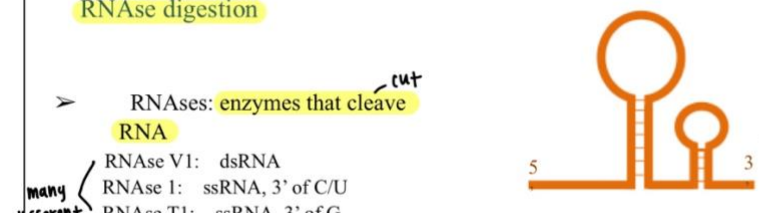
Exam 3: RNA notes

Class 1: The wonderful world of RNAs

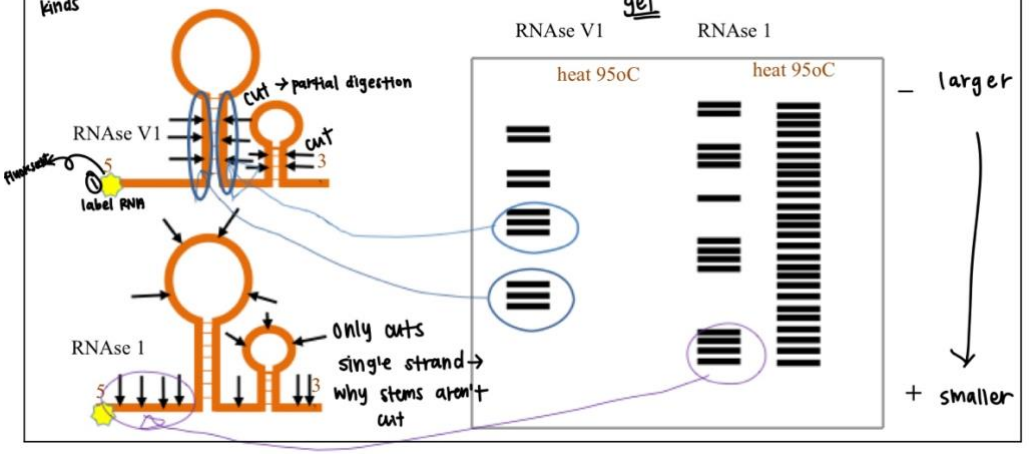
- Base pairs in RNA:
  - G-C: 3 hydrogen bonds → STRONGEST
  - A-U: 2 hydrogen bonds → WEAKER
  - G-U: 2 hydrogen bonds → WEAKEST
- Unusual bases in RNA
  - Post-transcriptional modification → different RNA forms arise post-transcript.
  - Most common in tRNAs but present in all RNA species
    - rRNA
    - non-coding RNAs
  - Pseudouridine → most common
  - Believed to stabilize RNA structure, but may affect functions
    - And interaction of RNA to keep the substrates together
  - May modify base pairing
- Non-Watson Creek base pairs
  - RNA 3D structures (tRNA)
  - Triple helix with DNA → happens in regulation of gene expression
  - Base pairing can serve to stabilize tertiary structures
  - Makes hydrogen bond in different ways
- RNA folding
  - Primary structures
    - Sequence
  - Secondary structure
    - Formed by canonical base-pairing (not really in 2D)
  - Tertiary structure
  - Preservation of the shape is very important → can get same shape with different sequences
  - In non-coding RNA, nucleotides don't have to be preserved very well – but shape must be maintained
  - SPECIFIC SEQUENCE AND 3D STRUCTURES ARE BOTH ESSENTIAL FOR RNA FUNCTION
- Representation of basic secondary structures in RNA
  - Secondary structures- easier to predict compared to tertiary
    - more stable than tertiary → minimizes free energy (ΔG)
    - ncRNA secondary structures→ more evolutionary conserved than sequences (STRUCTURES → FUNCTION)
    - in mRNA: lesser extent and less stable
  - the folding of secondary structure is based on minimizing the energy of each bond → each bond has lower energy especially at the stem → each H-bond lowers energy (significantly)
- Assay for RNA secondary structure: RNAse digestion, DMS-seq, SHAPE-seq,

Assay for RNA secondary structure: RNAse digestion

- RNAses: enzymes that cleave (cut) RNA
  - many different kinds
  - RNAse V1: dsRNA
  - RNAse 1: ssRNA, 3' of C/U
  - RNAse T1: ssRNA, 3' of G



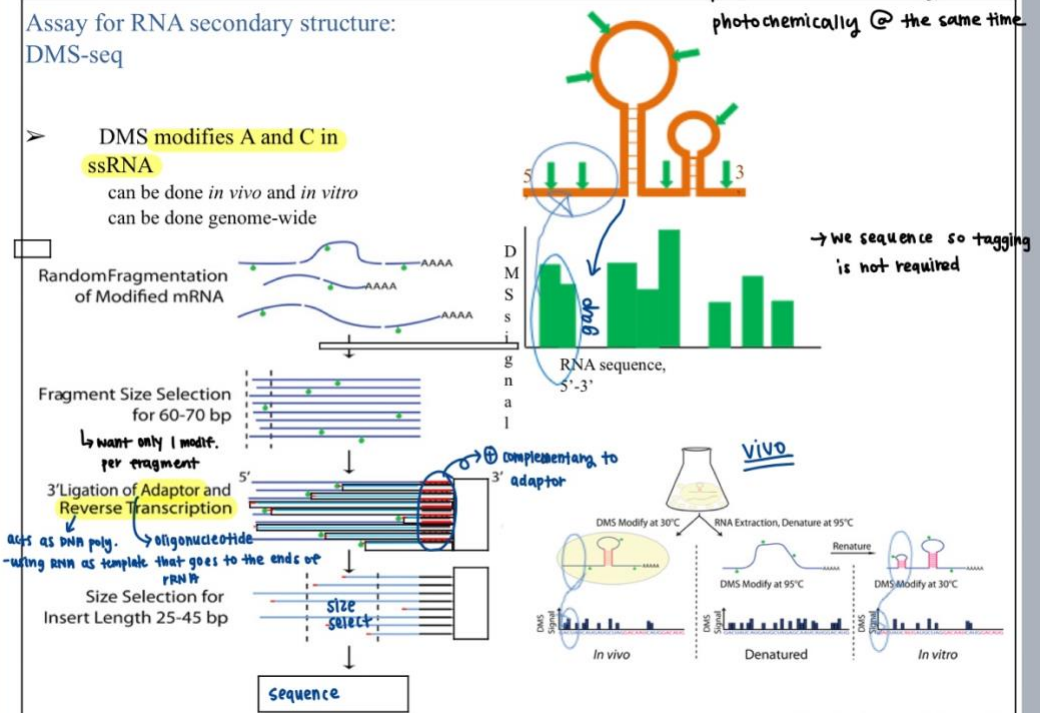
– know the structure contains stem but unsure about its sequence

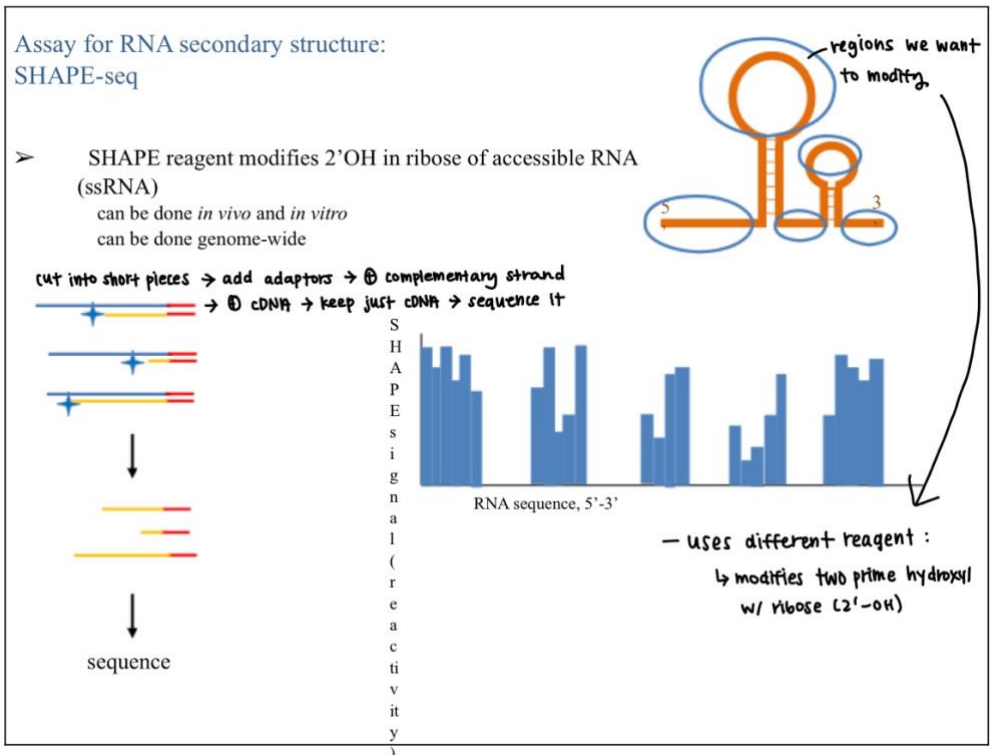


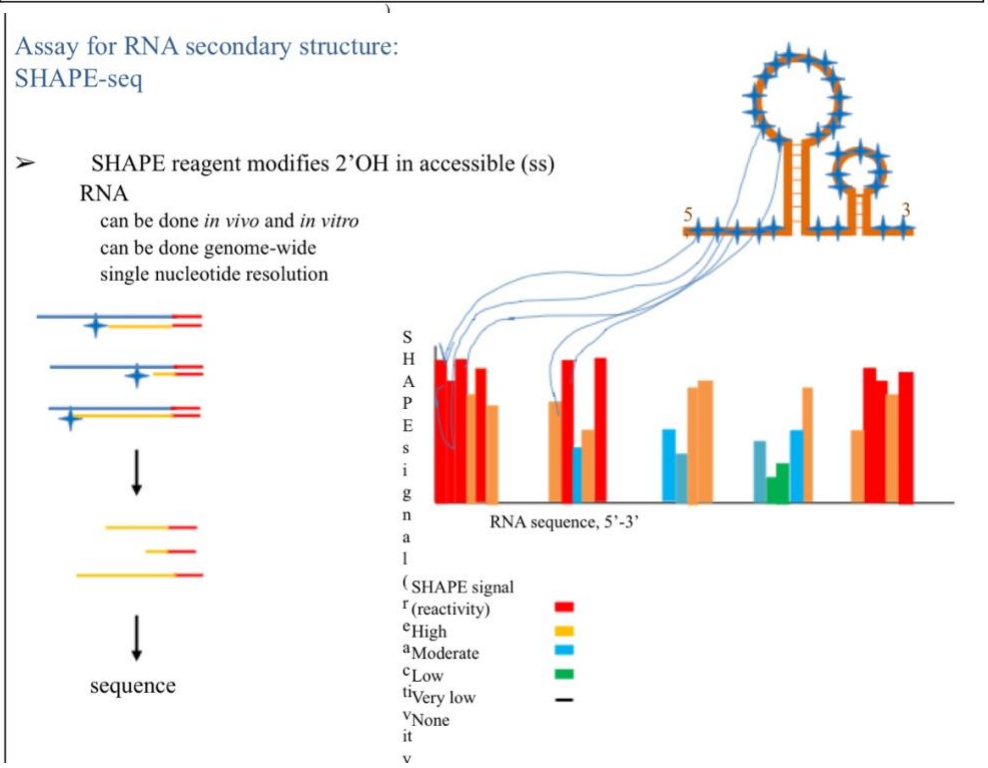
Assay for RNA secondary structure: DMS-seq

– performed in vivo entity or photochemically @ the same time

- DMS modifies A and C in ssRNA
  - can be done *in vivo* and *in vitro*
  - can be done genome-wide







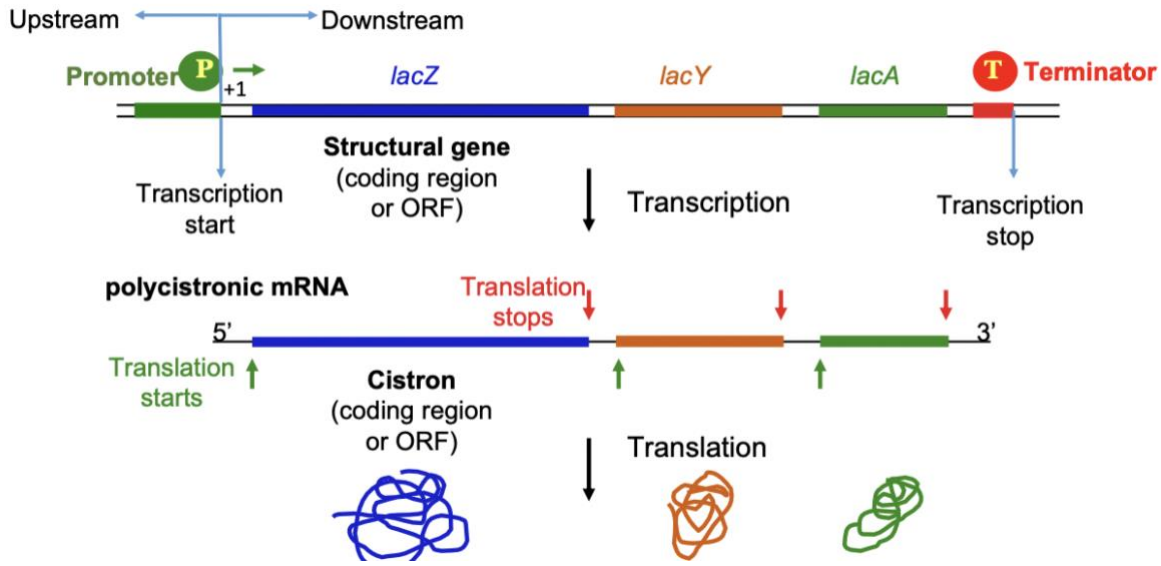
- Tertiary (3D) structures of RNA
  - RNA helix – wider than DNA
  - Stabilized by hydrogen bonds and stacking interactions (tertiary interactions)
    - Sometimes – metal ions (present in the middle to stabilize)
  - Base-pairing between distant nucleotides
  - Interactions between 2'OH groups
- RNA is a multifunctional molecule
  - >95% of RNA in the cell is NOT TRANSLATED into proteins
    - Non-mRNA RNAs
    - Non-coding proteins

- Signal Recognition Particle (SRP) couples with the synthesis of membrane proteins to their proper membrane localization (protein sorting RNA)
    - SRP delivers nascent membrane proteins to ER (eukaryotes) or plasma membrane (bacteria)
    - Eukaryotic SRP: composed of 7SL RNA and several proteins
- pRNA: drives the most powerful molecular motor in nature
    - pRNA: packing RNA – packs RNA into DNA into viral capsids
        - literally push the DNA of the viral genome into the capsule → VERY SMALL
- Ribozyme: some RNAs can catalyze chemical reactions
    - Peptide bond formation → rRNA (a.a in ribosome)
    - Phosphodiester bond cleavage → DNA or RNA backbone
    - RNA ligation
    - RNA phosphorylation
    - RNA aminoacylation
    - Glycosidic bond formation and others
- RNA catalyst: Hammerhead ribozyme
    - RNA that cuts other RNA
        - Hairpin ribozyme
        - Hepatitis Delta Virus Ribozyme
        - Varkud satellite ribozyme
        - RNAse P
        - Group I introns
        - Group II introns
- One sequence, two ribozyme
    - Ligase fold or HDV fold
        - Synthetic RNA molecule → capable to acquire 2 completely different secondary structures
            - Each structure performs different enzymatic activity → Ligase vs cleavage
            - Based on two different initial ribozymes with similar length
                - Ribozyme vs Riboswitch
- Riboswitch: sense environmental changes
    - Regulatory element → senses environmental change → switches from one to another conformation
    - Regulate gene expression → sense environmental change → stop of allow transcription
    - Mostly present in bacteria (Ex. Bacterial RNA thermometer)
        - At 30C: ribosomes cannot bind to RBS and AUG due to double strand or highly constructed RNA
        - At 42C: contains heat shock protein allows binding of ribosome to RBS and AUG site → allows translation of mRNA into protein
- RNAs that can catalyze chemical reactions or bind specific ligands have been selected in vitro
    - Aptamers: RNAs that bind specific ligands
        - Can fit like lock & key → like antibodies (advantages to antibodies)
    - Ribozymes: RNAs that catalyze specific chemical reactions
    - Synthetic ones (Diels Alder Ribozymes): very fast and cheap to make, but
        - usually less active
        - lower affinities
        - lower rate enhancement
            - But potential therapeutics/diagnostic applications
- The RNA World: RNA at the beginning of life

