Seminar of Genomic Study

2020 Spring Semester

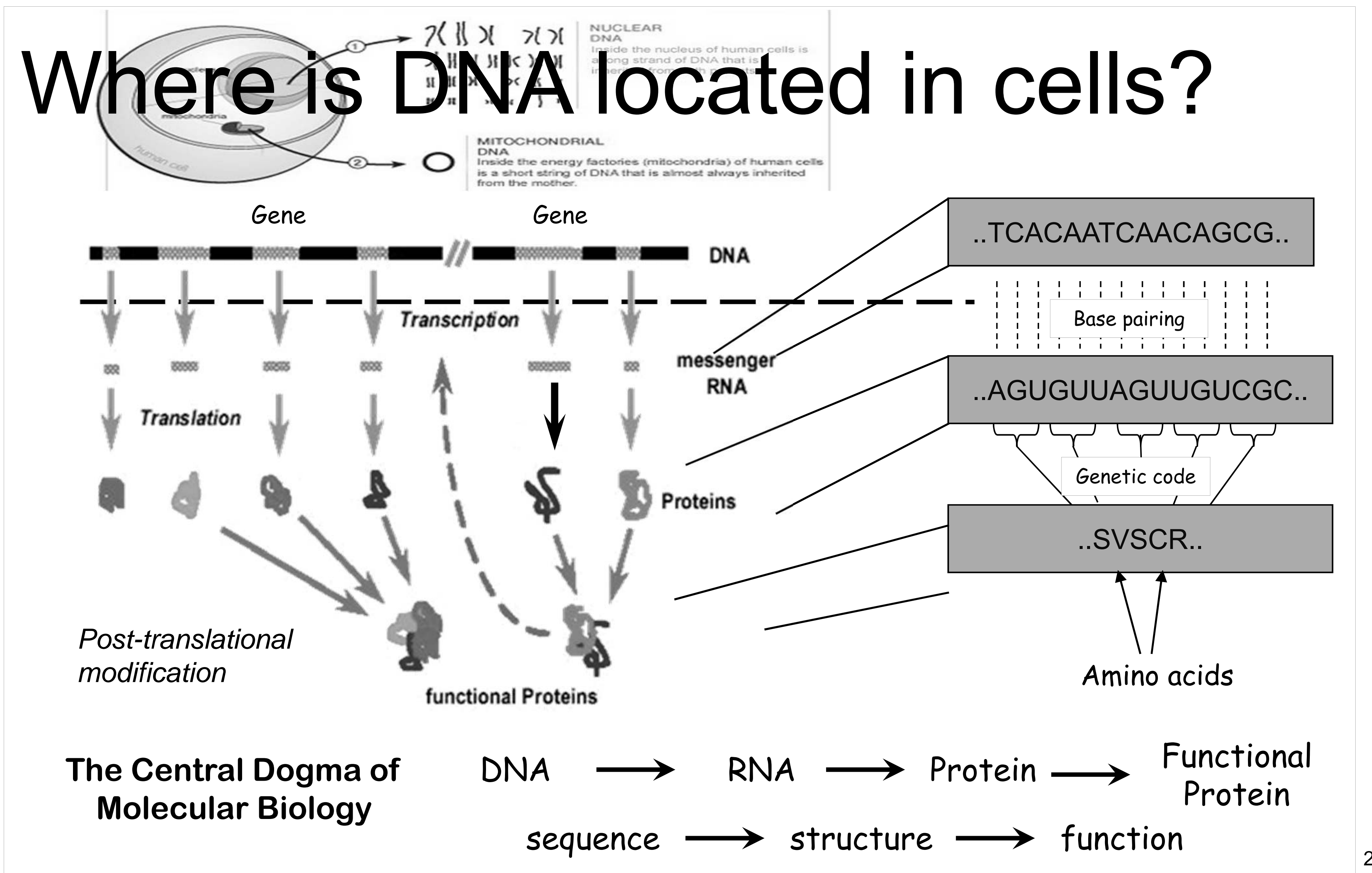
Liang-Chuan Lai

Email: llai@ntu.edu.tw

Ph: 23123456 ext. 88241

https://ceiba.ntu.edu.tw/1082Physio7017_

DNA: The Molecule of Life



Vertebrate Mitochondrial Codon

- Differences from the universal codon

DNA codons	RNA codons	This code (2)	Standard code (1)
AGA	AGA	STOP = Ter (*)	Arg (R)
AGG	AGG	STOP = Ter (*)	Arg (R)
ATA	AUA	Met (M)	Ile (I)
TGA	UGA	Trp (W)	STOP = Ter (*)

Second letter

First letter	U	C	A	G	Third letter
U	UUU Phe UUC Phe UUA Leu UUG Leu	UCU Ser UCC Ser UCA Ser UCG Ser	UAU Tyr UAC Tyr UAA Stop UAG Stop	UGU Cys UGC Cys UGA Trp UGG Trp	U C A G
C	CUU Leu CUC Leu CUA Leu CUG Leu	CCU Pro CCC Pro CCA Pro CCG Pro	CAU His CAC His CAA Gln CAG Gln	CGU Arg CGC Arg CGA Arg CGG Arg	U C A G
A	AUU Ile AUC Ile AUA Met AUG Met	ACU Thr ACC Thr ACA Thr ACG Thr	AAU Asn AAC Asn AAA Lys AAG Lys	AGU Ser AGC Ser AGA Stop AGG Stop	U C A G
G	GUU Val GUC Val GUA Val GUG Val	GCU Ala GCC Ala GCA Ala GCG Ala	GAU Asp GAC Asp GAA Glu GAG Glu	GGU Gly GGC Gly GGA Gly GGG Gly	U C A G

Gene Expression: Transcription

Central dogma of biology: DNA → RNA → protein

5' ...A T G G C C T G G A C T T C A... 3'   Sense strand of DNA
3' ...T A C C G G A C C T G A A G T... 5'   Antisense strand of DNA

↓ Transcription of antisense strand

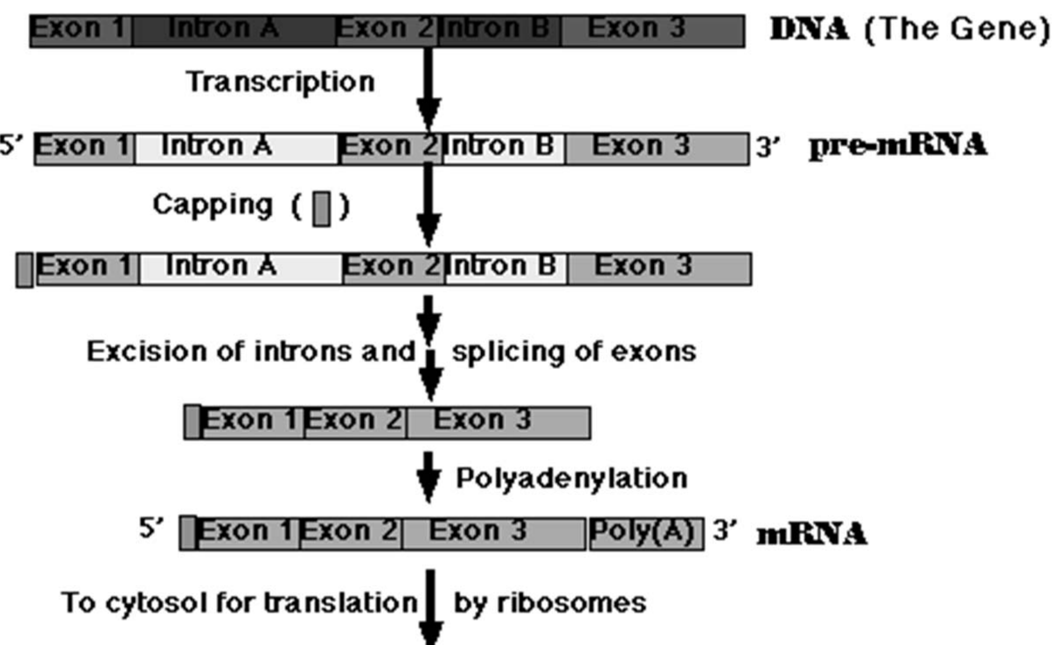
5' ...A U G G C C U G G A C U U C A... 3'   mRNA

