

Recombinant DNA technology

Genomics

Proteomics

**Genetic engineering, recombinant DNA technology,
encompasses an array of molecular techniques
that can be used to
ANALYZE
ALTER
and
RECOMBINE
virtually any DNA sequence**

Recombinant DNA technology used:

- In many other fields:
 - Biochemistry
 - Microbiology
 - Developmental biology
 - Neurobiology
 - Evolution
 - Ecology
- To create commercial products like, drugs, hormones, enzymes, and crops- **BIOTECHNOLOGY**
- To study the nature of cancer and several diseases- **MOLECULAR MEDICINE**
- In genetic diagnosis and infectious diseases- **DIAGNOSIS**
- To treat genetic disorders- **GENE THERAPY**

Recombinant DNA techniques

Locating specific sequences

Isolating a specific sequence

Amplifying a particular DNA sequence

Finding genes

Transferring DNA sequences into recipient cells

Typical situation:

- **Isolate** a particular human gene
- Place it **inside** bacterial cells
- Use the bacteria to **produce** large quantities of the encoded human protein

– **If located and isolated**

- **Insert** in a stable form
- Successfully **replicate**
- Ensure it is properly **transcribed** and **translated**

Isolating a specific sequence

- Tools
 - Enzymes for DNA modification
 - DNA polymerases
 - Nucleases
 - DNA ligases
 - End-modification enzymes
 - Restriction enzymes
 - Methods for visualizing and resolve DNA fragments
- Nucleic acids purification
- PCR

Amplifying a particular DNA sequence

- ***In vivo***

- Cloning

- Vectors

- Plasmids
 - Cosmids
 - Fagemids
 - Phages
 - Expression vectors
 - retrovirus

Prokaryotic
Eukaryotic
Shuttle

Ori
Selection marker

- Introducing DNA in host cells
 - Selection

- ***In vitro***

- PCR (several applications)

Table 18.4 Considerations in developing a cloning strategy

Step in Gene Cloning	Considerations
1. Isolation of DNA fragment	<ul style="list-style-type: none">a. The purpose of cloning (is <u>expression required?</u>). Is the <u>entire</u> sequence needed?b. What is known about the gene and the protein (if any) that it encodes?c. The size of the gene.d. Is the chromosomal location of the gene known?e. Size of the genome from which the gene is isolated.
2. Joining DNA fragment to vector	<ul style="list-style-type: none">a. Type of cloning vector used.<ul style="list-style-type: none">i. The <u>size</u> of the gene.ii. The <u>organism</u> into which the gene will be cloned.iii. The need for a selection mechanism.iv. Whether <u>expression</u> is required.v. Efficiency of transfer to host cell required.vi. The purpose of cloning.b. Method of joining the gene to vector.<ul style="list-style-type: none">i. Simplicity of method.ii. Availability of restriction sites.iii. The need to retrieve the fragment from the vector.iv. Whether expression is required.v. The purpose of cloning.
3. Transfer of recombinant vector to host cell	<ul style="list-style-type: none">a. Type of cloning vector used.
4. Identification of cells carrying recombinant molecule	<ul style="list-style-type: none">a. Known information about the gene.b. Type of cloning vector used.c. Efficiency of transfer.d. Purpose of cloning.

