Cellular and Molecular Biology I

BIO-205-8

Camille Goemans

Recap: DNA methylation

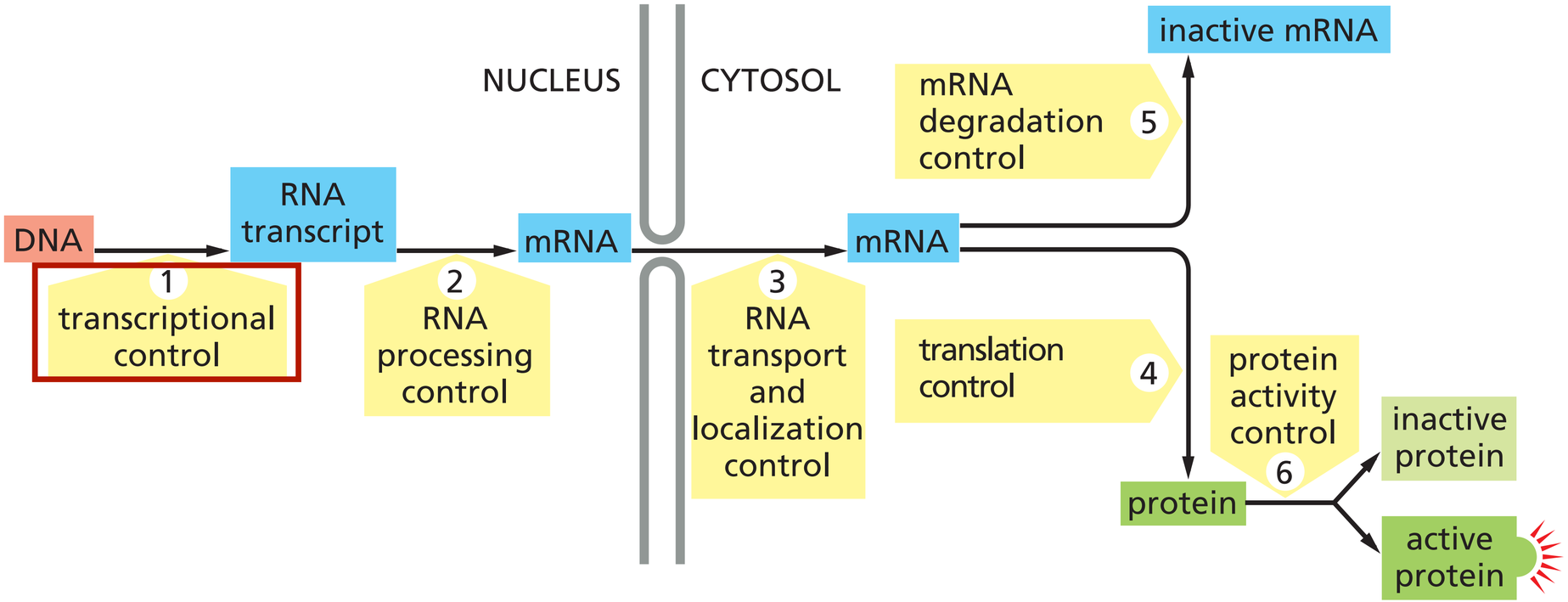
- DNA methylation = addition of a methyl group ($–CH_3$) to DNA, typically to the base cytosine (C) or adenine (A).
- DNA methylation occurs in both eukaryotes and prokaryotes, but it serves different functions and involves different enzymes, targets, and biological roles.

Feature	Eukaryotes	Prokaryotes
Main methylated	Cytosine; sometimes adenine	Adenine and cytosine
Common sequence	CpG dinucleotides (5'-CG-3')	Specific short recognition sequences (e.g., GATC, CCWGG)
Enzymes	DNA methyltransferases	Restriction–modification system methyltransferases
Main biological role	Epigenetic gene regulation, genomic imprinting, X-chromosome inactivation,…	Protects host DNA from restriction enzymes; distinguishes self from foreign DNA (e.g., phage DNA)
Inheritance	Heritable through cell division; maintained during replication	Methylation pattern maintained by methyltransferases but not used for long-term gene regulation
Effect on gene expression	Typically represses transcription when in promoter regions	Can regulate DNA replication and occasionally affect transcription or mismatch repair
Reversibility	Dynamic — can be actively or passively demethylated	Generally stable; loss can affect protection against restriction enzymes

Erratum: Trans-sequences

- **Cis-Regulatory Sequences (or Elements)** = segments of noncoding DNA located on the same molecule (“cis”) as the gene they regulate
    - They serve as binding sites for transcription factors and other proteins that control transcription
    - They control when, where, and how much a gene is transcribed
    - They include:
        - ✓ **Promoters** — right upstream of the transcription start site; where RNA polymerase and general TFs bind
        - ✓ **Enhancers** — increase transcription; can be far away from the gene (upstream, downstream, or even within introns
        - ✓ **Silencers** — decrease transcription
        - ✓ **Insulators** — block enhancer–promoter communication or separate chromatin domain
- **Trans-Regulatory Sequences (or Elements)** = segments of DNA located on another molecule (not on the same DNA molecule as the gene being regulated)
- Trans-regulatory elements are usually genes that encode diffusible molecules — such as proteins or RNAs — that bind to *cis*-regulatory sequences on DNA to regulate transcription.

Levels of gene expression regulation



IV. Control of gene expression

1. Transcriptional control

a. Transcription regulators (activators, repressors)

b. Understanding different control systems

c. Combinatorial gene control and cell types

d. Genomic imprinting

e. X-chromosome inactivation

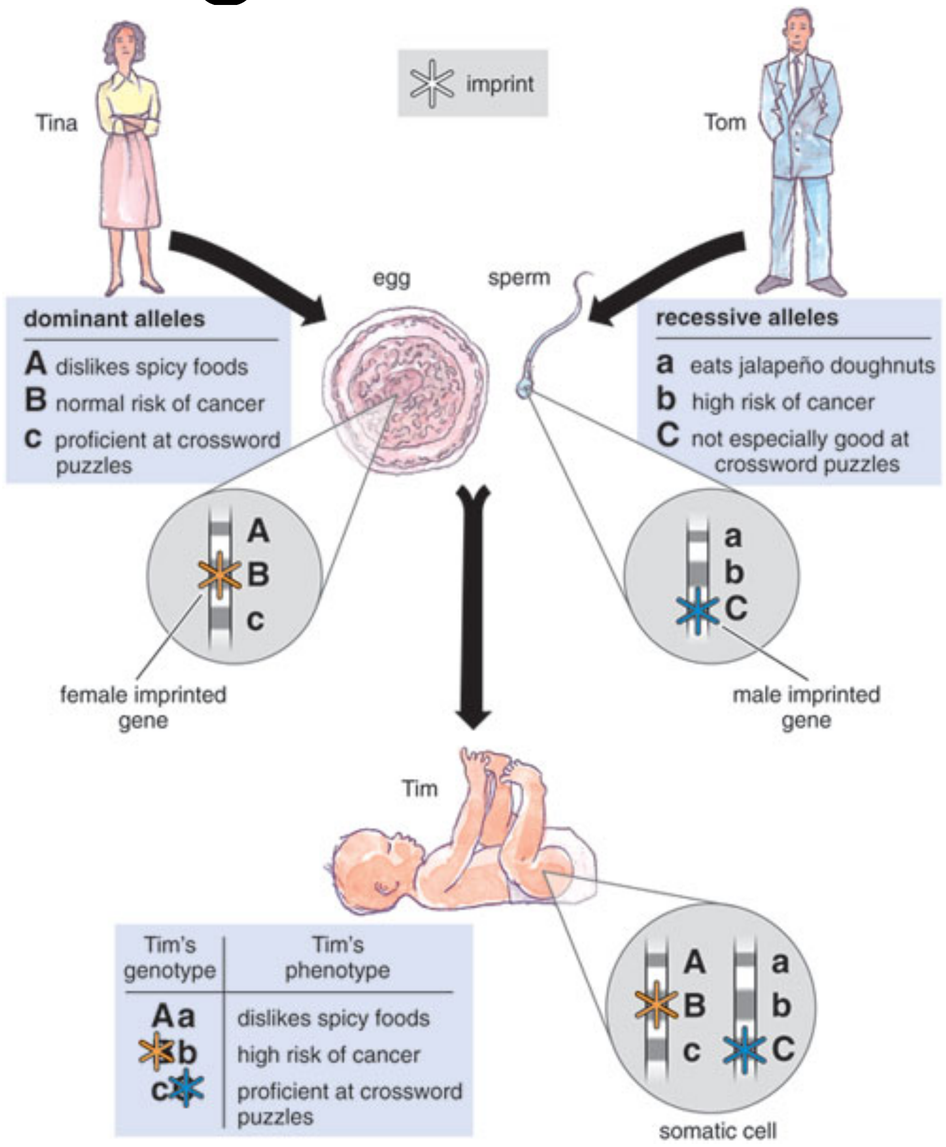
f. Epigenetic inheritance

Genomic imprinting

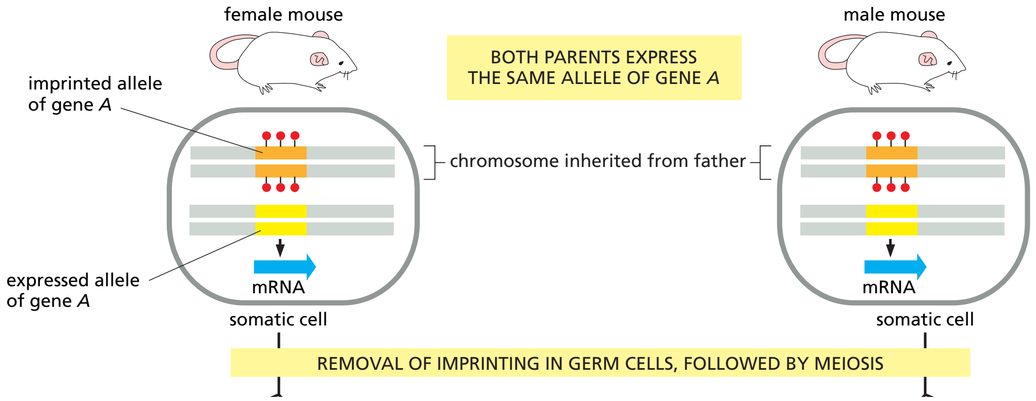
- Mammalian cells are **diploid** (one set of gene from the father and one from the mother)
- For a small subset of genes, expression depends on whether they have been inherited from the mother or the father
  - **when one copy is active, the other one is silent,** and *vice versa*
- This phenomenon is called **genomic imprinting**
- **~300** genes in humans

Genomic imprinting is an epigenetic phenomenon that causes genes to be expressed or not, depending on whether they are inherited from the female or male parent.