- What are the requirement for life?
  - Self-replicating RNA → replicase ribozyme (not found in nature can only be synthesized)
- RNA: only macromolecule that can store genetic information AND catalyze reactions
- Ribozyme: likely the first enzymes, protein enzymes took over later (more efficient)
- Similar for DNA with storing genetic information → probably by RNA then switched to DNA due to higher efficiency

Class 3: Transcription in Bacteria

- Transcription: the process in which RNA is synthesized on the DNA template
  - Performed by RNA polymerase
  - Similarity to DNA polymerase
    - New chain synthesis is guided by a template
    - 5' → 3' – adds single nucleotide to the new chain
  - Difference
    - Does NOT require primer
    - Not restricted to S- phase
    - Only template DNA strand is used in transcription
    - More error-prone
      - RNA pol is an enzyme (does not last)
      - DNA pol: has proofreading quantities (more preserved than RNA)
    - Different genes transcribed at different times with different efficiency
    - RNA transcription: happens all the time → 1 gene can make several mRNA



E. coli RNA polymerase is composed of five subunits

TABLE 24.1 Subunit composition of *E. coli* RNA polymerase

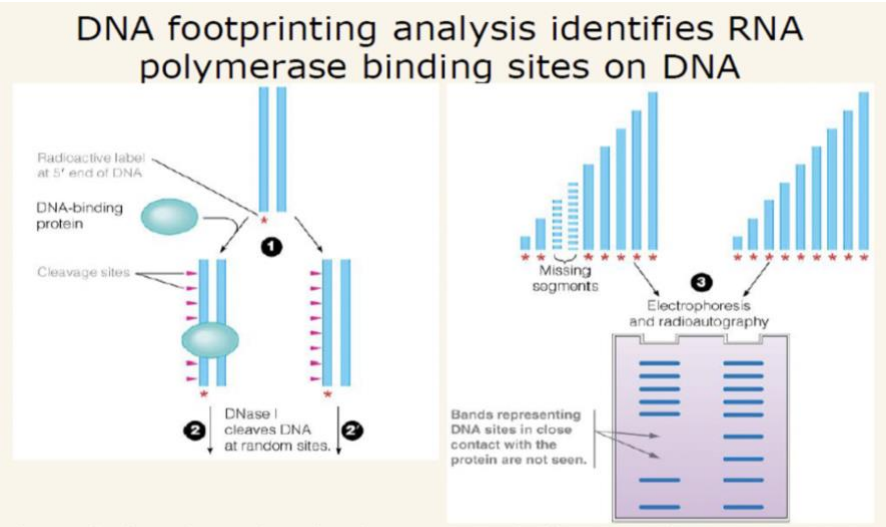
Subunit	M_r	Number per Enzyme Molecule	Function
α	36.5 kDa	2	Chain initiation, interaction with regulatory proteins and upstream promoter elements
β	151.0 kDa	1	Chain initiation and elongation
β'	155.0 kDa	1	DNA binding
σ	70.0 kDa[a]	1	Promoter recognition
ω	11.0 kDa	1	Promotion of enzyme assembly

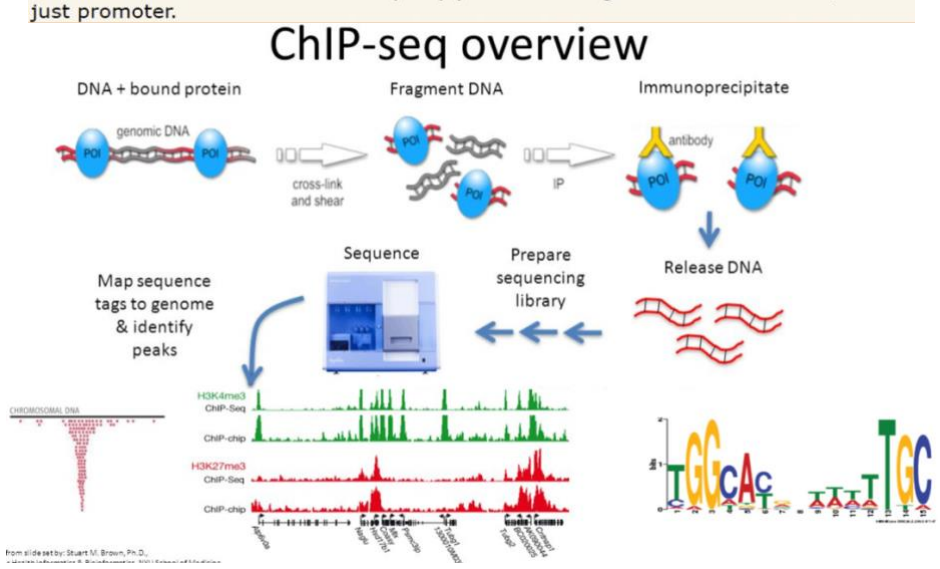
Different σ factors for different promoters.

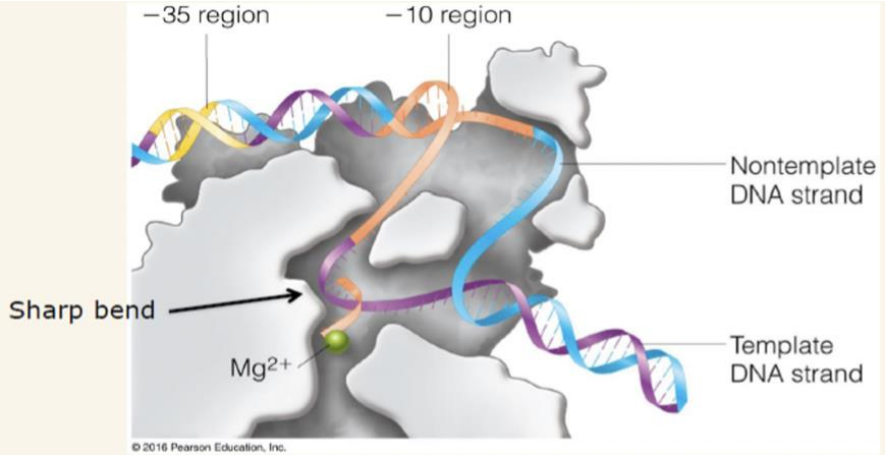
Structure of bacterial RNA polymerase

- Core enzyme (without sigma factor) has the ability to transcribe but it is NOT specific
  - Sigma factor: recognizes the promotor sequence and directs RNA polymerase to the promotor

- Three steps in RNA polymerization
    - 1. Initiation
        - RNA polymerase recognizes and binds the promoter (closed complex)
            - Holoenzyme binds to the promoter sequence in DNA
                - At -35 and -10 region (consensus sequencing sigma 70)
                    - Gaps between -35 and -10 is present due to the geometry of the protein
            - Sigma70: variation in promoter sequences affects the strength of promoter
                - Strongest promoters match consensus
            - Beneficial to have multiple sigma factors → allow different genes to be expressed @ different times
            - Deviation (good) → allow control of the expression → ↑ strength of the promoter → ↑ expression, ↑ # of copies
            - Some genes have upstream promoter (A-T rich)
                - A flexible C-terminal domain of one of the a-subunits binds the UP element → strengthens the promoter → present only in highly expressed genes







- RNA polymerase first forms a closed complex in which DNA is not unwound
- Next, the stable, open complex forms where the -10 region is unwound
- Lastly, an Mg^{2+} dependent isomerization further unwinds the DNA from -12 to +2

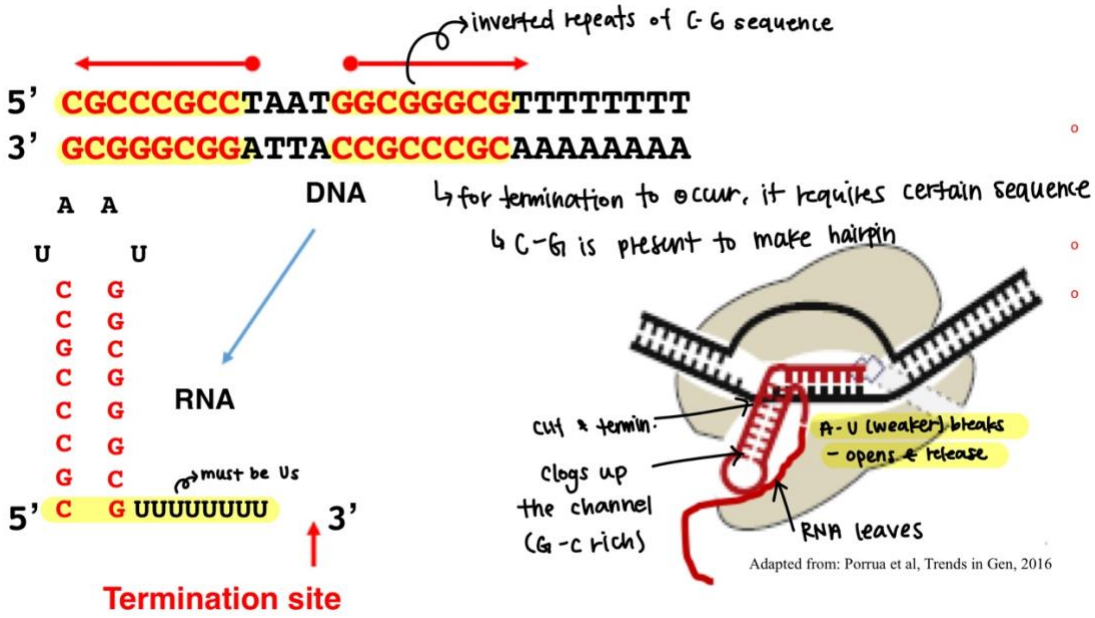
▪ RNA polymerase unwinds DNA [transcription bubble (open complex)]
▪ First phosphodiester bonds between rNTPs (unstable ternary complex)
▪ Release of sigma factor [after ~10 nucleotide (stable ternary complex)]
▪ First rounds of polymerization are non-productive
- Abortive initiation
  - Sigma factor partially blocks the RNA exit channel – needs to be displaced → unstable ternary complex
  - Short strands of RNA (~10nt) are synthesized and released repeatedly before productive transcription start
  - Promoter escape (formation of stable ternary complex)
- STRONG PROMOTER: good binding w/ polymerase + pairs

○ 2. Elongation
▪ Chain elongation continues in 5' → 3' direction
▪ E. coli elongation rate in vivo: 40-80 nt/s (slower than DNA pol.)
▪ Positive supercoils form ahead and negative supercoils behind the transcription bubble (resolved by topoisomerase)
▪ Elongation: processive – RNA polymerase does not come off the DNA until termination
- Ternary complex of RNAP + DNA + RNA is very stable
- Not a smooth process
- RNA pol. Usually pause due to:
  - RNA secondary structure
  - Difficult sequences
  - Backtracking
  - Limiting NTP
- RNAP: has proofreading activity but more error prone than DNA Pol I
  - Proofreading activity is little → not as much as DNA pol

▪ Transcription in bacteria → coupled with translation (no nucleus) → translation starts on nascent mRNA before transcription is finished
- As transcription happens, ribosome grabs the end and starts translation

○ 3. Termination
▪ Transcription stops after terminator sequence
▪ RNA transcript is released, and RNA polymerase detaches from DNA rho-dependent termination

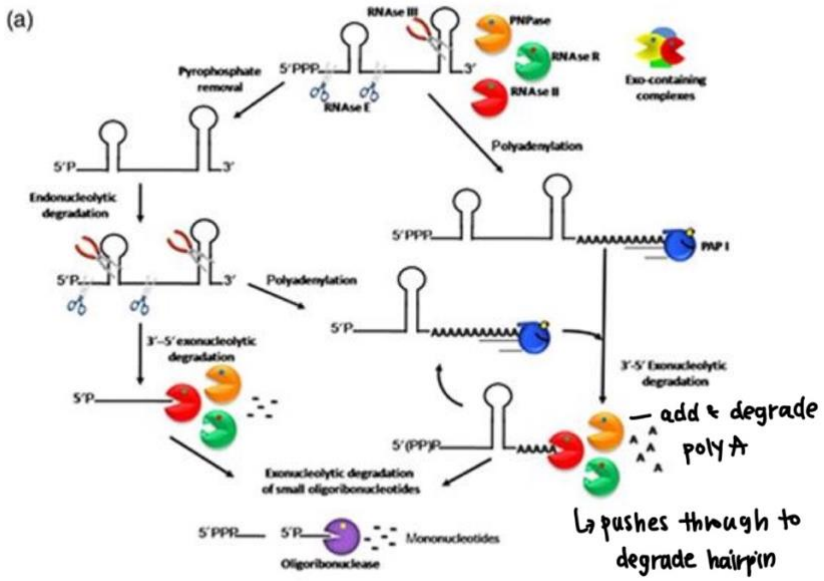
- **Intrinsic (rho-independent) termination → does not require ATP**
  - DNA template contains GC-rich inverted repeats followed by ~8A in the template strand
  - transcribed RNA forms a hairpin loop (strong GC bonds) that stalls polymerase
    - hairpin destabilizes 5' end of hybrid & RNA-RNAP contact
    - Us destabilize hybrid
  - RNA:DNA duplex (all AU pairs) weakens, RNA transcript is released and transcription stops



- **Rho-Dependent termination of transcription → requires ATP**
  - Rho-factor: ATP-dependent helicase that unwinds DNA:RNA hybrid
  - Rho binds to a specific site on nascent RNA
  - C- rich Rho-factor moves 5'-3' on the transcript hydrolyzing ATP
  - Stalling (when encountered a hairpin (generally) or GC rich sequence) of the polymerase facilitates Rho termination of transcription

- After termination: post-transcriptional RNA processing in bacteria
  - b/c transcription is coupled with translation, post-transcriptional modifications do not happen due to instantaneous binding of the ribosome
  - some modifications may be more prevalent than was initially suggested (polyadenylation)
  - Poly A: facilitates RNA degradation in bacteria
    - In eukaryotes increases stability
  - rRNA transcripts are processed into mature forms (cleavage, base modification)
  - ALL rRNAs are derived from a single precursor in prokaryotes
- Stability and degradation of bacterial RNAs
  - Bacterial mRNA = unstable, half-life in minutes
  - Degraded by endonuclease (cut in the middle) – RNAse E, RNAse III and exonuclease (cuts ends) – RNAse II, PNPase
  - RNA degradosome (complex of RNAses and RNA helicase
    - Degradosome contains several endonuclease and exonuclease
    - Helicase has secondary structures have duplex which unwinds to allow cutting of RNA → ↑ degradation
  - Exonuclease: cannot degrade secondary structure due to presence of hairpin structure made during the termination process

- Hairpin: may also protect 3' end
  - Strong hairpin bonding → intrinsic termination
  - Weak hairpin bonding → Rho-dependent termination



Arraiano et al., FEMS 2010

Assay for relative RNA abundance:
(q)RT-PCR

RT-PCR (reverse transcriptase-PCR)

Most common way to assay RNA abundance (qualitative) of a limited number of RNAs

low throughput – one RNA at a time

Sensitive – can detect just a few copies of RNA
- can be done from a single cell

- can quantitate the differences
- very sensitive — single cell detect. possible

Based on the PCR (polymerase chain reaction) technique
- commonly used
- quick check to see if RNA is present or not

How does PCR (polymerase chain reaction) work?