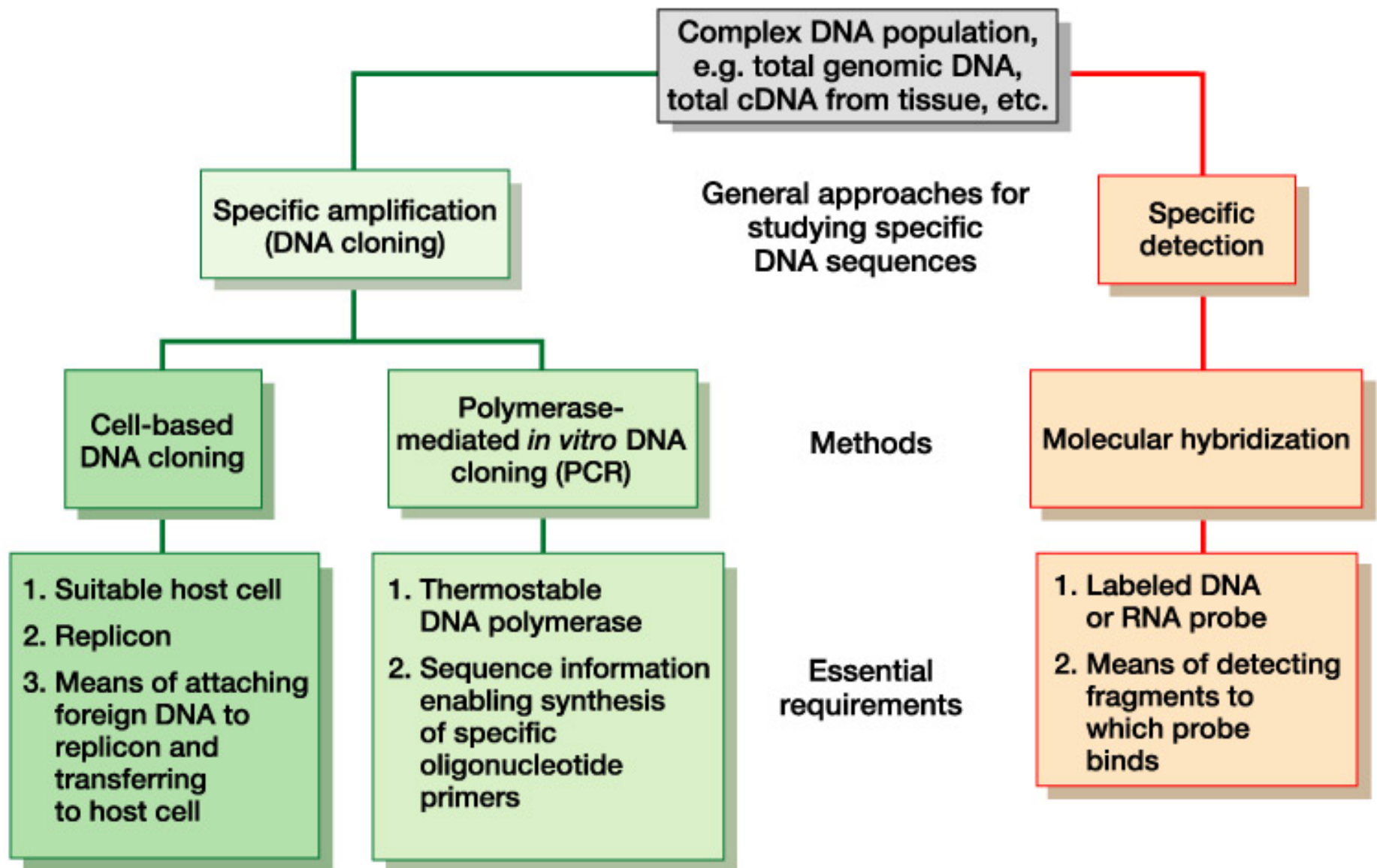


Figure 5-1 Human Molecular Genetics, 3/e. (© Garland Science 2004)

Locating specific sequences

- **Probe construction**
 - DNA or RNA
 - Radioactive or nonisotopic labeling
 - Synthetic or cloned, isolated and purified
- **Techniques**
 - hybridization assays
 - **Filter hybridization**
 - Southern (DNA)
 - Northern (RNA)
 - Western (proteins)
 - **Solution hybridization**
 - ***In situ* hybridization**

Nucleic Acid Hybridization

- Nucleic acid hybridization is a fundamental **tool** in molecular genetics which takes advantage of the ability of individual **single-stranded** nucleic acid molecules to form **double stranded** molecules (that is, to **hybridize** to each other)

Standard nucleic acid hybridization assays

- A labeled nucleic acid - a **probe** - to identify related DNA or RNA molecules
- Complex mixture of unlabeled nucleic acid molecules- the **target**
- Base complementarity** with a high degree of similarity between the probe and the target.

