CAP 5510: Introduction to Bioinformatics
CGS 5166: Bioinformatics Tools

Professor Giri Narasimhan

ECS 254; Phone: x3748

giri@cs.fiu.edu

www.cis.fiu.edu/~giri/teach/BioinfF18.html

CAP 5510: Introduction to Bioinformatics
CGS 5166: Bioinformatics Tools

Professor Kalai Mathee

BSc (Genetics), MSc (Molecular Biology),
PhD 1992 (Microbiology & Immunology),
MPH 2018 (Health Policy & Management)
AHC4 - 225; Phone: x0628
Kalai.Mathee@fiu.edu
www.faculty.fiu.edu/~matheek

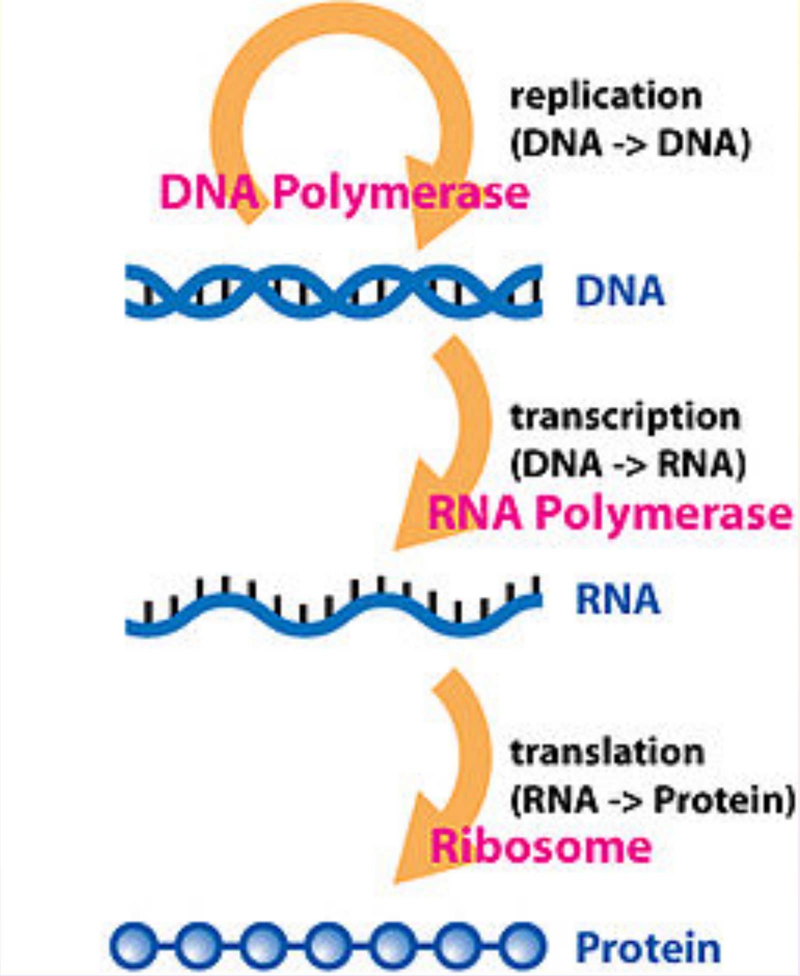
Session Learning Objectives

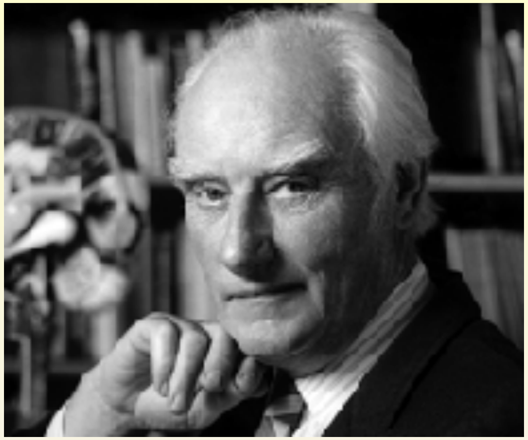
- Central Dogma
- Transcriptional Unit
  - Prokaryotic
    - Operon
    - Promoter
    - Operator
    - Enhancer etc
  - Eukaryotic
    - TATA Box
    - UAS
    - Silencer
    - Insulator etc
- Gene Regulation
  - Single
  - Multigene
    - Regulon
- Gene Expression
- Tools to study gene expression
  - Single gene
  - Multi gene
  - Genome-dependent
  - Genome-independent

Why Should You Care?

- Promoter Prediction Methods
  - https://scholar.google.com/scholar?q=promoter+prediction+methods&hl=en&as_sdt=0&as_vis=1&oi=scholart
- Gene Prediction Methods
  - https://scholar.google.com/scholar?hl=en&as_sdt=0%2C10&as_vis=1&q=gene+prediction+methods&btnG=
- Gene Expression Analysis
  - https://scholar.google.com/scholar?hl=en&as_sdt=0%2C10&as_vis=1&q=gene+expression+bioinformatic+methods&btnG=
- Sequence and Genome Analyses
  - https://scholar.google.com/scholar?hl=en&as_sdt=0%2C10&as_vis=1&q=Sequence+and+genome+analysis+&btnG=

Central Dogma

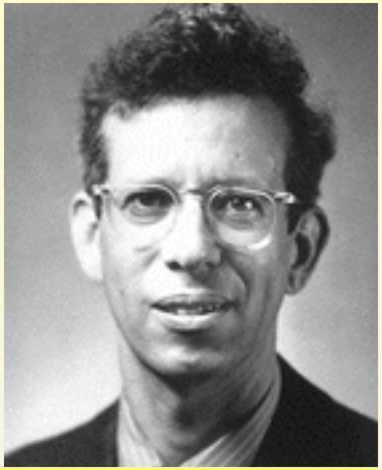




Francis Crick
(1916 - 2004)

- ❑ Nobel Prize — 1962 Med
- ❑ Central Dogma — 1958

Paradigm Change

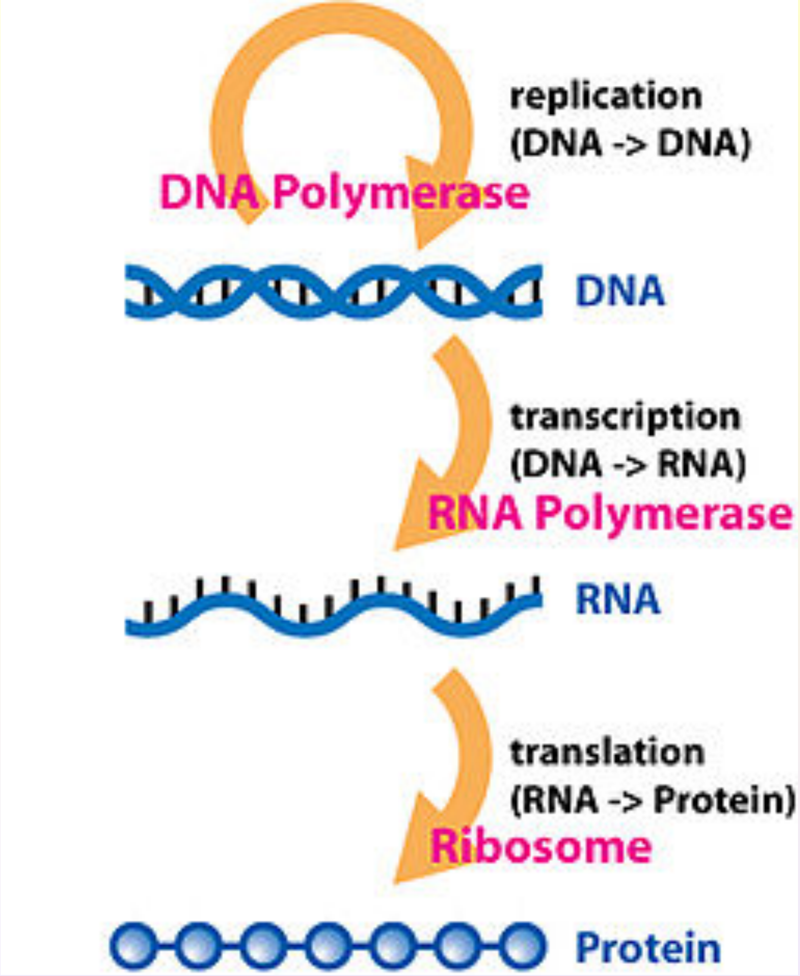


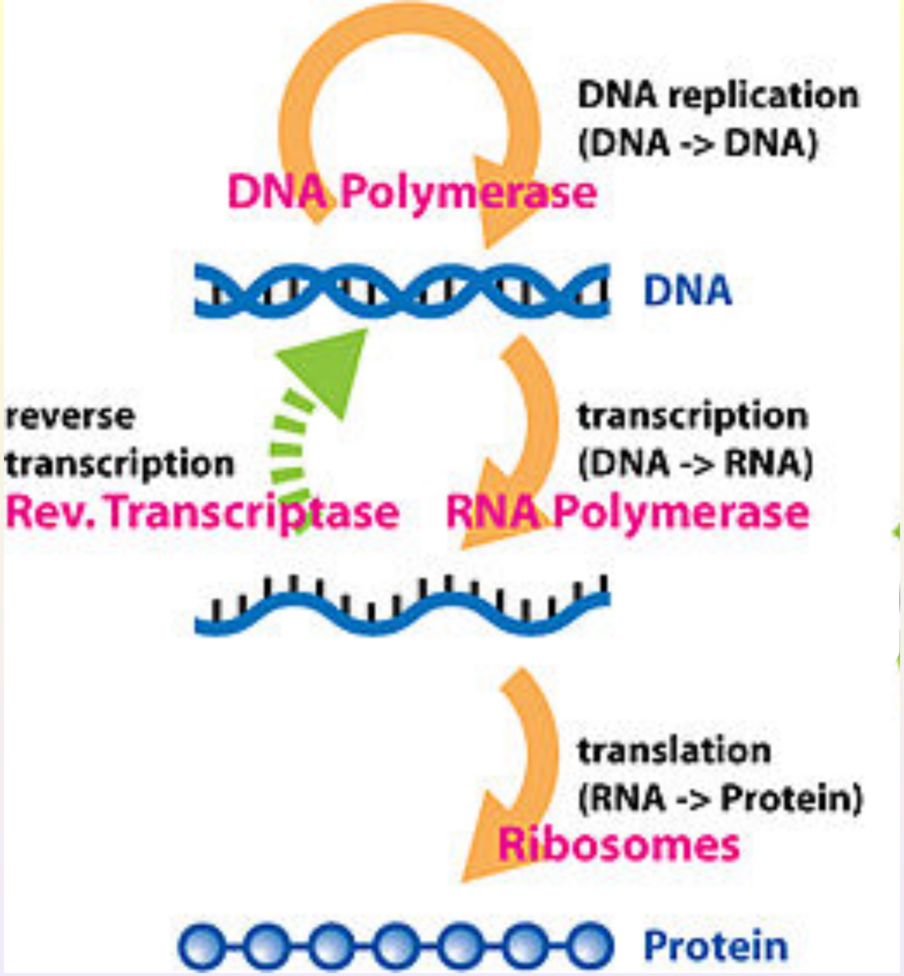
Howard M Temin
(1934 - 1994)

David Baltimore
(1938 -)