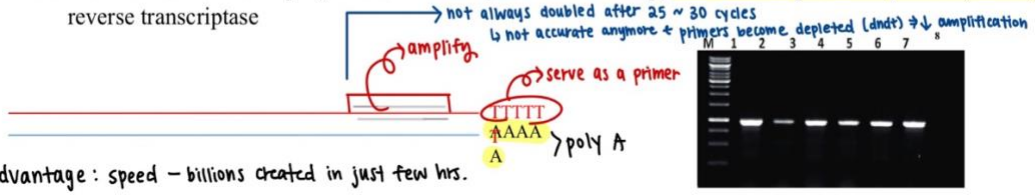
Polymerase Chain Reaction (bio-rad.com)

→ a process used to make copies (amplify) fragments of DNA

Template for PCR amplification is DNA

Hallmark of PCR → convenient method to seperate DNA strands w/o extra enzymes
- special heat-resistant DNA polymerase
  ↳ allow PCR samples to be heated w/o destroying DNA polymerase

- To be able to use PCR to amplify and detect RNA, we must first copy RNA to complementary DNA (cDNA)
- reverse transcriptase



> advantage : speed – billions created in just few hrs.

PCR cycle

1. FIRST Cycle

   Denature Cycle (95°C)

   - ↑ temp → DNA strand separate (Denaturing)

   Anneal primers (60°C)

   - ↓ temp to allow primers (short ssDNA molecules) to attach – (anneal) to the strands of DNA

   Extend Primers (95°C)

   - ↑ temp → optimum environment for special heat-resistant type of DNA polymerase ⇒ "*Thermus aquaticus*" or Taq polymerase
   - DNA polymerase extends the primers by ⊕ complementary nucleotides

2. Repeats until 5th cycle
3. After 5 cycles, 22 copies of target DNA sequence has been produced
   - about 30 cycles are used to produce enough copies
     ↳ how many copies would that be?
     ↳ check by gel (agarose)

☆ poly A tail in bacterial mRNA ↑ rate of mRNA degradation

Biotechnology Explorer — BIO-RAD

Assay for relative RNA abundance:

(q)RT-PCR
↳ quantitive - RT-PCR

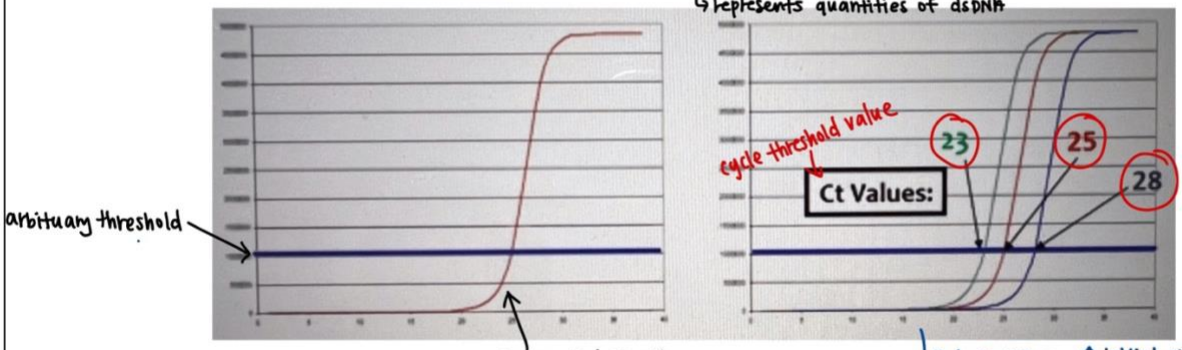
- similar to RNA seq, difficult to determine the quantities it began with but can compare quantities between control & sample (treated)

qRT-PCR (real-time RT-PCR)

Quantitative – (usually) relative quantification (normalized to housekeeping RNA)
↳ normalization required

Monitoring DNA amplification in real time, cycle by cycle → unlike RT-PCR (where 25-30 cycles are required for comparing)

one of the methods: intercalating fluorescent dye (more dsDNA, more fluorescence)
↳ represents quantities of dsDNA



↳ ↓ Ct value, ↑ initial starting value
↳ ↑ initial value, crosses the threshold faster

Record when the sample trace crosses arbitrary threshold value

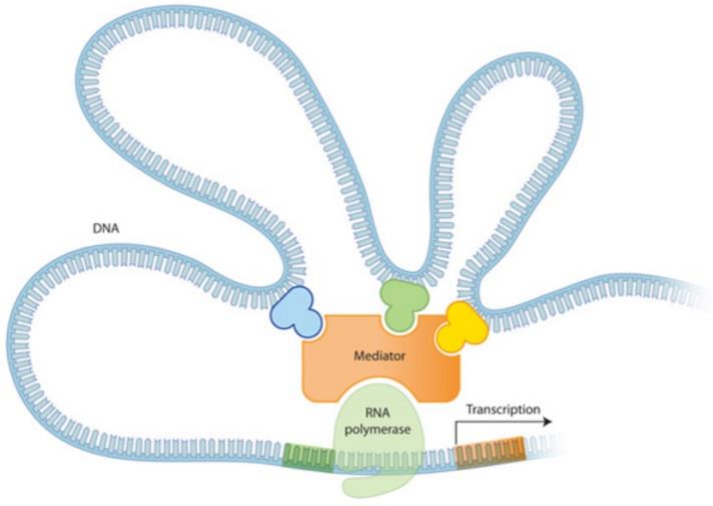
During exponential phase, product doubles every cycle (DNA amount ~ 2cycle number)

Starting quantity is proportional to Ct value (quantity = 2^{Ct})
↳ 2^{Ct} ⇒ exponential growth

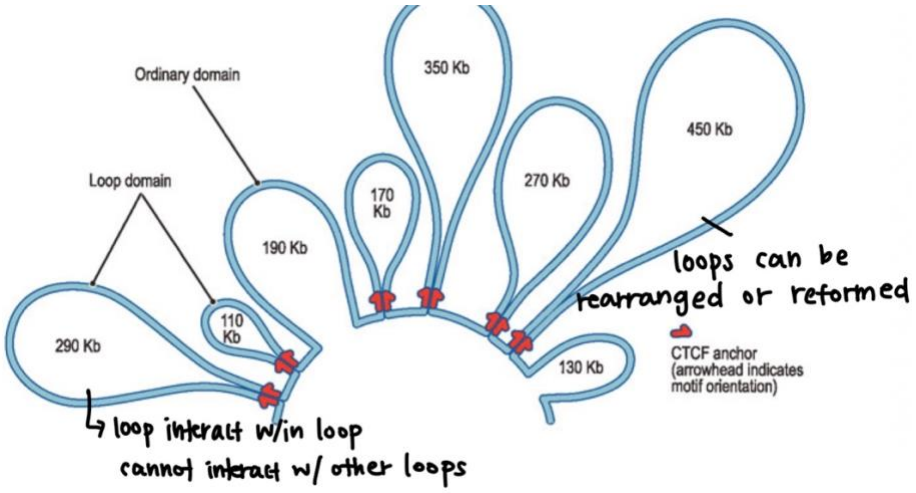
Class 4: Eukaryotic Transcription

RNA Polymerase	Types of genes transcribed	No. of different transcripts	% of total RNA
RNA Polymerase I	5.8S, 18S, 28S, rRNA	1	~80%
RNA Polymerase II (most diverse)	Protein-coding genes, lncRNA, snoRNA, miRNA, siRNA, most snRNA (all mRNA)	>20000	1-5%
RNA Polymerase III	tRNA, 5S rRNA, some snRNA, other small RNA	~100	~15%

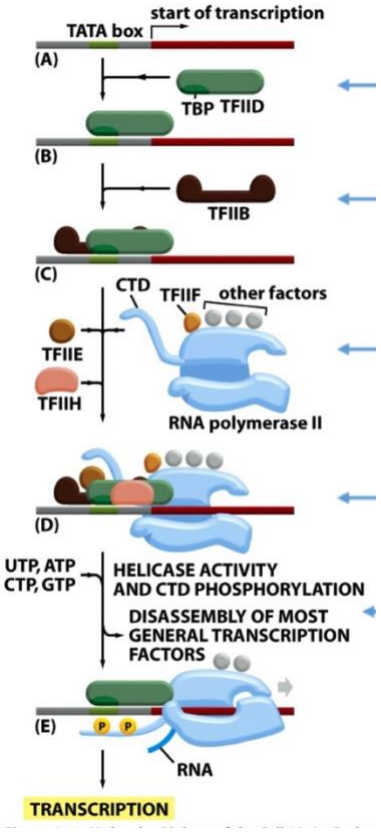
- Mitochondrial and chloroplast RNA polymerases are similar to bacterial
  - Only 1 RNA polymerase for bacteria
- RNA Polymerase II has **12 subunits**
  - Shares structural homology with bacterial RNA polymerase → similar structure, similar function
  - 5 essential transcription factors → essential for initiation of transcription
    - 2 most important
      - TFIID: like sigma factor → recognize TATA box b/c it contains TBP
      - TFIIH: helicase activity (unwinds DNA) phosphorylates Ser5 of RNA Pol CTD → P- of CTD releases RNA pol from the promoter
- Pol II promoter sequences
  - Core promoter (minimal promoter): ~40bp around TSS (transcription start site)
  - Proximal promoter: ~200-400 bp from TSS
  - Distal promoter: kbs from TSS → also an enhancer
- Distal promoters: regulatory regions several kbp upstream → also called enhancers
  - Activator proteins bind to enhancers
  - DNA looping brings in contact → activators and basal transcription factors through mediator



- Enhancer/promoter loops
  - Enhancers and promoters that interact in transcription may be thousands of bp apart
  - Large DNA loops encompass genes and their regulatory regions
  - Insulators establish chromatin boundaries
    - Insulators: red things in the picture



- RNA polymerization
  - Initiation: general mechanisms similar to prokaryotic initiation
    - more complex
    - basal (general) transcription factors required
    - transcriptional activators, mediator, chromatin modifying proteins also needed



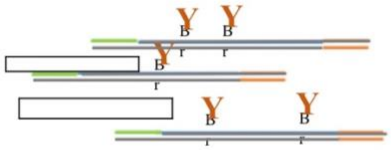
- TFIID containing TBP: binds to the TATA box
- TFIIB: recruited to the TBP-TATA box complex
- TFIID + TFIIB: recruits RNA polymerase with TFIIF to the promoter
- TFIIE and TFIIH: then joins the complex last.
  - Pre-initiation complex (closed complex) is complete.
  - TFIIH: helicase activity → unwinds DNA and phosphorylates CTD of Rpb1
    - Stimulates release of the mediator
- General transcription factors are released from the complex when transcription starts → elongation factors join
- TFIID: TBP binding to the promoter induces a sharp bend in the DNA
  - Important for promoter selection
-

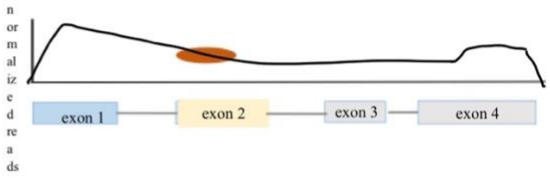
  - Elongation
    - General mechanism similar to prokaryotes
    - Involves elongation factors
    - Pol II proximal pausing
    - As nascent mRNA transcript emerges, 7mGTP cap is added to the 5' end
    - Splice junctions are marked by binding of splice factors
    - Splicing occurs co-transcriptionally

Assay for rate of transcription:
GRO-seq (Global Run-On sequencing) > does tell us the rate of transcript.

Assay for RNA pol that are actively transcribing
Strand-specific information
Isolate nuclei or permeabilize cells → all nucleotide leaves
Replace UTP with BrUTP (transcription continues)
↳ labels
- then isolate the RNA
↳ to check the rate of transcript.

actual results





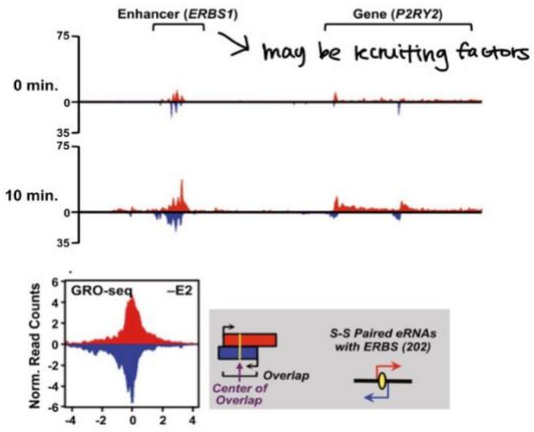
antibodies bind to modified RNA → remove antibodies → ⊕ adaptors for primers → ⊕ primers (on opposite side) → RT-RNA → double stranded RNA

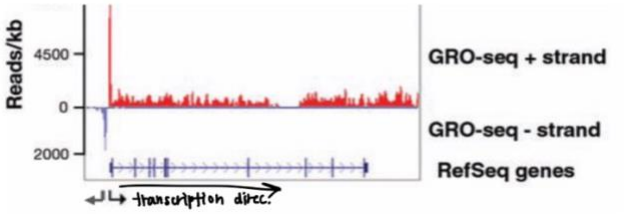
data from Wang et al, Sci Rep, 201

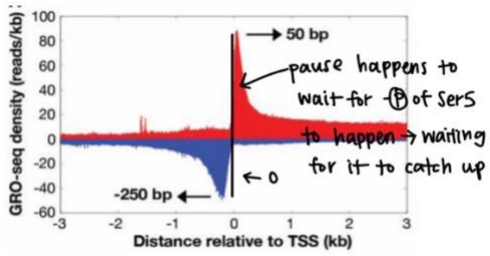
The surprising results of GRO-seq

→ genome wide ; does not need to choose a promotor
→ measure rate of transcript.

- For many transcripts, Pol 2 pauses after starting → waiting for Ⓟ-of Ser5 to reach CJ. transcrip. (promoter-proximal pausing)
- transcrip. in eukaryotes is bidirectional for many (maybe all?)







