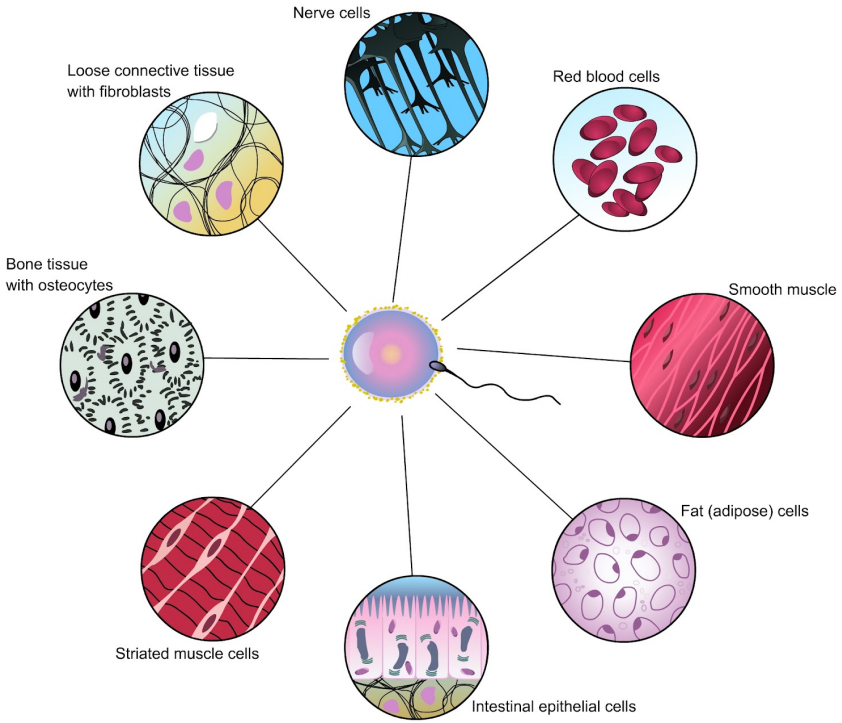
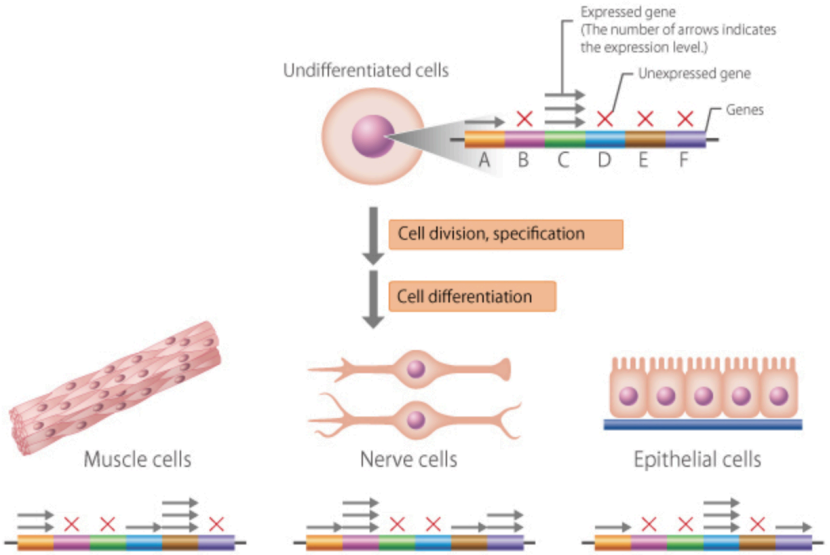
Principles of Gene Expression

I. Introduction

Regulation of gene expression is intrinsically linked to multicellularity and cell-specification in higher eucaryotes





II. Principles of gene regulation

Principle #1 : constitutive vs regulated gene expression

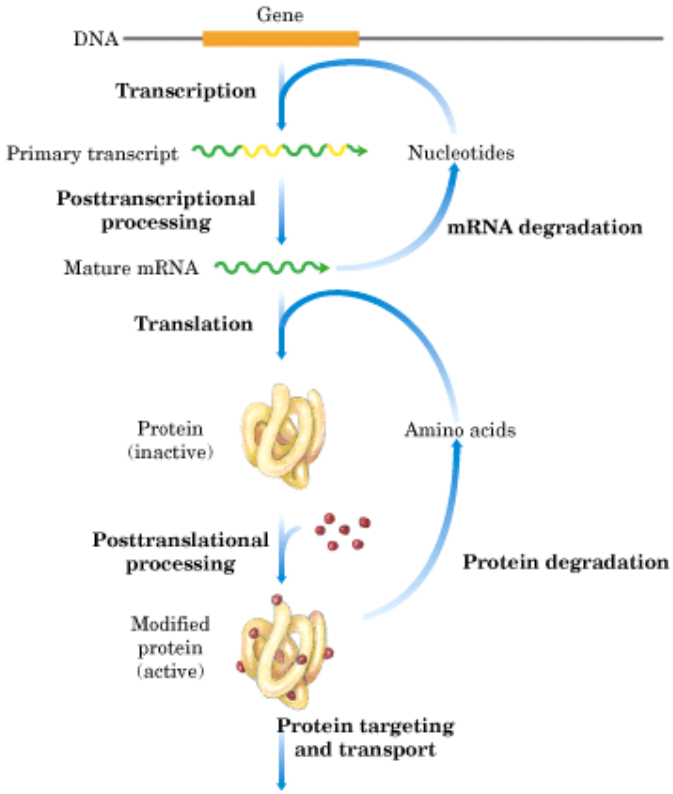
There are approx 10.000 genes expressed in an eukaryotic cell
(4.000 in yeast)

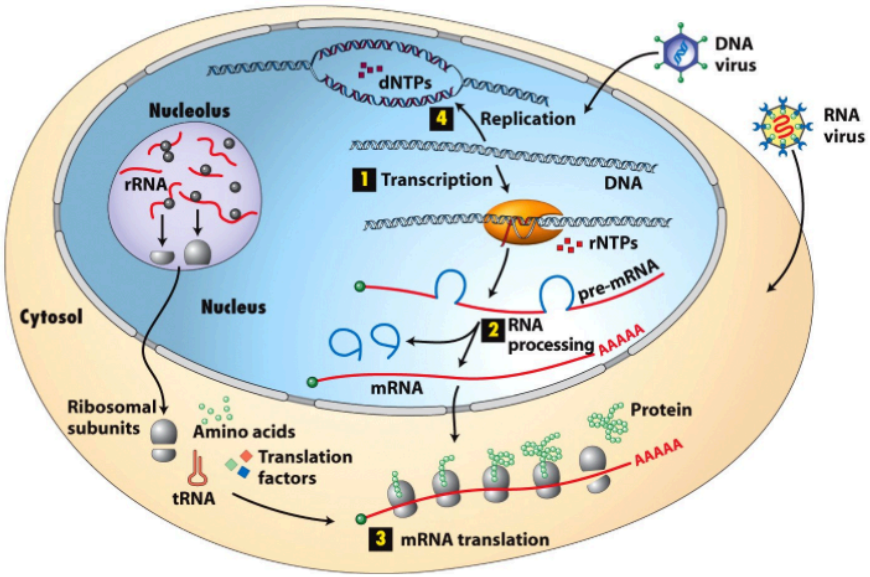
Most of the genes code for polypeptides that are required at all times at a more or less constant level, expressed at low levels, common to different cell types (*housekeeping genes*)

A minority of mARNs are expressed at high levels, they are specific to the cell type or conditions (*luxury genes*)

II. Principles of gene regulation

Principle #2 : Gene regulation can be acheived at several levels along polypeptide synthesis (different in prokaryotes and eukaryotes).





II. Principles of gene regulation

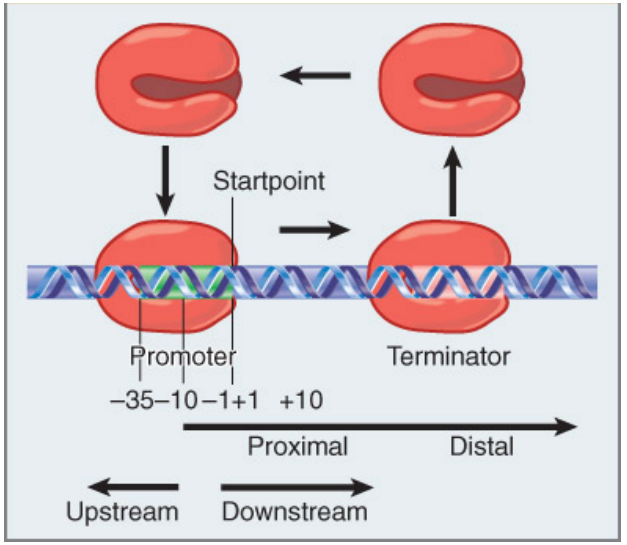
Principle #3 : Transcription proceeds through a series of defined processes : inititation, elongation and termination

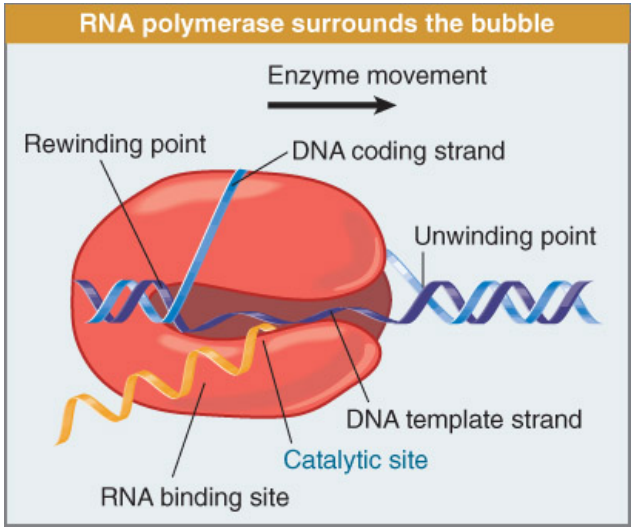
RNA polymerase recognizes **promoter sequences** (**initiation**).

RNA polymerase synthetizes RNA complementary to the template DNA (**elongation**)

RNA synthesis proceeds through terminator sequences (**termination**)

The promoter, the transcribed region and the terminator define a **transcription unit**





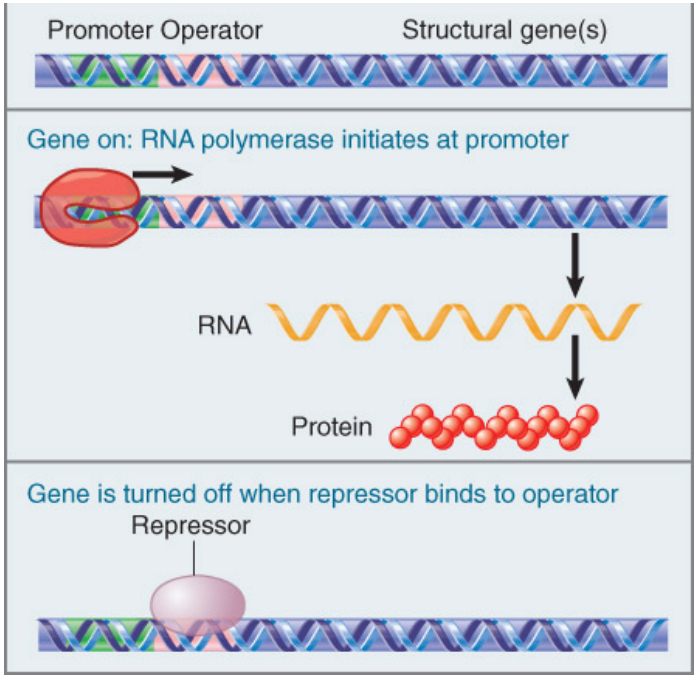
II. Principles of gene regulation

Principe #4 : Transcription initiation is regulated **by trans-acting proteins** (activators or repressors) that bind **to cis-acting sites** in or near promoters

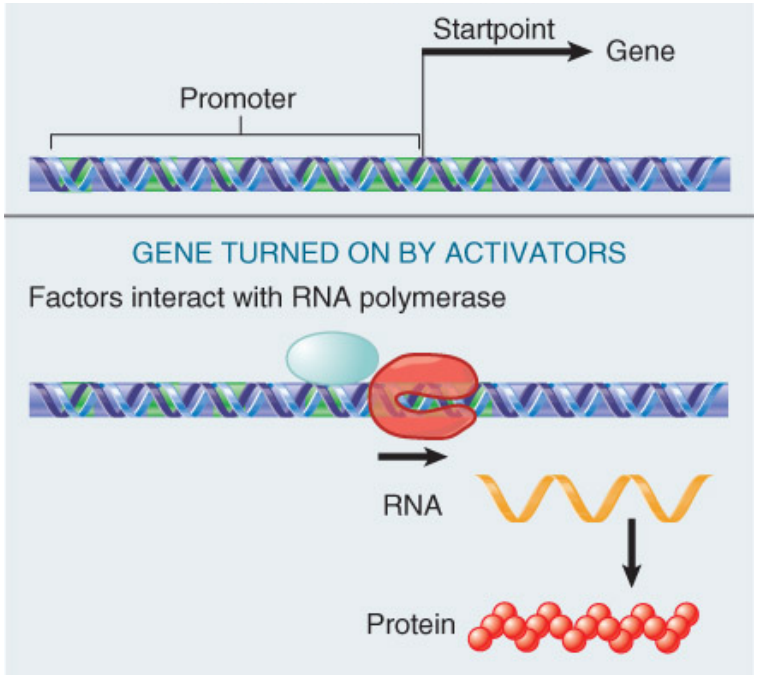
Repressors bind to operators in DNA to prevent or reduces the RNA polymerase activity : (*i.e.*, negative regulation)

Activator binds to DNA sites called enhancer to enhance the RNA polymerase activity. (*i.e.*, positive regulation)