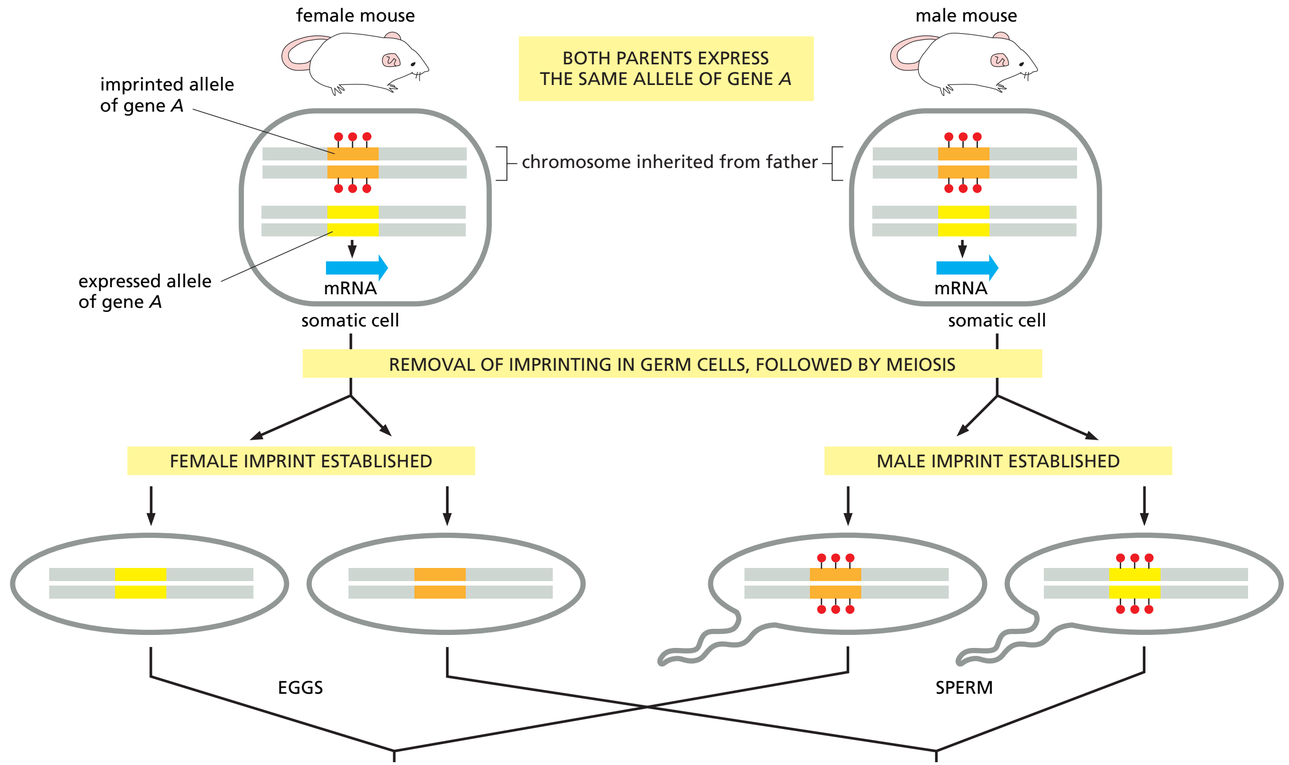
Genomic imprinting



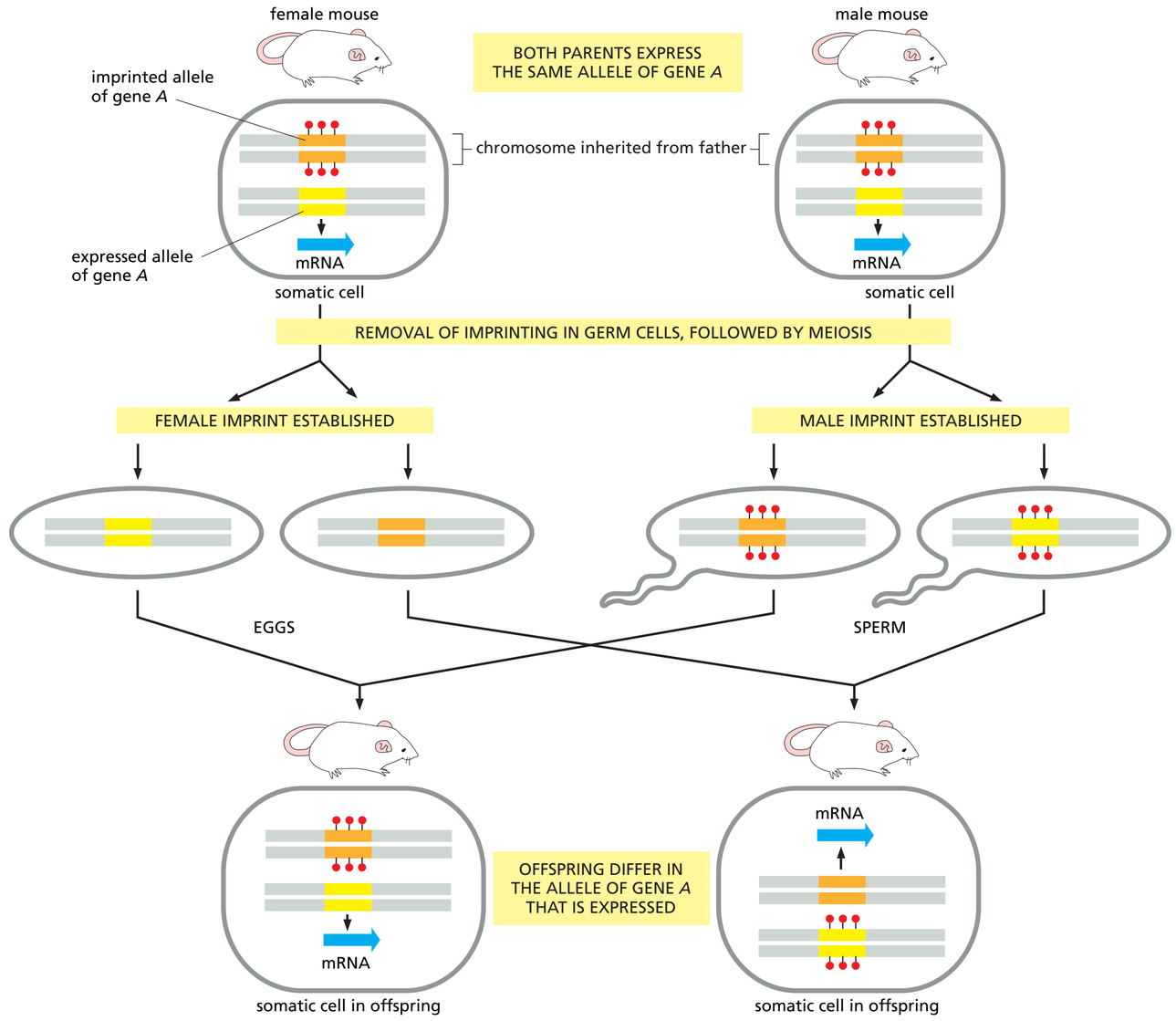
Genomic imprinting



Genomic imprinting



Genomic imprinting



IV. Control of gene expression

1. Transcriptional control

a. Transcription regulators (activators, repressors)

b. Understanding different control systems

c. Combinatorial gene control and cell types

d. Genomic imprinting

e. X-chromosome inactivation

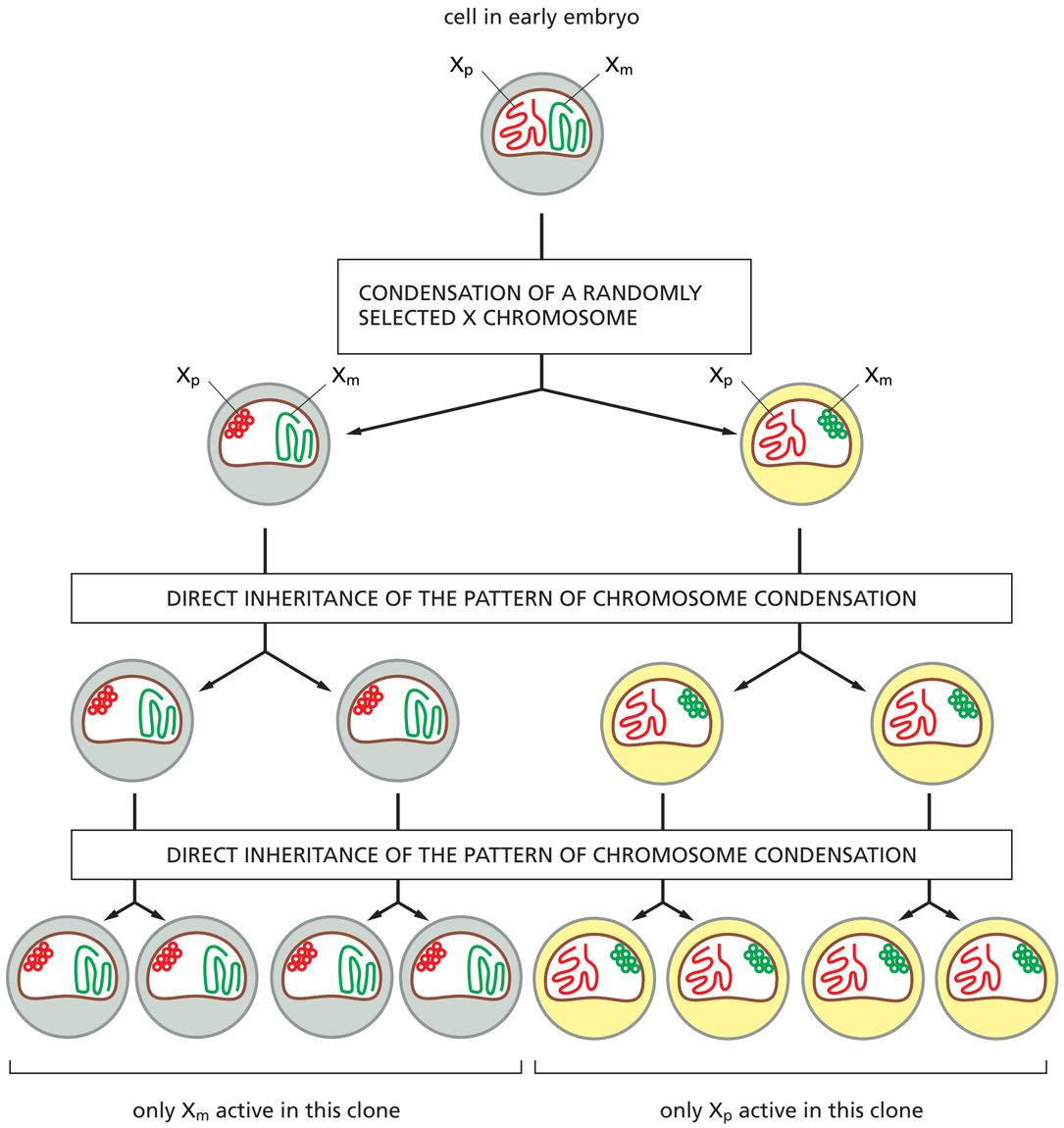
f. Epigenetic inheritance

X chromosome inactivation

- Males and females differ in their **sex chromosomes** (XX and XY)
- Female cells contain **twice the amount** of X genes as do male cells
- The X chromosome contains **> 1000 genes**, whereas the Y **<100 genes**
- Mammals achieve dosage compensation by the transcriptional **inactivation of one of the two X chromosomes in female somatic cells**

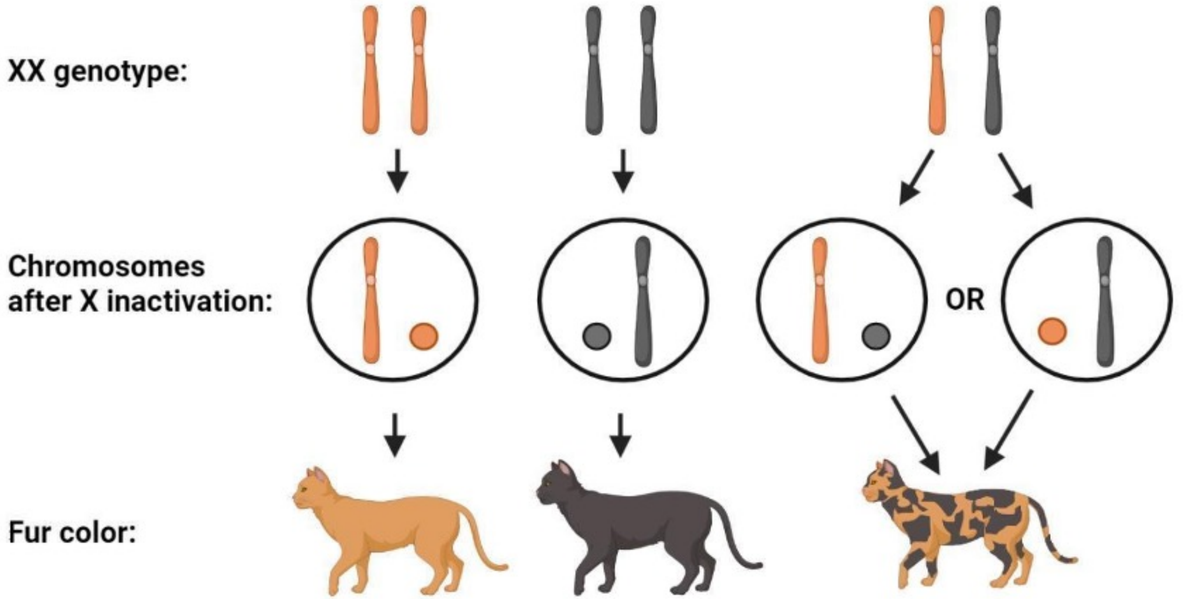
X-inactivation, a process occurring in female mammals to ensure dosage compensation, randomly silences one of the two X chromosomes in each somatic cell. This ensures that females (XX) have the same effective dose of X-linked genes as males (XY), who have only one X chromosome.

X chromosome inactivation



- In early embryo, one **X chromosome** (random) becomes highly **condensed**
- Once inactivated, it remains **silent** for all subsequent divisions
- It happens after several hundred cells are formed in the embryo, so **every female is a mosaic of clonal groups** of cells with either X chromosome silenced
- These clonal groups tend to remain **close together** during development
- Reversed during **germ-cell** formation

X chromosome inactivation

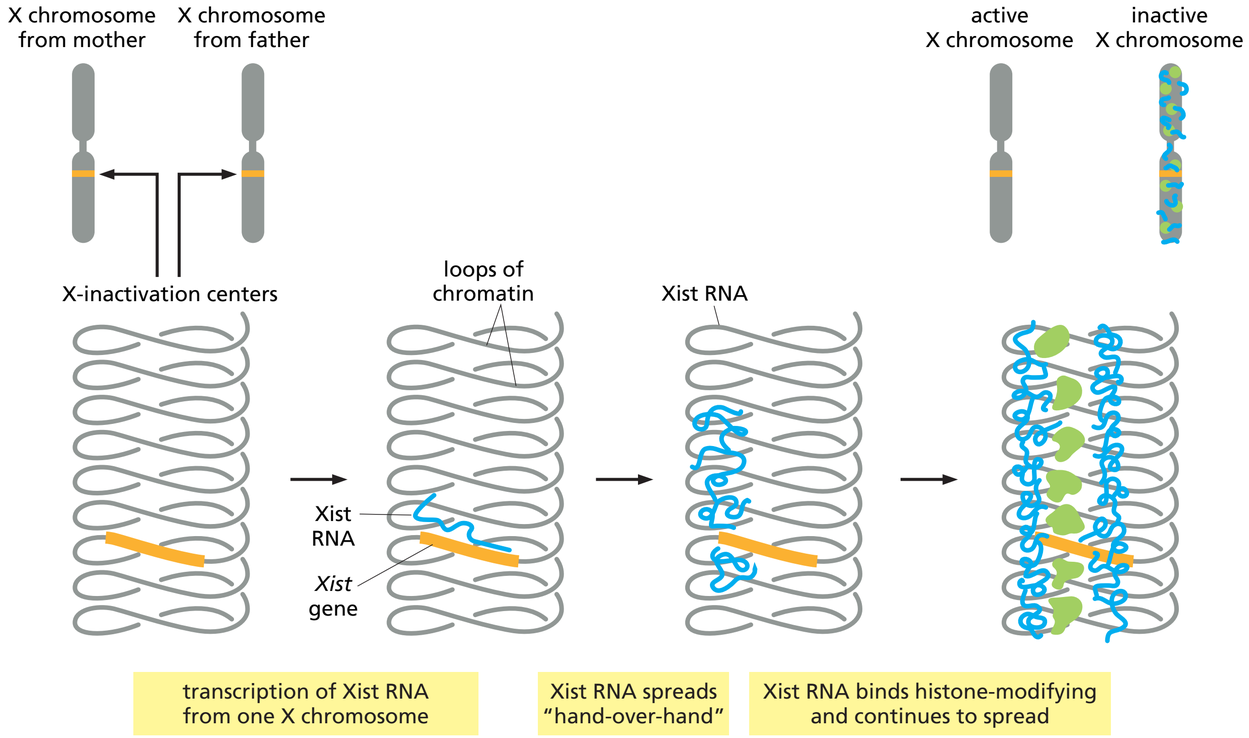


https://www.nature.com/scitable/blog/student-voices/

X chromosome inactivation

How is the **entire chromosome** inactivated?

- initiated from a single site near the middle of the X = the **X-inactivation center** (XIC)
- Within the XIC, a **lncRNA** (long non-coding RNA) is transcribed = **Xist**
- Xist **spreads over the chromosome** and silences genes
- Xist recruits **histone-modifying enzymes**
- 10% of the genes (including Xist) **escape** the silencing



Both imprinting and X-inactivation lead to **monoallelic gene expression** - only one of the two copies of a gene is expressed

X chromosome inactivation

The exact **mechanism by which the X chromosome to be inactivated is chosen** is still not fully understood, but several factors contribute:

- **Counting Mechanism**: The cell has a counting mechanism that ensures only one X chromosome remains active. If more than two X chromosomes are present (as in some cases of aneuploidy, like XXX individuals), all but one X will be inactivated.

- **Mutual Repression of Xist and Tsix**: Xist has an antisense partner called Tsix, which is also expressed from the XIC and acts to suppress Xist expression on the chromosome that remains active. Tsix and Xist are in a mutually exclusive relationship—only one X chromosome will end up expressing Xist (and become inactivated), while the other expresses Tsix (and remains active).

- **Stochastic Expression and Epigenetic Factors:** The random choice may involve stochastic (random) fluctuations in the levels of Xist and Tsix. Once one chromosome "wins" by expressing higher levels of Xist or Tsix, this choice is stabilized by epigenetic modifications (such as DNA methylation and histone modification) to ensure only one X remains active.