- enhancers are actively transcribed, in both direction (non-coding eRNA transcript.) — w/o enhancers, transcrip. does not happen quickly(?)

data from: Wang et al, Sci Rep, 2014; Core LJ at al, Science, 2008, Hah et al, Gene Res, 2015

- Termination
  - Involves termination factors
  - Specific CTD phosphorylation pattern
  - Polyadenylation of pre-mRNAs
- Termination by Pol II
  - Pol II usually continues transcription long after passing one or more polyadenylation signals
  - Pre-mRNA is cleaved 11-30 nt after the polyadenylation signal
  - Poly A polymerase adds poly A tail (up to 300 residue)
  - Longer poly A tail, longer half-life of mRNA
  - 3 models in termination
    - Allosteric model: binding of 3' processing factors leads to rearrangement of elongation complex and termination

- Torpedo model: nuclease (exonuclease) degrades nascent RNA from 5’ end, catches up with the elongation complex and displaces Pol II from DNA
- Combination model: both allosteric and torpedo model

RNA Polymerase	**Localization**	**Sensitivity to a-Amanitin**	**Types of genes transcribed**	**Promoter**
RNA polymerase I	Nucleolus	Insensitive	5.8S, 18S, 28S RNA	Bipartitie
RNA Polymerase II	Nucleus	Highly sensitive	Protein-coding genes, lncRNA, snoRNA, miRNA, siRNA, most snRNA	Mostly upstream, some elements in introns
RNA Polymerase III	Nucleus	Intermediate	tRNA, 5S rRNA, some snRNA, other small RNA	Internal

RNA Polymerase	**RNA Polymerase I**	**RNA polymerase II**	**RNA polymerase III**
Core transcription factors	UBF, SL1	TFIID, TFIIB, TFIIF, TFIIE, TFIIH	TFIIIA (only for 5S), TFIIIB, TFIIIC

Class 5: RNA processing and Degradation (PBL part 2)

- Eukaryotic mRNAs must be processed after transcription
  - Many modifications are performed co-transcriptionally:
    - 5' cap → happens as soon as it leaves form RNA polymerase
      - Without 5' cap, 5' gets exonucleased
    - Splicing (starts co-transcriptionally)
      - They either get spliced or labeled co-transcriptionally
    - Polyadenylation → protects degradation by 3' exonuclease on 3' site
    - Co-transcriptional modifications are facilitated by CTD of Pol II
- The CTD cycle
  - CTD of Rpb1 contains 15 to 50+ heptapeptide repeats: YSPTSPS
  - Phosphorylation patterns of CTD facilitate timely modifications of nascent transcripts by sequential recruitment of processing factors
    - Different -P patterns recruit different processing factors
  - S5-P: recruits capping and splicing factors to CTD
    - Phosphorylated by TFIIH
    - Phosphorylation of S5 releases the mediator → to start back from the beginning, all repeats have to be un-phosphorylated (mediator comes to bind to RNAP II)
  - S2-P: recruits polyadenylation and termination factors to CTD
  - S7-P: important for transcription of snRNAs

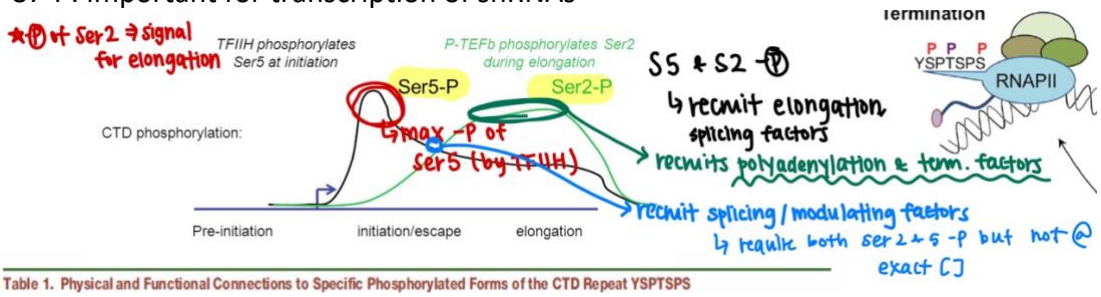
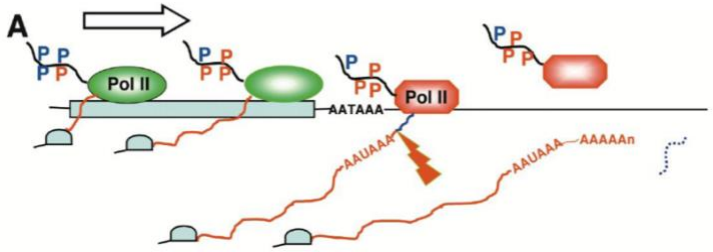


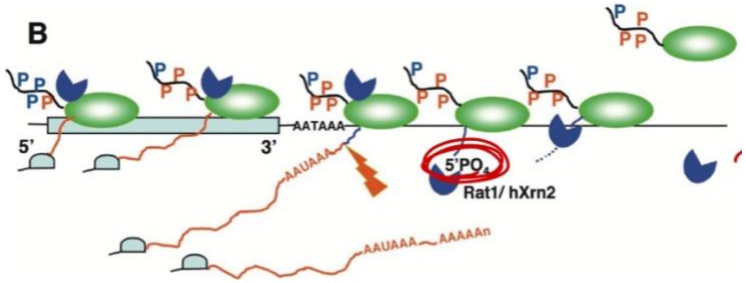
Table 1. Physical and Functional Connections to Specific Phosphorylated Forms of the CTD Repeat YSPTSPS

    - The concentration of phosphorylation of S5 and S2 is important. Sometimes, the ribosome halts to wait for elongation (S2 concentration) to reach its point to start elongation.
- 7-methylG cap is added to the 5' end of all Pol II transcript → 1st modification
  - Catalyzed by guanylyl transferase and methyl transferase
  - 5'-5' phosphodiesterase linkage between the 7mG cap and the first nucleotide (cap-0- present in all Pol II transcript)
  - In some RNAs the ribose of the first and/or second nucleotide is methylated (cap1, cap2 – increases the half-life of mRNA)
  - 7mG cap helps with initiation of translation
  - Increases RNA stability (protection from 5' exonuclease)
  - Facilitates splicing of the first intron → helps with initiation of transcription
- Almost all Pol II transcripts are polyadenylated at the 3' end
  - Protects 3' end from 3' end exonuclease except: all snRNAs, some mRNAs (histones) – no poly A tail
  - PAB: no access to 3' end by 3' exonucleases (increases half-life of mRNA)
  - Helps with initiation of translation
  - Increases mRNA stability (longer tail, increased stability)
  - Facilitates splicing
  - Facilitate termination of transcription
  - Pol II usually continues transcription long after passing one or more polyadenylation signals

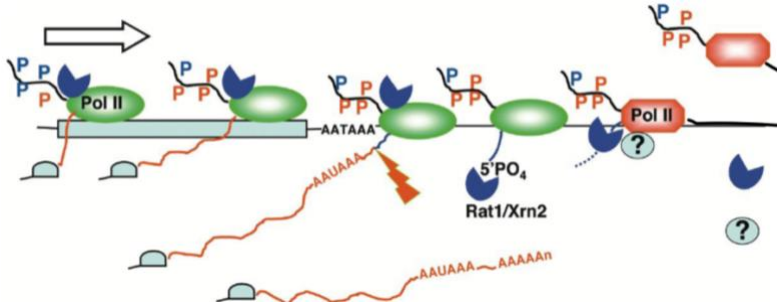
- 3 models of Pol II transcription termination for eukaryotes
    - Allosteric model: binding of 3' processing factors leads to rearrangement of elongation complex and termination



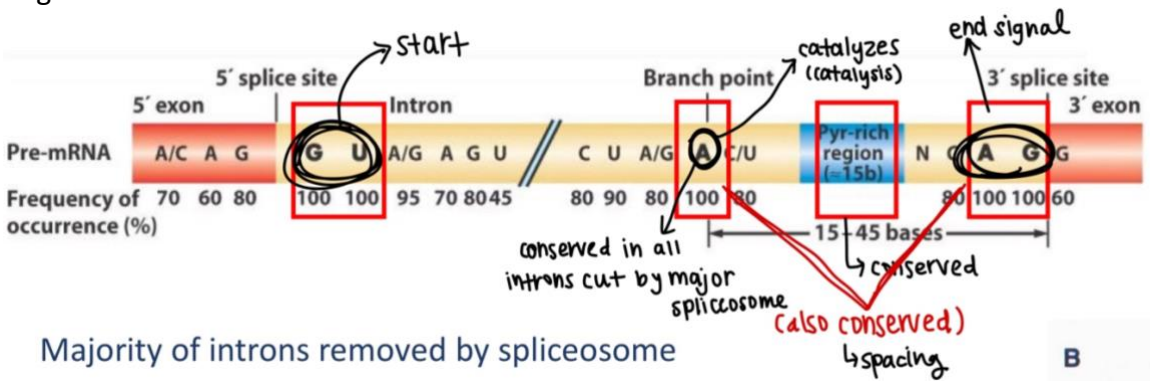
    - Torpedo model: exonuclease degrades nascent RNA from 5' end, catches up with the elongation complex, and displaces Pol II from DNA



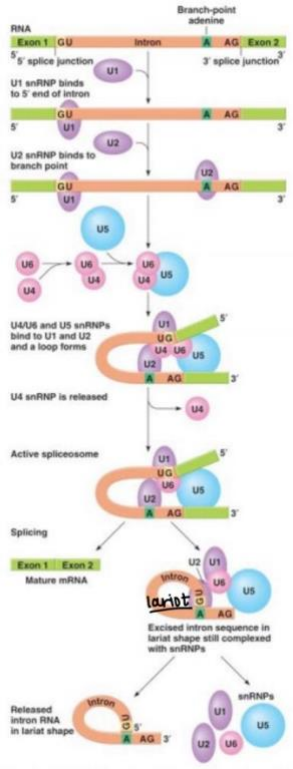
    - Combination Model: exonuclease → cause conformation change → release Pol II from DNA

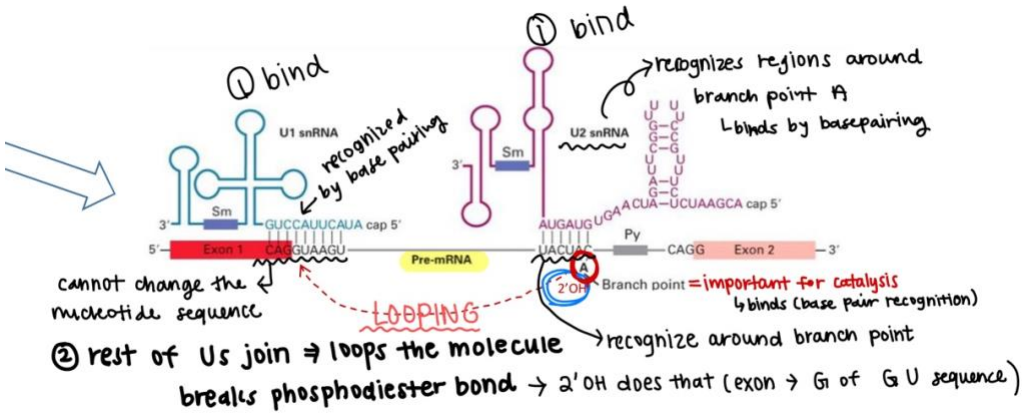


- BOTH 7mG cap and polyadenylation promote translation initiation and increase mRNA stability
- Splicing Site



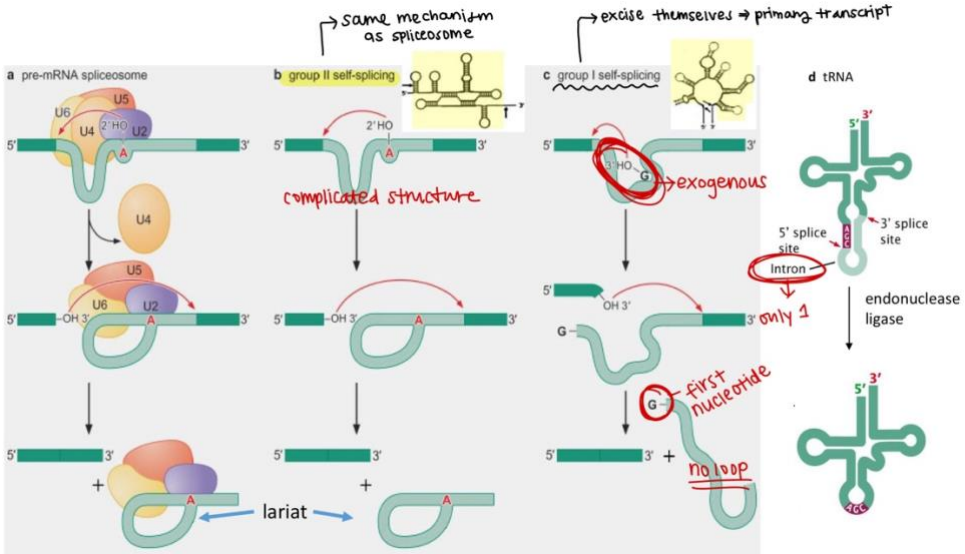
    - Major spliceosome: U1, U2, U4, U5, U6
        - Intron boundaries: GU (5') and AG (3')
    - Minor spliceosome: U11, U12, U4atac, U5, U6atac
        - Intron boundaries: AU (5') and. AC (3')
    - Exon/intron boundaries are marked co-transcriptionally (CTD – splicing factors Rpb-1)





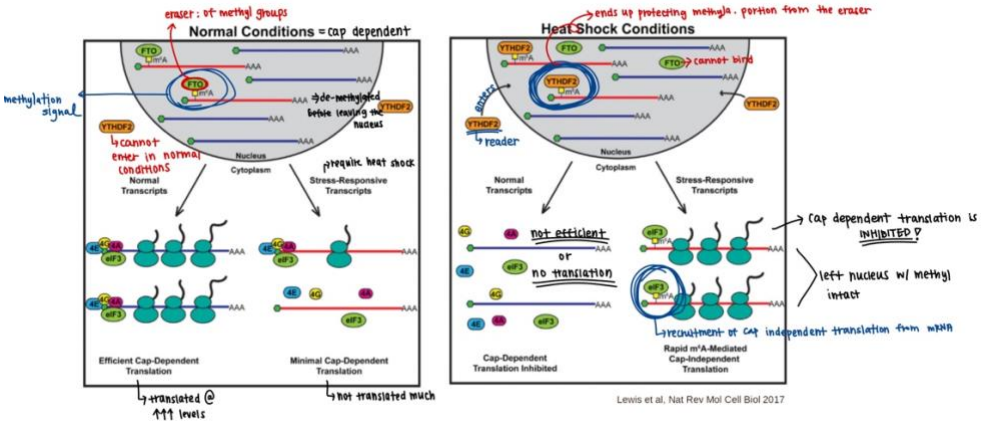
- Splicing sites are recognized by base-pairing with snRNAs
- Catalysis performed by U6 RNA (rybozime)
- After splicing, Exon Junction Complexes (EJC) mark the spliced sites → shows where introns were → important for quality control of mRNA surveilence
- THE SEQUENCES THAT WERE SPLICED WERE LABELED CO- TRASCRNIPTIONALLY THEN SPLICED

- Some introns are spliced using different mechanisms

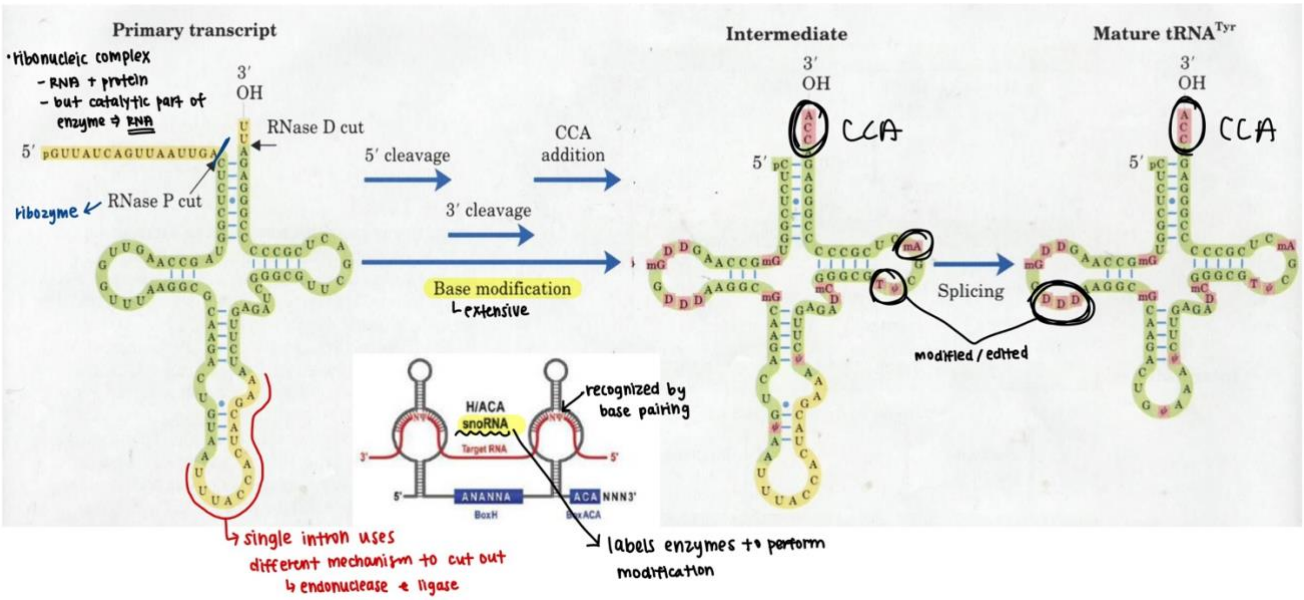


Class	Abundance	Mechanism	Catalytic Machinery
Nuclear pre-mRNA	Very common, used for most eukaryotic genes	Two transesterification reactions; branch site A	Major spliceosome; minor spliceosome
Group II introns	Rare, some eukaryotic genes from organelles, and prokaryotes	Two transesterification reactions, branch site A	Self-splicing, ribozyme
Group I introns	Rare, nuclear rRNA in some eukaryotes, organelle genes, few prokaryotic genes	Two transesterification reactions; exogenous G	Self-splicing ribozyme
tRNA introns → use different mech.	tRNA ONLY	Endonuclease, ligase; sometimes in cytoplasm	Endonuclease, ligase

