Lecture 8: Analysis of Gene Expression (Part 1, Global Gene Expression)

Global Gene Expression

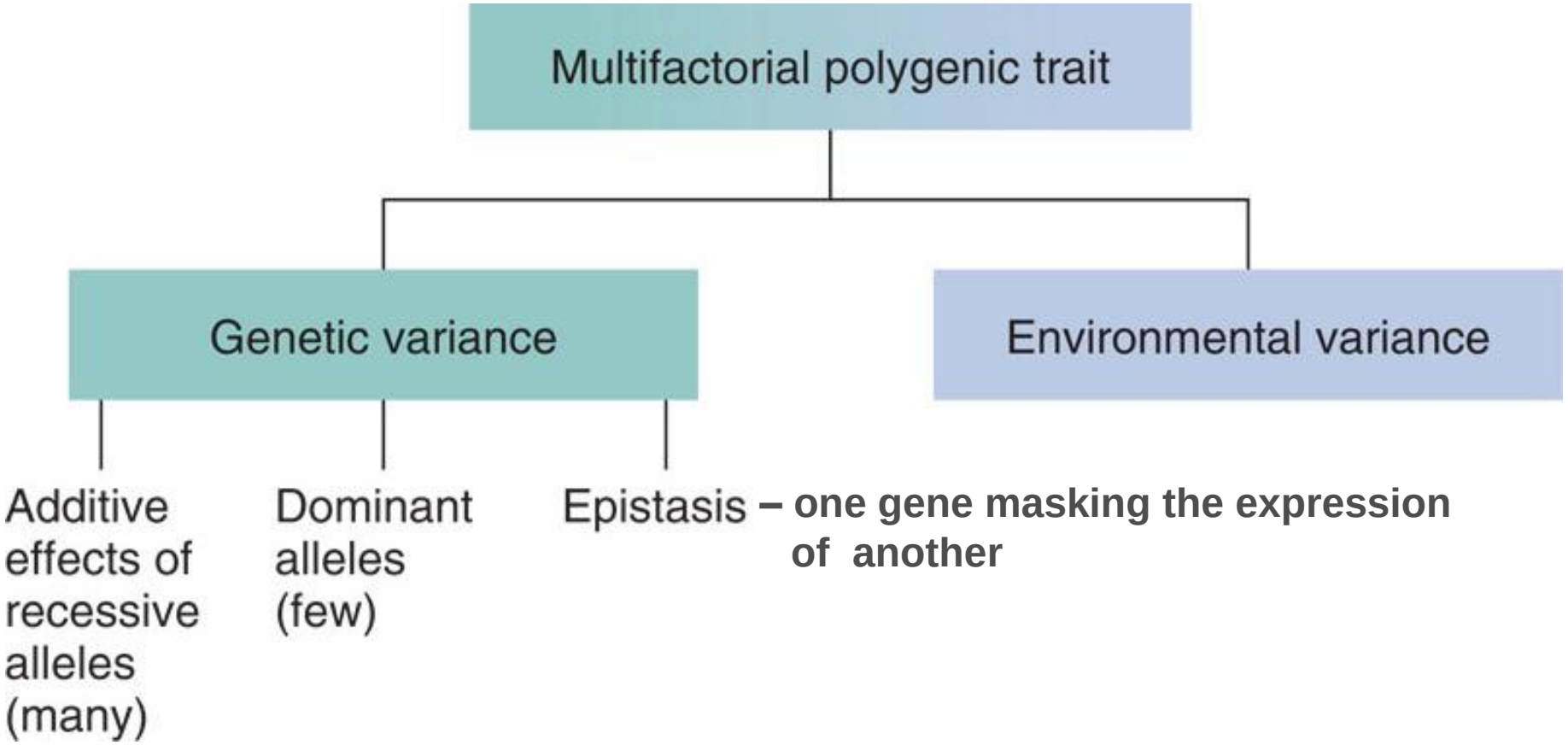
- A new field of genetic analysis called global gene expression has emerged in recent years, driven by the realization that traditional techniques of linkage and association analysis can be applied to thousands of transcript levels measured by microarrays.
- Genetic dissection of transcript abundance has shed light on the architecture of quantitative traits, provided a new approach for connecting DNA sequence variation with phenotypic variation and improved our understanding of transcriptional regulation and regulatory variation.

Global Gene Expression

- Studies in a number of species have documented abundant heritable variation in gene expression among individuals and strains.
- Linkage analysis has been used to map thousands of loci that affect gene expression.
- Expression traits consistently show complex inheritance, explicable only by multiple underlying loci and possibly interactions among the loci.
- Many types of genetic complexity are observed across the thousands of expression traits.
- A locus that affects gene expression can be classified according to its location as "local" (near the genomic location of the gene) or "distant" (elsewhere in the genome).
- Many gene expression traits are affected by local regulatory variation.
- It appears that most, but not all, local regulatory variations function in cis, with perhaps a quarter to a third acting in trans.
- Most transcripts link to loci that are distant from the genomic locations of the genes that encode the corresponding transcripts.
- One common feature of these loci is "hot spots": individual loci that affect large numbers of transcripts.
- Future research will focus on high-throughput identification and characterization of polymorphisms that affect expression, extending studies to population samples and applying the approach to other global molecular phenotypes.

Heritability (H)

H estimates the proportion of the phenotypic variation in a population due to genetic differences.

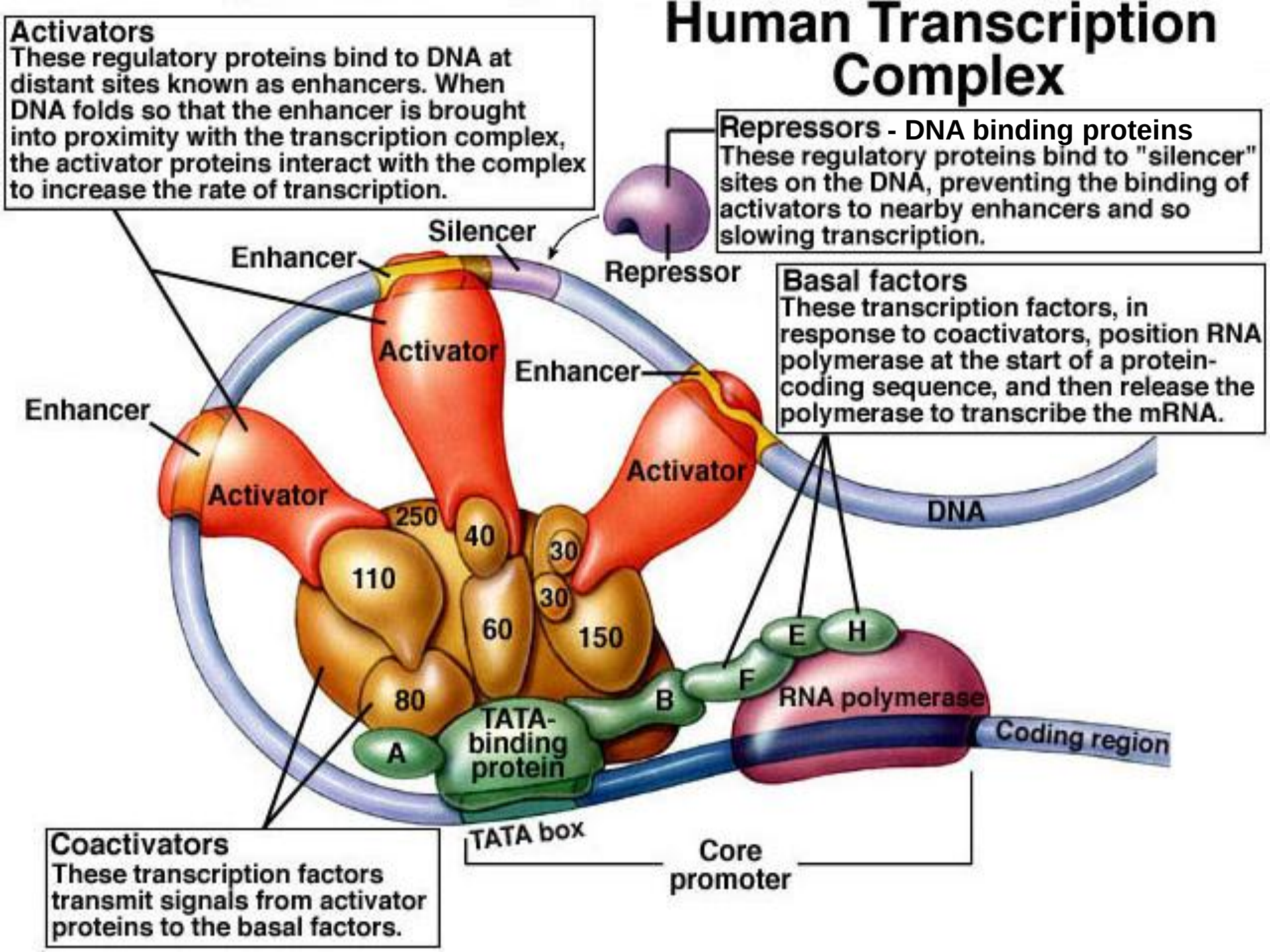


- **Genetic variants are determined mostly by the additive effects of recessive alleles of different genes.**
- **Variants also can be influenced by the effects of a few dominant alleles and by epistasis.**