Negative regulation

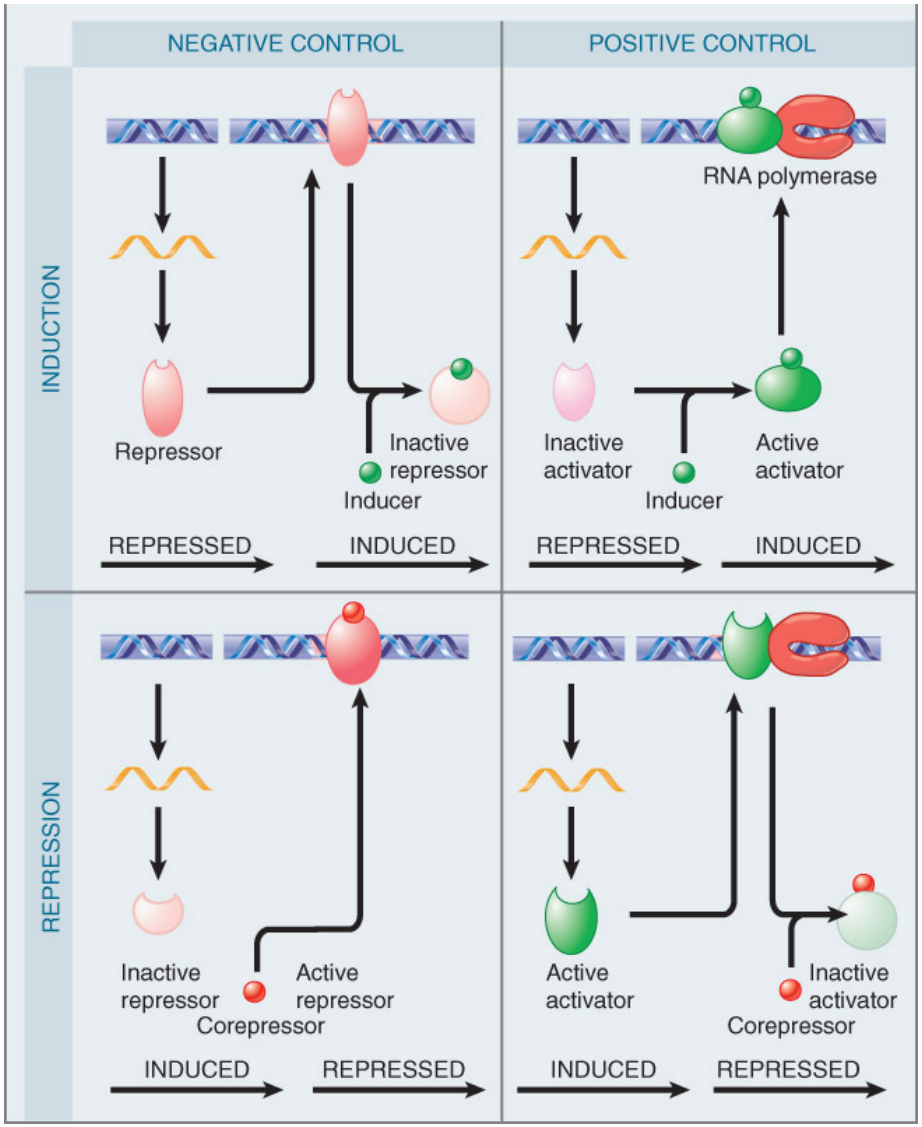


Positive regulation



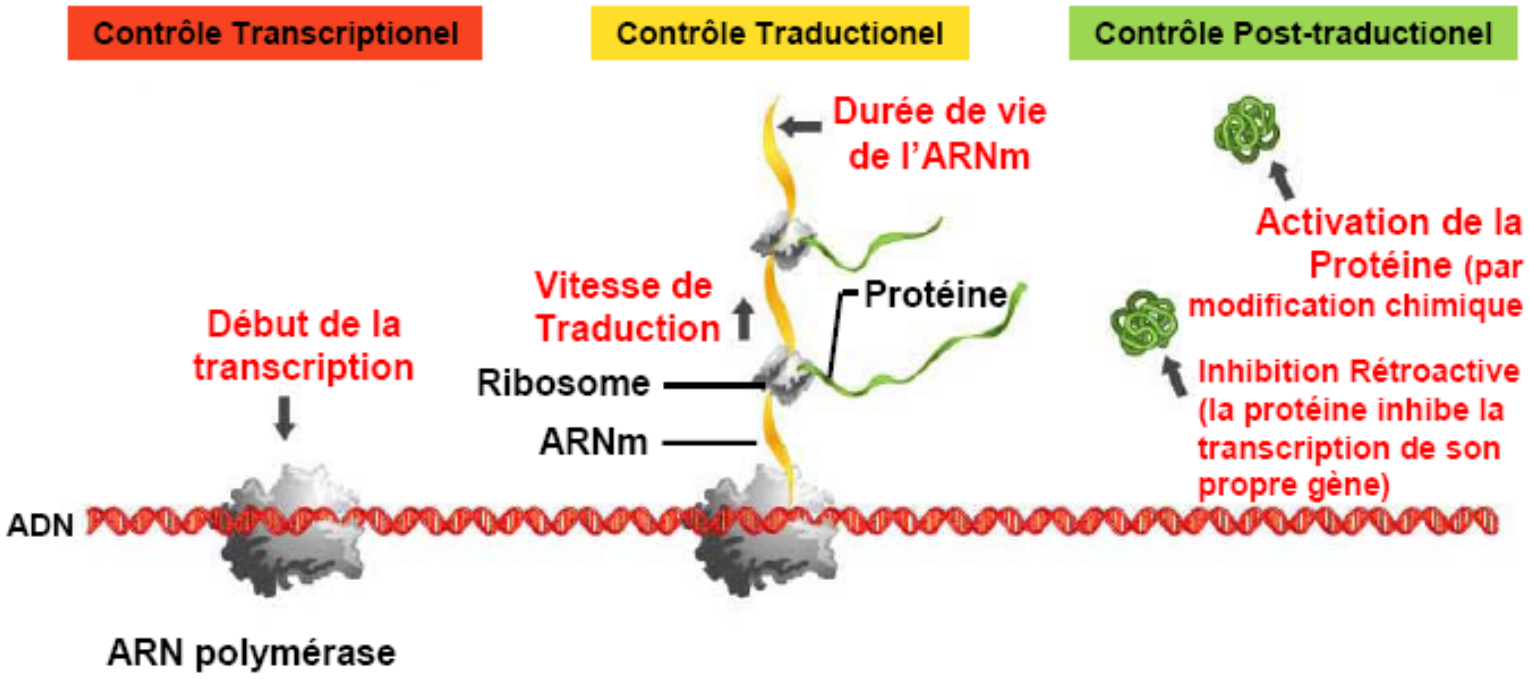
II. Principles of gene regulation

Principle #5 : Activation or repression by activators or repressors can be regulated by **co-regulators (**small molecules in procaryotes**)**



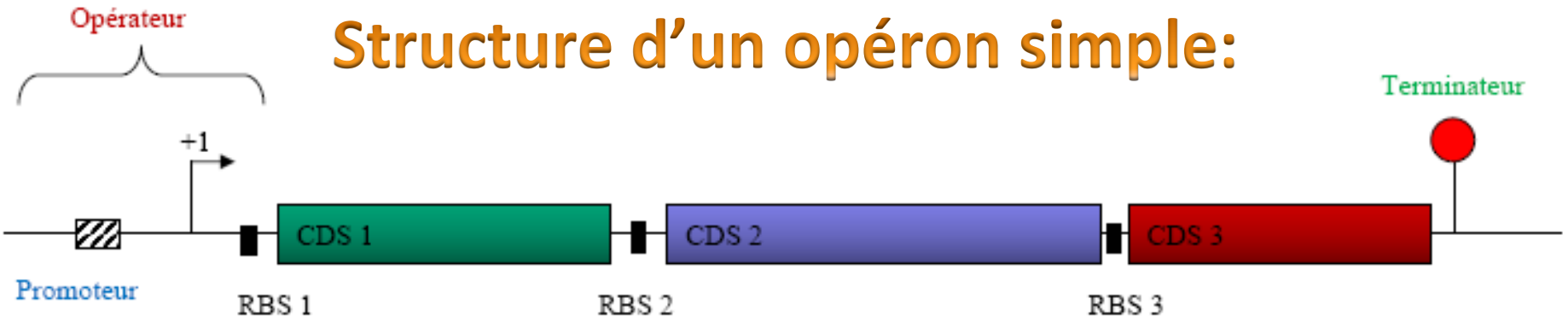
III. Gene Regulation in Prokaryotes

Principle : Gene regulation in Prokaryotes is achieved at the **transcription initiation level**

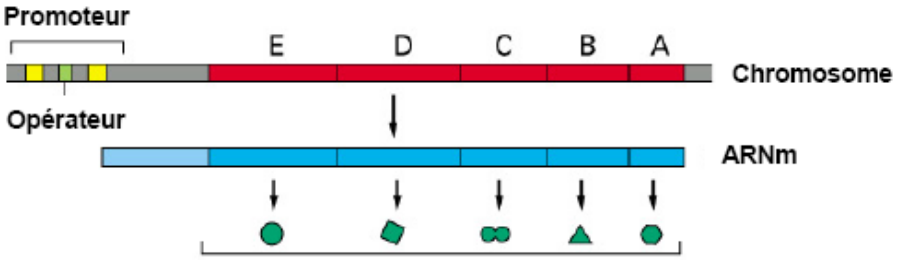


III. Gene Regulation in Prokaryotes

Prokaryotic genes are expressed and regulated **as operons**



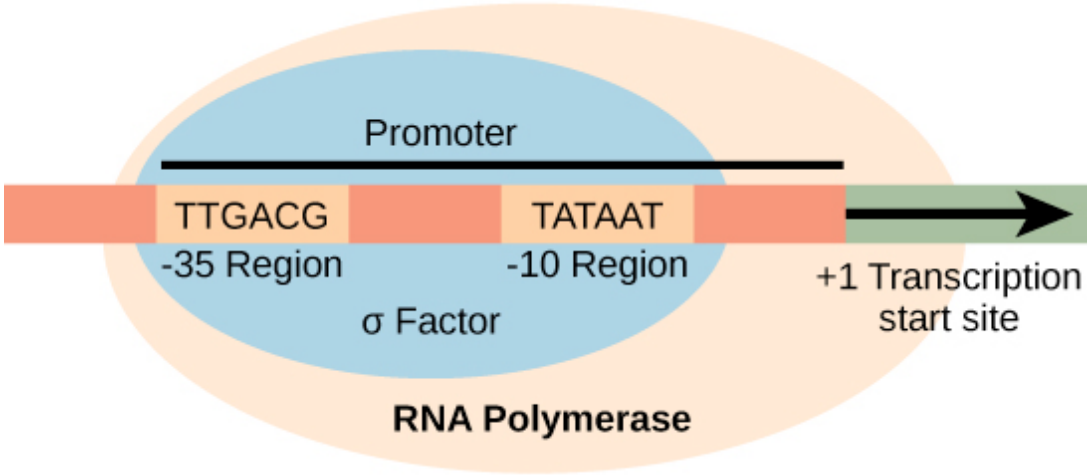
Operator : DNA sequences (binding sites) involved in transcription initiation control
Promoter : binding of RNA polymerase
+1 : transcription start
RBS : ribosome binding site
CDS : coding sequence
Terminator : ends transcription

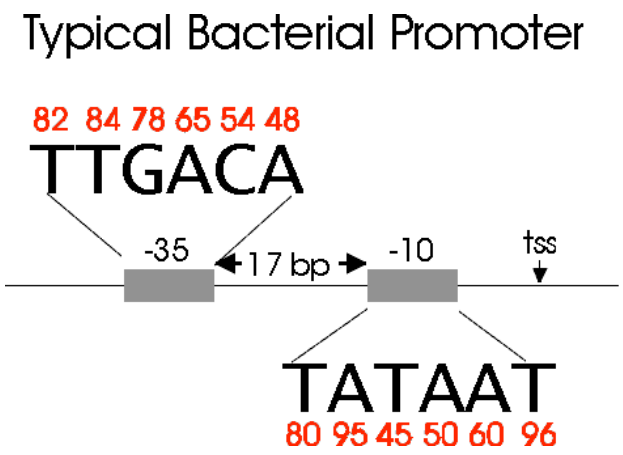


III. Gene Regulation in Prokaryotes

Bacterial promoters are defined by 4 core elements

TSS : transcription start site (purine in >90%)
TATA box at around -10 from the TSS
-35 Box
An insert region between the boxes of 16-18 nucleotide

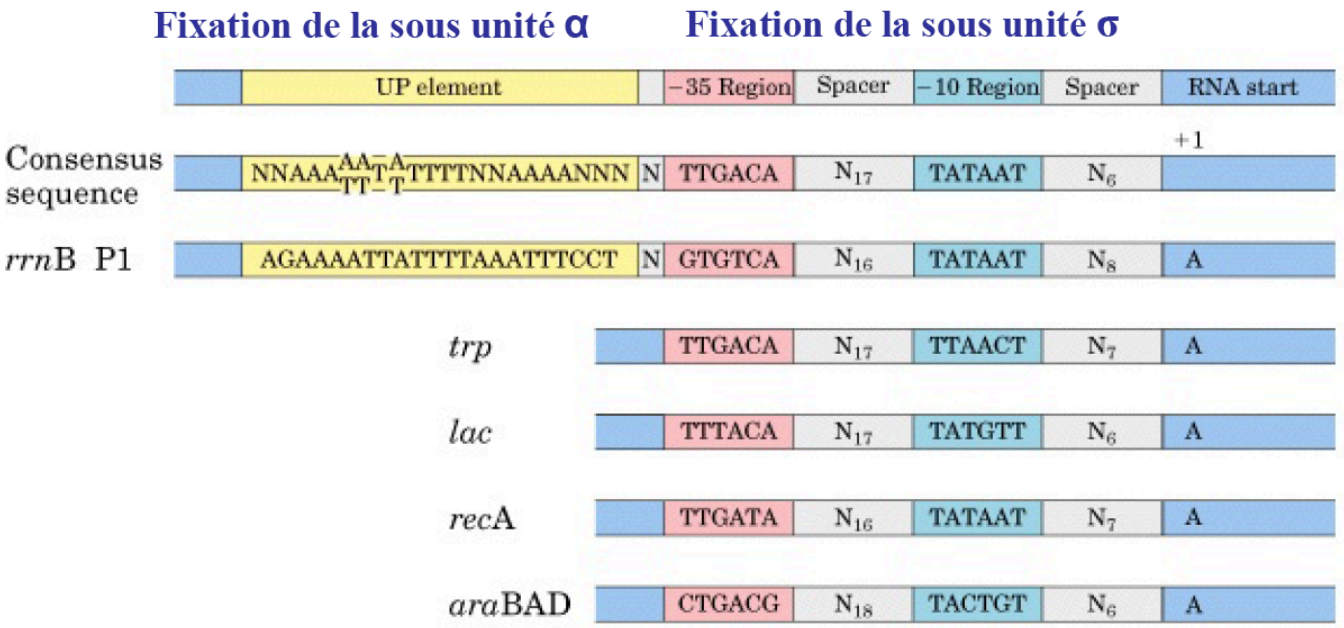




III. Gene Regulation in Prokaryotes

Divergence from canonical sequences and structure defines how efficiently a promoter will initiate transcription (promoter strength : weak or strong).