- ❑ Nobel Prize — 1975 Med
- ❑ Reverse Transcriptase

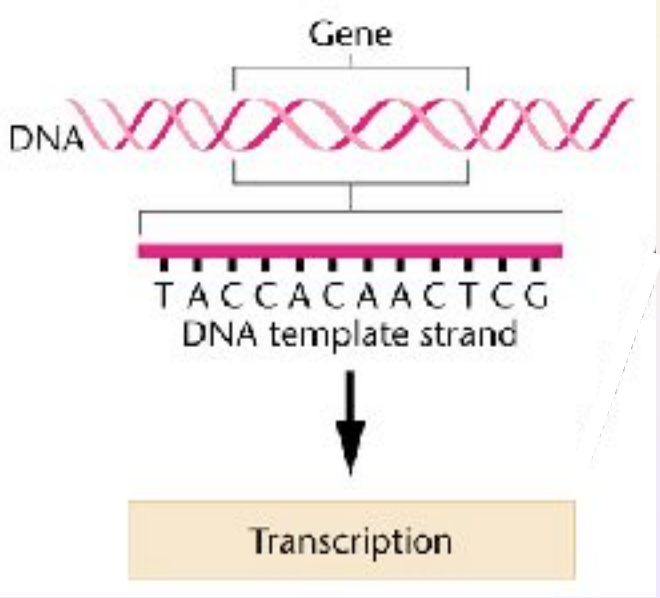
Central Dogma





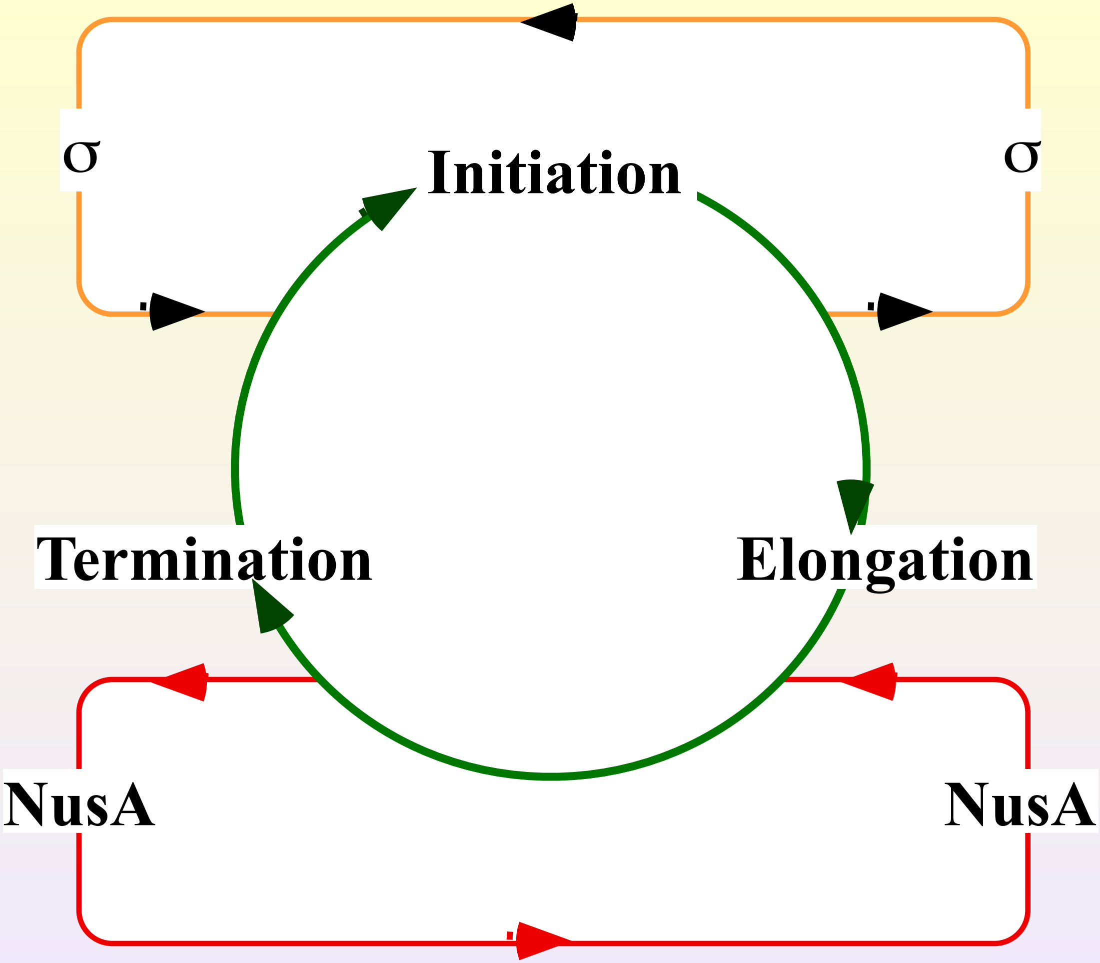
Transcription

- Process by which a selected segment of DNA corresponding to a gene is used to synthesize an RNA molecule
- Catalyzed by an enzyme called DNA-dependent RNA polymerase - (RNAP)



- Gene
- Open Reading Frame (ORF)
- https://www.youtube.com/watch?v=cXlv21NCGxQ
- https://www.youtube.com/watch?v=DKgJPhvCDU8

Transcription



σ

Initiation

σ

Termination

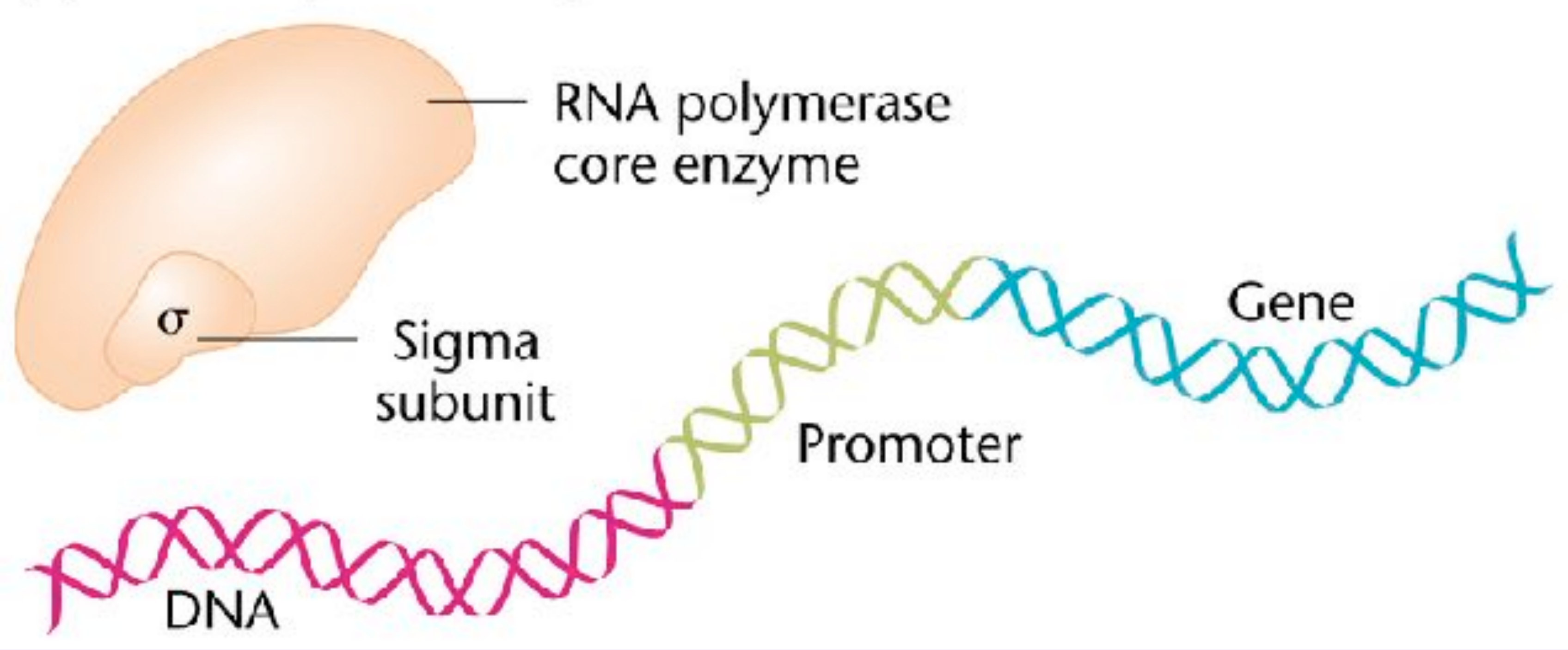
Elongation

NusA

NusA

Basic Transcription Components

- **Prokaryotes (Eubacteria)**
- **RNAP Core — 2α, β, β', ω**
- **RNAP Holoenzyme — Core + σ**

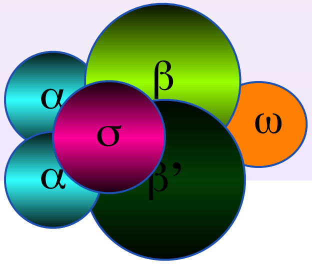


Eubacterial RNAP

Gene	Product	Subunits	Function
rpoA	α subunit 40 kD	α α	Enzyme assembly Promoter recognition Binds some activators
rpoB	β subunit 155 kD	β	Catalytic center
rpoC	β' subunit 150 kD	β'	
rpoZ	ω subunit 10 kD	ω	RNAP assembly
rpoD	σ subunit 32-90 kD	σ	Promoter specificity Binds to -10 and -35

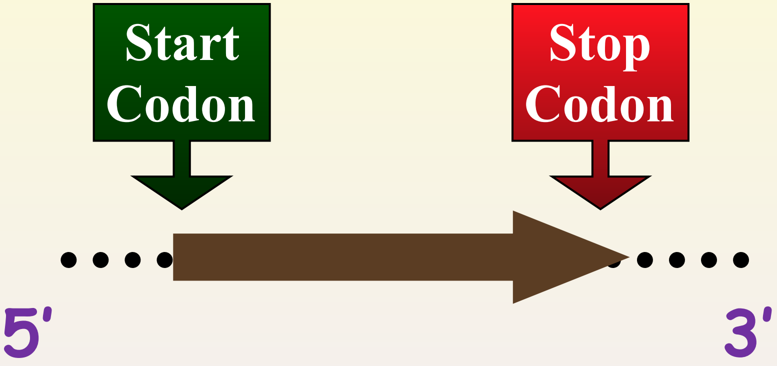
Core enzyme

Holoenzyme

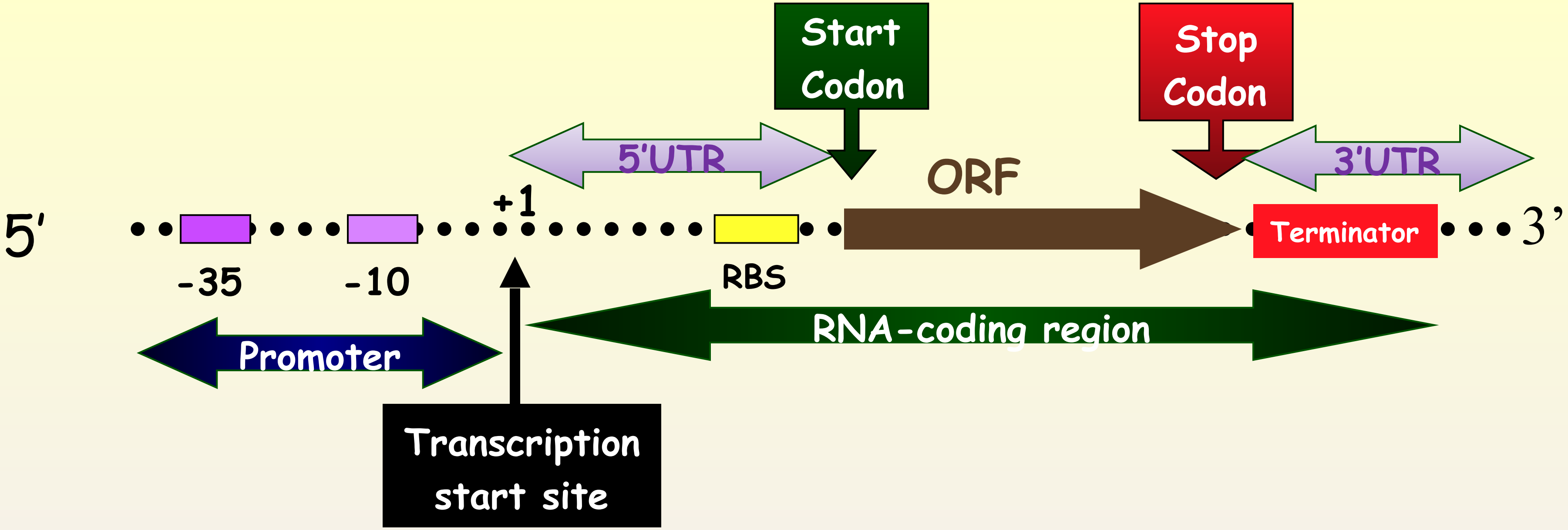


Genes/ORF – Eubacteria

- Start codon
  - AUG [DNA: ATG]
    - Met or Modified Met (fMet)
    - 83%
  - GUG [DNA: GTG]
    - 14 %
  - UUG [DNA: TTG]
    - 3 %
- Stop/Termination codons (3)
  - UAG [DNA: TAG] — Amber
  - UAA [DNA: TAA] — Ochre
  - UGA [DNA: TGA] — Opal
- Almost universal
  - Exception — Mitochondria

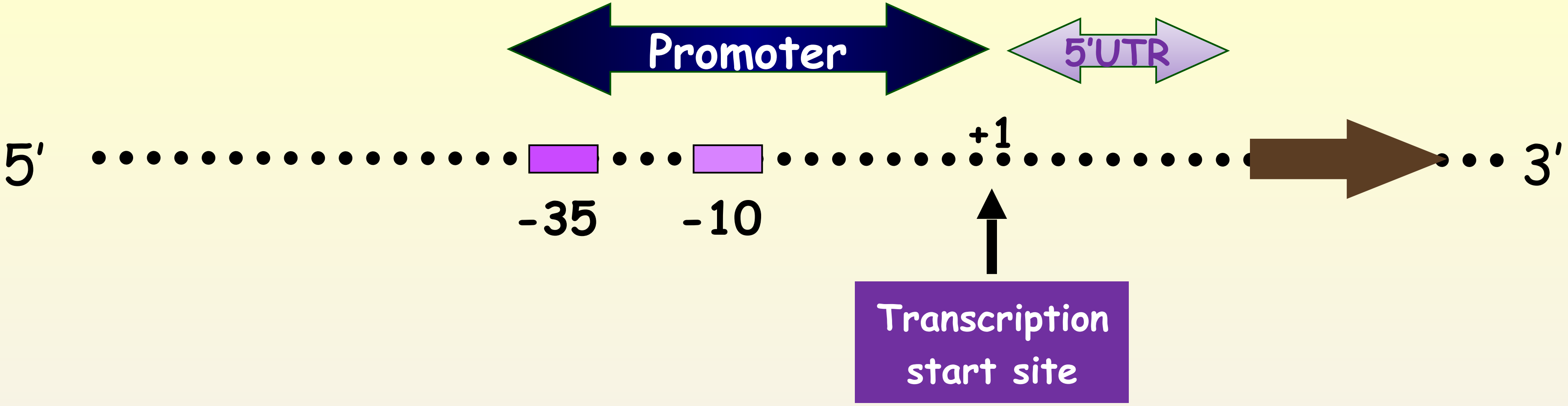


Eubacteria Transcription Unit



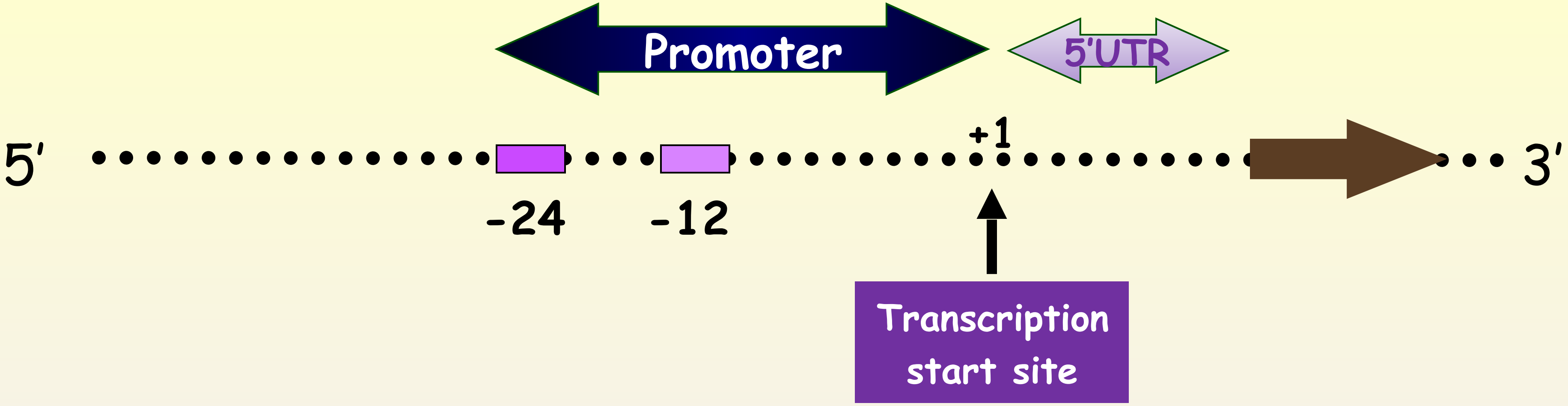
- Do eubacteria have introns?
  - https://www.ncbi.nlm.nih.gov/pmc/articles/PMC177115/pdf/1773897.pdf
- Transcription start site has to be verified → +1
- +1 Start to the start codon
- Region upstream: Promotor (negative value)

