Chapter IV:

Techniques for analyzing the genome and its modifications, and gene amplification

Techniques for analyzing the genome and its modifications

- Genomic DNA and complementary DNA (cDNA) libraries
- Polymerase chain reaction (PCR)
- DNA microarrays

DNA libraries

DNA libraries (gene libraries): A DNA library is a POPULATION of vectors containing diverse DNA fragments representative of a cell type, a genome, a particular stage of development.

DNA libraries : 2 types

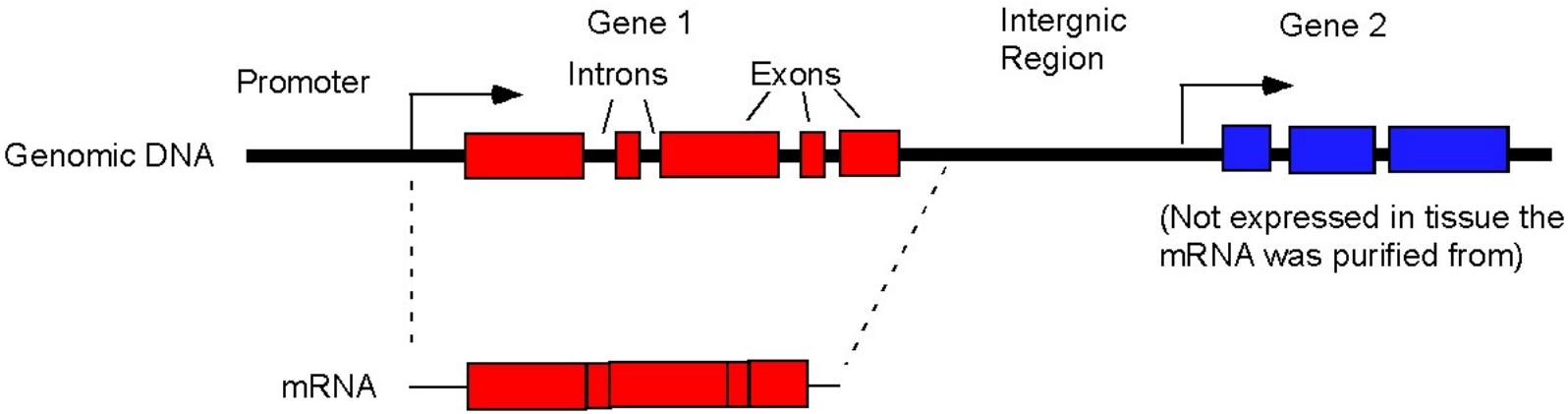
1) Genomic DNA libraries

(made from genomic DNA)

2) cDNA libraries

(made from cDNA copied from mRNA)

Differences between genomic DNA and cDNA libraries



<u>Genomic DNA libraries:</u>
Promoters
Let's get started
Intergenic Sequences
Unexpressed genes

<u>cDNA libraries:</u>
Genes Expressed
Transcription Sites (start)
Open reading frames (ORFs)
Splicing points

Differences between genomic DNA and cDNA libraries

Genomic DNA banks	cDNA banks
Made from total genomic DNA	Made from mRNA
It contains gene and regulatory regions	It contains only gene sequences
The gene cannot be expressed in the cloning host	The gene can be expressed in the cloning host
The gene can be expressed in the cloning host	The DNA insert is obtained using reverse transcriptase
All genes are present at the same frequency in a genomic DNA library	The frequency of a gene in a cDNA library depends on the abundance of the corresponding mRNA
The gene collection is independent of the cells or tissue used to isolate the DNA.	The collection of the cloned gene sequence depends on the cells or tissues used to isolate the mRNA.
Identifying the gene of interest is difficult due to the large number of recombinants.	Relatively simple to identify the gene based on the expressed protein
There will be a clone of each gene for an organism	There will be a clone of each gene for a particular cell or tissue
Genomic libraries are preferred for small genomes such as those of prokaryotic organisms.	cDNA libraries are preferred for larger genomes because only exons are cloned.
Useful for genome sequencing, promoter analysis and studies	Useful for coding region analysis

Genomic DNA libraries

✓DNA is extracted from cells

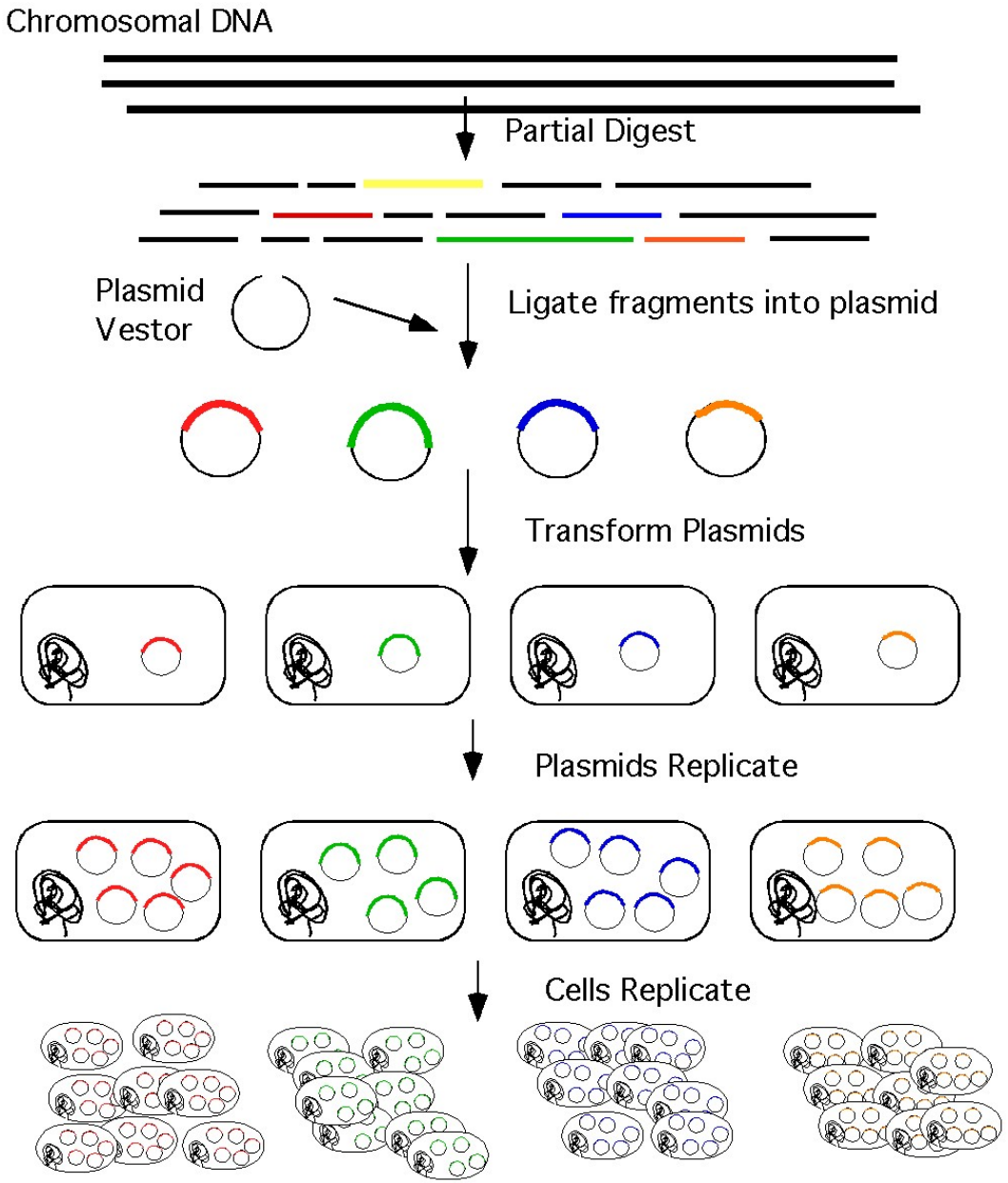
✓Digested with a restriction enzyme to cut the DNA into fragments of a specific size.

✓The fragments are inserted into the vector using DNA ligase,

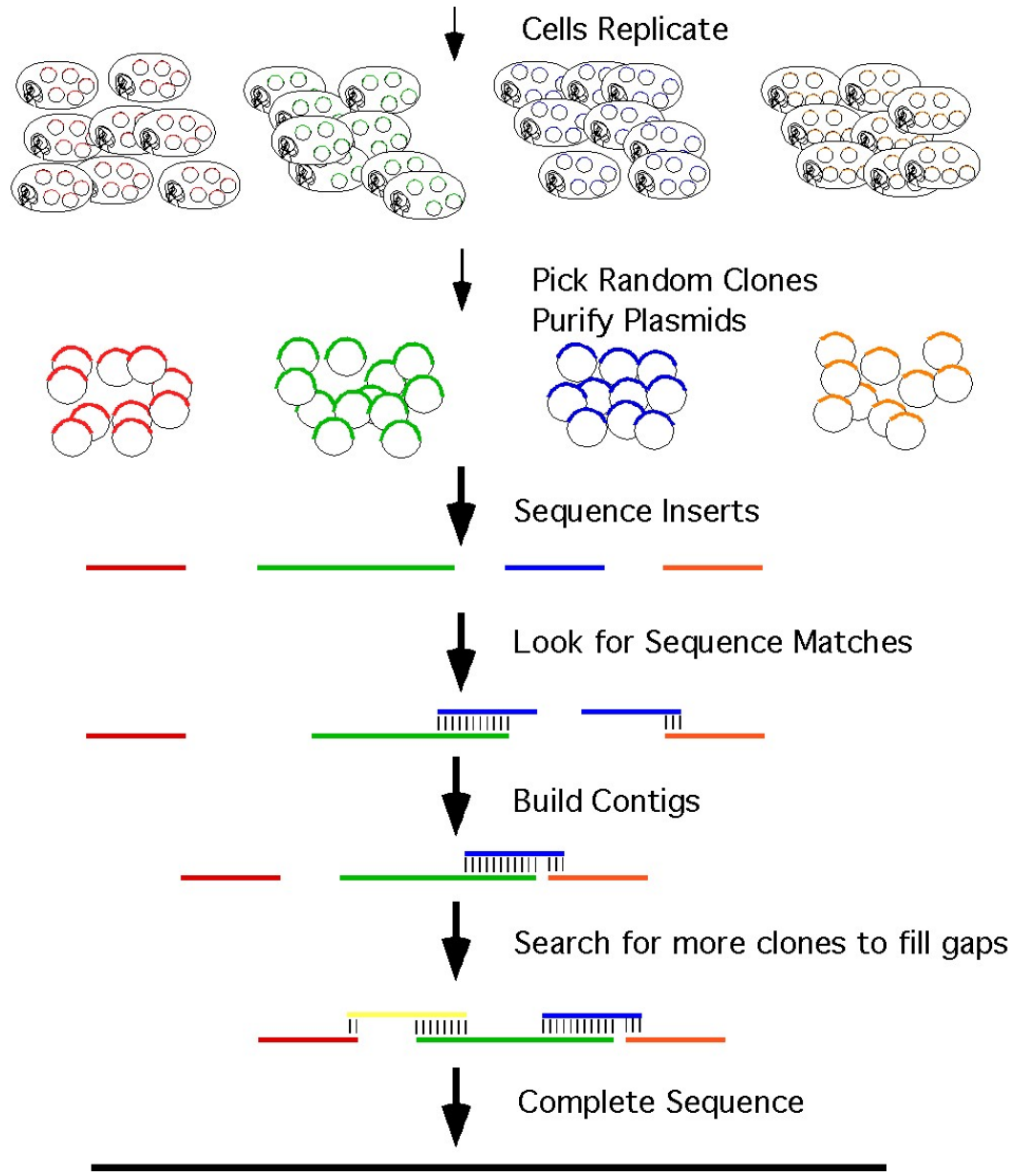
✓The recombinant vector can be transformed in a host organism - usually a population of *Escherichia coli* or yeast - each cell containing only one vector molecule.