Cis-acting Elements and Trans-acting Factors

- The transcription initiation complex is composed of promoter sequences and DNA binding proteins.
- These two components of transcription are normally described as cis-acting elements and trans-acting factors.
- Cis-acting elements are DNA sequences in the vicinity of the structural portion of a gene that are required for gene expression.
- Trans-acting factors usually are considered to be proteins that bind to the cis-acting sequences to control gene expression.

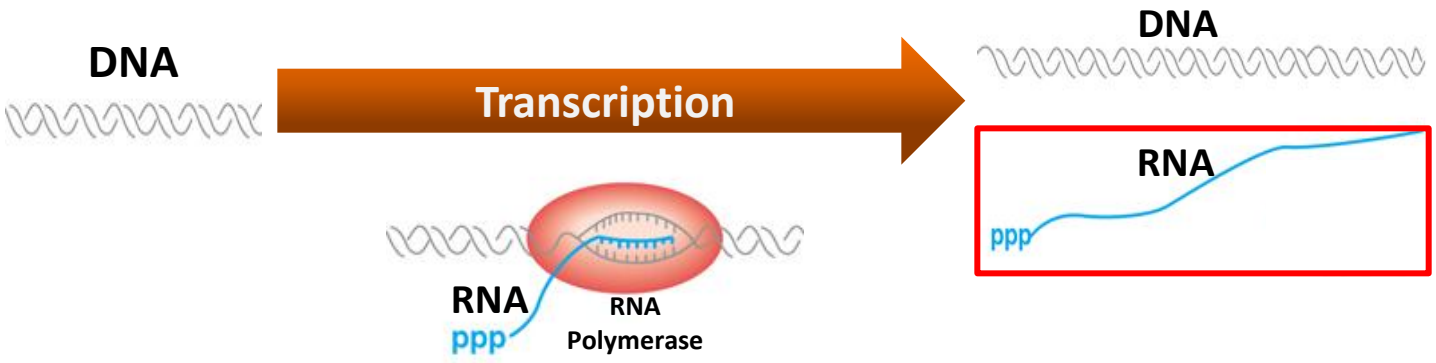
Human Transcription Complex



Analysis of Transcription

Two general approaches:

1. **Measure the amount of a specific transcript produced by a cell or a group of cells.**
2. **Measure the transcriptional activity of a particular gene or genes under certain conditions.**



Cautions for Handling RNA

Preparation for northern analysis:

1. **EVERYTHING must be thoroughly cleaned of RNAses, and gloves must be worn at all times.**
2. **Use RNAse-ZAP or some other commercially available RNAse cleaner to clean the work area and all equipment that may come into contact, directly or indirectly, with the RNA.**
3. **If no cleaners are available, use a dilute solution of bleach.**
4. **Clean everything with 95% ethanol, which will facilitate evaporation of any residual RNAse-ZAP.**
5. **Place signs around the work area that stress the importance of no contamination.**
6. **Use only DEPC-treated (diethylpyrocarbonate) chemicals.**
7. **Chemicals can be DEPC-treated by adding two ml of DEPC per liter of solution, followed by autoclaving.**

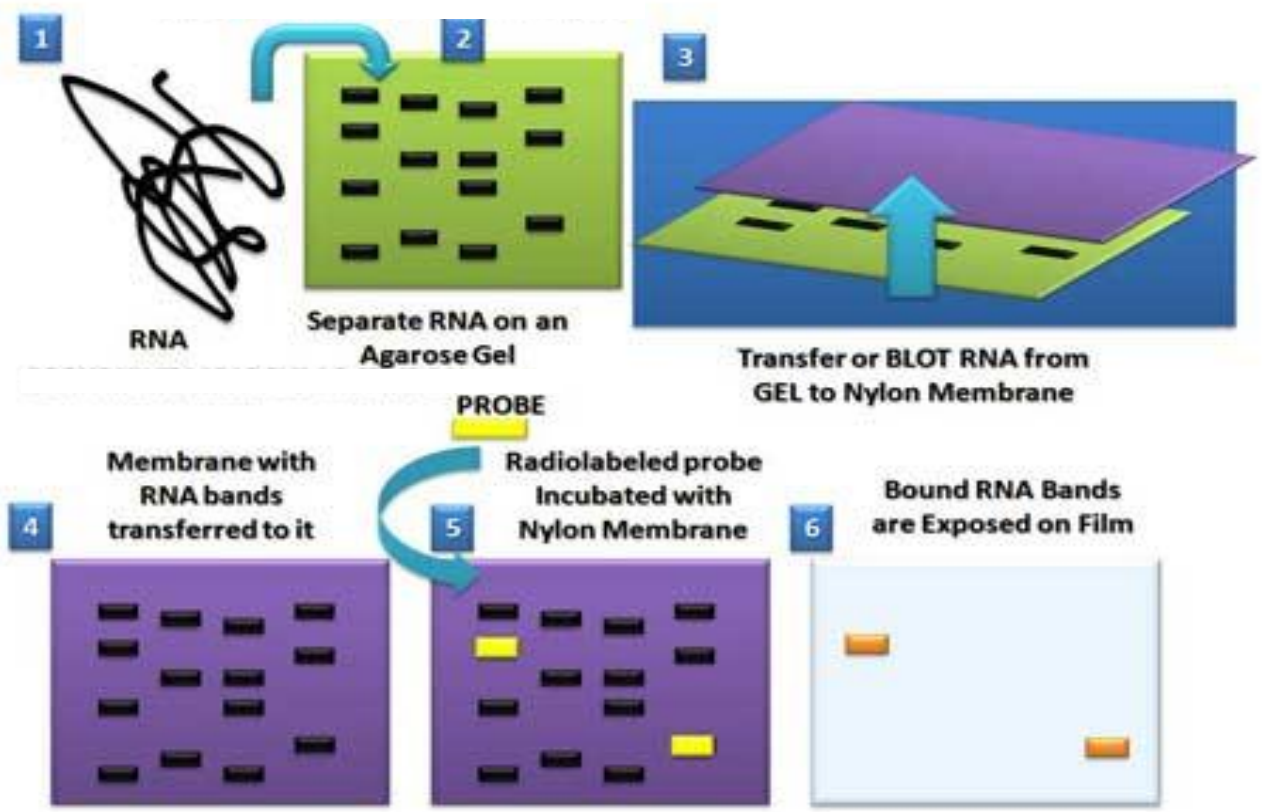
Protocol for RNA Isolation by TRIzol and RNeasy Column Binding for Affymetrix Small Sample Preparation

- The basis for this protocol is that TRIzol is used to initially lyse the cells and solubilize nucleic acids and proteins and other biomolecules.
- The organic mixture (plus chloroform) is separated using a phase-lock gel tube, which ensures that virtually all of the aqueous phase is collected without contamination from the interface or organic layer.
- RNA is then purified from the aqueous phase by RNeasy (Qiagen) purification.
- This procedure is preferred over alcohol precipitation for smaller amounts of RNA because of the risk of losing or not seeing a miniscule nucleic acid pellet.
- It also reduces the risk of organic solvent contamination of the RNA solution.
- Any of the RNeasy-style technologies can be used : the RNeasy mini-extraction kit, MiniElute kit or the Affymetrix sample clean-up kit.