Basic Promoter — Eubacteria



- Recognized by the house keeping sigma factor
  - E. coli σ^{70}
- Specific sequence
  - -10 — TATAAT
  - -35 — TTGACA
  - Spacing — 17 +/- 1 bp

Alternate Sigma Factors



Promoter

5'UTR

5′ 3′

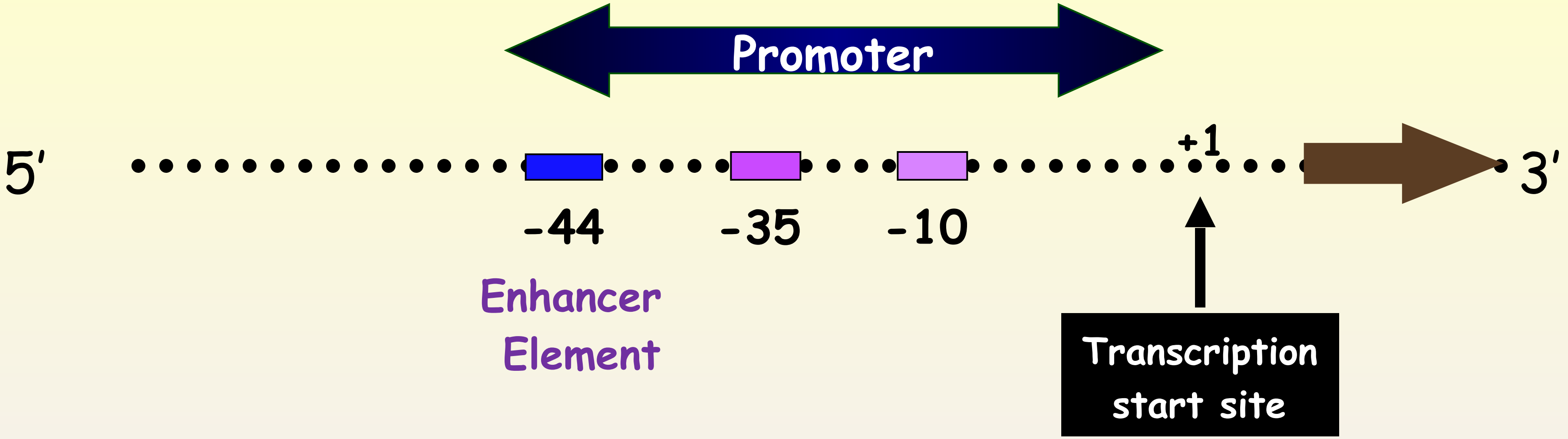
+1

-24 -12

Transcription start site

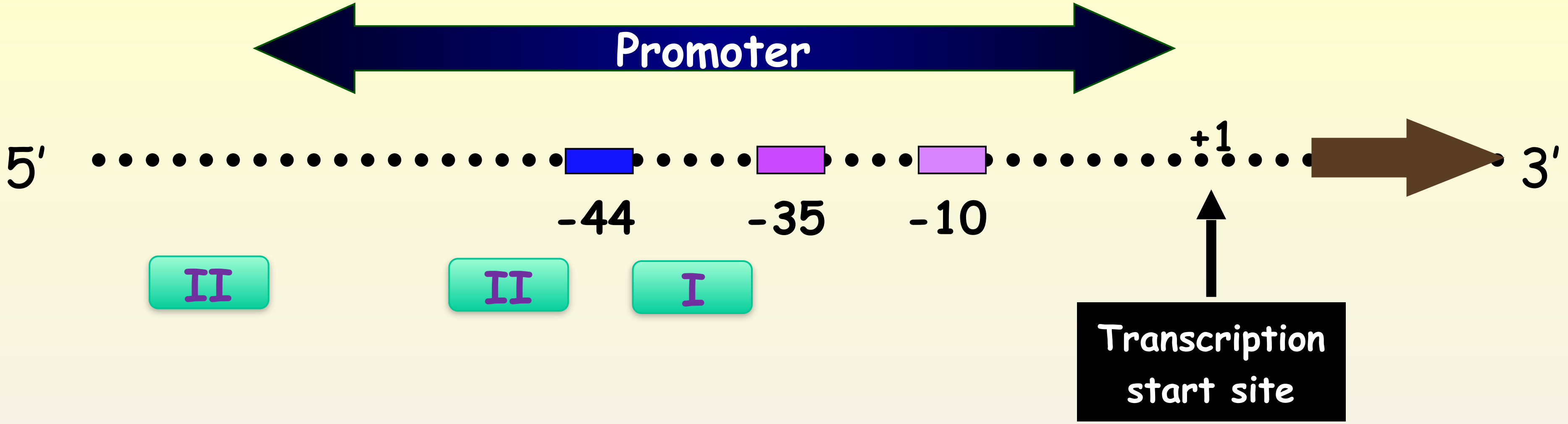
- Many in each bacteria
- Different -10 and -35
- Different location altogether
- RpoN or σ^{54}
  - -12 and -24

Promoter — Eubacteria



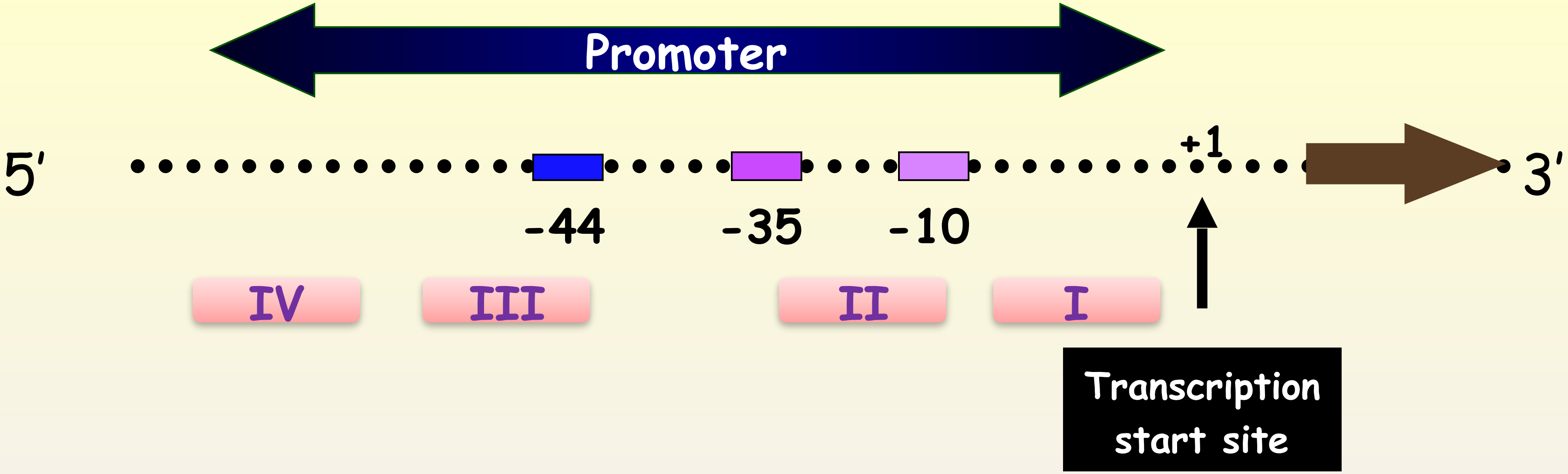
- ❑ Enhancer element – promotes transcription
- ❑ Highly expressed genes

Activator Binding Sites



- Activators —
  - Positive regulators
  - Activate transcription
  - Sequence varies with the regulators
- Modulators

Repressor Binding Sites

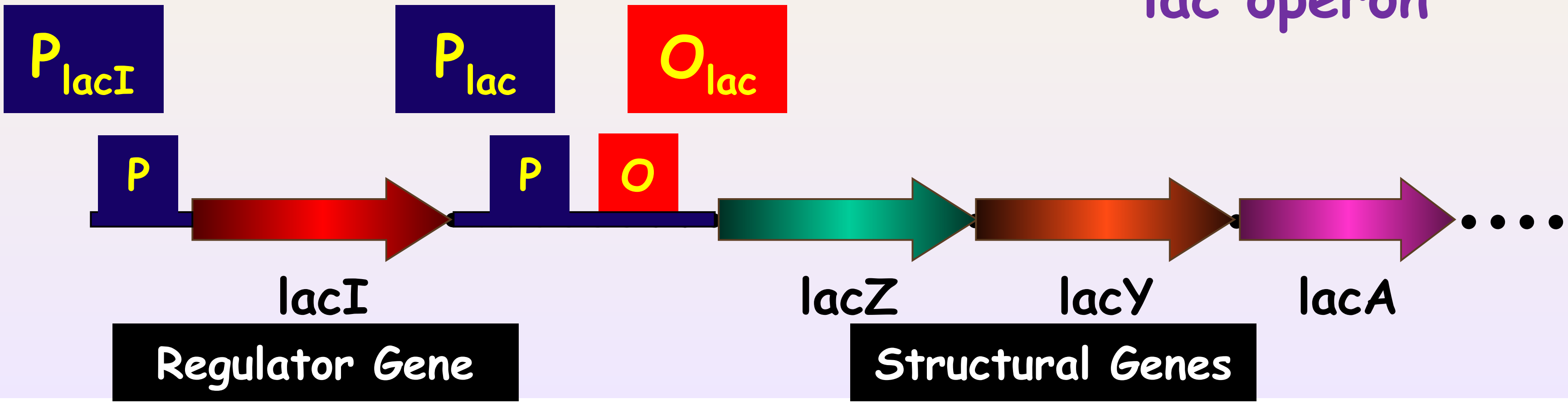


- Repressors
  - Negative regulators
  - Prevent transcription
- Operator sites

Regulatory Elements

- Genes
  - Regulator
  - Structural
- Cis Elements
  - Promoters
  - Operators
- Trans Elements/Transacting factors
  - Diffusible
  - Repressors
  - Activator