- Alternative splicing forms – multiple mature RNAs derived from same pre-mRNA by alternative splicing
    - >60% mammalian genes have alternative splicing → also can be tissue specific – specific to each phase during development
    - Exons code for functional domain → faster evolution of new function
- RNA editing: post-transcriptionally
    - Base editing: alter RNA sequence or change protein sequence
        - Mechanism is similar in noncoding RNA, tRNA, rRNA
        - most common in tRNA, less in rRNA, present in mRNA (prevalently high)
    - A to I editing
        - ADARs – enzymes that perform A to I editing
        - Synaptic transmission (glutamate, serotonin) → in nervous system
            - Only modify A when it is a part of a stem and double stranded
        - miRNA processing
            - modifies @ the stem to get mature miRNA
            - processing is essential A-I editing
        - important for development (embryonic lethal when ADARs function is lost)
        - neuronal development on mRNA in the nervous system → mRNA codes for synaptic transmission receptor
        - important for innate immunity → important for cell differentiation
    - C to U editing
        - APOBECs (apolipoprotein, HIV) → enzymes that does C – U editing
        - Most studied → C to U edit in intestine has different function than the one in liver → intestine edit absorb cholesterol from gut
    - Uracil to pseudouracil (most common uncommon RNA)
    - mRNA methylation
        - transient (reversible)
        - m6A: most common site
        - readers (recognize and bind to the modified base) → recruit factors who carry effect of the modification
        - writers (modify the bases by methylating)
        - erasers (reverse the modification which is transient)
        - effects:
            - increase/decrease mRNA stability
            - increase/decrease translation → also modify to make it cap dependent vs. independent or inhibit translation
            - modulate splicing
        - stress-responsive translation



- tRNAs are processed by part by RNP ribozyme (RNase P → cut 5' end)

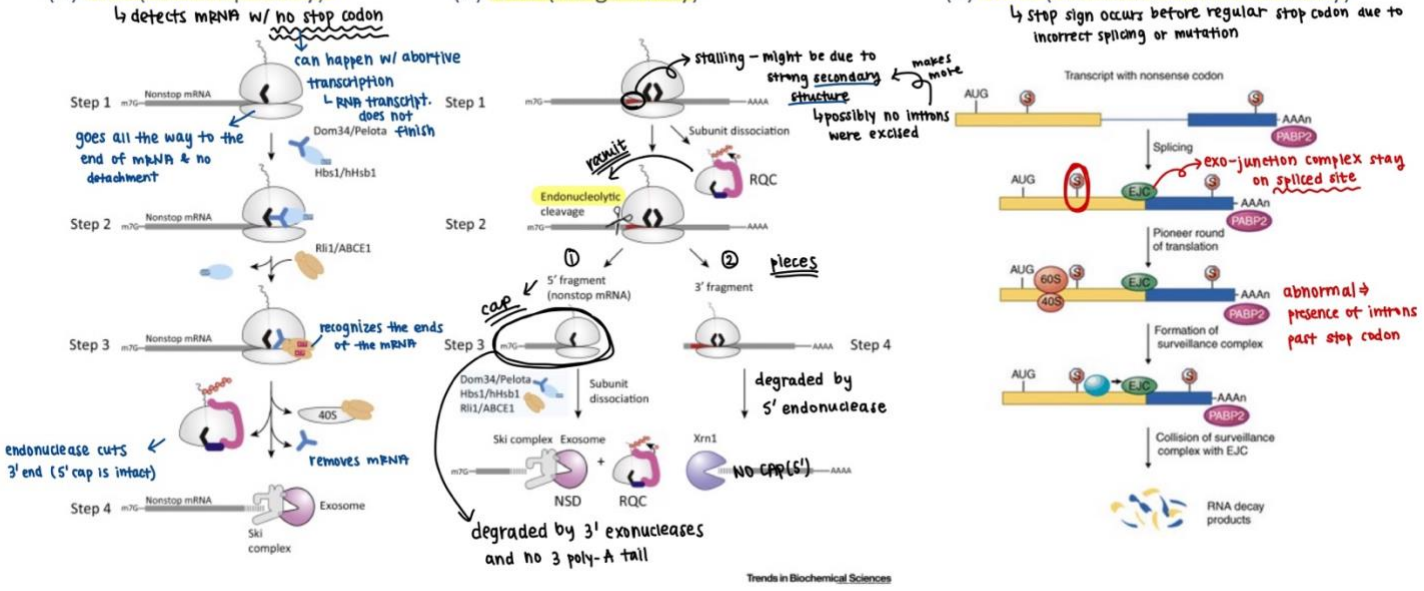


  - mature eukaryotic mRNAs are exported form the nucleus through the nuclear pore → pore regulates by increase/decreasing efficiency
- mRNA quality control

after being transported through the pores in to the cytosol, mRNA goes under 1st round of translation
like proofreading – detect if mRNA is fit for translation

mRNA quality control: *nonsense, nonstop, no-go*

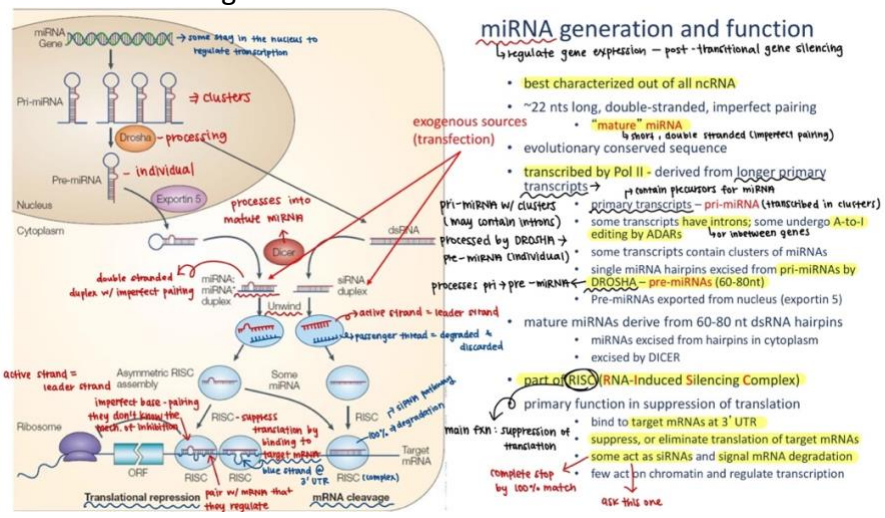
mRNA surveillance



- there are many pathways for RNA degradation
  - similar to bacteria
  - exosome contains RNAses (cuts RNA)
  - decapping and deadenylation exposes 5' and 3' end to exonucleases
  - degradation by endonucleases

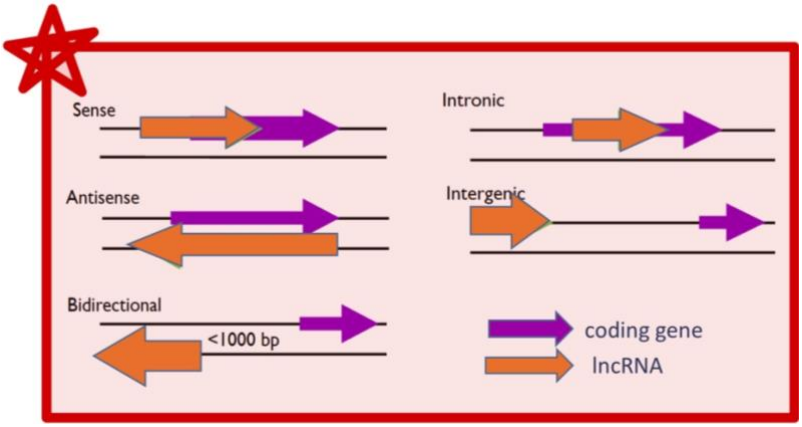
Class 6: ncRNAs – Dark Matter of the Genome

- Non-coding RNA: RNA that is not translated into protein
  - >95% of RNA in the cell is NOT translated into proteins → NON-coding RNA
  - Small vs long ncRNAs: Long (>200nt)
- Housekeeping: expressed at a higher level, higher quantity, essential for cell survival
  - rRNA (ribosomal)
  - tRNA (transfer)
  - snRNA (small nuclear)
  - snoRNA (small nucleolar) → DNA rep.
  - RNAse P
  - 7SL RNA
- Regulatory (gene expression):
  - miRNA, siRNA
    - post transcriptional gene silencing
    - regulation of transcription (act. Or repress.)
  - piRNA
    - Piwi-interacting RNAs
    - Epigenetic/post-transcriptional silencing in germline cells
  - eRNA, pRNA
    - enhancer (e), promoter (p) RNA
    - regulation of enhancer(upstream)/promoter
  - asRNA
    - antisense RNA → transcribed from non-template strand of coding region
    - sequestering of the sense transcript
  - lncRNA
    - long non-coding RNA; also lincRNA (long intervening RNA)
    - regulation of gene expression
    - chromatin remodeling
    - genomic imprinting; X-chromosome inactivation
    - circRNA → circular RNA
    - tRF RNA (TRS)
      - tRNA fragments



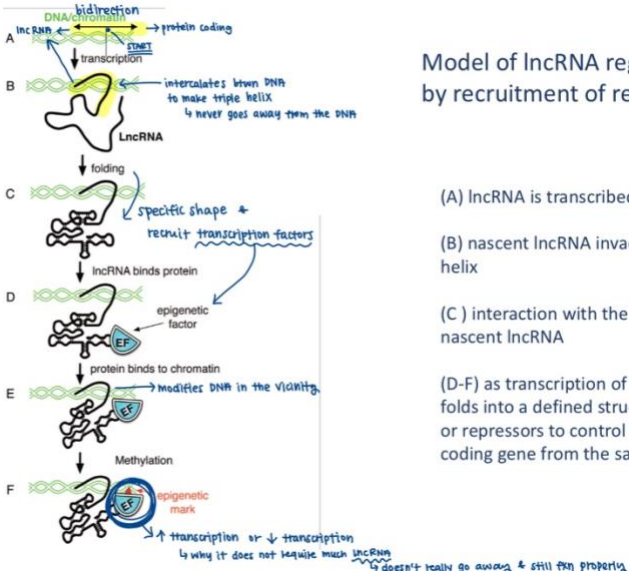
- mRNA gene → pri-miRNA (hairpin) in clusters → DROSHA process → single pre-miRNA → exportin 5 exports into cytoplasm → dicer makes it into a mature mRNA → dsRNA → siRNA duplex → unwind to active & passenger (degraded) → processed by RISC → binds to target mRNA 100% → cleavage → degradation

- **OR** mRNA gene → pri-miRNA (hairpin) in clusters → DROSHA process → single pre-miRNA → exportin 5 exports into cytoplasm → dicer makes it into a mature mRNA → miRNA (gap is present) → unwind to active & passenger (degraded) → processed by RISC → imperfect base pairing still present → degradation
- Imperfect pairing with target mRNA usually at 3' UTR → why single miRNA targets many (regulation)
- Mechanism of action in translation suppression not completely clear
    - Repression of translation at initiation → how miRNA prevents by not allowing ribosome binding
    - Repression of translation post-initiation → may start, but no translation → stops in the middle
    - Targeting mRNA for degradation → when its 100% binding
    - mRNA and miRNA → sequestered into p-bodies → membrane less organelle, like nuclear protein complexes contain RNA that is targeted for degradation
        - RNA degradation
        - Nonsense mediated decay
- LncRNA: noise or function
    - Noise:
        - Generally expressed at very low levels
        - Sequences not well conserved between species
        - Low affinity binding of Pol II to random sequences
        - Eukaryotic promoters are bidirectional
    - Function: most are functional
        - Some have well established function
        - No demand for struct sequence conservation
        - lncRNA promoter sequences conserved → but sequences are not conserved due to the fact that it is not coding for anything
        - expression is regulated → tissue specific
        - secondary/tertiary structure may be conserved → tertiary is important for the function of RNA not the sequence
        - position in the genome is conserved
- lncRNAs are transcribed from different locations → some are located in the introns of protein or overlap with protein coding region (anti or bi)

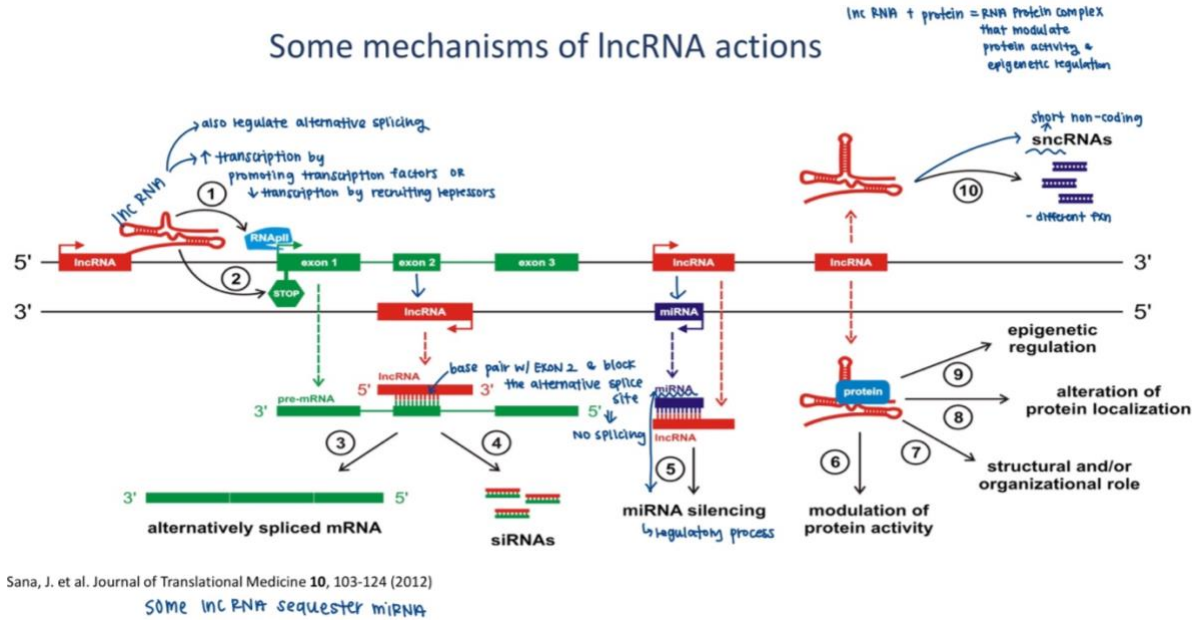


- lncRNA function:
    - regulation of chromatin structure
        - cis-and trans-silencing by recruiting epigenetic modifiers
    - regulation of gene expression by lncRNA transcription (activating or silencing)
        - either recruit or stop transcription factors from binding
    - regulation of transcription in cis (overlapping transcripts) or trans
        - activation or repression of transcription factors and accessory proteins
    - post-transcriptional regulation