Quantification of Transcripts

- Northern Blot Analysis
- RNase Protection Assay
- Primer Extension Assay
- Quantitative PCR
- Microarray
- Serial Analysis of Gene Expression
- Differential Display
- Subtractive Hybridization
- RNA sequencing

Northern Blot Analysis



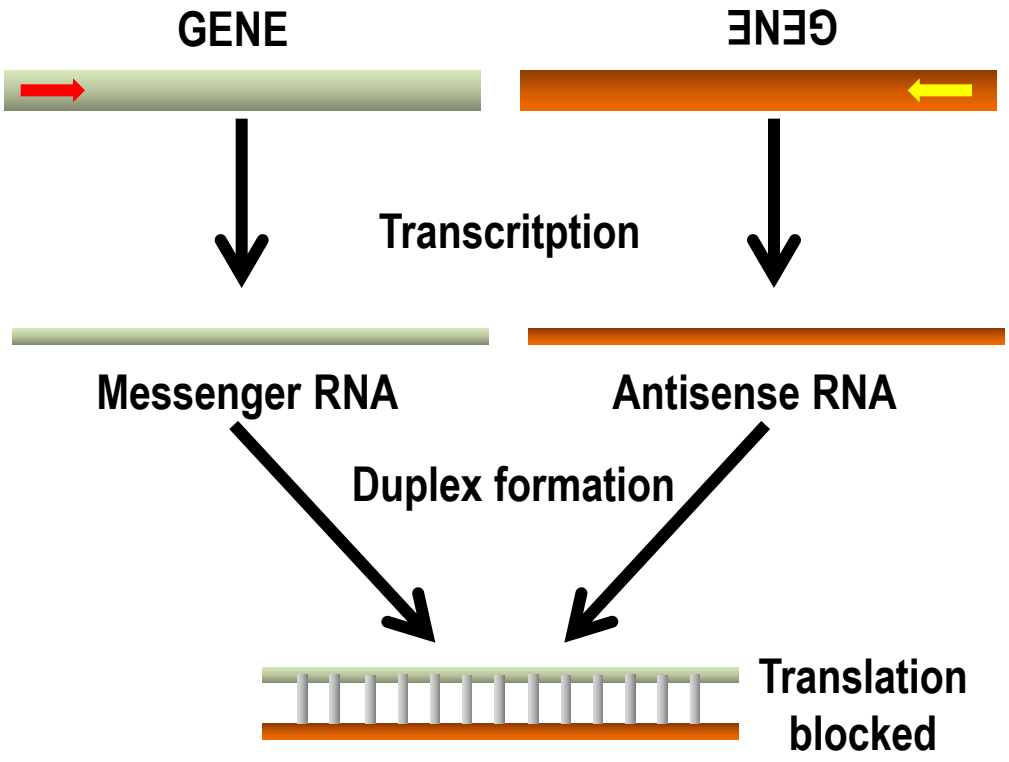
- It is the gold standard for quantifying RNA molecules.
- It facilitates direct measurement of RNA molecules from samples.
- It provides information about a RNA transcript, e.g., size and abundance, alternative splicing variants, etc.)

RNase Protection Assay

- **RNase protection assay is used to identify individual RNA molecules in a heterogeneous RNA sample extracted from cells.**
- **The extracted RNA is first mixed with anti-sense RNA probes that are complementary to the sequence or sequences of interest and the complementary strands are hybridized to form double-stranded RNA.**
- **The mixture is then exposed to ribonucleases that specifically cleave only single-stranded RNA but have no activity against double-stranded RNA.**
- **When the reaction runs to completion, susceptible RNA regions are degraded to very short oligomers or to individual nucleotides.**
- **The surviving RNA fragments are those that are complementary to the added anti-sense strand and contain the sequence of interest.**
- **Separation of the hybridization mixture on a polyacrylamide gel and exposure to an X-ray film will identify the RNA of interest.**
- **This technique can identify one or more RNA molecules of known sequence at low concentration.**
- **RNase protection assay is more sensitive and quantitative than northern blotting and is useful to map introns and 5′ and 3′ ends of transcribed gene regions.**

Antisense RNA

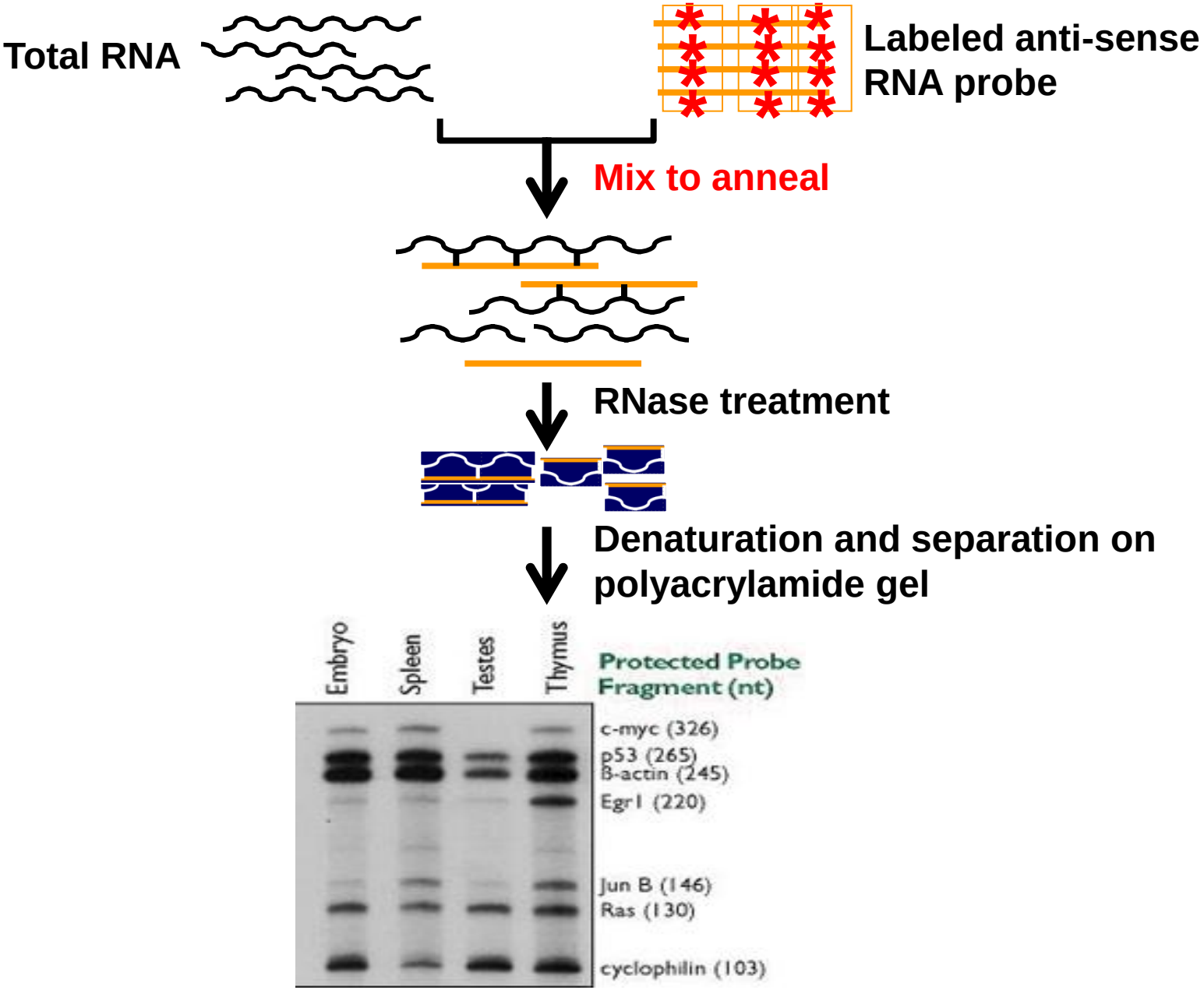
- **Antisense transcripts represent a class of long noncoding RNAs (ncRNAs).**
- **Antisense transcripts are transcribed from the strand opposite to that of the sense transcript of either protein-coding or non-protein-coding genes.**
- **RNA can form duplexes with its complementary molecule.**
- **Cells contain genes that are naturally translated into antisense RNA molecules capable of blocking the translation of other genes in the cell.**
- **Apparently, this is another method of regulating gene expression.**
- **Antisense transcripts preferentially accumulate in the nucleus.**



cRNA = synthetic transcripts of a specific DNA molecule or fragment, made by an *in vitro* transcription system; RNA derived from cDNA through standard RNA synthesis