lac operon



P_{lacI} P_{lac} O_{lac}

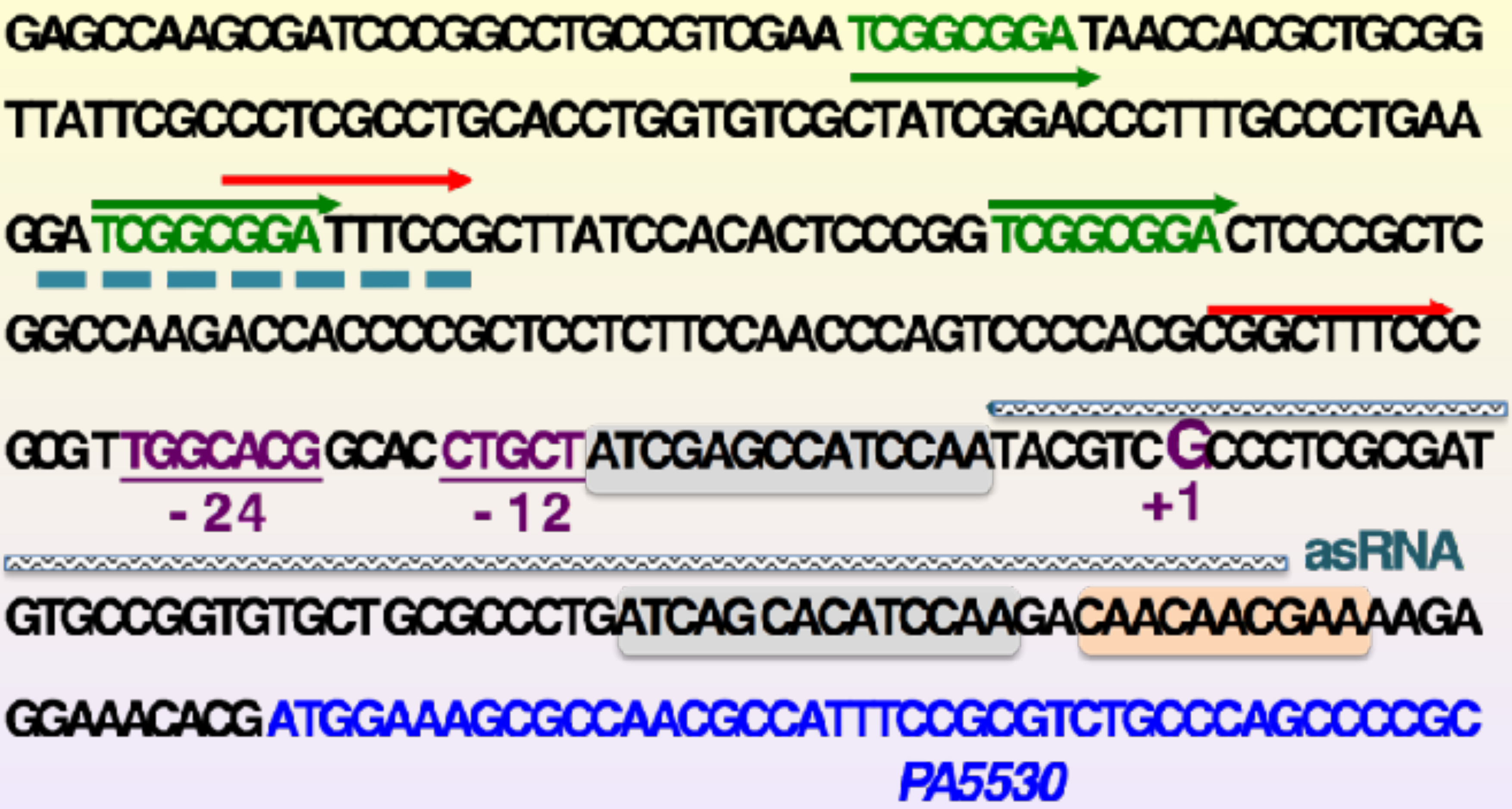
P P O

lacI lacZ lacY lacA

Regulator Gene

Structural Genes

Promoter — Eubacteria



GAGCCAAGCGATCCCGGCCTGCCGTCGAA TCGGCGGA TAACCACGCTGCGG

TTATTCGCCCTCGCCTGCACCTGGTGTCGCTATCGGACCCTTTGCCCTGAA

GGA TCGGCGGA TTTCCGCTTATCCACACTCCCGG TCGGCGGA CTCCCGCTC

GGCCAAGACCACCCCGCTCCTCTTCCAACCCAGTCCCCACGCGGCTTTCCC

GCGT TGGCACG GCAC CTGCT ATCGAGCCATCCAATACGTC G CCCTCGCGAT

- 24 - 12 +1

asRNA

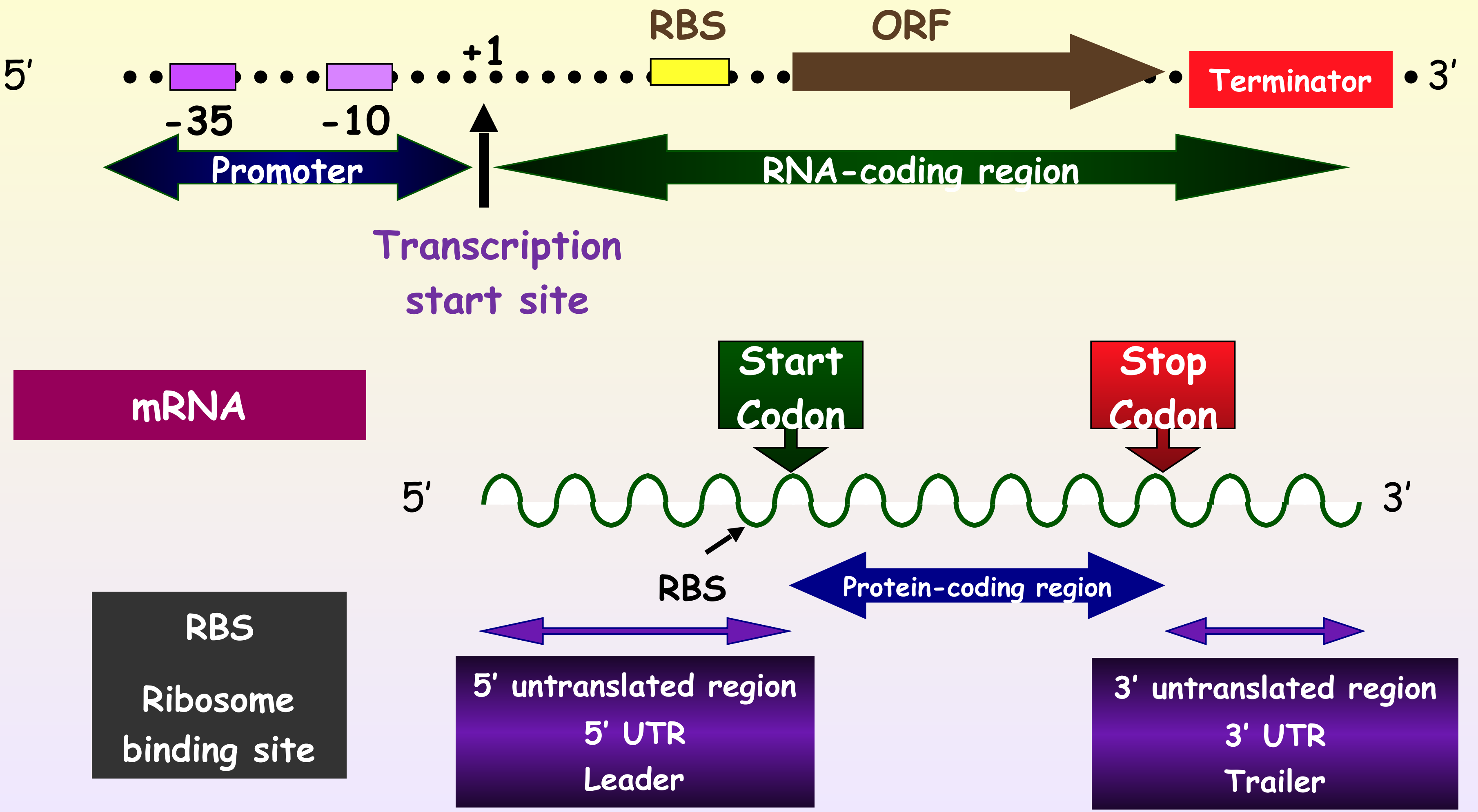
GTGCCGGTGTGCT GCGCCCTGATCAG CACATCCAAGACAACAACGAAAAGA

GGAAACACG ATGGAAAGCGCCAACGCCATTTCCGCGTCTGCCCAGCCCCGC

PA5530

Mature RNA — Eubacteria

Transcriptional unit



5′ 3′

RBS

ORF

+1

-35

-10

Terminator

Promoter

RNA-coding region

Transcription start site

mRNA

Start Codon

Stop Codon

5′ 3′

RBS

Protein-coding region

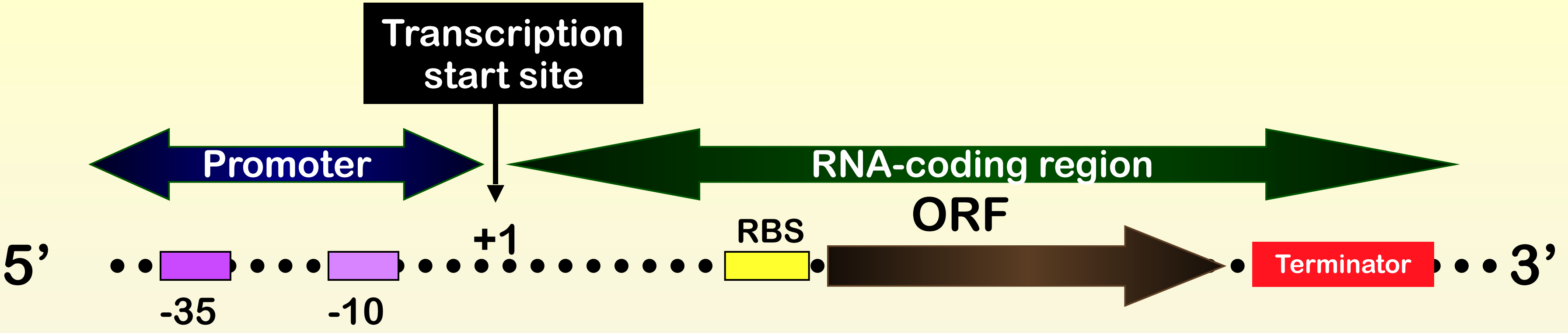
RBS

Ribosome binding site

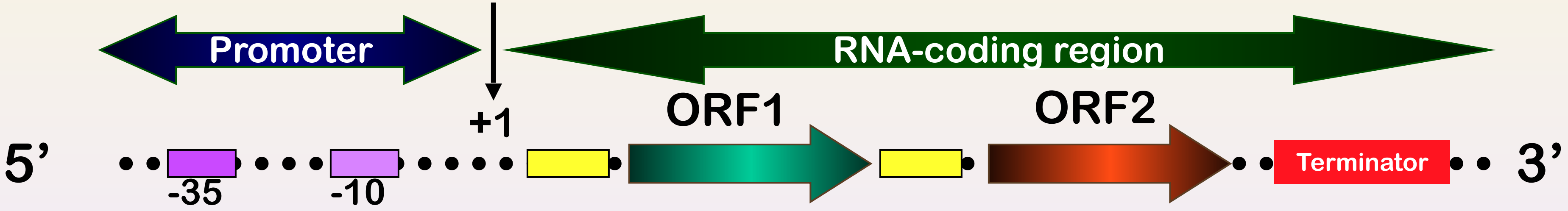
5′ untranslated region
5′ UTR
Leader

3′ untranslated region
3′ UTR
Trailer

Cistron — Mono vs Poly



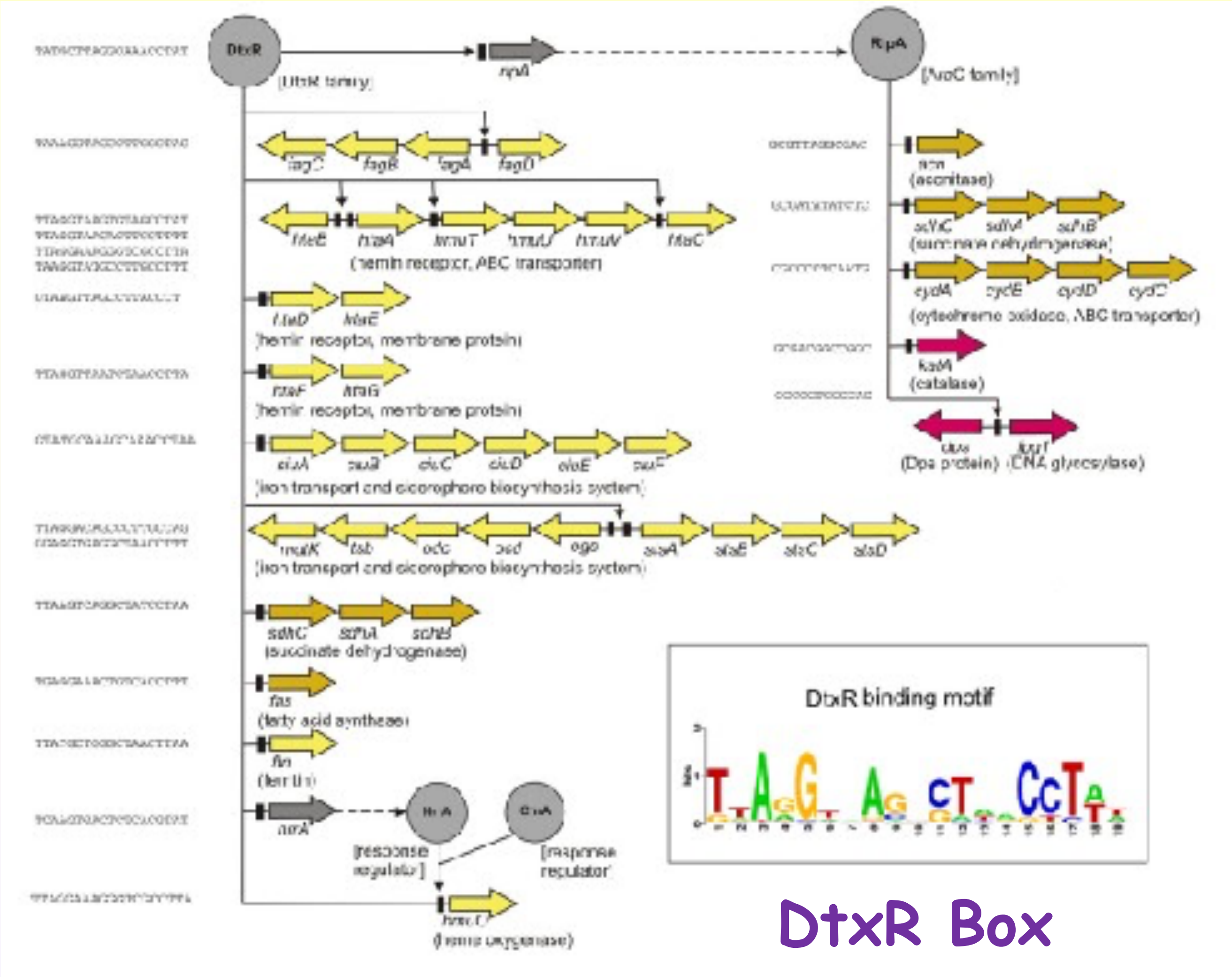
- ❑ Transcriptional unit - Single gene operon



- ❑ Transcriptional unit - Two-gene operon
- ❑ Under the control of single promoter