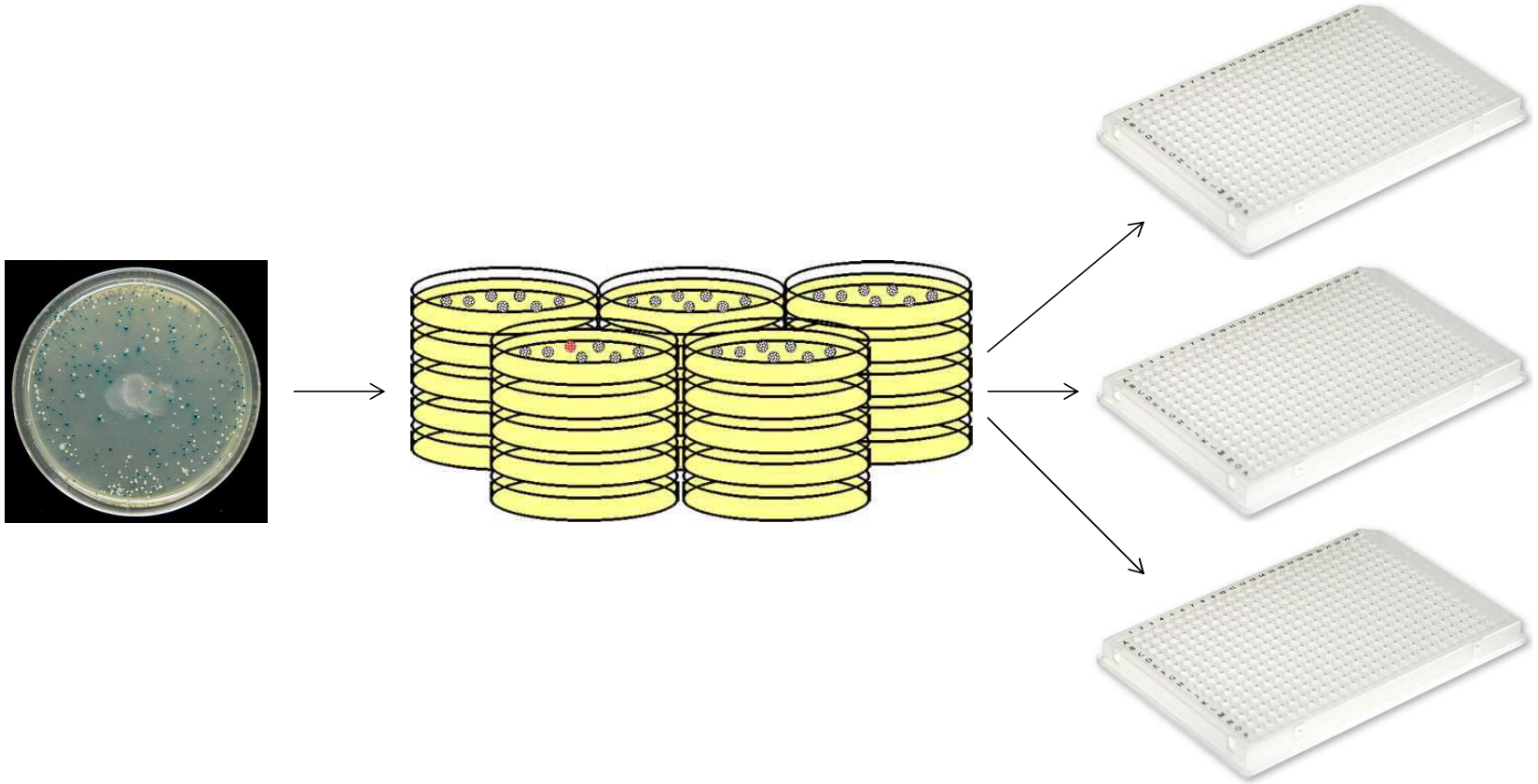
✓Selection and screening of clones for analysis

Random cutting into fragments of limited size, which are then cloned and sequenced; to obtain the broadest possible coverage of a genome

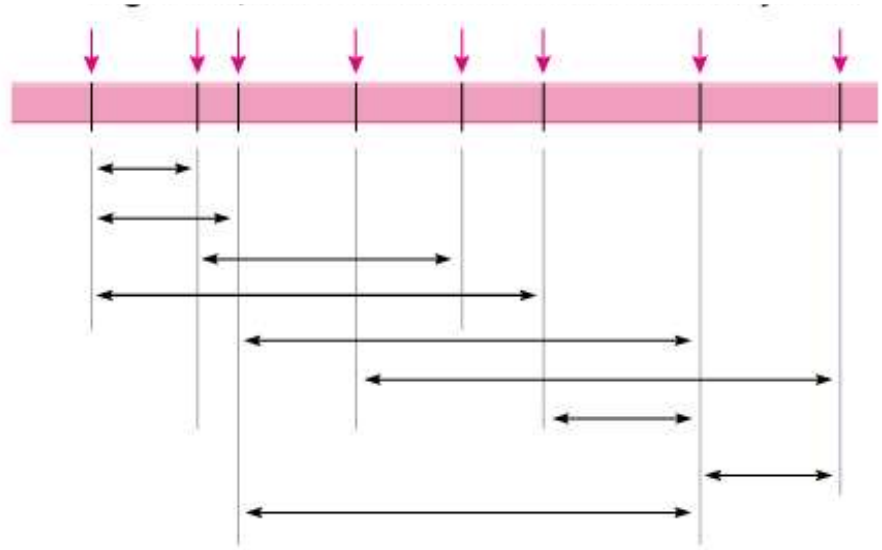


Assembling the obtained sequences requires efficient computer processing, the final sequence then being annotated to locate genes, repeated sequences, SNPs ("Single Nucleotide Polymorphism")....





Overlap of DNA fragments of different sizes in a library.



cDNA libraries

1. cDNA is derived from mature mRNA and does not include introns.

2. cDNA may contain less information than the coding region.

3. The cDNA library reflects the gene activity of a cell at the time the mRNAs are isolated (varies from tissue to tissue and over time).

4. mRNA degrades rapidly after cell death, and usually requires immediate isolation (cryoprotectants can increase yield if immediate freezing is postponed by fieldwork).

