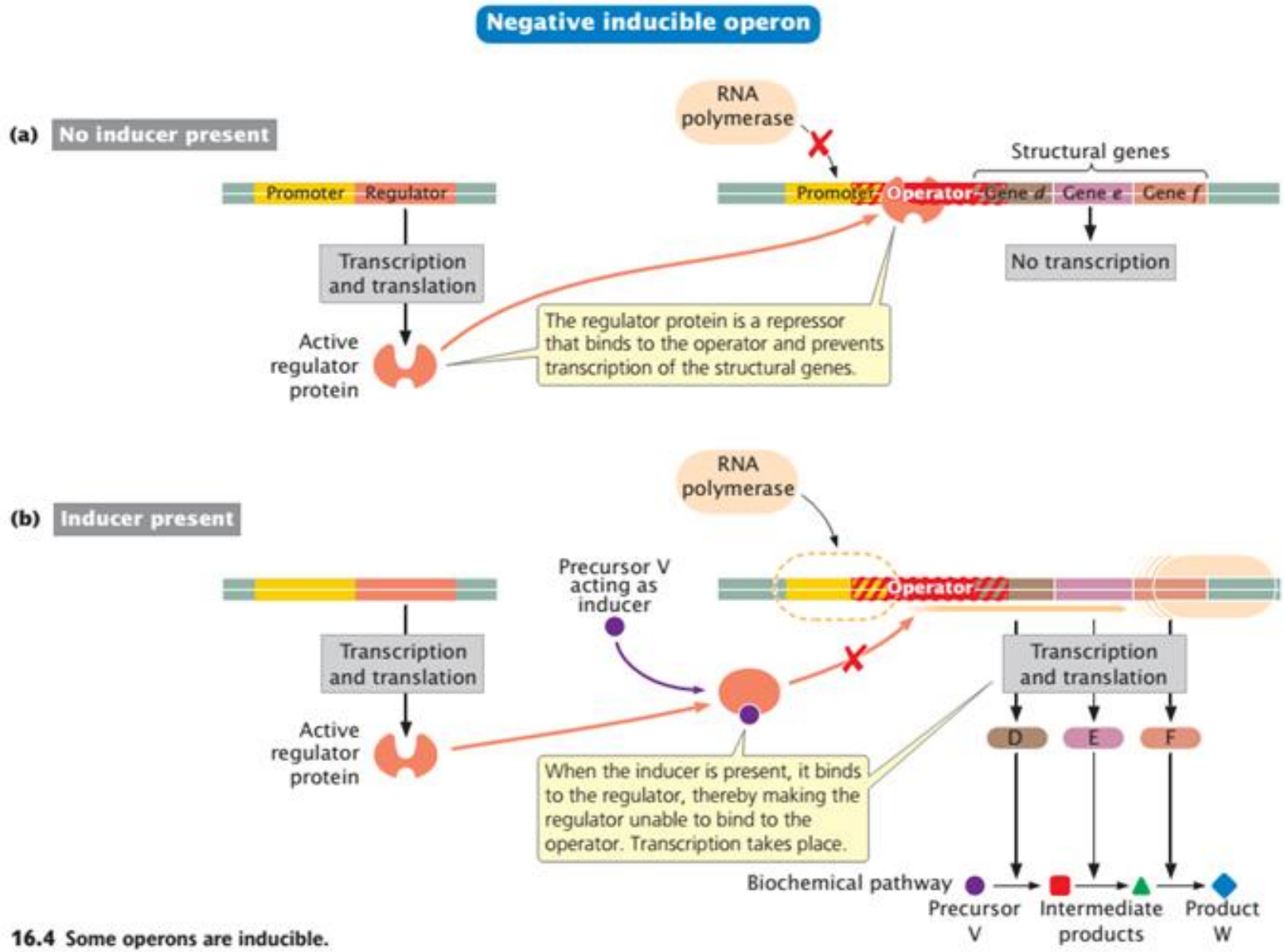
Bacterial genes are organized into operons, or clusters of coregulated genes. In addition to being ***physically close in the genome, these genes are regulated such that they are all turned on or off together***. Grouping related genes under a common control mechanism ***allows bacteria to rapidly adapt to changes in the environment***.

The organization of genes into an operon allows for ***simultaneous expression of all the genes that are located in cis*** (i.e., on the same contiguous piece of DNA) in the operon. Several features contribute to this characteristic of operons. First, all of the operon's genes are downstream of a single promoter.

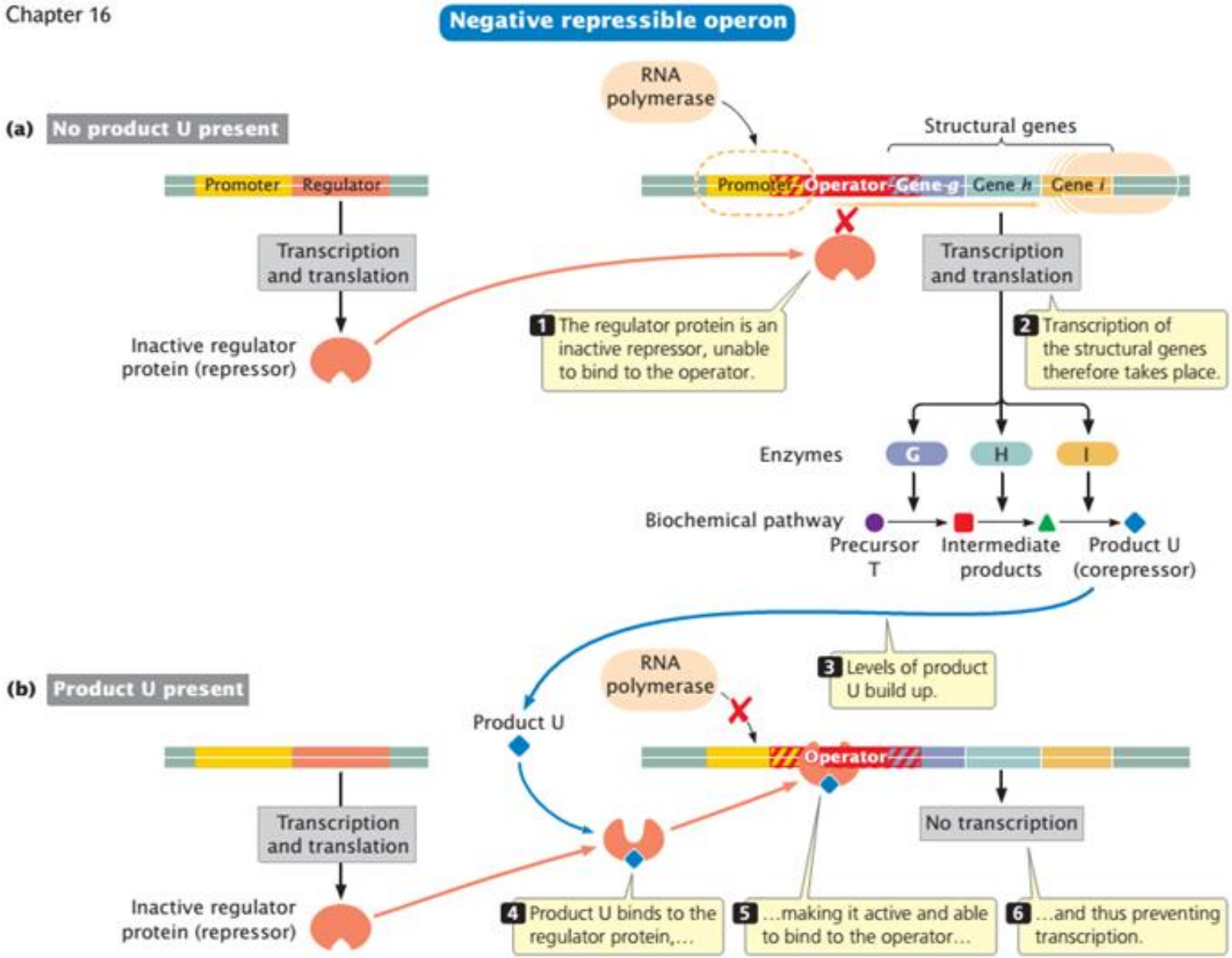
Thus, although bacteria may be considered simpler organisms than humans, it is clear that ***bacterial gene regulation is extremely efficient and that the bacterial genome is highly organized. Bacteria appear to be perfectly adapted to a variety of environments, and they are ready to respond to whatever environmental changes they encounter by employing elegant and complex regulatory mechanisms***.



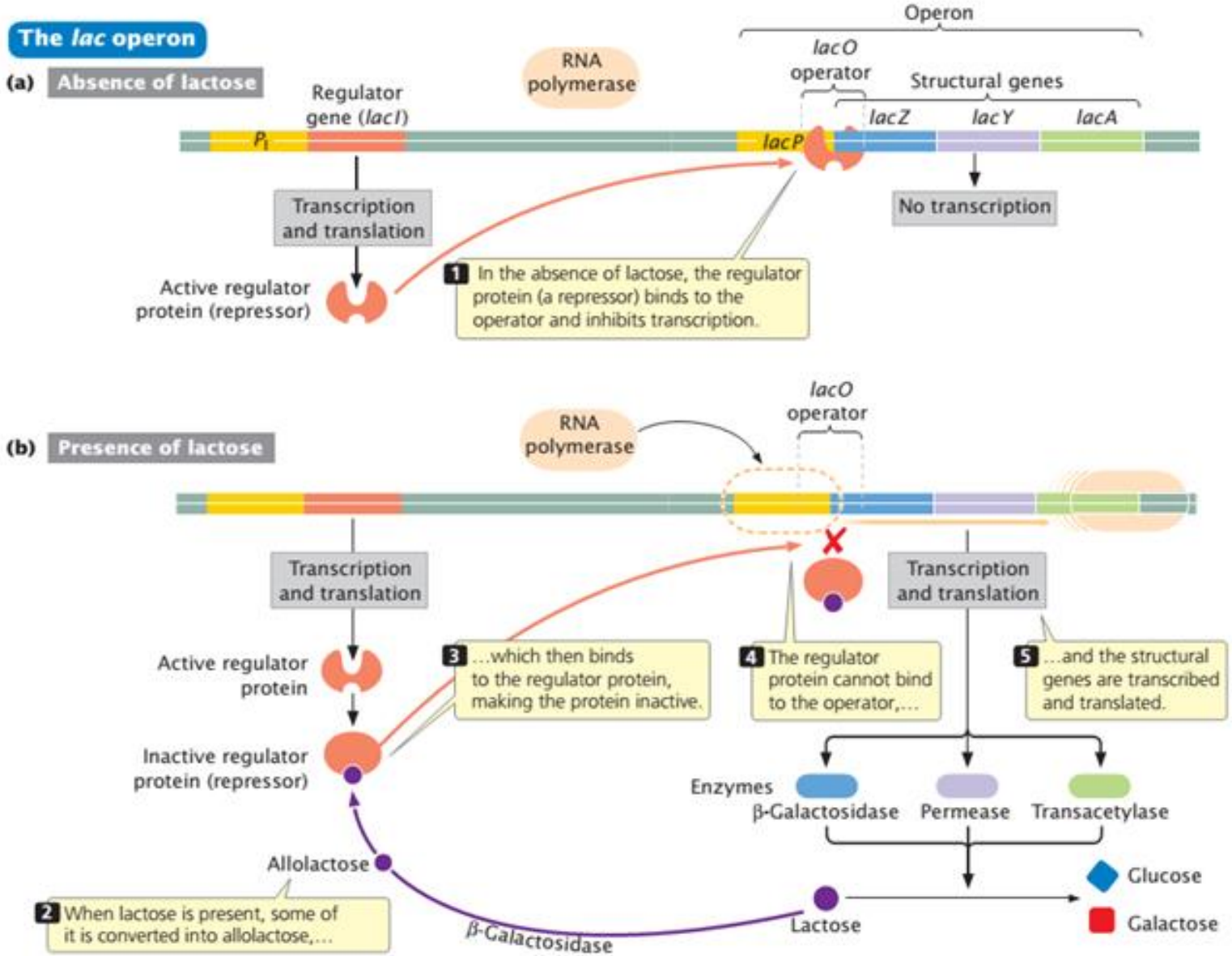
16.3 An operon is a single transcriptional unit that includes a series of structural genes, a promoter, and an operator.



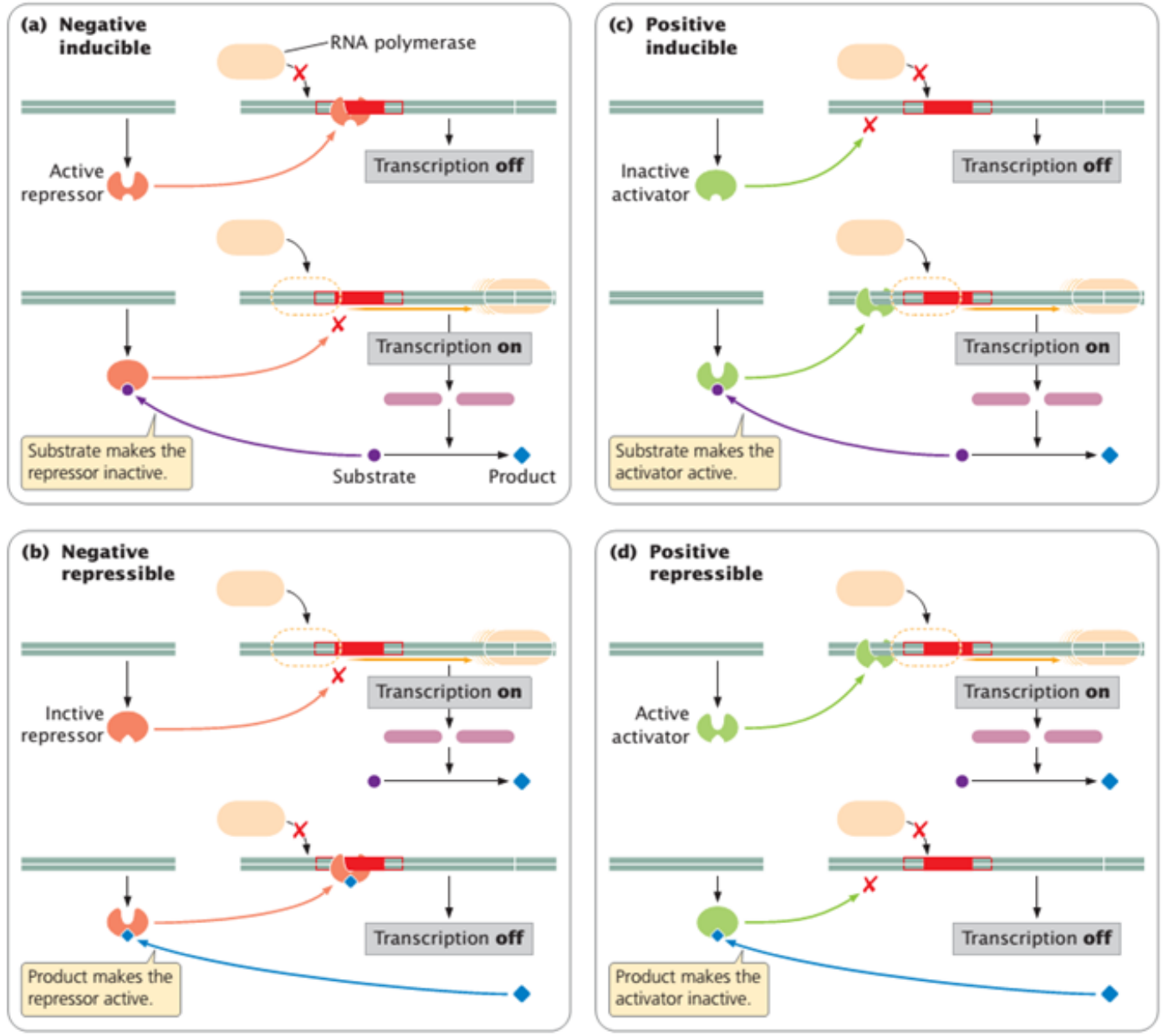
16.4 Some operons are inducible.



16.5 Some operons are repressible.



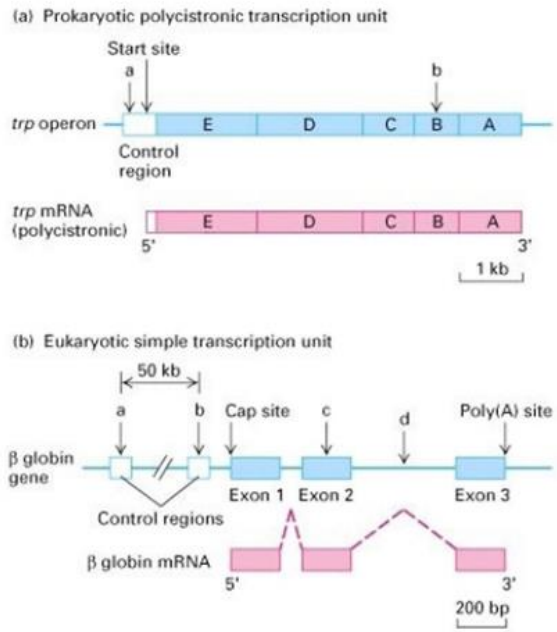
16.8 The *lac* operon regulates lactose metabolism.

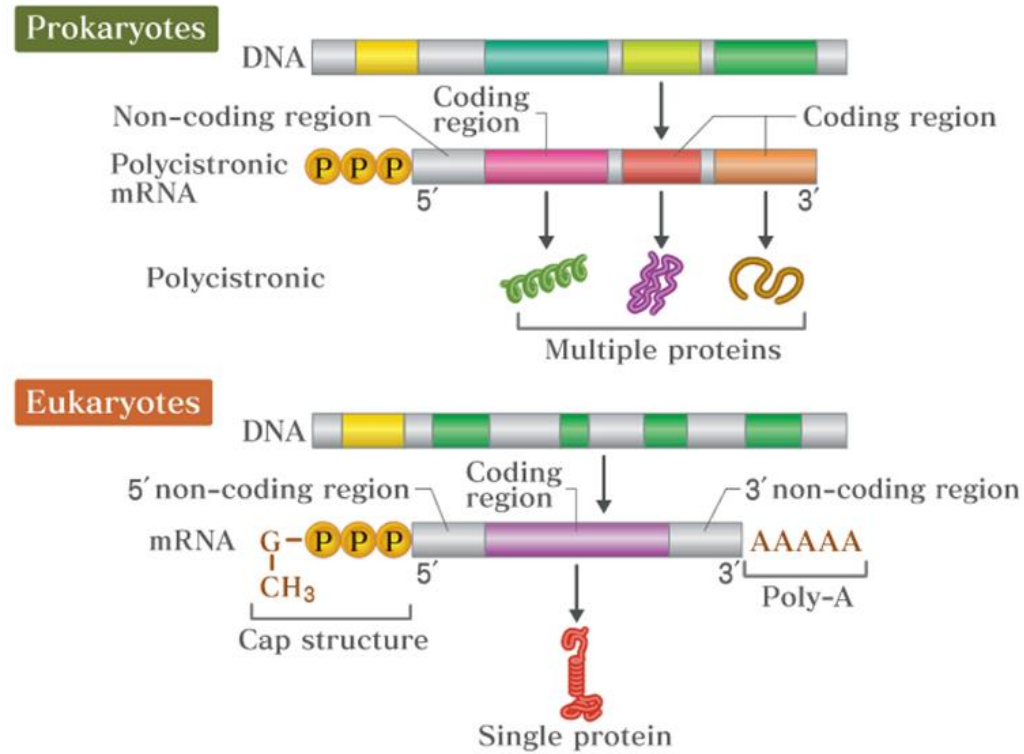


16.6 A summary of the characteristics of positive and negative control in inducible and repressible operons.

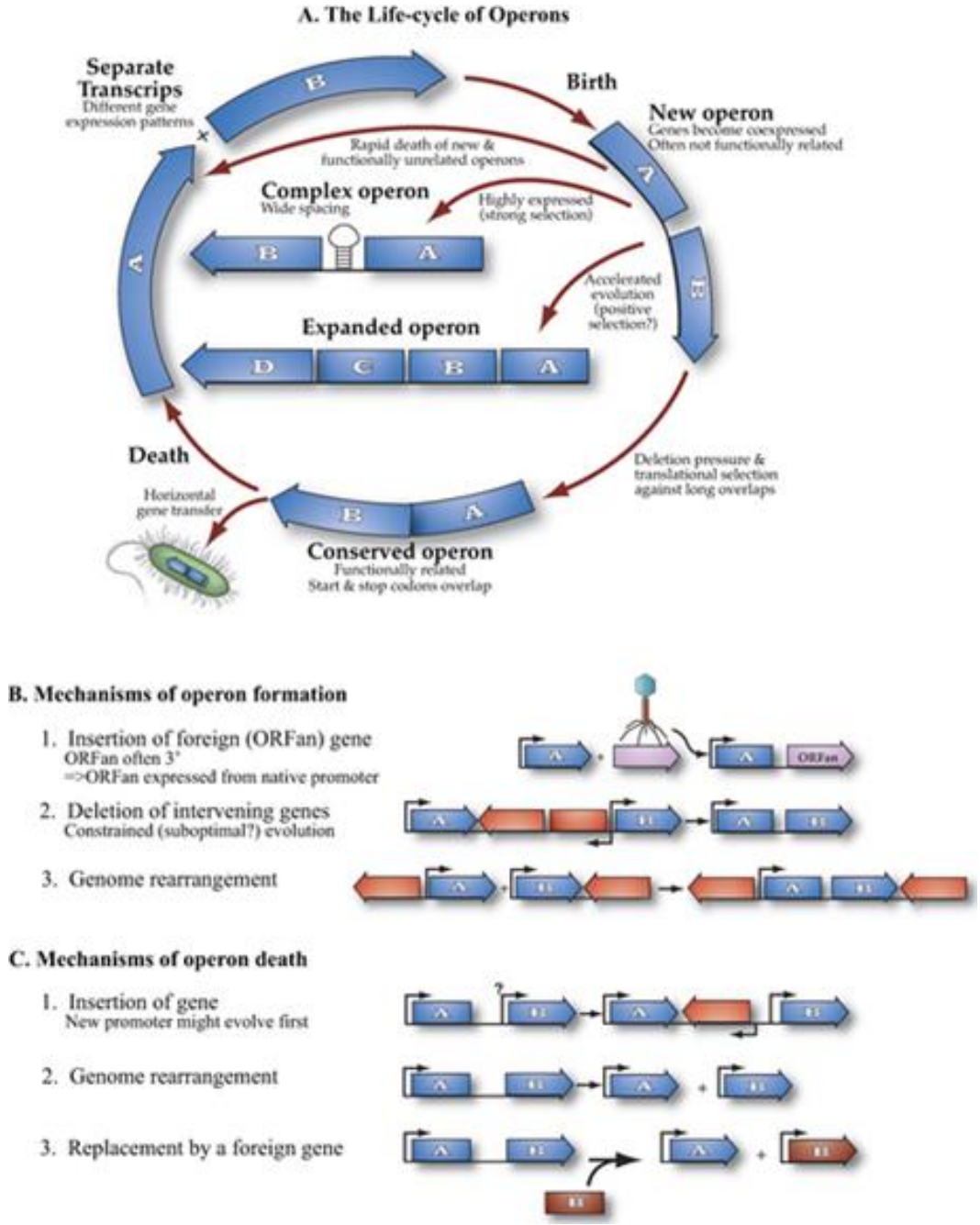
Gene Structure: Polycistronic vs monocistronic