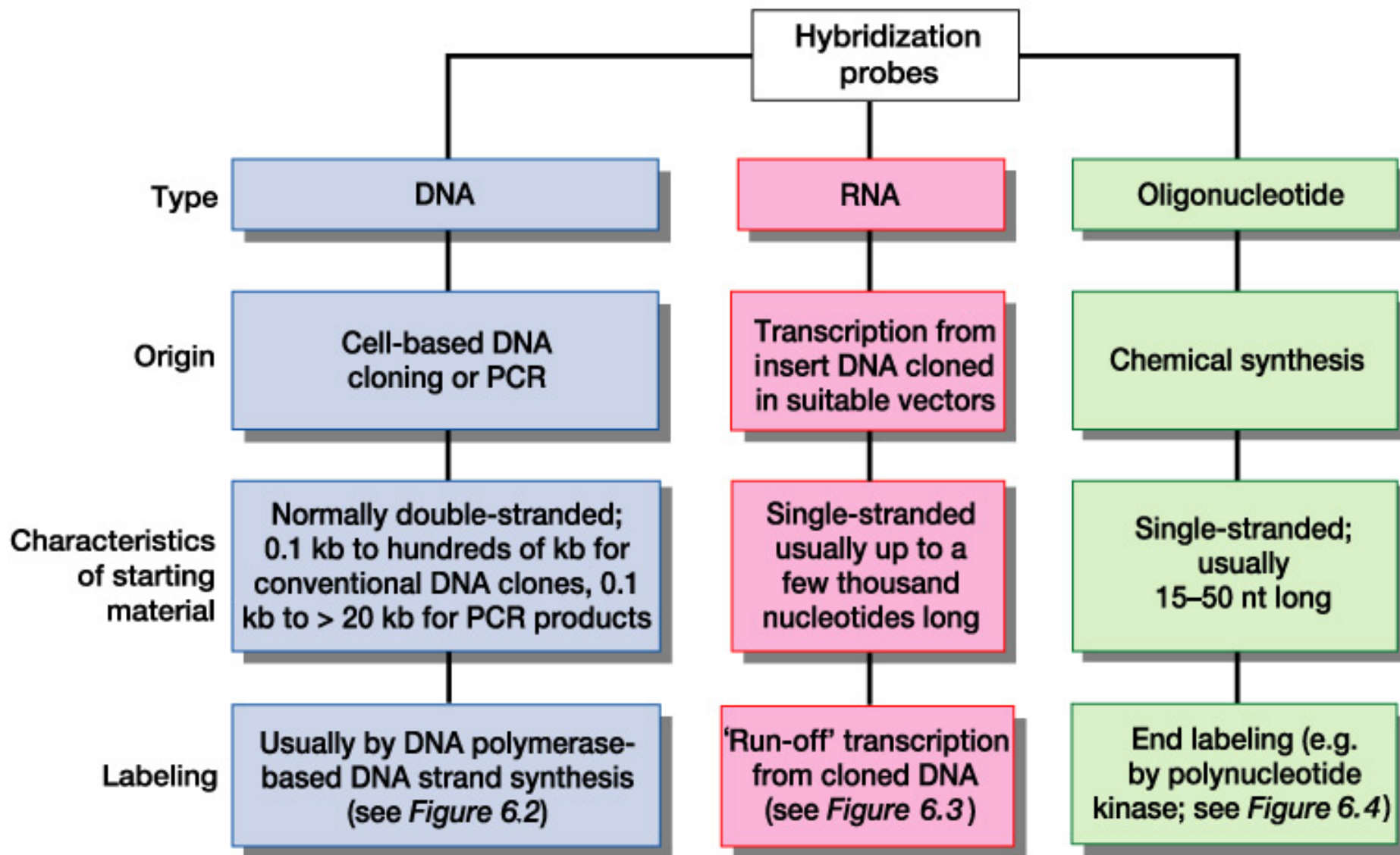