Mutations in promoter regions can affect promoter strength (mutations up/down)



III. Gene Regulation in Prokaryotes

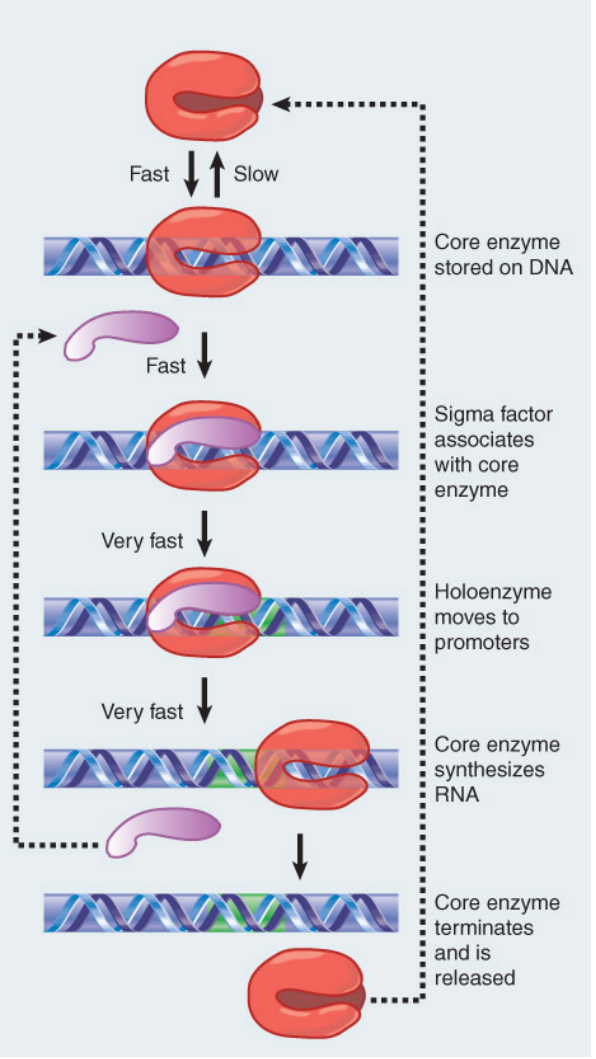
RNA polymerase requires σ-factors for efficient binding to DNA-sequences at TSS

The σ-factor has specific domains that recognize canonical elements in the promoter

The σ-factor controls the affinity of the polymerase for the promoter sequences

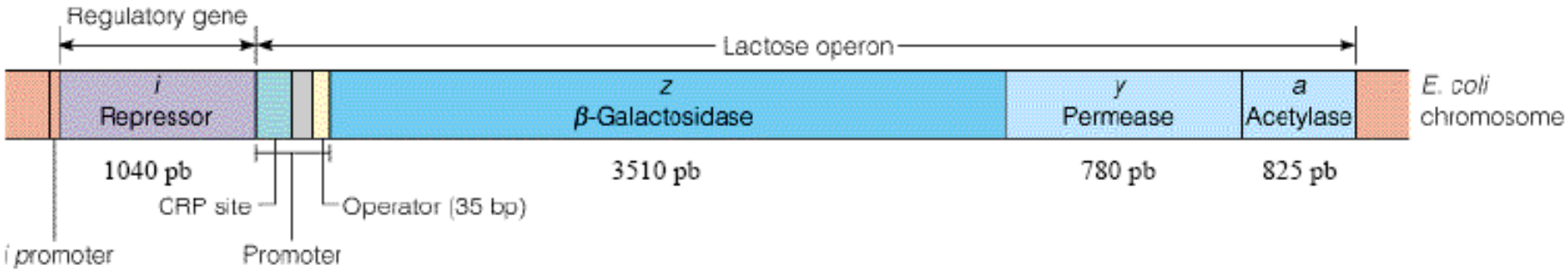
Recruitment of RNA polymerases to TSS is regulated by activating and repressing trans-acting factors

Most cis-acting elements are close to the TSS



III. Gene Regulation in Prokaryotes

Lactose operon encodes enzymes involved in **β-galactosides catabolism**



LacZ : β-galactosidase (hydrolyzes β-galactoside)

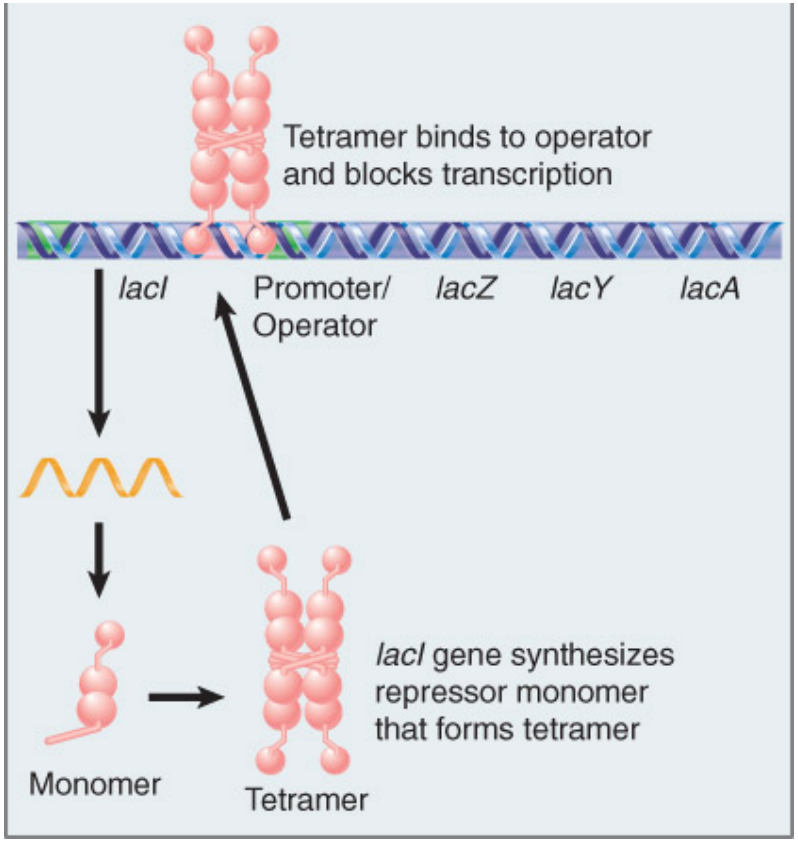
LacY : Permease (imports β-galactosides inside *E. coli*)

LacA : trans-Acetylase (acetylates β-galactosides)

III. Gene Regulation in Prokaryotes

The *lac repressor* is transcribed from its own promoter Pi

LacI binds the Plac (operator) and prevents transcription initiation



III. Gene Regulation in Prokaryotes

LacI is itself inactivated by an inducer (allolactose)

Lactose paradox

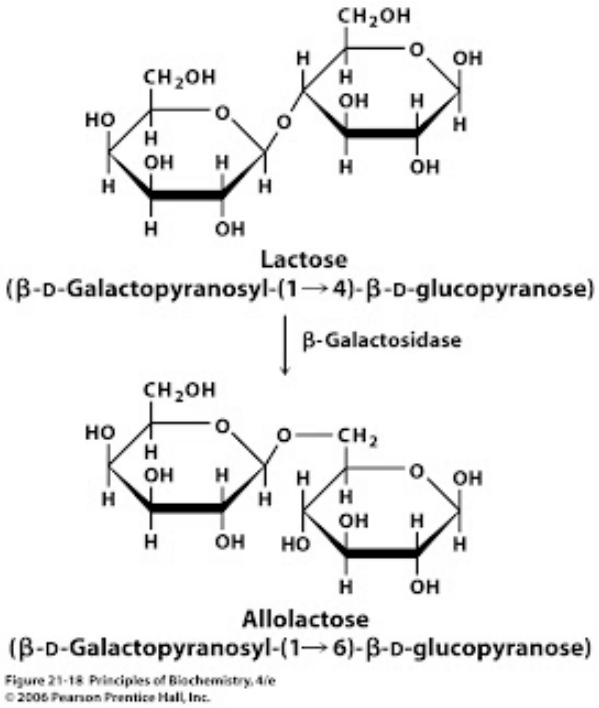
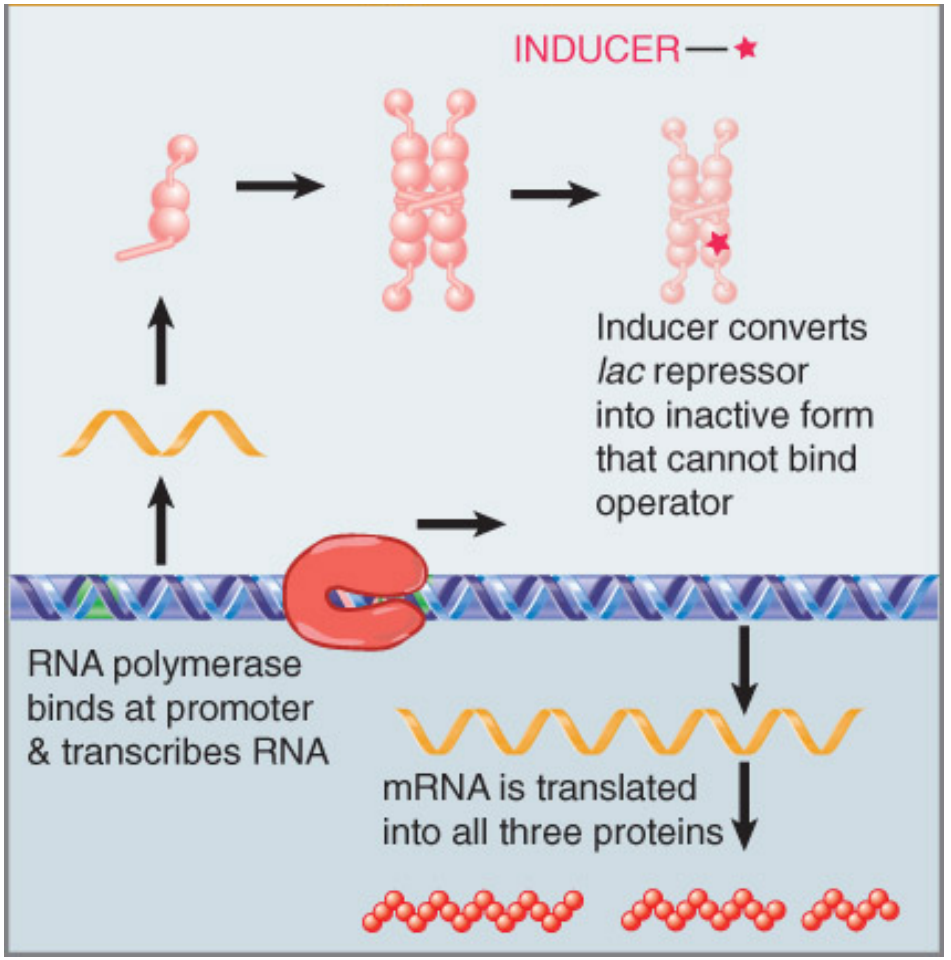
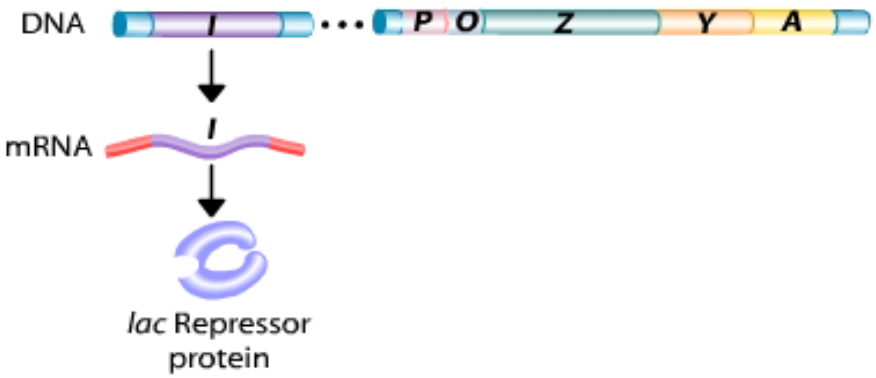


Figure 21-18 Principles of Biochemistry, 4/e
© 2006 Pearson Prentice Hall, Inc.

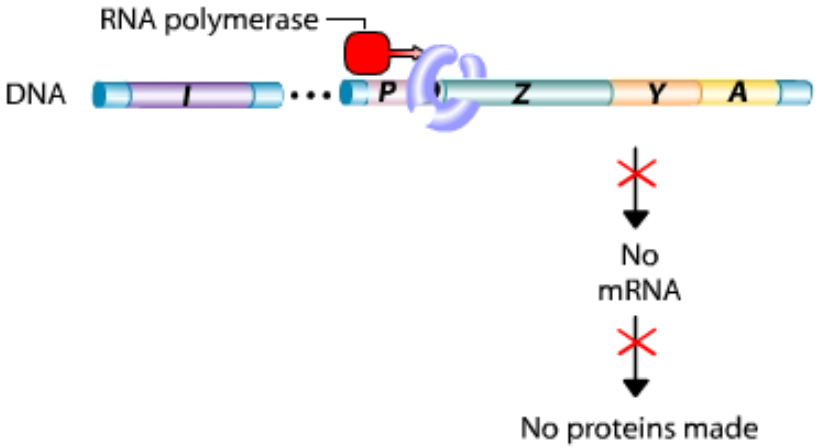


III. Gene Regulation in Prokaryotes

Glucose/ No Lactose



The Lac repressor protein, encoded by the *I* gene, is expressed in the absence or presence of lactose.