- **Cistron**: an old name for a gene
- **Polycistronic**: 1 promoter directs synthesis of 1 mRNA that can be translated to more than one polypeptide
  - Prokaryotic genes
- **Monocistronic**: 1 promoter directs synthesis of 1 mRNA that **usually** translates to only 1 protein
  - Eukaryotic genes

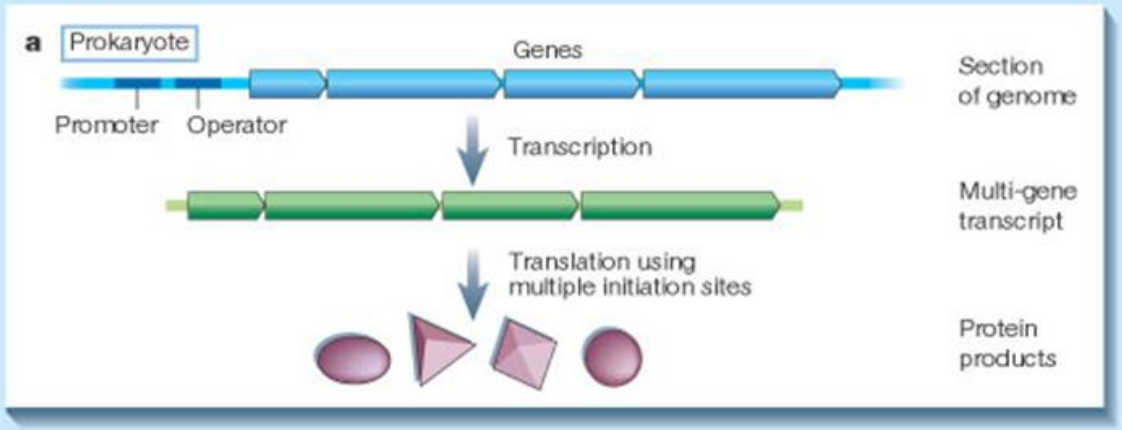


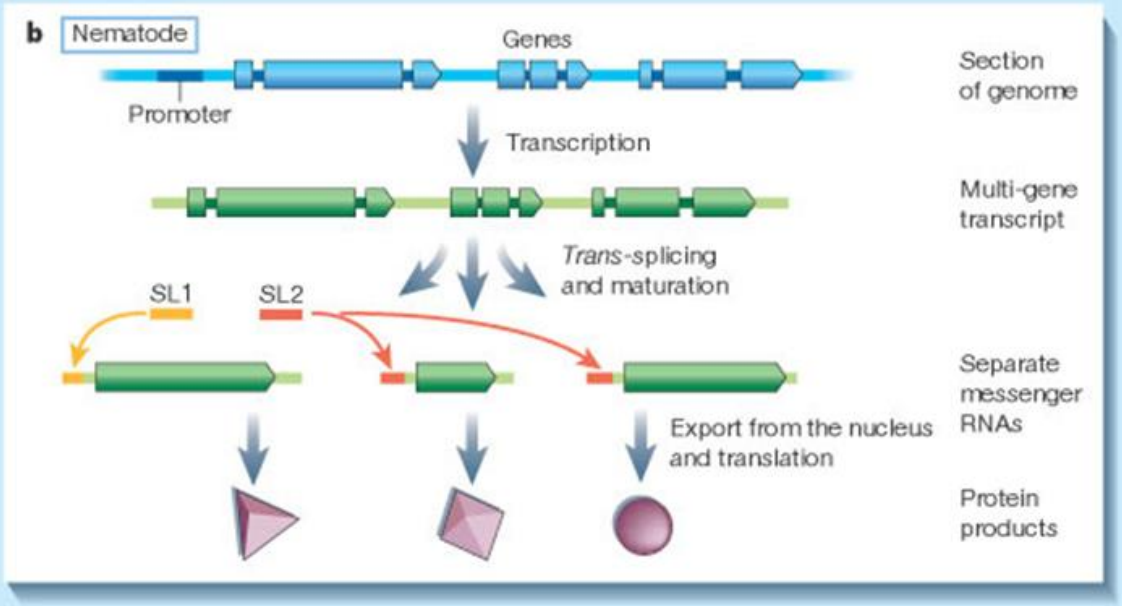


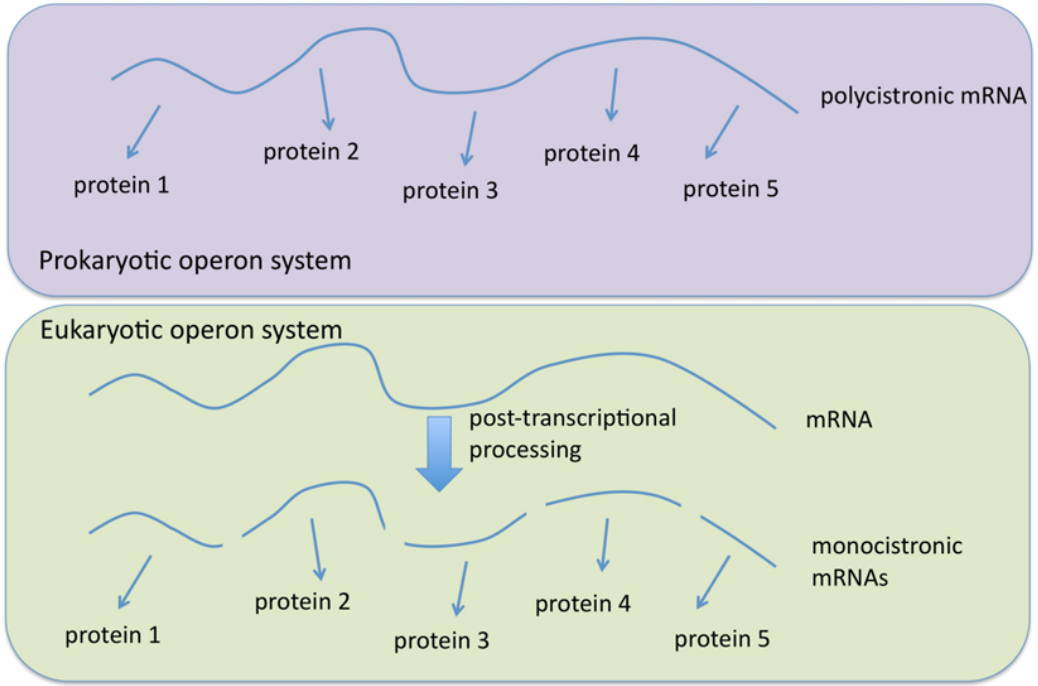
A Model for Operon Evolution



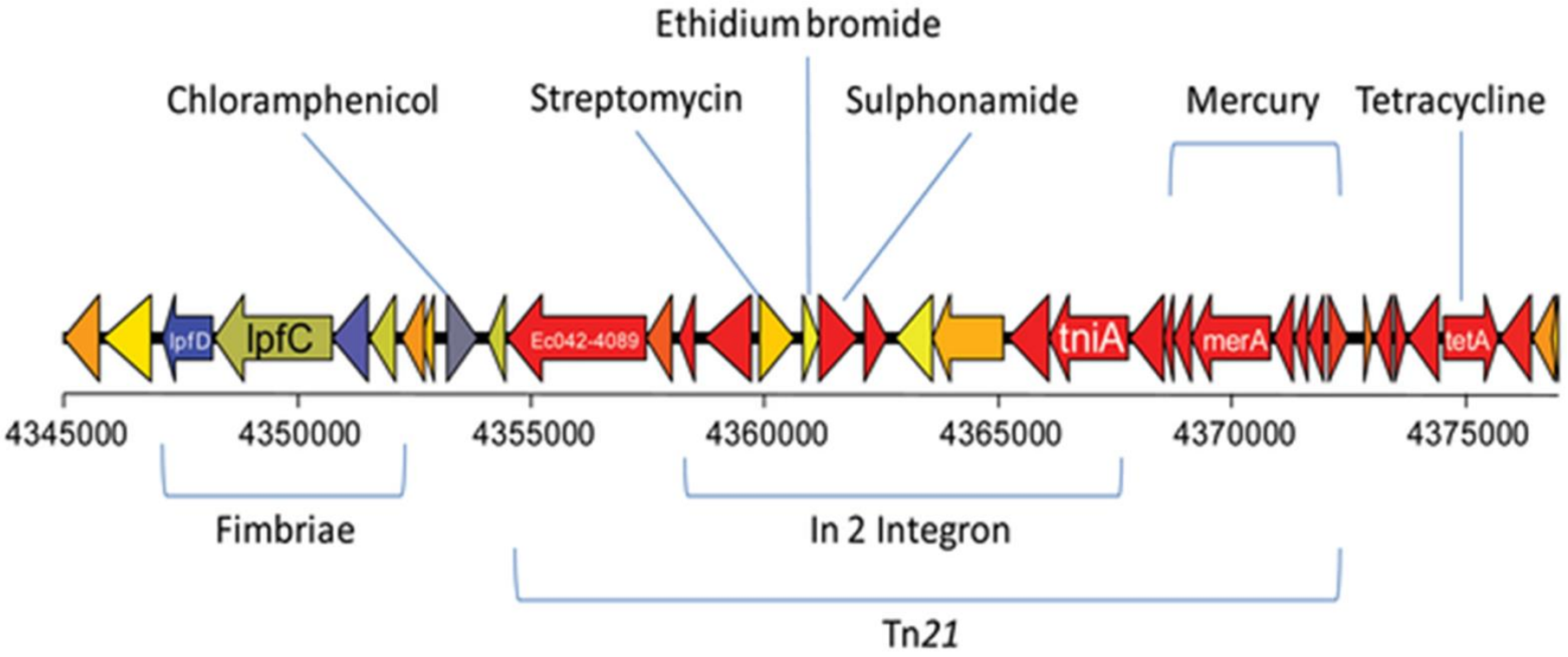
Operons in prokaryotes and eukaryotes





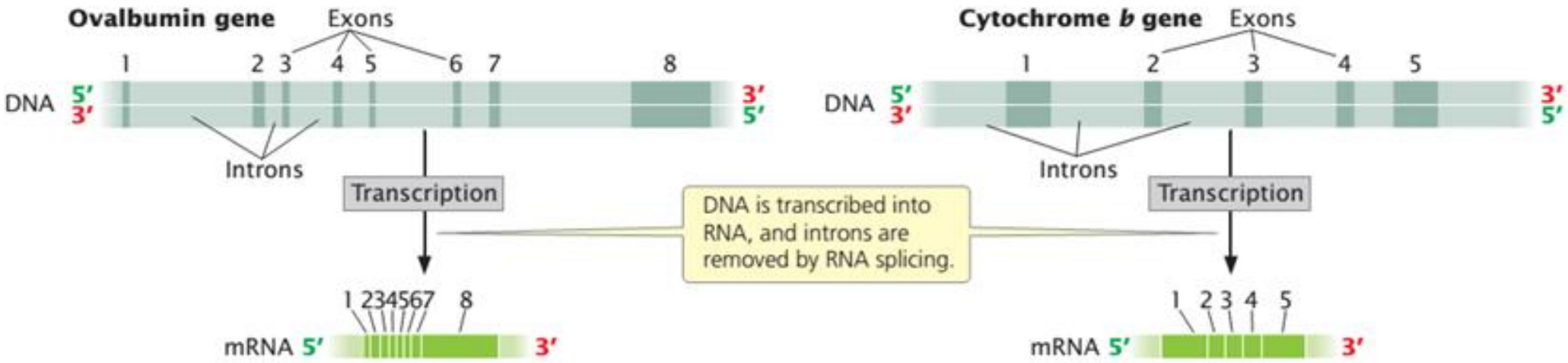


Gene organization of the Tn21 element containing loci encoding antibiotic resistance



Enteroaggregative *Escherichia col*i Strain 042 (EAEC)

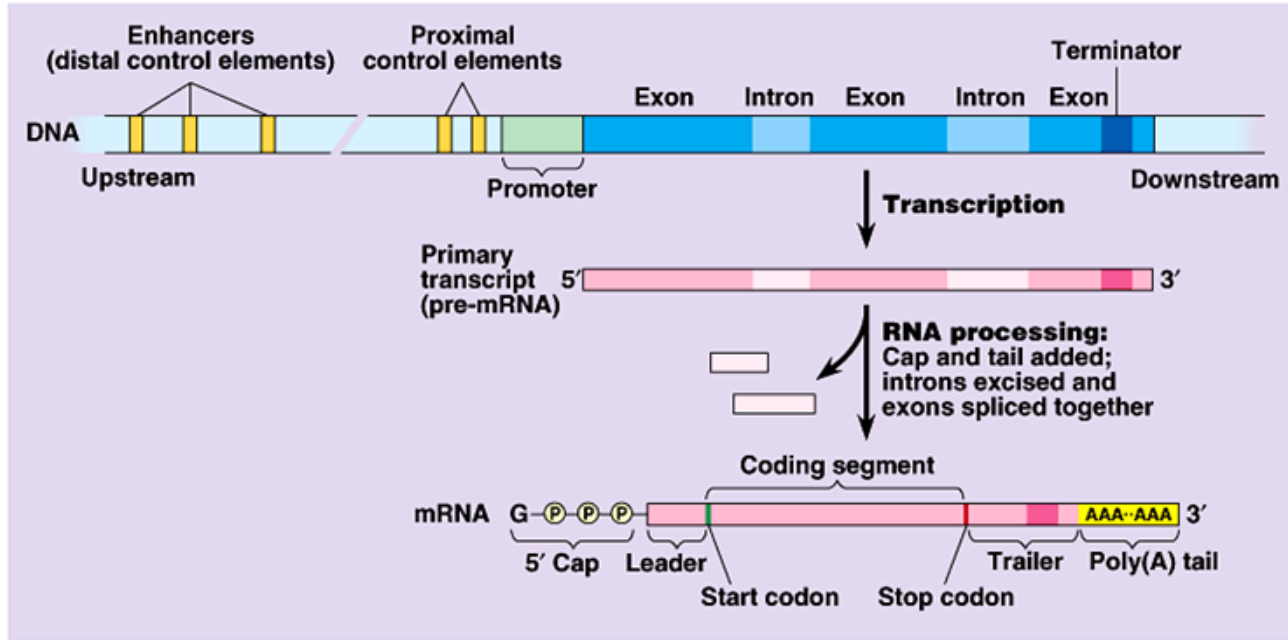
Struktura genov pri eukariontih: eksoni in introni, RNA (intron) splicing in alternativni splicing

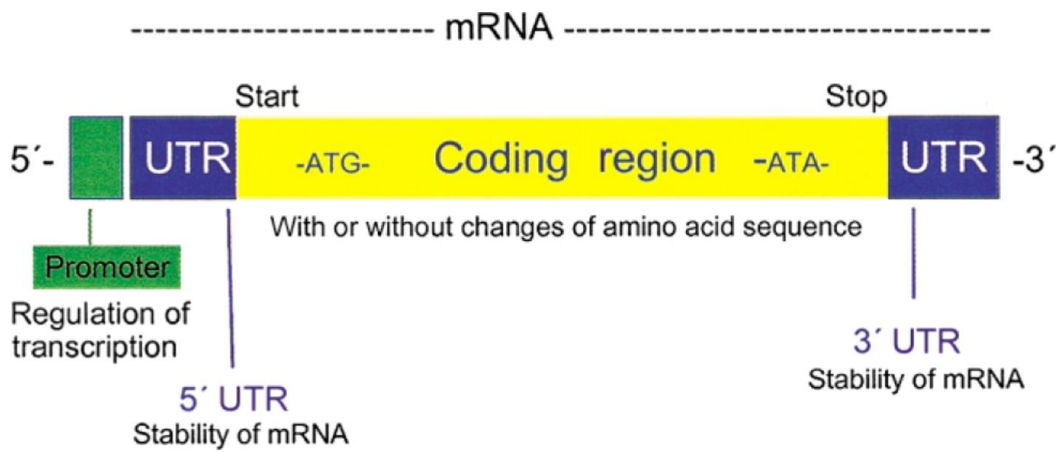


14.3 The coding sequences of many eukaryotic genes are disrupted by noncoding introns.

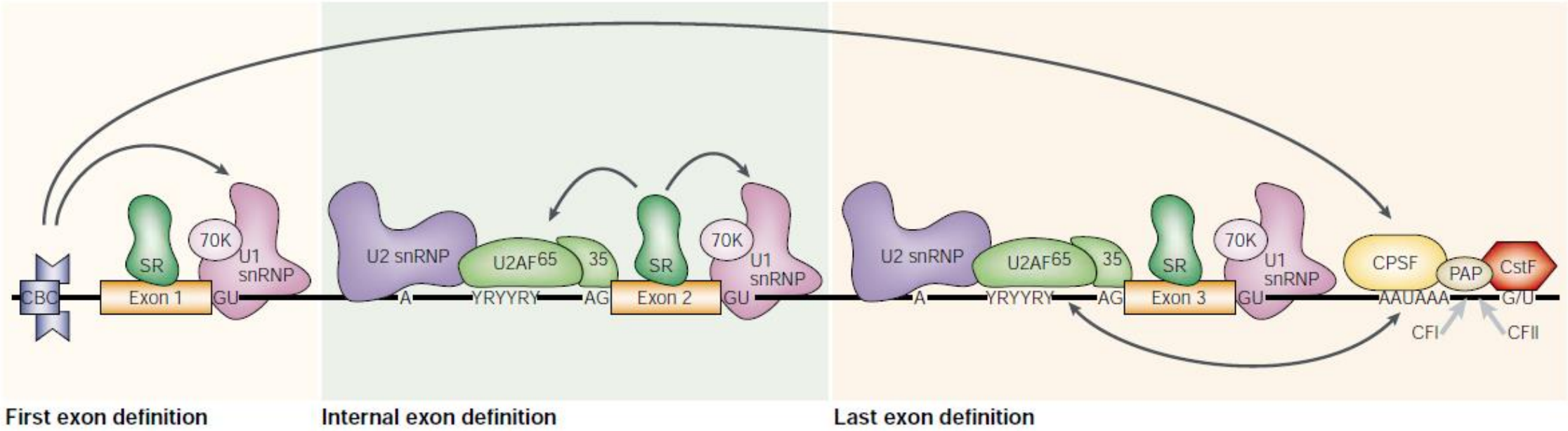
INTRON: An intervening non-coding sequence that interrupts two exons and that must be excised from pre-messenger RNA transcripts before translation.

EXON: The segment of a pre-messenger RNA transcript that contains protein-coding sequence and/or the 5' or 3' untranslated sequences, which must be spliced together with other exons to produce a mature messenger RNA.



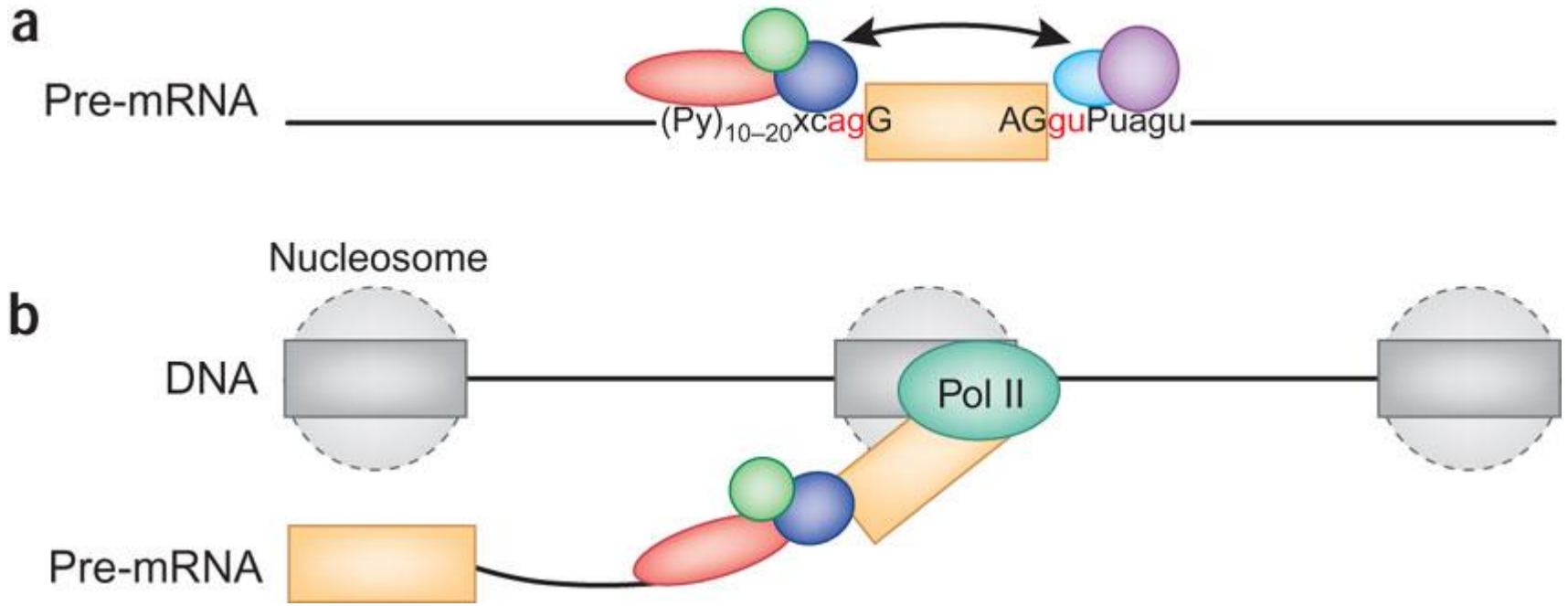


Kako so definirani (prepoznani) eksoni in introni, splicing signals

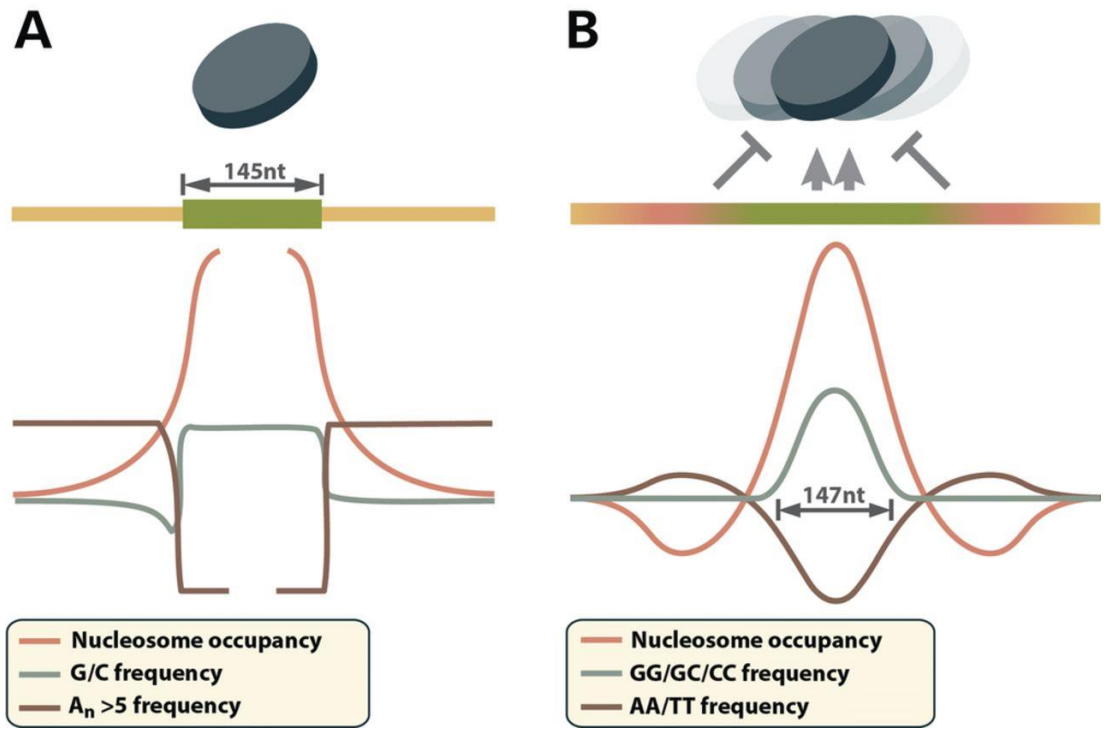


Exon-definition model. Typically, in vertebrates, exons are much shorter than introns. **According to the exon-definition model, before introns are recognized and spliced out, each exon is initially recognized by the protein factors that form a bridge across it. In this way, each exon, together with its flanking sequences, forms a molecular, as well as a computational, recognition module** (arrows indicate molecular interactions). CBC, cap-binding complex; CFI/II, cleavage factor I/II; CPSF, cleavage and polyadenylation specificity factor; CstF, the cleavage stimulation factor; PAP, poly(A) polymerase; snRNP, small nuclear RNP; SR, SR protein; U2AF, U2 small nuclear ribonucleoprotein particle (snRNP) auxiliary factor.