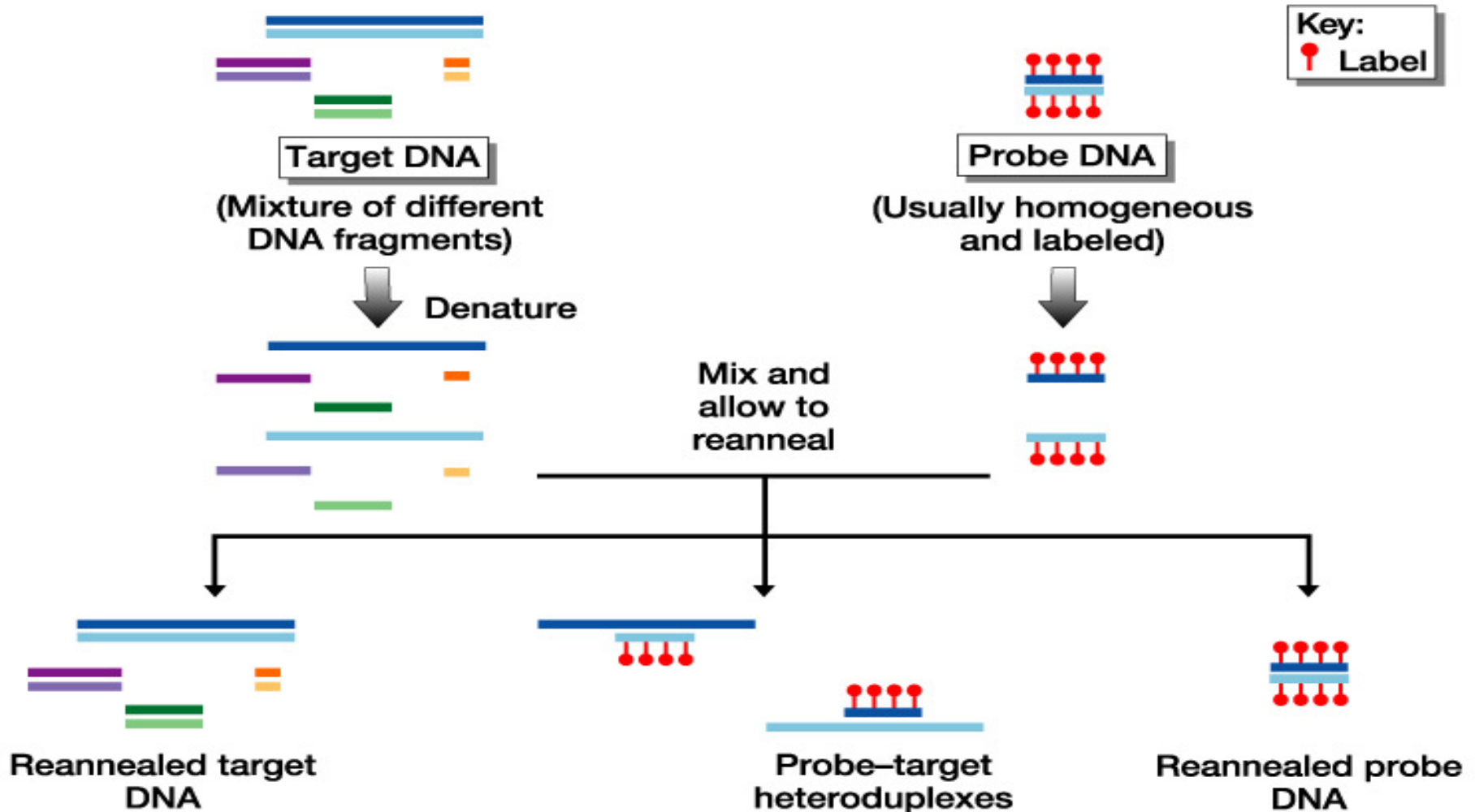