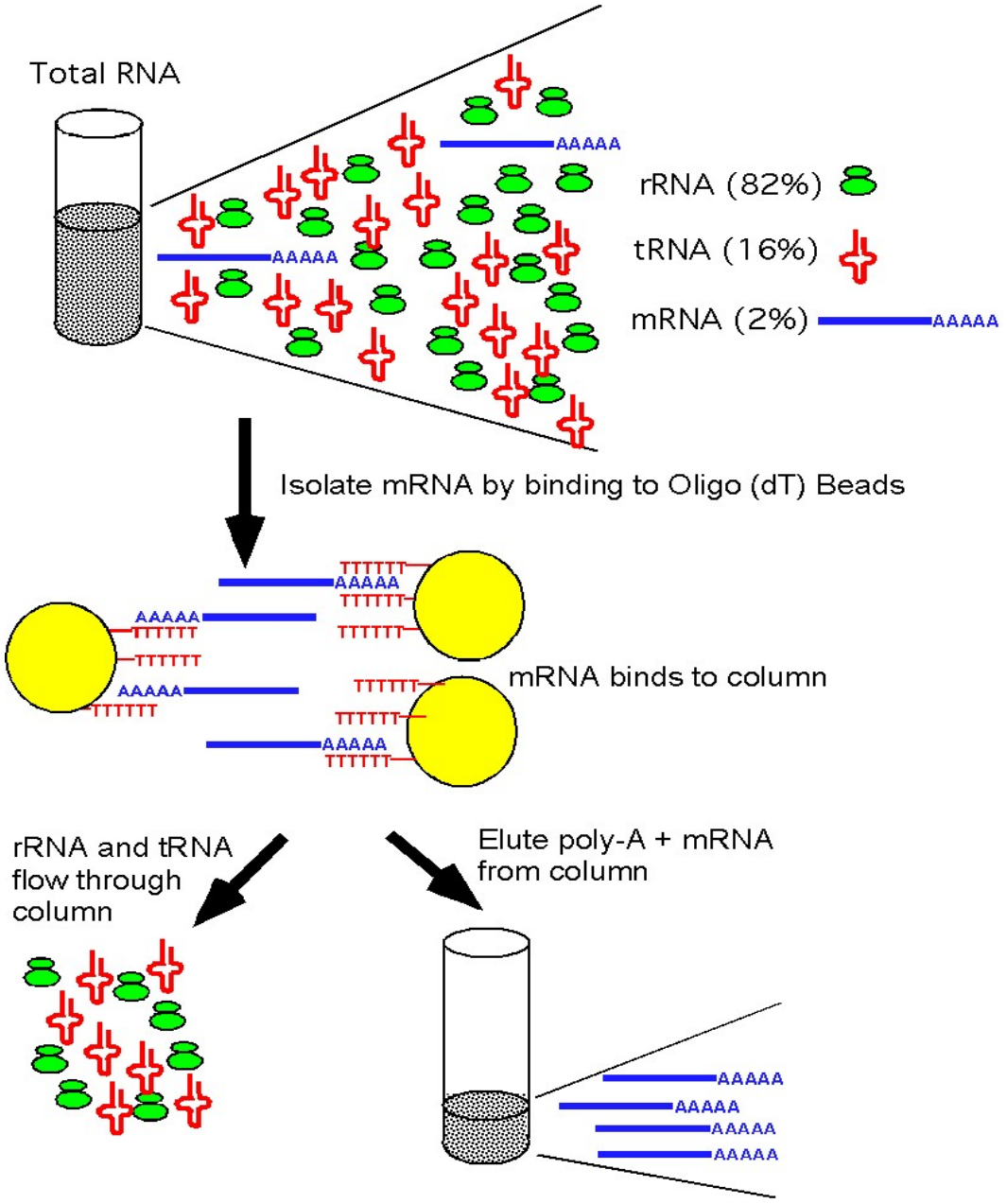
5. **Construction of a cDNA library:**

    1. Isolate mRNA

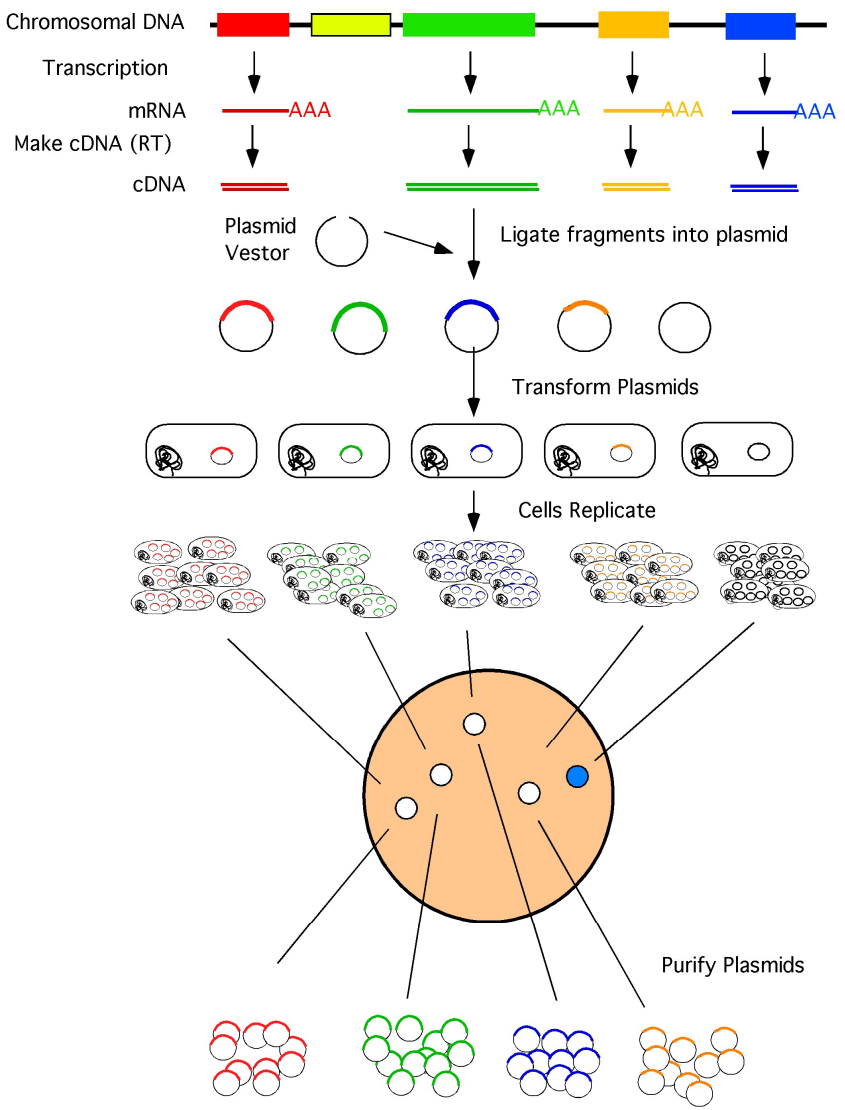
    1. Synthesize cDNA

    1. cDNA clone

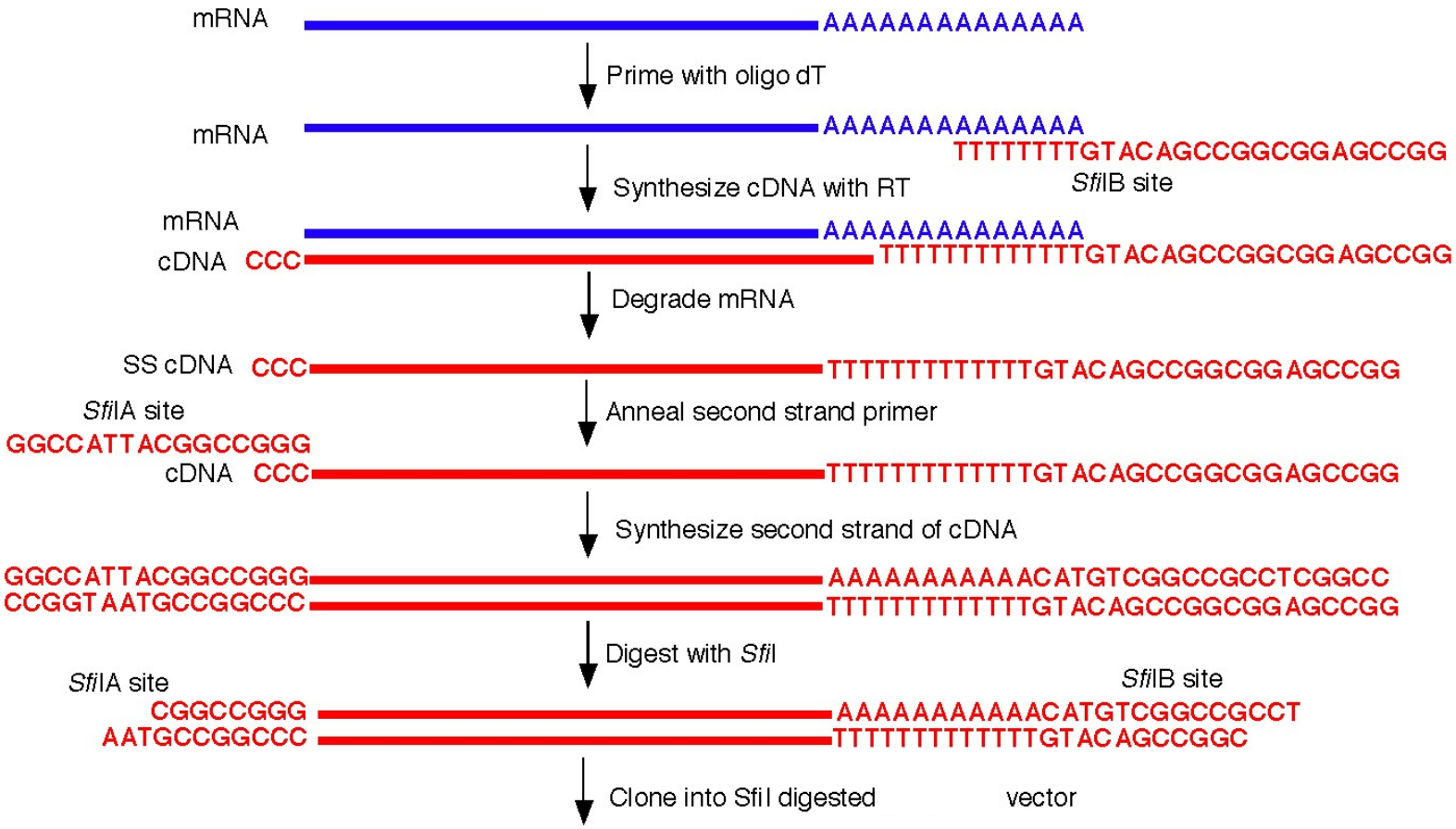
mRNA purification from total RNA



cDNA libraries construction



Synthesis of cDNA from mRNA
Use of adapters



Screening a cDNA library

- cDNA libraries are used to detect or sequence protein-coding genes because cDNAs are generated from genes that are transcribed.

- If you know the DNA sequence of the gene coding for the protein you want to find, a homologous DNA probe can be used

- Or the cDNA library can be sequenced directly using the universal M13 primers in the plasmid vector .

Screening a cDNA library

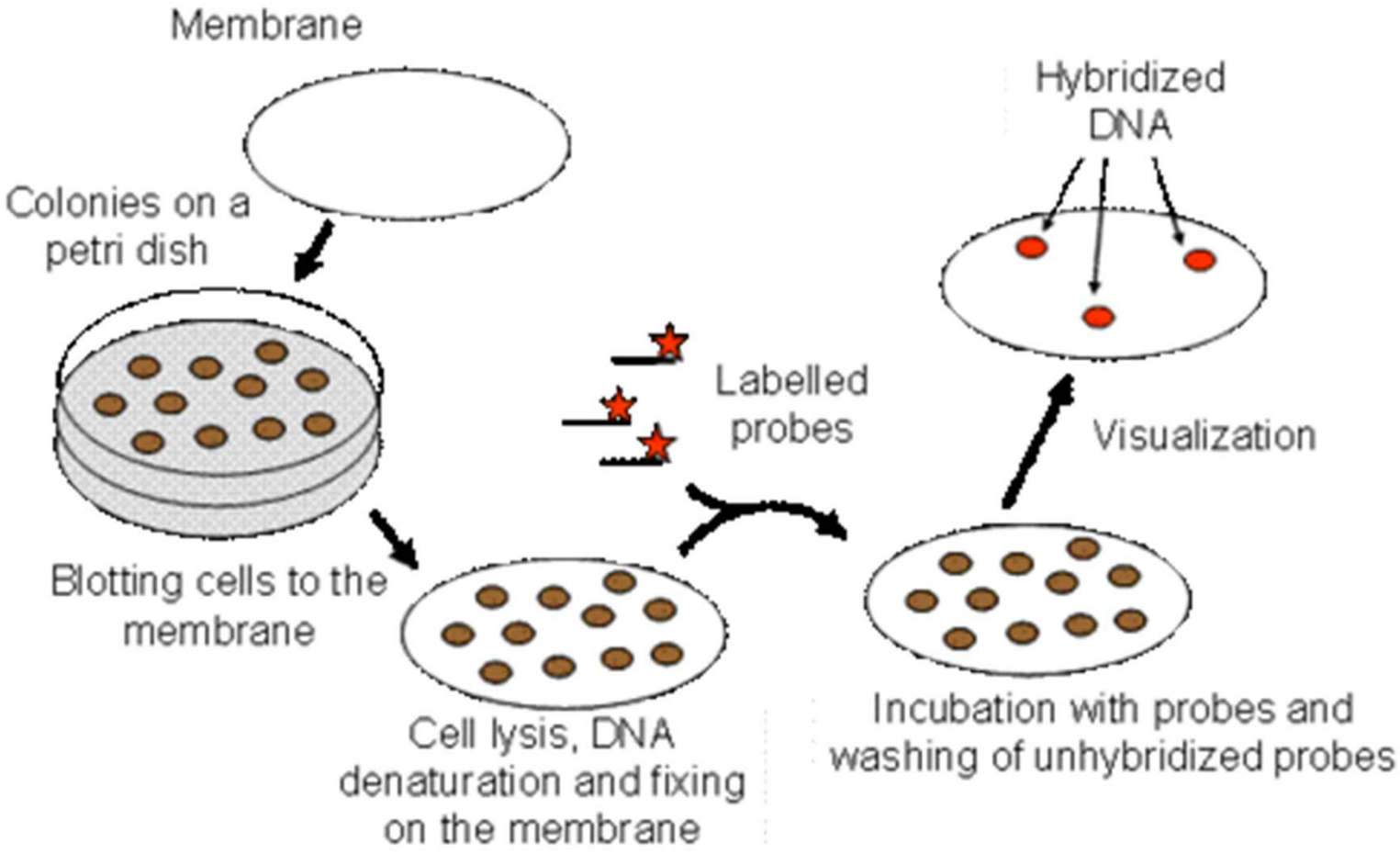
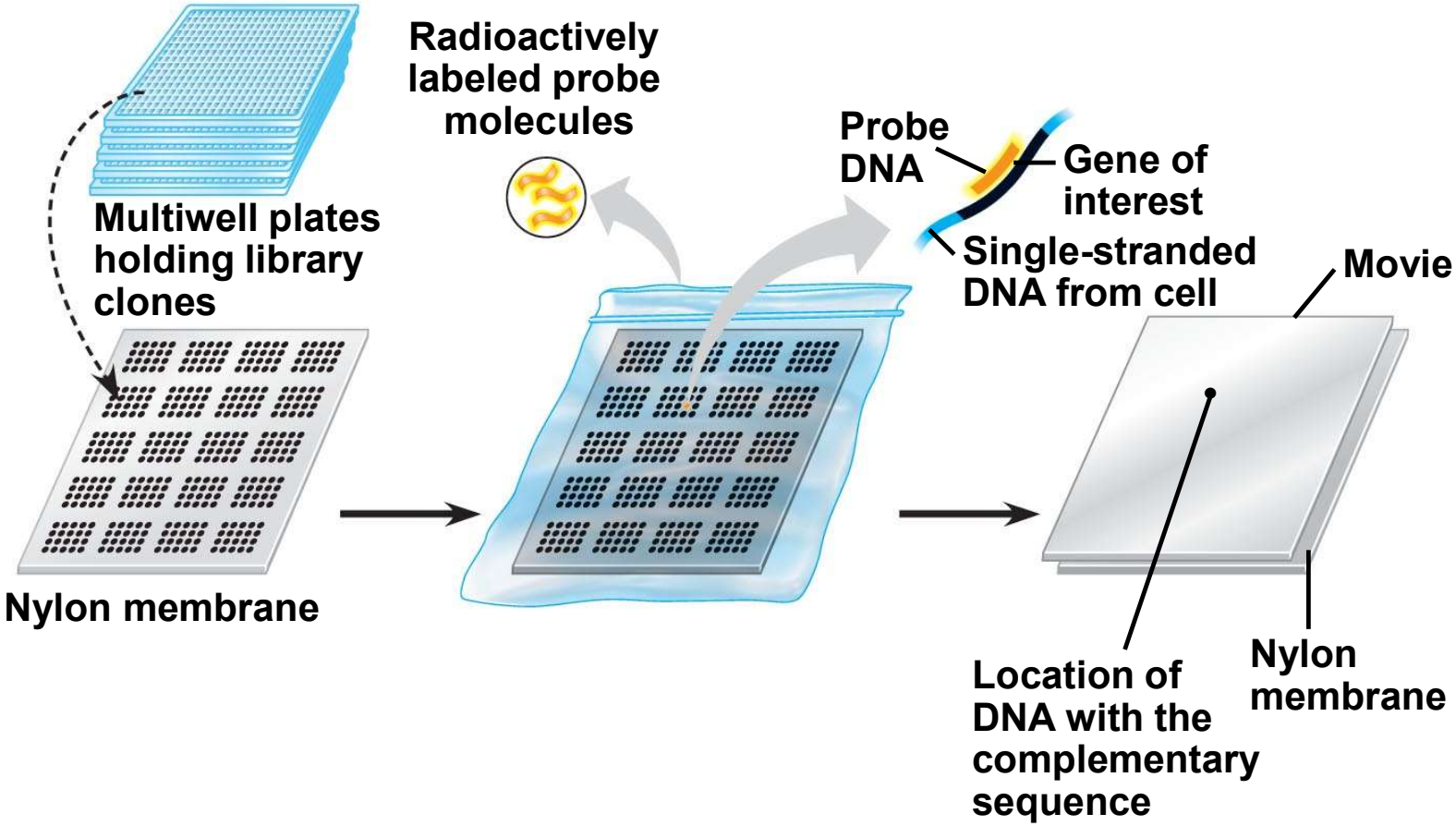
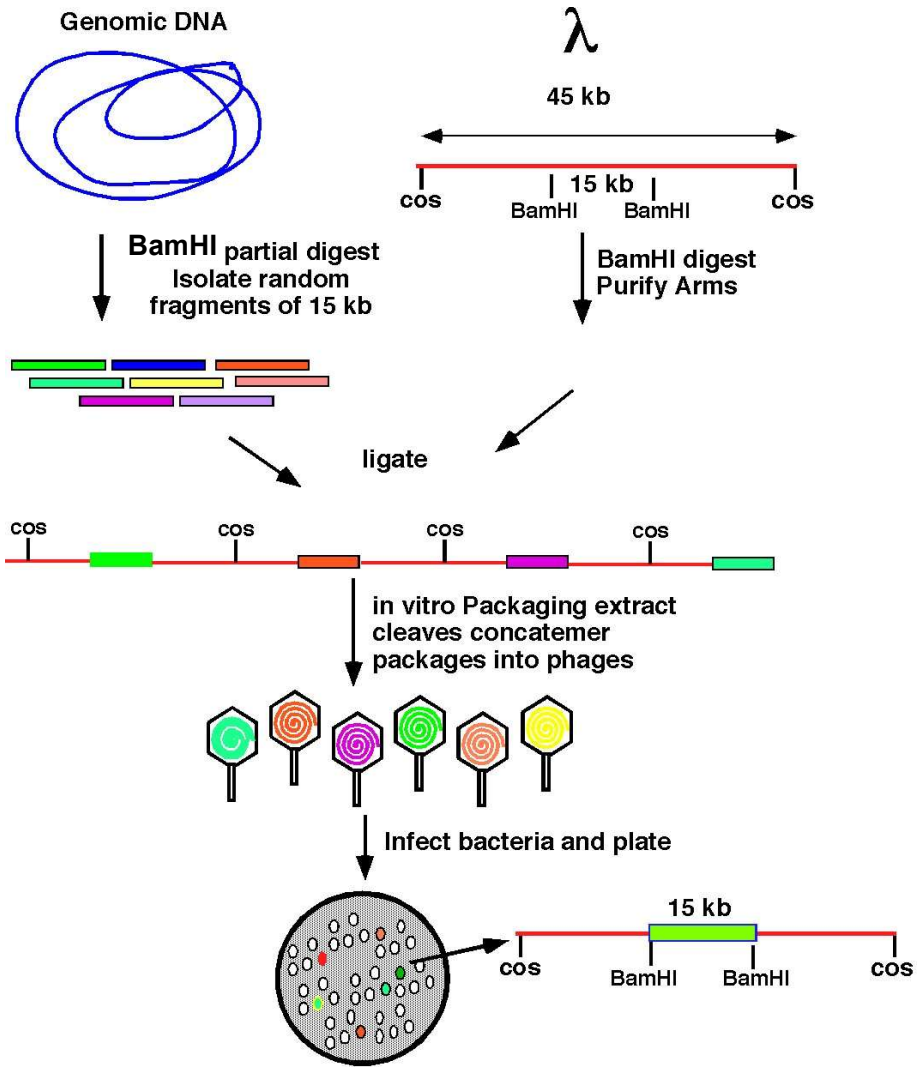