Recognition of the terminal exons. ***Definition of the 5′-terminal exon*** requires the cap-binding protein complex (CBC) to communicate with the 5′ss. On the other hand, ***definition of the 3′-terminal exon*** needs the communication between the 3′ss and the poly(A) factors (among which the cleavage/polyadenylation specificity factor (CPSF), cleavage stimulation factor (CstF), cleavage factor (CF), and poly(A) polymerase (PAP)).

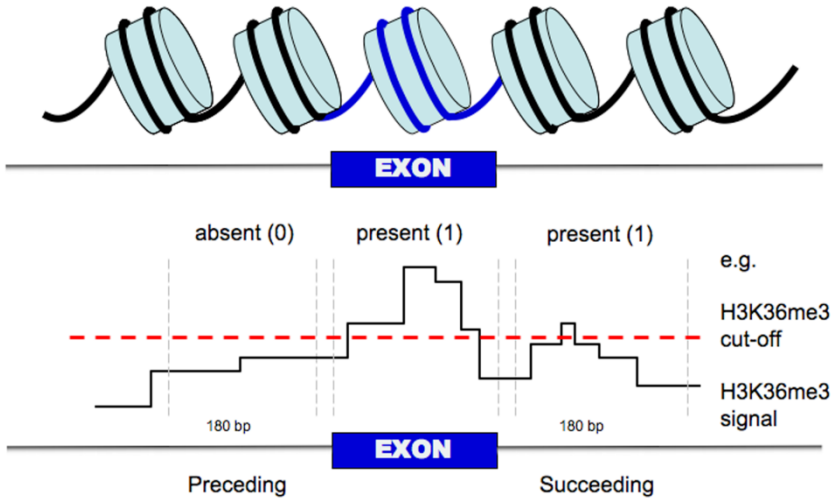


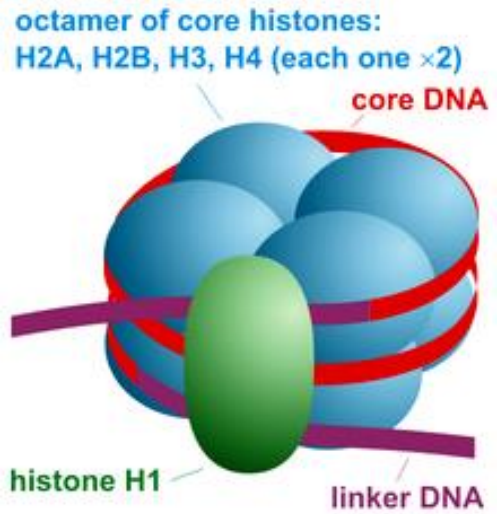
Possible mechanisms for exon definition. (a) ***Exon definition is achieved at the pre-mRNA level***. ***Spliceosomal and auxiliary factors are recruited to the splice sites flanking an exon on the mRNA precursor.*** Direct and indirect interactions between the 3' and 5' complexes favor exon recognition and splicing and exert selective pressure for a conserved exon length of 140–150 bp. Pu, purine; Py, pyrimidine. (b) **Nucleosomes (broken circles) are preferentially bound to exons**, whereas introns are mostly devoid of nucleosomes. ***Exons are therefore marked at the DNA level by nucleosome positioning***, which may act as 'speed bumps' for RNA polymerase II, helping in the co-transcriptional recruitment of splicing factors to the nascent pre-mRNA and improving exon definition. ***As nucleosomes accommodate DNA stretches of approximately 147 bp, their preferential location on exons may act as the selective pressure factor for the conservation in exon length.***



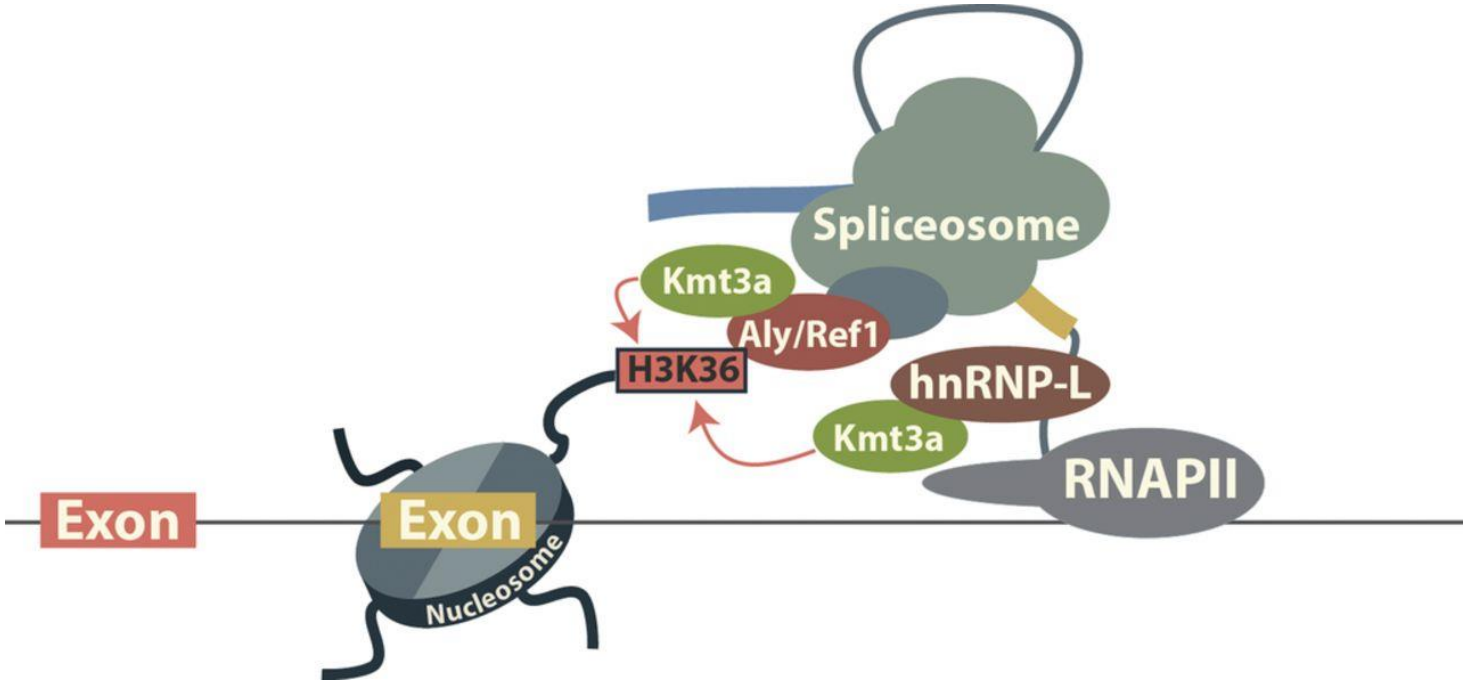
Nucleotide composition of exons and nucleosome 'container sites'. ***(A) Exons contain elevated GC content relative to introns, which favors nucleosome binding***, while nearby intron sequences are enriched in nucleosome repellant poly-dA:dT stretches. This sequence arrangement mirrors the recently identified ***strong nucleosome-positioning signals termed 'container sites' (B)***, which consist of relatively G/C-rich nucleosome binding core, surrounded by nucleosome repellant A/T-rich sequence that prevents nucleosomes from sliding out of position.

Exons are characterized by increased levels of nucleosome positioning, DNA methylation and certain histone modifications.

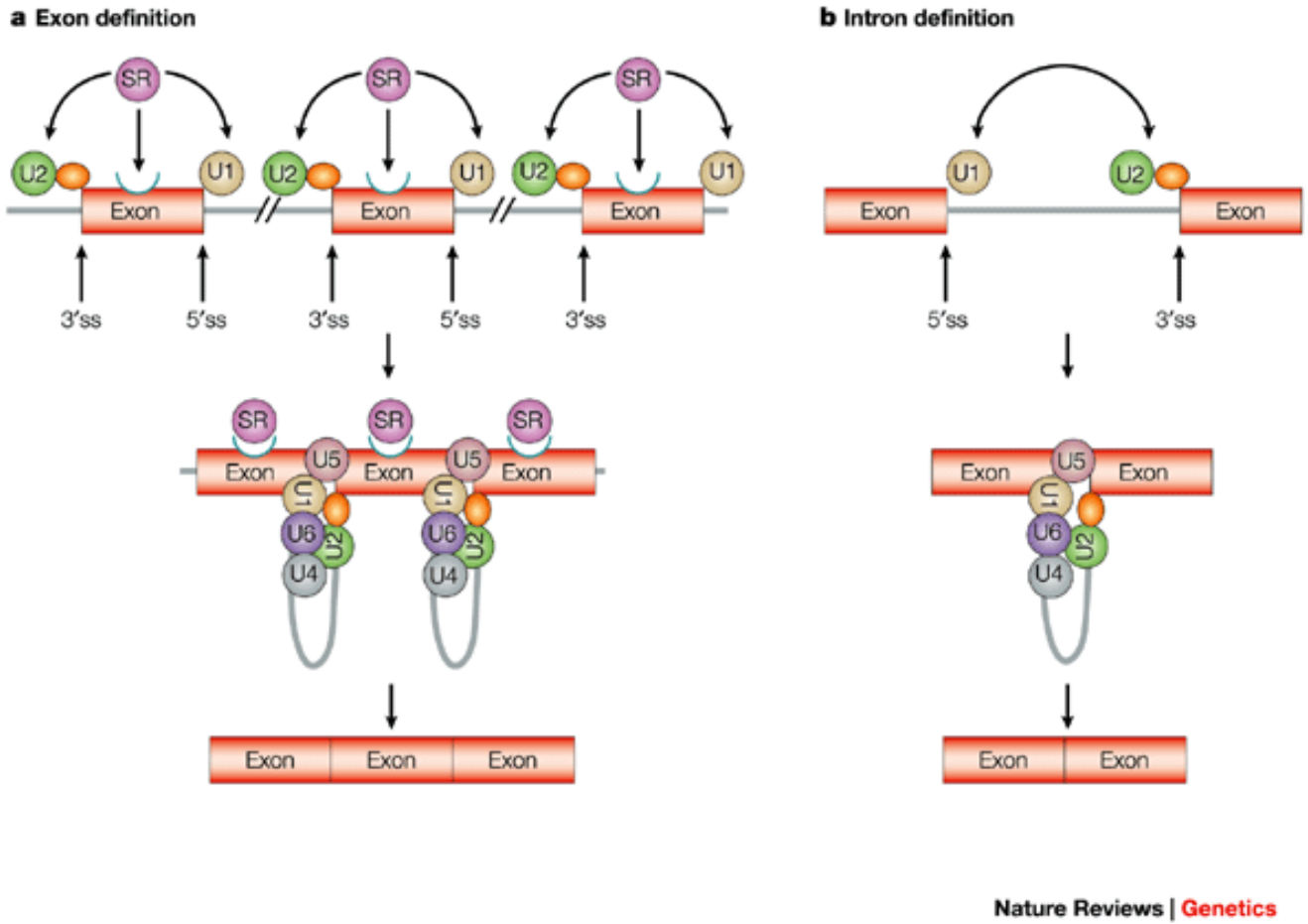




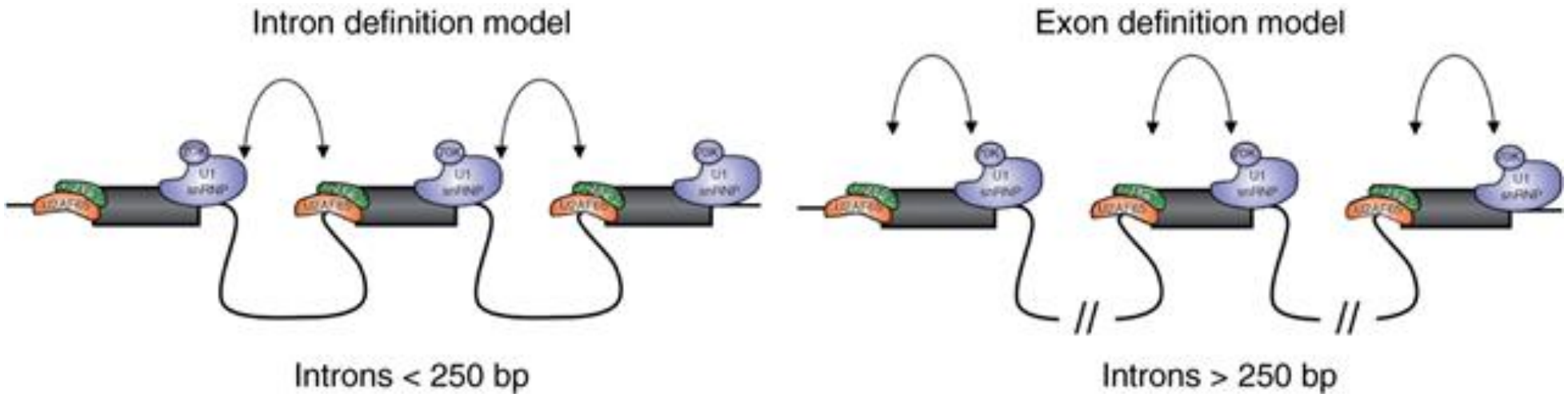
Nucleosome



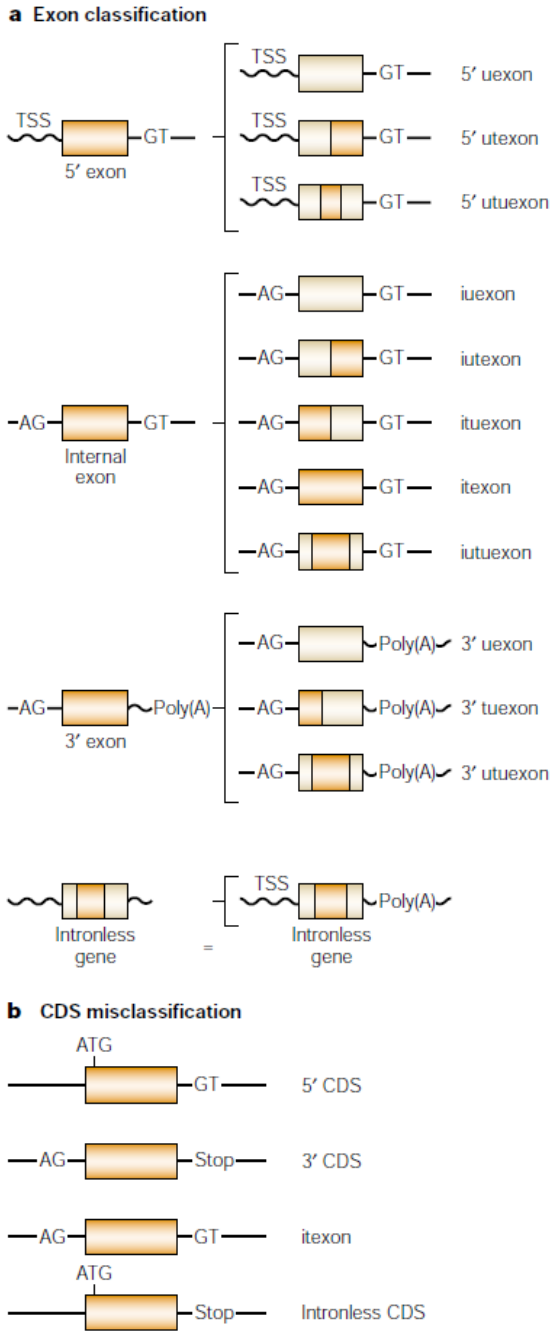
Deposition of H3K36me3 over exons. hnRNP-L directly interacts with RNA elements that place it in close proximity to the exons. Aly/Ref1 is recruited to the pre-mRNA during spliceosome assembly. In turn, hnRNP-L and Aly/Ref1 cooperate with RNAPII to bring Kmt3a to the exons where it methylates H3K36.



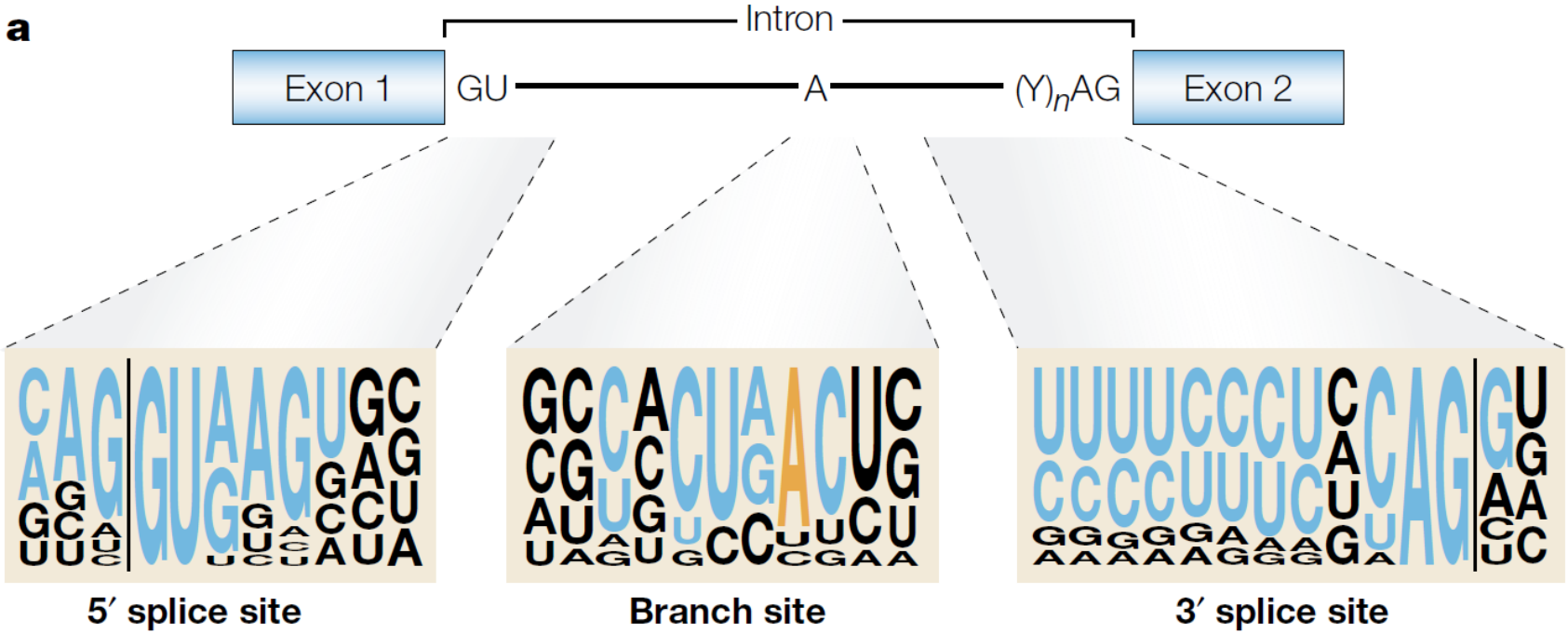
Exon and intron definition. a | ***Exon definition***: SR proteins (purple) bind to exonic splicing enhancers (ESE; blue), recruiting U1 to the downstream 5'ss and the splicing factor U2AF (orange) to the upstream polypyrimidine tract and the 3'ss. U2AF then recruits U2 to the branch site. Therefore, **when the SR proteins bind the ESE, they promote formation of a 'cross-exon' recognition complex by placing the basal splicing machinery in the splice sites that flanked the same exon**. b | ***Intron definition***: the binding of U1 to the upstream 5' splice site (ss) and U2AF and U2 to the downstream polypyrimidine tract and branch site, respectively, of the same intron. Therefore, **intron definition selects pairs of splice sites located on both ends of the same intron**, and SR proteins can also mediate this process.



Exon and Intron definition models. The left panel depicts the ***Intron definition model*** according to which pairing between the splice sites takes place across an intron when ***long exons are separated by short (<250 bp) introns***. On the other hand the right panel shows the ***Exon definition model*** where the ***splice site communication occurs across exons when they are separated by long (>250 bp) introns***.