Figure 6-1 Human Molecular Genetics, 3/e. (© Garland Science 2004)

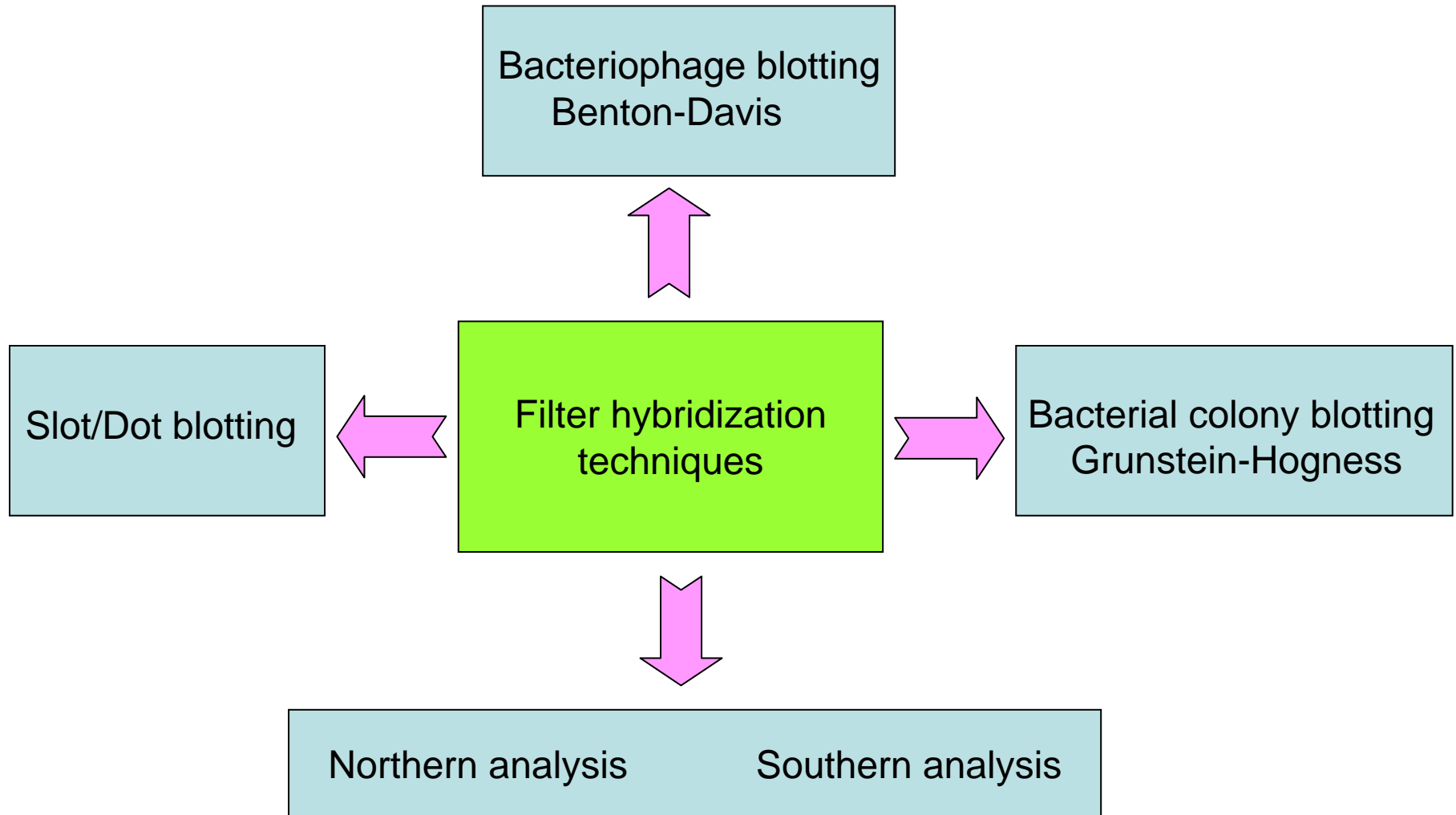
Probes

- **DNA labelling**
 - 5'
 - 3'
 - Uniform labeling
 - Nick translation
 - Random primer
 - PCR-mediated labeling
- **RNA labelling**
 - *In vitro* transcription of a cloned DNA insert
- **Different probes**
 - Radioactive labeling or isotopic labeling
 - Nonradioactive labeling or nonisotopic labeling