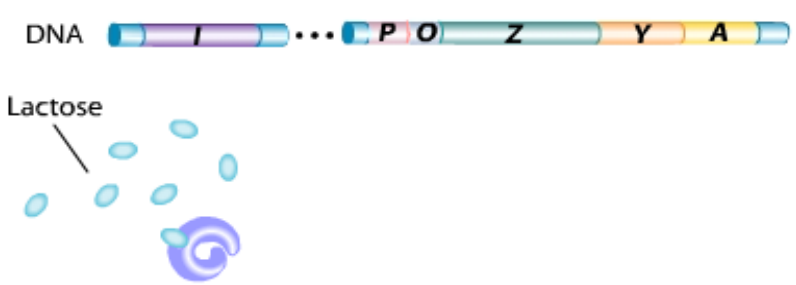


In the absence of lactose, the Lac repressor binds to the *lac* operator site.

Since RNA polymerase is unable to transcribe the *lac* structural genes, the corresponding proteins are not made.

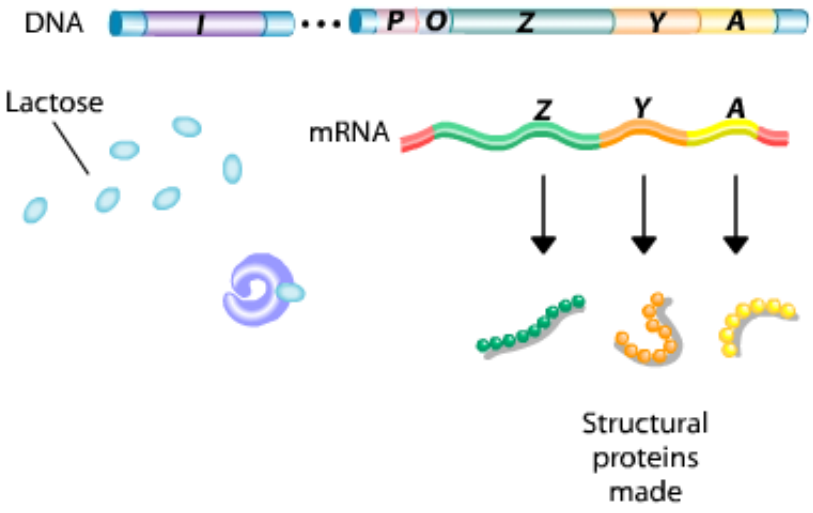
III. Gene Regulation in Prokaryotes

Lactose



When lactose is present in the cell medium, it binds to the allosteric site of the Lac repressor. This changes the conformation of the repressor.

In this conformation, the repressor can no longer bind to the *lac* operator site.

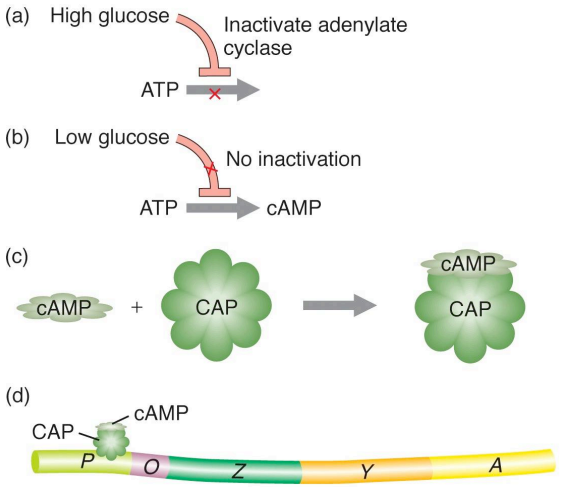


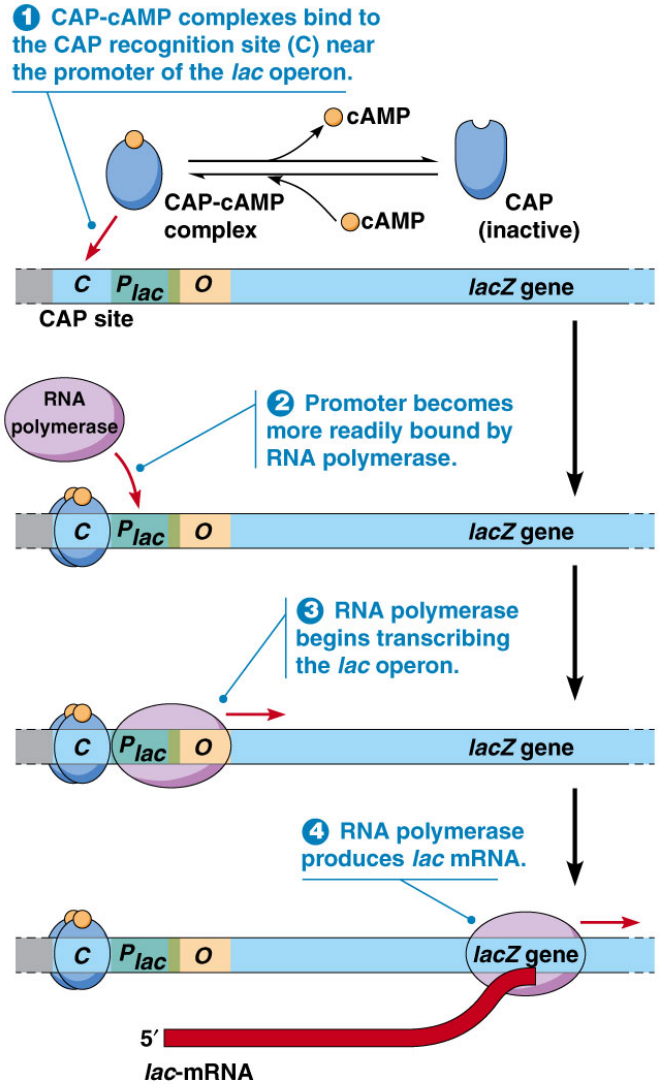
Without the repressor blocking its way, RNA polymerase is able to transcribe the structural genes.

Thus, in the presence of lactose, the *lac* structural genes are expressed. The proteins encoded by the *Z* and *Y* genes are required for the metabolism of lactose.

III. Gene Regulation in Prokaryotes

Preferential use of Glucose over lactose is acheived through the **cAMP Receptor Complex (CRP)**





III. Gene Regulation in Prokaryotes

Please refer to : https://www.youtube.com/watch?v=sDPRl5CbcfA

For detailled explanations of slides 13 to 19 about the Lac operon

IV. Gene Regulation in Eukaryotes

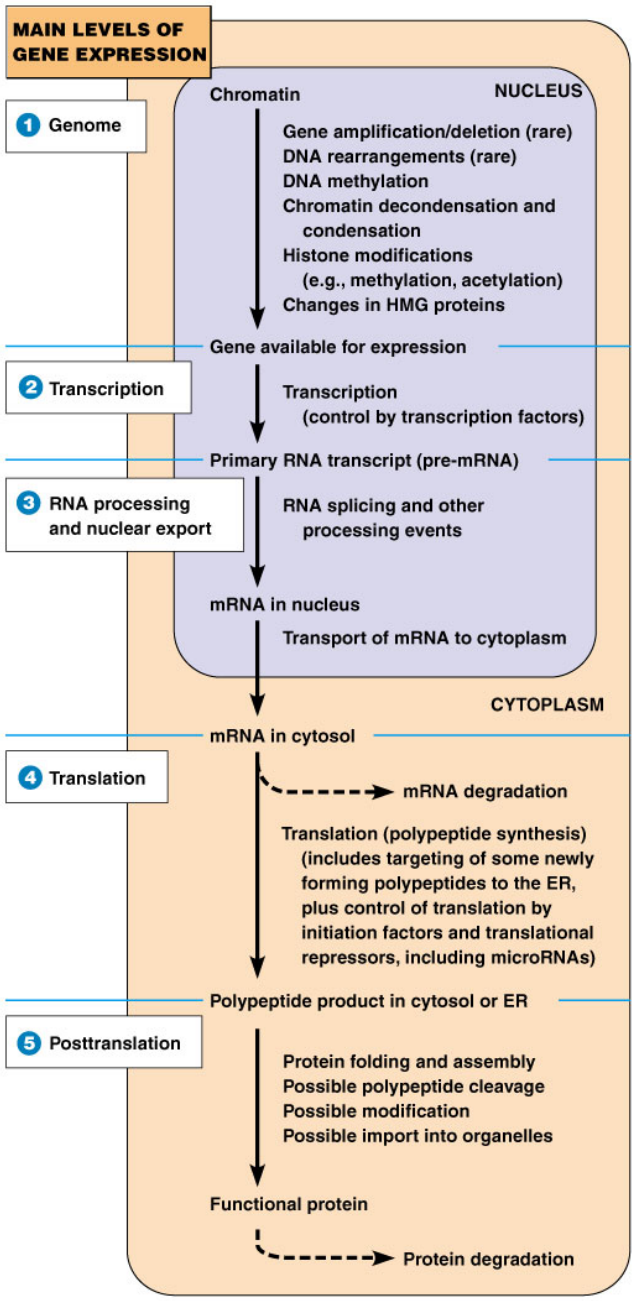
Genome Level (Chromatin Remodeling and DNA rearrangements)

Transcriptional Control

Post-Transcriptional Control

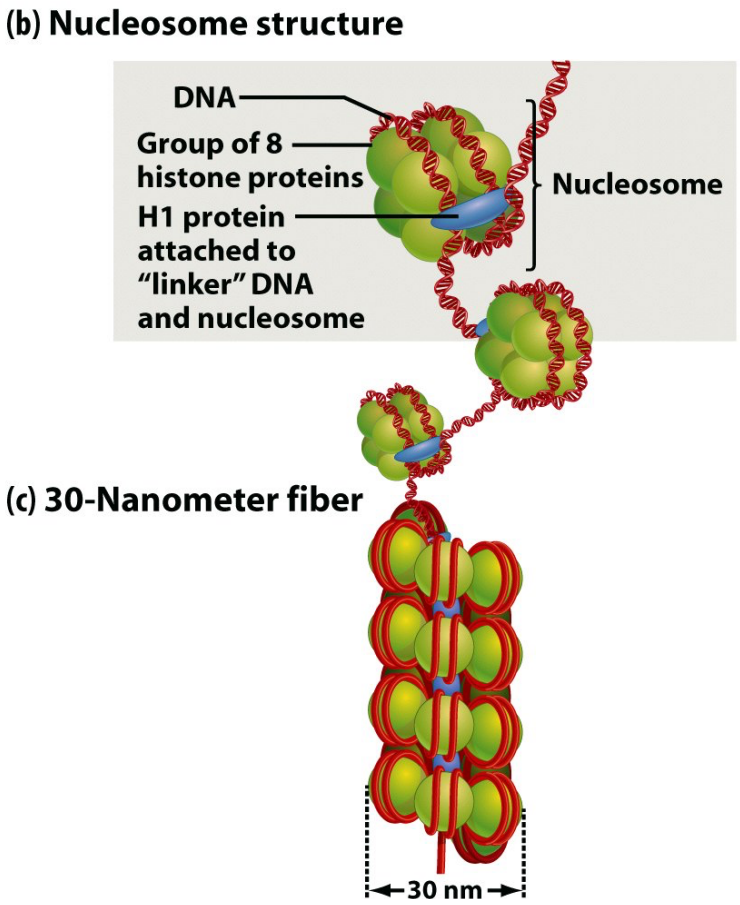
Translational Control

Post-Tranlational Control (protein activity control)



IV. Gene Regulation in Eukaryotes

IV.1 Chromatin structure



IV. Gene Regulation in Eukaryotes

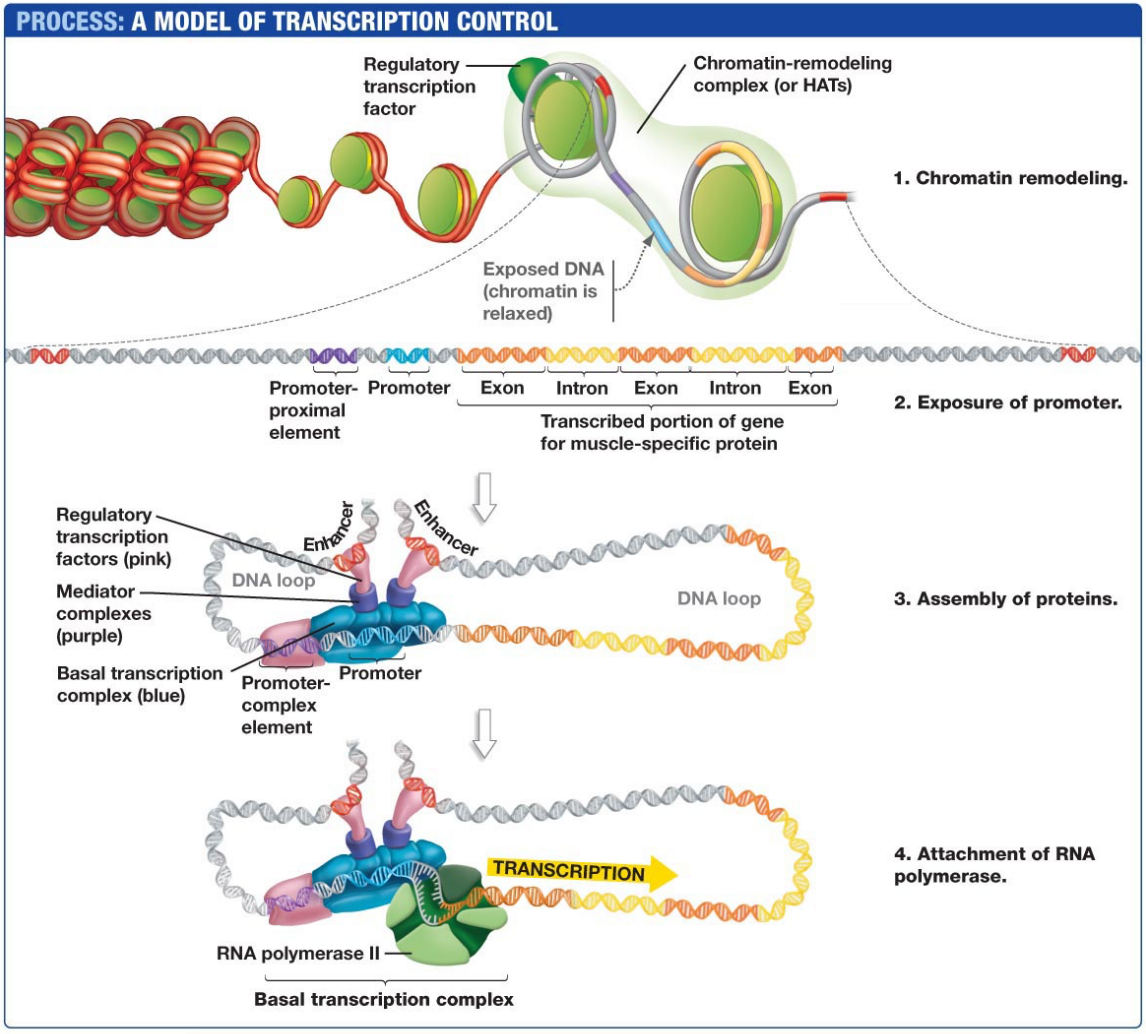
IV.2 Transcriptional control

Three Classes of Proteins Are Involved in Transcriptional Activation

basal transcription factors (TBP)

DNA-binding specific TF,

Coactivators (Mediator).