IV. Control of gene expression

1. Transcriptional control

a. Transcription regulators (activators, repressors)

b. Understanding different control systems

c. Combinatorial gene control and cell types

d. Genomic imprinting

e. X-chromosome inactivation

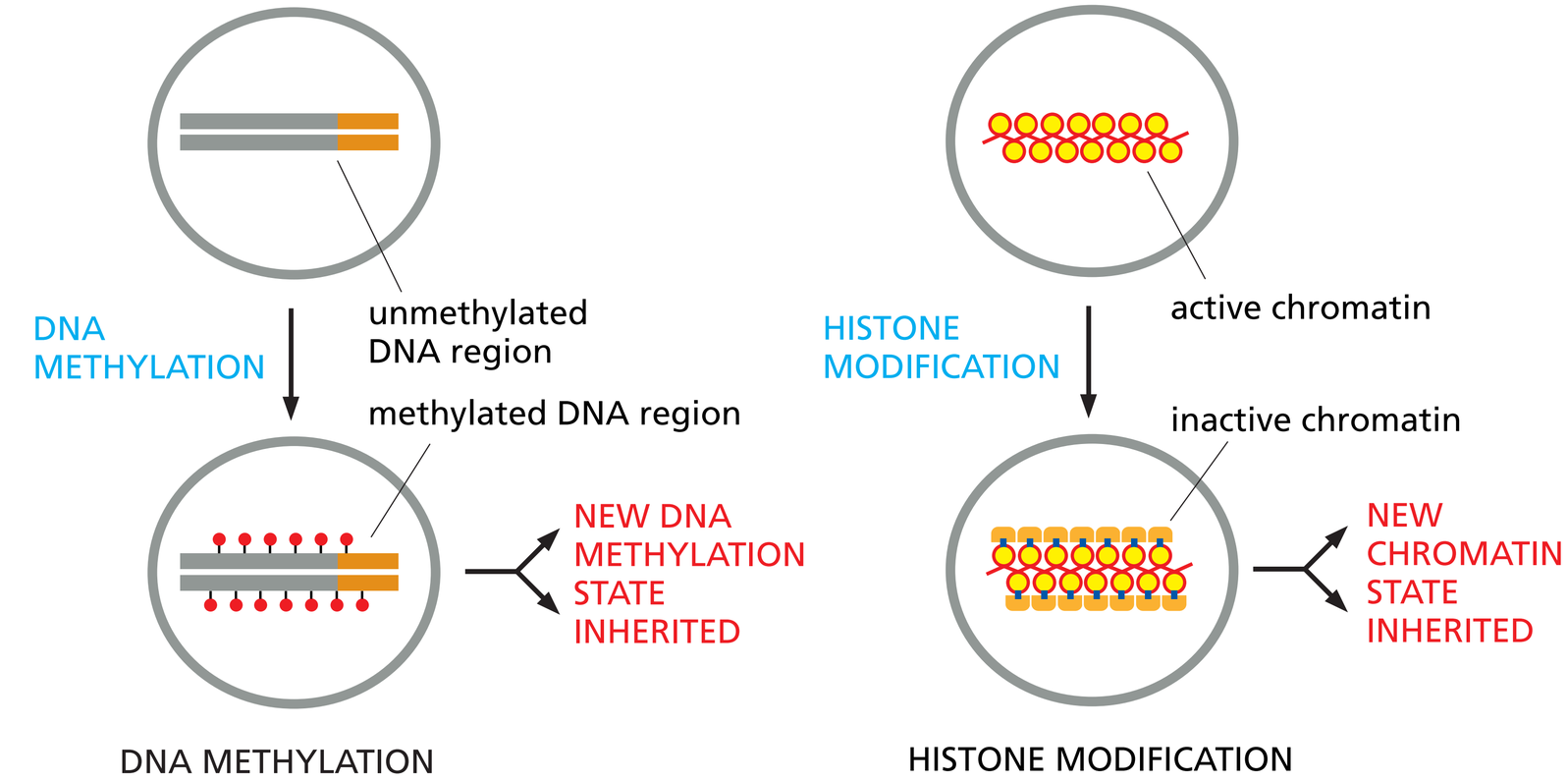
f. Epigenetic inheritance

Epigenetic inheritance

= The ability of a daughter cell to **retain a memory** of the gene expression patterns that were present in the parent cell

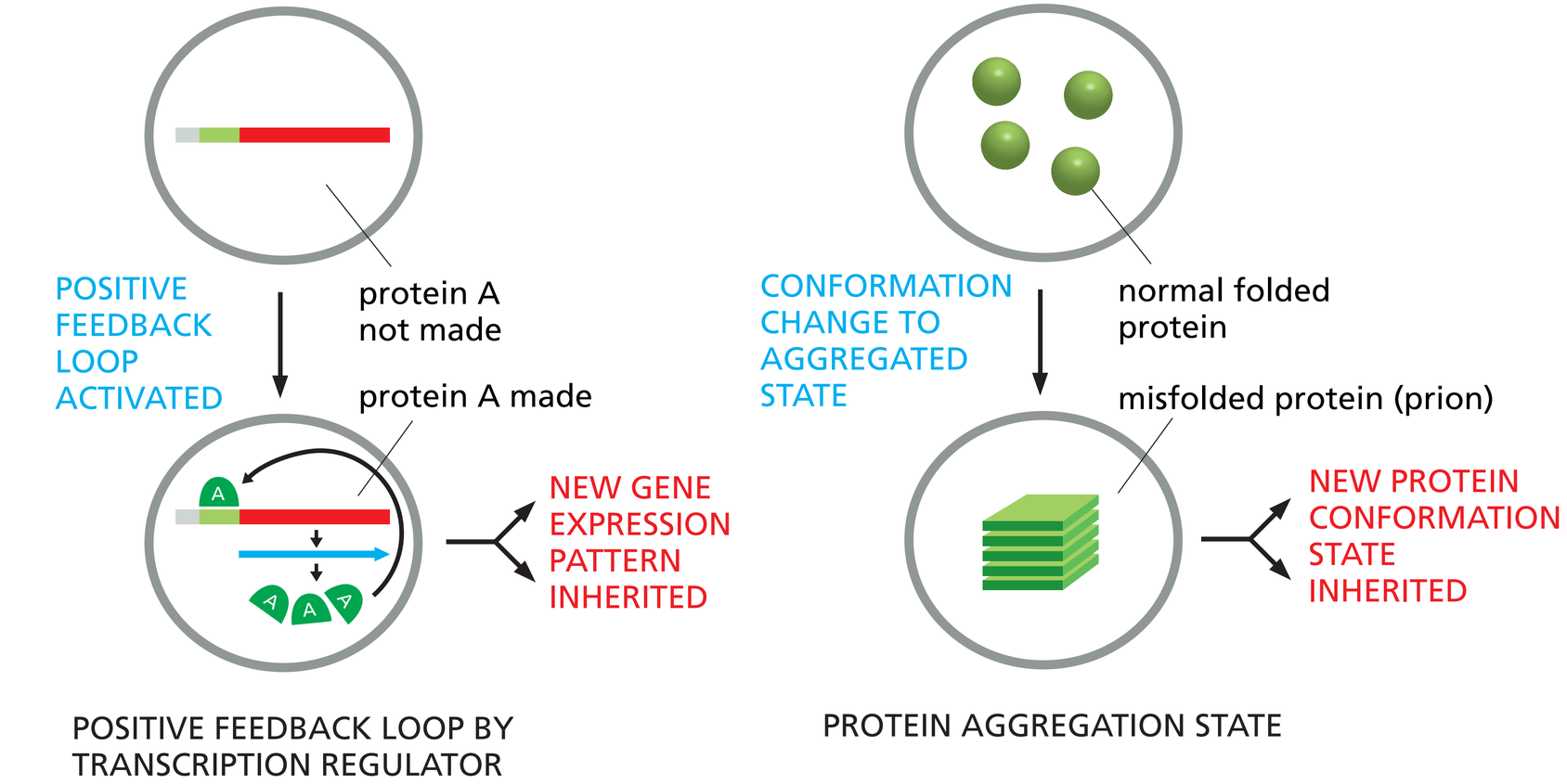
- Heritable alteration in a cell's **phenotype**
- Does not result from changes in the **sequence of DNA**
- **4 mechanisms** of epigenetic inheritance

Epigenetic inheritance



(A) EPIGENETIC MECHANISMS THAT ACT IN *CIS*

Epigenetic inheritance



POSITIVE FEEDBACK LOOP BY TRANSCRIPTION REGULATOR

PROTEIN AGGREGATION STATE

(B) EPIGENETIC MECHANISMS THAT ACT IN *TRANS*

IV. Control of gene expression

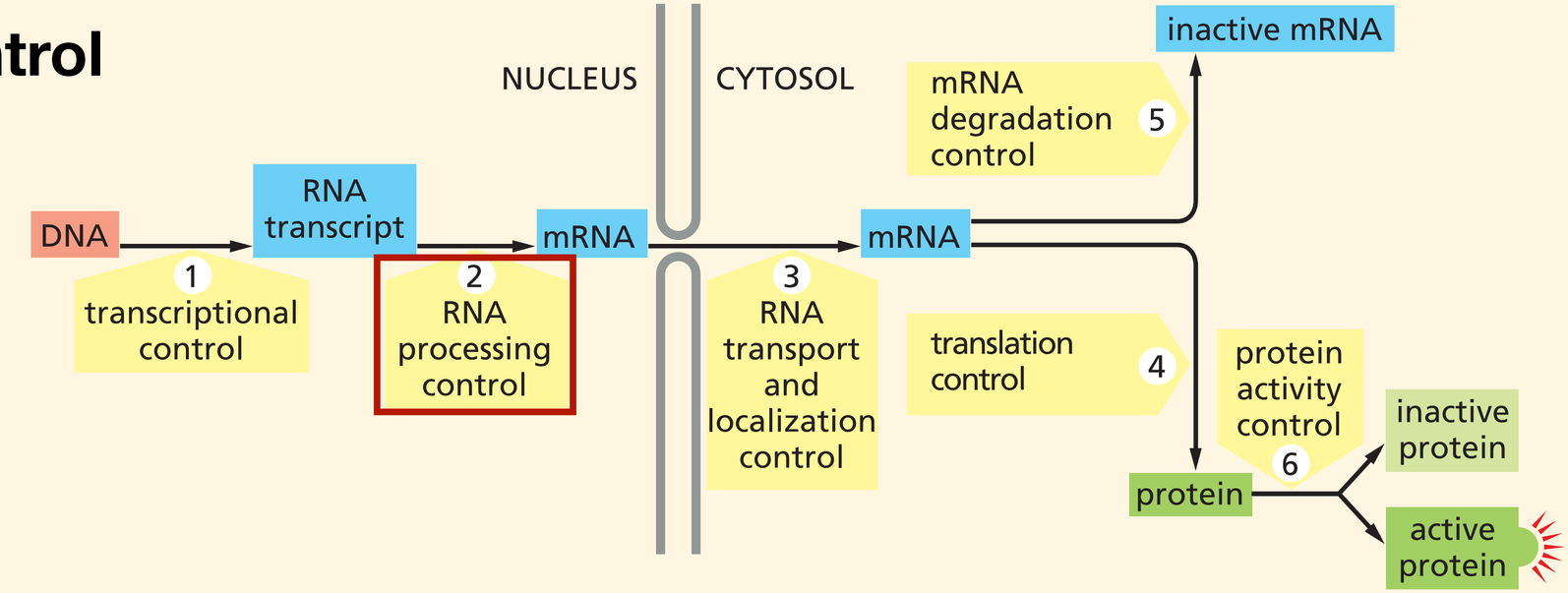
2. Post-transcriptional control

a. RNA processing

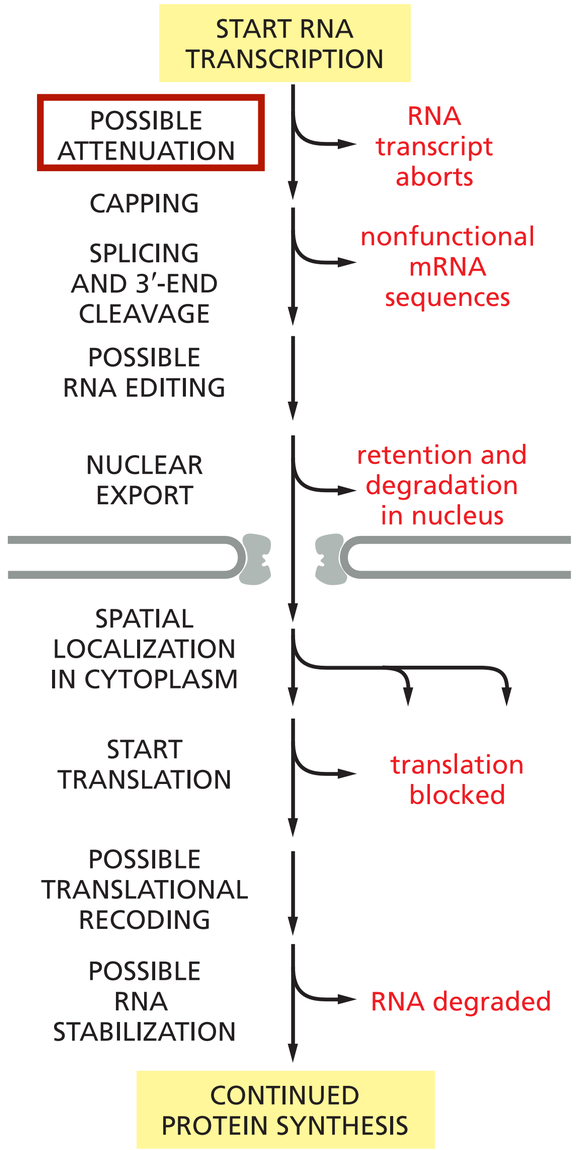
b. RNA export

c. Translational control

d. mRNA stability



Post-transcriptional controls



Transcription attenuation

The expression of some genes is inhibited by **premature termination of transcription**

- Only in **prokaryotes**
- relies on the capacity for prokaryotes to undertake **transcription and translation** at the same time (as the DNA is cytosolic)
- The nascent RNA chain can adopt a **structure** that causes it to interact with the **RNA polymerase**
- This leads to the **abortion of transcription**
- When the gene is needed, proteins bind to the RNA and **remove the attenuation**
- Example: the trp operon - trp repressor system blocks the initiation of transcription, attenuation blocks the completion of transcription