- regulation of mRNA transcripts (binding of antisense transcripts)
  - checks for stability as part of regulation
  - antisense transcript pairs with target mRNA to block transcription
- gene silencing pathways



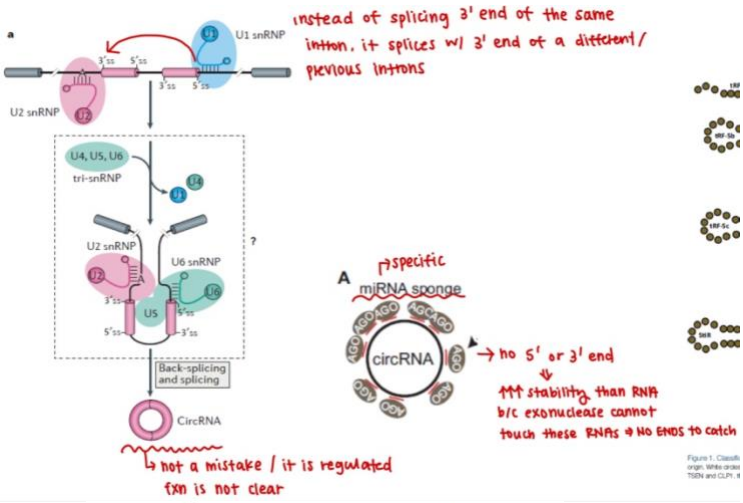
- lncRNA is transcribed from a bidirectional promoter
  - Bidirectionality – only in eukaryotic
  - One way codes for protein other does not
- Nascent lncRNA invades the DNA duplex to form a triple helix
- Interaction with the DNA promotes folding of the nascent lncRNA
- As transcription of the lncRNA continues, the lncRNA folds into a defined structure that binds protein activators or repressors to control the transcription of the protein coding gene from the same promoter



Sana, J. et al. Journal of Translational Medicine **10**, 103-124 (2012)

some lnc RNA sequester miRNA

- Some mechanisms of lncRNA actions in cytoplasm
  - Most lncRNA act in the nucleus and regulation transcription OR many act locally @ their site of transcription (but not always)
  - In cytoplasm: increase or decrease in stability of mRNA or inhibit/activate transcription or switch to cap- independent translation
- XIST: inactivation of X chromosome in cis by chromatin remodeling
- HOTAIR: HOX Trascrnipt Antisense RNA: oncogenic in many cancers. Regulates expression of HOX gene cluster through epigenetic modification → epigenetic modification are oncogenic in cancer
- MALAT1: metastasis associated lung adenocarcinoma transcript 1. ALSO called NEAT2
  - Overexpression (in cancer) linked to increase in cell proliferation in colorectal and lung cancer
  - May be implicated in breast cancer
  - Regulates SR splicing factors and controls repression of some genes
- Newest members of ncRNA family



- circRNA (circular RNA)
  - produced by back splicing mostly or from intron lariat after normal splicing
  - function is unclear; miRNA sponges (sequestering miRNA); some translated – rolling circle translation

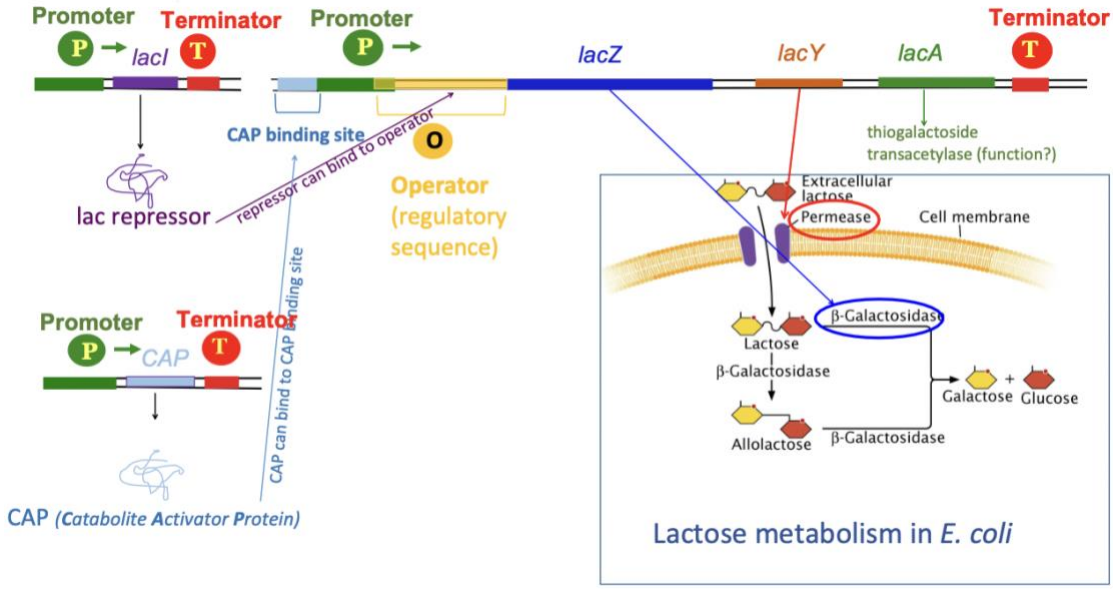
- tRF (tRNA fragment)
  - defined fragments of tRNA → parts of degradation of tRNA
  - induced by stress and starvation
  - regulate translation in stress conditions
  - suggested roles in cell proliferation, tumor suppression

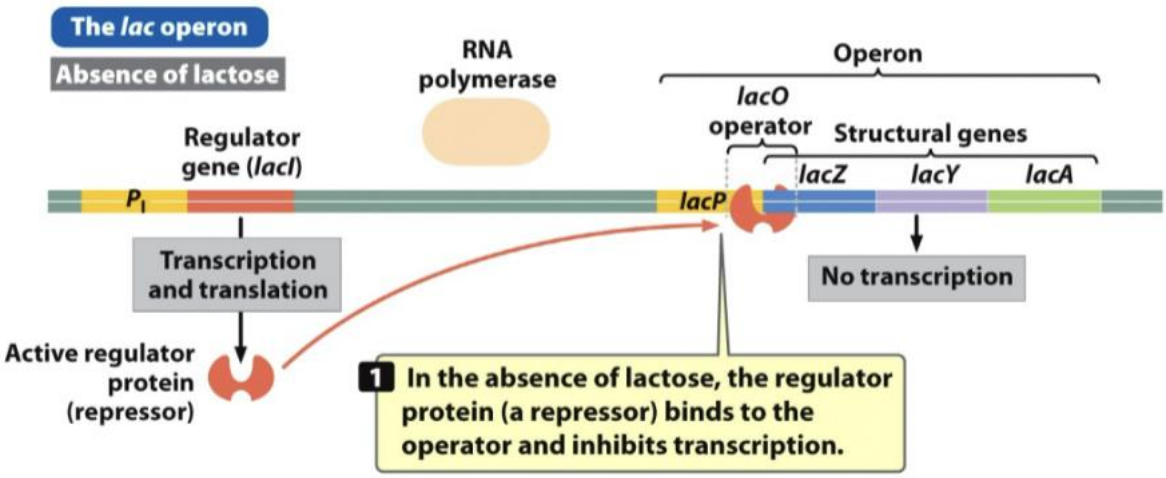
- CRISPR
  - How bacteria save themselves
  - CRISPR RNA serves as a key foreign DNA destructor complex
  - CAS 9 complex (type II), crRNA, tracrRNA
    - CAS 9: binds short motif
    - PAM: protospacer Adjacent Motif in DNA
      - Present next to protospacer recognize Cas9 and binds → cleavage of protospacer occur
  - RNA can be engineered to recognize protospacer sequence
    - Turns CRISPR 9 into powerful genome editing tool
      - CRISPR 9 cleaves specific sequences in the DNA and is repaired
      - Can also modify to bind DNA without cleaving them → CAS 9 fuse with fragments of other proteins → non-natural fusion protein complexes alter gene expression, chromatin modification, fluorescent protein (marks genome)
    - Breaks DNA by altering or adding new DNA sequences
- miRNA as a therapeutic
  - some miRNA: oncogenes or tumor suppressors
  - miR-34 for primary metastatic liver cancer – terminated due to side effects
    - miR-122 is required for virus replication, miravirsen blocks from binding → treatment for chronic HepC
  - miRNA (most ncRNA) expression is tissue-specific

    - cancer have different miRNA profiles form normal cells (cancer specific)
    - cancer-specific miRNA → detected in patient serum very early in tumor development
        - miRNA is double-stranded – more stable which can be isolated from blood → presence of double stranded indicate presence of cancer
        - **Potential for early, non-invasive cancer diagnosis (liquid biopsy)**
    - Cancer-derived exosomes (membrane-bound vesicles excreted from cells) present in serum carry miRNA, lncRNA, circRNA and tRF
        - >200 clincial trials currently looking at miRNAs as biomarkers for early detection of the disease or biomarkers of response to therapeutics
- siRNA application in research
    - RNA interference (RNAi): downregulation of expression of individual genes
    - Identify functions of a gene of interest
    - Therapeutic: transthyretin-mediated amyloidosis, hypercholesteremia, macular degeneration, truncated dystrophin protein for patients with DMD, covid vaccine (moderna) – mRNA
- Major challenges of RNA therapeutics
    - Stability (half-life)
        - Locked nucleic acid; backbone modifications and base modification
            - Prolong therapeutic life
    - Delivery – charged molecules do not pass easily
        - Lipid nanoparticle: does not require much modification of nucleotides
            - Increases the therapeutic life of materials inside the lipid nanoparticles
                - Fuses to the membrane and release cargo inside
                - If not, fuse with endosome
            - Have defined concentration of lipid with protein which act as an anchor to target specific cells → deliver therapeutics where it needs to be
        - Conjugation to carrier: liver cell delivery ONLY
            - Therapeutic RNA → conjugated to this molecule to target liver

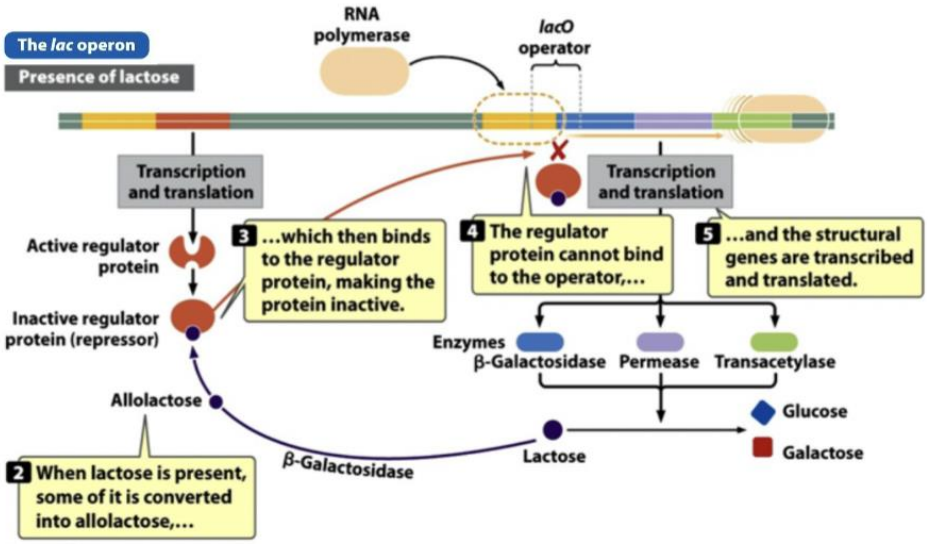
Class 7: Regulation of Gene expression in Bacteria

- Gene expression in bacteria is regulated
  - E. coli has ~4200 protein-coding genes but not all genes are expressed all the time. WHY?
    - Conservation of resources
    - Response to change in environmental conditions
  - Regulation of gene expression in bacteria:
    - Constitutive expression: expressed all the time
    - Repressible: normally on, but can be turned off by a repressor
    - Inducible: normally off, but can be turned on (induced) by an inducer
    - Regulation of transcription: how much of the mRNA is made → **most important mechanism for majority of the genes**
    - Regulation of mRNA stability: how fast the mRNA is degraded
    - Regulation of translation: how efficiently the mRNA is translated
- Examples of regulation of gene expression in bacteria
  - Lac operon (inducible system – negative and positive regulation

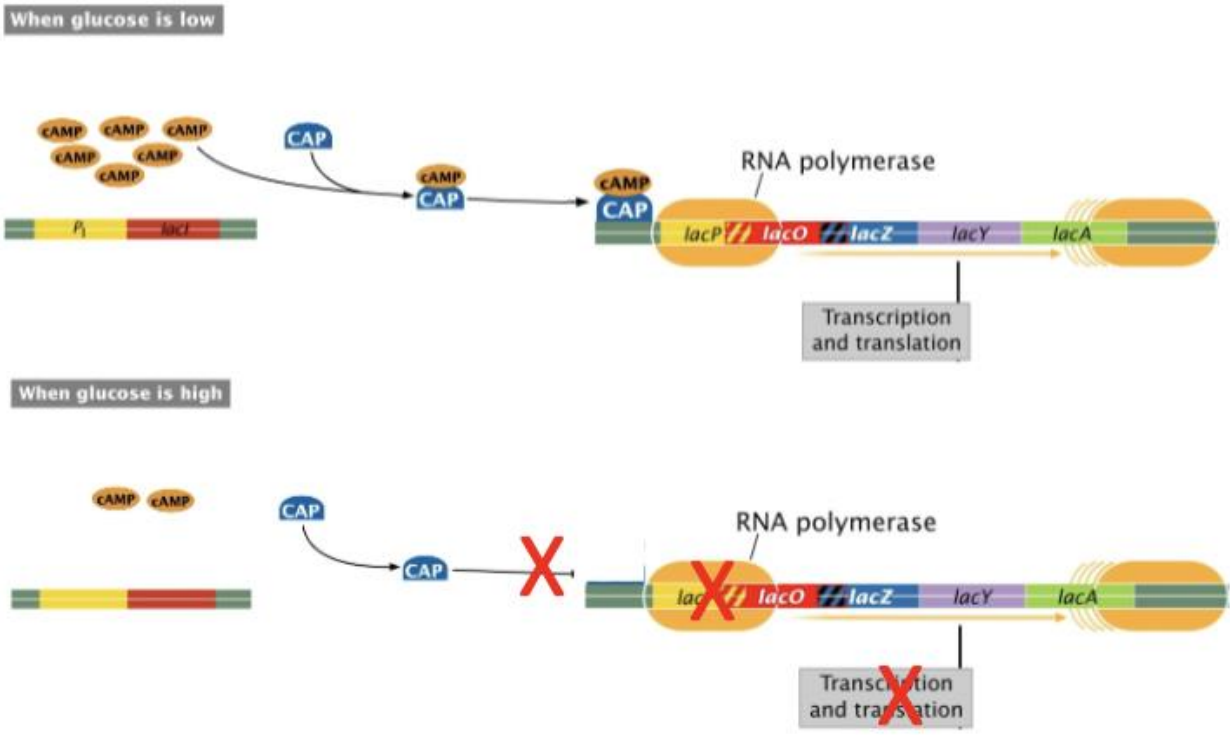




- When lactose is NOT available:
  - Lac repressor protein in made
  - Lac repressor binds to the lac operator sequence
  - RNA pol cannot bind to the promoter
  - Binding of the repressor to the operator inhibits transcription of lac genes