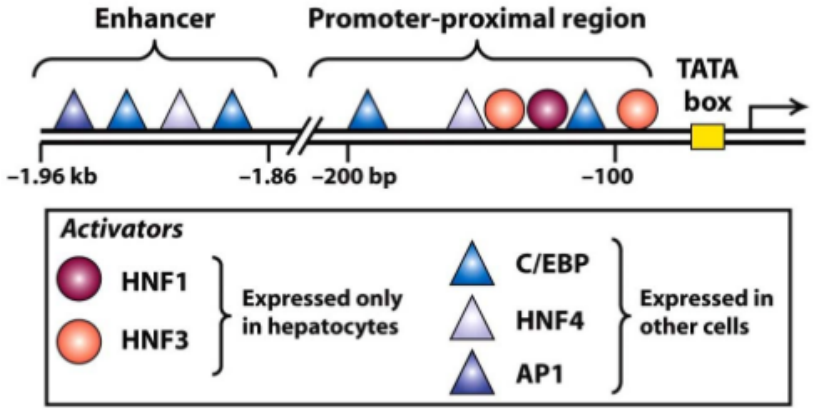
IV. Gene Regulation in Eukaryotes

Unlike bacterial cells and most single cell eukaryotes, cells in multicellular organisms have relatively few genes that are directly and reversibly regulated by environmental conditions

Instead multicellular organisms have many different, specialized cells. Hence, **tissue-specific gene control** is important for development and differentiation.

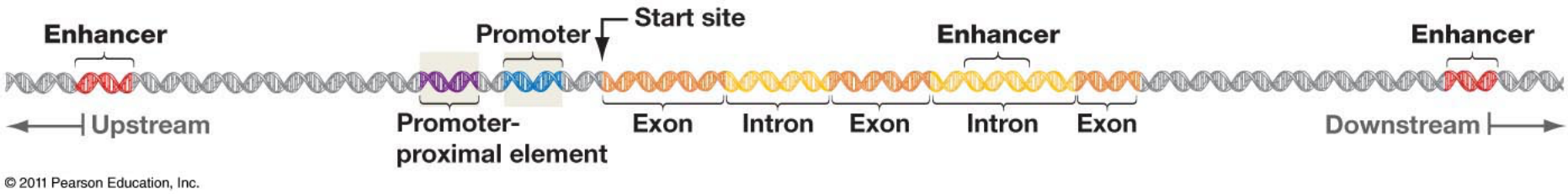
every gene has more than one gene regulator (all of which must be on for the gene to function).

Eukaryotic regulatory elements are usually several Kb away from the promoter



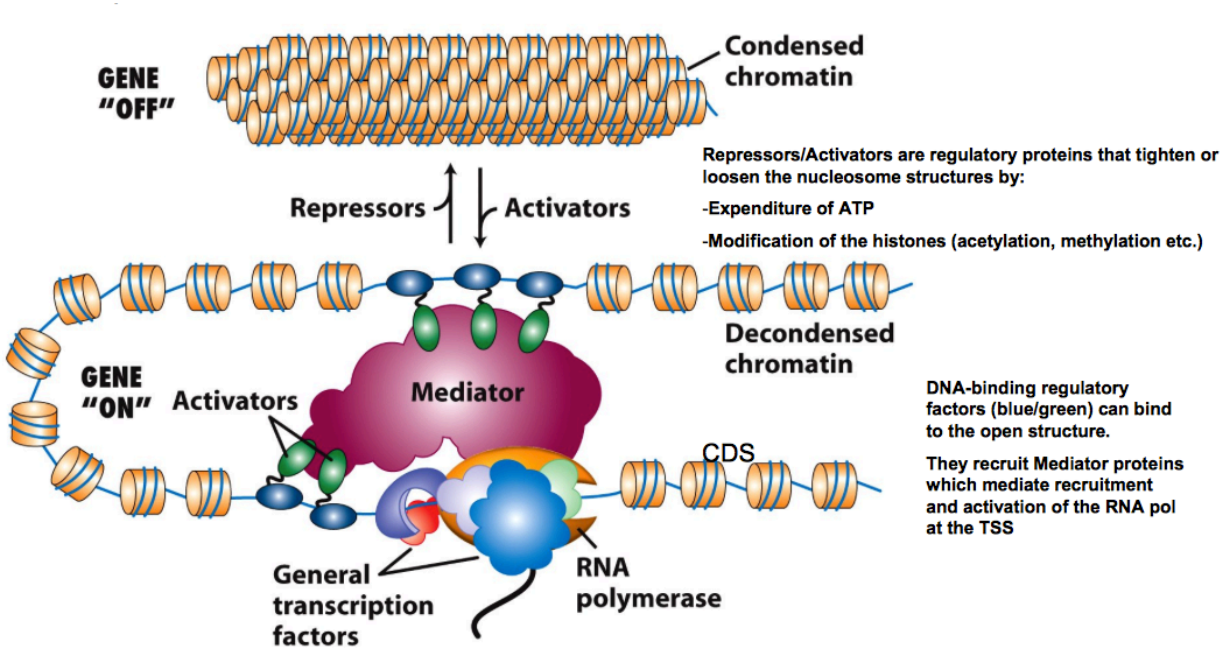
IV. Gene Regulation in Eukaryotes

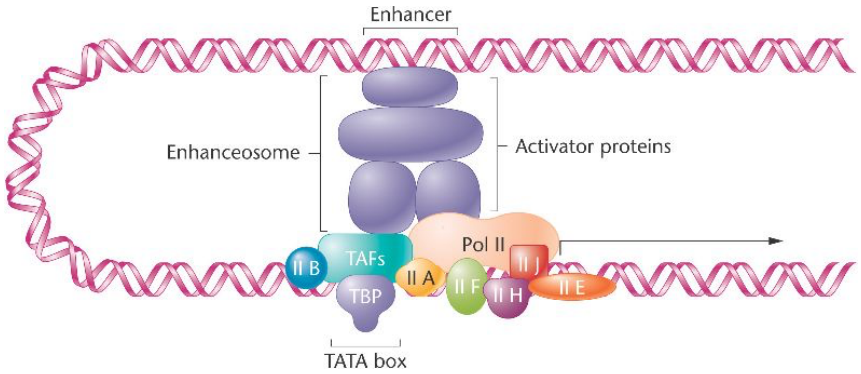
Eukaryotic regulatory sequences can be close or far from the core promoter



IV. Gene Regulation in Eukaryotes

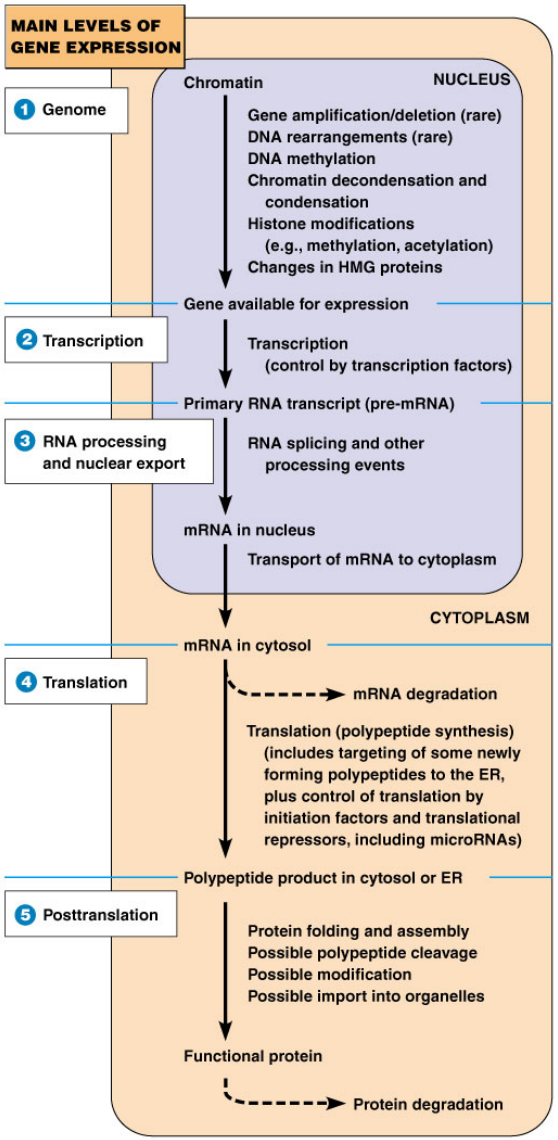
Transcription factors are key regulators of eukaryotic gene expression

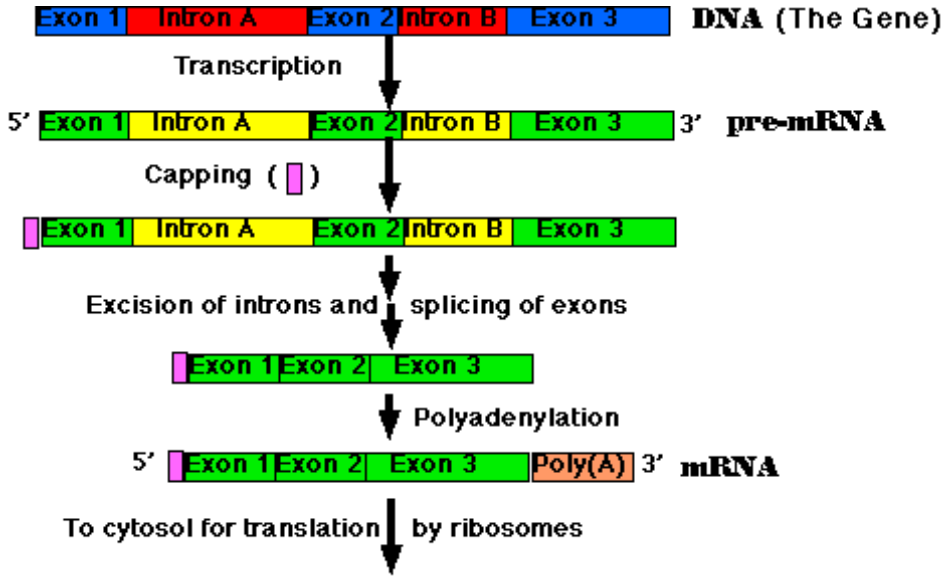




III. Gene Regulation in Eukaryotes

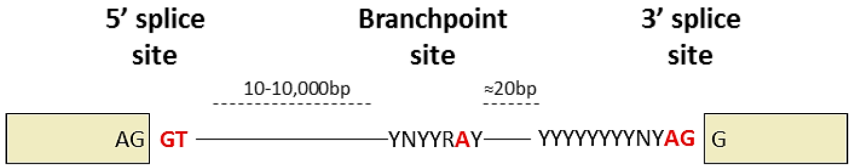
IV.3 Post-transcriptional control of gene expression

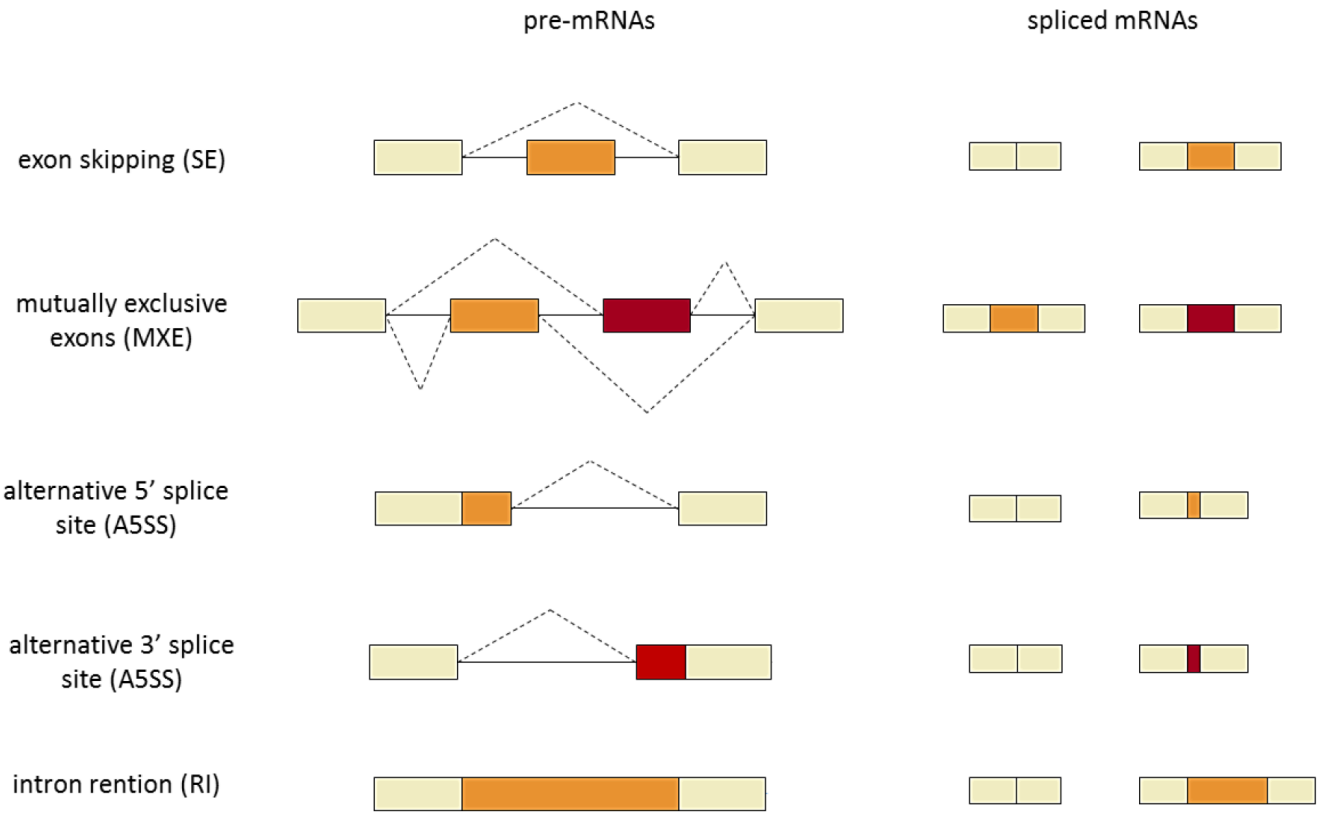




III. Gene Regulation in Eukaryotes

IV.3 Post-transcriptional control of gene expression : splicing





IV. Gene Regulation in Eukaryotes

IV.3 Post-transcriptional control of gene expression : non coding RNAs

only 1% of the genome is translated into proteins but much more transcribed

A great portion of the (**procaryotic and eukaryotic**) genome is transcribed into non coding RNAs. Some of which can be involved in gene regulation :