Fig. 20-7

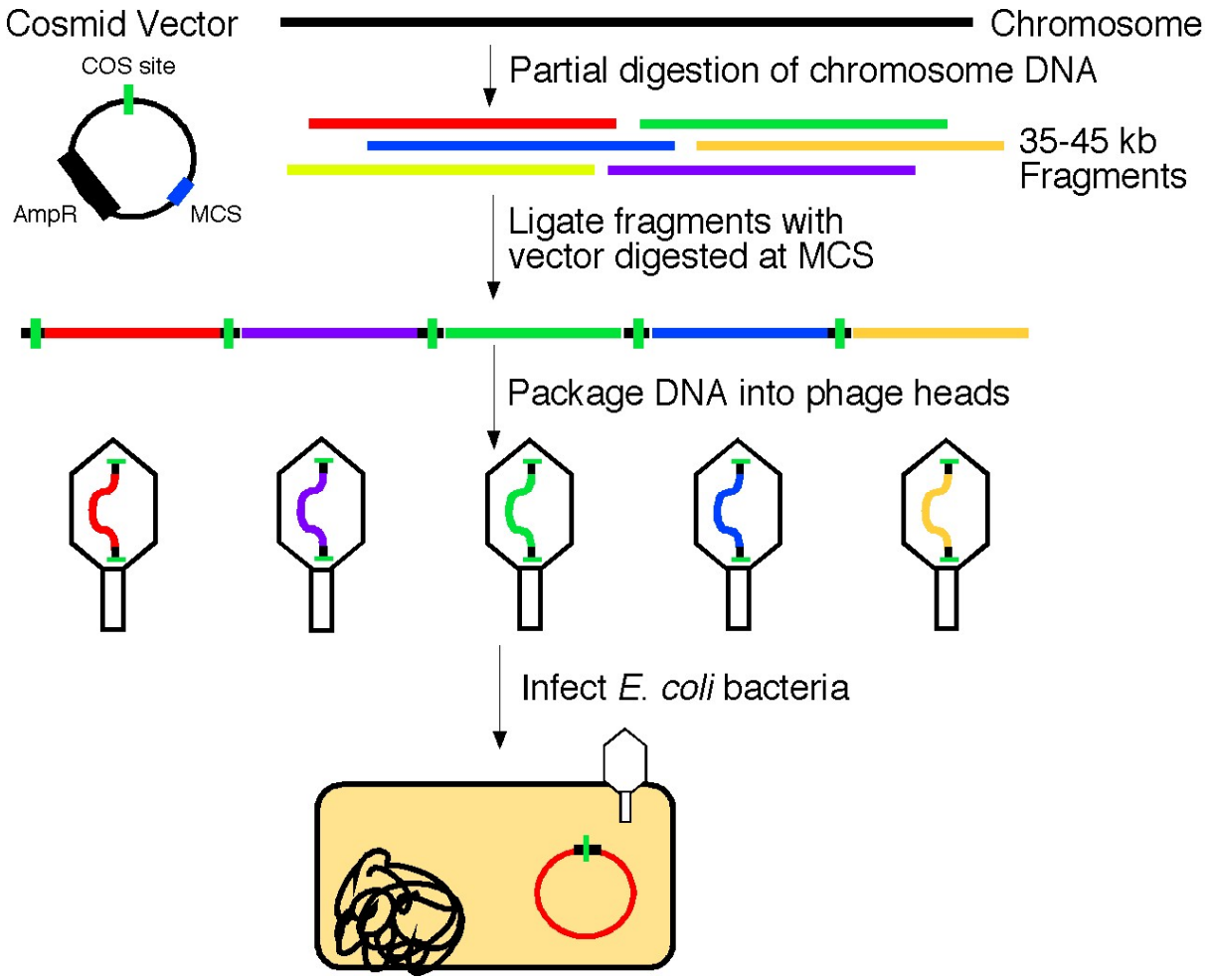
TECHNICAL



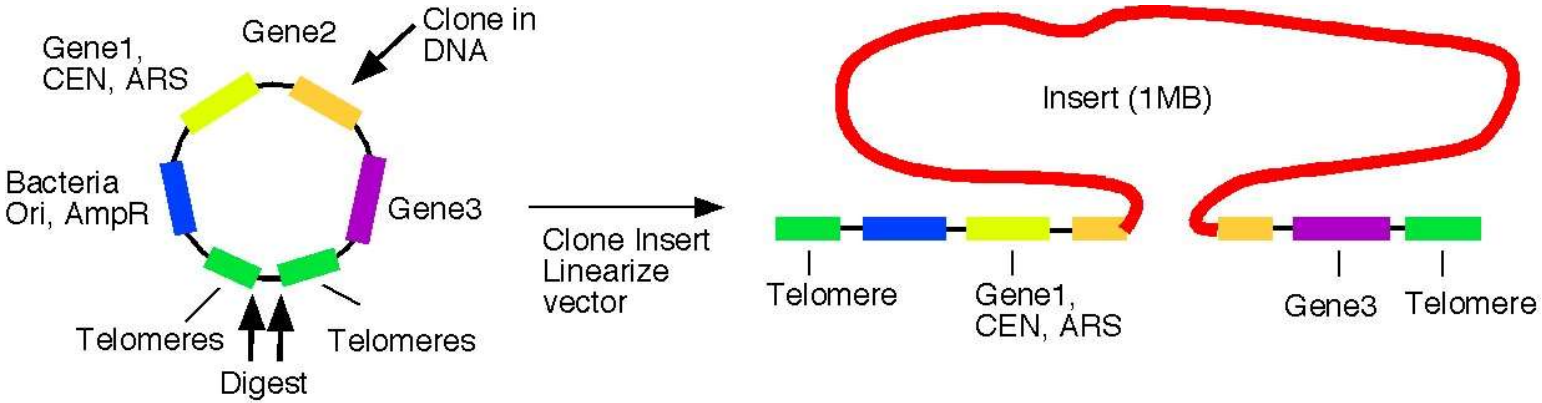
Construction of DNA libraries in phage □



Construction of DNA libraries in a cosmid



Building DNA libraries in a YAC



New strategies for preparing DNA libraries

- Building DNA libraries without cloning:
  - DNA or cDNA fragmentation
  - Ligation of fragment-specific adapters to enable PCR amplification.
  - High-throughput sequencing (see chapter 5)

Techniques for analyzing the genome and its modifications

- Genomic DNA and complementary DNA (cDNA) banks
- Polymerase chain reaction (PCR)
- DNA microarrays

Polymerise Chain Reaction (PCR)

The polymerase chain reaction (PCR) is a rapid technique for amplifying a DNA fragment in vitro.

- it can create millions of copies of a target DNA region starting from a very small initial amount.
- It was invented by **Kary Mullis in 1983**, earning him the **Nobel Prize in Chemistry (1993)**.

The purpose of PCR is to:

- Make copies of a specific region of DNA
- Identify the presence of particular genes
- Make the genetic material sufficient for future work

Applications of PCR

- **PCR is commonly used to…**
  - Identify the species
  - Identify alleles/genotypes to assess variability in a population
  - Create sequences for phylogenies to determine taxonomic relationships
  - Conduct forensic investigations

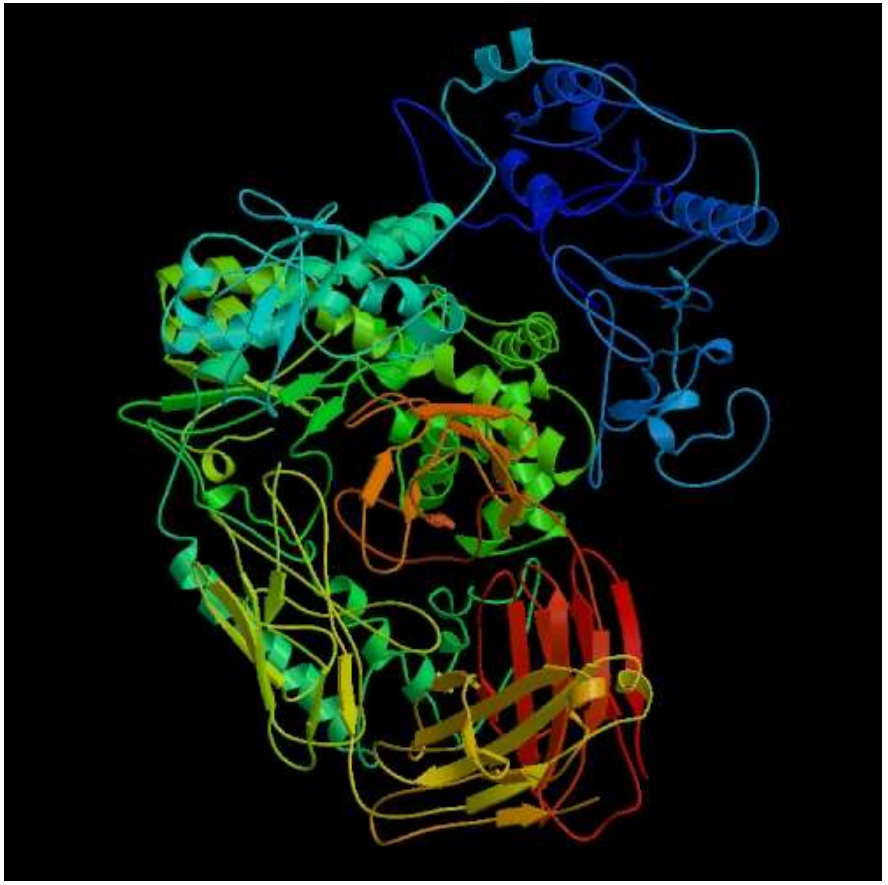
PCR reaction reagents

- A DNA template containing the target sequence
- A pair of primers
- DNA polymerase (e.g. Taq)
- dNTP
- Buffer containing Mg2+

Roles of PCR reagents

- **Primers**
  - **Hybridizes to single-stranded DNA template**
  - **Provide an initiation site for the extension of new DNA**
  - **primer meaning**
    - **Hybridizes to the antisense strand**
  - **antisense primer**
    - **Hybridizes to the sense strand**
- **DNA matrix**
  - **In this case, the product of our DNA extraction**

Taq polymerase



PCR reaction reagents

Component	Function
Template DNA	The DNA containing the target sequence to amplify.
Primers (Forward & Reverse)	Short single-stranded DNA sequences (~20 bases) that define the region to amplify.
DNA Polymerase (Taq polymerase)	Heat-stable enzyme (from *Thermus aquaticus*) that synthesizes new DNA strands.
dNTPs (deoxynucleotide triphosphates)	Building blocks (A, T, G, C) for DNA synthesis.
Buffer solution	Maintains pH and optimal ionic conditions.
$MgCl_2$	Cofactor for DNA polymerase activity.