Regulon

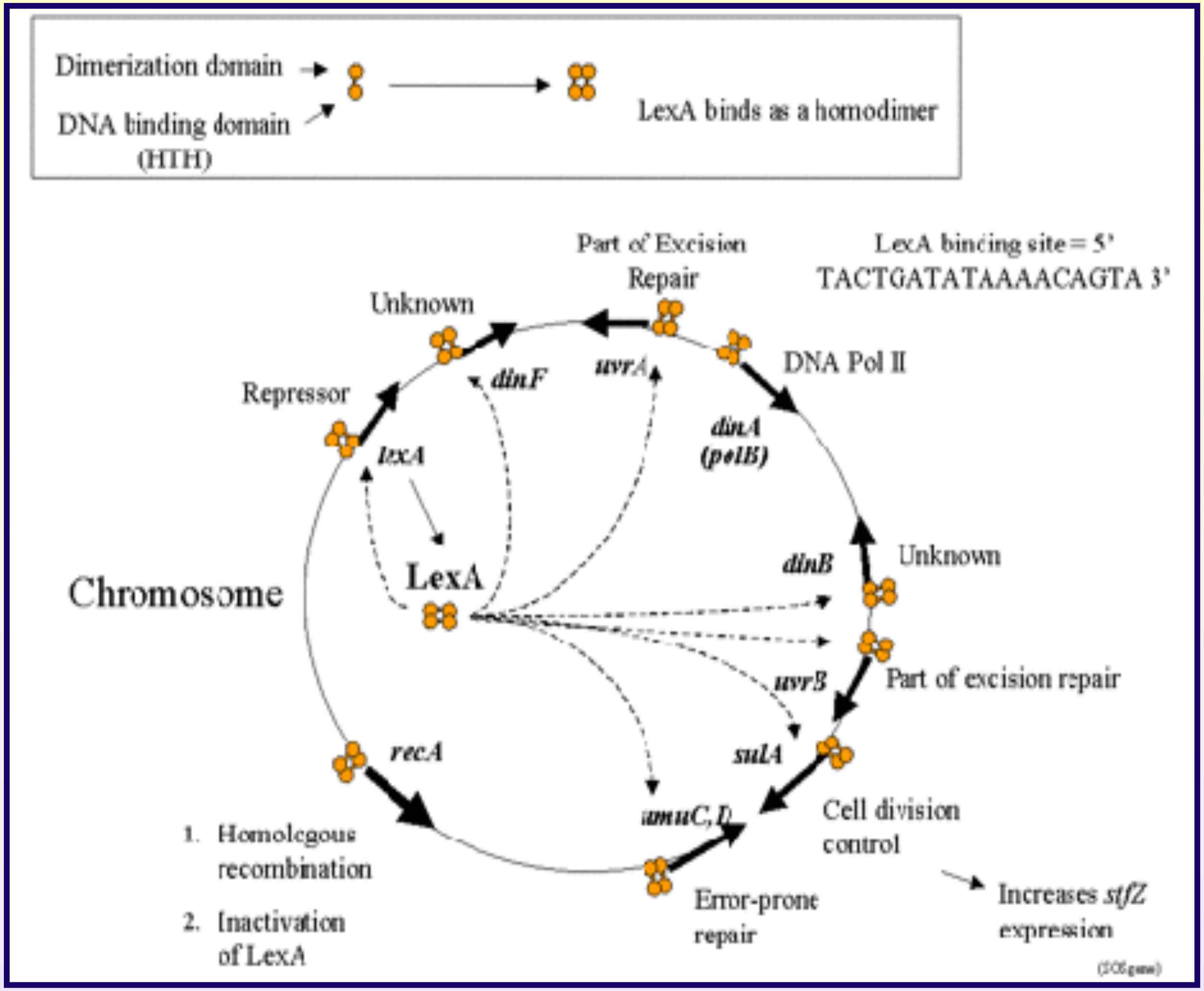
- A group of operons regulated as a unit
- Modulator
  - Repressor
  - Activator
- Consensus binding site



DtxR Box

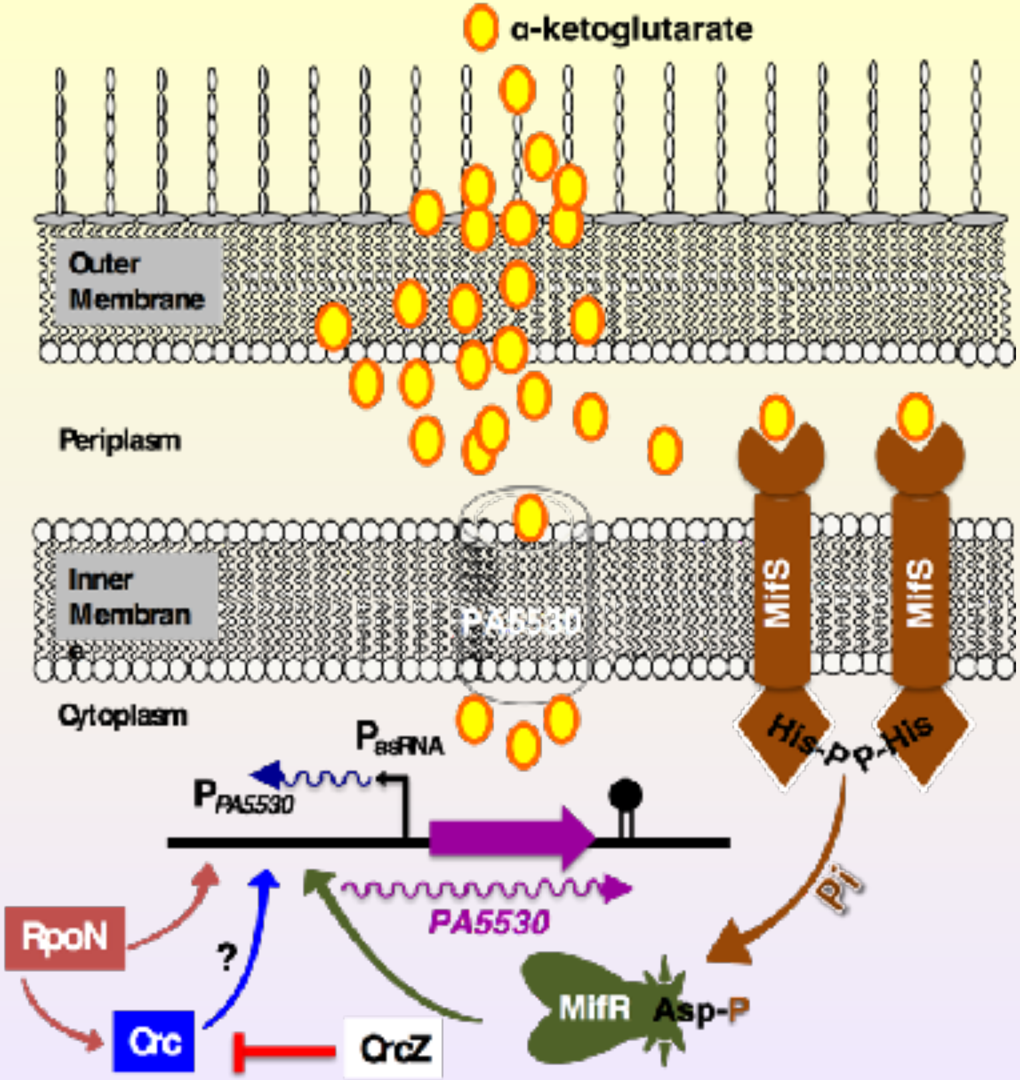
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3022926/

LexA Regulon



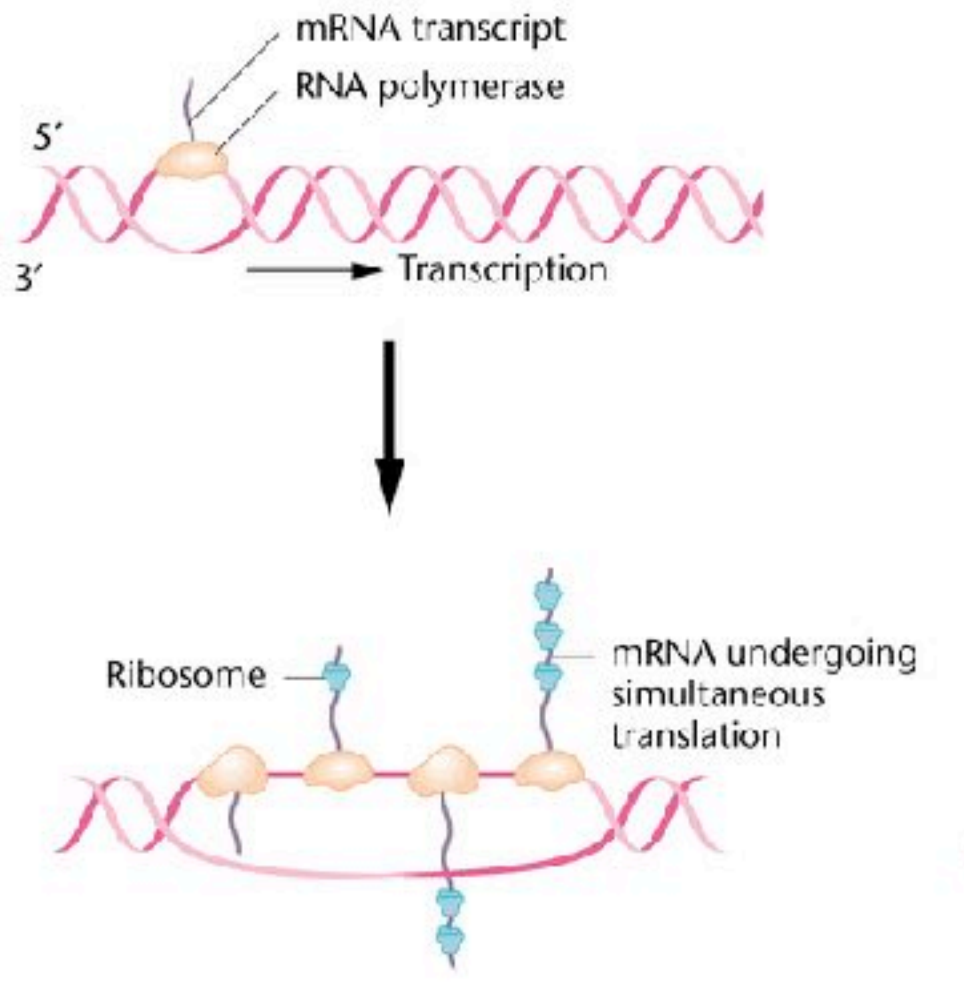
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3022926/

Complex Regulation



Tatke, Mathee et al 2017

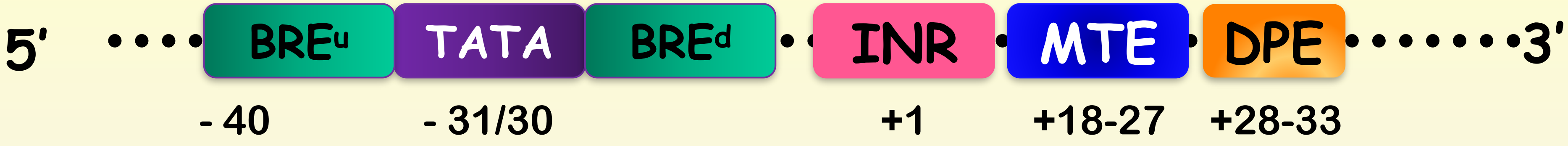
Coupling


mRNA transcript
RNA polymerase
5′
3′
Transcription
Ribosome
mRNA undergoing simultaneous translation

Eukaryotic RNAPs

- Three: RNAPI, RNAP II and RNAP III
- Multisubunit
- RNAP I
  - Exclusive to ribosomal RNA (28S, 18S & 5.8S)
- RNAP II
  - Most of the genes in eukaryotes
  - Respond to environmental cues
- RNAP III
  - Short non-coding RNAs such as tRNAs, 5S rRNA, U6 snRNA & a limited number of others

Core Promoter — Eukaryote



- ☐ TATA box
- ☐ BRE - Transcription factor IIB Recognition Element
- ☐ INR — Initiator
- ☐ MTE — Motif Ten Element
- ☐ DPE — Downstream Promoter Element

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1219314/
https://epd.vital-it.ch/index.php

Basic Promoter — Eukaryote

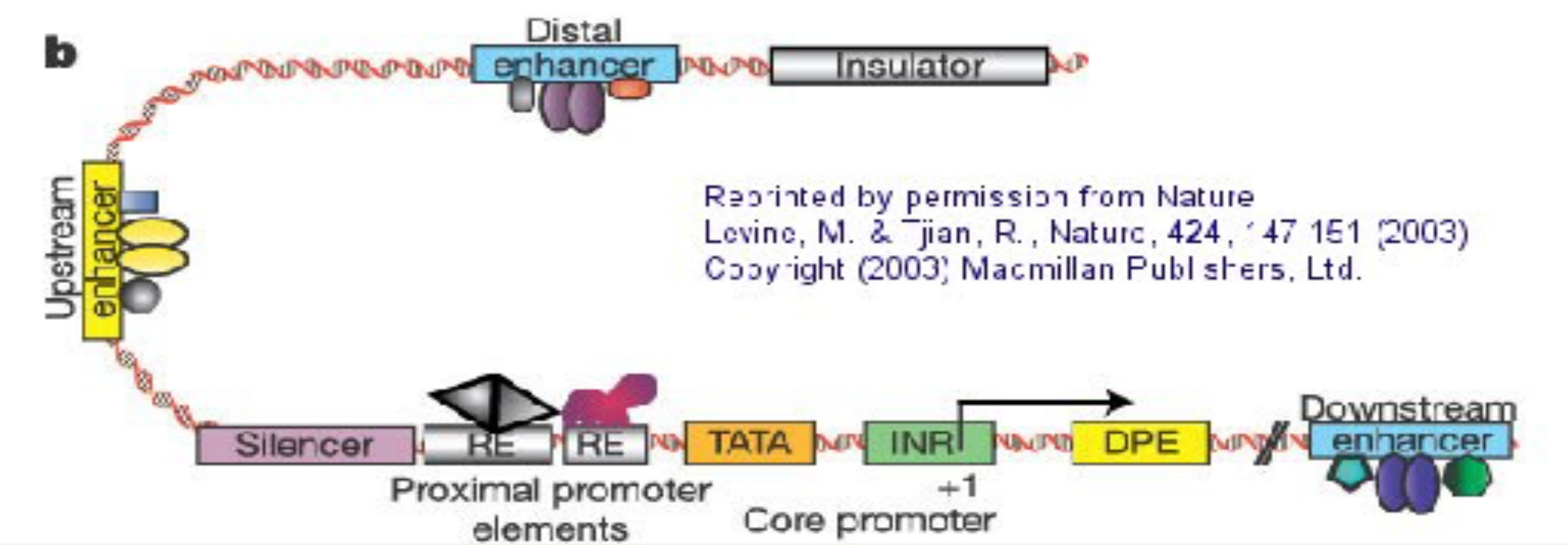

5' Silencer // BRE^u TATA BRE^d INR MTE DPE 3'