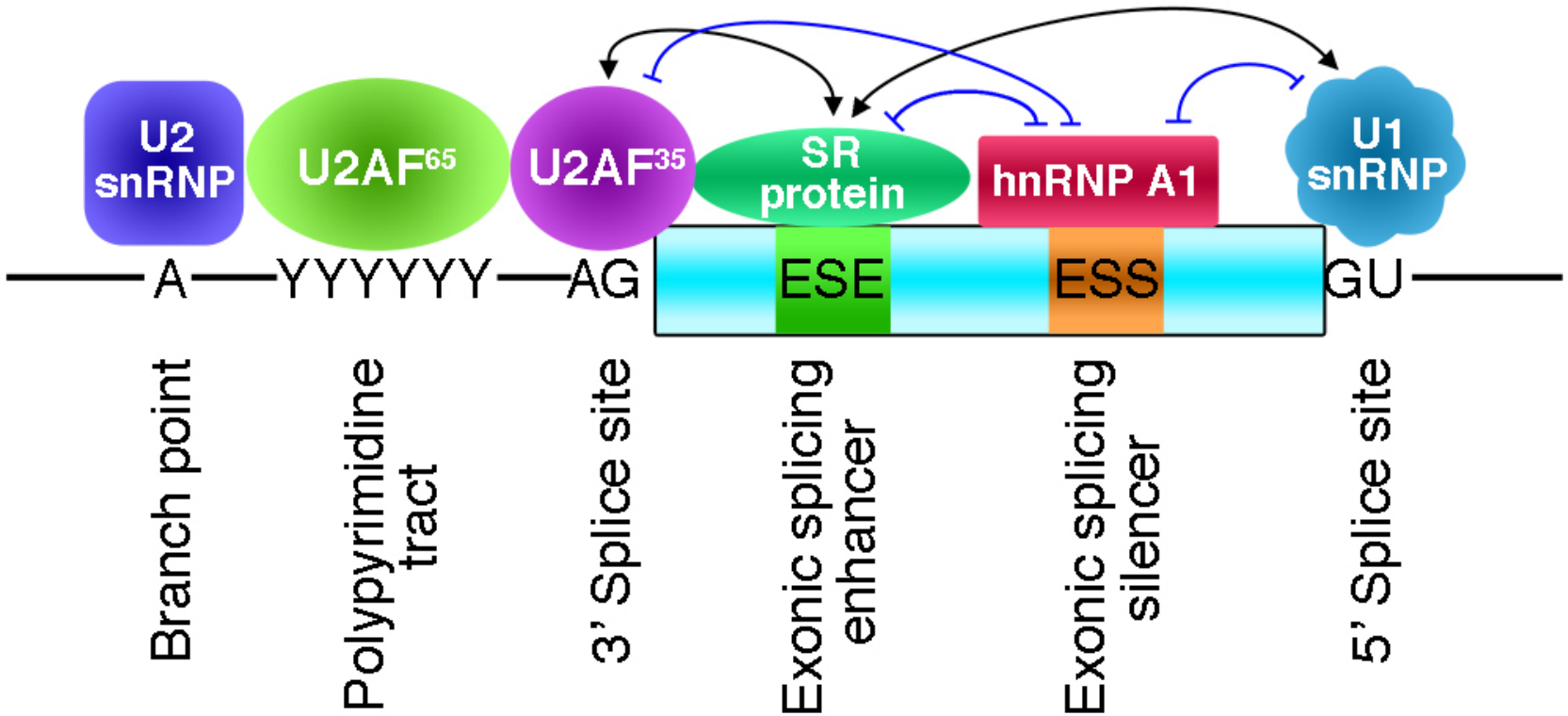


Exon classification. a | Exons can be classified into ***four classes and 12 subclasses***, as shown. b | Coding sequence (CDS) 'exons'. Four classes of exon-coding regions. These regions are not whole exons, except for the internal coding exons (itexons). i, internal; poly(A), polyadenylation; t, translated; TSS, transcription start site; u, untranslated.

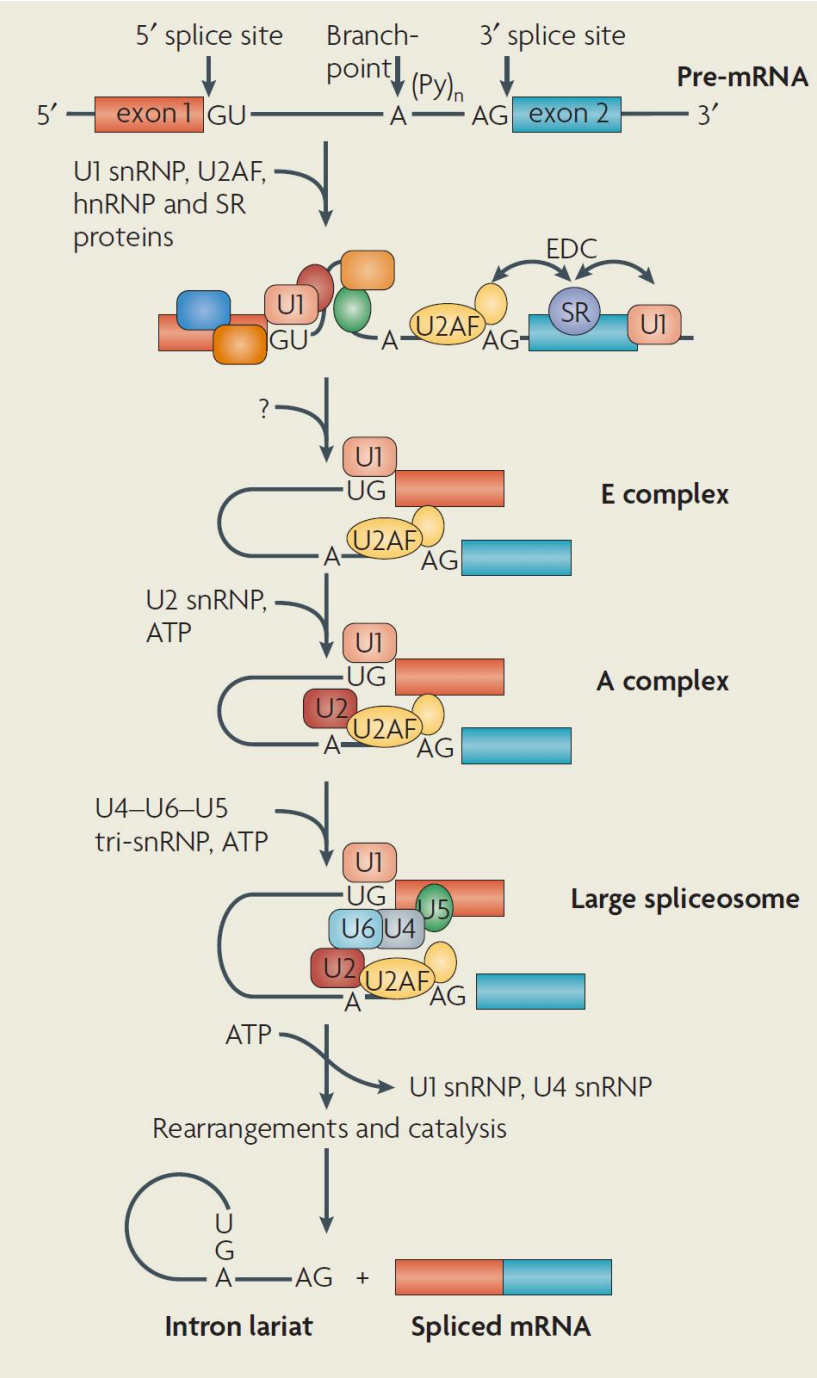


Classical splicing signals. a | ***Conserved motifs at or near the intron ends***. The ***nearly invariant GU and AG dinucleotides at the intron ends, the polypyrimidine tract (Y)n preceding the 3′ AG***, and the A residue that serves as a branchpoint are shown in a two-exon pre-mRNA. The sequence motifs that surround these conserved nucleotides are shown below. For each sequence motif, the size of a nucleotide at a given position is proportional to the frequency of that nucleotide at that position in an alignment of conserved sequences from 1,683 human introns. Nucleotides that are part of the classical consensus motifs are shown in blue, except for the branch-point A, which is shown in orange. The vertical lines indicate the exon–intron boundaries.

Splicing cis-elements and trans-factors



RNA splicing and alternative splicing



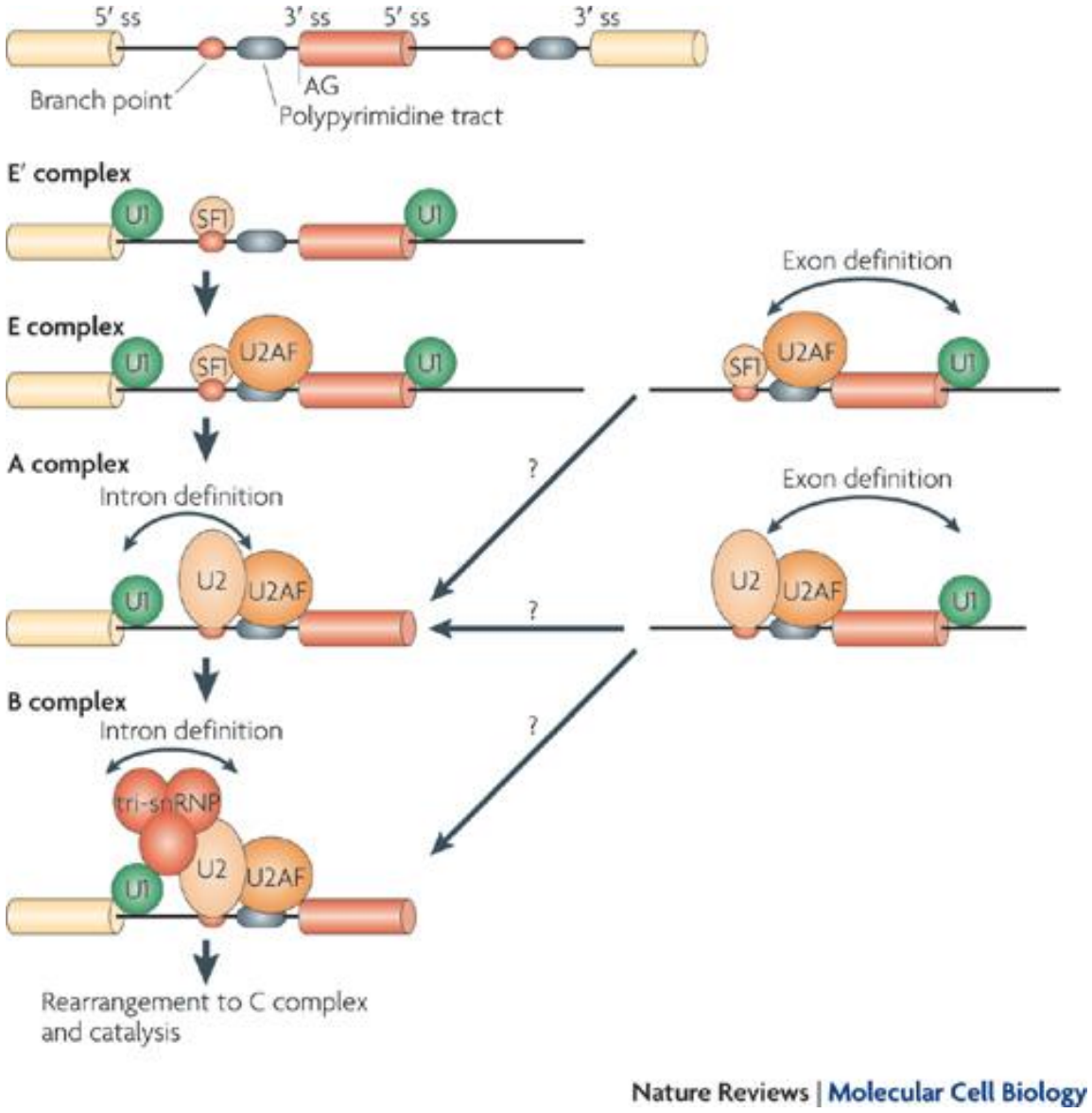
The pathway of spliceosome assembly in vitro

The ***5′ splice site of an intr***on is recognized by the U1 small nuclear ribonucleoprotein (snRNP), through complementary pairing of the RNA bases; the ***3′ splice site***, with its upstream polypyrimidine tract ((Py)n), is bound by the protein U2 auxiliary factor (U2AF).

These factors, together with additional proteins, form a ***complex called the E (for early) or commitment complex***, which ***bridges the intron*** and brings the splice sites that are to be cleaved and joined into juxtaposition. U2AF recruits the U2 snRNP, and an ATP-dependent step allows the RNA portion of the U2 snRNP to base pair with a sequence called the branchpoint, which lies upstream of the 3′ splice site. The pairing of the U2 snRNA at the branchpoint completes the pre-spliceosomal A complex. Subsequent, ATP-dependent steps lead to the binding of the U4–U5–U6 tri-snRNP and the formation of the large spliceosome; rearrangements that detach the U1 and U4 snRNPs then follow. This forms the catalytic spliceosome, which performs two transesterification reactions on the splice sites. ***These reactions result in the ligation of the exons and the excision of the intron***, ***in the form of a lariat RNA*** that has its 5′ phosphate joined to the 2′ hydroxyl at the branch point. *Spliceosome assembly and its excision chemistry must be performed on each intron for them to be excised from an mRNA.*

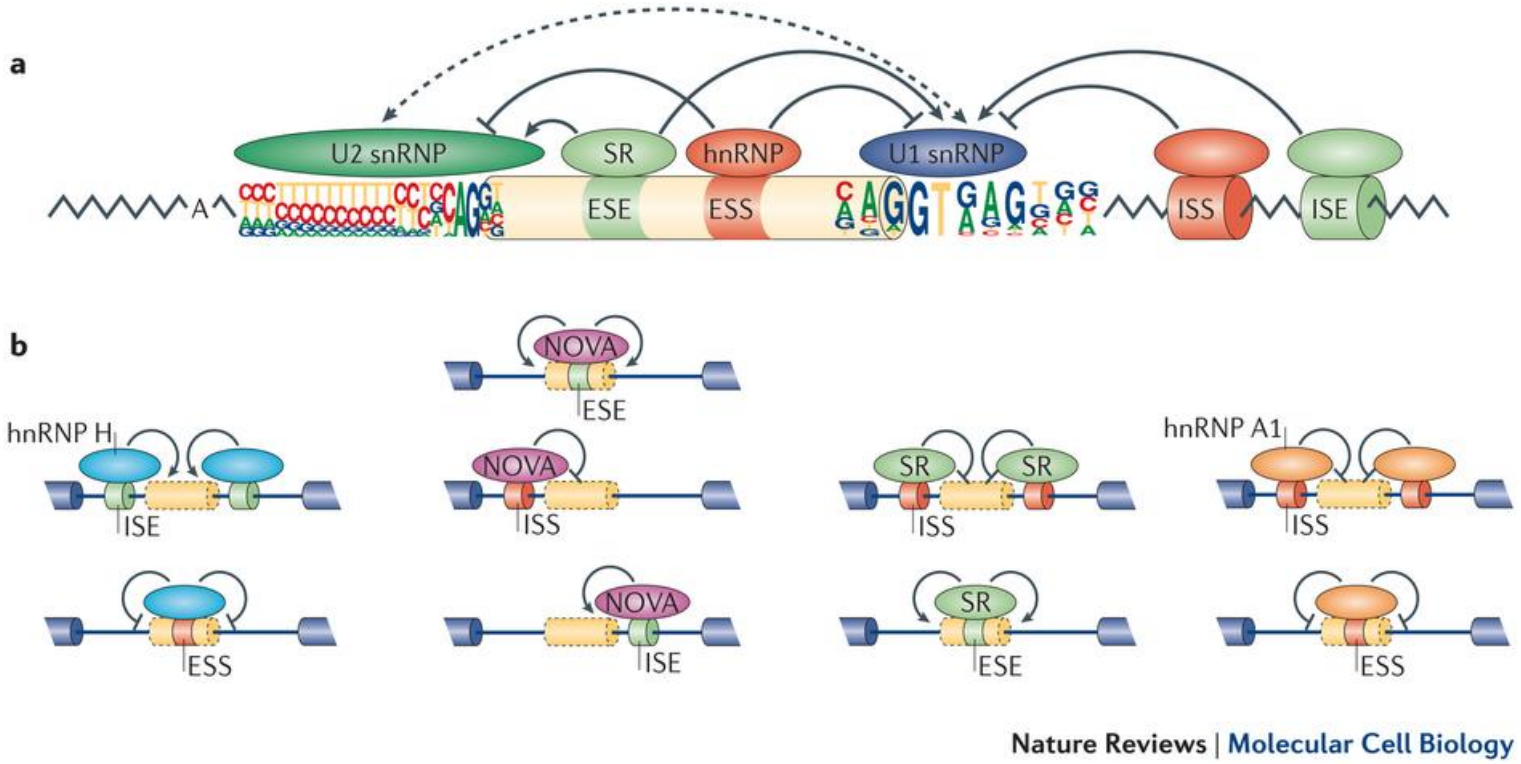
Before the spliceosome forms across the large length of an intron (introns are typically comprised of between 500 and 50,000 nucleotides), splice sites are usually defined through interactions between splice-site-bound components across the short (typically between 50 and 300 nucleotides) exons.

The assembly of these ‘exon definition’ complexes (EDCs) is stimulated by proteins that bind to special sequences in the exon itself (such as the serine–arginine (SR) proteins) and in the adjacent introns.

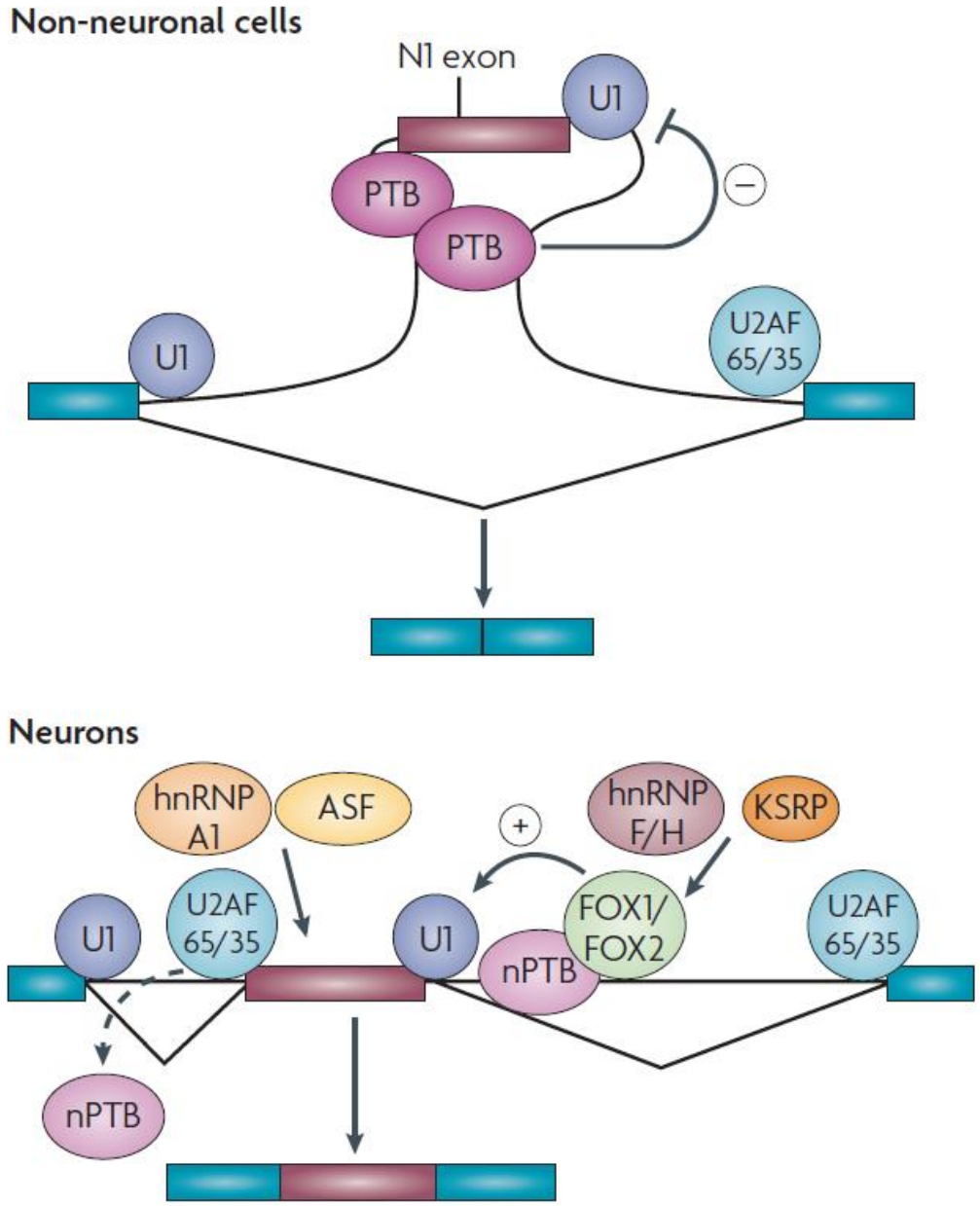


Splicing and spliceosome assembly. Pre-mRNA splicing is a process in which intervening sequences (introns) are removed from an mRNA precursor. ***Splicing consists of two transesterification steps, each involving a nucleophilic attack on terminal phosphodiester bonds of the intron.*** In the first step 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. 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 point3 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 AG158. 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***.

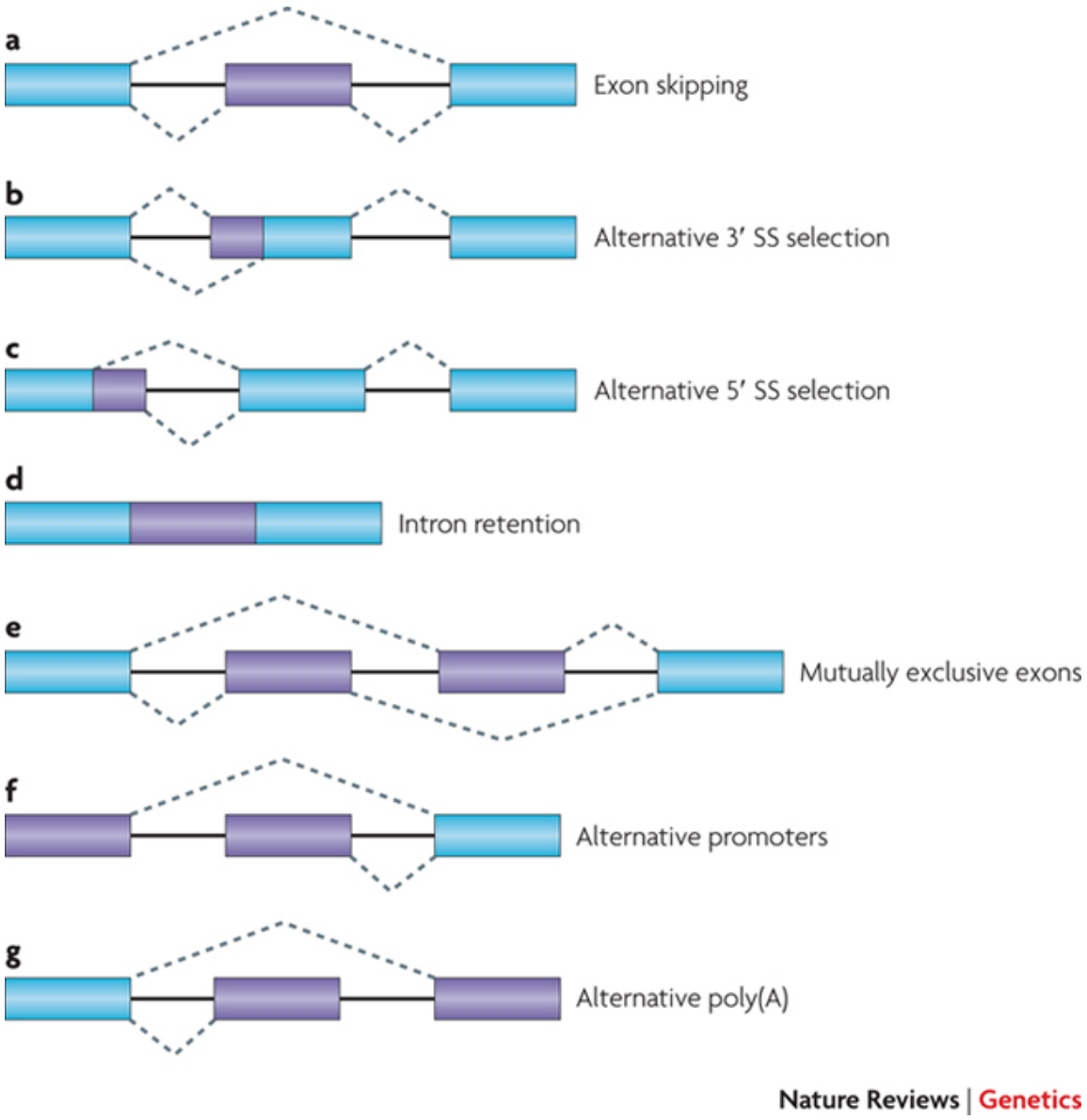


Regulation of alternative splicing. a | ***Splice site choice is regulated through cis-acting splicing regulatory elements (SREs) and trans-acting splicing factors.*** On the basis of their relative locations and activities, SREs are classified as exonic splicing enhancers (ESEs), intronic splicing enhancers (ISEs), exonic splicing silencers (ESSs) or intronic splicing silencers (ISSs). These SREs specifically recruit splicing factors to promote or inhibit recognition of nearby splice sites. ***Common splicing factors include SR proteins, which recognize ESEs to promote splicing, as well as various heterogeneous nuclear ribonucleoproteins (hnRNPs), which typically recognize ESSs to inhibit splicing***. Both often affect the function of U2 and U1 small nuclear RNPs (snRNPs) during spliceosomal assembly. The consensus motifs of splice sites are shown in the colored pictograph, with the height of each letter representing nucleotide frequency in each position. The dashed arrow represents the formation of the exon definition complex. b| ***The activity of splicing factors and cis-acting SREs is context-dependent***. Four well-characterized examples are shown. Oligo-G tracts, recognized by hnRNP H, function as ISEs to promote splicing when they are located inside an intron (top), and as ESSs when located within exons (bottom). YCAY motifs, recognized by neuro-oncological ventral antigen (NOVA), act as ESEs when located inside an exon (top), as ISSs when located in the upstream intron of an alternative exon (middle) and as ISEs when located in the downstream intron (bottom). Binding sites for SR proteins and hnRNP A1 also have distinct activities when located at different regions on the pre-mRNA.