PCR Procedures: repeated 3 steps

95° C: denatured DNA

Tm° C: primer hybridization

72° C: primer elongation by Taq DNA polymerase

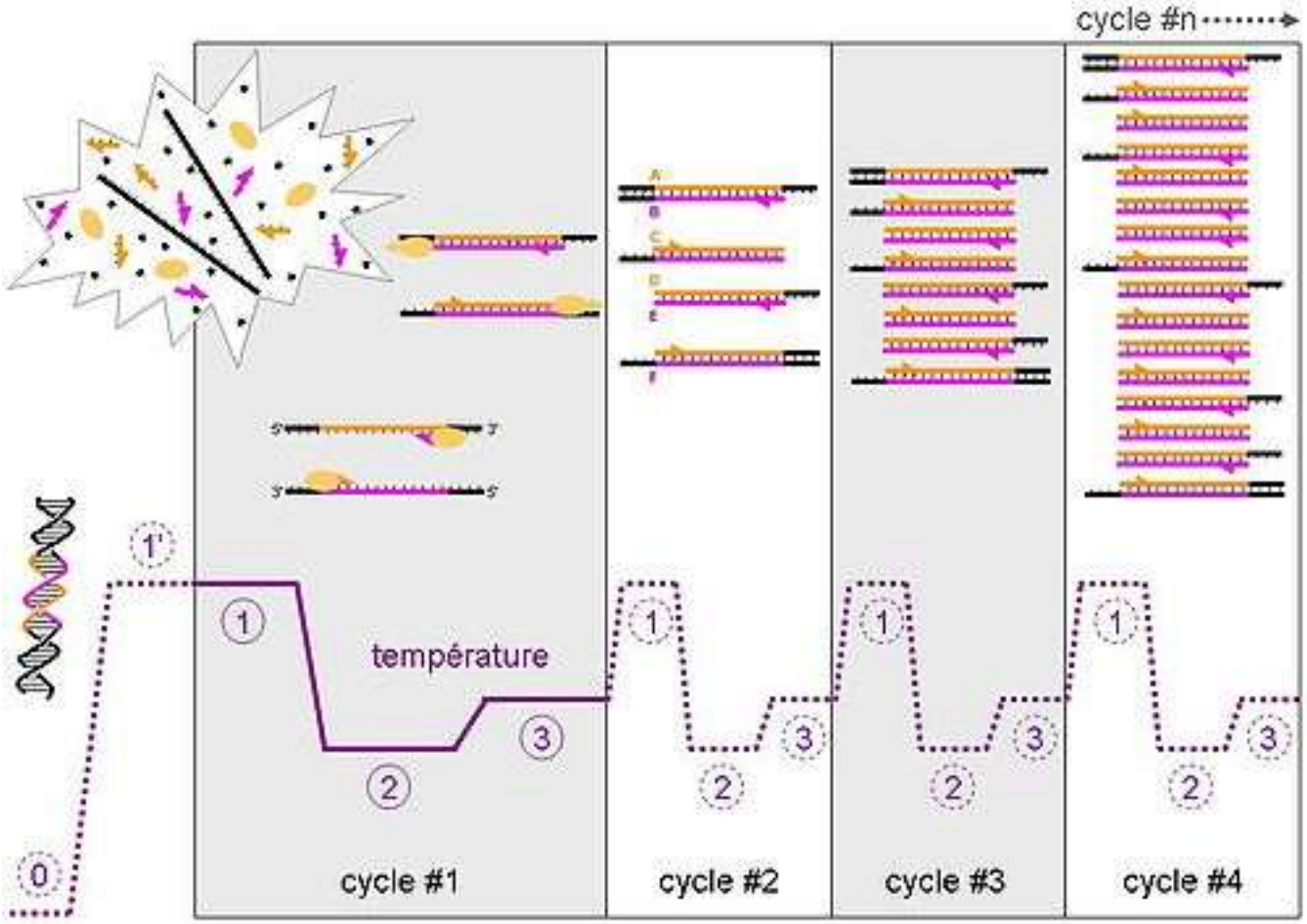
Repeat 30 cycles or more

PCR Procedures

- **Denaturation:** The DNA template is denatured into single-stranded DNA from double-stranded DNA by heating.
- **Annealing (hybridization):** This step allows the hybridization of the primers with the target DNA.
- **Elongation (Extension):** This process is the stage of DNA synthesis.

Schematic Overview

Step	Temperature	Duration	Process
Denaturation	94–98 °C	20–30 sec	DNA strands separate
Annealing	50–65 °C	20–40 sec	Primers bind
Extension	72 °C	30–60 sec	DNA synthesis by Taq
Final extension	72 °C	5–10 min	Completes any unfinished DNA strands
Hold	4 °C	∞	Preserves the product


cycle #n
température
0
1'
1
2
3
cycle #1
cycle #2
cycle #3
cycle #4

Amplifying DNA *in Vitro*: The Polymerase Chain Reaction (PCR)

- The **polymerase chain reaction, PCR**, can produce many copies of a specific target segment of DNA
- A three-step cycle—heating, cooling, and replication—brings about a chain reaction that produces an exponentially growing population of identical DNA molecules

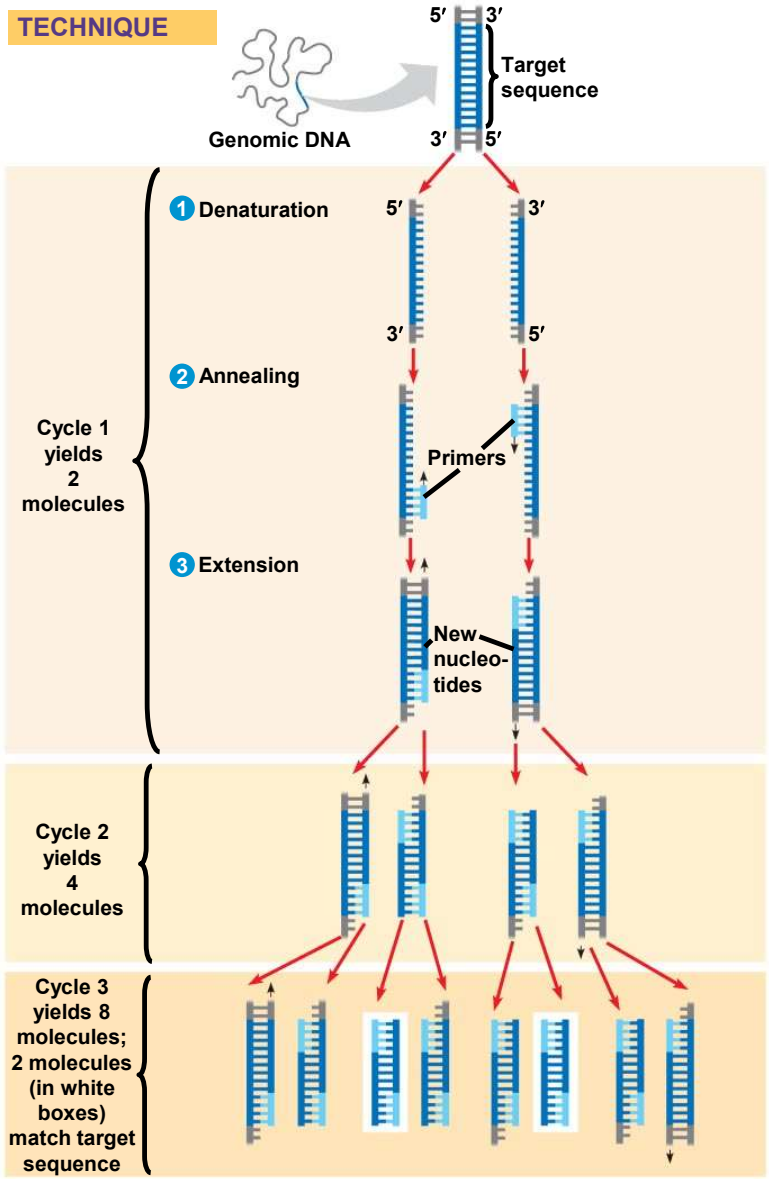
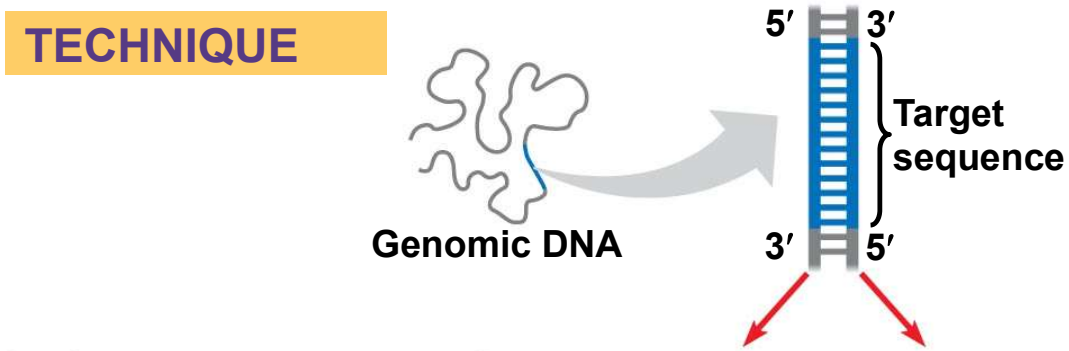