- 40   - 31/30   +1   +18-27   +28-33

- TATA box
- UAS — Upstream activator sequence
- Silencer
  - ± 100-200 bp upstream of TATA box
  - Position independent

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1219314/
https://epd.vital-it.ch/index.php

Complex Promoter — Eukaryote

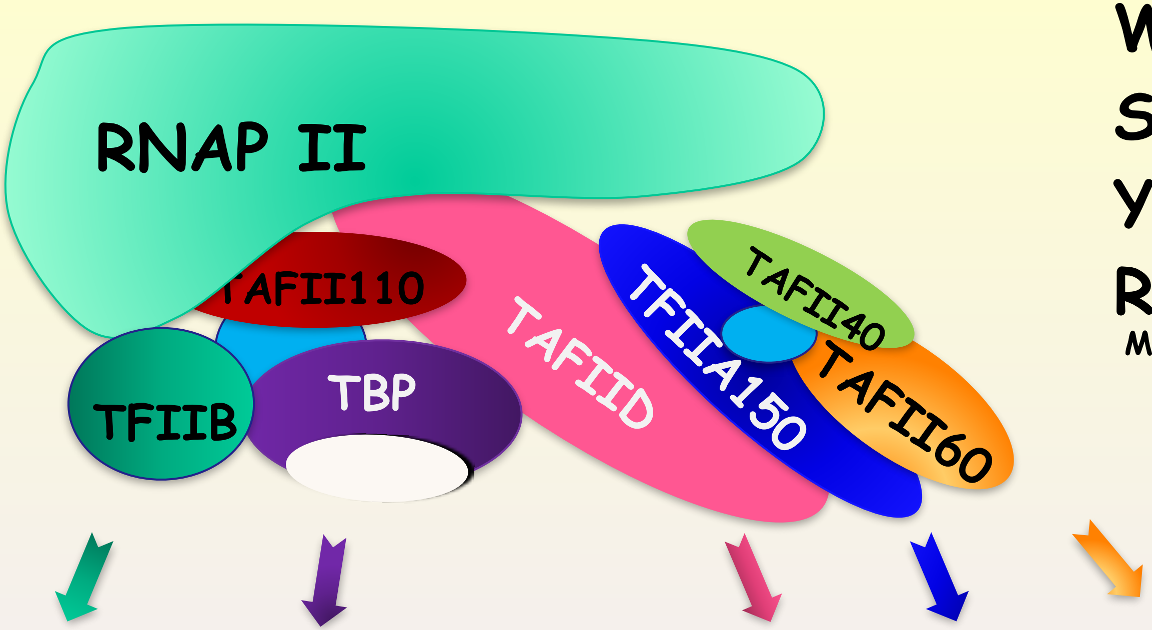


Reprinted by permission from Nature
Levine, M. & Tjian, R., Nature, 424, 147 151 (2003)
Copyright (2003) Macmillan Publishers, Ltd.

- Silencer & Insulator — 10 to 50 kb US/DS
- NRE — Negative Regulatory Sequence
  - Position-dependent

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2830304/ http://genesdev.cshlp.org/content/16/3/271.full

RNAPII — Core Promoter



W — A & T
S — C & G
Y — C & T
R — A & G
MTE – motif ten element

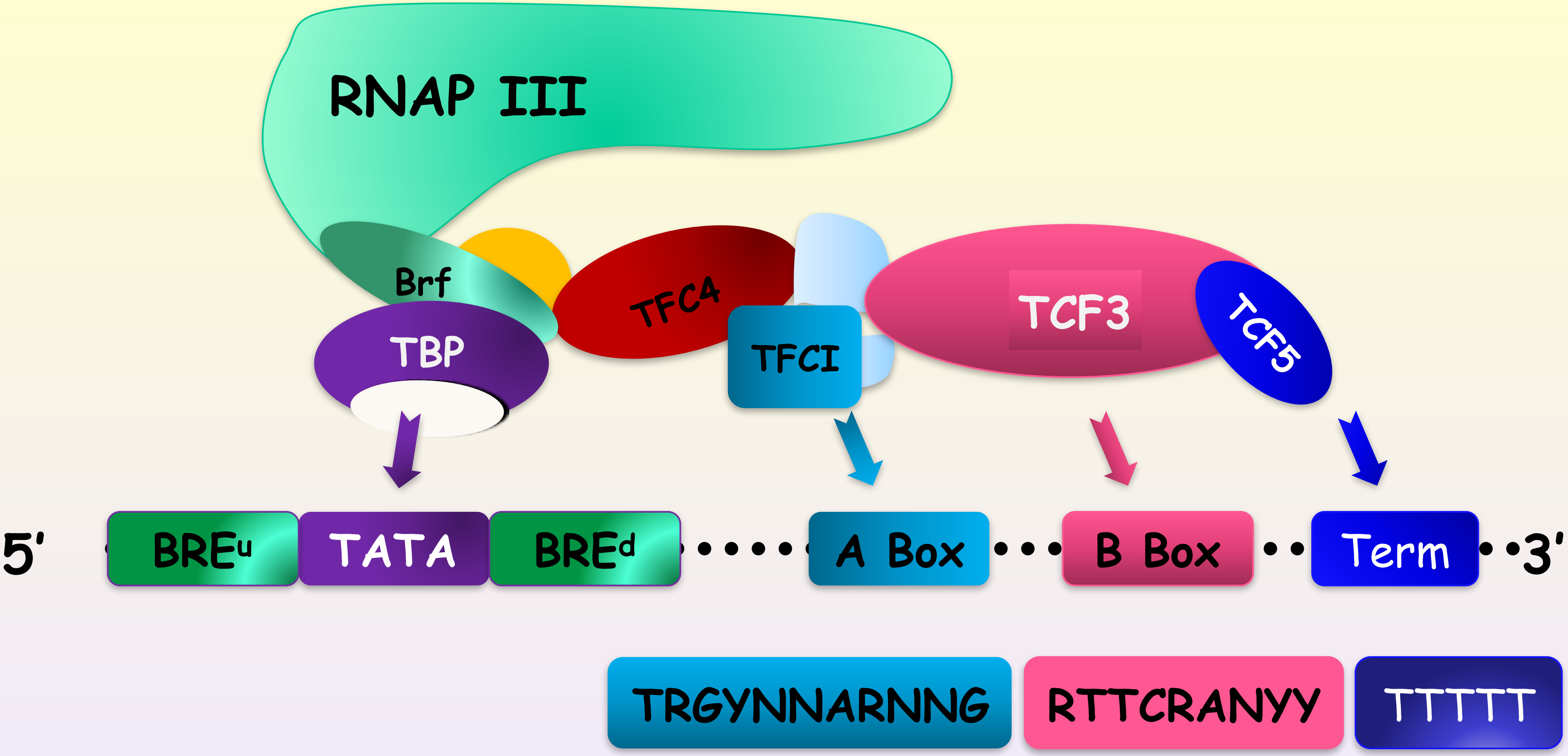
5' ••••BRE^u TATA BRE^d •• INR • MTE • DPE •••••••3'

- 31/30 +1 +18-27 +28-33

SSRCGCC TATAWAAR RTDKKK YYANWYY CGANCCGG WYGT

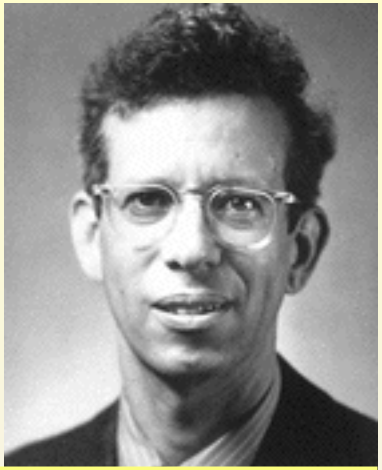
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4214234/
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2830304/

RNAPIII — Promoter



https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4214234/

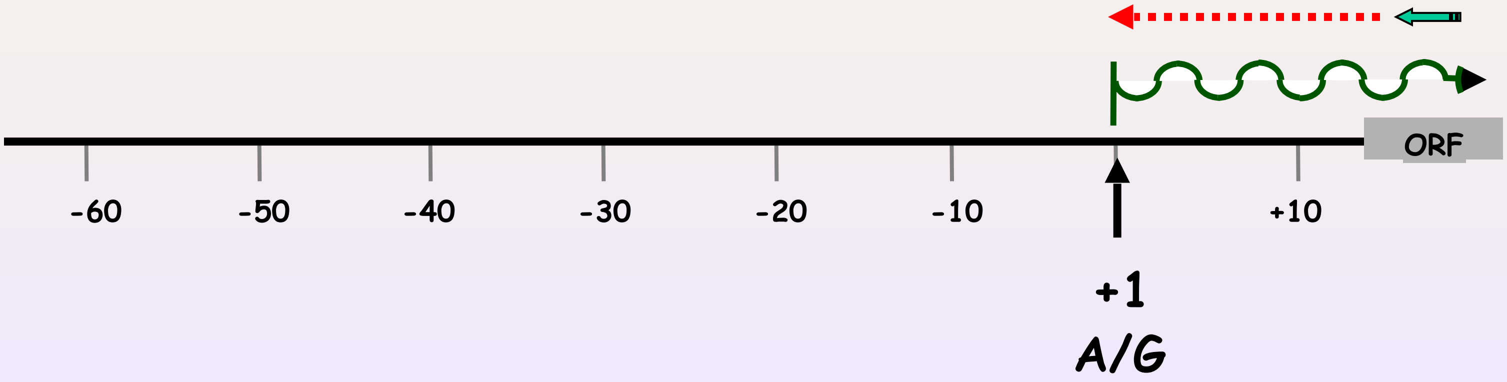
Promoter Start Sites



Howard M Temin
(1934 - 1994)

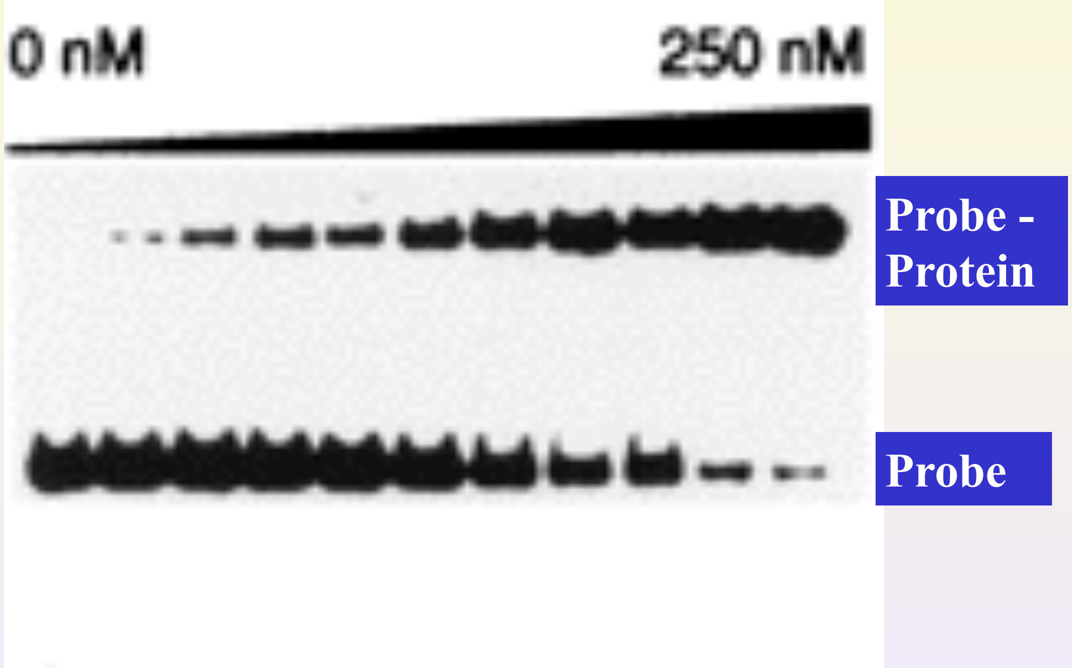
David Baltimore
(1938 -)

- Nobel Prize — 1975 Med
- Reverse Transcriptase
- Primer Extension Analysis
- RNA Sequencing
- Initiated vs processed RNA