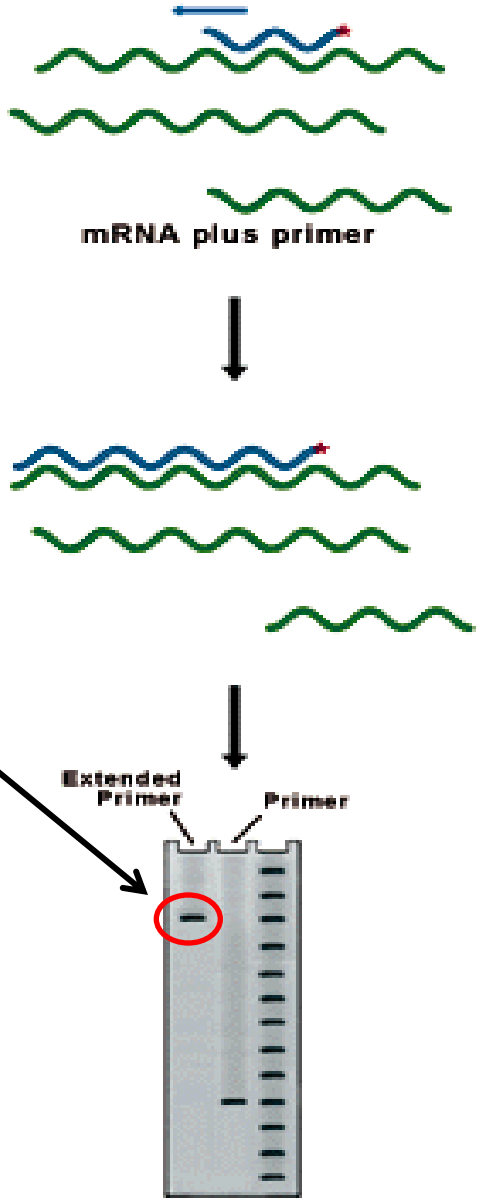
Nature Reviews Genetics **14: 880–893 (2013)**

RNase Protection Assay



Primer Extension Assay

- In primer extension, the primer introduced to the mRNA pool will hybridize with the RNA of interest if it is present.
- Hybrids are then extended by reverse transcriptase.
- The information gained through this method includes the confirmation of the presence of the RNA of interest, the location of the transcription start site, and, if an excess of primer is used, the amount of the particular RNA present in the sample.



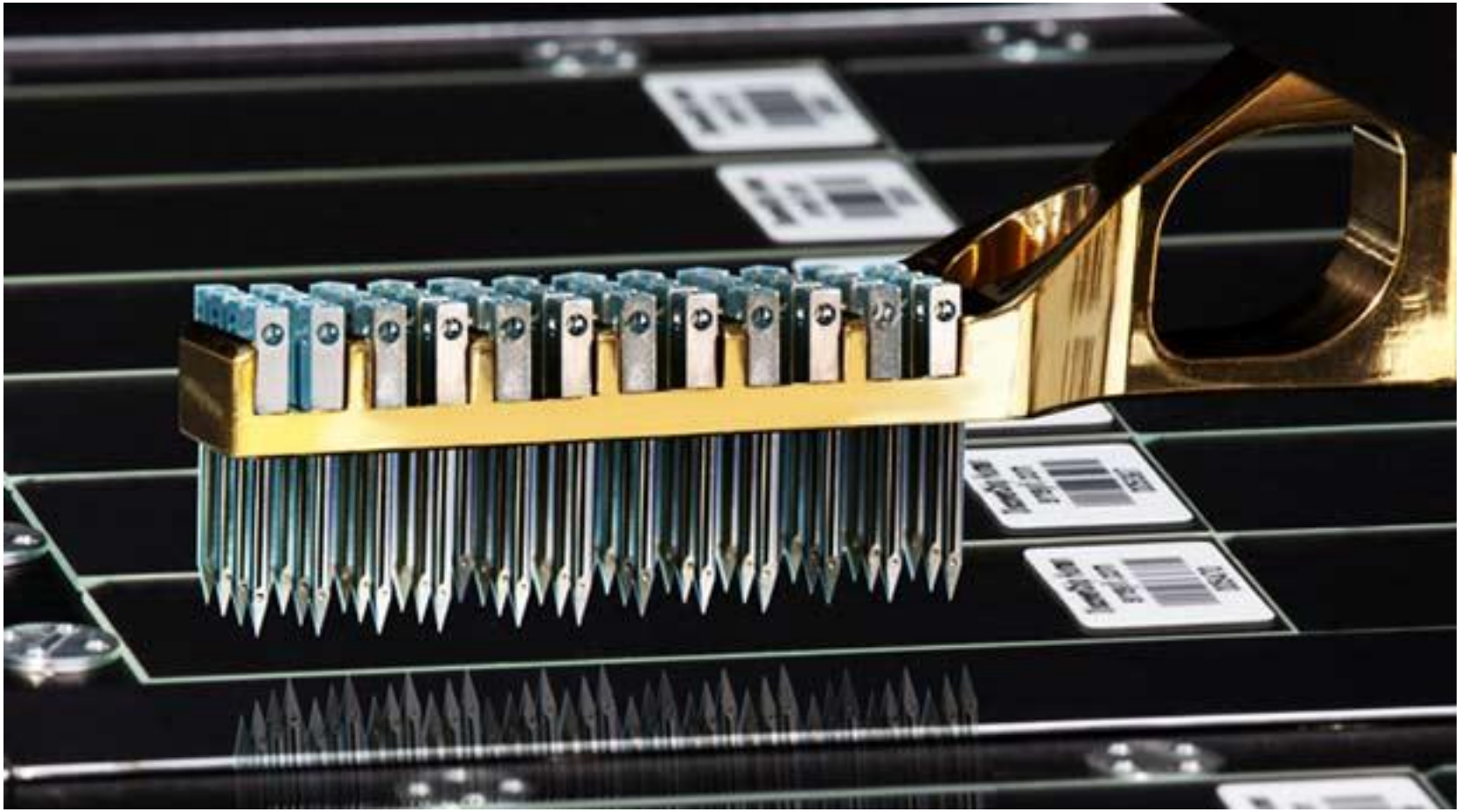
Analysis by Microarrays

- A microarray is a tool for analyzing gene expression that consists of a small membrane or glass slide containing samples of many genes arranged in a regular pattern.
- Microarrays are a significant advance both because they may contain a very large number of genes and because of their small size.
- Microarrays are useful when one wants to survey a large number of genes quickly or when the sample to be studied is small.
- Microarrays may be used to assay gene expression within a single sample or to compare gene expression in two different cell types or tissue samples, such as in healthy and diseased tissue.
- Because a microarray can be used to examine the expression of hundreds or thousands of genes at once, it has revolutionized the way scientists examine gene expression.

Types of Microarrays

- There are three basic types of samples that can be used to construct DNA microarrays, two are genomic and the other is "transcriptomic", i.e., it measures mRNA levels.
- What makes them different from each other is the kind of immobilized DNA used to generate the array and, ultimately, the kind of information that is derived from the chip.
- The target DNA used will also determine the type of control and sample DNA that is used in the hybridization solution.

Spotting a Glass Slide for Microarray Analysis



Colors of a Microarray

Serial Analysis of Gene Expression (SAGE)

- SAGE is used to produce a snapshot of the mRNA population in a sample of interest in the form of small tags that correspond to fragments of those transcripts.
- SAGE experiments include the following steps:
  - ✓Isolate mRNA from tissue.
  - ✓Use a retrotranscriptase to copy mRNA to cDNA.
  - ✓Extract a small amount of sequence from a defined position of each cDNA molecule.
  - ✓Link these small pieces of sequence together to form a long chain (or concatemer).
  - ✓Example of a concatemer:
    CATGACCCACGAGCAGGGTACGATGATACATGGAAACCTATGCACCTTGGGTA GCACATG
  - ✓Clone these chains into a vector which can be taken up by bacteria.
  - ✓Sequence the chains.
  - ✓Process this data with a computer to count the small sequence tags.