- sncRNA
- LncRNA
- miRNAs

IV. Gene Regulation in Eukaryotes

SUMMARY TABLE 18.1 **Regulating Gene Expression in Bacteria and Eukaryotes**

Level of Regulation	Bacteria	Eukaryotes
Chromatin remodeling	• Limited packaging of DNA • Remodeling not a major issue in regulating gene expression.	• Extensive packaging of DNA • Chromatin must be opened for transcription to begin.
Transcription	• Positive and negative control by regulatory proteins that act at sites close to the promoter • Sigma interacts with promoter.	• Positive and negative control by regulatory proteins that act at sites close to and far from promoter • Large basal transcription complex interacts with promoter. • Mediator complex required.
RNA processing	• None documented	• Extensive processing: alternative splicing of introns addition of 5′ cap and 3′ tail
mRNA stability	• Some RNA interference documented	• For many genes, RNA interference limits life span or translation rate.
Translation	• Regulatory proteins bind to mRNAs and/or ribosome and affect translation rate.	• Regulatory proteins bind to mRNAs and/or ribosome and affect translation rate.
Post-translational modification	• Folding by chaperone proteins • Chemical modification (e.g., phosphorylation) may change activity.	• Folding by chaperone proteins • Chemical modification (glycosylation, phosphorylation) • Ubiquination targets proteins for destruction by proteasome.

V. Principles of transcriptome analysis

V.1 SAGE Technology

SAGE : Serial *analysis of gene expression (1995)* :

Unique tags are used to identify each mRNA

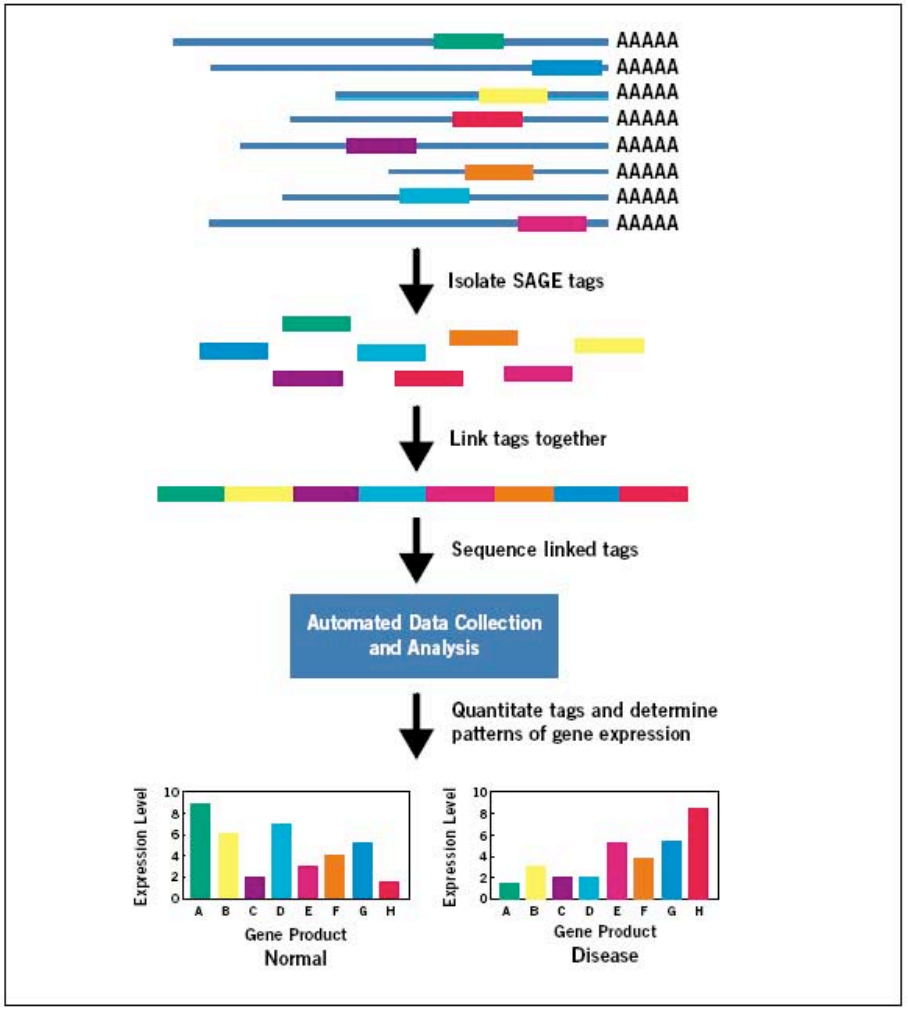
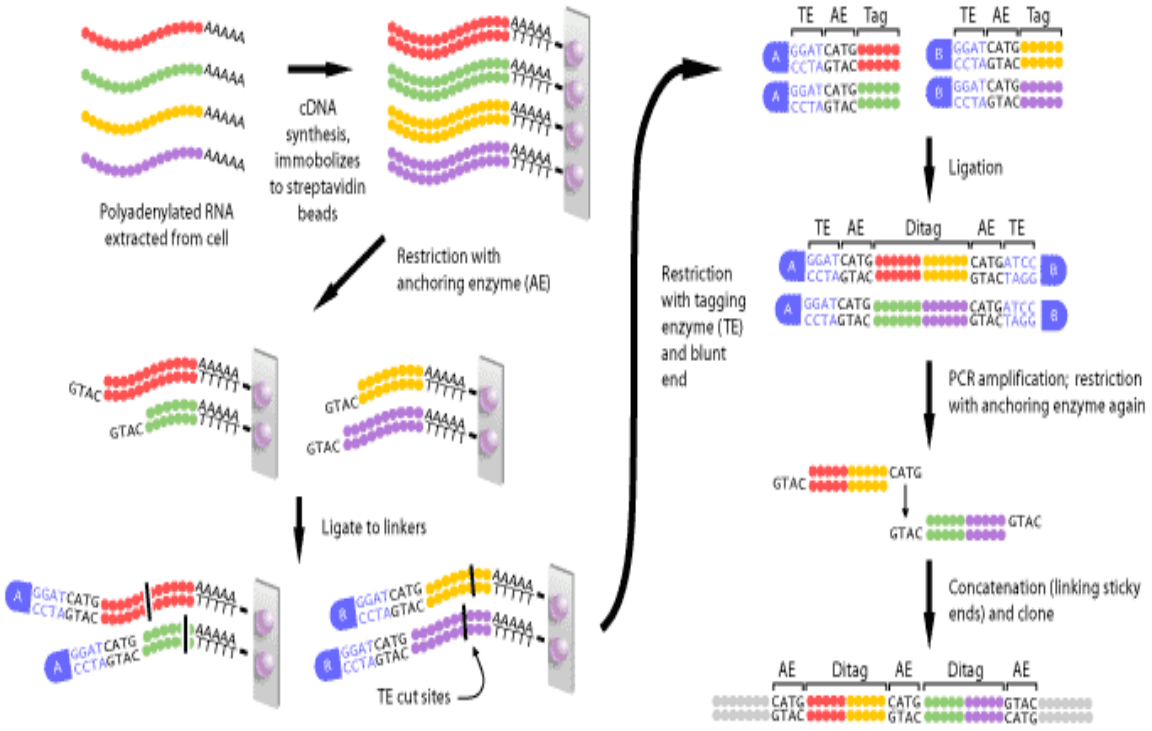


Figure 1. Schematic of SAGE method (Courtesy of sagenet.org)[3]

V. Principles of transcriptome analysis

V.1 SAGE Technology

1. Isolate the mRNA of an input sample (e.g. a tumour). Use Reverse Transcriptase to synthesize complementary DNA (cDNA) from mARN
2. Extract a small chunk of sequence from a defined position of each cDNA molecule.
3. Link these small pieces of sequence together to form a long chain (or concatamer).
4. Clone these chains into a **vector** which can be taken up by bacteria.
5. Sequence these chains using modern high-throughput DNA sequencers.
6. Process this data with a computer to count the small sequence tags.



V. Principles of transcriptome analysis

V.1 SAGE Technology

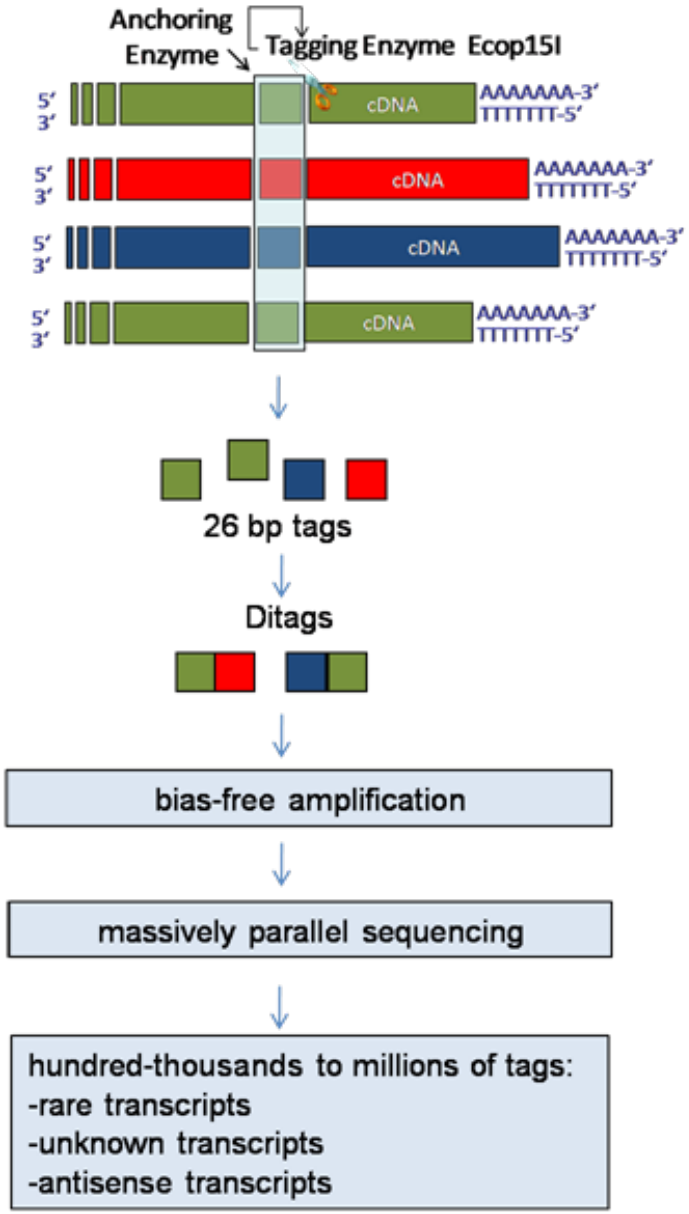
TAG	COUNT		TAG	COUNT		TAG	COUNT
CCCATCGTCC	1286		CACTACTCAC	245		TTCACTGTGA	150
CCTCCAGCTA	715		ACTAACACCC	229		ACGCAGGGAG	142
CTAAGACTTC	559		AGCCCTACAA	222		TGCTCCTACC	140
GCCCAGGTCA	519		ACTTTTTCAA	217		CAAACCATCC	140
CACCTAATTG	469		GCCGGGTGGG	207		CCCCCTGGAT	136
CCTGTAATCC	448		GACATCAAGT	198		ATTGGAGTGC	136
TTCATACACC	400		ATCGTGGCGG	193		GCAGGGCCTC	128
ACATTGGGTG	377		GACCCAAGAT	190		CCGCTGCACT	127
GTGAAACCCC	359		GTGAAACCCT	188		GGAAAACAGA	119
CCACTGCACT	359		CTGGCCCTCG	186		TCACCGGTCA	118
TGATTTCACT	358		GCTTTATTTG	185		GTGCACTGAG	118
ACCCTTGGCC	344		CTAGCCTCAC	172		CCTCAGGATA	114
ATTTGAGAAG	320		GCGAAACCCT	167		CTCATAAGGA	113
GTGACCACGG	294		AAAACATTCT	161		ATCATGGGGA	110

- Collect sequence records from GenBank
- Assign sequence orientation (by finding poly-A tail or poly-A signal or from annotations)
- Assign UniGene identifier to each sequence with a SAGE tag
- Record (for each tag-gene pair)
  - #sequences with this tag
  - #sequences in gene cluster with this tag
- Softwares : SAGE Genie or SAGE map