Binding

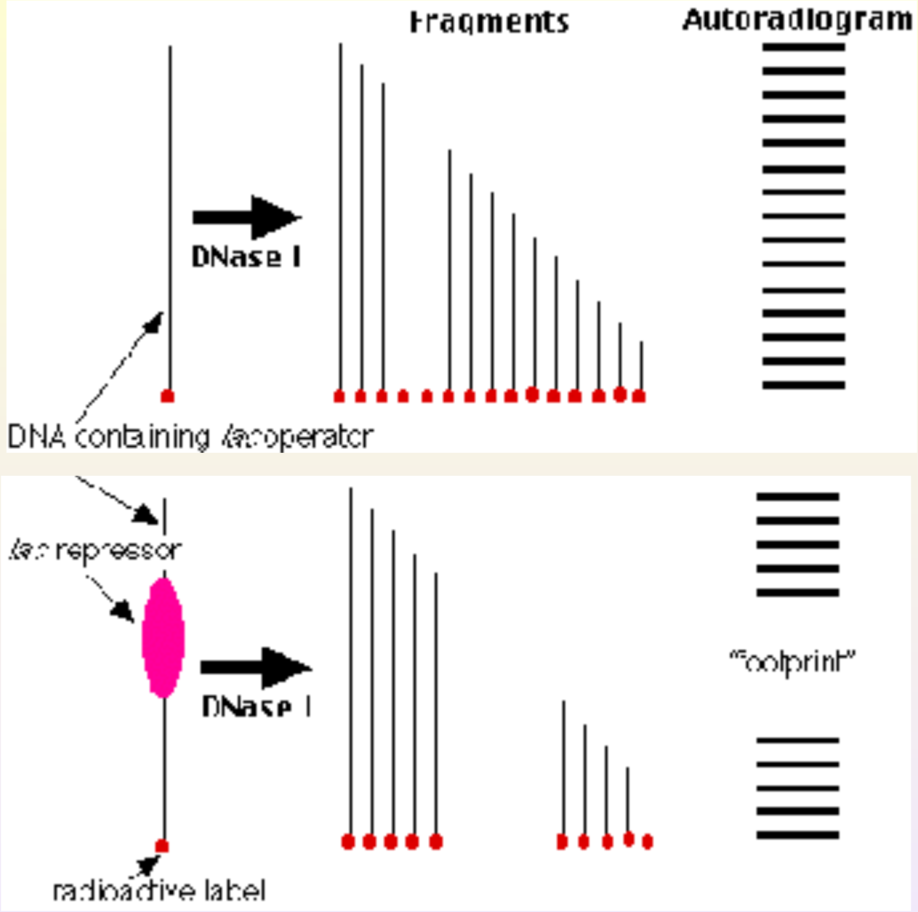
- Gel Shift Assay
  - Retardation assay

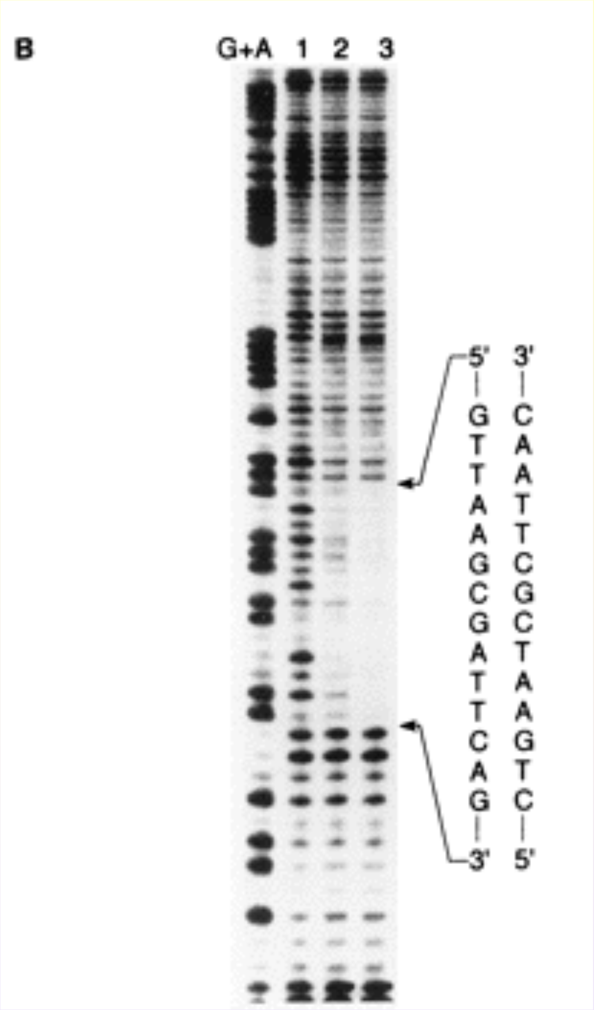


Cra-Dependent Transcriptional Activation of the *icd* Gene of *Escherichia coli* J. Bacteriol. 1999 181: 893-898

Binding Sequence

- **Footprinting**

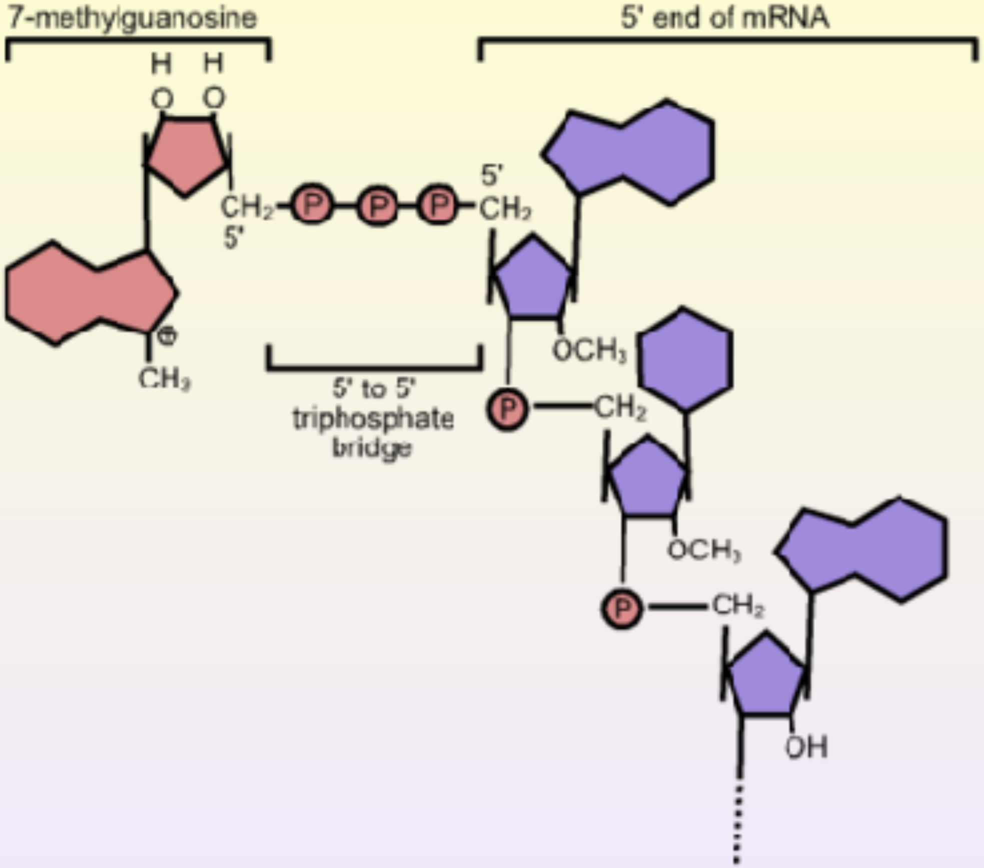




Cra-Dependent Transcriptional Activation of the *icd* Gene of *Escherichia coli* J. Bacteriol. 1999 181: 893-898

mRNA Capping

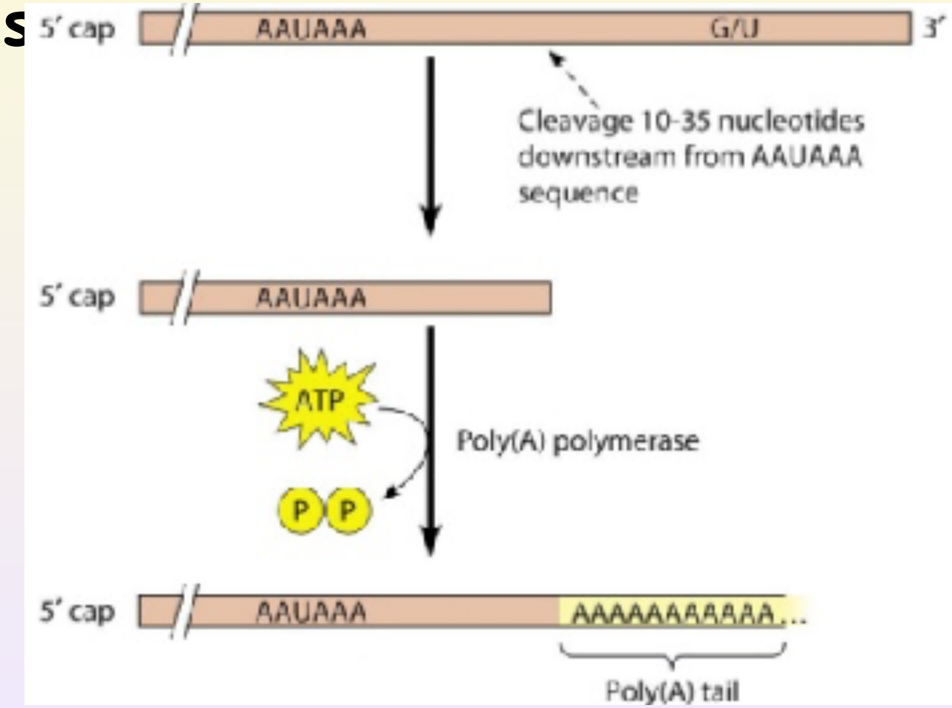
- *Capping occurs during transcription*



https://en.wikipedia.org/wiki/Five-prime_cap

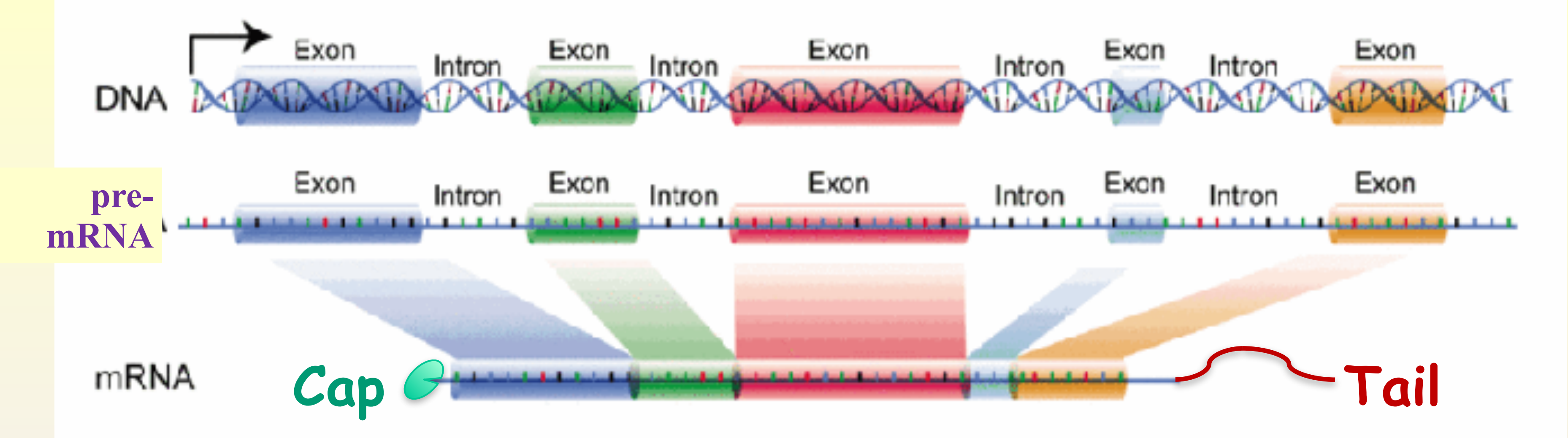
PolyA-Tailing

- ☐ Adding of string of As to the tail of RNA
  - https://www.youtube.com/watch?v=DoSRu15VtdM
- ☐ It can go up to ~250 As
- ☐ Template independent
- ☐ Prior to splicing



https://www.slideshare.net/DeeshmaKp/8-rna-types-85710809

RNA Processing



- mRNA is capped and tailed
- Spliced
  - Spliced junction promotes diversity in some genes
- mRNA moves out of nucleus to cytoplasm

https://fr.m.wikipedia.org/wiki/Fichier:DNA_exons_introns.gif

Why Regulate?