Velculescu et al. *Science*, 1995

Features of SAGE

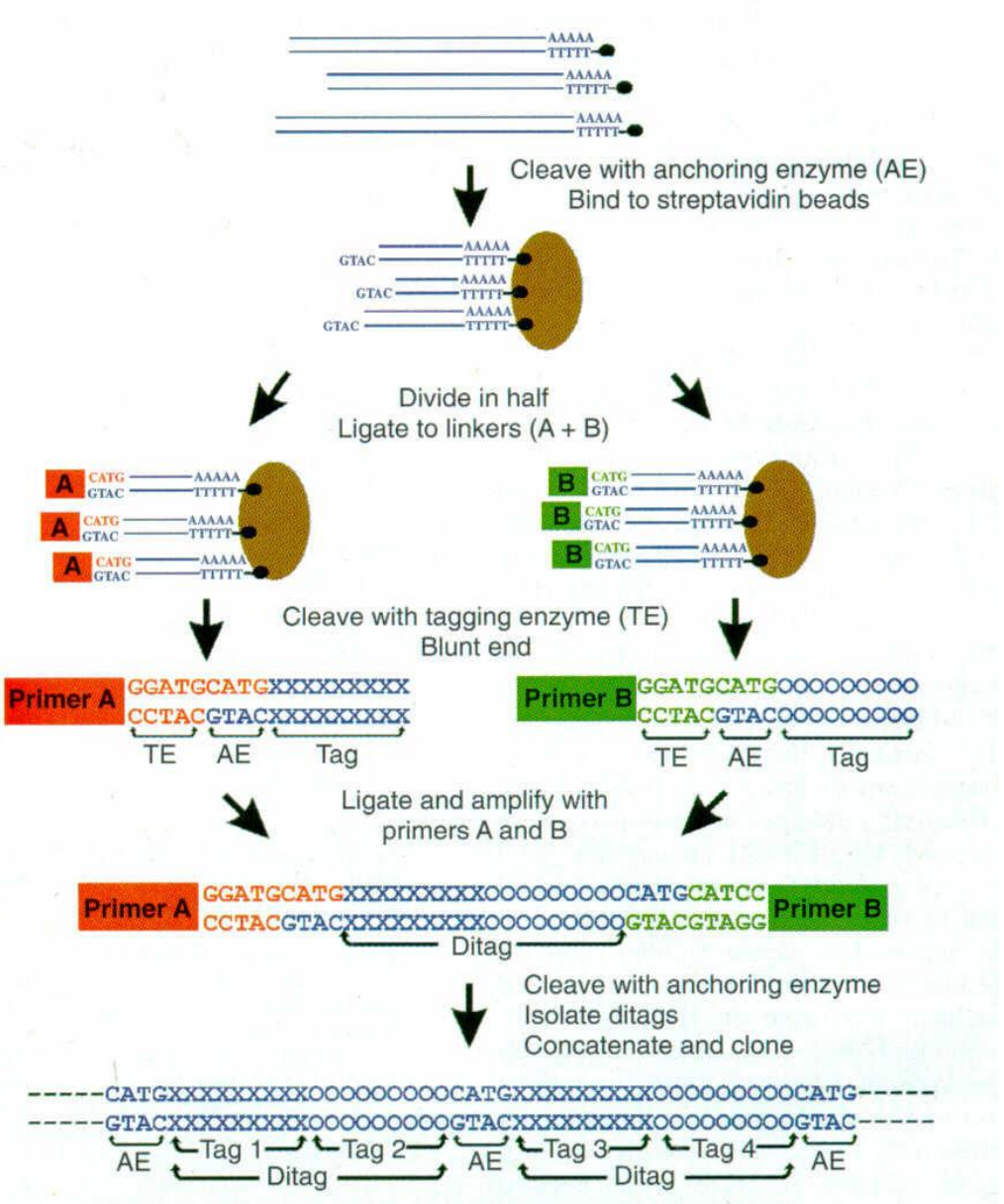
- **Provides researchers with a means to study expression of an entire genome rather than single genes.**
- **Facilitates comparison of gene expression differences between two samples with information about abundance of transcripts.**
- **Affords quantitative gene expression through library screening.**

Two Principles of SAGE

1. **A 9-bp tag is sufficient to unambiguously identify a gene.**
2. **Concatenation (linking together) of these short DNA sequences increases the efficiency of identifying unique transcripts in a serial manner.**

SAGE Protocol

1. mRNA cDNA
2. Cleave with anchoring enzyme (AE), such as *Nla*III.
3. Isolate 3′-most transcript of each cDNA by binding to streptavidin beads
4. Divide cDNA in half.
5. Ligate to one of two linkers, each with a tagging enzyme (TE) site.
6. Ligate the two pools of tags together.
7. Ligated linkers serve as primers for amplification.
8. Cleave PCR products with AE to isolate ditags.
9. Concatenate by ligation.
10. Clone.
11. Sequence.