Transcription attenuation

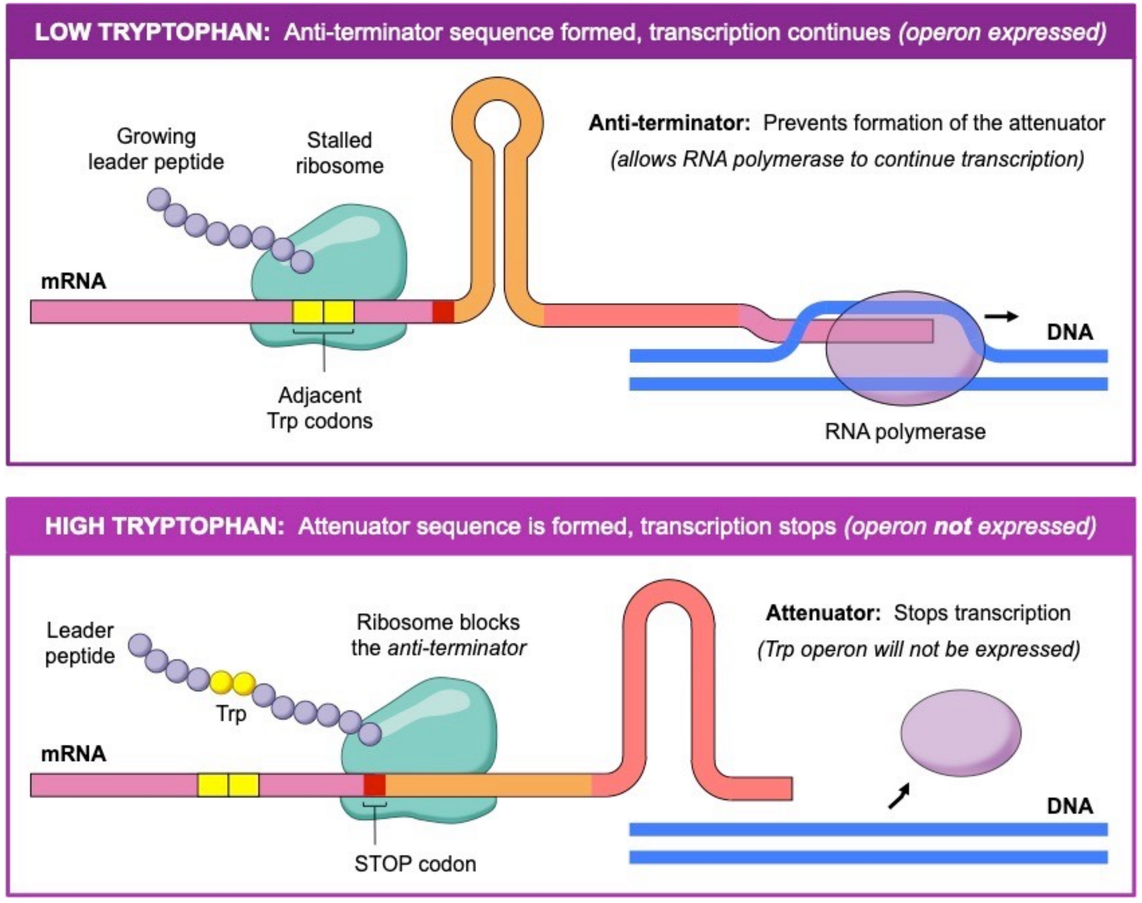
Located between the operator and trp genes is a **leader sequence that encodes a short polypeptide containing two Trp residues**

- When transcribed, this leader sequence contains several self-complementary sections that are capable of forming hairpin loops
- One such loop is called the attenuator and functions as a site of transcriptional termination when it is formed (transcription halts)
- Alternatively, an earlier loop called the anti-terminator may form, which prevents attenuator formation (so transcription continues)

The **presence of tryptophan determines whether the attenuator or anti-terminator** is formed within the leader sequence

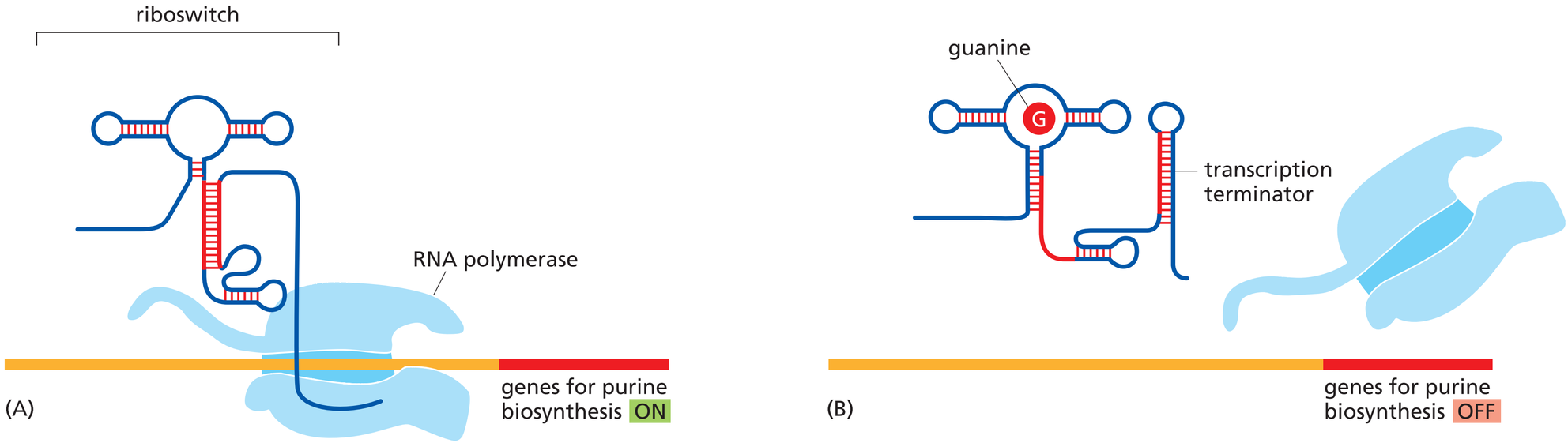
- If tryptophan is in low supply, the ribosome will pause at the two Trp codons within the leader sequence
- This pause allows time for the anti-terminator to form on the mRNA transcript, so transcription will continue (operon is expressed)
- If tryptophan is in high supply, the ribosome will not need to pause at the two Trp codons within the leader sequence
- This means the anti-terminator is unable to form and the attenuator will form instead (transcription halted ; operon not expressed)

Transcription attenuation



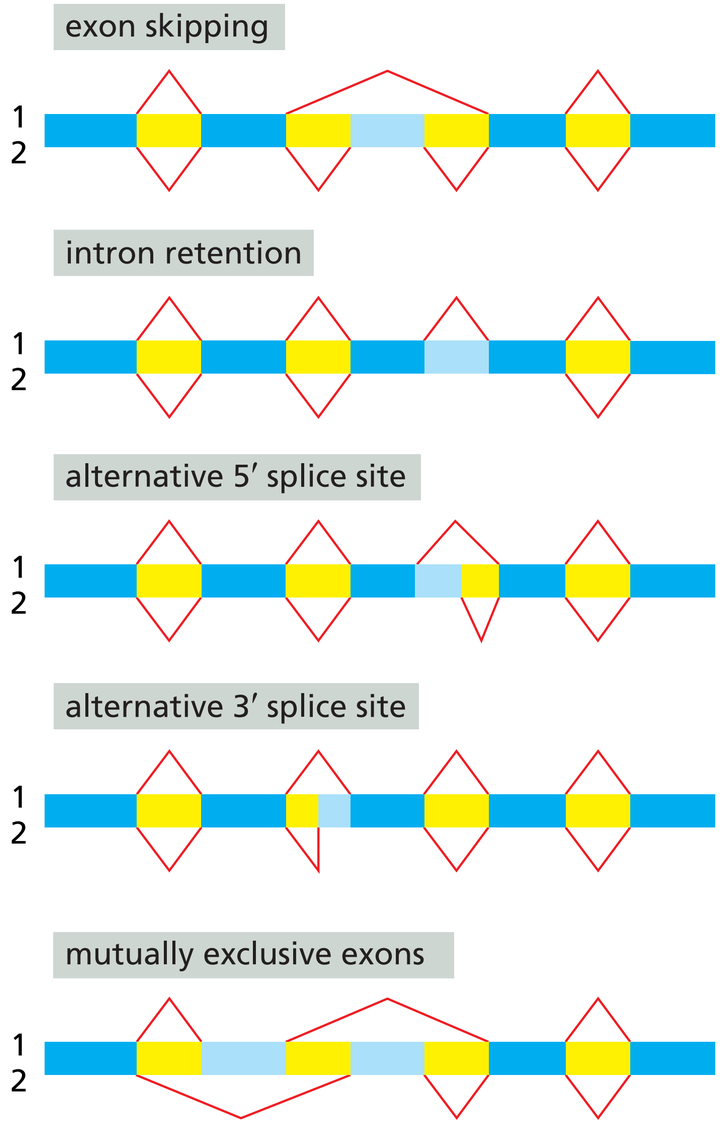
Riboswitches

- **Short sequences of RNA** that change conformation upon **metabolite binding**
- The **conformational change** regulates gene expression
- Often at **5’ end of mRNA**, folding during mRNA synthesis
- **Blocking or permitting** progress of the RNA polymerase



Alternative splicing

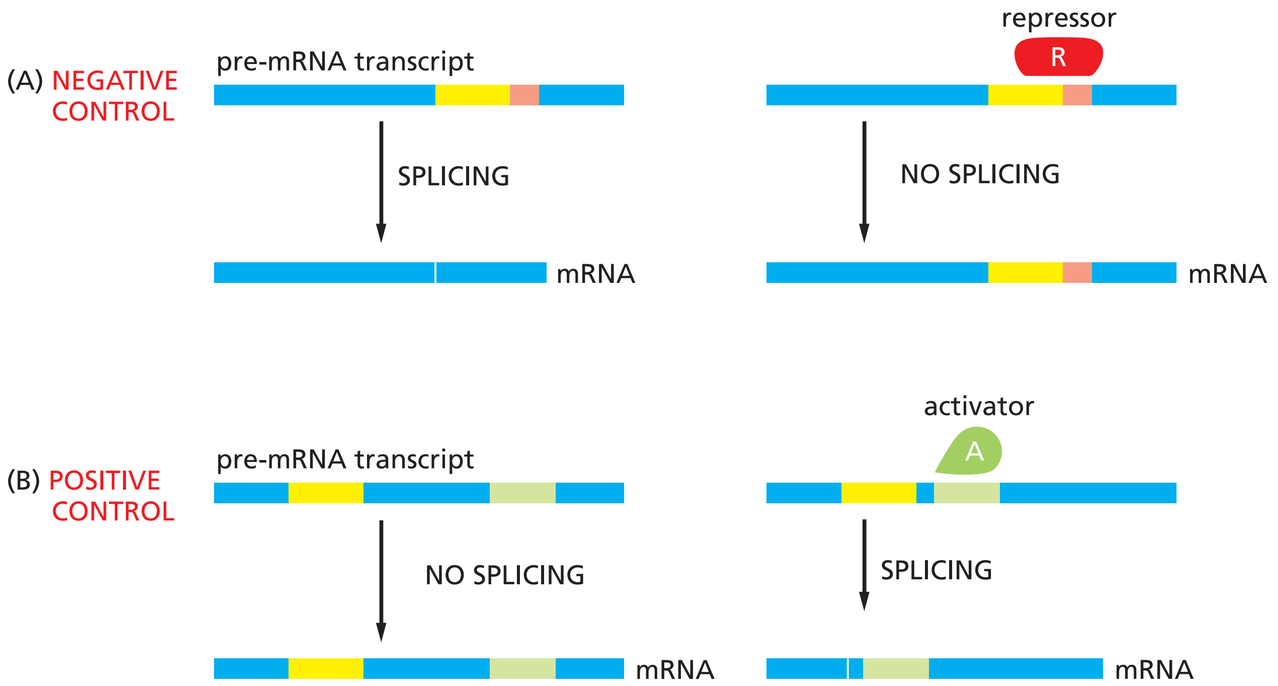
- **Splicing** = removes the introns from the mRNA precursor
- **Alternative splicing** allows to make different polypeptide chains from the same gene
- ~ **90% of human genes** produce multiple proteins this way



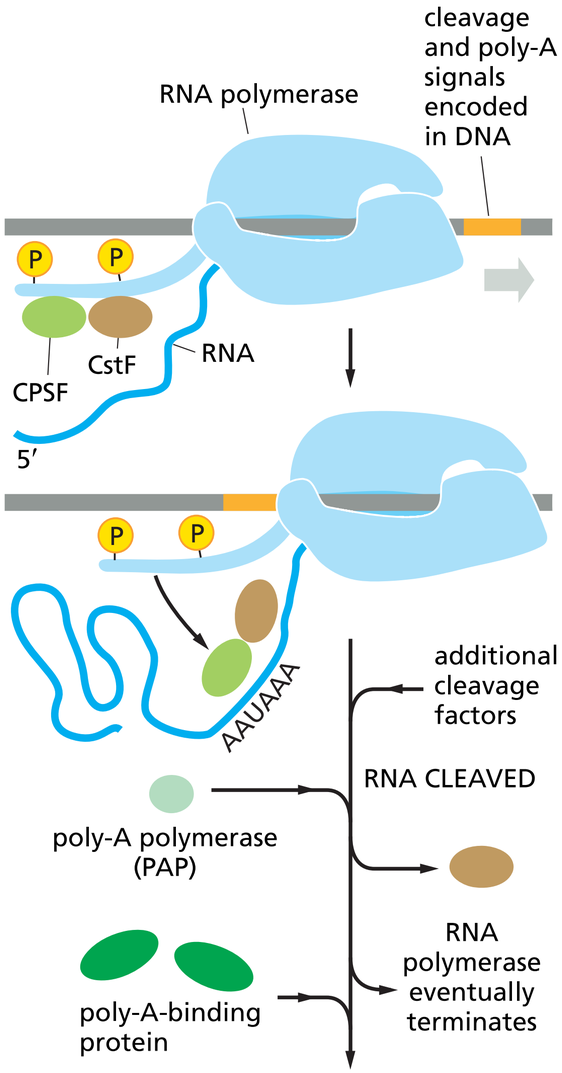
Alternative splicing

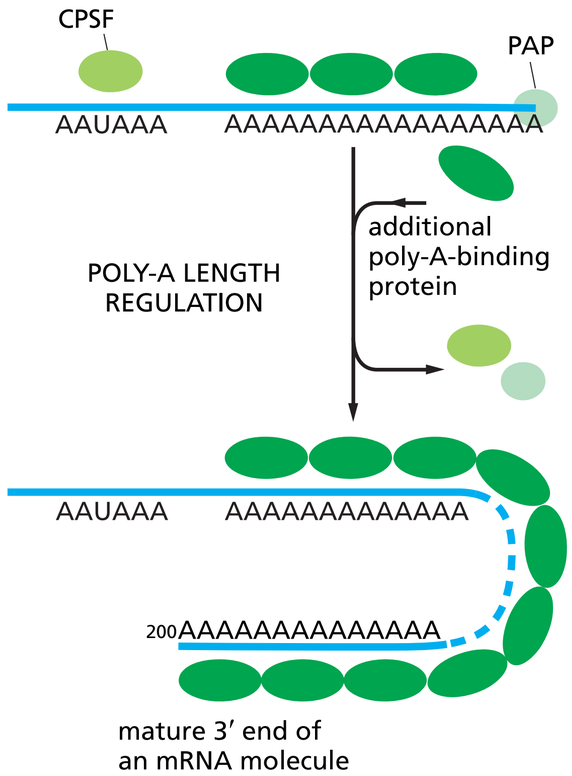
- RNA splicing can be regulated **negatively** (i.e. a regulatory molecule prevents the splicing machinery from accessing a site)