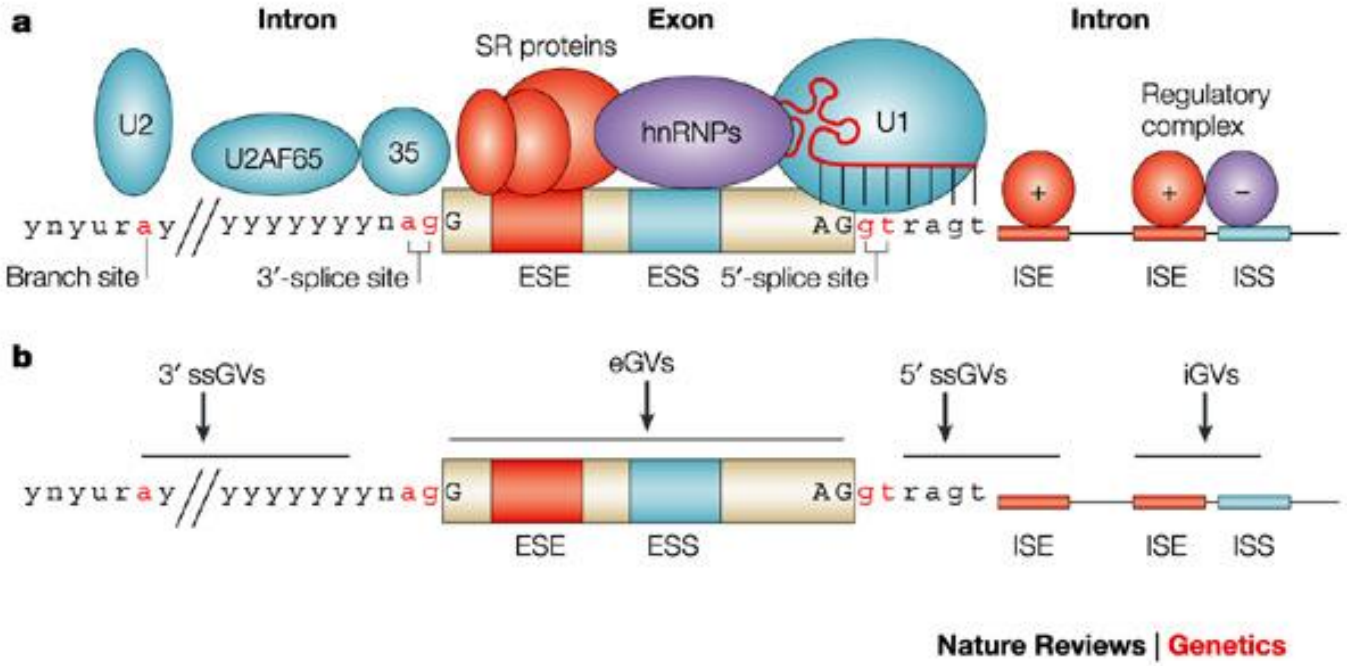


Alternative exons are controlled by multiple regulators. This figure illustrates the ***regulation of the splicing of the N1 exon of the SRC gene***. N1 is ***repressed in non-neuronal cells by polypyrimidine tract-binding protein (PTB)***, which binds to elements in the N1 3′ splice site and in the downstream intron. This binding blocks the assembly of an pre-spliceosomal E complex between the N1 exon's 5′ splice site and the downstream exon's 3′ splice site.

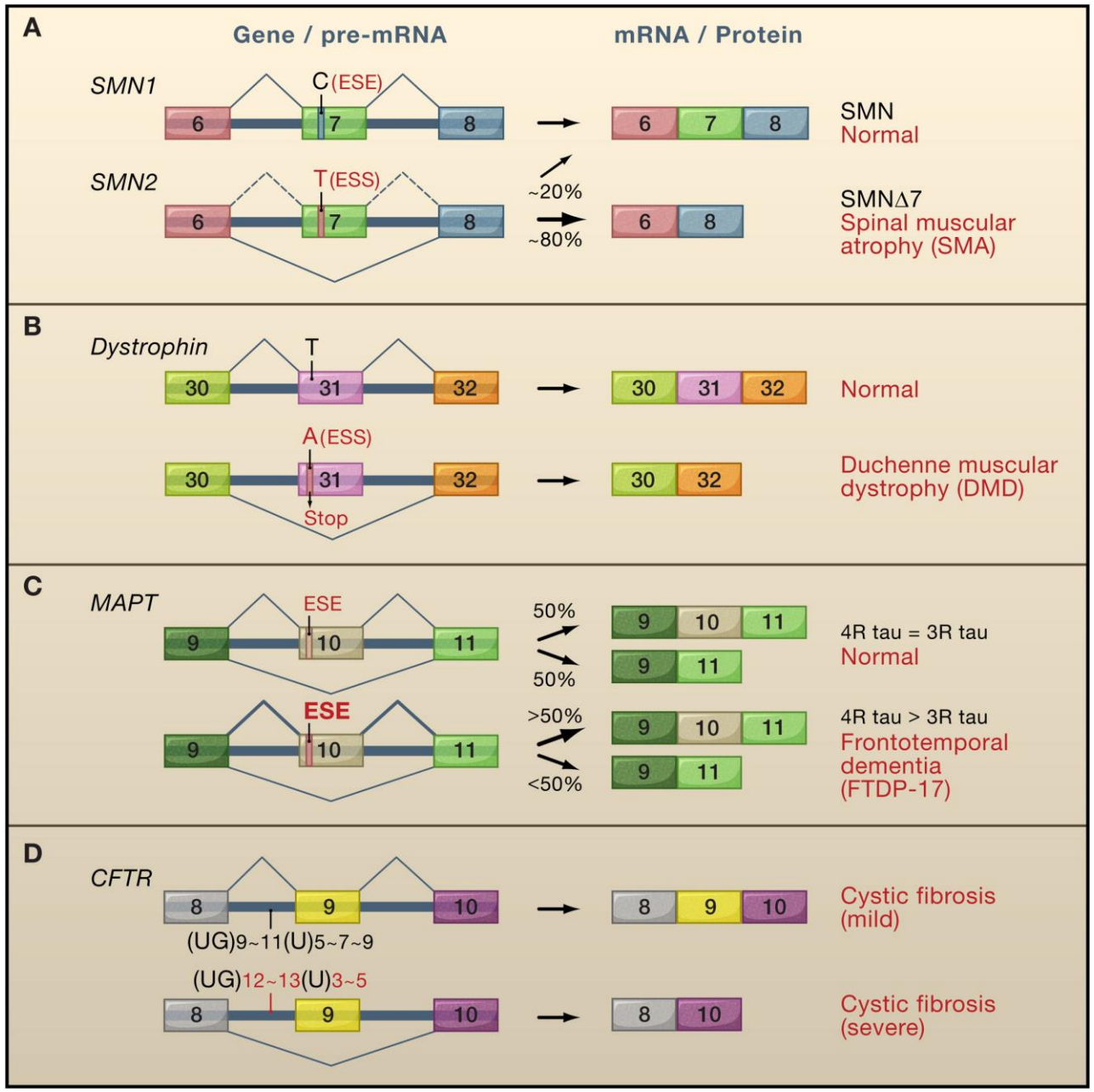
In neurons, PTB is replaced by neural PTB (nPTB), which binds to the PTB repressor elements but does not prevent splicing. ***Neurons also express splicing activators that are members of the Fox family***; these bind to enhancer elements downstream of the exon to stimulate its splicing. nPTB must be displaced from the N1 3′ splice site (shown as a dashed arrow) to allow its splicing, and it may or may not be displaced from its downstream binding site by the adjacent Fox protein. ***Other RNA-binding proteins that affect N1 exon splicing include alternate splicing factor (ASF), which binds to the exon, and heterogeneous nuclear ribonucleoprotein H (hnRNP H), hnRNP F and KH-type splicing regulatory protein (KSRP), which bind to the downstream intron. These proteins might modulate the function of the key regulators, or might allow regulation in additional cell types***.



Different types of alternative splicing. There are several different types of alternative splicing (AS) events, which can be classified into four main subgroups. The first type is **exon skipping**, in which a type of exon known as a cassette exon is spliced out of the transcript together with its flanking introns (see the figure, part a). Exon skipping accounts for nearly 40% of AS events in higher eukaryotes, but is extremely rare in lower eukaryotes. The second and third types are **alternative 3′ splice site (3′ SS) and 5′ SS selection** (parts b and c). These types of AS events occur when two or more splice sites are recognized at one end of an exon. Alternative 3′ SS and 5′ SS selection account for 18.4% and 7.9% of all AS events in higher eukaryotes, respectively. The fourth type is **intron retention** (part d), in which an intron remains in the mature mRNA transcript. This is the rarest AS event in vertebrates and invertebrates, accounting for less than 5% of known events. By contrast, intron retention is the most prevalent type of AS in plants, fungi and protozoa. Less frequent, complex events that give rise to alternative transcript variants include **mutually exclusive exons** (part e), **alternative promoter usage** (part f) and **alternative polyadenylation** (part g). Another rare form of AS involves reactions between two primary transcripts in trans (not shown).
In the figure, constitutive exons are shown in blue and **alternatively spliced regions in purple**. Introns are represented by solid lines, and dashed lines indicate splicing options.



Regulatory elements in pre-mRNA splicing and GVs that can affect them. a | The essential ***splicing signals that define the exon boundaries are relatively short and poorly conserved sequences. Only the GU and the AG dinucleotides that directly flank the exon (at the 3' and 5' ends, respectively) and the branch-point adenosine (all in red) are always conserved***. In most cases, there is also a polypyrimidine tract of variable length (the consensus symbol 'y' represents a pyrimidine base — cytosine or thymine) upstream of the 3'-splice site. The branch point is typically located 18–40 nucleotides upstream from the polypyrimidine tract. Components of the basal splicing machinery bind to the consensus sequences and promote assembly of the splicing complex. This multiprotein complex, known as a spliceosome, performs the correct identification of the splicing signals and catalysis of the cut-and-paste reactions. ***Five small nuclear ribonucleoproteins (snRNPs) and more than 100 proteins make up the spliceosome.*** The U1 snRNP binds to the 5'-splice site, and the U2 snRNP binds the branch site through RNA–RNA interactions. Additional enhancer and silencer elements in the exons (EXON SPLICING ENHANCER (ESE); EXON SPLICING SILENCER (ESS)) and/or introns (INTRON SPLICING ENHANCER (ISE); INTRON SPLICING SILENCER (ISS)) allow the correct splice sites to be distinguished from the many cryptic splice sites that have identical signal sequences. Trans-acting splicing factors can interact with enhancers and silencers and can accordingly be subdivided into two main groups: members of the serine arginine (SR) family of proteins and of the HETEROGENEOUS NUCLEAR RIBONUCLEOPROTEIN PARTICLES (hnRNPs). In general, SR protein binding at ESE facilitates exon recognition whereas hnRNPs are inhibitory. Protein–protein interactions in the spliceosome that modulate the recognition of the splice sites are the probable cause of splicing inhibition or activation. b | ***Genomic variants (GVs) can affect different splicing regulatory elements, leading to aberrant splicing. Exonic GVs (eGVs)*** can either change the amino acid, result in synonymous GVs in exons (sGVs) or introduce a nonsense codon. ***Intronic GV***s might be located within approximately 50 bp from the splice sites (that is, 3'-splice site GVs (ssGVs) and 5' ssGVs) or deep in the introns (intronic GVs (iGVs)). Term genomic variants (GVs) denotes single-nucleotide substitutions or small insertions and deletions.



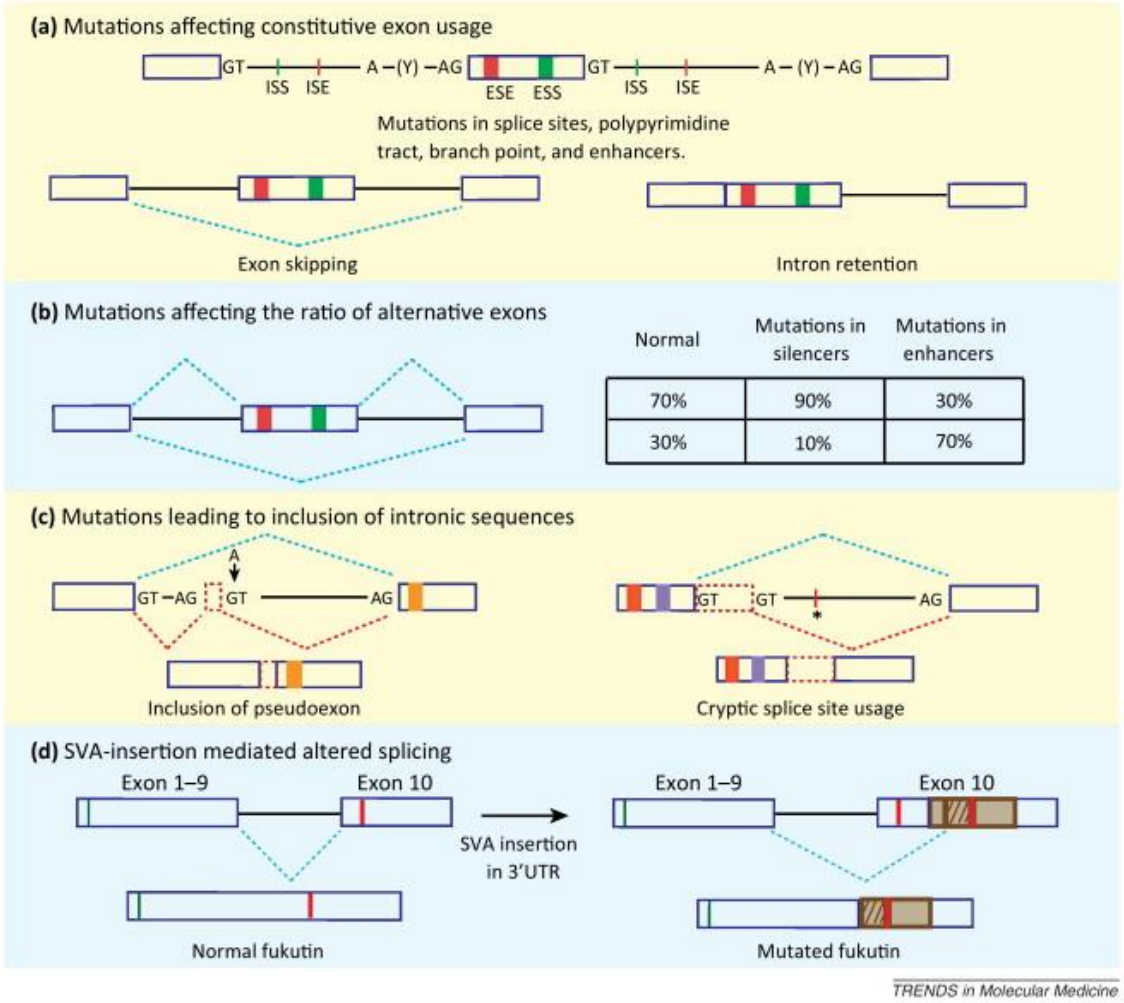
Mutations Disrupting the Cis-Acting Splicing Code Cause Splicing Defects and Disease

(A) The motor neuron degenerative disease SMA results from deletion of the SMN1 gene. The remaining SMN2 gene has a C→T substitution in exon 7 that **inactivates an exonic splicing enhancer (ESE), and creates an exonic splicing silencer (ESS), leading to exon 7 skipping and a truncated protein (SMNΔ7)**.

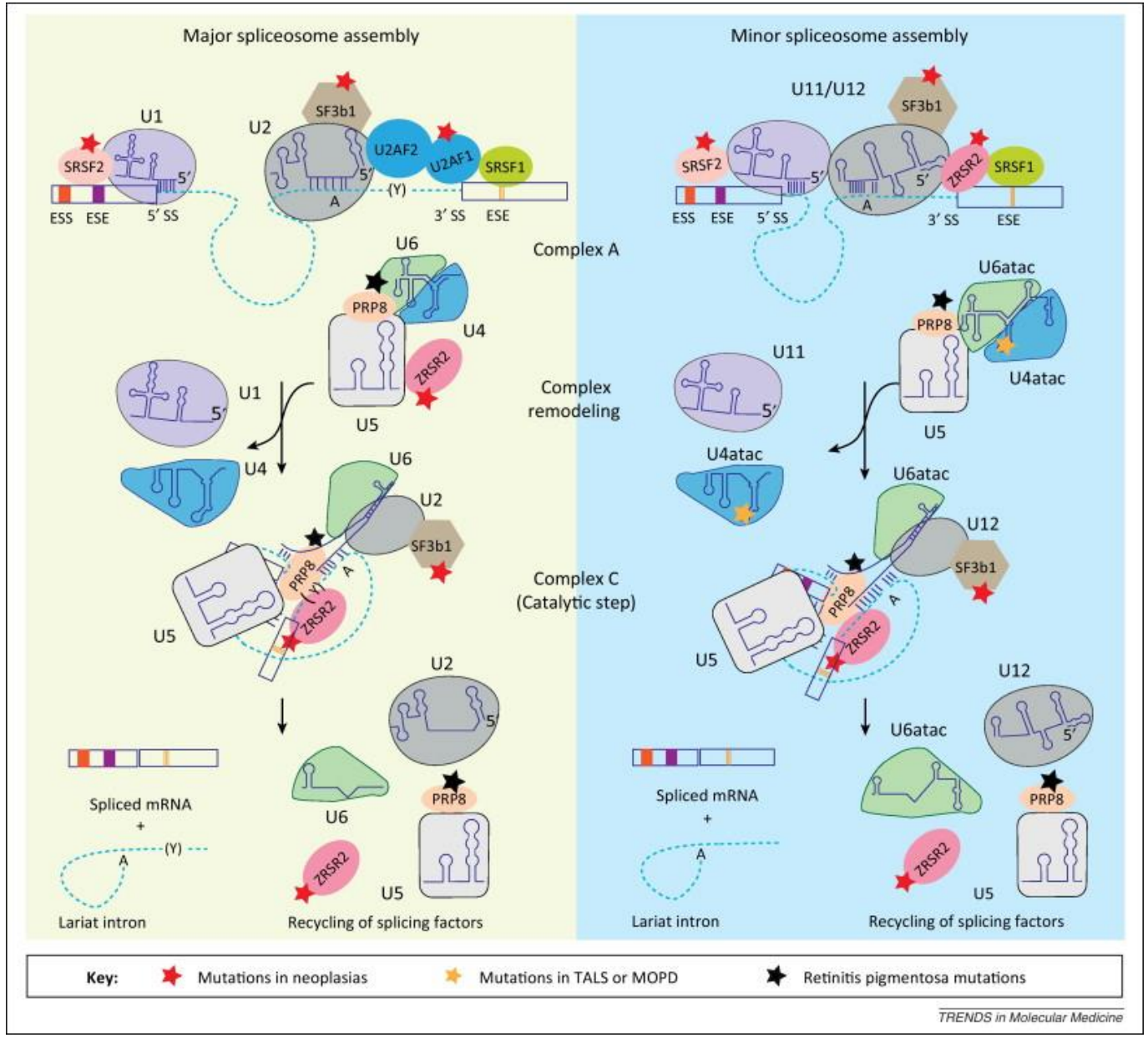
(B) A T→A substitution in exon 31 of the dystrophin gene simultaneously **creates a premature termination codon (STOP) and an ESS, leading to exon 31 skipping**. This mutation causes a mild form of DMD because the mRNA lacking exon 31 produces a partially functional protein.

(C) Mutations within and downstream of exon 10 of the MAPT gene encoding the tau protein **affect splicing regulatory element**s and disrupt the normal 1:1 ratio of mRNAs including or excluding exon 10. This results in a perturbed balance between tau proteins containing either four or three microtubule-binding domains (4R-tau and 3R-tau, respectively), causing the neuropathological disorder FTDP-17. The example shown is the N279K mutation, which enhances an ESE function, promoting exon 10 inclusion and shifting the balance toward increased 4R-tau.

(D) **Polymorphic (UG)m(U)n tracts within the 3′ splice site** of the CFTR gene exon 9 influence the extent of exon 9 inclusion and the level of full-length functional protein, modifying the severity of cystic fibrosis (CF) caused by a mutation elsewhere in the CFTR gene.