↓ Translation of mRNA

Met— Ala— Trp— Thr — Ser —   Peptide

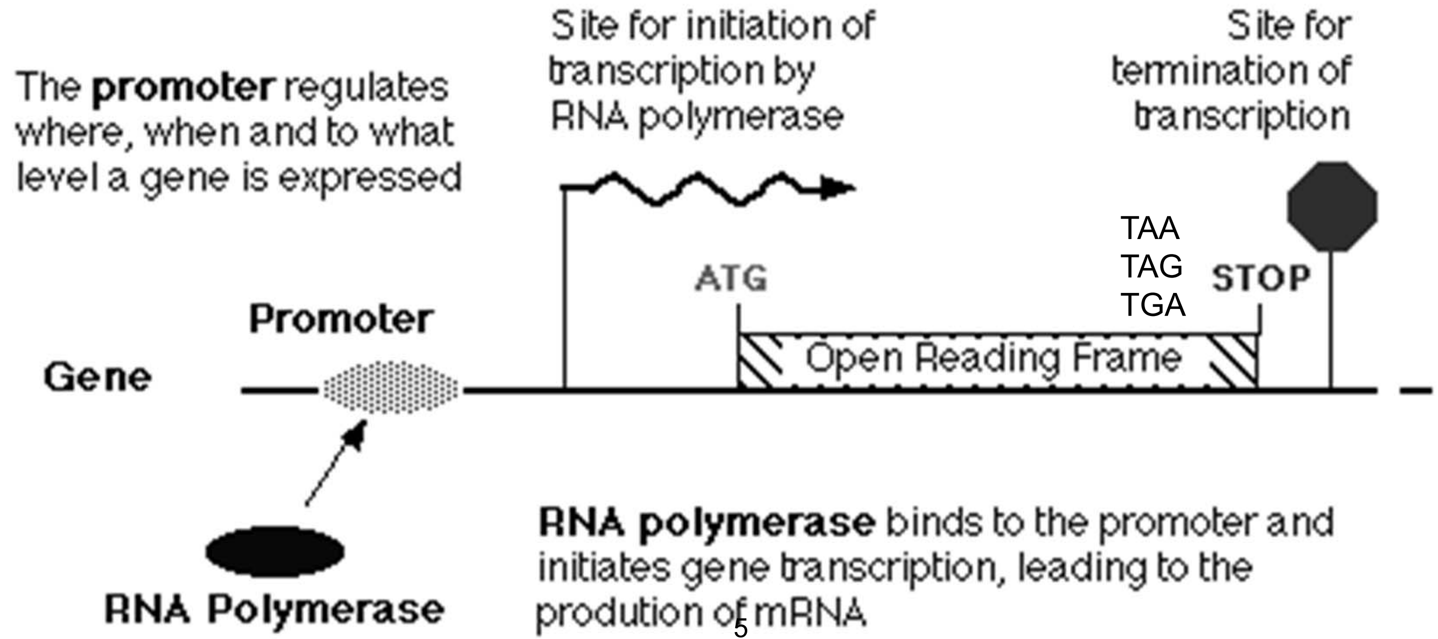
RNA Processing: pre-mRNA → mRNA

- Capping
- Splicing
- Polyadenylation



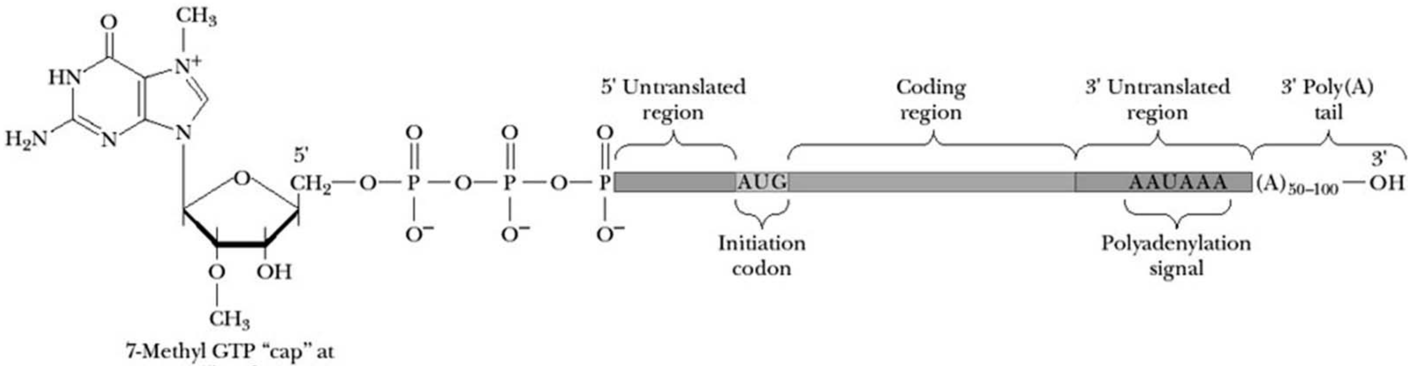
Gene Expression: Transcription

Gene is a portion of DNA that contains both "coding" sequences that determine what the gene does, and "non-coding" sequences that determine when the gene is active (expressed)



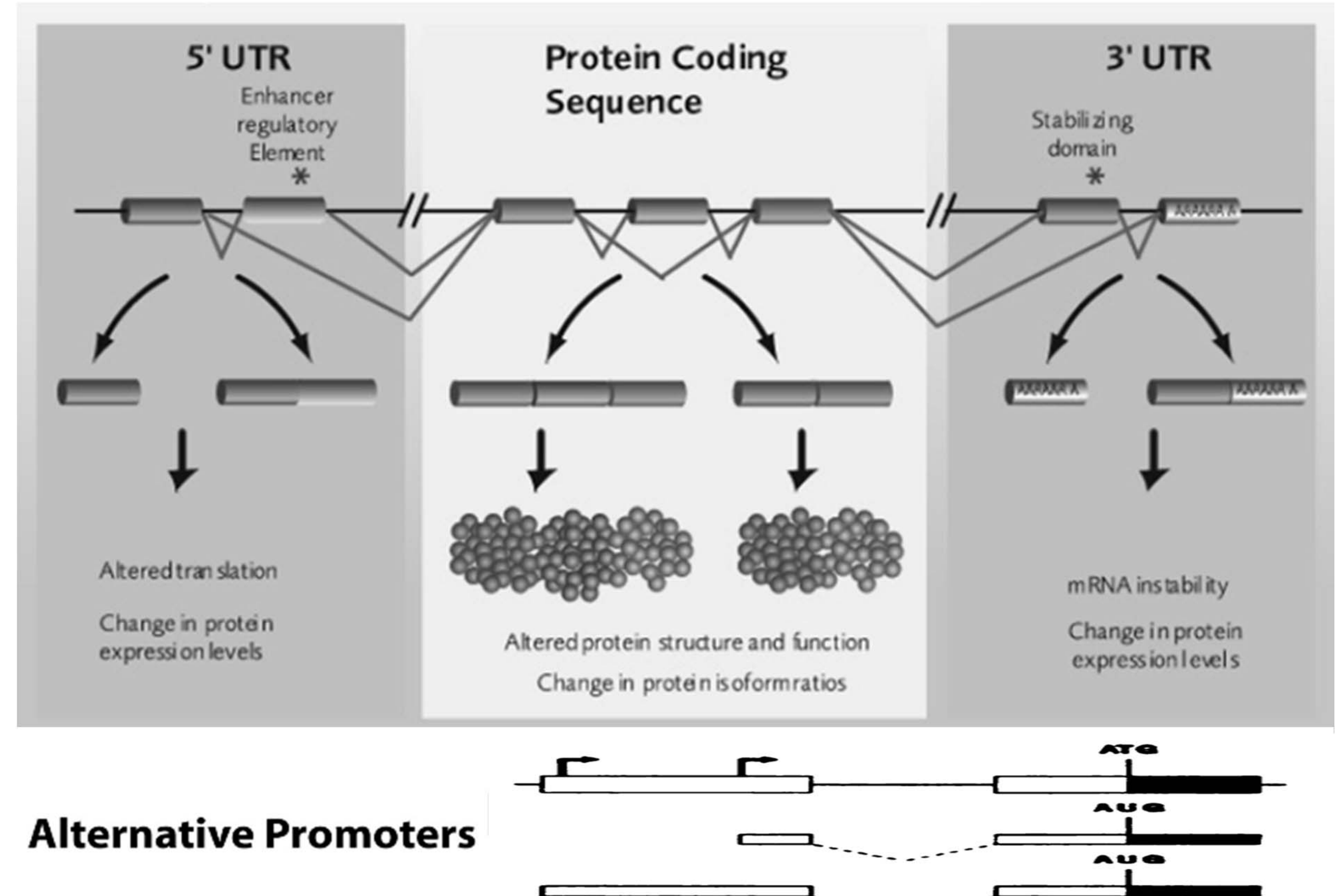
Structure of Mature mRNA

- Coding region
- 7-methyl-GTP cap
  - Bound by cap binding proteins
- Untranslated regions
  - 5′ UTR
    - Translation regulation
  - 3′ UTR
    - Stability elements
    - Subcellular localization (zip codes)
- Poly(A) tail