- RNA splicing can be regulated **positively** (i.e. a regulatory molecule helps direct the splicing machinery to a site)



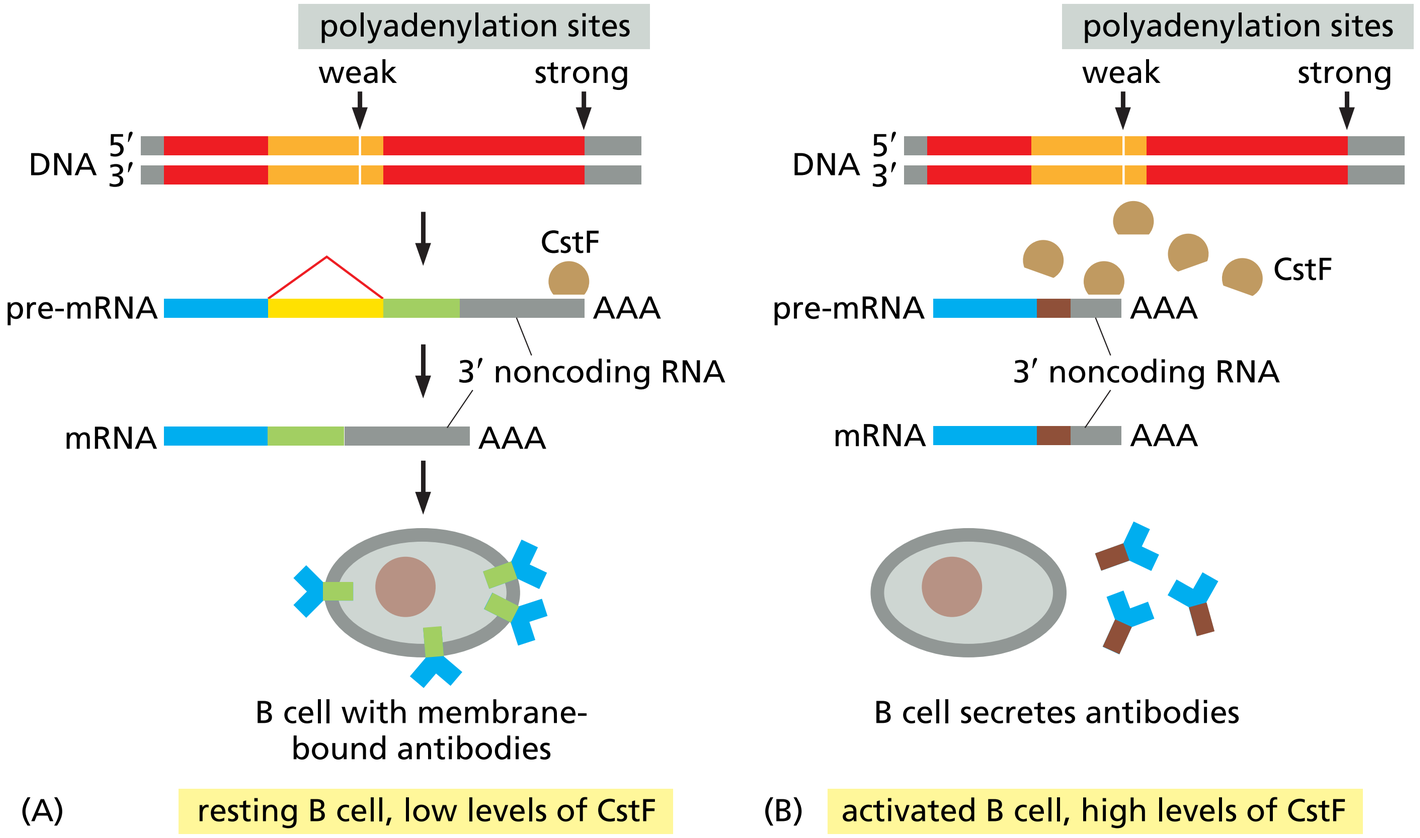
Transcript cleavage and polyA addition





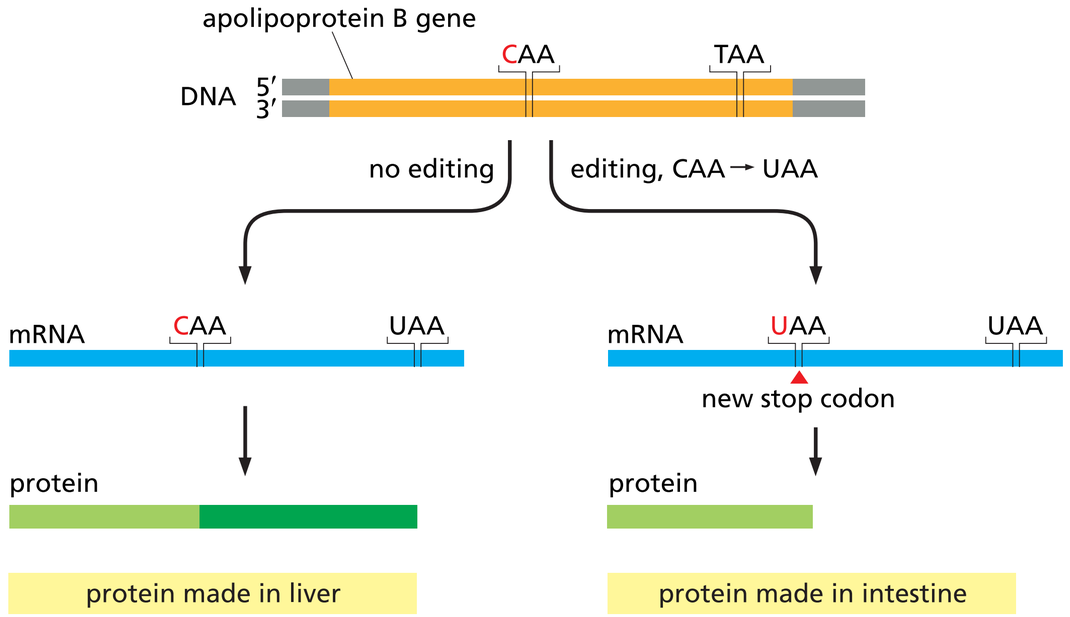
Transcript cleavage and polyA addition

- Cells can **control the site of cleavage and polyadenylation (polyA)** (leading to longer or shorter proteins)



RNA editing

- Alters the **nucleotide sequence** of RNA transcripts
  - **deamination of adenines** to produce inosine (A-to-I editing)
  - **deamination of cytosine** to produce uracil (C-to-U-editing)
- If it happens in a **coding region**, it may change the amino acid sequence of the protein
- If it happens **outside of a coding region**, it may affect splicing, transport of mRNA, efficiency of translation,…



IV. Control of gene expression

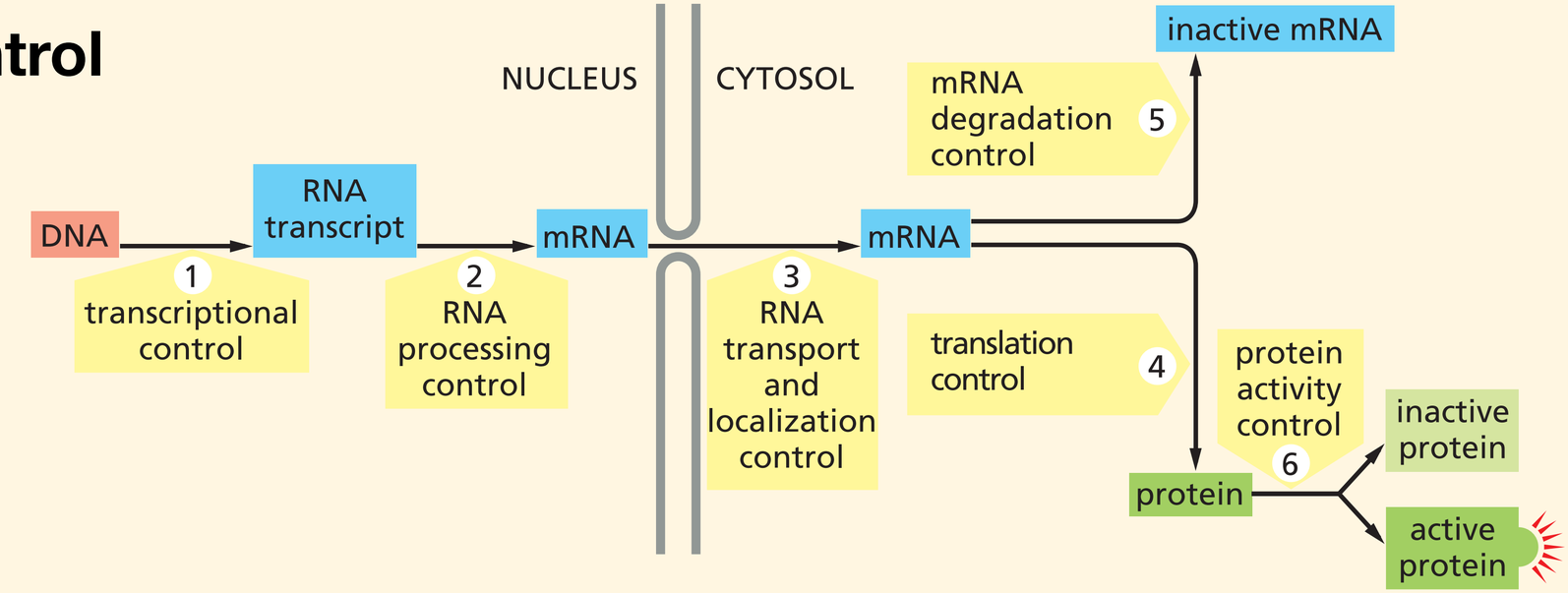
2. Post-transcriptional control

a. RNA processing

b. RNA export

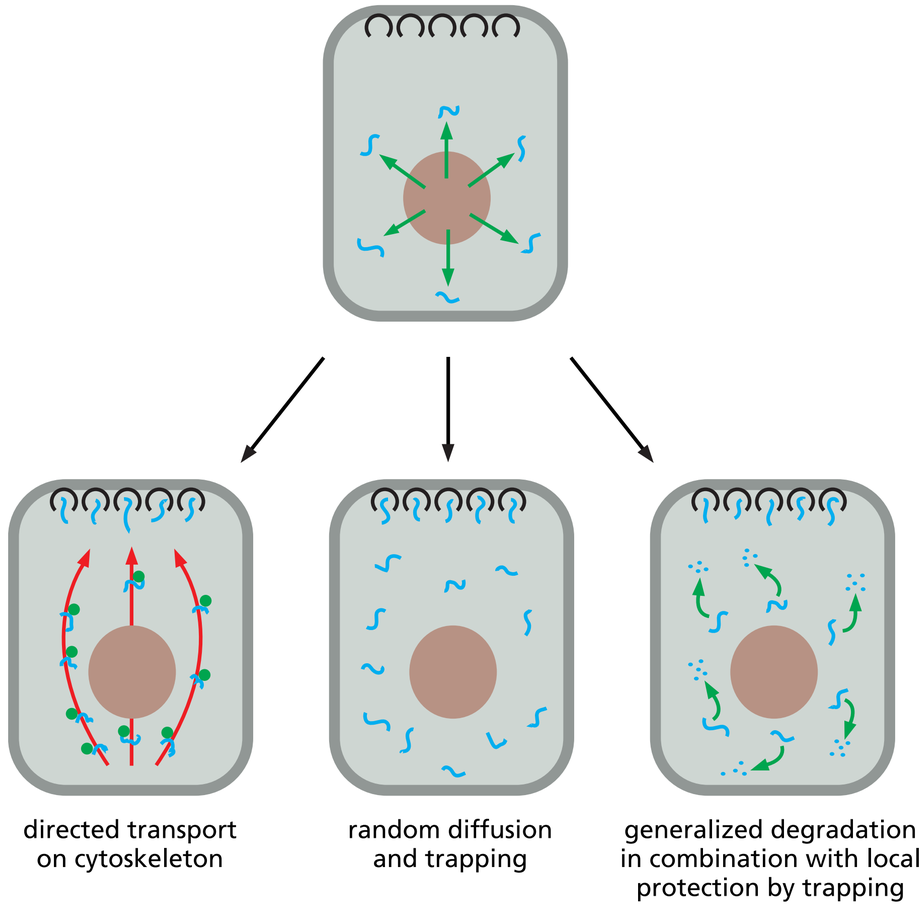
c. Translational control

d. mRNA stability



RNA export

- Only happens once the RNA has been **processed**
- Once out of the nucleus, the mRNA can reach **different locations** in the cell
- For membrane or secreted proteins, the mRNA is targeted to the **endoplasmic reticulum**
- Many mRNAs are directed to **specific intracellular locations**, close to sites where the protein is needed
- This allows the establishment of **asymmetries** in the cytosol of the cell