V. Principles of transcriptome analysis

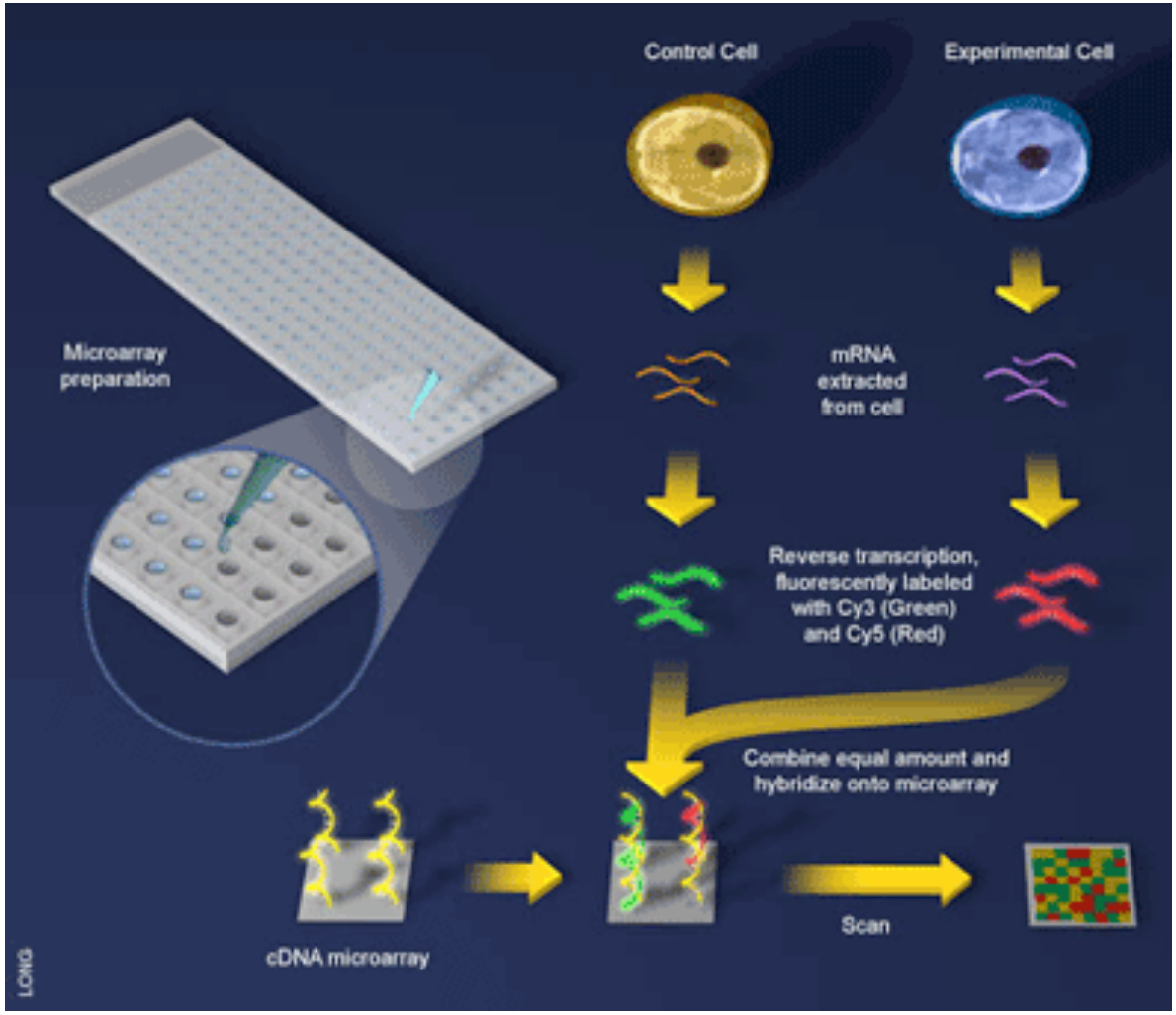
V.1 SAGE Technology

HT-SuperSAGE



V. Principles of transcriptome analysis

V.2 Microarray Technology



Microarray technology animation

V. Principles of transcriptome analysis

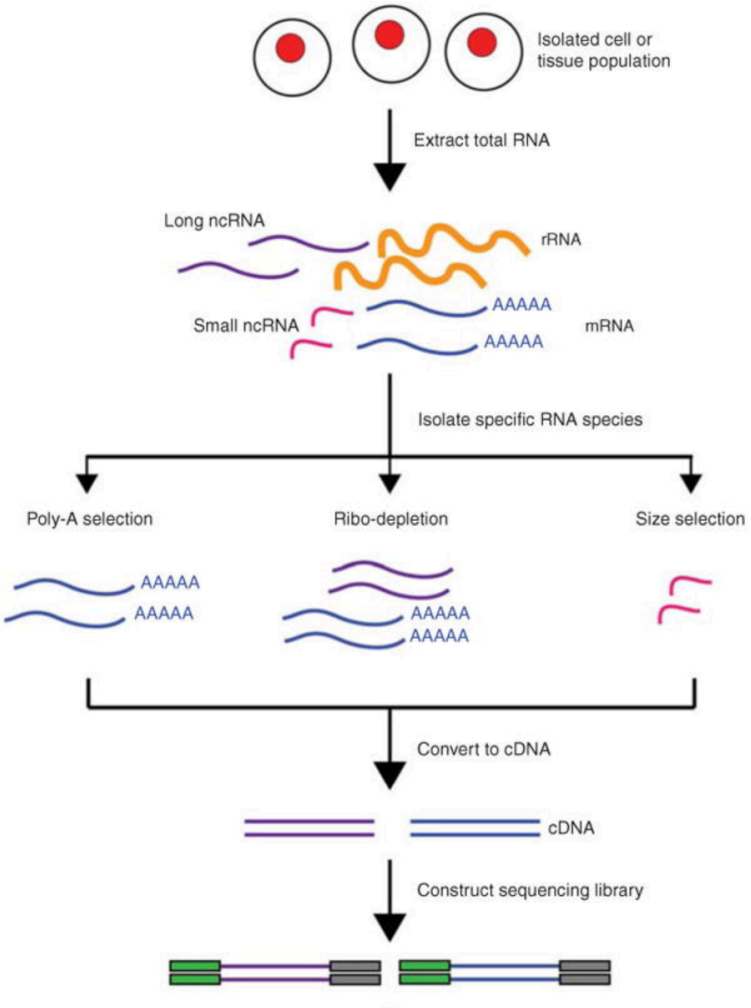
V.2 Microarray Technology

Features	*SAGE*	*Microarray*
Detects unknown transcripts	Yes	No
Quantification	Absolute measure	Relative measure
Sensitivity	High	Moderate
Specificity	Moderate	High
Reproducibility	Good for higher abundance transcripts	Good for data from intra-platform comparison
Direct cost	5-10X higher than arrays.	5-10 X lower than SAGE

V. Principles of transcriptome analysis

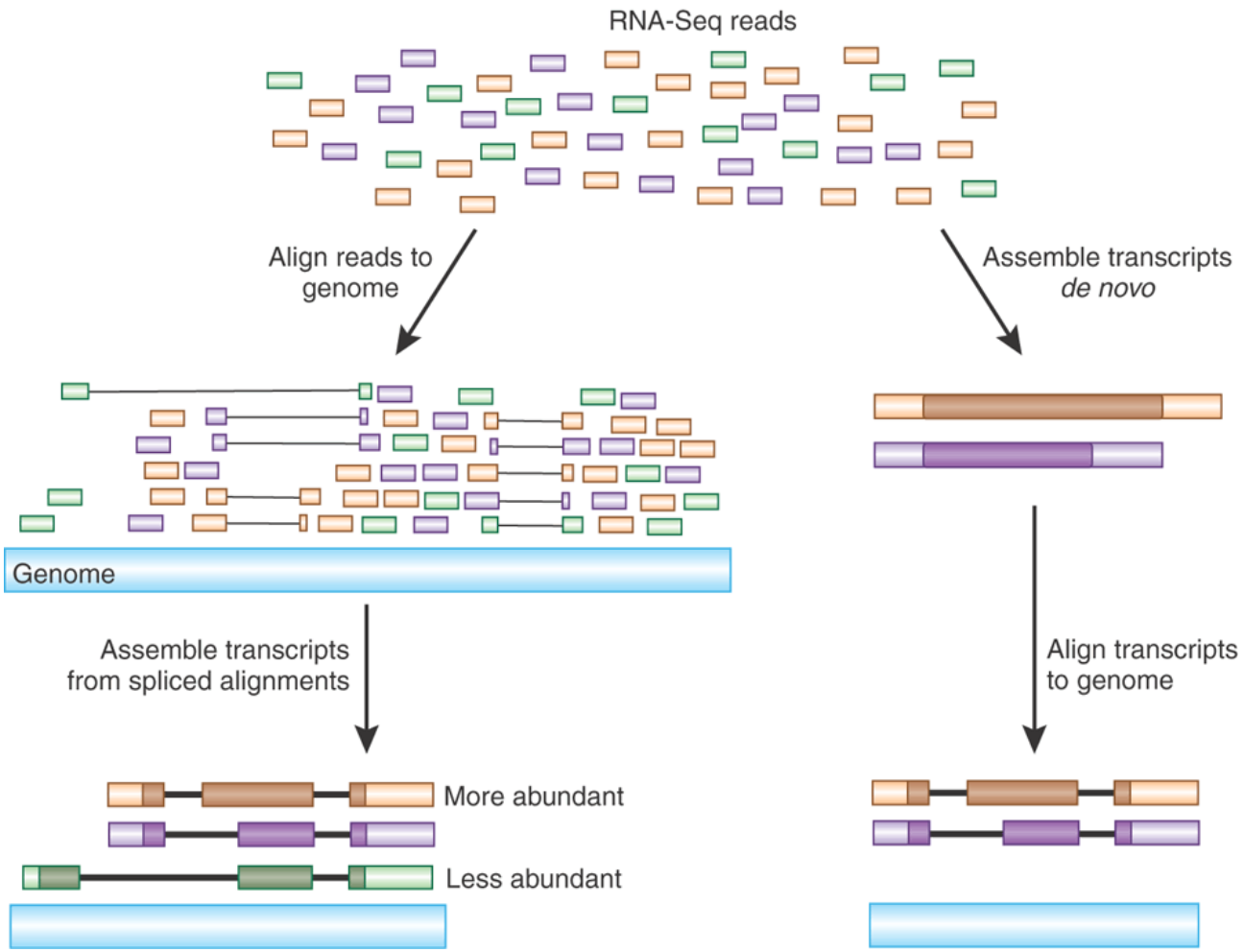
V.3 RNA-Sequencing (RNA-Seq) : methodology

1. **RNA extraction**
2. **Reverse transcription**
3. **Library preparation**
4. **Polymerase Chain reaction (on solid-Illumina or emulsion-454)**
5. **Sequencing (reversible terminator or pyrosequencing)**
6. **Bioinformatic processing**