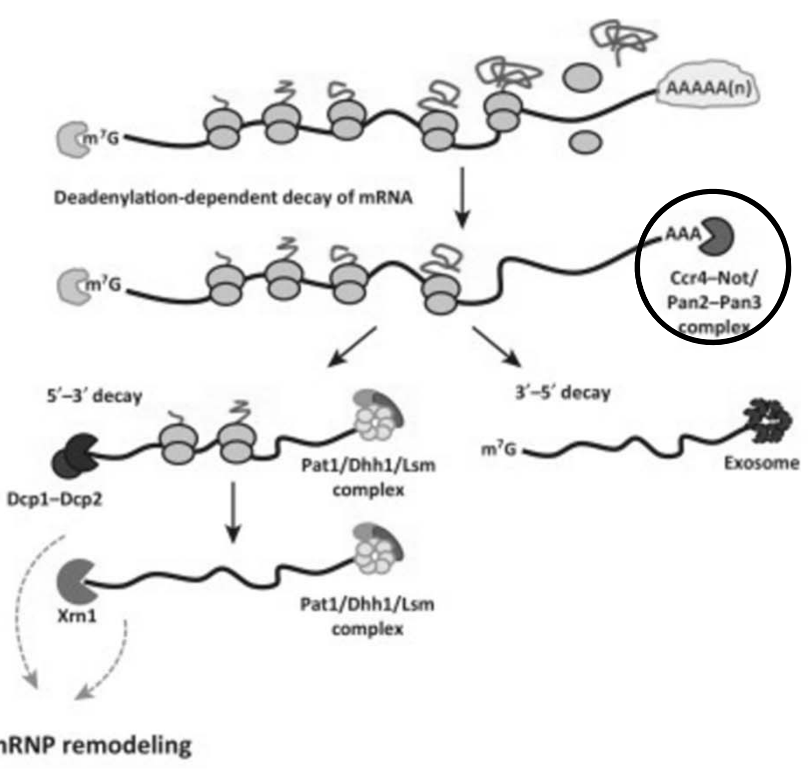


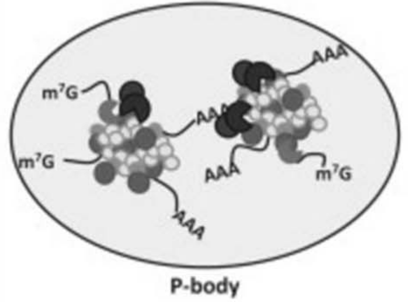
© 2006 Brooks/Cole - Thomson

The Impact of Alternative RNA Splicing



mRNA Decay





- Normal mRNA degradation pathways

1. Poly(A) shortening: catalyzed by the Ccr4–Not and poly(A)-specific deadenylases

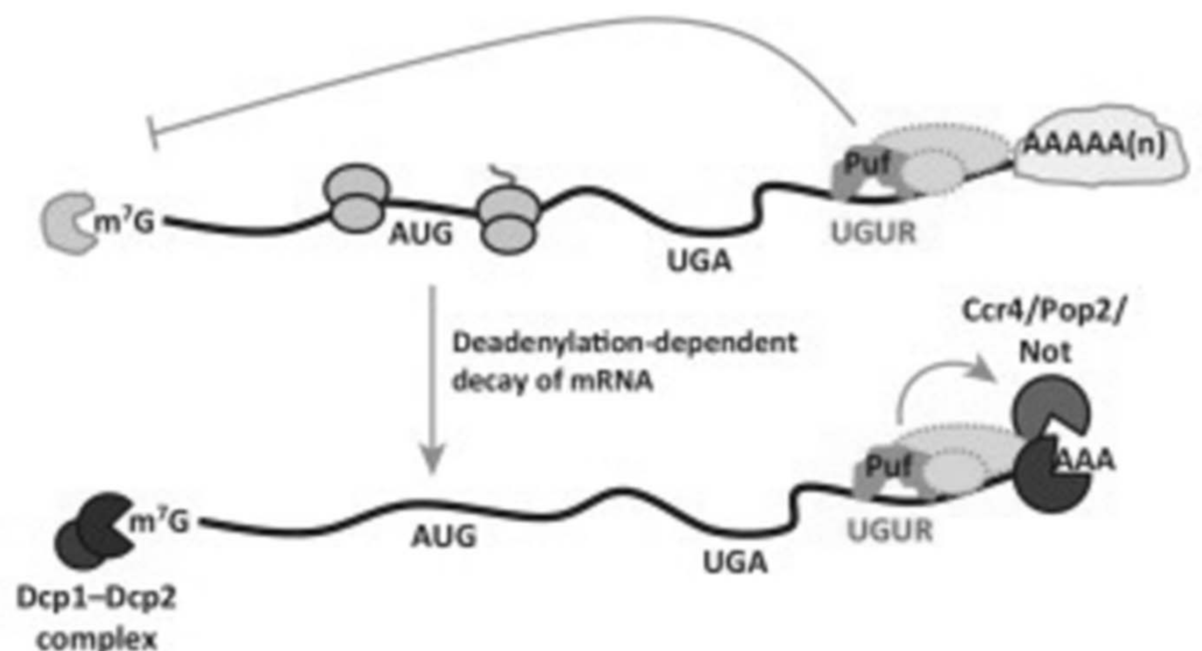
   1.1 Puf proteins

   1.2 RNA-induced silencing complex (RISC)

mRNA Decay

1.1 Puf proteins mediate translation repression and mRNA decay

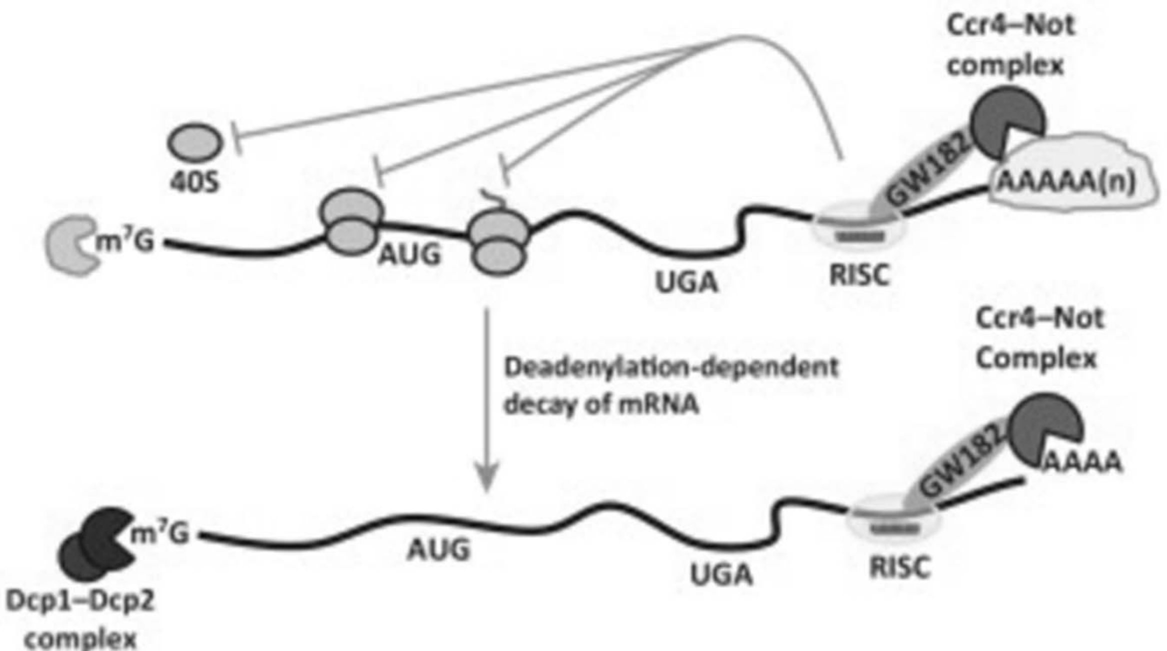
- Puf binding elements (UGUR) in the mRNA
- Recruitment of the Ccr4–Not deadenylase complex can trigger deadenylation-dependent mRNA decay