IV. Control of gene expression

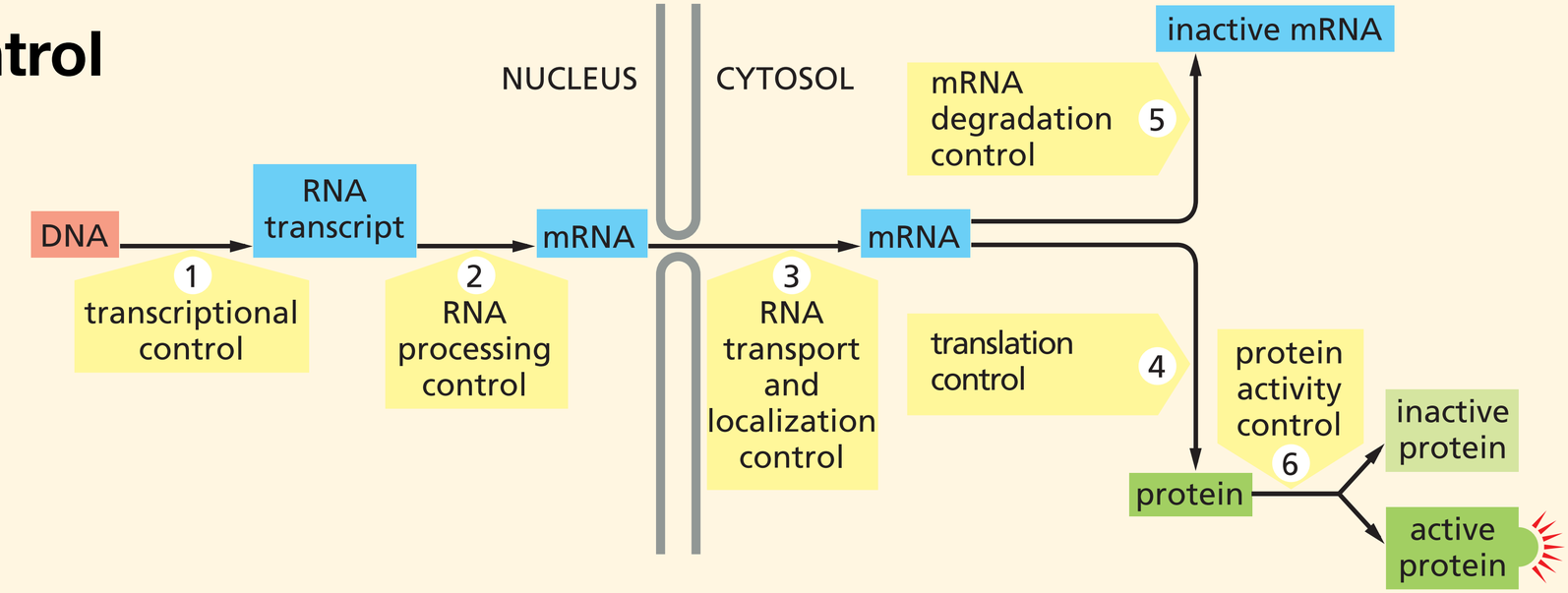
2. Post-transcriptional control

a. RNA processing

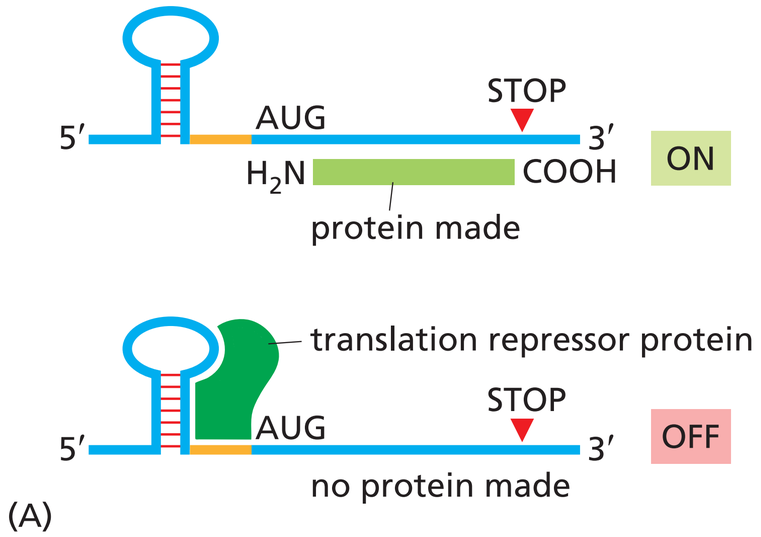
b. RNA export

c. Translational control

d. mRNA stability



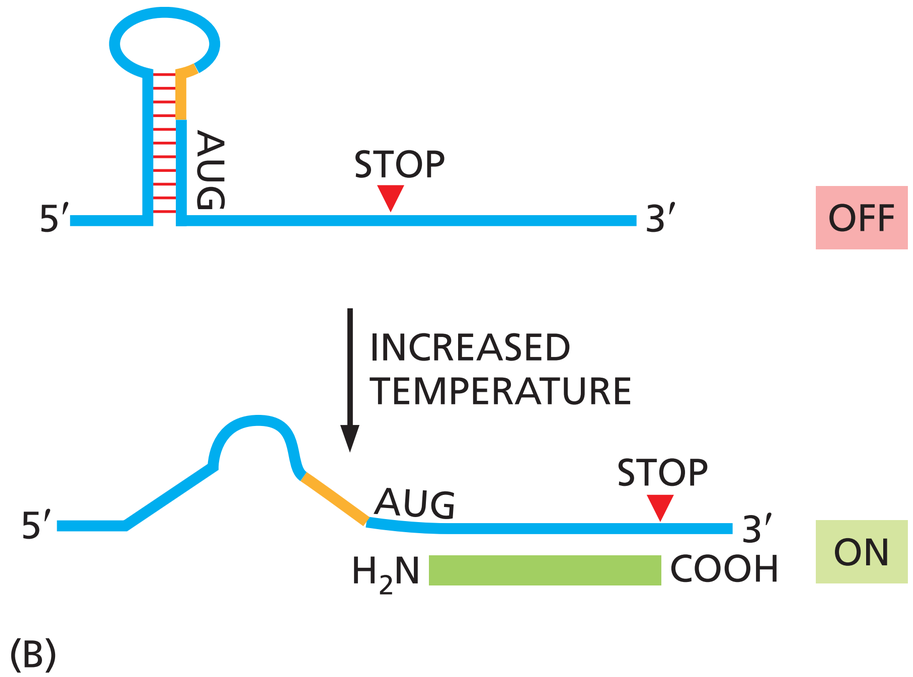
Translational control (I)



- 4 examples from bacteria - but similar principles apply to eukaryotes

- A translation repressor protein binds to a specific DNA sequence that blocks the access of the ribosome to the **Shine-Dalgarno** sequence (upstream of the AUG codon, in orange, only in bacteria)

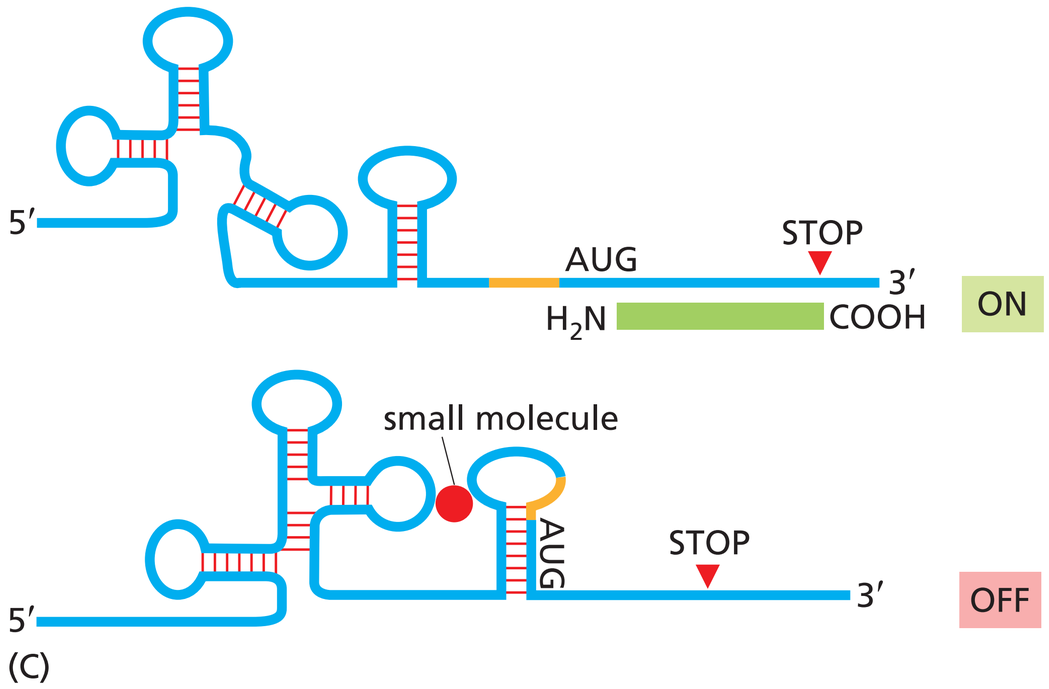
Translational control (II)



(B)

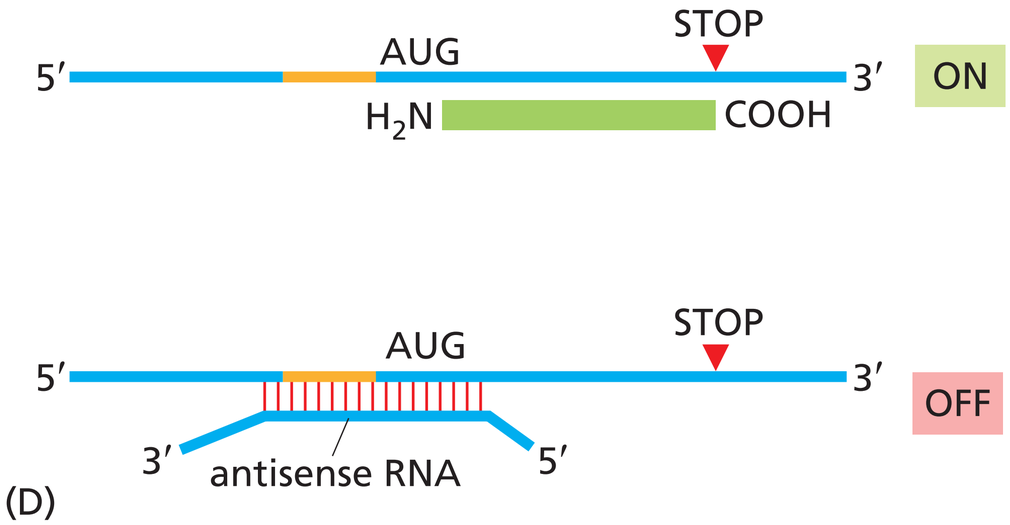
- An RNA “thermosensor” permits efficient translation only at **elevated temperatures**, when the stem loop has been melted.

Translational control (III)



- A small molecule binds to a **riboswitch** and causes the rearrangement of the RNA forming a different set of stem-loop which blocks the Shine-Dalgarno sequence

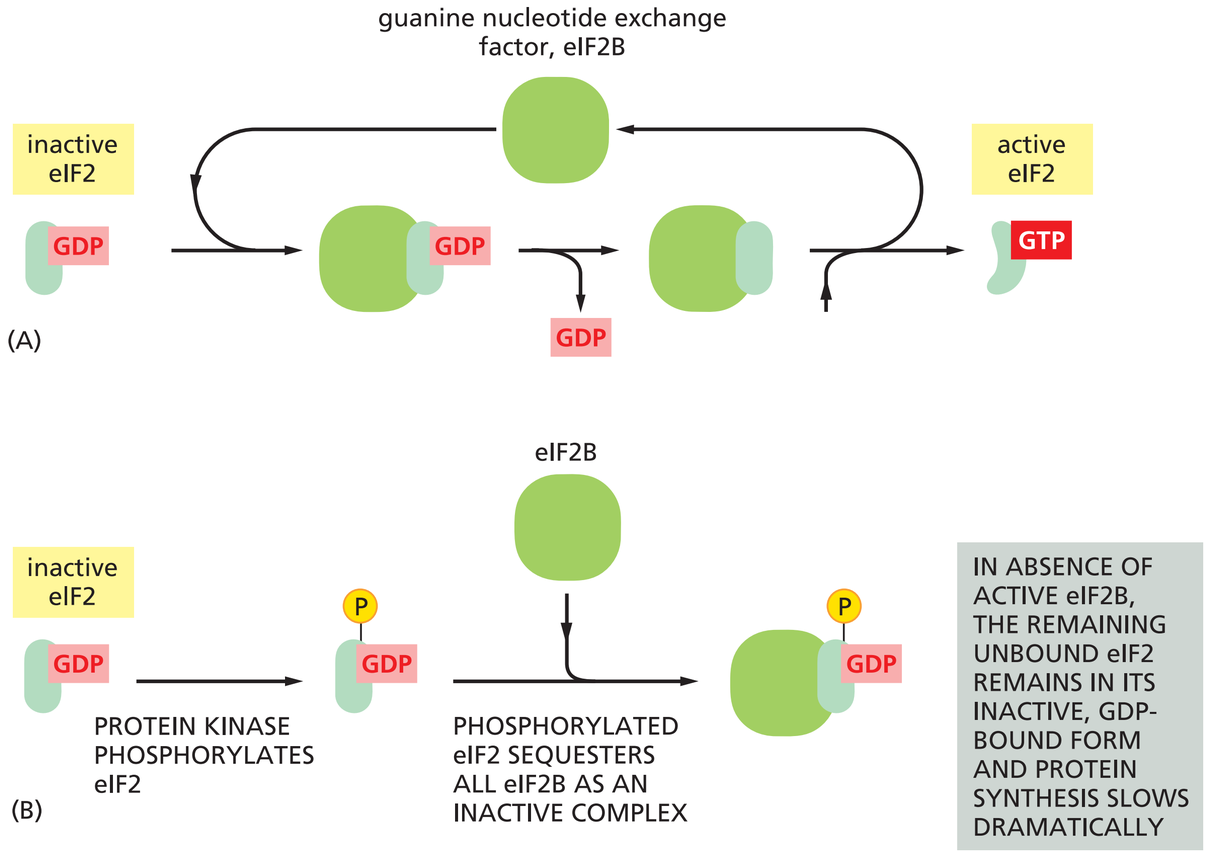
Translational control (IV)



- An **non-coding RNA** produced somewhere else in the genome base-pairs with a specific mRNA, which blocks translation

Translational control

- Cells decrease protein synthesis in response to **stresses** (low nutrients, infection, temperature increase, etc.)
- In Eukaryotes, this happens through the **phosphorylation** of the translation initiation factor eIF2