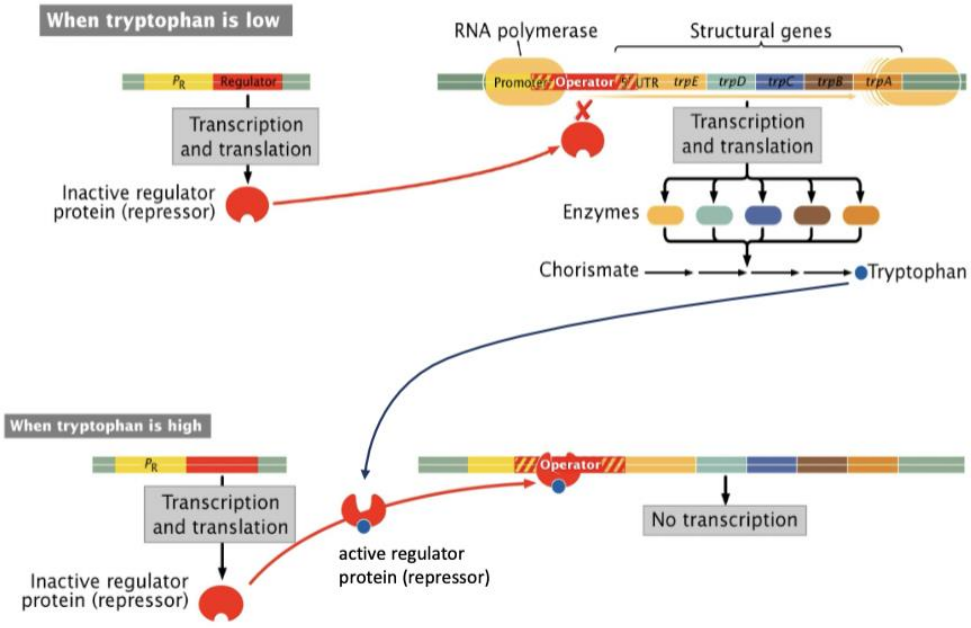


- When lactose IS available:
  - Lac repressor protein is made
  - Lac repressor binds to the allolactose (derived from lactose)
  - Allolactose binding changes conformation of the repressor → repressor + allolactose complex CANNOT bind to the operator
  - RNA pol can bind to the promoter and transcription of lac genes can proceed
- Lactose operon in E. coli → prefers glucose (bacteria) – more efficient
  - Presence of BOTH lactose + glucose → metabolically more efficient to utilize glucose
  - + glucose → lac genes are not efficiently expressed even with presence of lactose
    - RNA pol binding to the lac promoter is weak → very low level of transcription
    - Cap can bind upstream of lac promoter to the CAP-binding site
    - When CAP is bound, it can stabilize the RNA pol on the promoter → stronger binding → increase level of transcription
    - CAP binding is necessary for efficient transcription of lac genes
    - CAP needs to be in complex with cAMP to be able to bind to the CAP-binding site



Lactose	Glucose	cAMP	Activator (CAP)	Repressor	RNA Pol	Transcription of Lac
-	+	Low	No bound	Bound	Blocked by repressor	-
+	+	Low	No bound	Not bound	Weakly bound	+
-	-	High	Bound	Bound	Blocked by repressor	-
+	-	High	Bound	Not bound	Strongly bound	++++++

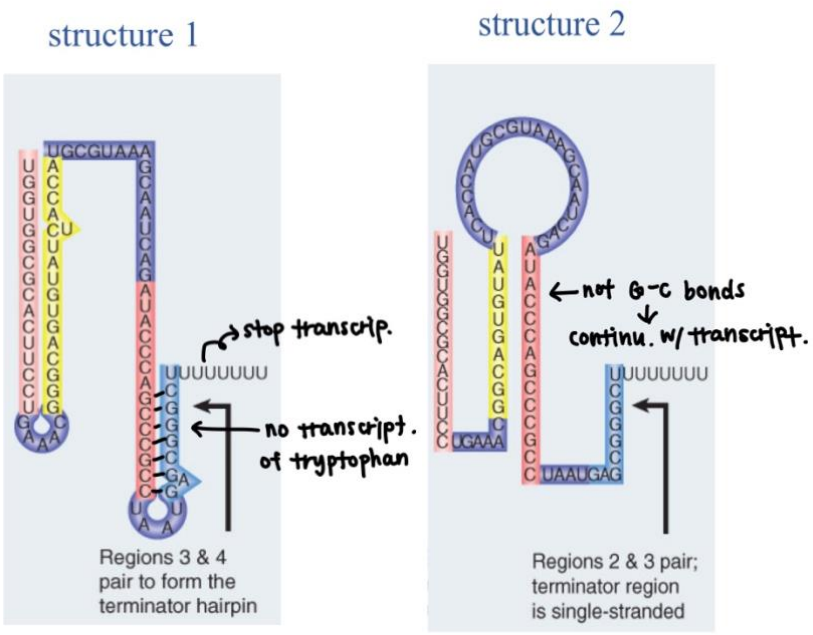
- TRP operon (repressible system – two mechanisms of negative regulation)
  - Two levels of control:
    - Repression
    - Attenuation
  - 5' structural genes coding for enzymes necessary to synthesize Trp → all they need to make the genes
  - Leader sequence in 5' UTR of the transcript
  - Leader sequence contains 14-codon Open Reading Frame (ORF) → code for a short peptide w/ 2 Trp residues



When Trp in the cell is LOW:
- Trp repressor protein made
- Trp repressor cannot bind to the Trp operator sequence
- RNA Pol can bind to the promoter and transcription of Trp genes can proceed

When Trp in cell is HIGH:
- Trp repressor protein is made
- Trp repressor protein binds Trp
  - Binding of Trp changes conformation of the repressor → repressor +Trp can now bind to the Trp operator
- RNA pol cannot bind to the promoter and transcription of Trp genes stop

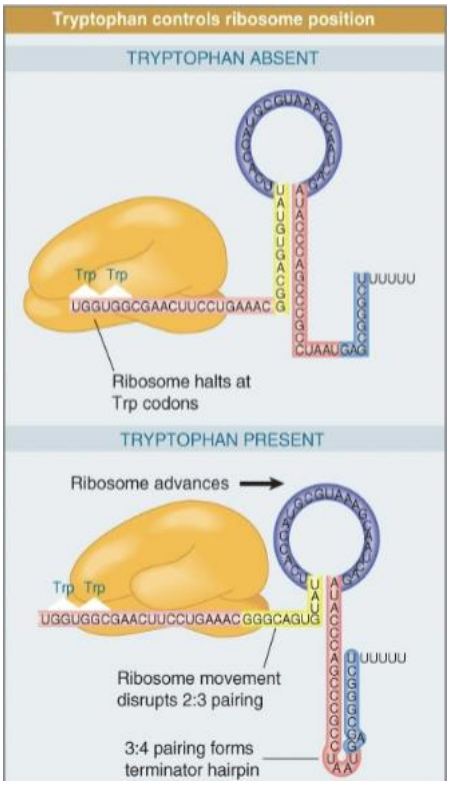
- Attenuation: additional level of control → useful b/c repression is not 100% effective



Structure 1: contains a terminator hairpin
- RNA Pol is released from the template DNA
- Trp genes downstream of the leader sequence are not transcribed

Structure 2: does not have a terminator hairpin
- RNA Pol continues transcription after the leader sequence
- Trp genes downstream of the leader sequence are transcribed

- Nascent mRNA of the leader sequence can adopt 2 different secondary structures
  - Secondary structure determined by translation
  - Regulated by the presence of Trp tRNA

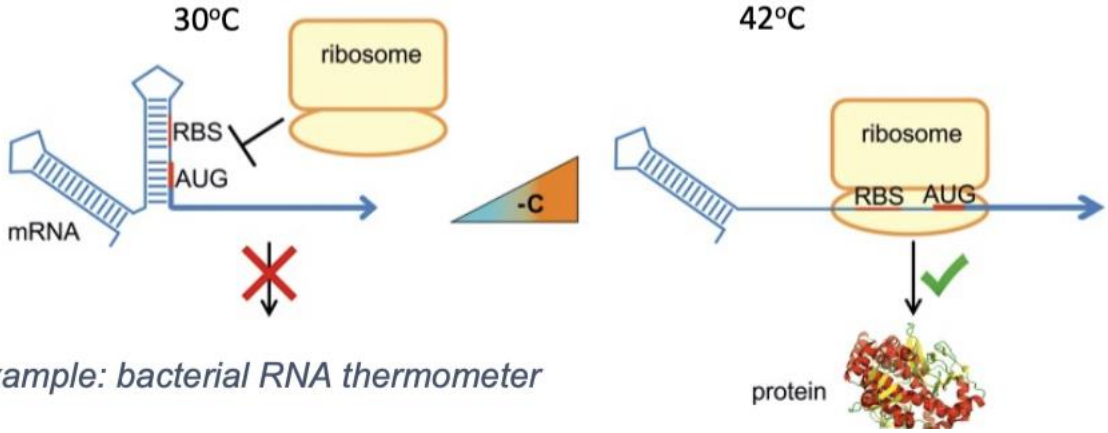


When Trp in the cell is LOW:
- Trp-tRNA is low
- Leader sequence is translated up to Trp codons
- Ribosome stalls at Trp codons
- Favors formation of the alternative loop
- Transcription continues, and Trp genes are transcribed

When Trp in the cell is HIGH:
- Trp-tRNA is high
- Leader sequence is translated successfully
- Ribosome passes Trp codons without stalling
- Favors formation of the terminator loop
- Transcription terminates, Trp genes are not transcribed

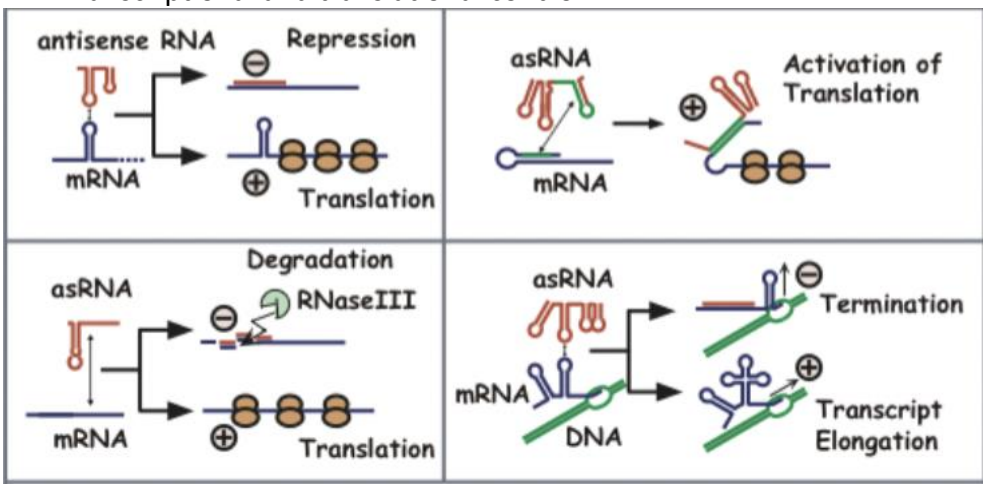
- Other examples of regulation of gene expression in bacteria
  - Riboswitches
    - Common in bacteria
    - Mostly in 5' UTR sequences
    - Direct binding of metabolites to RNA → changes in secondary structure

- Transcriptional control (activation, repression)
- Translational control (activation, repression)



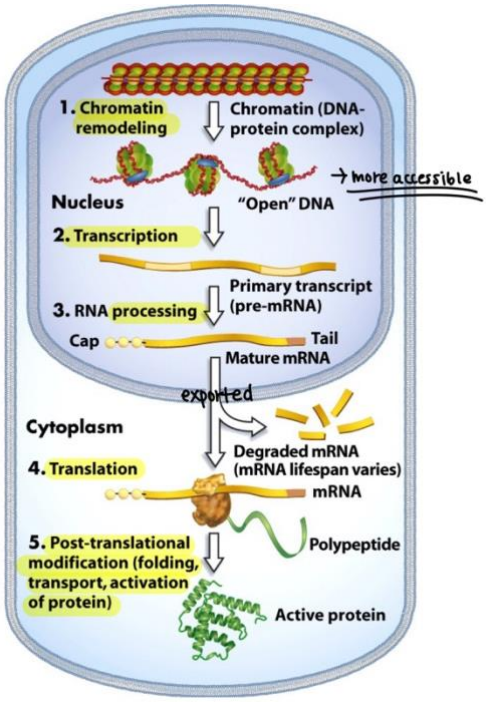
example: bacterial RNA thermometer

- Small RNA (sRNA)
  - ~80-100 bp
  - Act by complementary base pairing with mRNAs → pairing is not usually perfect (have gaps)
  - One sRNA can regulate multiple target genes
  - RNA chaperone Hfq often promotes interactions between sRNA and target mRNA
  - Transcriptional and translational control



Class 8: Regulation of gene expression in Eukaryotes

- Eukaryotic regulations
  - Genes are not organized in operons → each gene is transcribed separately
  - Eukaryotic DNA is organized into chromatin
  - Eukaryotic genes contain introns
  - Transcription and translation are separated
  - Mechanisms of transcriptional regulation essentially similar to regulation in bacteria → but much more complex
  - Regulation at various levels of gene expression
- Importance of regulated gene expression in multicellular eukaryotes
  - In multicellular organisms, regulation of gene expression is essential for cell specialization
    - All somatic cells in an organism have the same DNA, but they can be vastly different in size, shape, and function
    - Different types of cells express different sets of genes
  - Cells respond to external signals
    - Signals are communicated to transcriptional regulators through signal transduction pathways

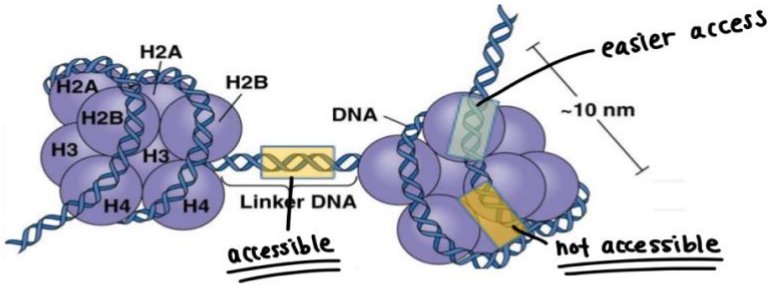


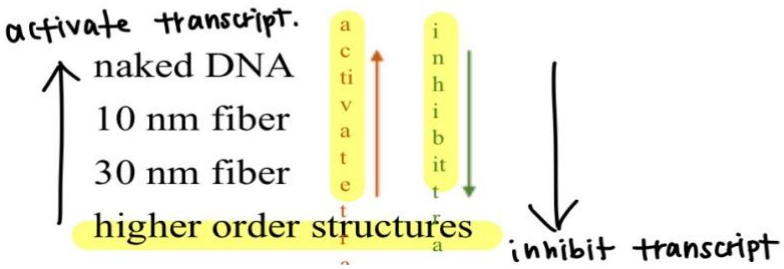
Gene expression in Eukaryotes: many steps from DNA to active Protein

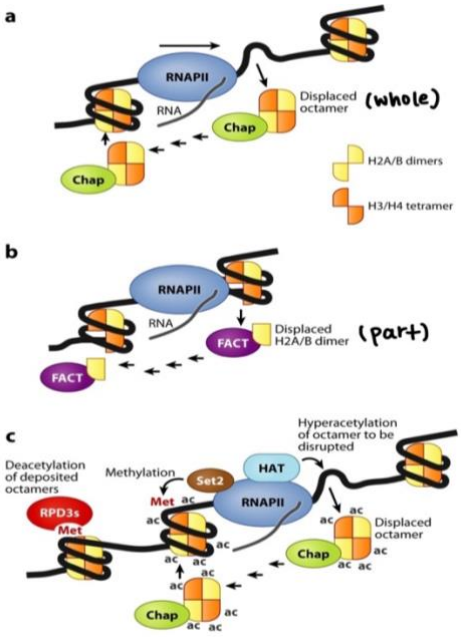
- Regulatory mechanisms exist at each of these steps

"Regulation of chromatin structure can regulate gene expression"

- Eukaryotic RNA polymerases have to deal with nucleosomes
- Nucleosomes/chromatin are a general inhibitor of eukaryotic transcription (limit access to DNA)
  - Transcription factors need to bind (need space)