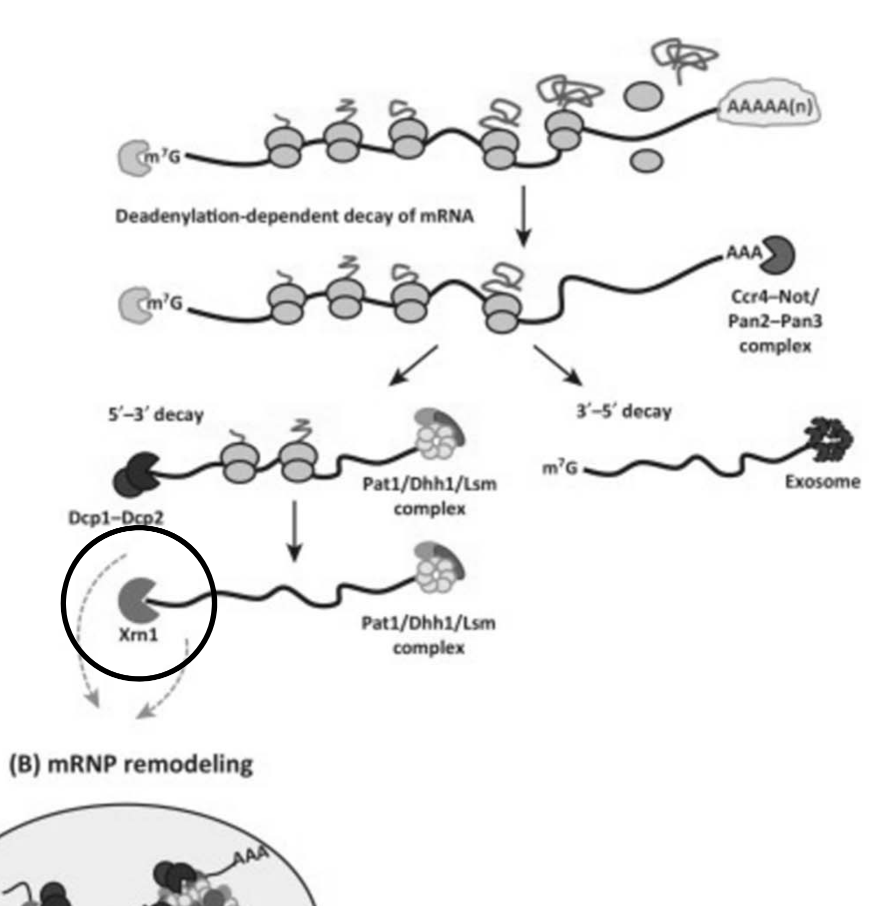
mRNA Decay

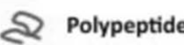
1.2 Binding of the RNA-induced silencing complex (RISC)

- inhibition of translation initiation: interfering with cap recognition, 40S recruitment, 60S subunit joining
- Interaction with the Ccr4–Not deadenylase complex triggers deadenylation-dependent mRNA degradation



mRNA Decay





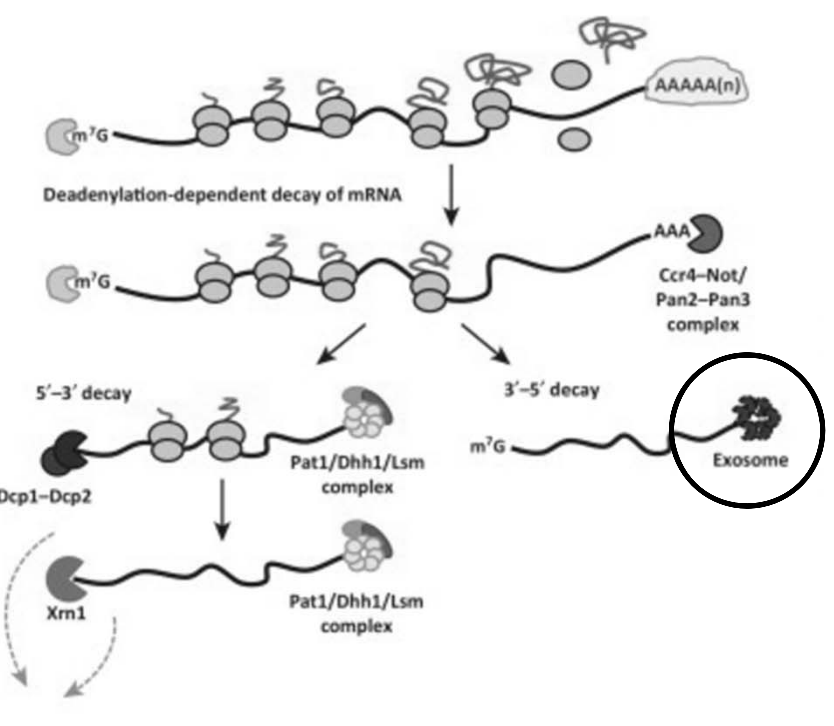
- Normal mRNA degradation pathways

2.1. 5′–3′ decapping-dependent decay: Xrn1-mediated exonucleolytic decay

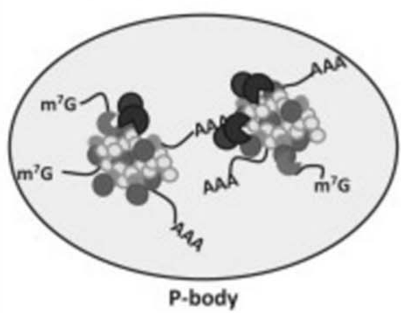
P-body

- RNA–protein cytoplasmic granules
- RNA targeted for decay
- Proteins:
  - Cap-binding complex: Dcp1–Dcp2, Xrn1
  - Pat1, Dhh1, and Lsm complex