Fig. 20-8a



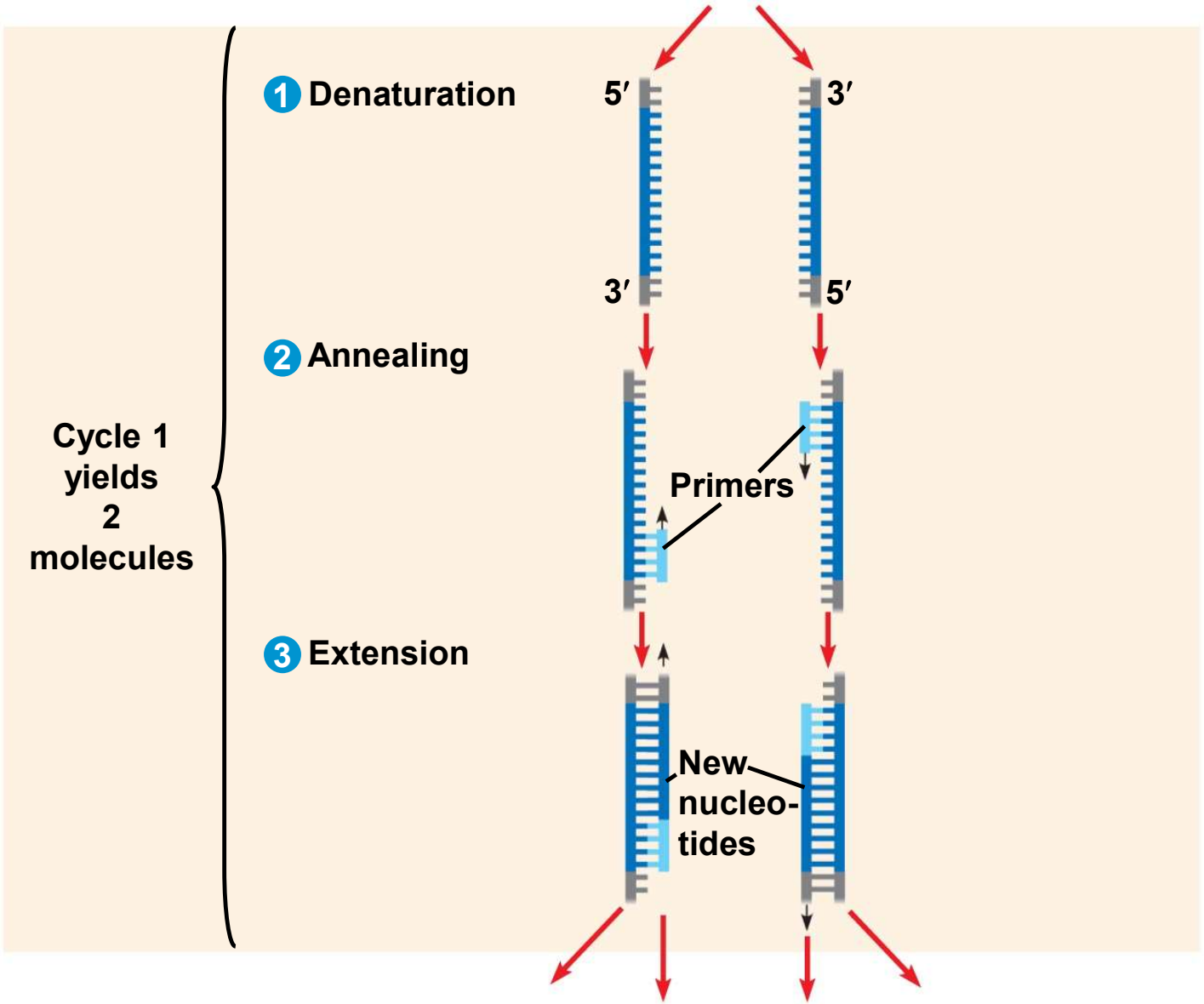
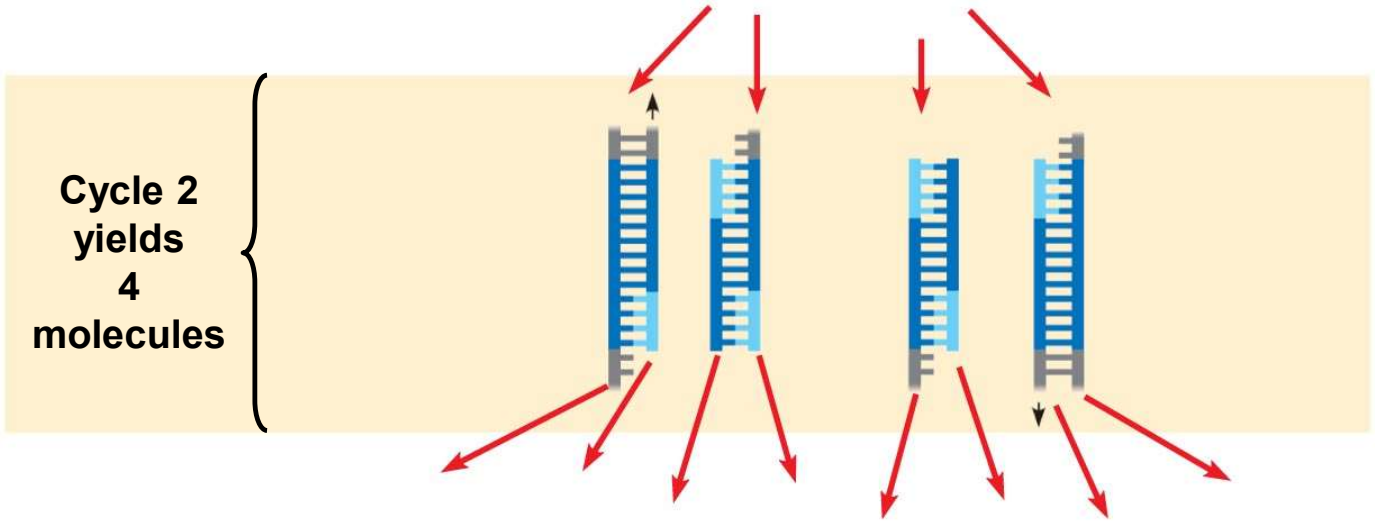
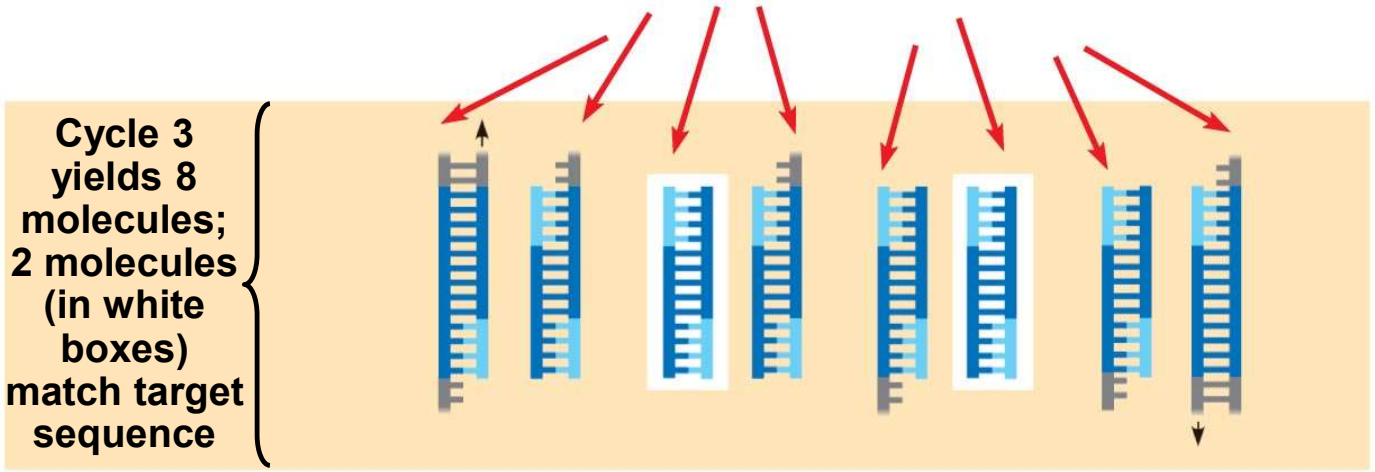


Fig. 20-8c





PCR Cycle Repetition

These 3 steps are repeated for **25–40 cycles** in a **thermocycler**.
Each cycle **doubles** the DNA quantity → exponential amplification.
Example:

- After 1 cycle → 2 copies
- After 2 cycles → 4 copies
- After 30 cycles → >1 billion copies

PCR Cycle Repetition

These 3 steps are repeated for **25–40 cycles**

The process is repeated along the reaction according to the following formula:

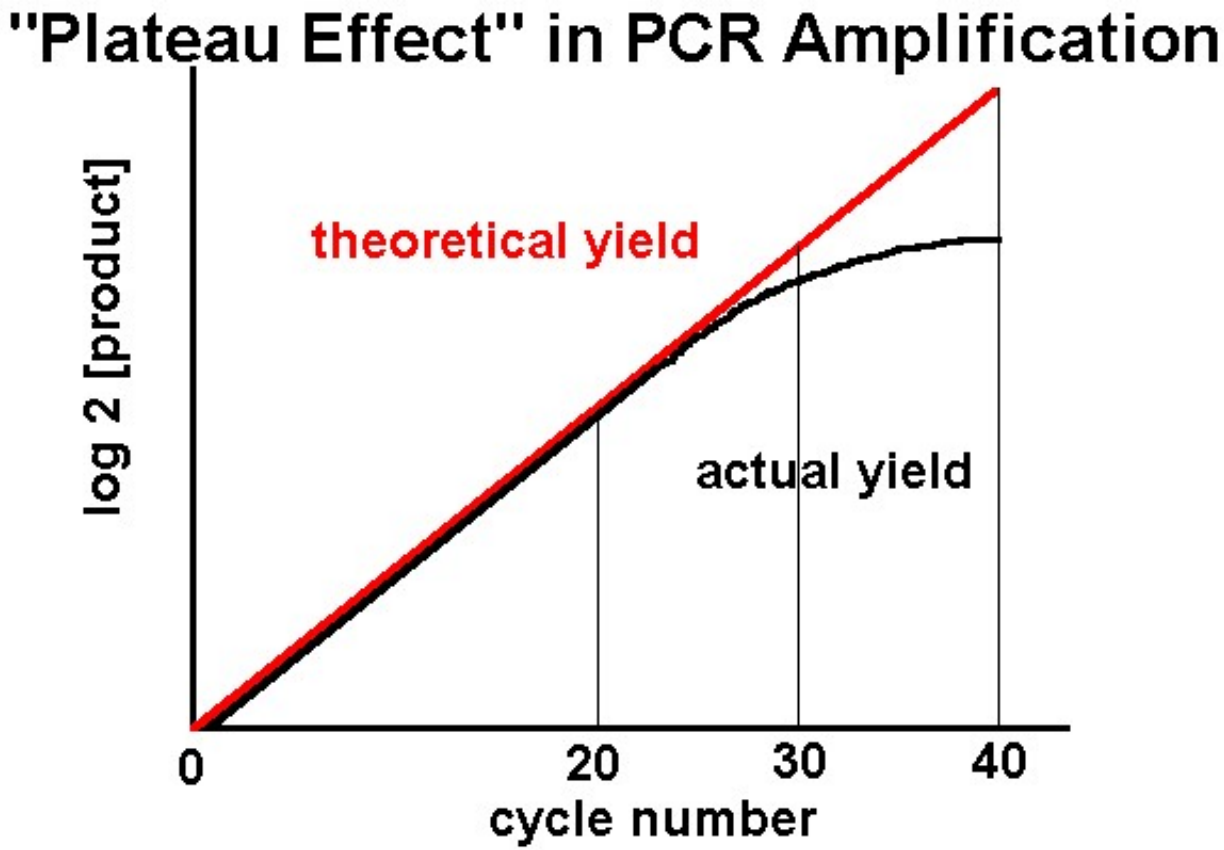
$$C = C_0(1 + E)^{\mathbf{n}}$$

where C is the final amount of DNA, C_0 is the initial amount of DNA, E is efficiency ($E = 10^{-1/S} - 1$; if $E = 1$ then s = –3.3219), n is the number of cycles and s is the slope of the exponential phase

Potentially, after 30 PCR cycles there will be around 2^{30} -fold amplification, assuming 100% efficiency during
each cycle.

The 'Plateau' Effect of PCR Amplification

Plateau Effect :
- Taq Denaturation
- PPi Inhibitor
- Competition/
- Rehybridization



PCR Product Revelation by Gel Electrophoresis

- One indirect method of rapidly analyzing and comparing genomes is **gel electrophoresis**
- This technique uses a gel as a molecular sieve to separate nucleic acids or proteins by size
- A current is applied that causes charged molecules to move through the gel
- Molecules are sorted into “bands” by their size