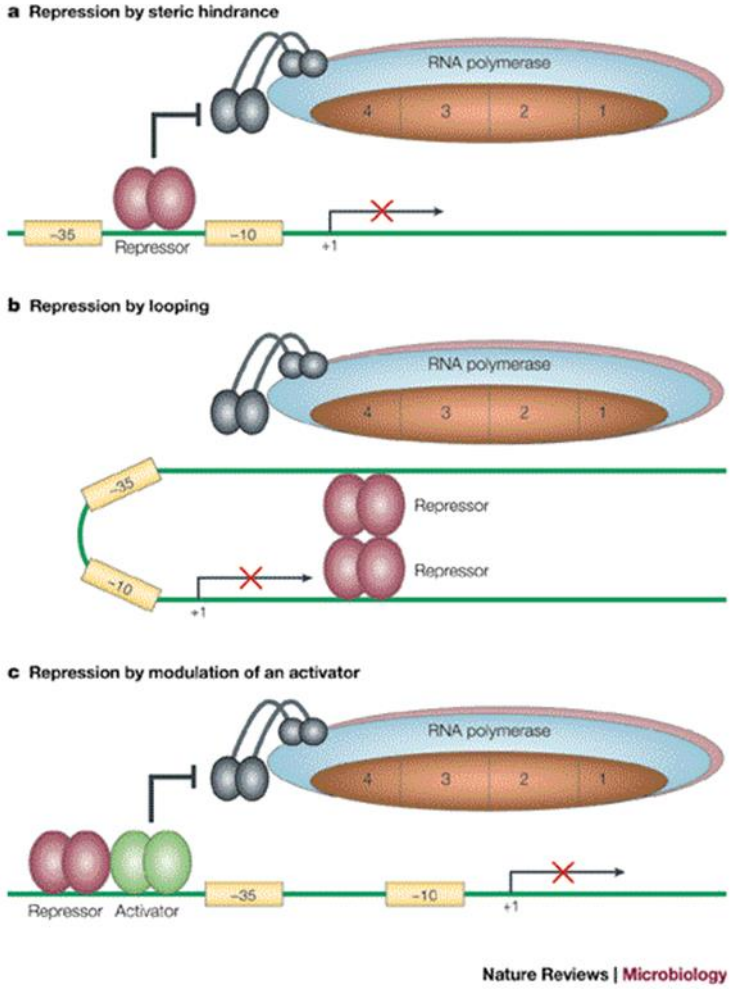


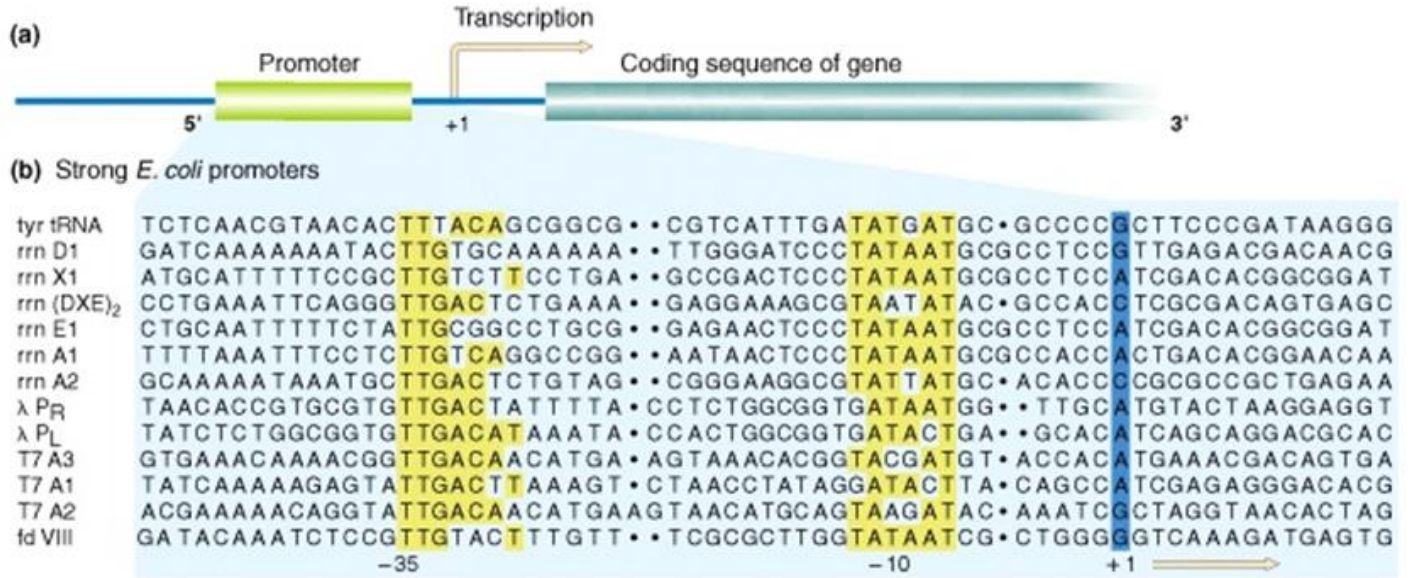
Four mechanistic categories of altered gene function by splicing mutations. (a) The basic cis elements of the splicing code are indicated: 5′ and 3′ splice sites (represented by GT and AG), polypyrimidine tract (Y), branch point (A), and exonic and intronic enhancers (ESEs and ISEs) and silencers (ESSs and ISSs). Mutations affecting splice sites, the polypyrimidine tract, the branch point, or splicing enhancers lead to exon skipping or intron retention. (b) ***Mutations in enhancer or silencer elements*** can change the ratio of isoforms containing alternative exons. (c) ***Mutations within introns*** can lead to inclusion of intronic sequences (indicated by red broken rectangles) by creating a splice site/pseudoexon (indicated by arrow) and/or by creating an enhancer element (indicated by asterisk), allowing recognition of a cryptic splice site. The blue broken lines indicate the normal splicing pattern, whereas the red broken lines indicate the splicing pattern caused by the mutation. (d) ***Insertion of transposable elements*** (SVAs, represented by a brown rectangle) in the 3′ untranslated region (UTR) of the fukutin gene leads to the alternate use of splice sites producing a protein with a different carboxy-terminal sequence (patterned brown rectangle). The green lines indicate the start codon of both the normal and mutated fukutin, whereas red lines indicate stop codons.

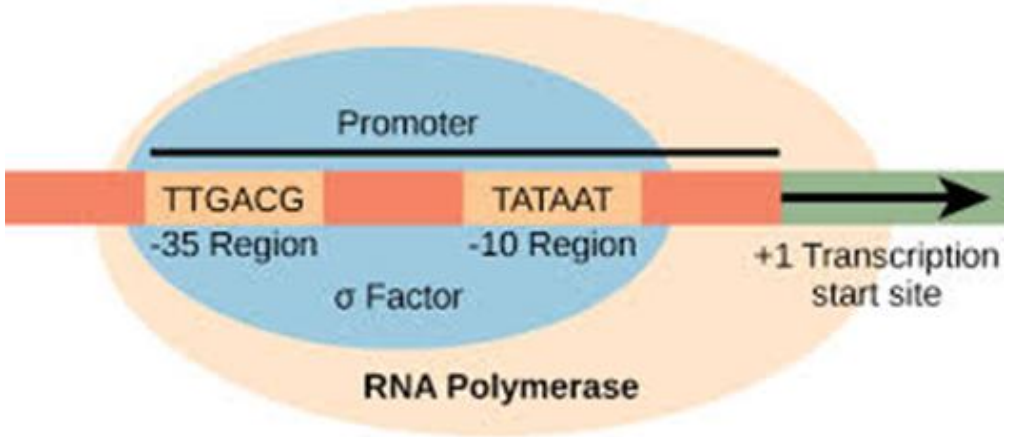


Spliceosome assembly and disease-associated mutations in spliceosome components. The broken lines and rectangles represent introns and exons, respectively. The left panel shows the assembly of the ***major (U2 type) spliceosome***. U1 and U2 small nuclear ribonucleoproteins (snRNPs) are recruited to the consensus 5′ splice site (5′ SS) and branch point (A), respectively. The U2-auxiliary factor heterodimer (U2AF2/U2AF1) interacts with the polypyrimidine track (Y) and 3′ splice site (3′ SS), forming complex A. The U4/6 and U5 snRNPs join the assembling spliceosome followed by remodeling of the complex leading to removal of the U1 and U4 snRNP and formation of the catalytic complex (complex C). Two transesterification reactions join the exons and release an intron lariat that is subsequently degraded and the spliceosome components are recycled for subsequent rounds of splicing. The right panel shows the assembly of ***the minor (U12 type) spliceosome***, in which U1, U2, U4, and U6 are replaced by homologous U11, U12, U4atac, and U6atac snRNPs, respectively. The red star indicates the components that are mutated in neoplasias. The black star indicates the components that are mutated in retinitis pigmentosa. The ***orange star indicates the mutation in U4atac that is associated with microcephalic osteodysplastic primordial dwarfism type 1 (MOPD1)***. Abbreviations: ESEs, exonic splicing enhancers; ESSs, exonic splicing silencers.

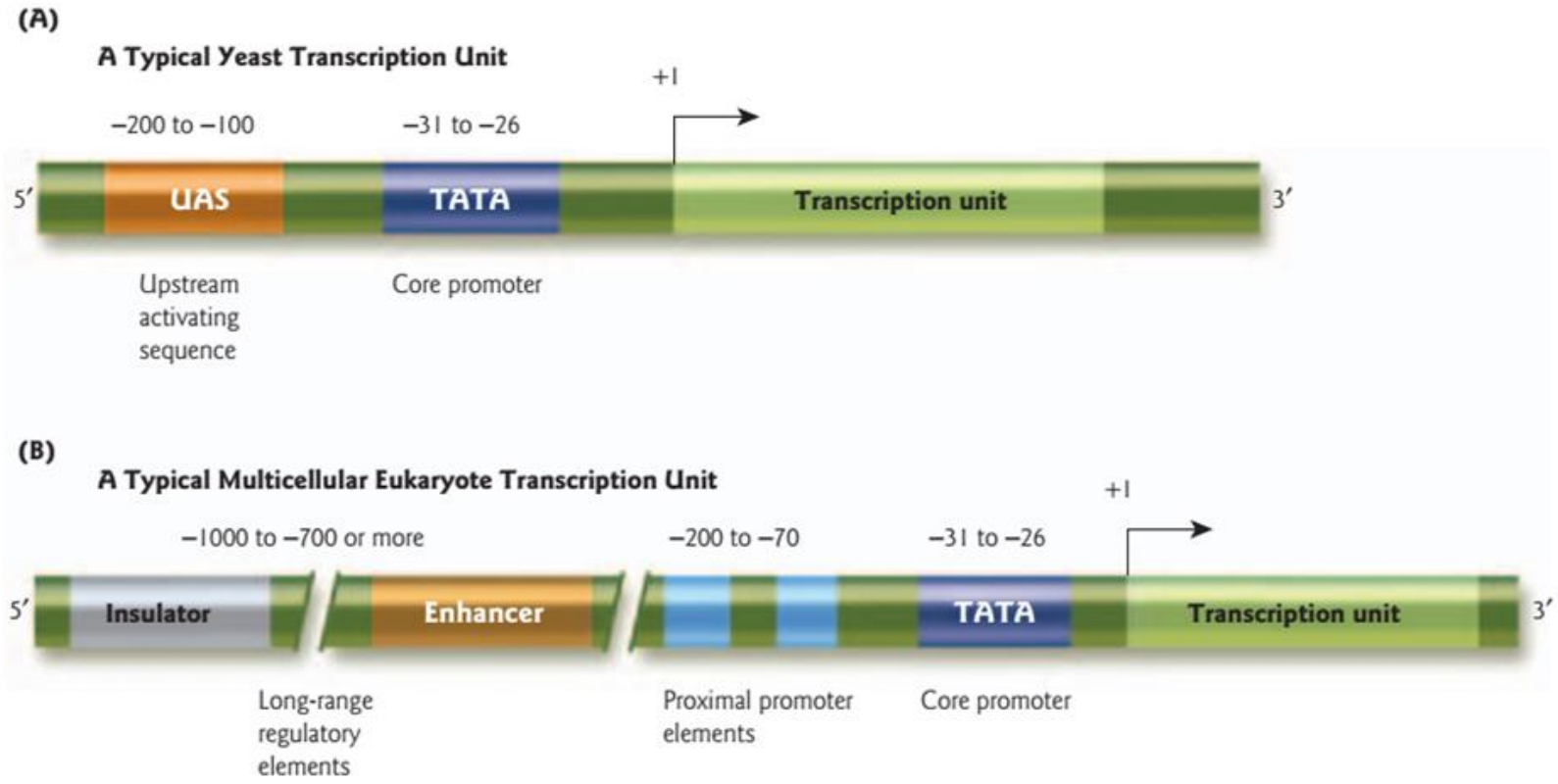
Regulatorne regije v genih: promotorji in neprevedene regije (5‘UTR in 3‘UTR)



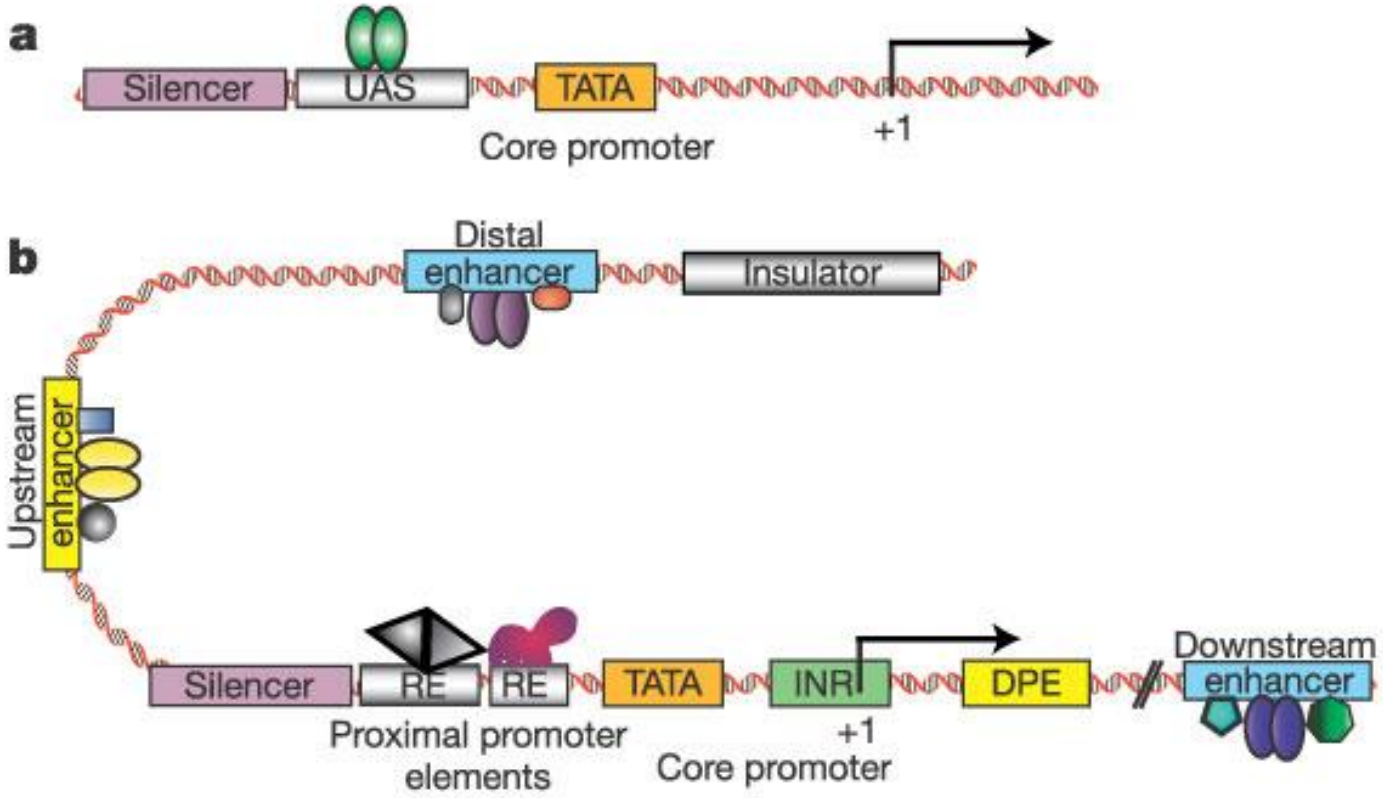




The regulation of bacterial transcription initiation



Comparison of a simple and complex RNA pol II transcription unit. (**A**) A typical yeast (***unicellular eukaryote***) transcription unit. The start of transcription (+1) of the protein-coding gene (transcription unit) is indicated by an arrow. (**B**) A typical ***multicellular eukaryote*** transcription unit with clusters of proximal promoter elements and long-range regulatory elements located upstream from the core promoter (TATA). There is ***variation in whether a particular element is present or absent, the number of distinct elements, their orientation relative to the transcriptional start site, and the distance between them***. Although the figure is drawn as a straight line, the binding of transcription factors to each other draws the regulatory DNA sequences into a loop.



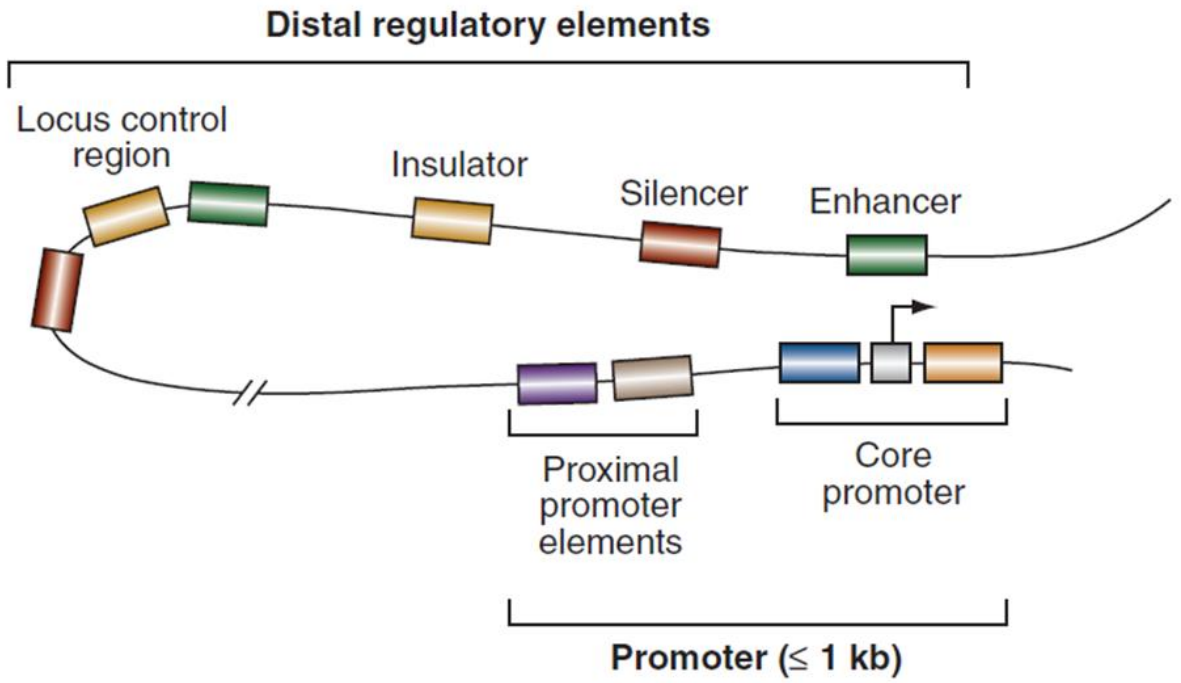
Comparison of a simple eukaryotic promoter and extensively diversified metazoan regulatory modules. a, Simple eukaryotic transcriptional unit. A simple core promoter (TATA), upstream activator sequence (UAS) and silencer element spaced within 100–200 bp of the TATA box that is typically found in unicellular eukaryotes. b, Complex metazoan transcriptional control modules. A complex arrangement of multiple clustered enhancer modules interspersed with silencer and insulator elements which can be located 10–50 kb either upstream or downstream of a composite core promoter containing TATA box (TATA), Initiator sequences (INR), and downstream promoter elements (DPE).

Table 1 | Selected major categories of non-coding functional element

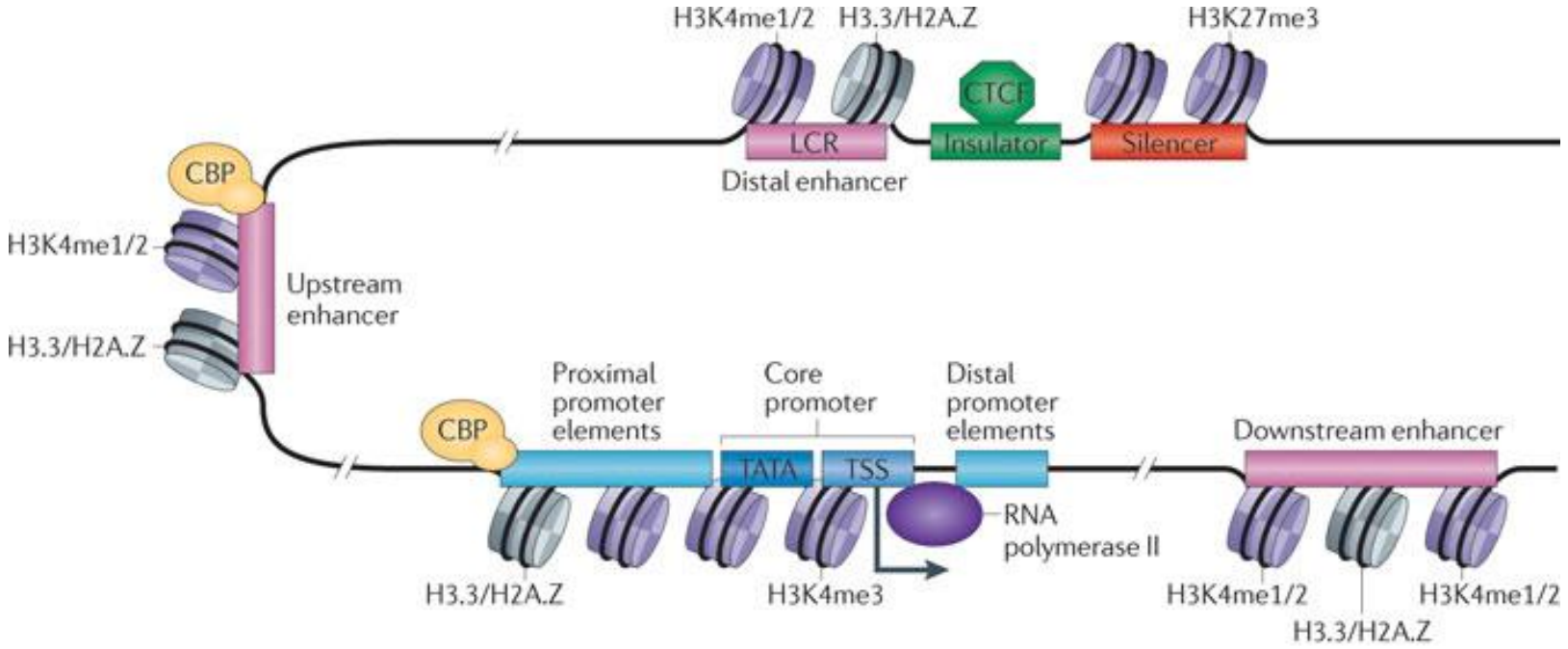
Category	Function	Selected associated chromatin marks*
Promoter	Region that is located immediately upstream of a protein-coding gene, and binds to RNA polymerase II; where transcription is initiated	RNA polymerase II[44], H3K4me3 (ref. 40) (active promoters)
Enhancer	Region that activates transcription, often in a temporally and spatially restricted manner, by acting on a promoter. Enhancers can be located far from target promoters and are orientation independent	p300 (refs 40, 56), H3K4me1 (ref. 40)
Insulator	Separates active from inactive chromatin domains and interferes with enhancer activity when placed between an enhancer and promoter	CTCF[44,53]
Repressor/ silencer	Negative regulators of gene expression	REST[45], SUZ12 (refs 69, 70)

*Many additional chromatin marks were found to correlate with one or several of these categories of regulatory element. Detailed descriptions of these markers and their respective binding characteristics at different types of regulatory sequence element can be found in refs 40, 41, 44, 51 and 55.