mRNA Decay



(B) mRNP remodeling



Key: Cap-binding complex 80S ribosome Polypeptide

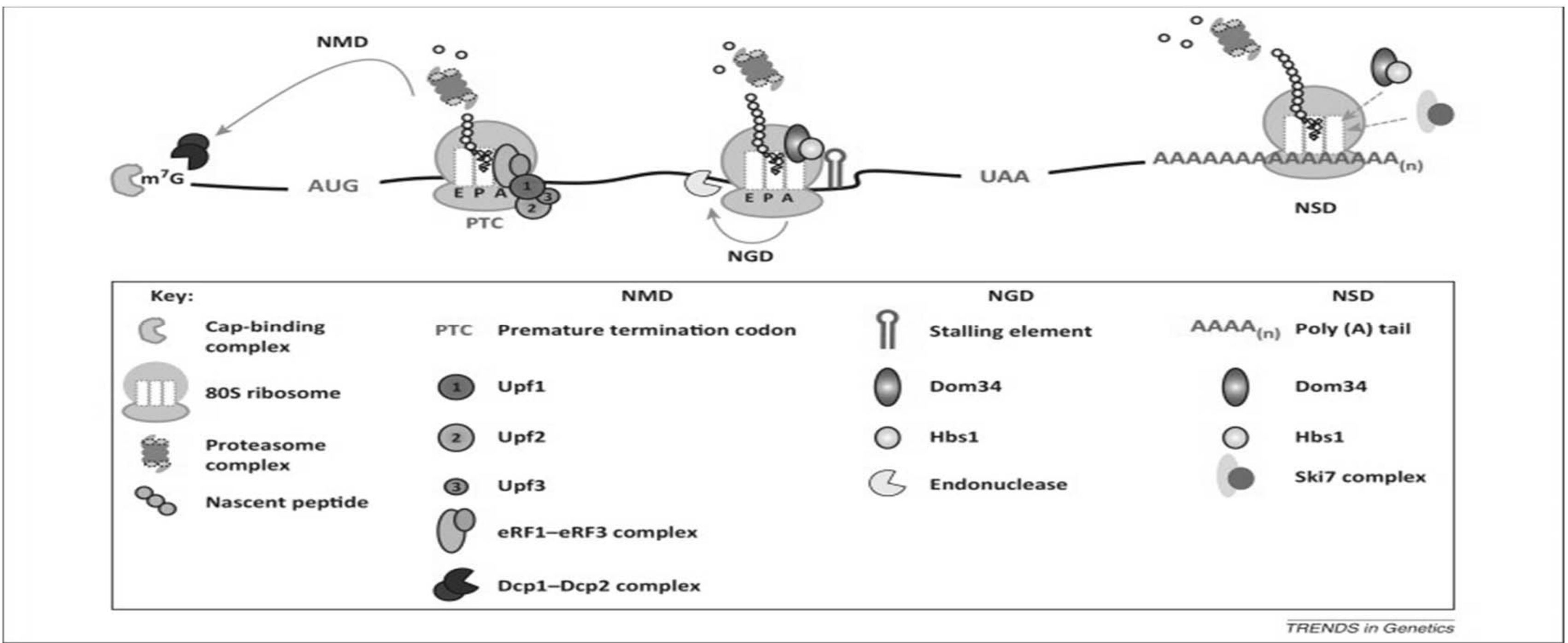
- Normal mRNA degradation pathways

2.2. 3′–5′ exosome-mediated exonucleolytic decay

Abnormal Translational Events Leading to Accelerated mRNA Decay

Abnormal translation events

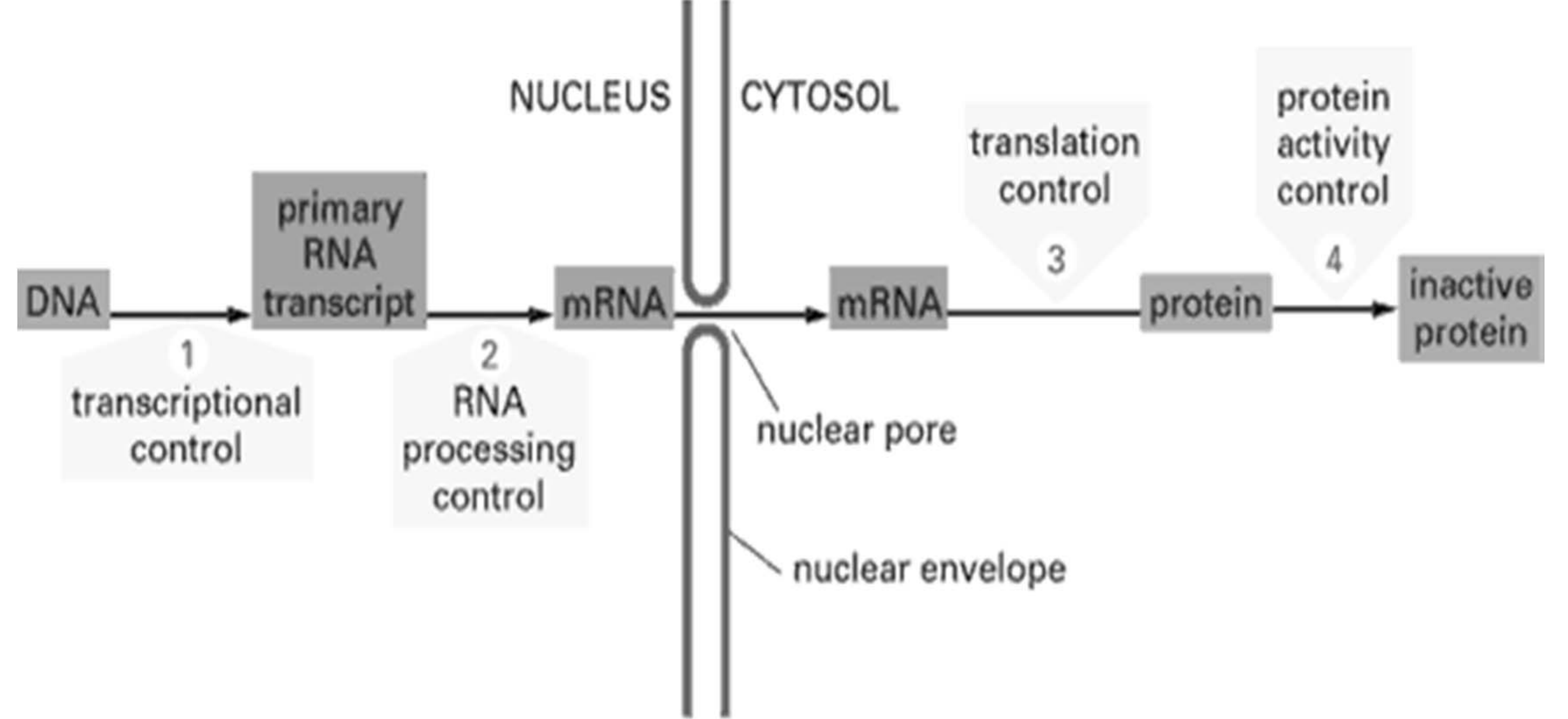
- A premature termination event → nonsense-mediated decay (NMD)
- An elongation stall → no-go decay (NGD)
- Poly(A) translation → nonstop decay (NSD)



To Explore Gene Regulation

4 steps to control gene expression:

- Transcriptional control: actinomycin D
- RNA processing control (5' capping, splicing, poly A)
- Translation control: cycloheximide vs. MG132
- Protein activity control



Loss-of-function vs. Gain-of-function

- Loss-of-function analysis (artificial gene interference): an operation that results in reduced or abolished protein function

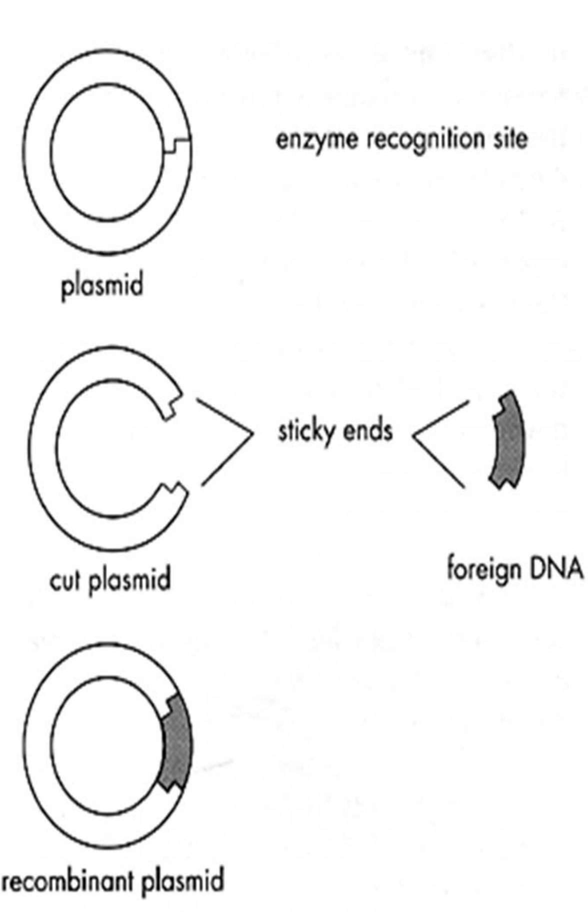
  e.g. siRNA library

- Gain-of-function analysis (ectopic expression): an operation that results in increased protein function

  e.g. cDNA library

Gene Manipulations

- Developed in 1970s and 1980s
- Recombinant DNA technology
  - Also known as genetic engineering or cloning
  - The ability to combine the DNA of one organism with the DNA of another organism
  - Use enzymes to cut (restriction enzyme) and paste (ligase) DNA into "recombinant" molecules
- Clone: organisms or cells of nearly identical genetic makeup derived from a single source



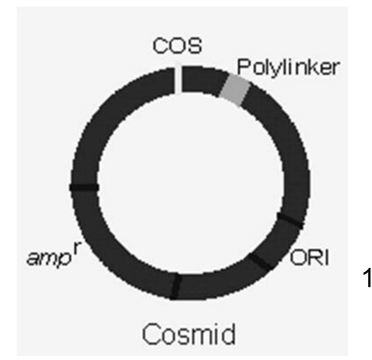
Creating recombinant plasmids

Gene Manipulations

Vector: a DNA molecule used as a vehicle to transfer foreign genetic material into another cell

- Major types of vectors:
  - Plasmids: a DNA molecule that is separate from the chromosomal DNA, and can replicate independently (1–20 kb of DNA)
  - Cosmids: cos sites + plasmid (40–50 kb of DNA)
  - Artificial chromosomes (150–350 kb of DNA): e.g., bacterial artificial chromosome (BAC), yeast artificial chromosome (YAC)
  - Virus (8–10 kb of DNA)

COS:
cohesive end **s**ite (sticky ends of two single-stranded segments)



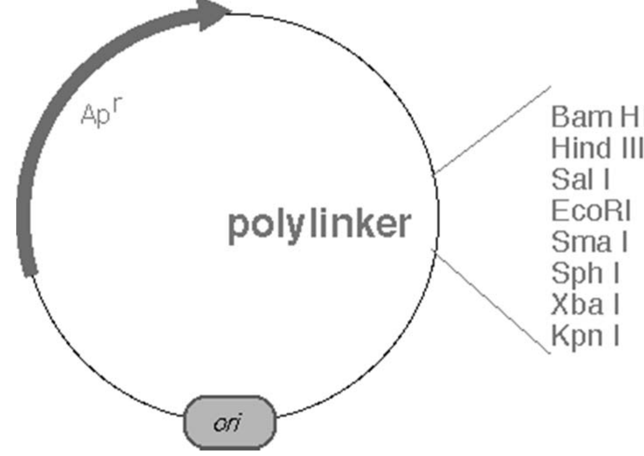
Gene Manipulations

- Major types of vectors:
  - Viruses: Retroviruses only infect dividing cells

	Adeno-associated virus (AAV)	Adenovirus	Lentivirus
Genome	ssDNA	dsDNA	ssRNA
Relative transduction efficiency	70%	100% (except blood cells)	70%
Host genome interaction	Non-integrating (transient)	Non-integrating (transient)	Integrating (long-lasting)
Tropism	Dividing and non-dividing cells	Dividing and non-dividing cells	Dividing and non-dividing cells
Immune response	Very low	High	Low

Gene Manipulations

- Common to all engineered vectors:
  - an origin of replication
  - a multi-cloning site
  - a selectable marker



- Transformation: insertion of a vector into the bacterial cells
- Transfection: insertion of a vector into the eukaryotic cells
- Transduction: insertion of virus into bacteria or cells

Factors Influencing Transfection Efficiency

- Transfection method

Chemical	Biological	Physical
Chemical methods that use carrier molecules to neutralize or impart a positive charge to the negatively charged nucleic acids and include: • Cationic lipid transfection • Calcium phosphate transfection • DEAE-dextran transfectopm • Delivery by other cationic polymers (e.g., polybrene, PEI, dendrimers)	Biological methods that rely on genetically engineered viruses to transfer non-viral genes into cells (also known as transduction) and include: • Viral delivery	Physical methods directly deliver nucleic acids into the cytoplasm or the nucleus of the cell and include: • Electroporation • Biolistic particle delivery (particle bombardment) • Direct microinjection • Laser-mediated transfection (phototransfection)