IV. Control of gene expression

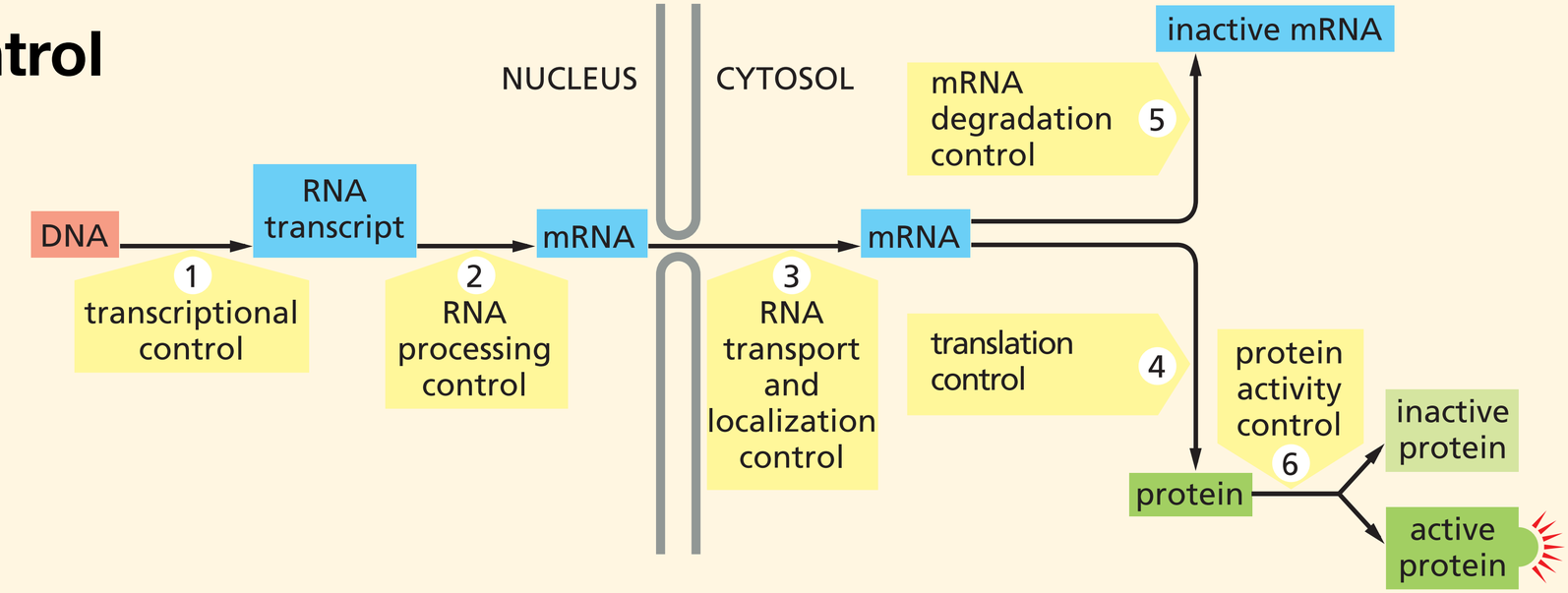
2. Post-transcriptional control

a. RNA processing

b. RNA export

c. Translational control

d. mRNA stability



mRNA stability

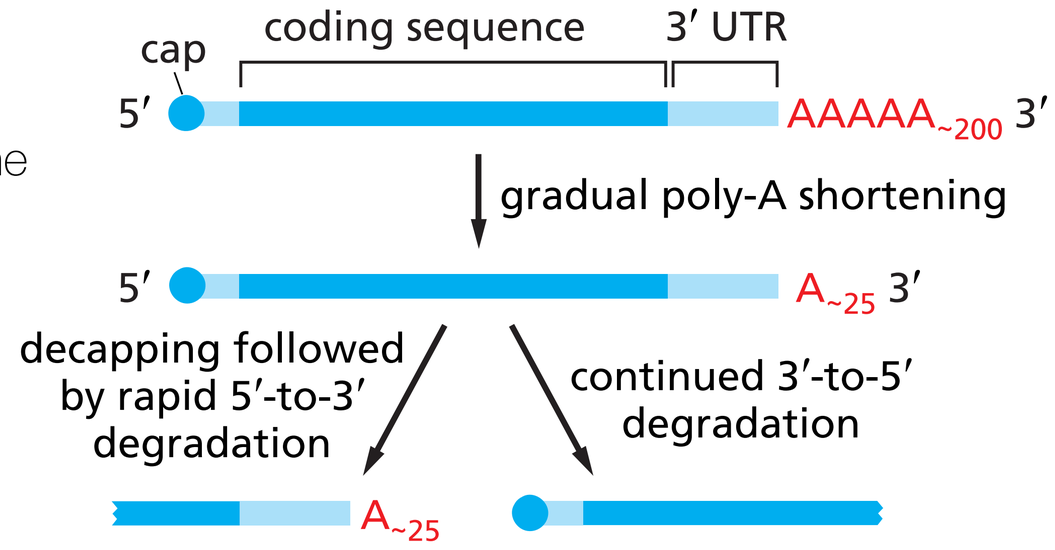
In **bacteria**,

- Most mRNA are very **unstable** (half-life of <few minutes)
- **Exonucleases** are responsible for the degradation of mRNAs

mRNA stability

In **Eukaryotes**,

- mRNAs are more **stable** (half-lives from 30min to 10h)
- shortening of the **polyA tail by exonucleases** (count down of the mRNA half-live)
- when the **polyA reaches ~ 25 nt**
  - **cap** is removed and mRNA is degraded from 5'
  - mRNA continues to be degraded **from 3'**



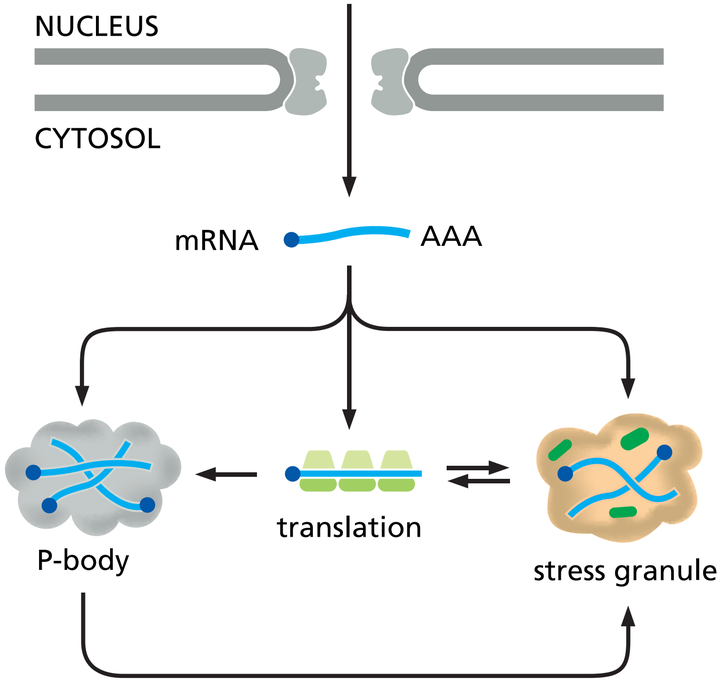
mRNA stability

P-bodies, or processing bodies

- dynamic, membrane-less granules within the cytoplasm of eukaryotic cells
- hubs for mRNA metabolism, particularly for mRNA decay and storage
- P-bodies are involved in post-transcriptional regulation, influencing gene expression by controlling the stability, translation, and degradation of mRNA

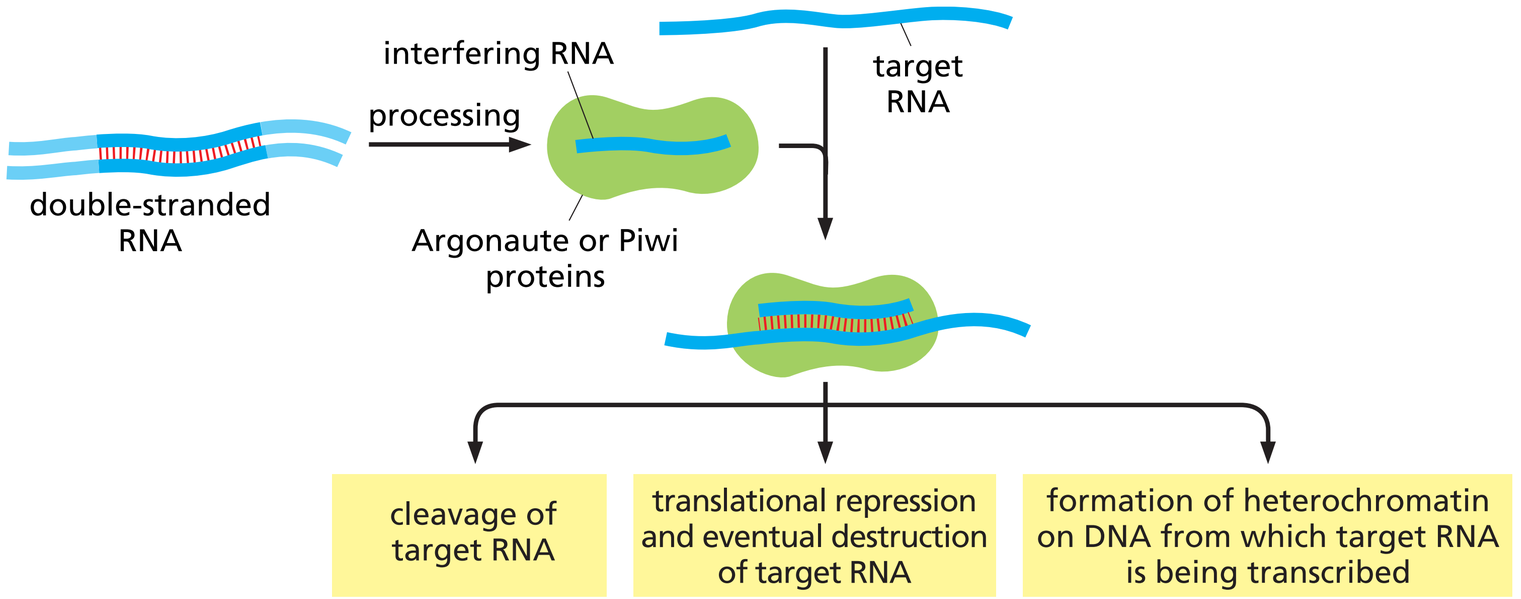
Stress granules

- dynamic, membrane-less cytoplasmic aggregates of RNA and protein that form in response to cellular stress, such as oxidative stress, heat shock, or nutrient deprivation
- crucial role in helping cells survive by temporarily halting translation and preserving select mRNAs under adverse conditions
- helps conserve energy and resources, allowing the cell to prioritize immediate stress response needs while protecting non-essential mRNAs for future use



Noncoding RNAs - RNA interference (RNAi)

- **Short single-stranded RNAs** (20-30 nucleotides) serving as guide RNAs that bind other RNAs in the cell
- When the target is an **mRNA**, it **prevents** its translation or catalyzes its **degradation**
- When the target is an **RNA being transcribed**, it can direct the formation of **repressive chromatin**

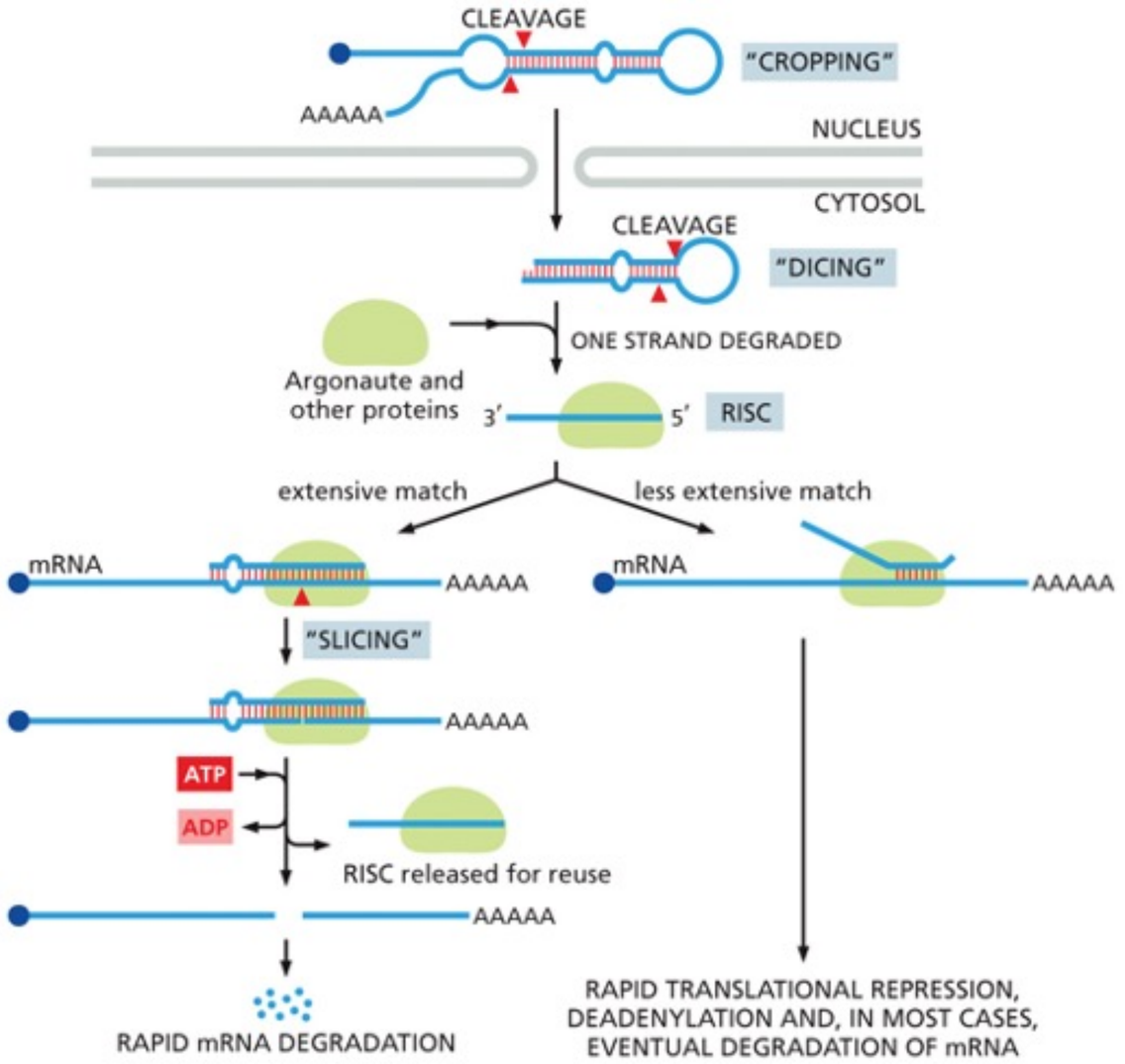


miRNA, siRNA and piRNA

MicroRNA (miRNA), small interfering RNA (siRNA), and piwi-interacting RNA (piRNA) are all small RNA molecules involved in gene regulation, primarily by silencing or down-regulating target genes. However, they differ in their origins, structures, mechanisms, and biological roles.