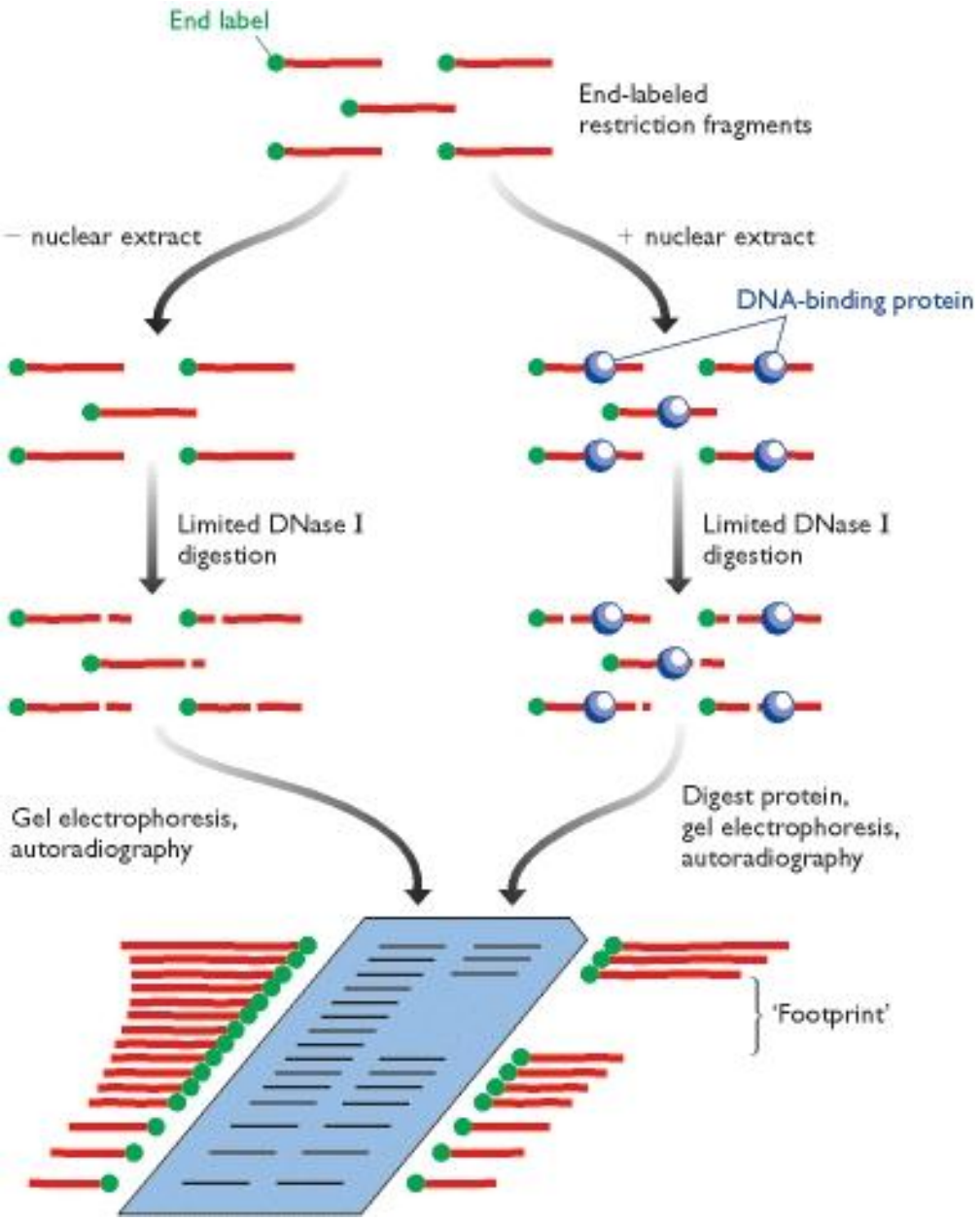
Measurement of Transcriptional Activity

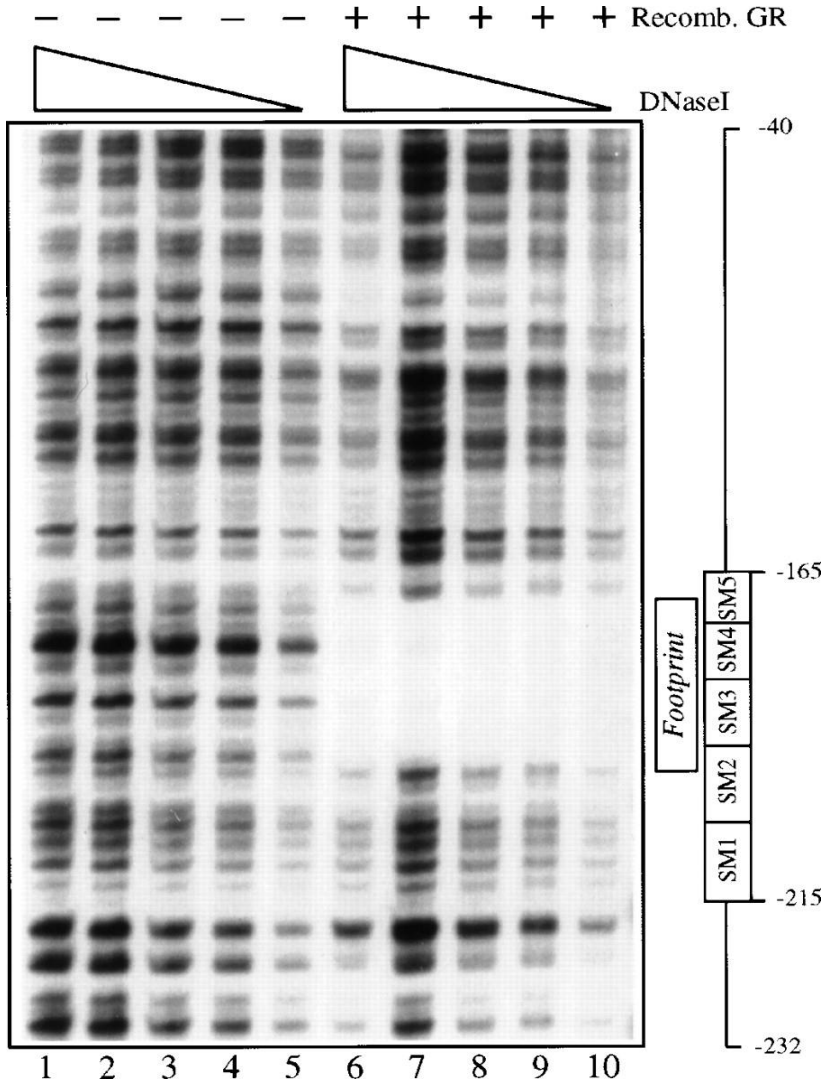
- **Dnase I Footprint Assay**
- **Gel Retardation or Electrophoretic Mobility Shift Assay (EMSA)**
- **Yeast One-hybrid Assay**
- **Reporter Assay**
- **Mutagenesis of DNA Sequences**
- **Chromatin immunoprecipitation (ChIP)**
  - ✓ **ChIP**
  - ✓ **ChIP-chip and ChIP-sequencing**
  - ✓ **Chromosome conformation capture (3C)**

DNase I Footprinting

- DNase I footprinting is a method of studying DNA-protein interactions and identifying the DNA sequence to which a protein binds.
- A target DNA fragment about 100-300 bp in length is either PCR-generated or cut from a vector and then uniquely labeled (at only one end) and incubated with protein (usually nuclear extract), followed by controlled digestion with DNase I, which cuts the probe randomly but only once.
- The digested DNA is recovered from the reaction and resolved on a polyacrylamide gel along with G+A chemical sequencing reaction (Maxam–Gilbert sequencing based on nucleobase-specific partial chemical modification of DNA and subsequent cleavage of the DNA backbone at sites adjacent to the modified nucleotides), which uses the same probe as the template.
- The regions bound by proteins will be protected from DNase I digestion and will be shown as a blank area on the gel track, whereas the exact protein-bound sequence can be read out by comparing the location of the blank with the sequencing reaction.

DNase I Footprinting



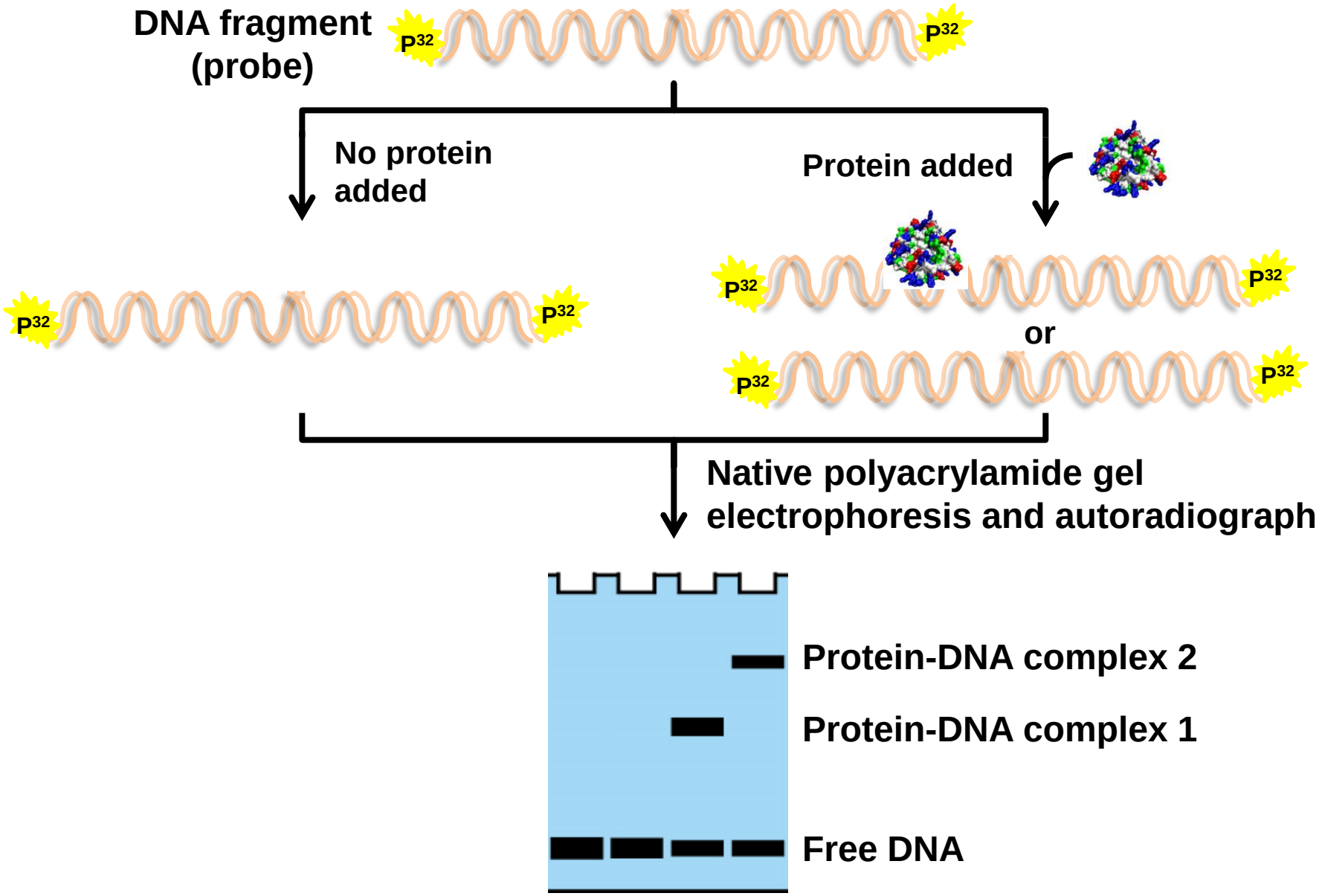


Li et al. ***American Journal of Physiology - Gastrointestinal and Liver Physiology*** **276:G1425-G1434.** ***1999.***