Diethylaminoethyl (DEAE)-dextran: a polycationic derivative of the carbohydrate polymer dextran

https://www.thermofisher.com/tw/zt/home/references/gibco-cell-culture-basics/transfection-basics/gene-delivery-technologies/deae-dextran-mediated-delivery.html

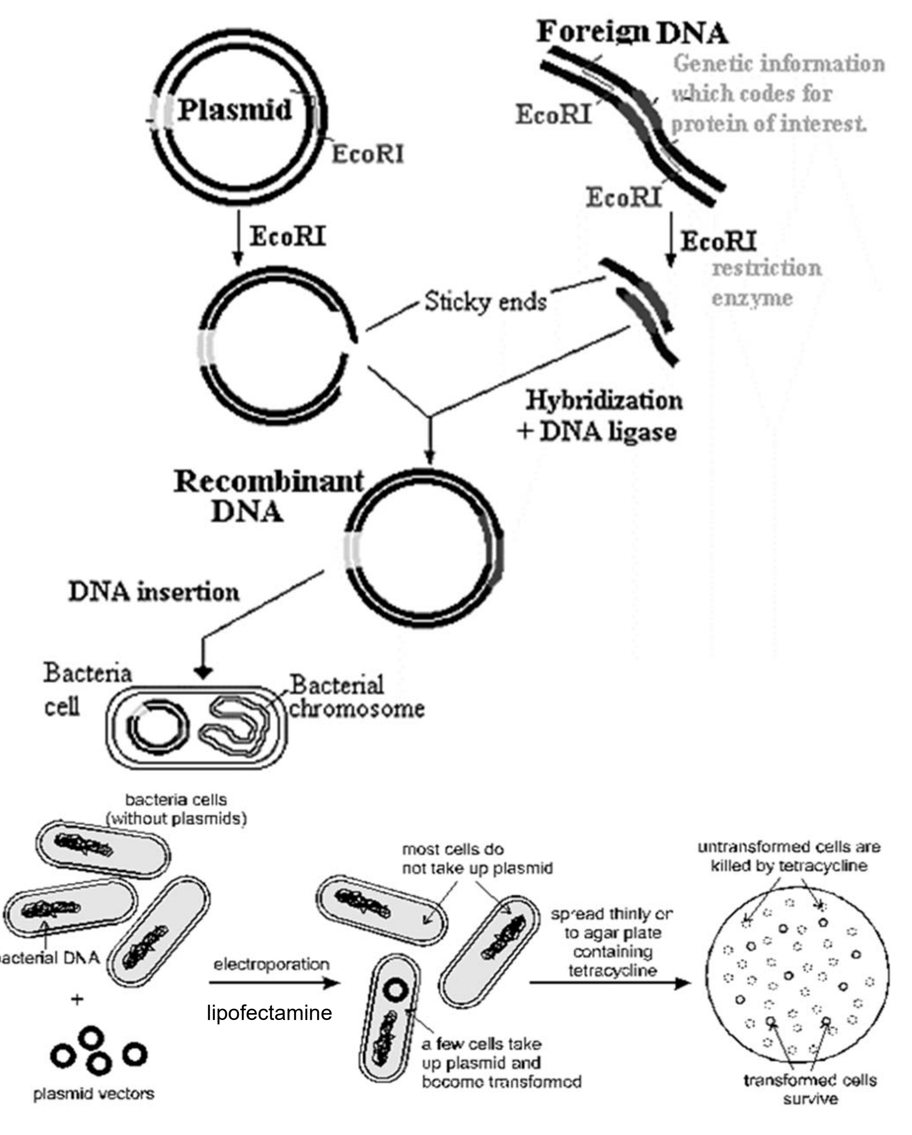
Factors Influencing Transfection Efficiency

- Cell type
- Cell health and viability
  - at least 90% viable prior to transfection
  - have had sufficient time (24 h) to recover from passaging
- Confluency
  - 70–90% confluency for adherent cells
  - 5×10^5 to 2×10^6 cells/mL for suspension cells
- Media: fresh medium, especially if any of the components are unstable

Factors Influencing Transfection Efficiency

- Antibiotics
  - cationic lipid reagents increase cell permeability → antibiotics resulting in cytotoxicity
- Type of molecule
  - supercoiled plasmid DNA: most efficient
  - linear DNA: lower DNA uptake but yields optimal integration of DNA into the host genome

Recombinant Bacteria



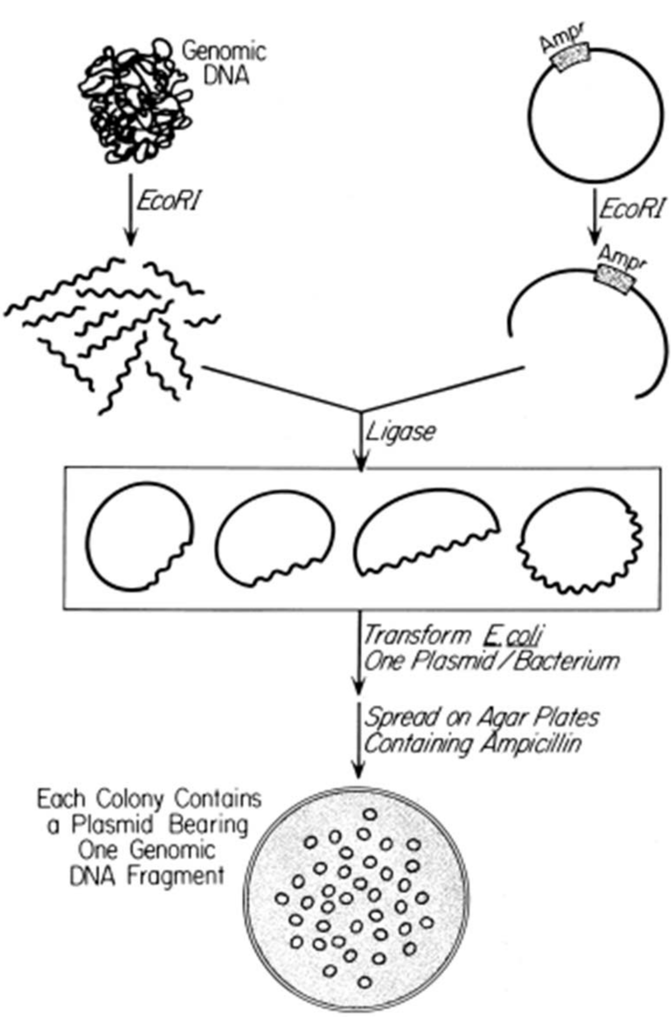
1. Remove bacterial DNA (plasmid)
2. Cut the Bacterial DNA with "restriction enzymes"
3. Cut the DNA from another organism with "restriction enzymes"
4. Combine the cut pieces of DNA together with ligase and insert them into bacteria
5. Reproduce the recombinant bacteria
6. The foreign genes will be expressed in the bacteria

Question

What is the meaning of genomic library?

- A genomic library is a collection of the total genomic DNA from a single organism
- The DNA is stored in a population of identical vectors, each containing a different insert of DNA

Genomic Library



- A population of host bacteria, each of which carries a DNA molecule that was inserted into a cloning vector, such that the collection of cloned DNA molecules represents the entire genome of the source organism (e.g. genomic library)

Restriction Enzyme

- Exonucleases: enzymes that work by cleaving nucleotides from the end (exo) of a polynucleotide chain
- Endonuclease: enzymes cleaves phosphodiester bonds in the middle (endo) of a polynucleotide chain

- Cuts DNA at or near restriction sites (specific recognition nucleotide sequences)
  - Between 4 and 8 bases (occur once every 256 (4^4) – 65536 (4^8) bp)
  - Palindromic sequence
    - Mirror-like: GTAATG
    - Complementary: GTATAC
  - Sticky or blunt end

上海自來水來自海上
花蓮白種人種白蓮花

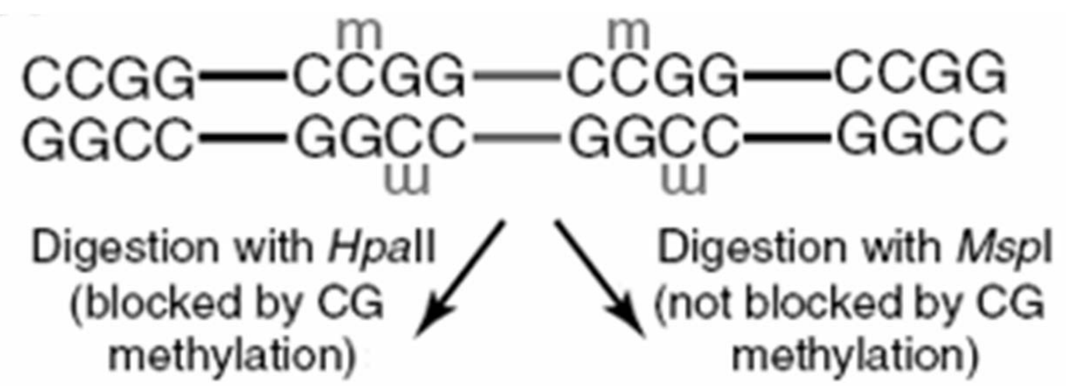
G|AATTC
CTTAA|G

CCC|GGG
GGG|CCC

Restriction Enzyme

- Neoschizomer: different enzymes that recognize the same location but cleave in different location
- Isoschizomer: different enzymes that recognize and cleave in the same location
  - Methylation Isoschizomer