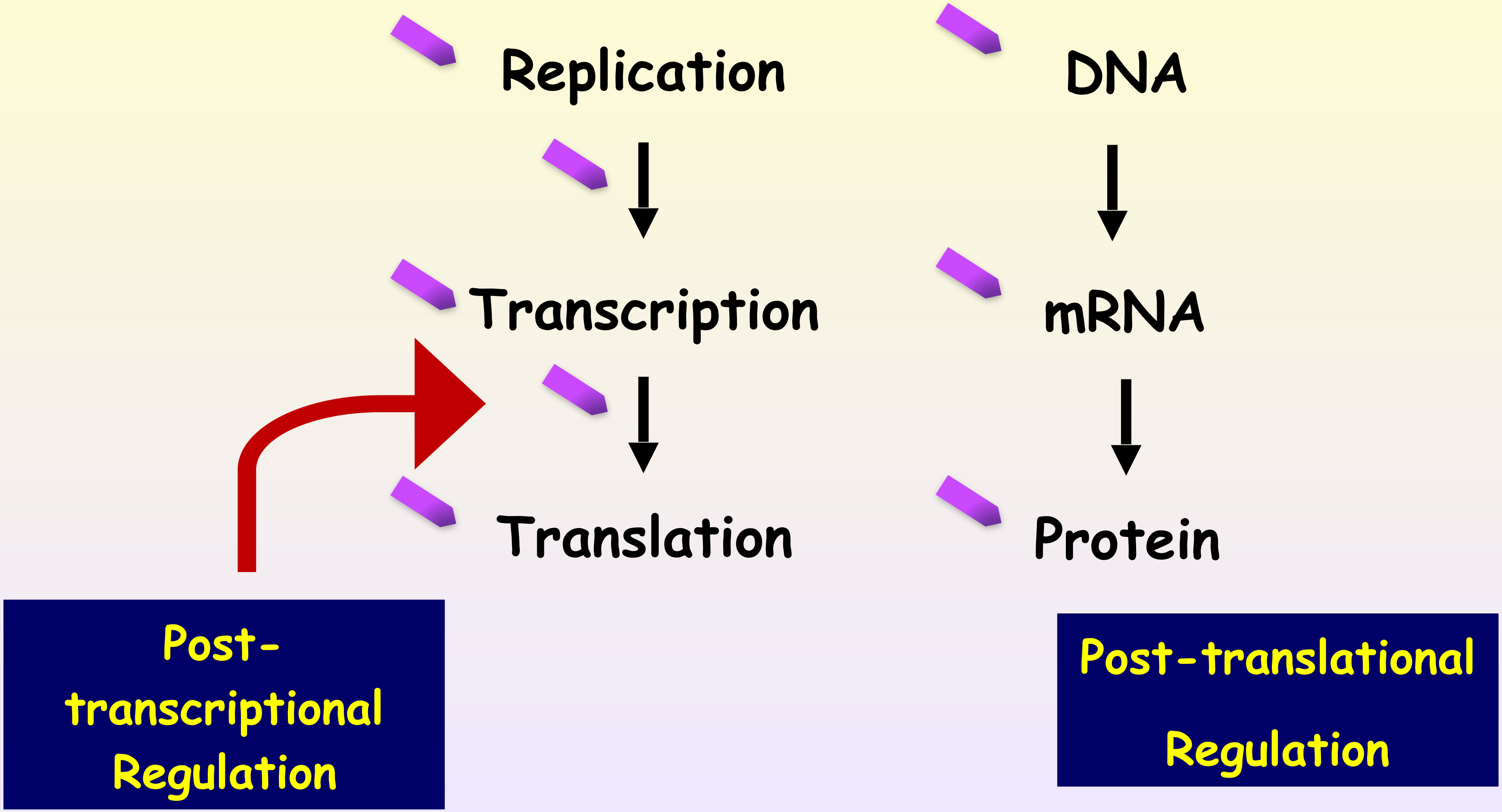
- Control the quantities of gene products
    - Some are needed in large quantities: Ribosomes
    - Some are needed in small quantities: Enzymes
- Synthesize when the need arises (inducible)
    - Respond to environment, eg., heat shock, UV radiation
- Need for temporal expression
    - Phage development
    - Growth
- Internal flexibility
    - Provides adaptability

Types of Regulation

- Constitutive
  - Unregulated (housekeeping genes)
- Inducible
  - Basal levels → Induced in response
- Repressible
  - Normal levels → Repressed in response to signal
- Temporal
  - Growth, phage
- Coordinated Regulation
  - Several products that act in a single pathway

Points of Regulation

From genotype to phenotype



- Replication
- Transcription
- Translation

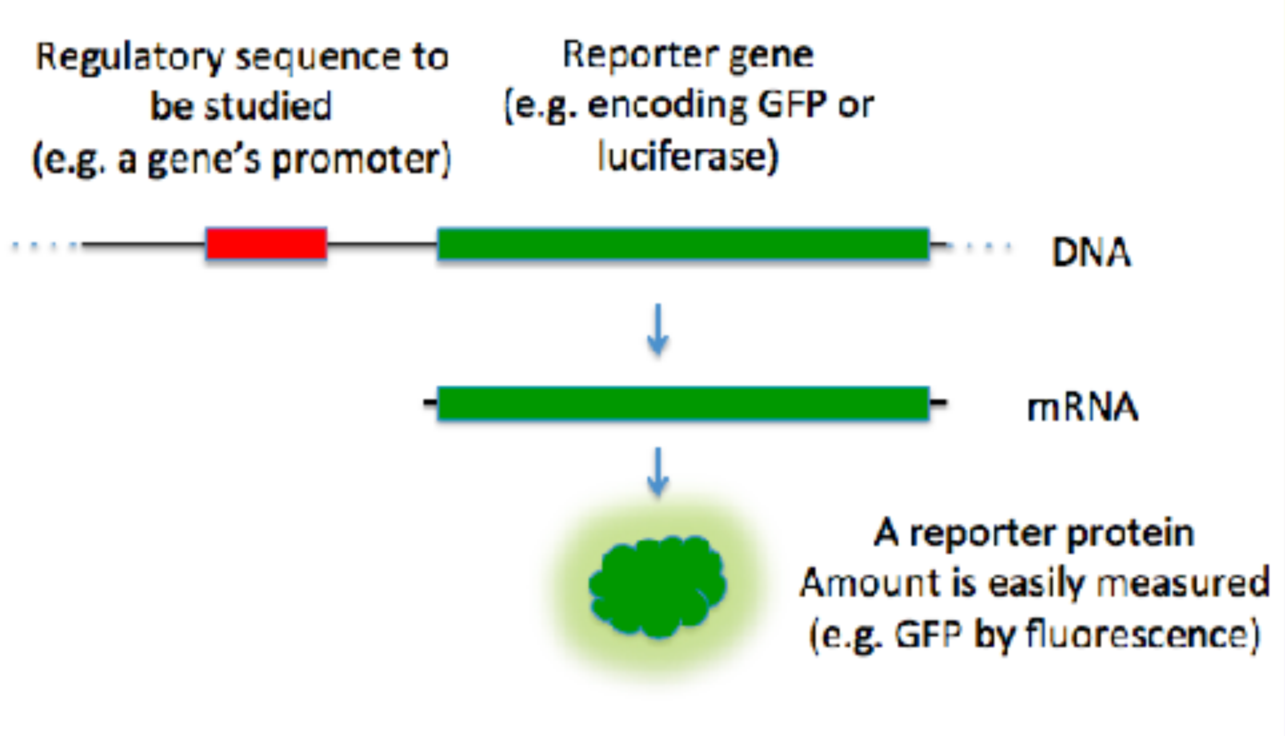
- DNA
- mRNA
- Protein

Post-transcriptional Regulation

Post-translational Regulation

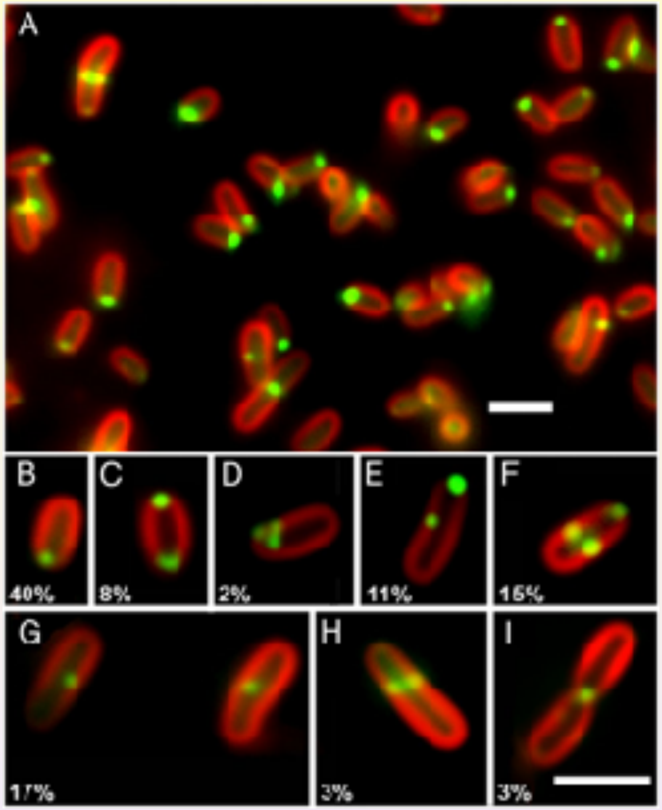
Gene Expression

- Promoter fused with a reporter
  - Reporters: lacZ, lux, GFP etc



https://en.wikipedia.org/wiki/Reporter_gene

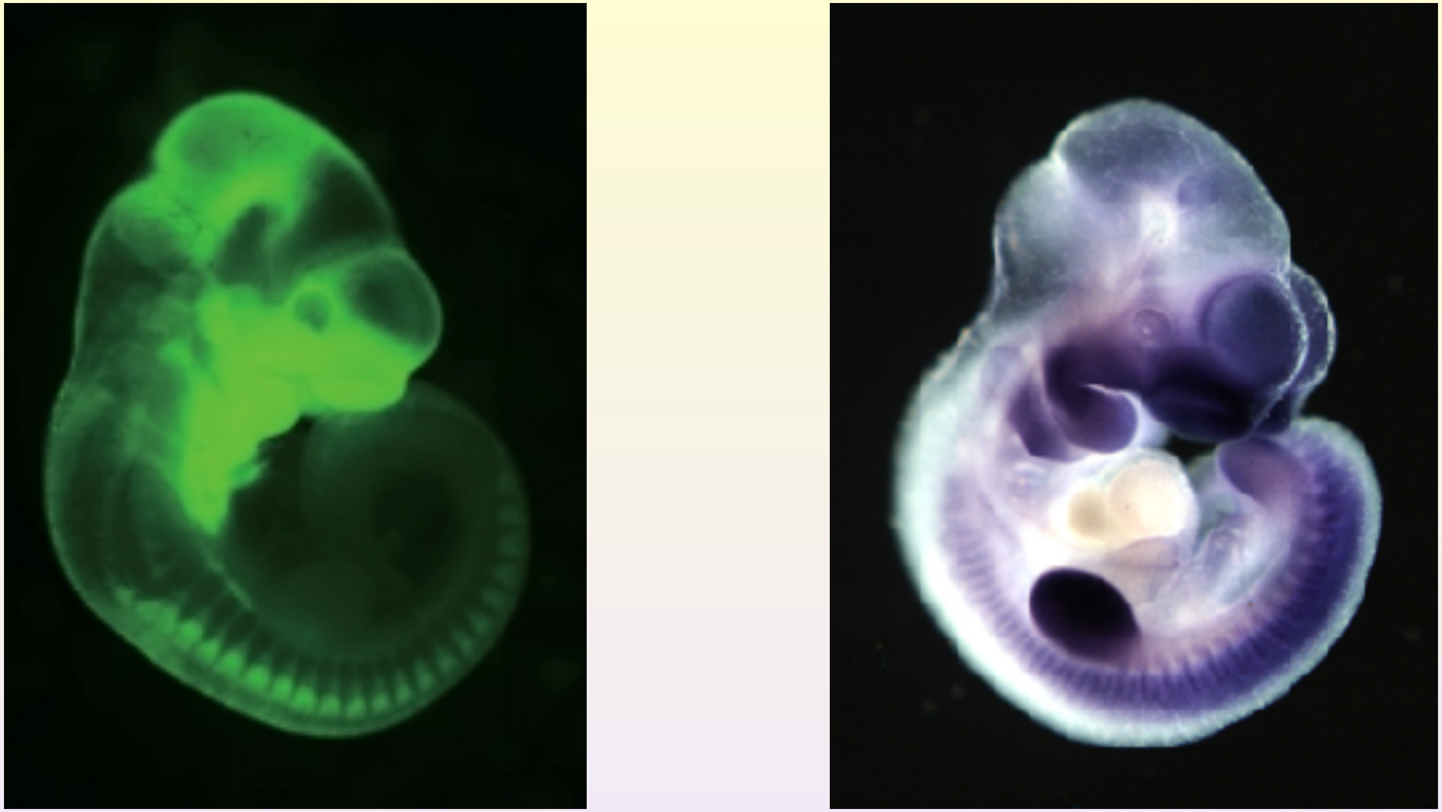
Gene Expression in E. coli



https://fphoto.photoshelter.com/image/I00005Qdkpj7izeM
http://www.pnas.org/content/110/22/9060
https://www.sigmaaldrich.com/technical-documents/articles/biology/blue-white-screening.html

Gene Expression in Embryo

- ❑ **Wnt - reporter**



https://www.tcd.ie/Zoology/research/groups/murphy/WntPathway/tcf3.php#tcf3e105
https://www.researchgate.net/figure/a-Neural-crest-lineage-labeling-mouse-Wnt1-Cre-GFP-clearly-demonstrates-green_fig1_295883105
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4104936/ & https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3967567/ - Reporters

Whole Genome Expression

- Microarray
  - A grid of DNA segments of known sequence
  - Test if any of them are expressed by hybridizing with RNA
  - Genome/platform-dependent

  https://www.youtube.com/watch?v=pWk_zBpKt_w
  https://www.youtube.com/watch?v=xoxUWGl8WFs

- RNA Sequencing
  - Genome/platform-independent
  - Mostly indirect — cDNA
  - Direct — Nanopore sequencing

  https://www.nature.com/articles/nature08390
  https://www.ncbi.nlm.nih.gov/pubmed/28334071

Whole Genome Expression

- To be continued……………

Session Learning Objectives

- Central Dogma
- Transcriptional Unit
  - Prokaryotic
    - Operon
    - Promoter
    - Operator
    - Enhancer etc
  - Eukaryotic
    - TATA Box
    - UAS
    - Silencer
    - Insulator etc
- Gene Regulation
  - Single
  - Multigene
    - Regulon
- Gene Expression
- Tools to study gene expression
  - Single gene
  - Multi gene
  - Genome-dependent
  - Genome-independent