Gel Retardation Assay

- **The gel retardation or electrophoretic mobility shift (EMSA, or gel shift) assay is a sensitive technique for studying protein–DNA interactions.**
- **It is widely used to investigate the formation of protein-DNA complexes in a crude nuclear protein extract.**
- **The method relies on the stability of protein DNA complexes when subjected to non-denaturing polyacrylamide gel electrophoresis.**
- **The DNA is radiolabeled to enable rapid detection and, in its native state, it migrates quickly through the gel matrix.**
- **Protein binding generates slower mobility protein-DNA complexes that resolve as discrete bands.**

Gel Retardation Assay



Yeast One-hybrid Assay

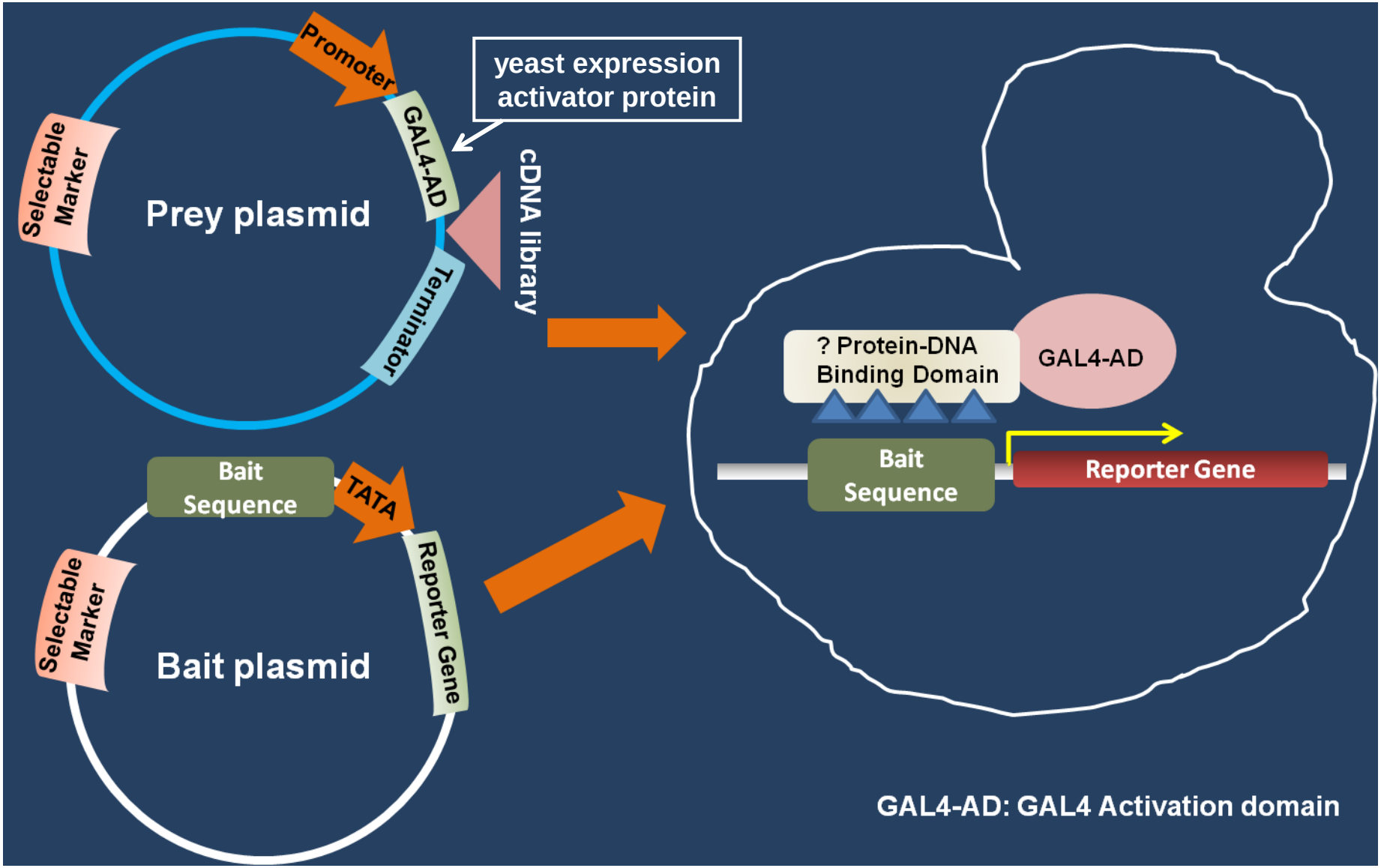
- **One-hybrid screening in yeast is a powerful method to rapidly identify heterologous transcription factors that can interact with a specific regulatory DNA sequence of interest (bait sequence).**

- **In this technique, the interaction between proteins (prey) and the target DNA sequence (bait) is detected by *in vivo* reconstitution of a transcriptional activator that "turns on" expression of a reporter gene.**

- **To ensure that DNA binding results in reporter-gene activation, cDNA expression libraries are used to produce hybrids between the prey and a strong trans-activating domain.**

- **The advantage of cloning transcription factors or other DNA-binding proteins by one-hybrid screenings, compared to biochemical techniques, is that the procedure does not require specific optimization of *in vitro* conditions.**

Yeast One-hybrid Assay



Reporter Gene Assay

- **Reporter gene assay – an assay used to study regulatory promoter and enhancer sequences as well as transcription factors.**
- **Regulatory sequences of interest are cloned into a reporter gene-containing vector and are introduced into cells.**
- **Rationale is that the level of reporter gene expression will correlate with the transcriptional activity of the DNA fragment of interest.**
- **Expression of reporter genes can be measured by:**

  1. **Enzyme activity assay of the expressed enzyme encoded by the reporter gene using chromo-, fluoro- or luminogenic substrates.**
  2. **Immunological assay of the expressed protein encoded by the reporter gene (reporter gene ELISA).**
  3. **Histochemical staining of cells or tissues typically to localize enzymatic activity ectopically expressed from reporter gene constructs in transformed cells.**