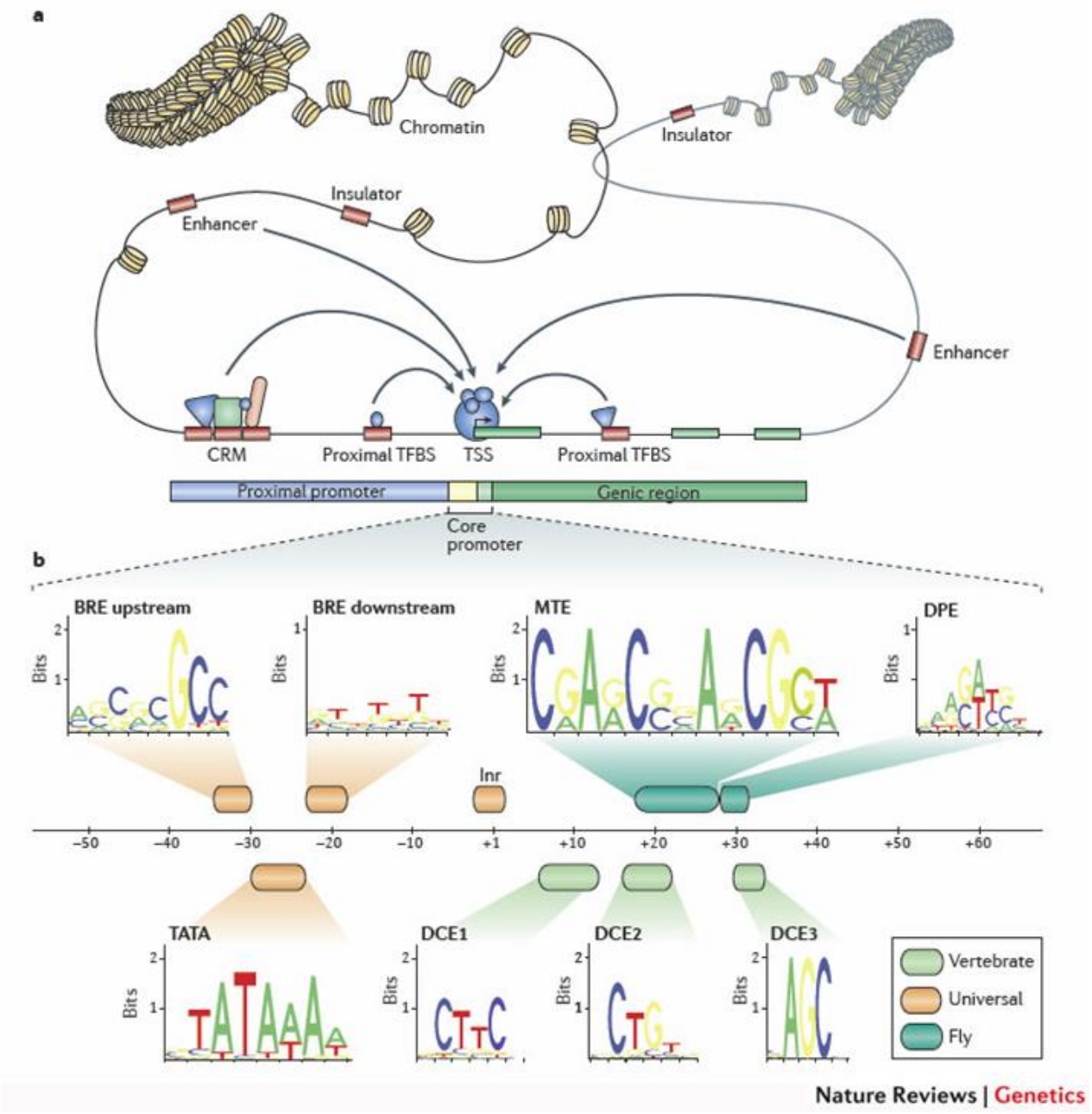
Enhancer = ojačevalec; Insulator = izolator



Schematic of a typical gene regulatory region. The promoter, which is composed of a core promoter and proximal promoter elements, typically spans less than 1 kb pairs. Distal (upstream) regulatory elements, which can include enhancers, silencers, insulators, and locus control regions, can be located up to 1 Mb pairs from the promoter. These ***distal elements may contact the core promoter or proximal promoter through a mechanism that involves looping out the intervening DNA***.

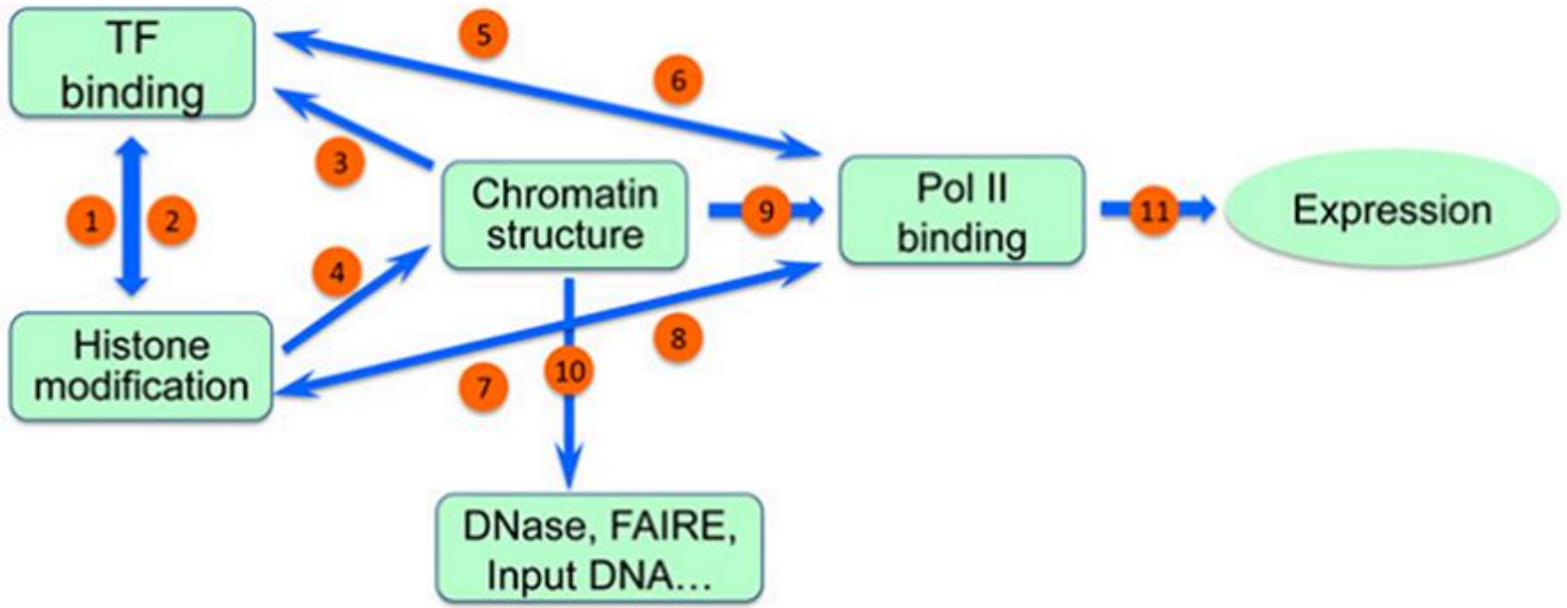


Transcriptional regulatory elements in metazoans. The promoter is typically comprised of proximal, core and downstream elements. ***Transcription of a gene can be regulated by multiple enhancers that are located distantly and interspersed with silencer and insulator elements, which are bound by regulatory proteins such as CCCTC-binding factor (CTCF).*** Recent genome-wide data have revealed that ***many enhancers can be defined by unique chromatin features and the binding of cyclic AMP-responsive element-binding (CREB) protein (CBP).***
H3K4me1/2, histone H3 mono- or dimethylation at lysine 4; H3K4me3, histone H3 trimethylation at lysine 4; H3K27me3, histone H3 trimethylation at lysine 27; H3.3/H2A.Z, histone variants H3.3 and H2A.Z; LCR, locus control region; TATA, 5′-TATAAAA-3′ core DNA sequences; TSS, transcription start site.

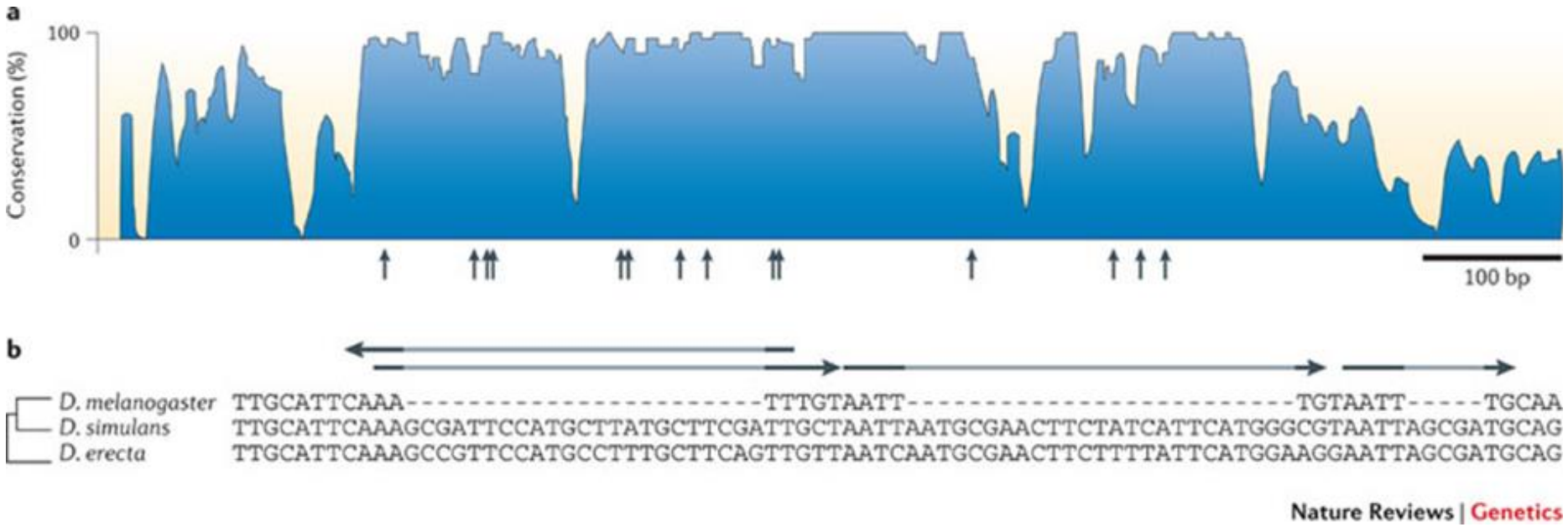


Regulation of transcription. a | ***A summary of promoter elements and regulatory signals.*** Chromatin is comprised of DNA wrapped around histones to form nucleosomes. The structure of chromatin can be tightly wrapped or accessible to proteins. Boundaries between these states may be marked by insulators. The region around the transcription start site (TSS) is often divided into a larger proximal promoter upstream of the TSS and a smaller core promoter just around the TSS. The exact boundaries vary between studies. ***To recruit RNA polymerase II (RNAPII) and to activate transcription of the gene, sequence-specific regulatory proteins (transcription factors) bind to specific sequence patterns (namely, transcription factor binding sites (TFBSs)) that are near to the TSS (proximal elements) or that are far away from it (enhancers). TFBSs can also occur in clusters, forming cis-regulatory modules (CRMs).*** b | ***Sequence patterns in core promoters.*** The region around the TSS has several over-represented sequence patterns; the TATA box and initiator (Inr) are the most studied and most prevalent. The location of patterns relative to the TSS and their sequence properties are shown as boxes and as associated sequence logos based on the JASPAR database. The Inr pattern is not shown as it varies considerably between studies, ranging from a TCA(G/T)TC(C/T) to a single dinucleotide (pyrimidine (C/T)–purine (A/G)). Importantly, ***most promoters only have one or a few of these patterns, and some patterns are mostly found in certain species.*** BRE, B recognition element; DCE, downstream core element; DRE, DNA recognition element; MTE, motif ten element.

Regulatory mechanism of TF (transcription factor) binding, histone modification, and other chromatin features on gene expression



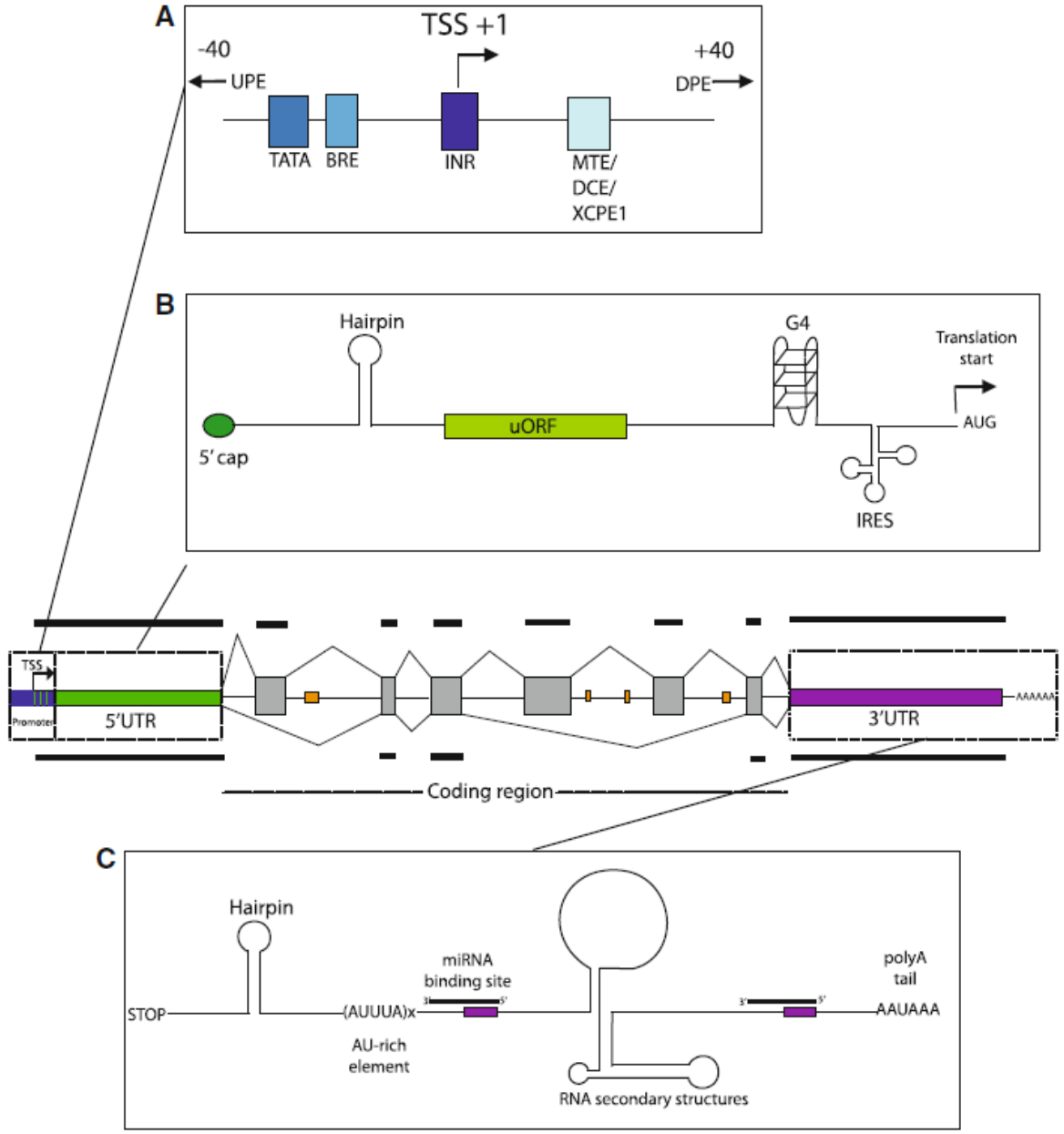
(1) Recruiting histone modifiers
(2) Recruiting TFs
(3) Accessibility
(4) Remodeling
(5) Recruiting general TFs
(6) Interacting with TFs
(7) Recruit general TFs
(8) Interacting with histone modifiers
(9) Accessibility
(10) Accessibility
(11) Transcription



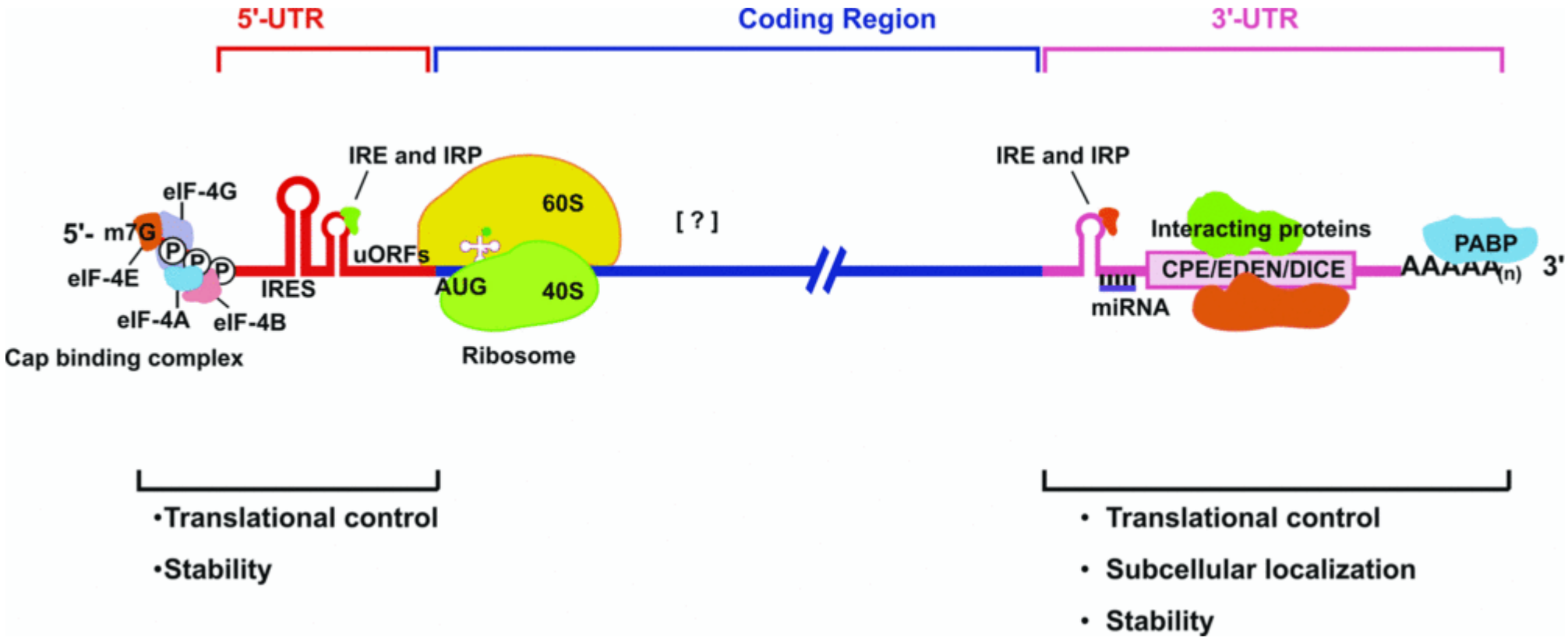
Examples of point mutations and deletions causing divergent cis-regulatory activity. a | A **1,007 bp enhancer** located 5′ of the shavenbaby promoter contains 14 nucleotide differences that are specific to Drosophila sechellia. A measure of sequence conservation for this enhancer is plotted, and **vertical arrows indicate the location of the 14 D. sechellia-specific alleles**. b | **Multiple deletions in an enhancer** for the desatF gene (also known as Fad2) (only part of which is shown) have occurred in Drosophila melanogaster and create copies of the hexamer sequence AATTTG — three on one strand and one on the other (indicated with horizontal arrows). The D. melanogaster gene sequence is compared with those of Drosophila simulans and Drosophila erecta. Although no specific TF has yet been shown to bind to these sites, mutating them shows that they are essential for enhancer activity.

Mechanisms through which noncoding variants influence human disease

Promoter	Promoter regions are an essential component of transcription initiation and the assembly of RNA polymerase and associated regulators. Mutations can affect binding of activators or repressors, chromatin state, nucleosome positioning, and also looping contacts of promoters with distal regulatory elements. Genes with coding disease mutations can also harbor independently associated regulatory variants that correlate with expression, are bound by proteins in an allele-specific manner, and disrupt or create regulatory motifs[168].	Mutations in the promoter of the HIV-1 progression–associated gene *CCR5* are correlated with expression of the receptor it encodes and bind differentially to at least three transcription factors[169,170].
		APOE promoter mutations are associated with Alzheimer's disease[171,172].
		Heme oxygenase-1 (*HO1*) promoter mutations lead to expression changes and are associated with many diseases[173].
Enhancer	Enhancers are distal regulatory elements that often lie 10,000–100,000 nucleotides from the start of their target gene. Mutations within them can disrupt sequence motifs for sequence-specific transcription factors, chromatin regulators and nucleosome positioning signals. Structural variants including inversions and translocations can disrupt their regulatory activity by moving them away from their targets, disrupting local chromatin conformation, or creating interactions with insulators or repressors that can hinder their action. Although it is thought that looping interactions with promoter regions play a role, the rules of enhancer-gene targeting are still poorly understood.	The role of distal enhancers in disease was suggested even before the development of GWAS by the many Mendelian disorders for which some patients had translocations or other structural variants far from the promoter[174–176].
		In one early study, point mutations were mapped in an unlinked locus in the intron of a neighboring gene, a million nucleotides away from the developmental gene *Shh*[177]; this distal locus acted as an enhancer of *Shh* and recapitulated the polydactyly phenotype in mouse.
		A number of GWAS hits have been validated as functional enhancers[178]: for example, common variants associated with cancer susceptibility map to a gene desert on chromosome 8, with one SNP demonstrated to disrupt a TCF7L2 binding site and to inhibit long-range activation of the oncogene *MYC*[179–181].