TECHNIQUE

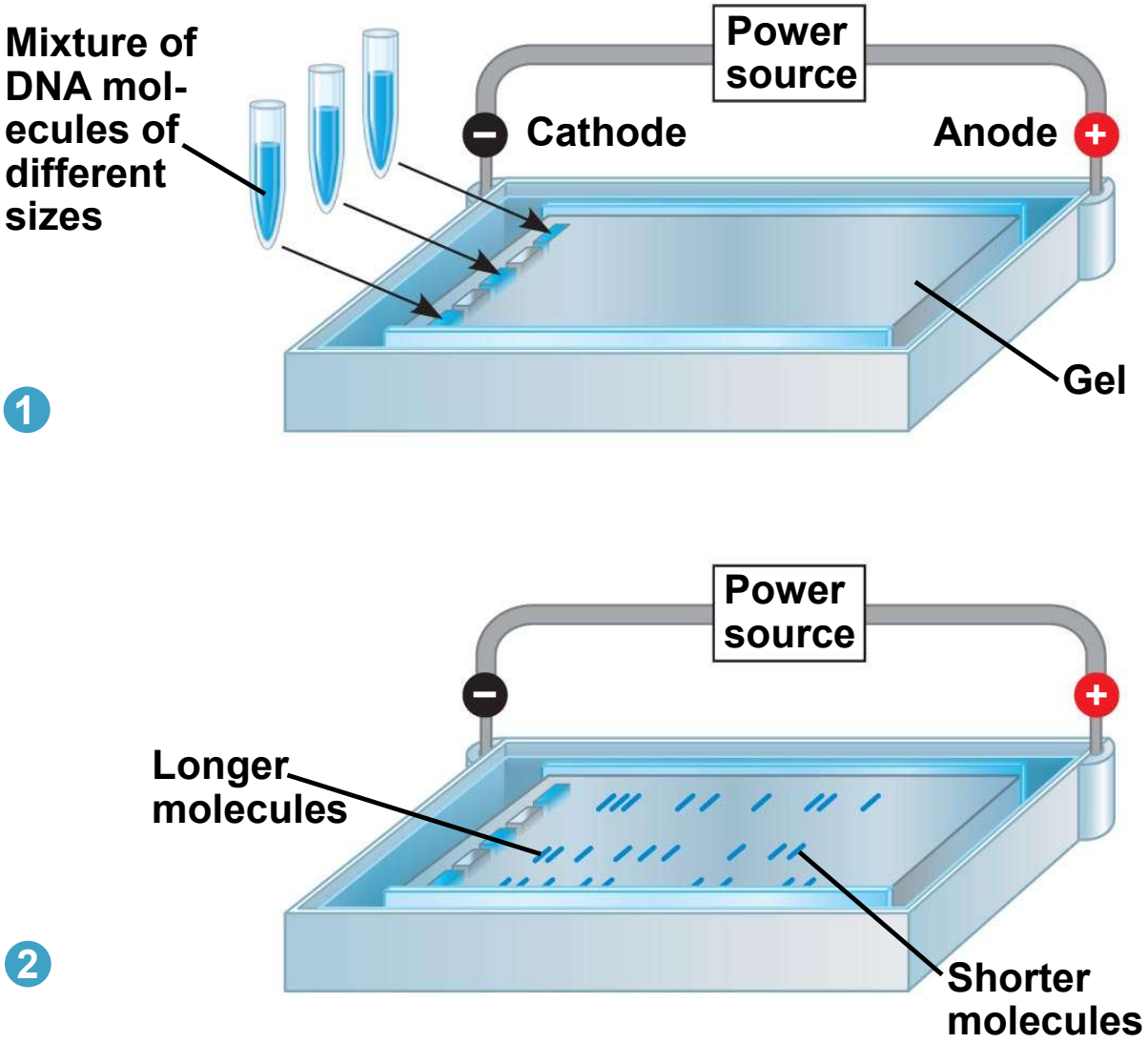
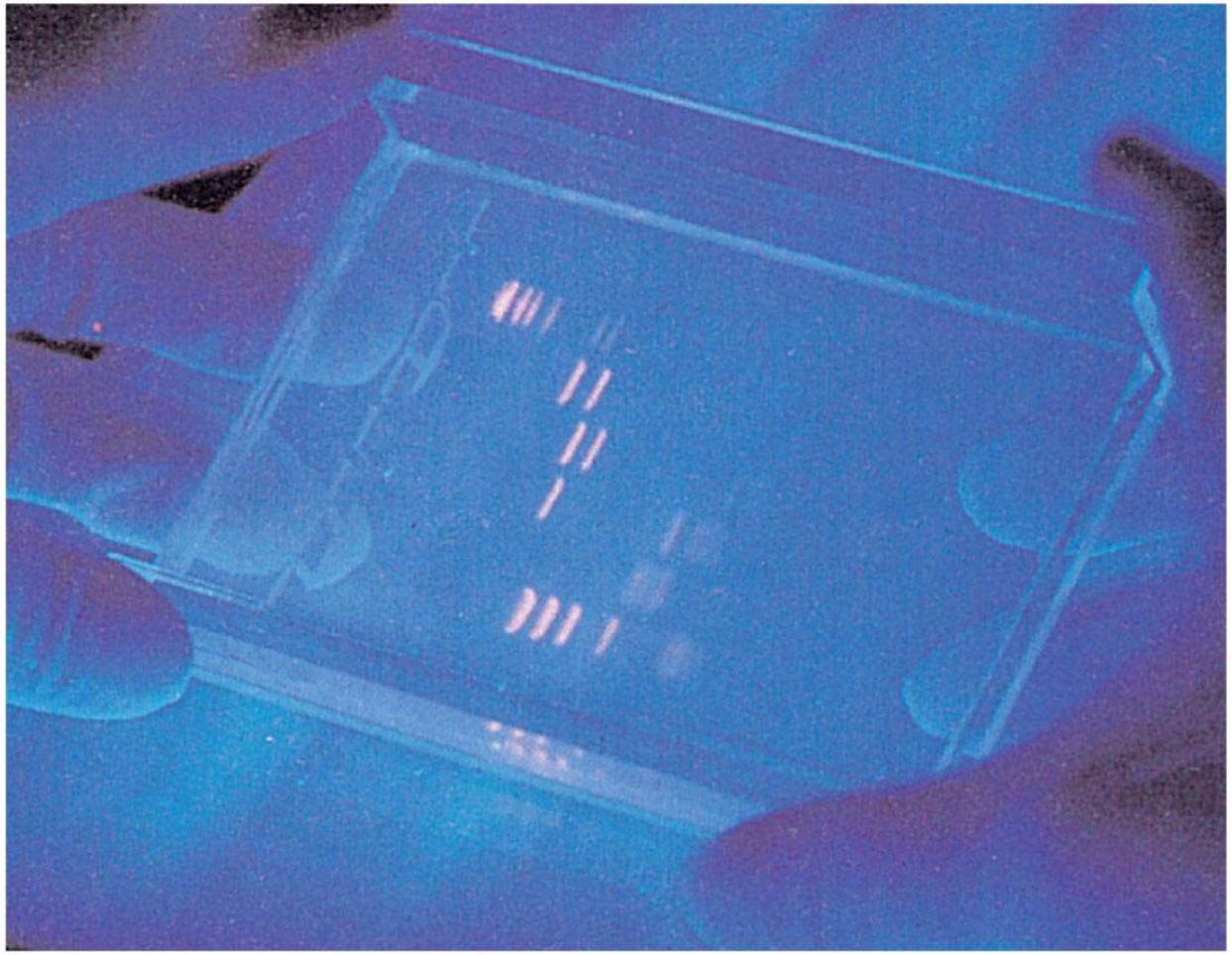


Fig. 20-9b

RESULTS



PCR and Contamination

- The most important consideration in PCR is contamination
- Even the smallest contamination with DNA could affect amplification
- For example, if a technician in a crime lab set up a test reaction (with blood from the crime scene) after setting up a positive control reaction (with blood from the suspect) cross contamination between the samples could result in an erroneous incrimination, even if the technician changed pipette tips between samples. A few blood cells could volitilize in the pipette, stick to the plastic of the pipette, and then get ejected into the test sample
- Modern labs take account of this fact and devote tremendous effort to avoiding cross-contamination

Optimizing PCR protocols

- While PCR is a very powerful technique, often enough it is not possible to achieve optimum results without optimizing the protocol

- Critical PCR parameters:

- - Concentration of DNA template, nucleotides, divalent cations (especially Mg2+) and polymerase

- - Error rate of the polymerase (Taq, Vent exo, Pfu)

- - Primer design

Primer design

- General notes on primer design in PCR:
  - Perhaps the most critical parameter for successful PCR is the design of primers
- Primer selection : Critical variables are:
  - - primer length
  - - melting temperature (Tm)
  - - specificity
  - - complementary primer sequences
  - - G/C content
  - 3’-end sequence
  - - oligonucleotides between 20 and 30 (50) bases are highly sequence specific

Primer design

- Complementary primer sequences:
  - primers need to be designed with absolutely no intra-primer homology beyond 3 base pairs. If a primer has such a region of self-homology, “snap back” can occur
  - another related danger is inter-primer homology: partial homology in the middle regions of two primers can interfere with hybridization. If the homology should occur at the 3' end of either primer, primer dimer formation will occur
- G/C content
  - ideally a primer should have a near random mix of nucleotides, a 50% GC content
  - there should be no PolyG or PolyC stretches that can promote non-specific annealing

Primer design

- 3’-end sequence :
  - the 3' terminal position in PCR primers is essential for the control of mis-priming
  - inclusion of a G or C residue at the 3' end of primers helps to ensure correct binding (stronger hydrogen bonding of G/C residues)

Techniques for analyzing the genome and its modifications

- Genomic DNA and complementary DNA (cDNA) banks
- Polymerase chain reaction (PCR)
- DNA microarrays

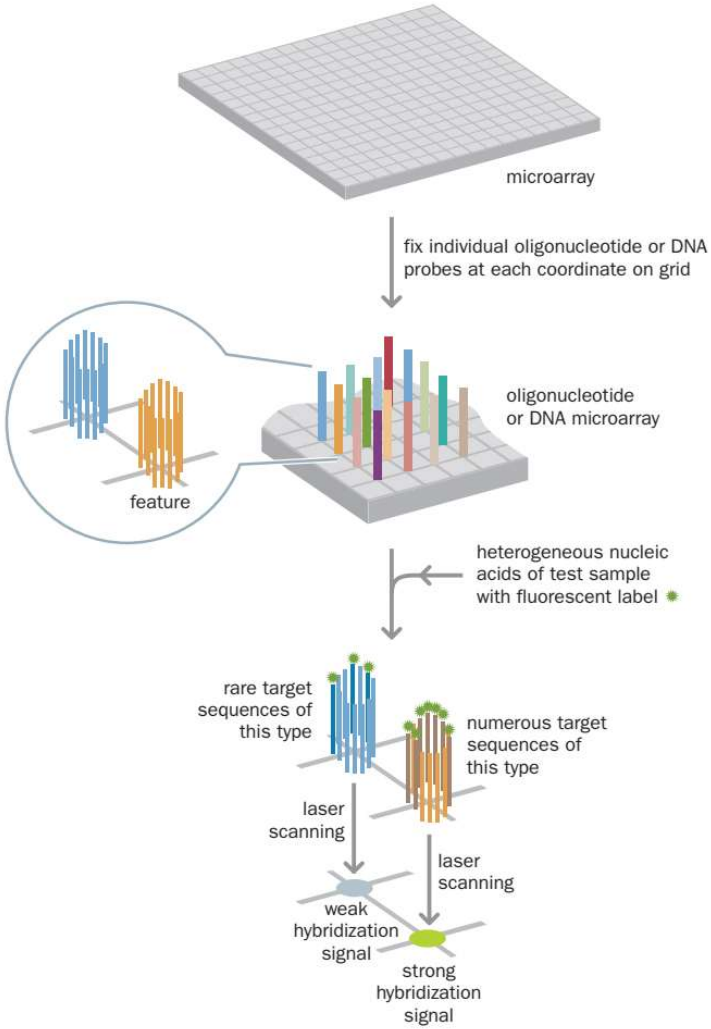
DNA Microarray (DNA chips)

DNA Microarray

- Hybridization technique enabling comparative genomic analysis of the expression of a large number of mRNA patterns
    - **Principle:**
- Specific oligonucleotides (probes) targeting different known genes or cDNAs are immobilized on a solid support (matrix), whose role is to detect labeled and complementary targets.
- The probes are grafted onto the support.
- The total RNA from both cell populations (e.g., healthy cell *versus* diseased cell) is extracted, converted into cDNA , and each labeled with a different fluorochrome.
- The two samples are then deposited simultaneously onto the biochip; this is called competitive hybridization.
- Hybridization signals are detected by fluorescence and quantified.