*Hpa*II vs *Msp*I



Restriction Enzyme

- Classification:
  - Type I: cleave at sites remote (> 1,000 bp) away from recognition site; require both ATP and S-adenosyl-L-methionine
  - Type II: cleave within (IIP) or at short specific distances [IIS, IIG (or IIC)] from recognition site; require magnesium (e.g., *Eco*RI)
  - Type III: cleave at sites a short distance (20-30 bp) from recognition site; require ATP (tagging enzyme)
  - Type IV: target modified DNA (e.g., methylated DNA)
  - Type V: utilize guide RNAs to target specific non-palindromic sequences found on invading organisms (e.g., the cas9-gRNA complex from CRISPRs)

Restriction Enzyme

- Nomenclature based on bacterial genus, species and strain

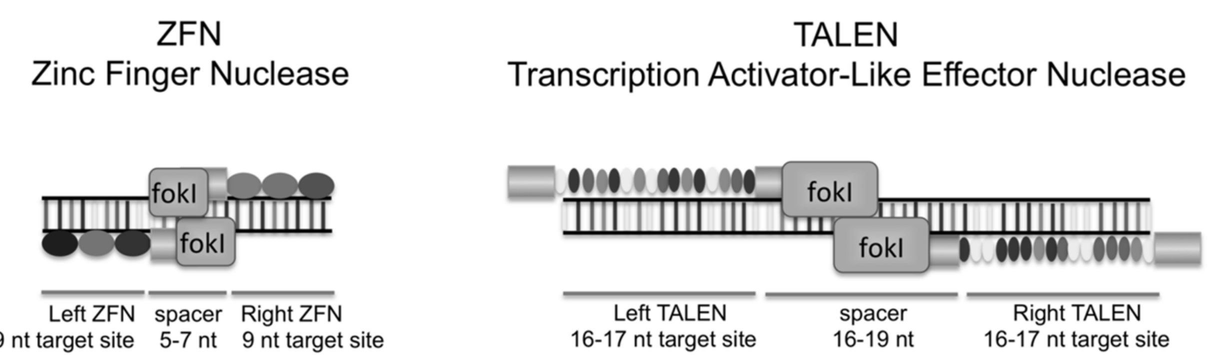
  E.g., *Eco*RI

  E: Escherichia *co: coli*

  R: RY13 I: first identified

- Artificial restriction enzymes: fusing a natural or engineered DNA binding domain [Zinc finger nuclease (ZFN); TAL (transcription activator-like) effector nuclease (TALEN)]

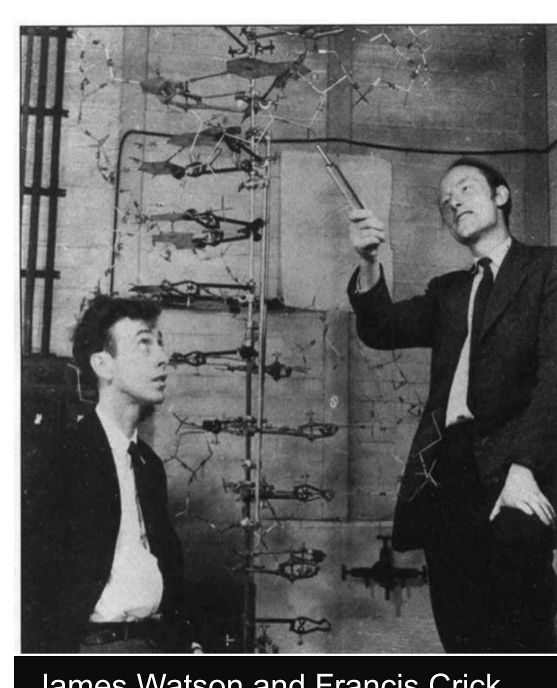
  to a nuclease domain (cleavage domain of the type IIS restriction enzyme *Fok*I)

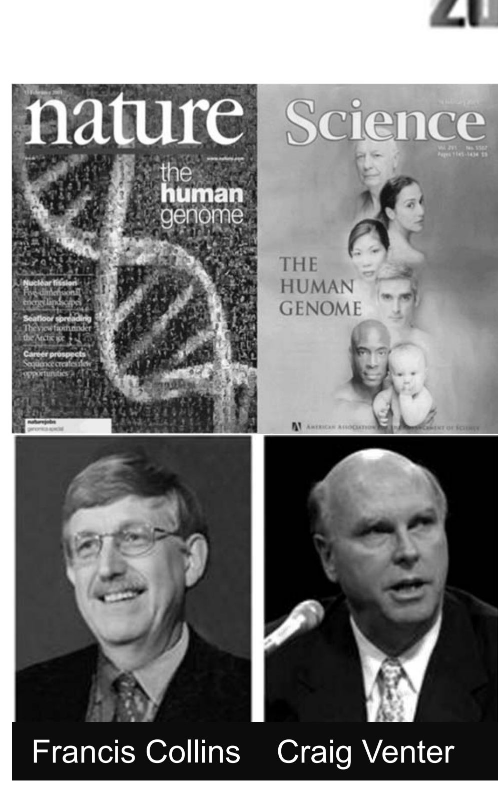


Kingdom → Phylum → Class → Order → Family → Genus → Species

Human Genome Project

- Started in 1990 by 20 centers of six nations, led first by Watson and after 1992 by Collins
- The completed sequence of the human genome (3×10^9 bp) was published in April 2003


James Watson and Francis Crick



Francis Collins Craig Venter

Question:

Which technique can be applied to measure gene expression?

A) Hybridization-based techniques
B) PCR-based techniques
C) Sequence-based techniques
D) All of them

Comparative Gene-expression Analysis

- Hybridization-based techniques
  - Northern blot
  - Microarrays
- PCR-based techniques
  - Differential display
  - Rapid amplification of cDNA ends (RACE)
  - Subtractive hybridization
- Sequence-based techniques
  - EST (expressed sequence tags)
  - SAGE (serial analysis of gene expression)
  - CAGE (capped analysis of gene expression)
  - Next generation sequencing (NGS)

Question:

Which hybridization technique can be used to detect the amount of polysaccharide?

A) Eastern blotting
B) Western blotting
C) Southern blotting
D) Northern blotting

Agarose vs. Polyacrylamide Gels

- Agarose gels
  - to resolve large fragments of DNA
- Polyacrylamide gels
  - to separate shorter nucleic acids, generally in the range of 1−1000 base pairs, based on the concentration used
- Gels without a denaturant (e.g., SDS): native gels
  - Secondary structure affects migration at different rates
  - Secondary structure will not form in denaturing gels → only the length of the DNA will affect mobility

Agarose Gels		Polyacrylamide Gels	
% agarose	Size Range for Optimum Resoultion (bp)	% acrylamide	Size Range for Optimum Resoultion (bp)
0.5	1,000-30,000	3.5	1,000-2,000
0.7	800-12,000	5	80-500
1.0	500-10,000	8	60-400
1.2	400-700	12	25-150
1.5	200-500	15	25-150
		20	6-100

Troubleshooting Gel Electrophoresis

- *Blurry bands*
    - Too much DNA (100−250 ng/mm well width)
    - Too much salt

- *Bands in the wrong place*
    - Heat nucleic acids before running on a native gel
    - Run gel >20 V/cm (run gel slowly → sharper bands)
    - Gel temp. >30 °C

- *Loading buffer floats away*
    - Some salts built up in the wells

  → Rinse wells with running buffer before loading

  → Add a little more glycerol to the dye

Question

Which gel is the best choice for running microRNA samples?