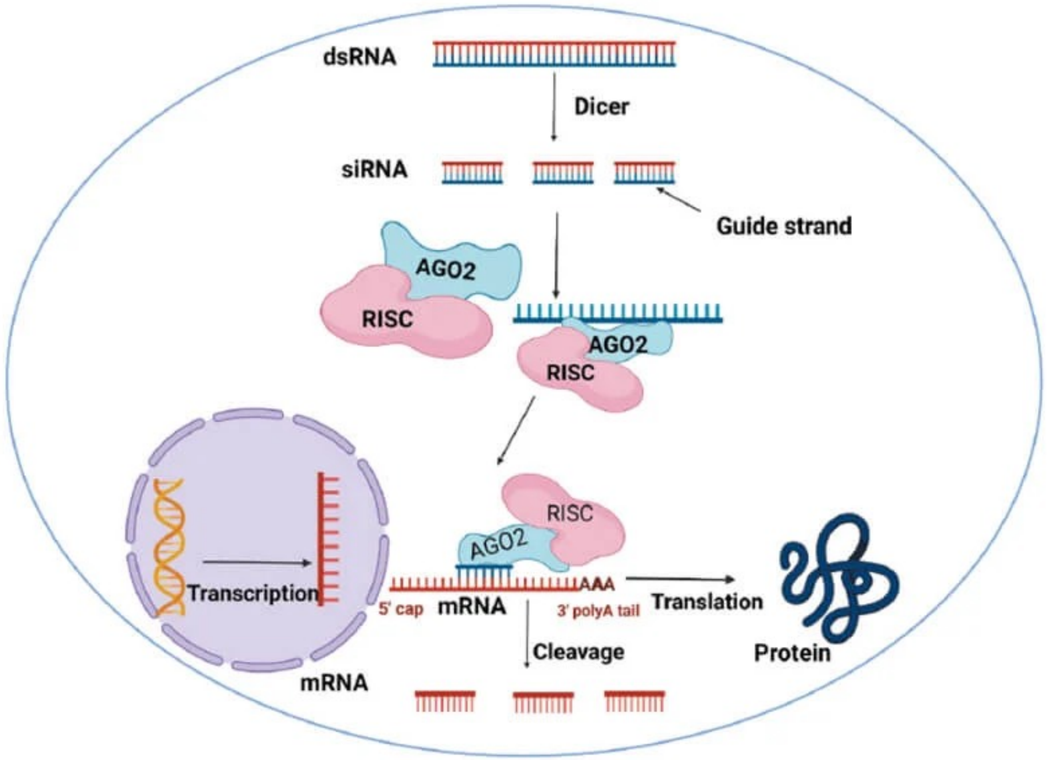
micro-RNAs (miRNAs)

- **>1000** miRNAs endogenously encoded in the human genome, controlling at least **half of the coding genes**
- regulate **gene expression** by binding to complementary sequences in the 3' untranslated region (3' UTR) of target mRNAs. They usually result in translational repression or mRNA degradation, although the exact mechanism depends on the level of sequence complementarity.
- made by **RNA polymerase II**, **capped** and **poly-adenylated**
- special processing and assembly with a set of proteins forming the **RNA-induced silencing complex (RISC)**
- The complex searches for **complementary sequences**
- a single miRNA can regulate a **whole set of different mRNAs**

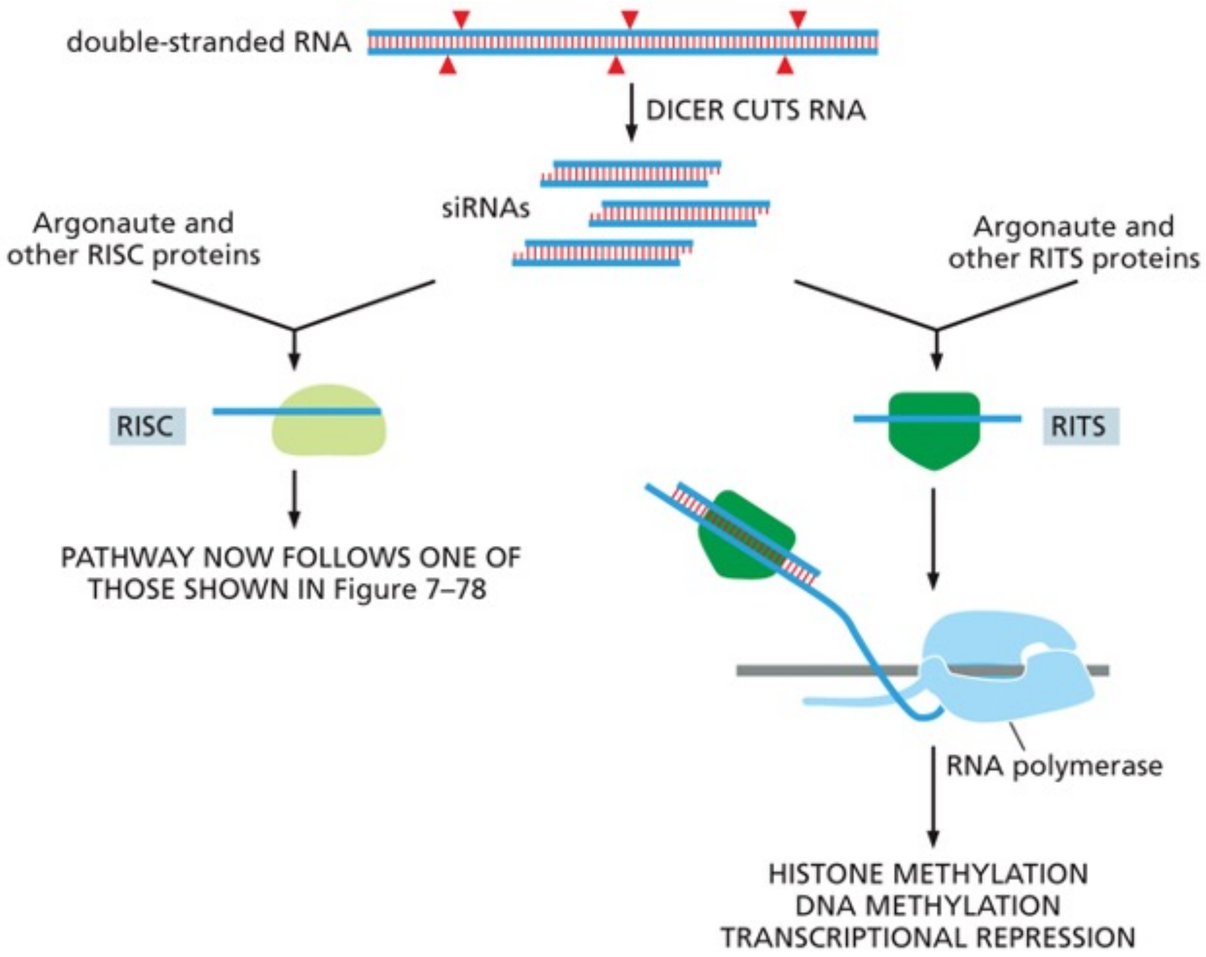


Small interfering RNA (siRNA)

- siRNAs can be endogenous (e.g., from transposons or viruses) but are often **exogenous**, such as those introduced experimentally
- siRNAs are typically 20-25 nucleotides long and have perfect or near-perfect complementarity to their target RNA. They are often derived from **longer double-stranded RNAs** (dsRNAs) that are processed by the enzyme **Dicer** into siRNA duplexes.
- siRNAs guide the **RNA-induced silencing complex** (RISC) to degrade specific mRNAs through a perfect base-pairing mechanism. This direct pairing leads to the target **mRNA's cleavage and degradation**, reducing gene expression.
- siRNAs are part of the **cellular defense against viral infection and transposons**. They are also widely used as a tool in research to **"knock down" genes** in experiments by selectively silencing specific mRNAs.



Small interfering RNA (siRNA)



- In some cases, RNAi shuts off **synthesis** of mRNA
- **RITS** = RNA-induced transcriptional silencing complex
- Causes the formation of **heterochromatin** (H3K9me3 mark)
- Maintains **transposable elements** in the silent form

Piwi-interacting RNA (piRNAs)

- piRNAs are derived from specific genomic regions called **piRNA clusters**, which are often transposon-rich. They do not require Dicer for processing but are instead processed through a **unique biogenesis pathway** (long single-stranded RNA)

- They do not form duplexes like siRNA and miRNA and do not have a defined secondary structure.

- piRNAs interact with **Piwi proteins (a subclass of the Argonaute family)** to silence transposable elements, particularly in germ cells, through both transcriptional and post-transcriptional silencing.

- play a crucial role in **protecting the integrity of the germline genome by silencing transposons and other repetitive elements**. They are essential for fertility and maintaining genomic stability in reproductive cells. In the germ-line, many **histone modifications are indeed erased,** releasing **transposons** from their “normal constraints

Arms race between cell's DNA and parasitic elements

- Even with our **defense mechanisms**, parasitic DNA (transposons, viruses, …) make up nearly **half of our DNA**
- **siRNA** and **piRNA** are surveillance systems (based on RNA-base pairing)
- Additional system with **sequence-specific DNA-binding proteins** (KRAB-ZFP proteins)
- They recognise viral or transposon **sequences**
- They recruit **histone writers** that place marks on the nearby histones
- They recruit **DNA methylases** that methylate the surrounding DNA
- They recognise sequences that are **crucial** for the virus replication or transposon transposition

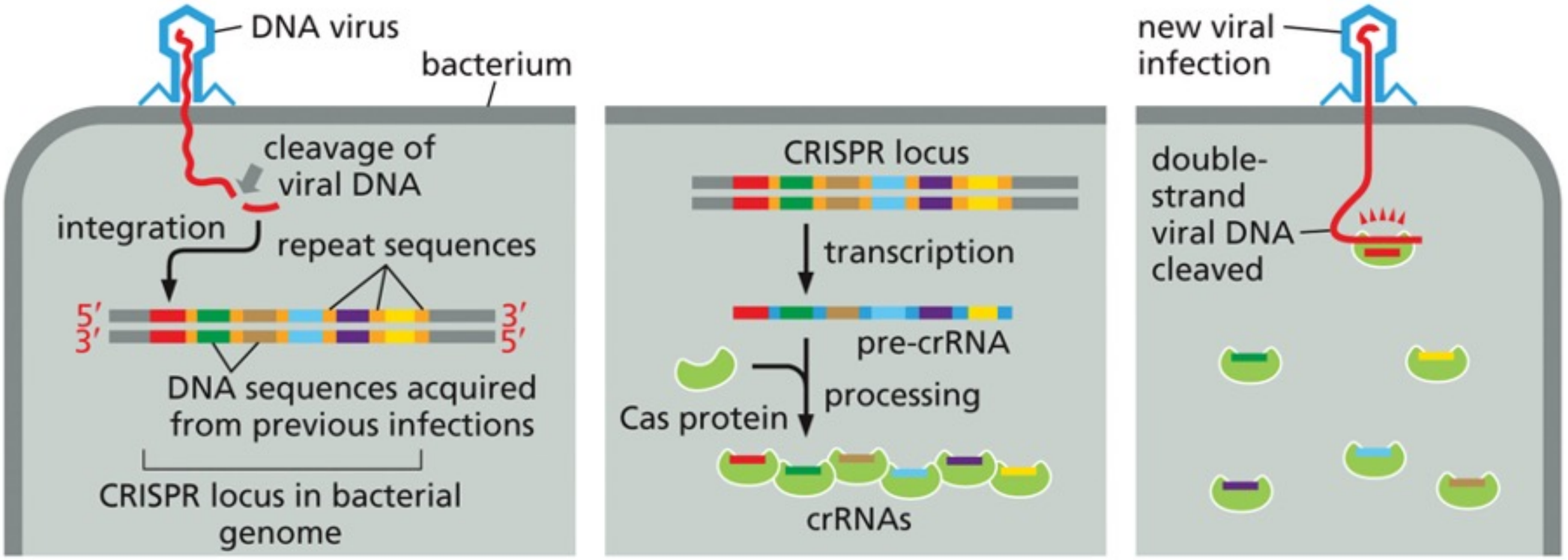
Recap: miRNA, siRNA and piRNA

Summary of Differences

Feature	miRNA	siRNA	piRNA
Source	Endogenous, genome-encoded	Endogenous or exogenous (e.g., viruses)	Endogenous, piRNA clusters
Length	~21-23 nucleotides	~20-25 nucleotides	~24-31 nucleotides
Structure	Imperfectly paired, stem-loop	Perfectly paired, double-stranded	Single-stranded, no defined structure
Processing	Dicer-dependent	Dicer-dependent	Dicer-independent
Protein Complex	Argonaute (AGO) family	Argonaute (AGO) family	Piwi subfamily of Argonaute
Function	Translational repression or degradation	mRNA cleavage	Transposon silencing
Primary Role	Gene regulation in development, differentiation	Antiviral defense, gene knockdown	Genome integrity in germ cells

Bacteria use CRISPR to protect themselves from viruses

- CRISPR = clustered regularly interspersed short palindromic repeat



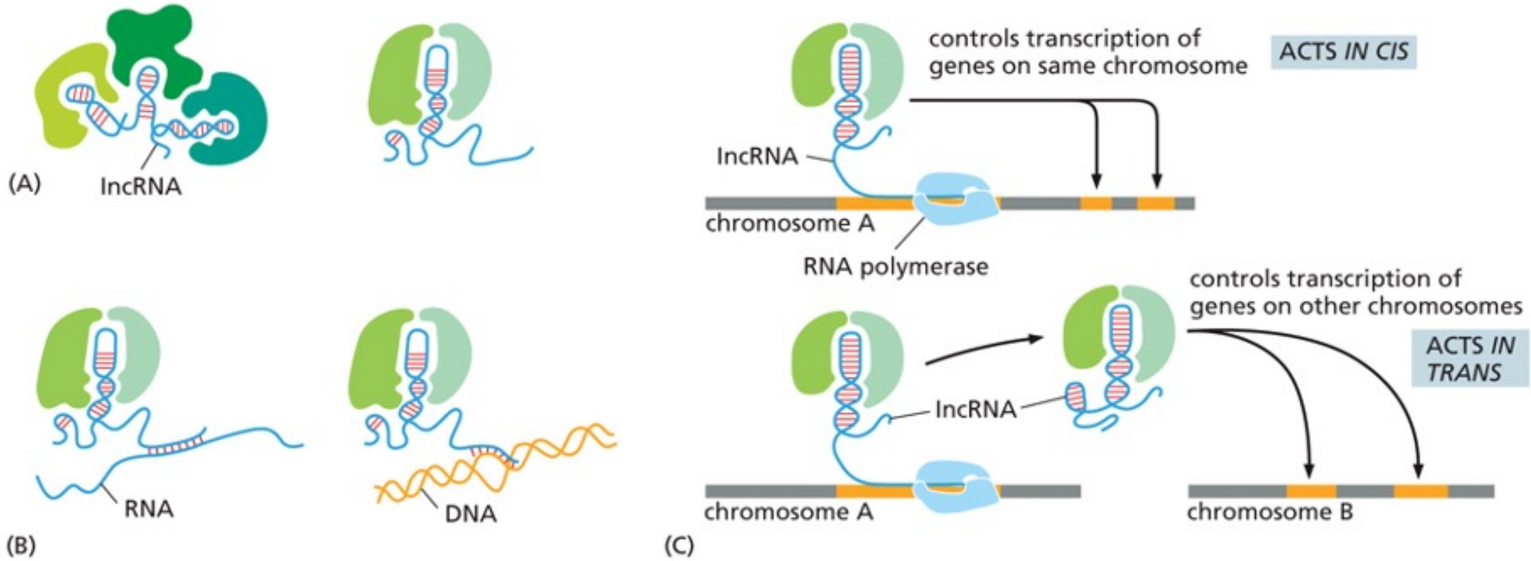
STEP 1: short viral DNA sequence is integrated into CRISPR locus

STEP 2: RNA is transcribed from CRISPR locus, processed, and bound to Cas protein

STEP 3: small crRNA in complex with Cas seeks out and destroys viral sequences

Long non-coding RNAs (lncRNA)

- >1000 nt
- function mostly unknown
- > 5000 in the human genome
- Examples: RNA in telomerase, Xist RNA, RNA involved in imprinting
- They can work as **scaffold RNA molecules, guide sequences to bind specific RNA/DNA**, or **affect transcription**



Have a nice day!