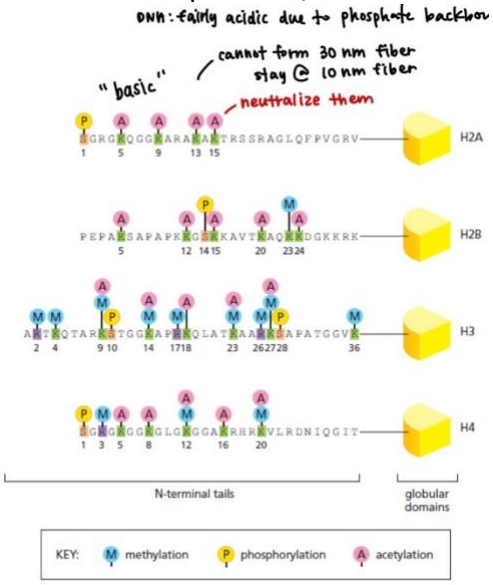




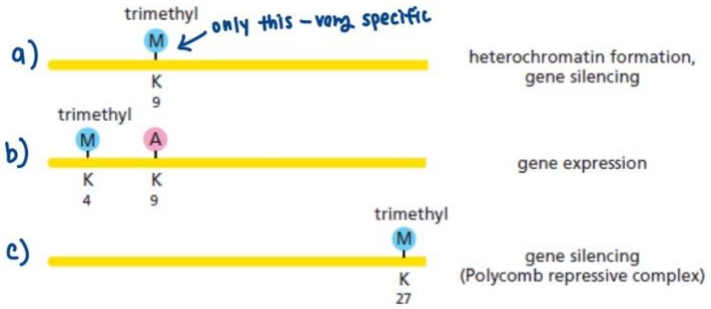


- Eukaryotic RNA polymerases have to deal with nucleosomes → nucleosomes/chromatin are a general inhibitor of eukaryotic transcription (limited access to DNA)
  - Chaperones can displace the octamer and replace it behind the polymerase
  - Polymerase may traverse through nucleosomes without fully displacing them
  - Co-transcriptional modification of histones can help with displacement, replacement, and recruitment of additional modifiers

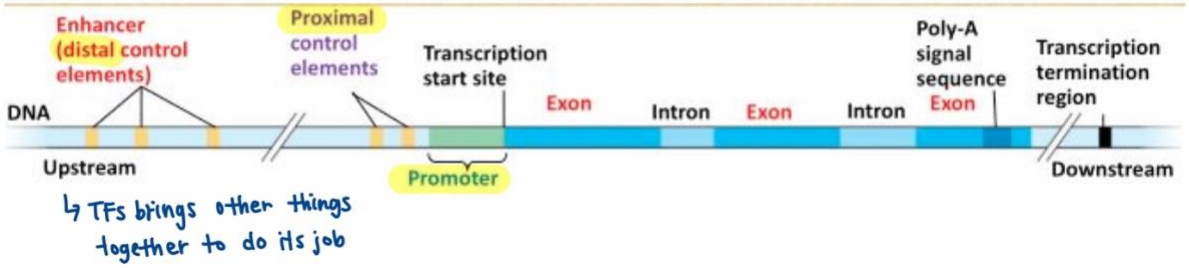
- Regulators of nucleosome function/location
  - Nucleosome remodeling complexes
    - dsDNA translocases (can shift the DNA around the nucleosome)
    - slide histone octamers on DNA
    - exchange histone subunits
      - different histone variants have different functions→ some preferentially present at transcription sites
  - Histone Modifying enzymes
    - Modify specific histones by adding a small or large molecule, typically on the N-terminal tail that is unstructed
      - Modifications are DYNAMIC
- Chromatin structure affects gene expression → many activators of transcription in eukaryotes act by changing accessibility of DNA
  - Heterochromatin restricts gene expression (X-Chromosome inactivation)
  - Chromatin remodeling complexes can alter DNA accessibility by changing nucleosome positions
  - Histone modifying enzymes can change gene expression by recruiting factors that promote/inhibit transcription



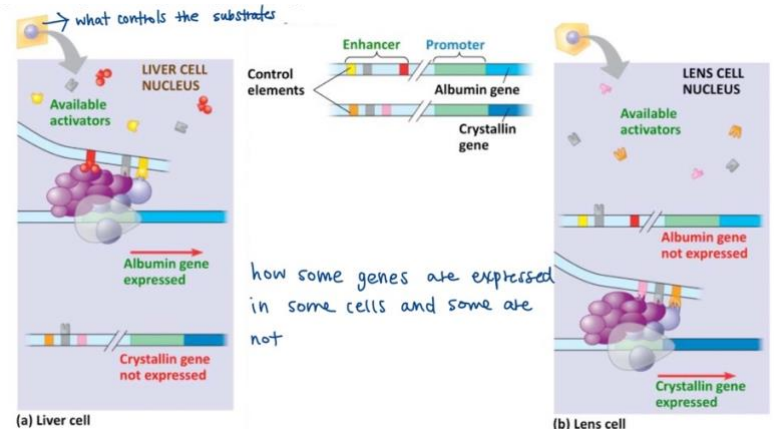
- Core histones are covalently modified at many different sites
  - Histone modification affect accessibility of DNA to transcription factors and RNA pol
    - Acetylation: increase transcription
    - Methylation: decrease transcription (not always)
    - Histone modification: dynamic (HAT, HDAC)
  - Specific combinations of histone modifications may also regulate gene expression by recruiting transcription regulators
    - "histone code" hypothesis
    - Histone readers, writers, erasers



- Signals from multiple regulatory elements of a eukaryotic gene are integrated to regulate its transcription
    - Remember the positive and negative regulator of lac operon – integration of signals from 2 regulatory elements

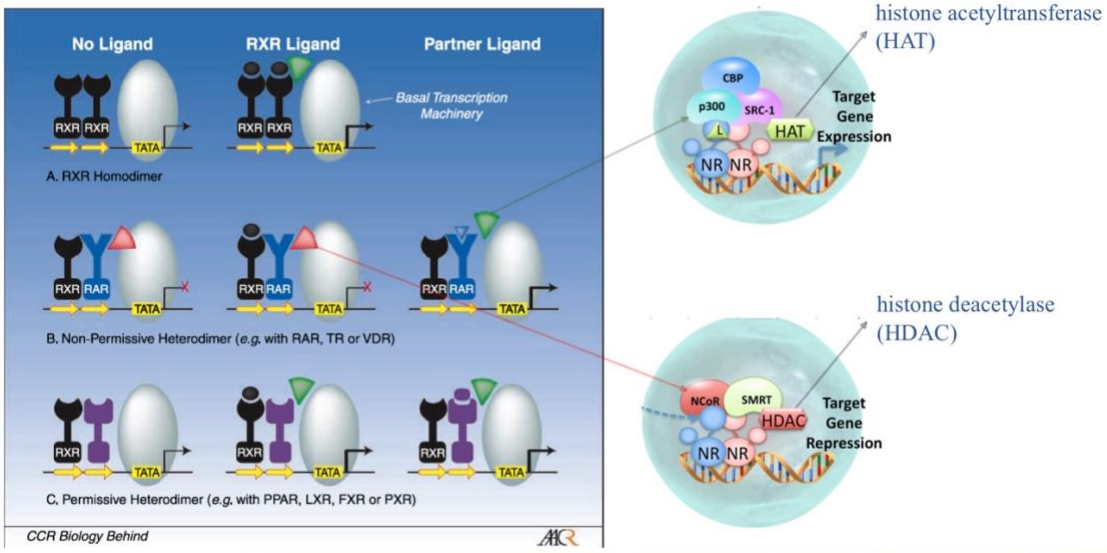


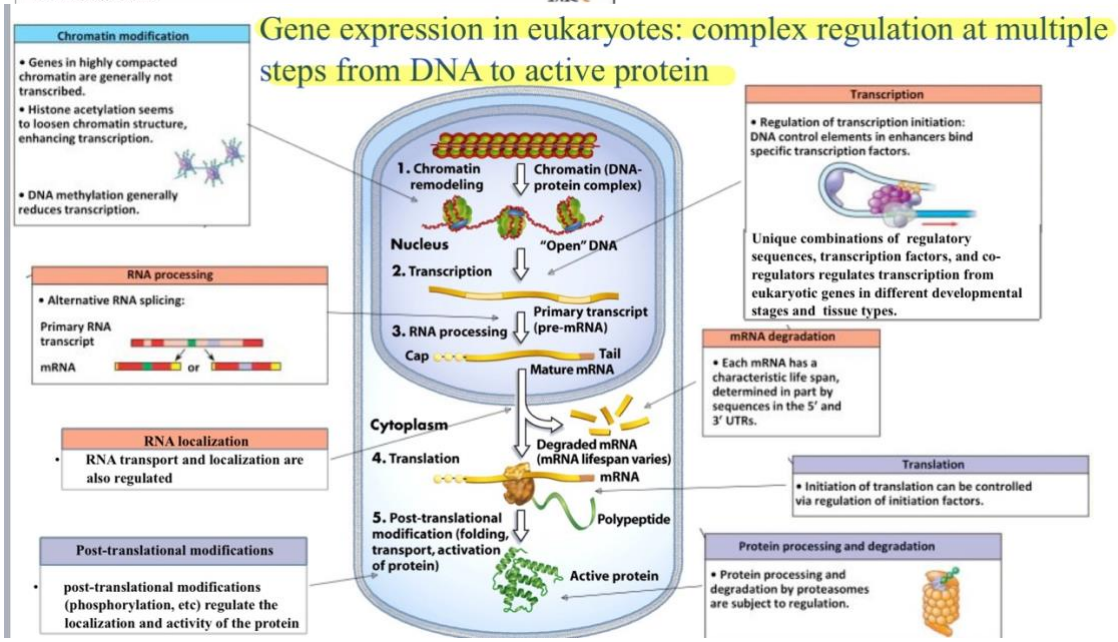
    - Transcription of each eukaryotic gene is regulated by multiple elements
    - Transcription regulators act in complexes with coregulators
    - One regulator can regulate transcription of many genes
    - One regulator can have opposite effects on transcription (either active or repress) depending on context (gene, cell type)
- Transcription regulators bind cooperatively to DNA → need multiple not single regulators
    - Binding of one protein facilitates binding of the other by destabilizing the nucleosome and making the DNA more accessible
    - Pioneer transcription factors can bind to nucleosomal DNA in closed chromatin
    - Protein-protein interactions between regulators can favor DNA binding
    - Different models of regulatory complex assembly
- Cell specific availability of activators, repressor, and co-regulators can regulate specificity of gene expression



- Master transcriptional regulators control specific gene expression programs → control expression of other transcription factors
    - Super-enhancers occupied by master transcription factors and mediator control mammalian cell identity
    - Ectopic expression of these key transcription factors reprogram differentiated cells

- Post-transcriptional regulation of gene expression
    - Alternative splice
    - miRNA generation and function
- Model of lncRNA regulation of local transcription by recruitment of repressors/activators
    - lncRNA is transcribed from a bidirectional promoter/enhancer
    - nascent lncRNa invades the DNA duplex to form a triple helix
    - interaction with the DNA promotes folding of the nascent lncRNA
    - as transcription of the lncRNA continues, the lncRNA folds into a defined structure that binds protein activators or repressors to control the transcription of the protein coding gene from the same promoter
- Eukaryotic gene expression is also regulated at the level of translation → post translational modifications regulate activity of many proteins
- RXR nuclear receptor – transcription factor with pleiotropic functions → regulation of gene expression in eukaryotes
    - Obligatory heterodimeric partner for other nuclear receptors
    - Can activate or repress transcription, depending on context (heterodimeric partner; presence/absence of ligand; tissue specific coregulators
    - Acts by recruiting coregulators (coactivators/corepressors)





Questions for RNA Modules

Which of the following is a difference between DMS-seq and SHAPE-seq techniques in mapping RNA secondary structures? Check all that applies

a. DMS reagent modifies unstructured regions of RNA, while SHAPE reagent modifies regions that are highly structured
b. SHAPE-seq is high-throughout (it can assess structures of many RNAs at the same time), while DMS-seq is low throughput (it can only assess one RNA in one experiment)
c. DMS-seq and SHAPE-seq have similar procedures, but use different reagents to assess RNA structure
d. SHAPE-seq can provide base pair-level resolution, while DMS-seq can not

Which method would be the most accurate for predicting RNA secondary structure?

a. Computational prediction, based on minimizing the free energy of the RNA fold of given RNA sequence
b. Partial RNAse digestion of the given RNA using RNAse T1 (cuts ssRNA after G)
c. Shape-seq

Which of the following statements about lncRNAs is NOT correct

a. lncRNAs are usually expressed at low levels
b. lncRNAs can act by recruiting chromatin modifiers
c. lncRNAs regulate expression only at the transcription level
d. lncRNAs can regulate gene expression both positively and negatively

Which polymerase is more error prone?

a. DNA polymerase
b. RNA polymerase

Most bacteria have

a. Single RNA polymerase, single sigma factor
b. Single RNA polymerase, several sigma factors
c. Several RNA polymerases, single sigma factor.
d. Several RNA polymerases, several sigma factors

What is the main role of the sigma factors?

a. It speeds up the elongation rate of RNA polymerase
b. It unwinds DNA template
c. It directs RNA polymerase to the promoter
d. It binds to the terminator sequence.

Promoter sequence of the bacterial gene is the perfect match to the consensus sequence for sigma factor. It two adenines in the sequence are mutated to guanines, what would be the most likely outcome?

a. The gene will be transcribed more, because GC pairs bond more strongly than AT
b. The gene will be transcribed less, because sigma factor will have less affinity for the changed sequence.
c. It will have no effect on gene transcription, because the coding sequence of the gene was not changed
d. It will have no effect on gene transcription, because the transcription start site will be moved to compensate for the mutations.

Presence of secondary structures in the nascent mRNA chain during transcription in bacteria can lead to (check all that applies)

a. Termination of transcription
b. Stalling of the polymerase
c. Polyadenylation of nascent mRNA
d. Editing of nascent mRNA

How many nuclear RNA polymerases are present in eukaryotes?

a. 1
b. 2
c. 3
d. 4

Which of the following statements about Pol II promoters is correct?

a. TBD binds to all Pol II promoters
b. Pol II cannot bind to the promoter sequence without the sigma-factor
c. Pol II promoters are always located upstream of the transcription start site
d. Pol II promoters usually have several consensus sequences to which transcription factors can bind.

MiRNAs are transcribed by:

a. Pol I
b. Pol II
c. Pol III
d. Pol IV

In addition to the A, U, G, and C, RNAs can contain some unusual bases. The most common unusual base in RNA is:

a. Thymine
b. Inosine
c. Methylguanine
d. Pseudouracil

The most abundant RNAs in the cell are:

a. mRNAs
b. tRNAs
c. miRNAs
d. rRNAs

7SL RNA is:

a. an example of a ribozyme
b. an example of riboswitch
c. present in eukariots
d. part of the RNai pathway

7-methylG cap is added to the 5’ end of all Pol II transcripts

a. True
b. False

All Pol II transcripts are polyadenylated at the 3’ end

a. True
b. False

Almost all introns are spliced by a ribozyme

a. True
b. False

Which of the following RNA secondary structures will most likely have the lowest ΔG upon folding?

a. Very large loops
b. Single-stranded regions
c. Very small loops
d. Double-stranded regions (stems)

Does the stability of any mRNA change after drug treatment? Select all that apply.

a. RNAseq
b. Groseq
c. ChIPseq

Does the drug affect the ability of bidirectional promoters to transcribe in both directions?

a. Groseq
b. RNAseq
c. ChIPseq

Nrf2 is a transcription factor expressed in skin cells. Does Nrf2 bind to unidirectional or bidirectional promoters in skin cells?

a. Groseq
b. RNAseq
c. ChIPseq