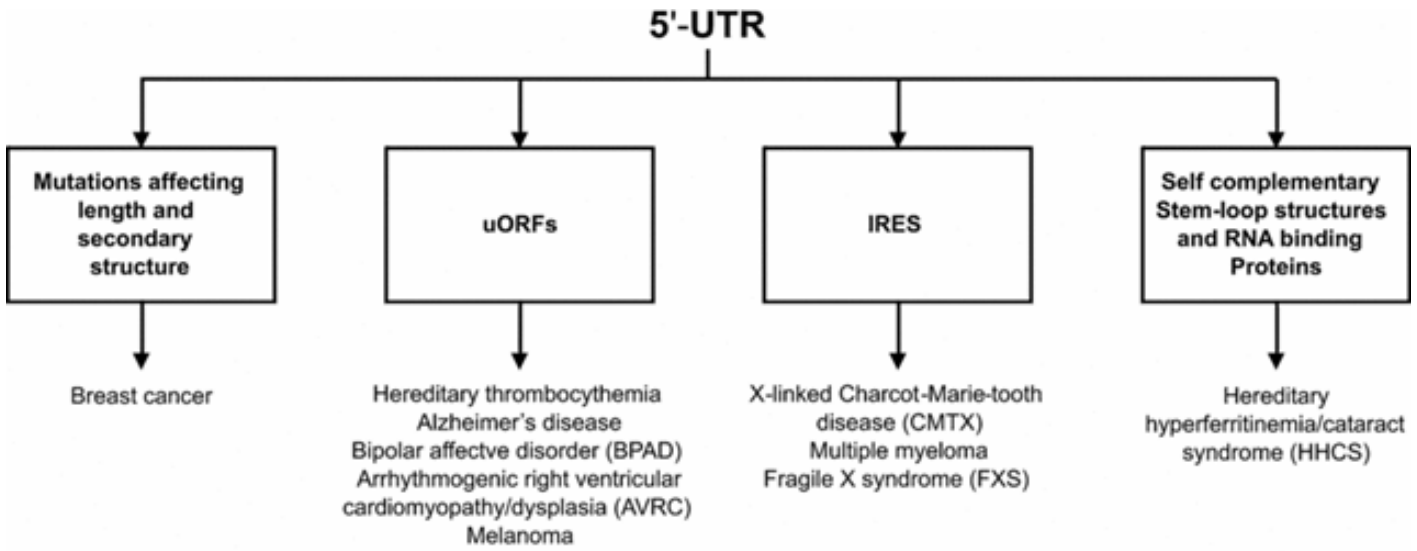
Regulatory elements within the noncoding gene regions. The centre image shows a typical gene, with exons indicated in grey. The orange rectangles indicate intronic enhancer elements. a ***Promoter region regulatory elements***. Upstream and downstream promoter elements situated outside of the core promoter region are indicated by the arrows. b ***Regulatory elements in the 5'UTR***. c ***Regulatory elements in the 3'UTR***.

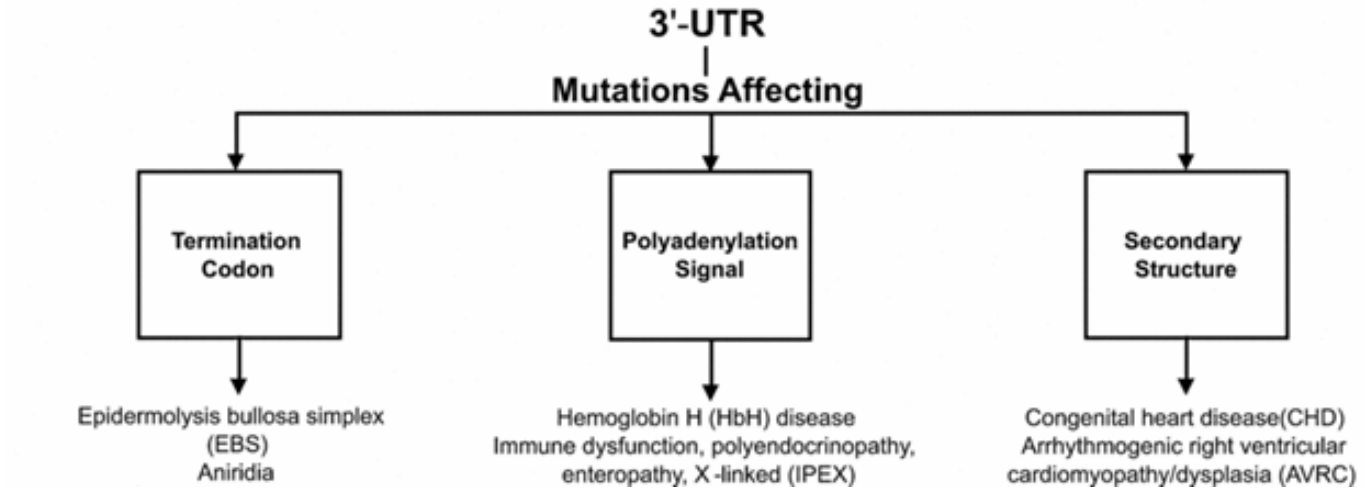


Structural organization of eukaryotic mRNA and the different points of possible regulation of translation through various trans-acting factors.

5′-m7G, cap structure; eIF, eukaryotic initiation factor; CPE, cytoplasm polyadenylation element; EDEN, embryonic deadenylation signal; DICE, differential control element; PABP, poly(A)-binding protein. [?], possible sites of interaction of transacting factors (yet unknown) in the coding sequence. Regions of mRNA involved in subcellular localization and stability are also indicated.

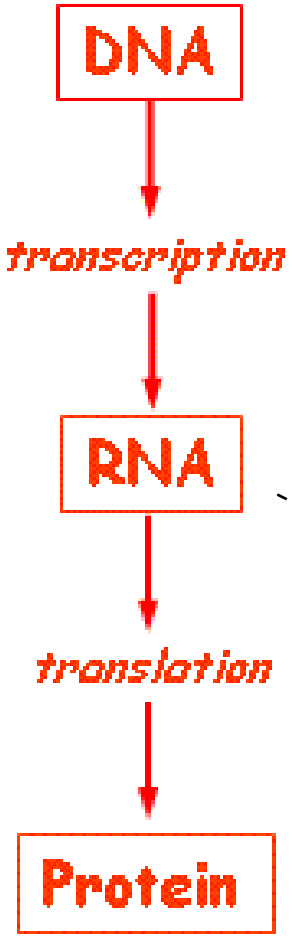
Involvement of various regulatory elements of 5′- and 3′-UTRs of mRNAs in various diseases

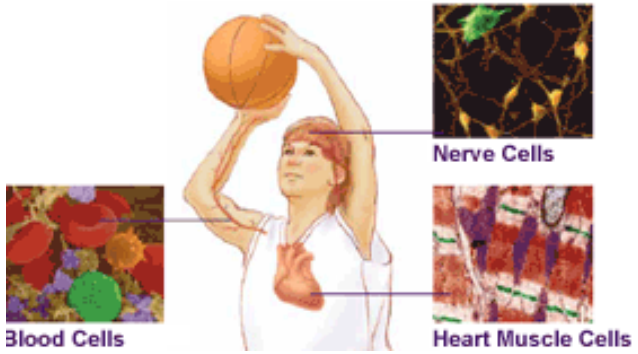


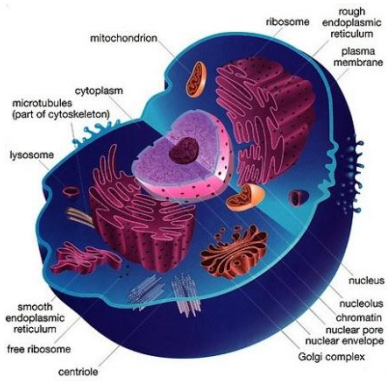


Gene regulation

Gene Regulation







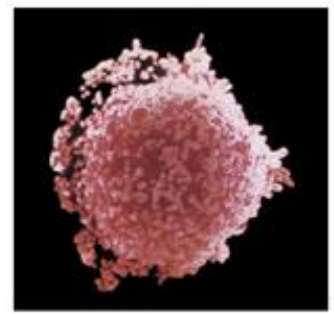
Different cells in our body hold copies of (essentially) the same genome.

Yet they express *very* different repertoires of proteins and non-coding RNAs.

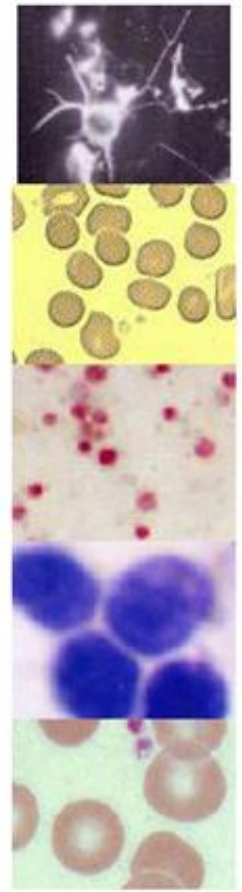
How do cells do it?

A: like they do everything else: using their proteins & ncRNAs…

How can only ~20,000 genes specify a complex mammal?

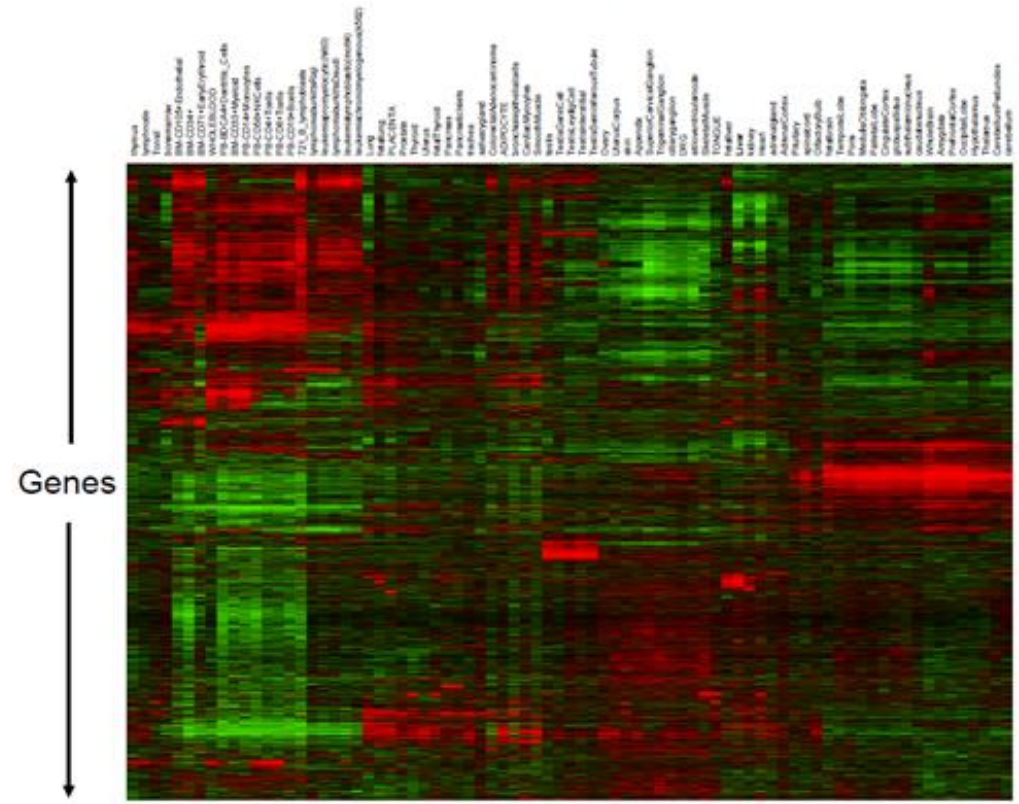


1 Cell



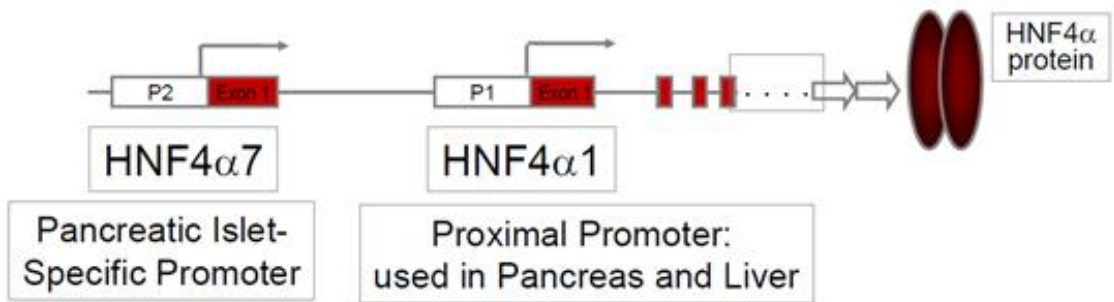
>200 Cell Types

How can only ~20,000 genes specify a complex mammal?
Cell-type specific gene expression



How can only ~20,000 genes specify a complex mammal?

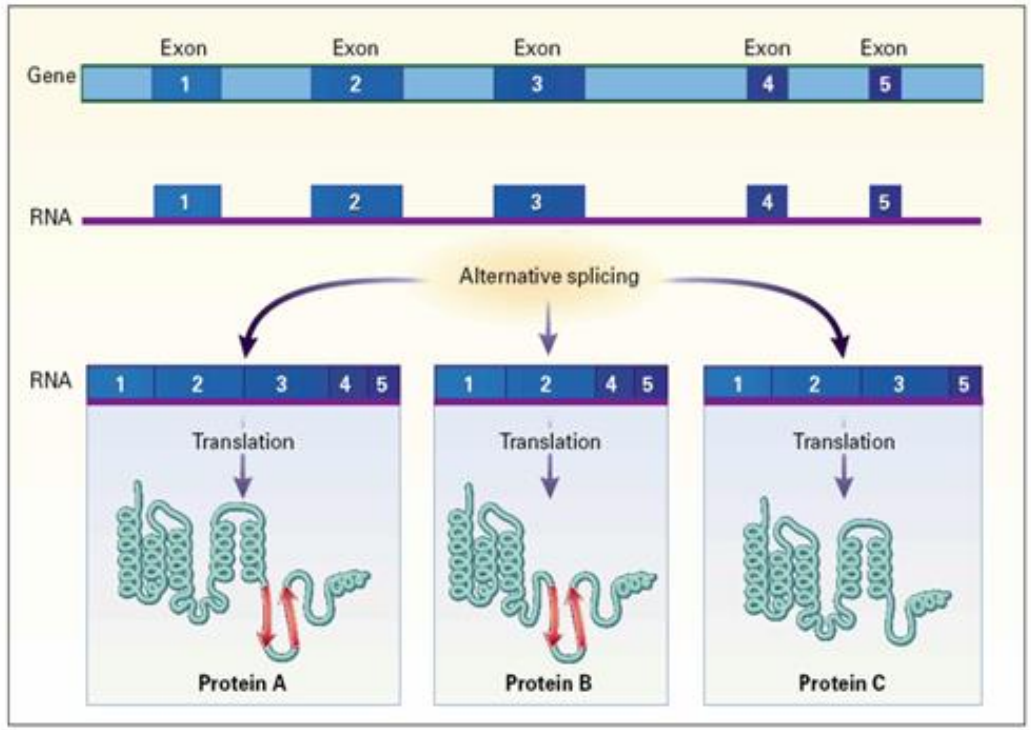
Tissue-specific Gene Transcription: Promoter Utilization in Pancreas and Liver



Thomas et al. *Human Mol Gen* **10** (2001) 2089

How can only ~20,000 genes specify a complex mammal?

Alternative Splicing



How can only ~20,000 genes specify a complex mammal?

Tissue-specific Protein Modification

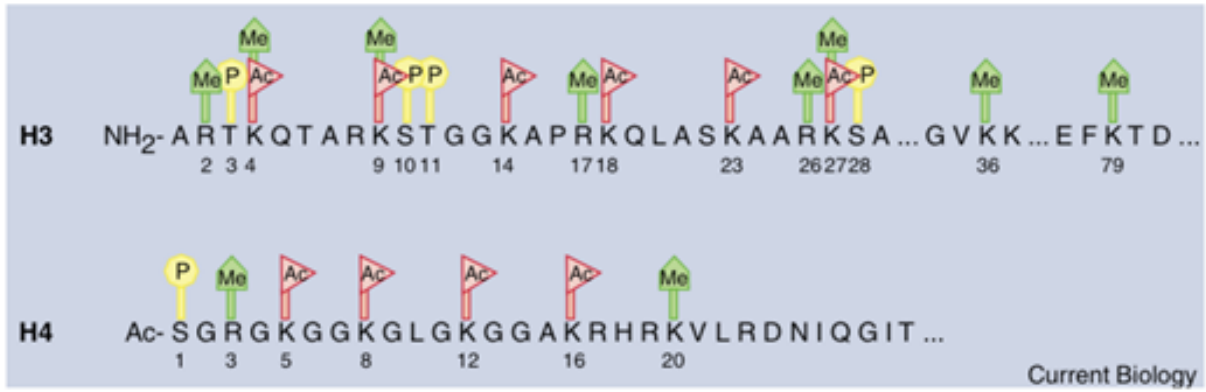
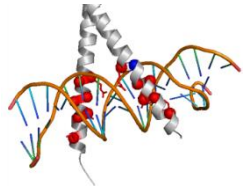


Figure adopted from Peterson and Laniel 2004

Cis-Regulatory Components

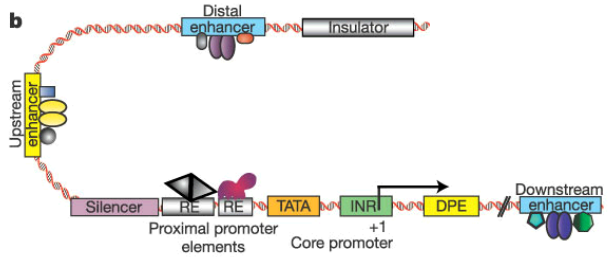
Low level (“atoms”):

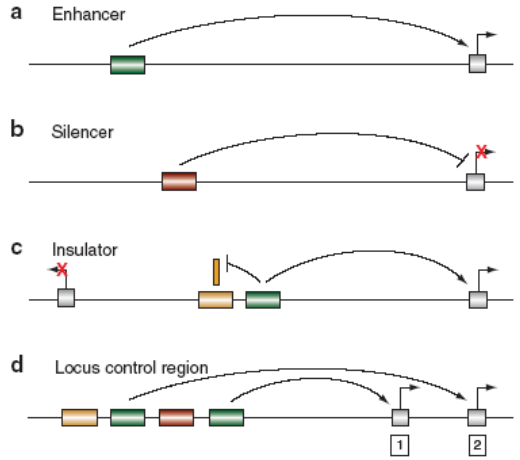
- Promoter motifs (TATA box, etc)
- Transcription factor binding sites (TFBS)



Mid Level:

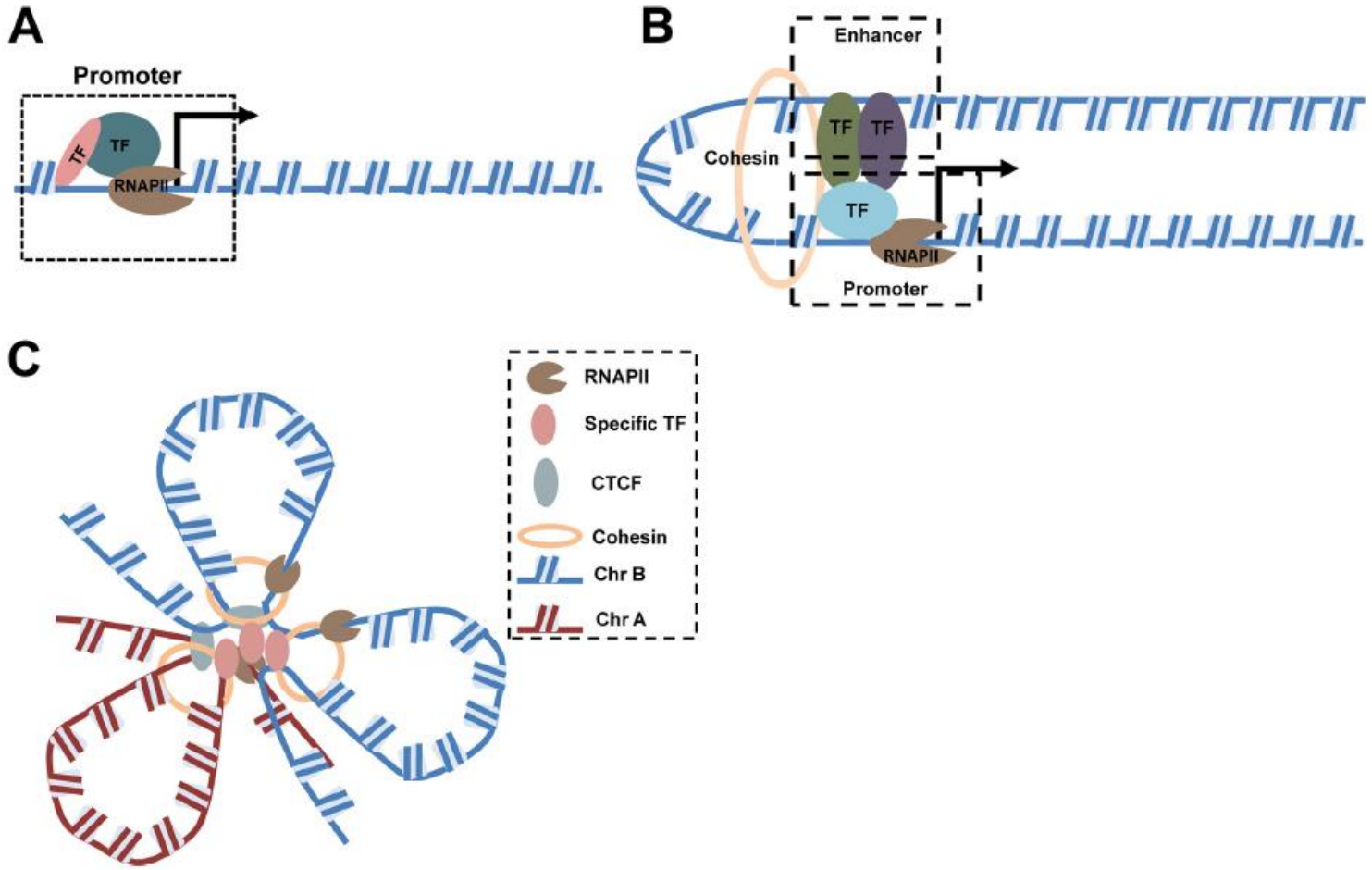
- Promoter
- Enhancers
- Repressors/silencers
- Insulators/boundary elements
- Cis-regulatory modules (CRM)
- Locus control regions (LCR)





High Level:

- Epigenomic domains / signatures
- Gene expression domains
- Gene regulatory networks (GRN)



Transcriptional architecture: from a one-dimensional linear model to a three-dimensional transcription factory model.

A, the simplest model of transcription organization considers ***chromatin structure as one-dimensional and posits that the transcriptional machinery acts mainly on promoters or enhancers immediately upstream of a transcription start site***. TF, transcription factor.

B, the **looping model** usually describes ***interaction between a promoter and a distant upstream enhancer. The cohesin complex is considered essential to maintain the looping structure***.

C, a **transcription factory** consists of ***immobilized and enriched RNAPII foci***. ***DNA loci in cis and in trans are recruited to the factory by specific transcription factors as well as by CTCF and the cohesin complex***. Different chromosomes are represented as Chr A and Chr B.