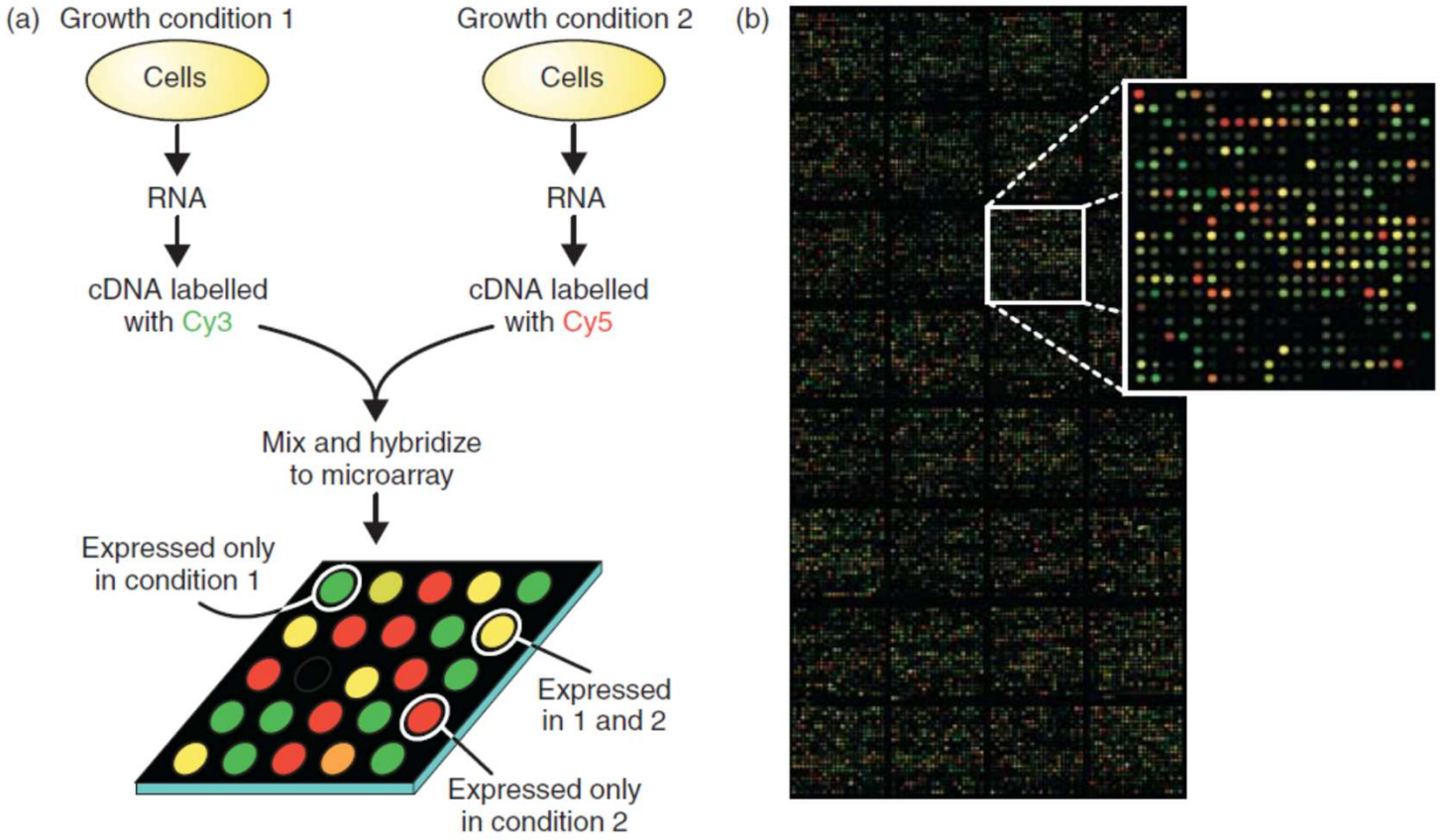
DNA Microarray (DNA chips)

DNA Microarray

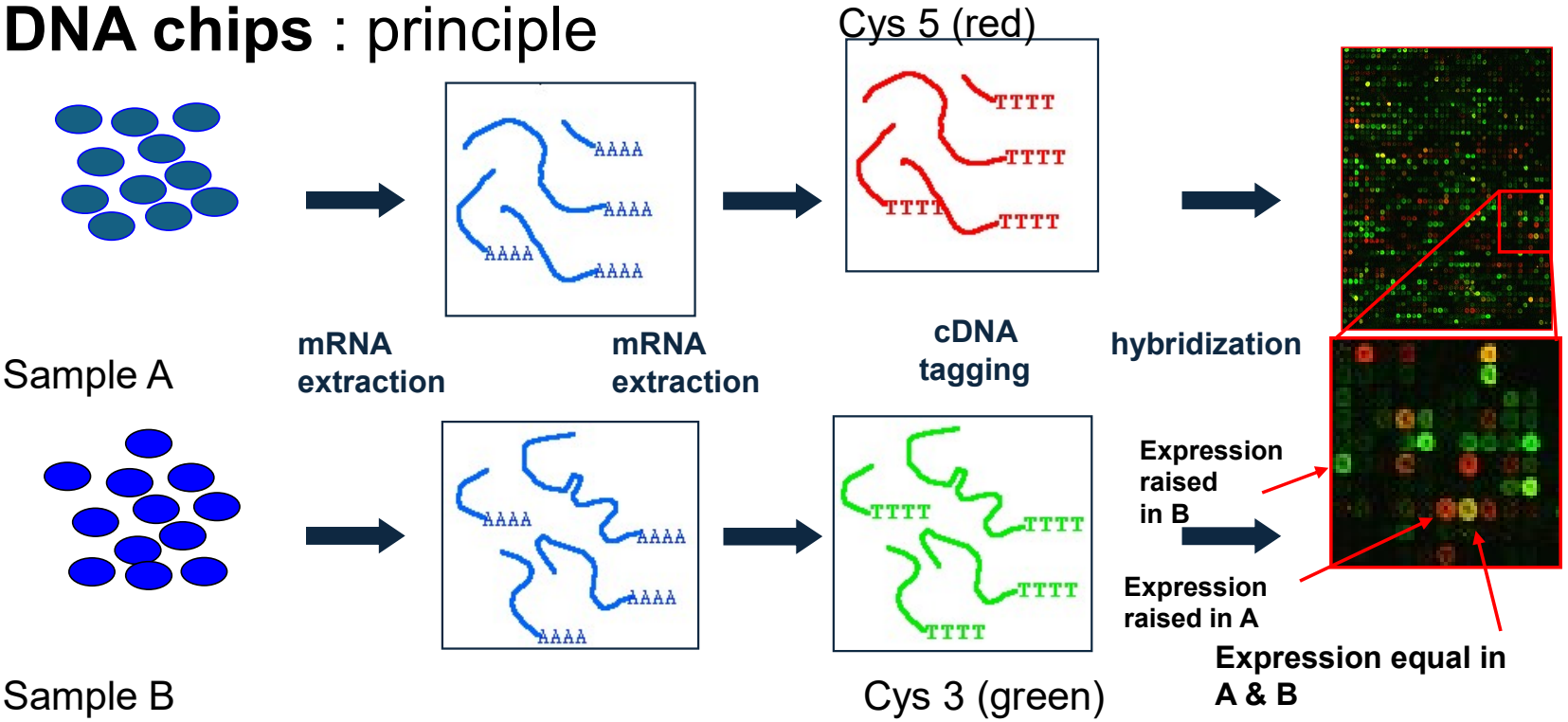


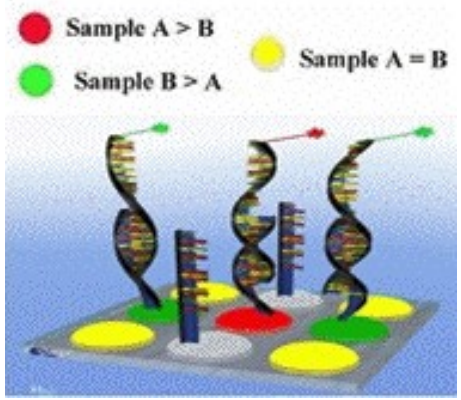
DNA Microarray (DNA chips)

DNA Microarray

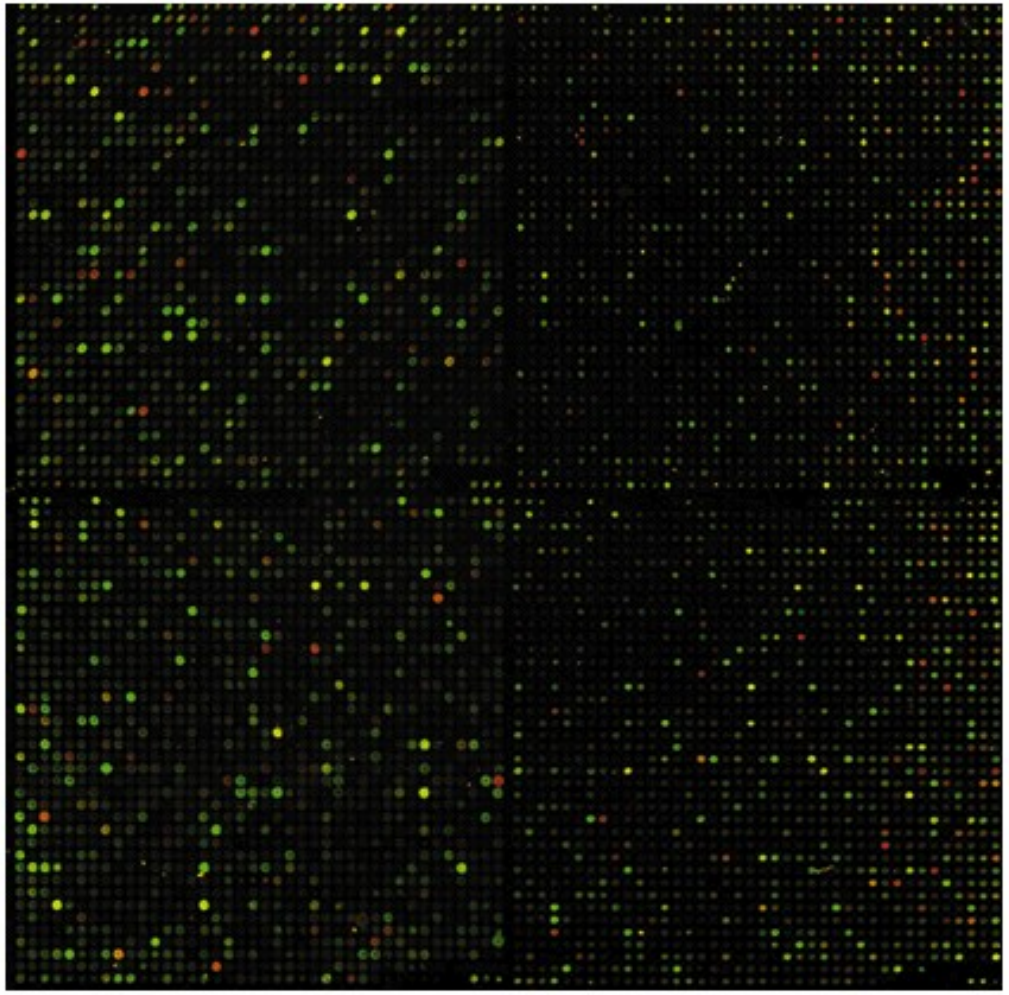


DNA chips : principle





Microarray results:



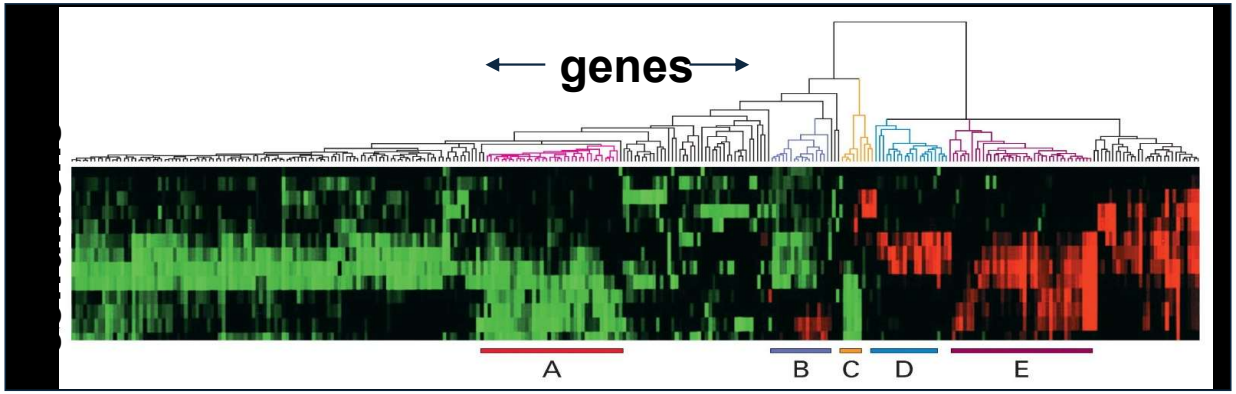
Data Analysis: Clustering

- Principal component analysis (PCA)
- “Although the mathematics is complex, the basic principles are straightforward. Imagine taking a three dimensional cloud of data points and rotating it so that you can view it from different perspectives.You might imagine that certain views would allow you to better separate the data into groups than other views. PCA finds those views that give you the best separation of the data.”

Data Analysis: Clustering

Hierarchical clustering

At the beginning, each gene is a cluster. In each subsequent step, the two closest clusters are merged until only one cluster remains. There are a few different ways of doing this.



- simple and widely used method
- in large clusters, can lose true representation of expression pattern
- cannot go back—early errors become fixed

Data Analysis: Clustering