V. Principles of transcriptome analysis

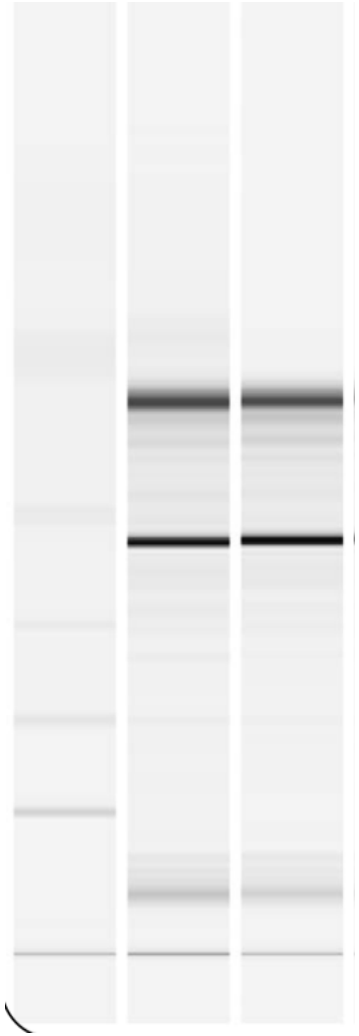
V.3 RNA-Sequencing (RNA-Seq) : basics

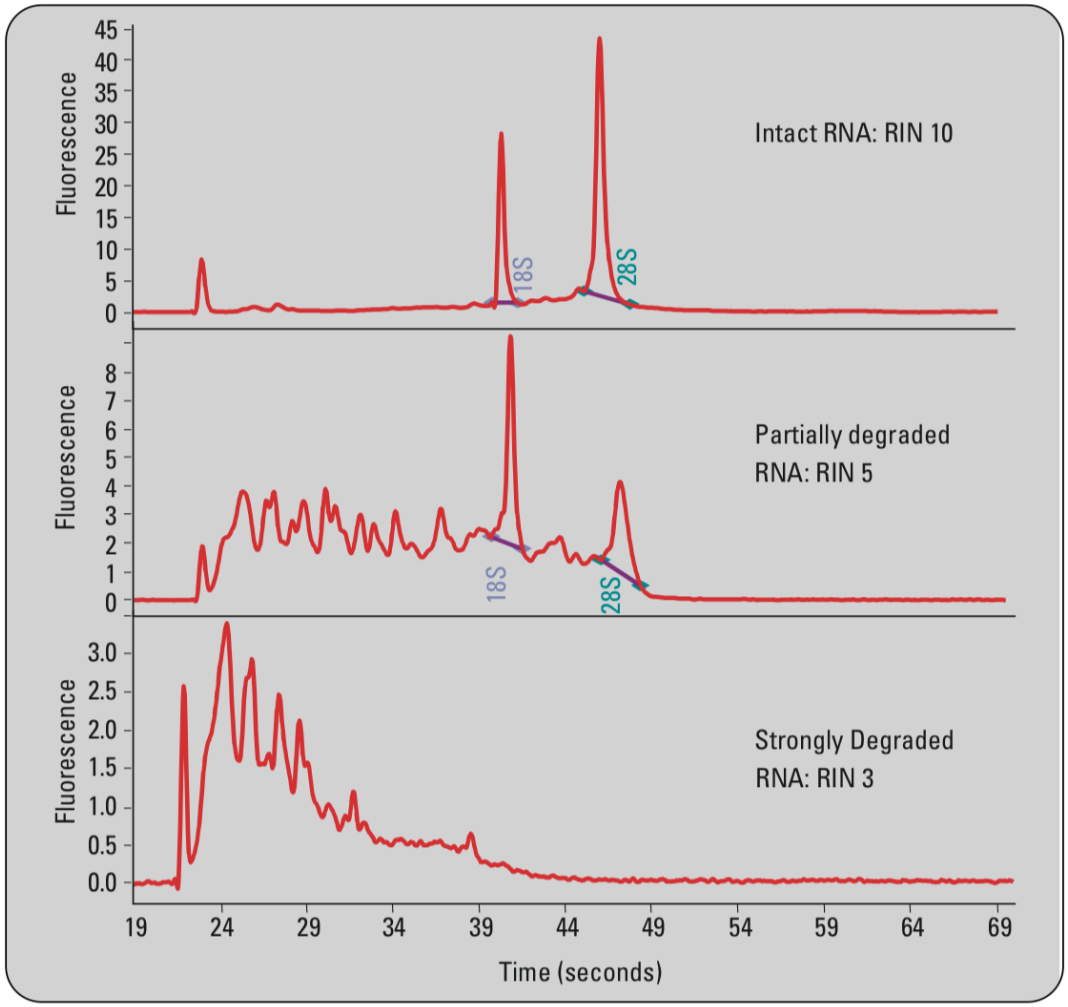


V. Principles of transcriptome analysis

V.3 RNA-Sequencing (RNA-Seq) : methodology

1. RNA extraction and QC (Agilent Bioanalyzer



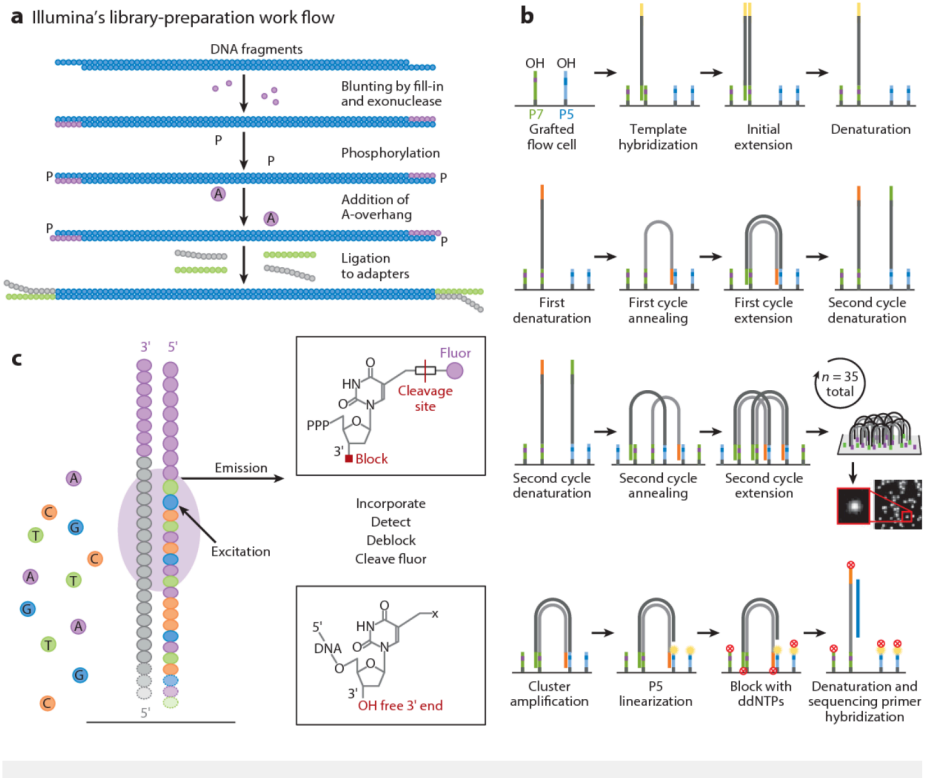


V. Principles of transcriptome analysis

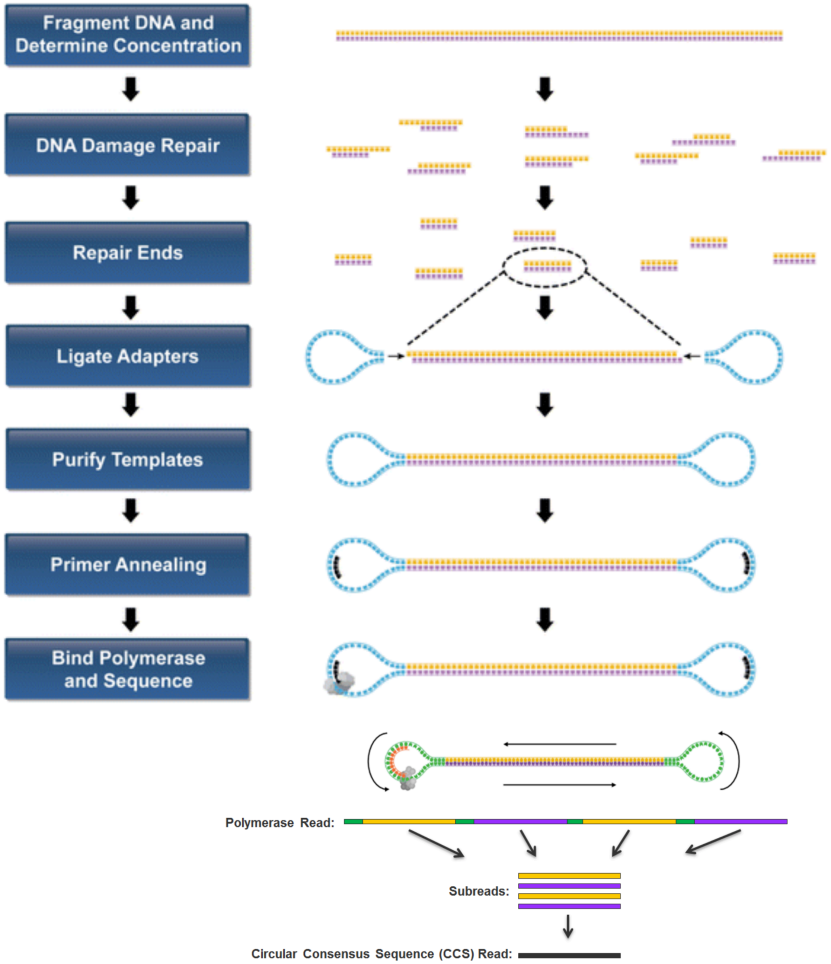
V.3 RNA-Sequencing (RNA-Seq) : methodology

2. Reverse transcription, library preparation, amplification, sequencing

ILLUMINA



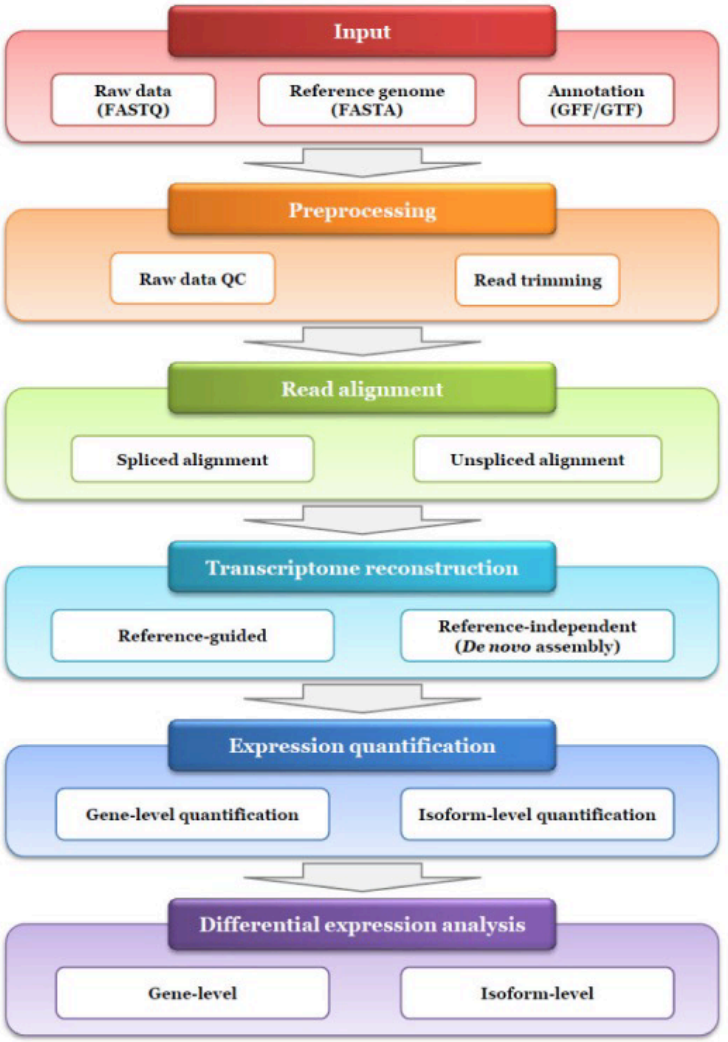
PACIFIC BIOSIENCES



V. Principles of transcriptome analysis

V.3 RNA-Sequencing (RNA-Seq) : methodology

6. Bio-informatics



Workflow	Category	Package
Preprocessing of raw data	Raw data QC	FastQC
		HTQC
	Read trimming	FASTX-Toolkit
		FLEXBAR
Read alignment	Unspliced aligner	MAQ
		BWA
		Bowtie
	Spliced aligner	TopHat
		MapSplice
		STAR
		GSNAP
RNA-seq specific quality control		RNA-SeQC
		RSeQC
		Qualimap 2
Transcriptome reconstruction	Reference-guided	Cufflinks
		Scripture
		StringTie
	Reference-independent	Trinity
		Oases
		transABySS
Expression quantification	Gene-level quantification	ALEXA-seq
		Enhanced read analysis of gene expression (ERANGE)
		Normalization by expected uniquely mappable area (NEUMA)
	Isoform-level quantification	Cufflinks
		StringTie
		RSEM
		Sailfish
Differential expression	Gene-level	NOIseq
		edgeR
		DESeq
		SAMseq
	Isoform-level	Cuffdiff
		EBSeq
		Ballgown

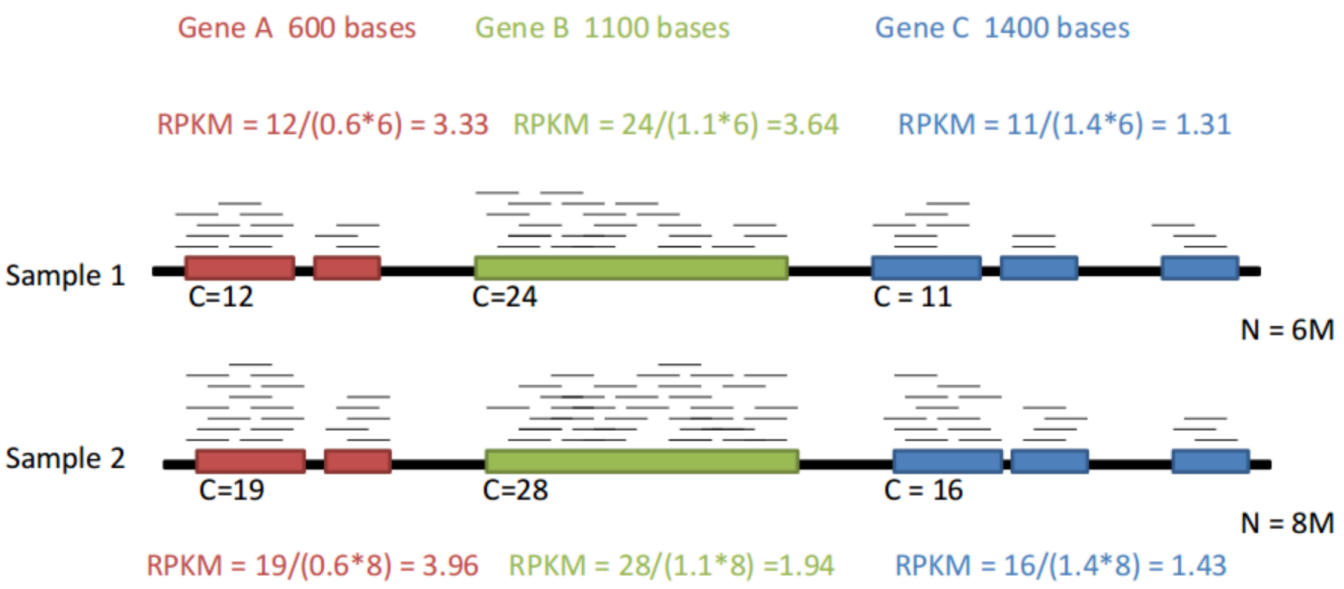
IV. Principles of transcriptome analysis

IV.3 RNA-Sequencing (RNA-Seq)

Gene expression levels

$$RPKM = \frac{number\ of\ reads\ of\ the\ region}{\frac{total\ reads}{1{,}000{,}000} \ x \ \frac{region\ length}{1{,}000}}$$

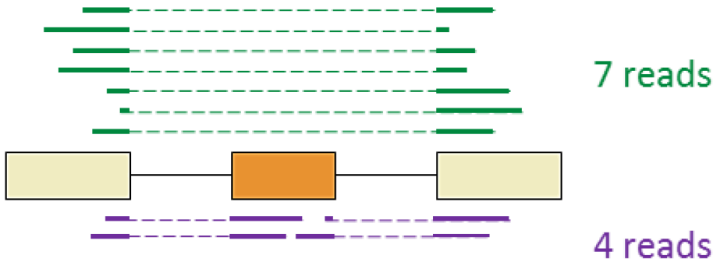
RPKM Example



IV. Principles of transcriptome analysis

IV.3 RNA-Sequencing (RNA-Seq)

Splicing analysis



$$\text{PSI} = \frac{\#\ inclusion\ reads}{(\#\ inclusion\ reads + \#\ skipping\ reads)} = \frac{4}{(4+7)} = 0.36$$

IV. Principles of transcriptome analysis

IV.3 RNA-Sequencing (RNA-Seq)

Detection of novel transcripts and other changes

Dynamic range

Specificity and sensitivity

Costly and requires high computational power