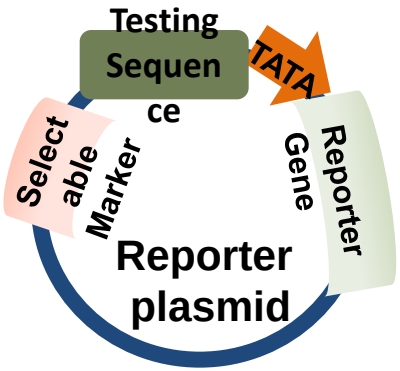
Reporter Gene Assay

1 kb
0.7 kb
0.5 kb
0.3 kb

Putative regulatory DNA sequence

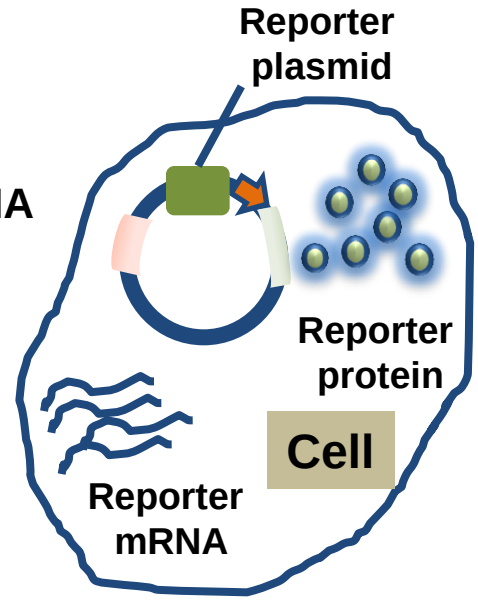
1

Cloning



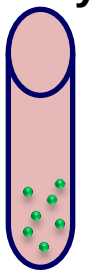
2

Introduce plasmid DNA into cells



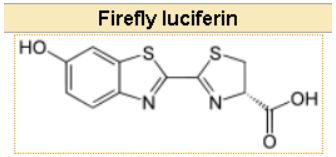
3

Prepare cellular lysate

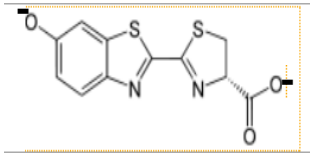


4

Assaying for reporter gene expression



+ ATP



+

Glo

+

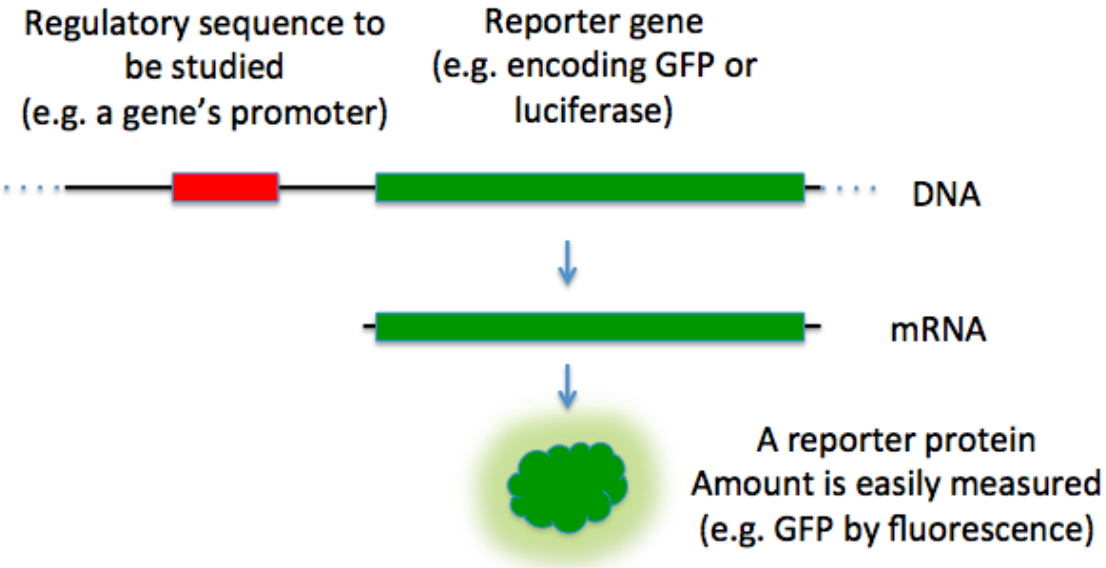
AMP + PPi + CO_2

Measure generated product

5

Plasmid No.	Reporter activity
1	++++
2	++++
3	+
4	+
5	-

Reporter Genes



- **Commonly used reporter genes that induce visually identifiable characteristics usually involve fluorescent and luminescent proteins.**
- **Examples include:**
  - ✓ **the gene that encodes jellyfish green fluorescent protein (GFP), which causes cells that express it to glow green under blue light**
  - ✓ **the enzyme luciferase, which catalyzes a reaction with luciferin to produce light**
  - ✓ **the red fluorescent protein from the gene dsRed.**

Western Blotting

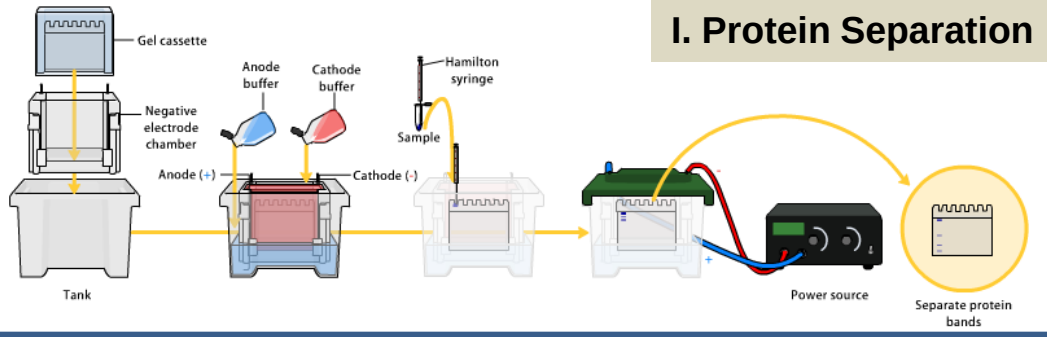
- **Proteins are separated according to size by SDS-polyacrylamide gel electrophoresis and transferred from the gel to a filter.**

- **The filter is incubated with an antibody directed against a protein of interest.**

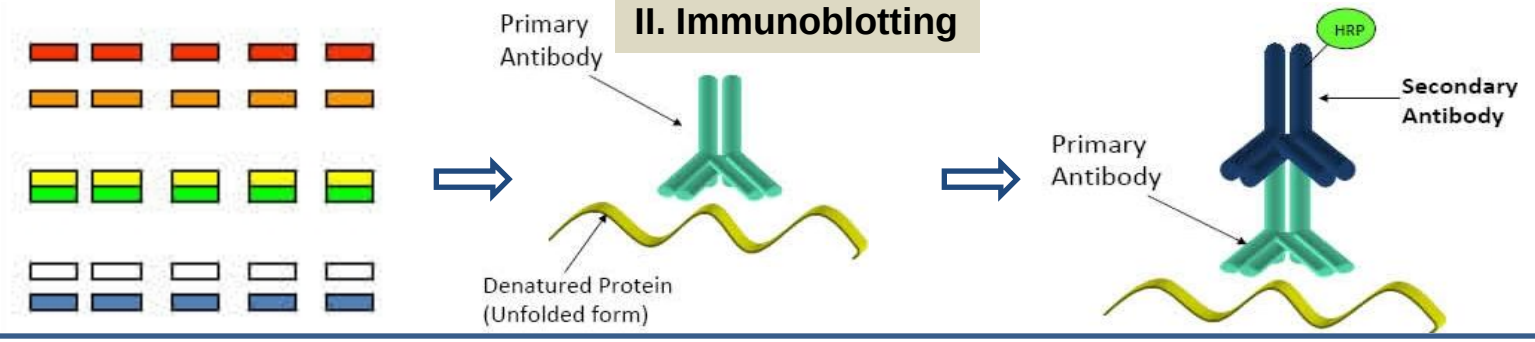
- **The antibody bound to the filter can then be detected by reaction with various reagents, such as a radioactive probe that binds to the antibody.**

Western Blotting

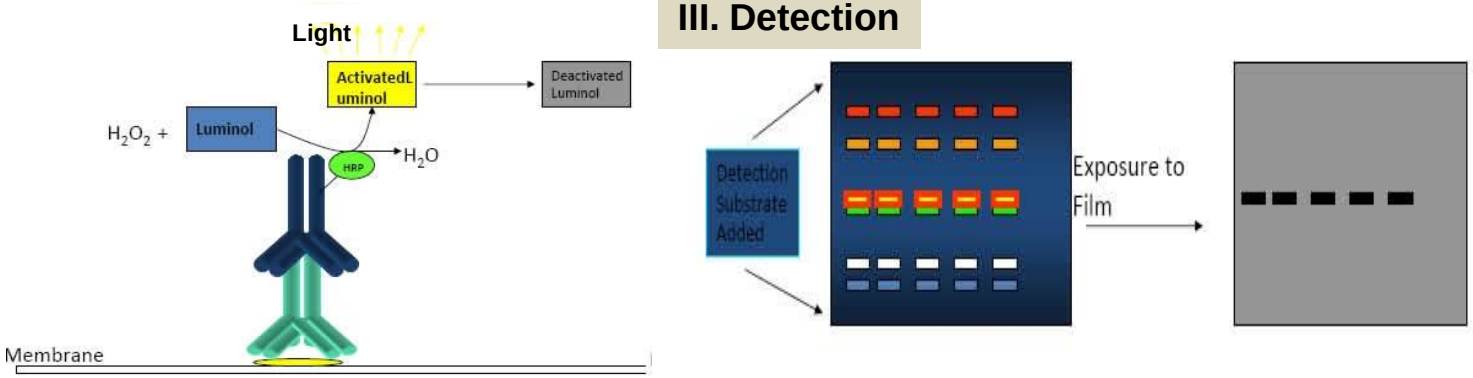
I. Protein Separation



II. Immunoblotting



III. Detection



South Western Blotting

- South Western blotting is a method for rapid characterization of both DNA-binding proteins and their specific sites on genomic DNA.
- Proteins are separated on a SDS-polyacrylamide gel, renatured by removing SDS in the presence of urea, and blotted onto nitrocellulose by diffusion.
- The genomic DNA region of interest is digested by restriction enzymes selected to produce fragments of appropriate but different sizes, which are subsequently end-labeled and allowed to bind to the separated proteins.
- The specifically-bound DNA is eluted from each individual protein-DNA complex and analyzed by acrylamide gel electrophoresis.
- Sequence-specific binding allows the purification of the corresponding selectively-bound DNA fragments and improves protein-mediated cloning of DNA regulatory sequences.

Anal. Biochem. 179: 299-303 (1989)