A)10% Polyacrylamide gels

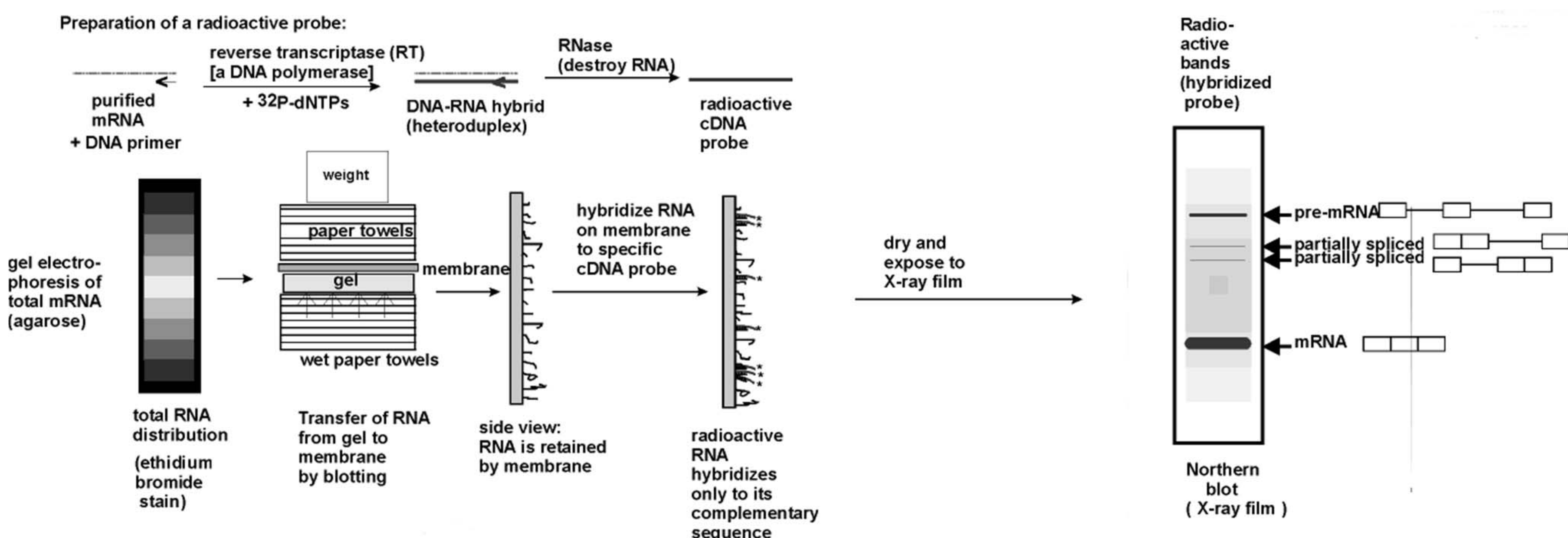
B) 1% Polyacrylamide gels

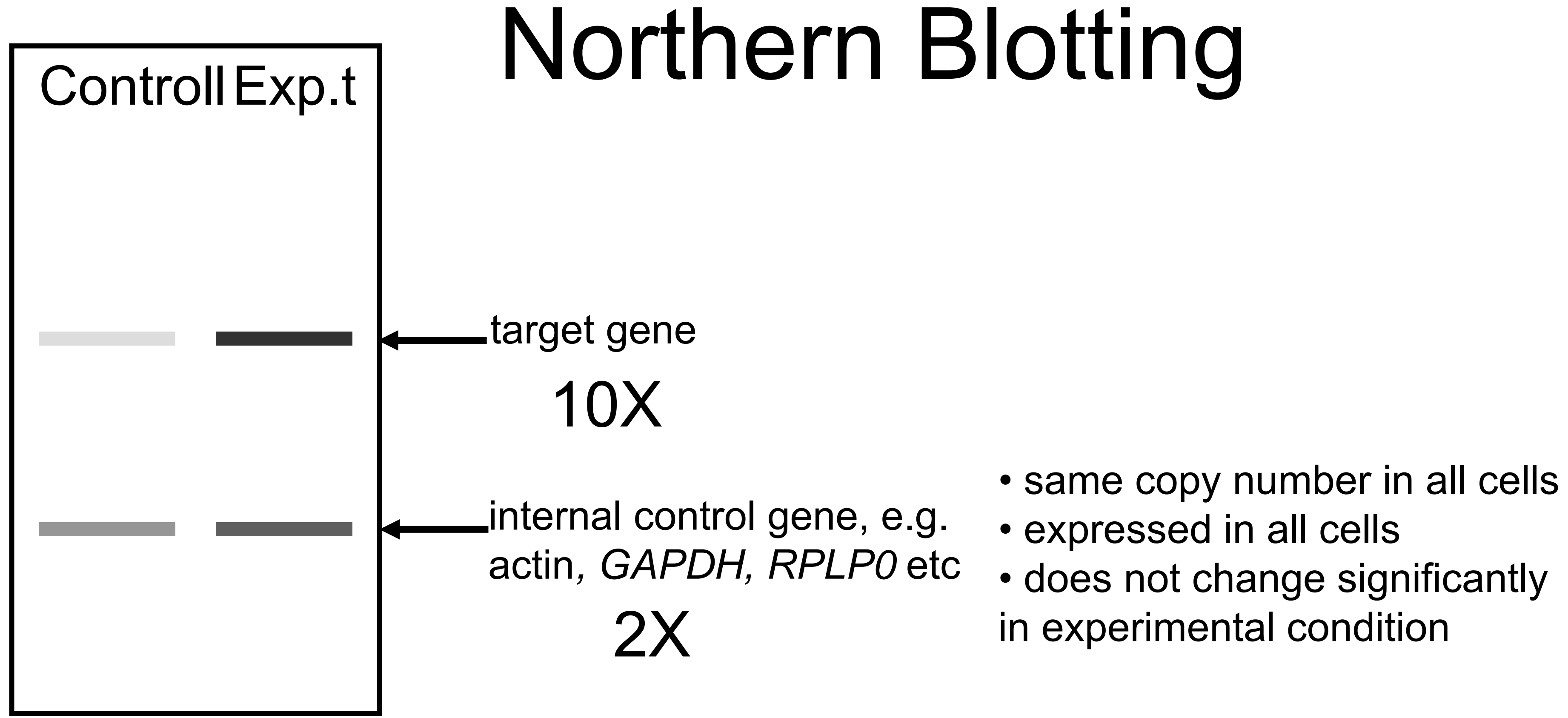
C) 1% Agarose gels

D) 0.5% Agarose gels

Hybridization-based techniques$_1$

- Northern blotting
  - Detect a specific human RNA sequence in total cell RNA (probe: labeled DNA; target: RNA on membrane)





Corrected fold increase = 10/2 = 5

Ratio target gene in experimental/control = $\dfrac{\textbf{fold change in target gene}}{\textbf{fold change in reference gene}}$

Hybridization-based techniques$_2$

- Microarray
  - Small, solid supports onto which the sequences from thousands of different genes or EST (Expressed Sequence Tag) are immobilized at fixed locations
  - Support: glass microscope slides, silicon chips or nylon membranes
  - Spots (probes): DNA, cDNA, or oligonucleotides (20-70mers)
    - cDNA: high sensitivity, low specificity
    - Oligomer: high specificity, low sensitivity
  - Tagged samples (targets): reverse-transcribed RNA into cDNA with a tracking molecule like a radioisotope, fluorescent dye, or an affinity molecule (e.g., biotin)

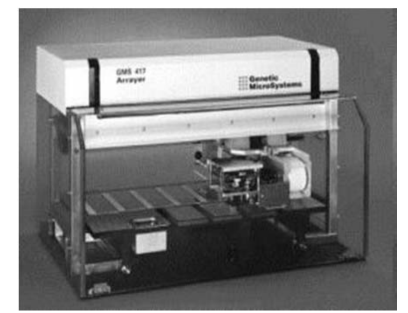
Microarray Platforms

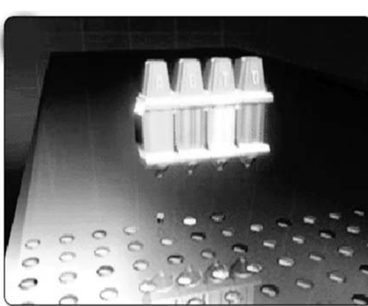
Two-Color Platform:

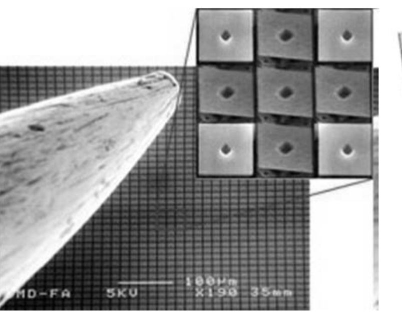
- In-house spotted arrays
- Prespotted array—inkjet print (Agilent)

One-Color Platform:

- GeneChip – photolithograph (Affymetrix)
- Maskless array synthesis – Roche (NimbleGen)
- BeadChip – Illumina







Pharmacogenomics 7:123, 2006

Advantages & Applications of Microarray

- Advantages:
  - Favor small sample size
  - High throughput: gather data on thousands of genes (genome) in a single experiment
  - Quantitative analysis
- Application:
  - Gene expression
  - Genotyping-polymorphisms (SNP) and copy number variation
  - Binding site identification: ChIP-on-chip
  - Epigenetics: methylation chip
  - Epigenetics: non-coding RNA, e.g., microRNA, lncRNA

High Throughput of Traditional Experiment