Nucleic acid hybridization- formation of heteroduplexes



Filter hybridization methods



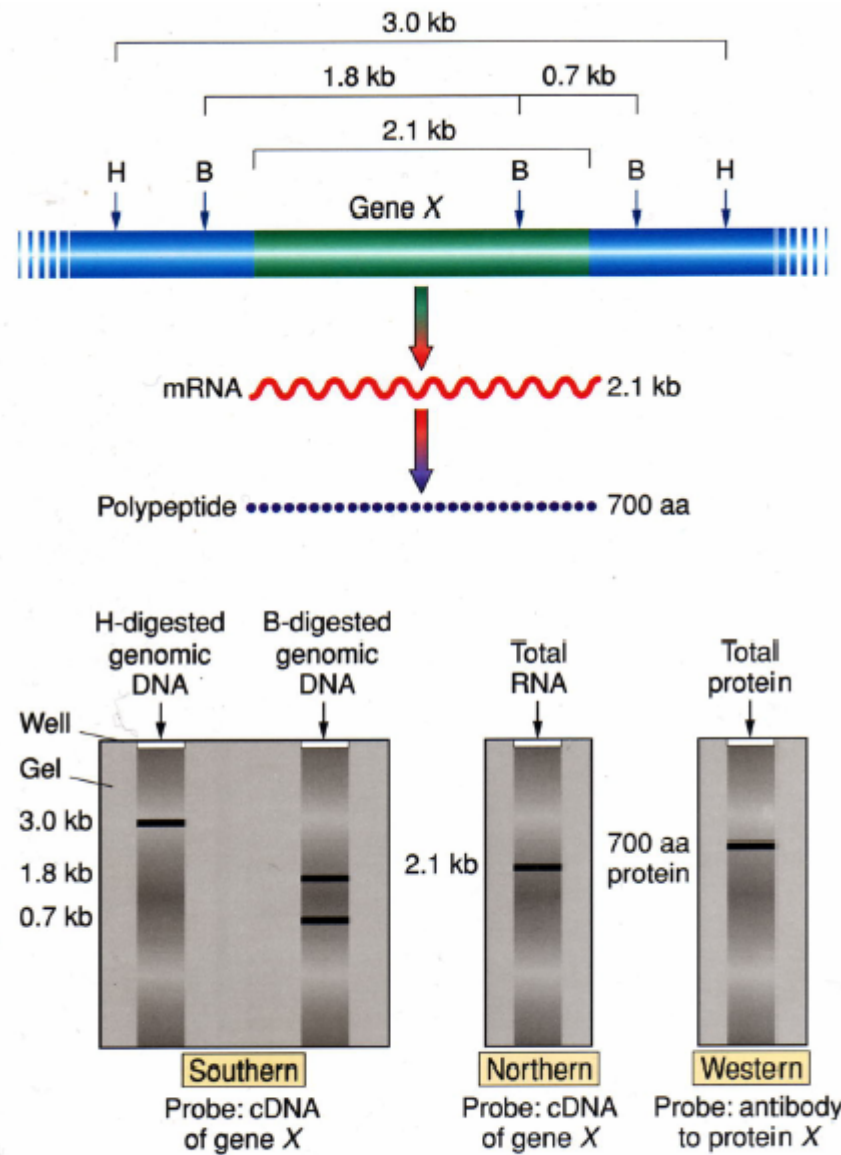
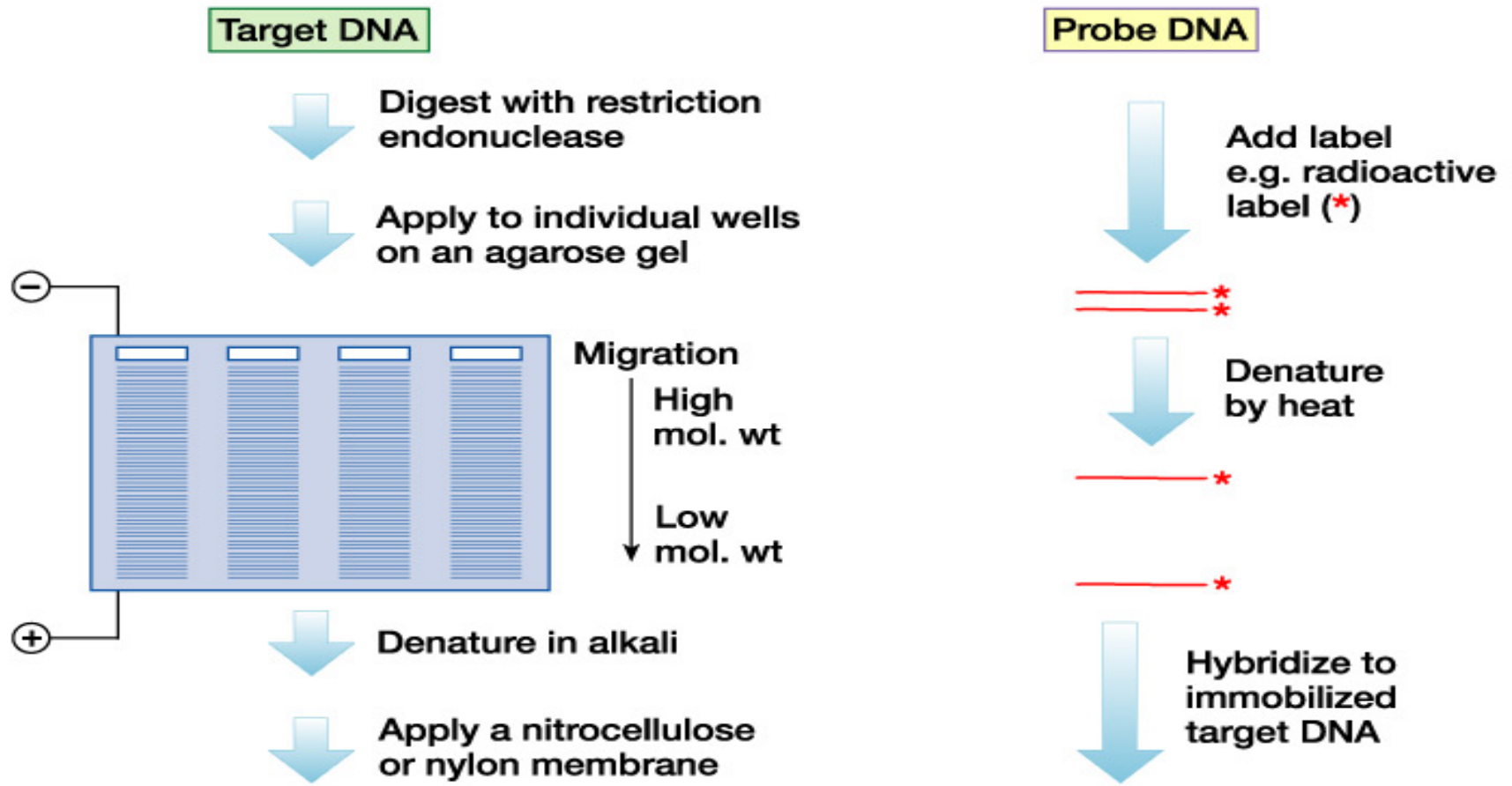
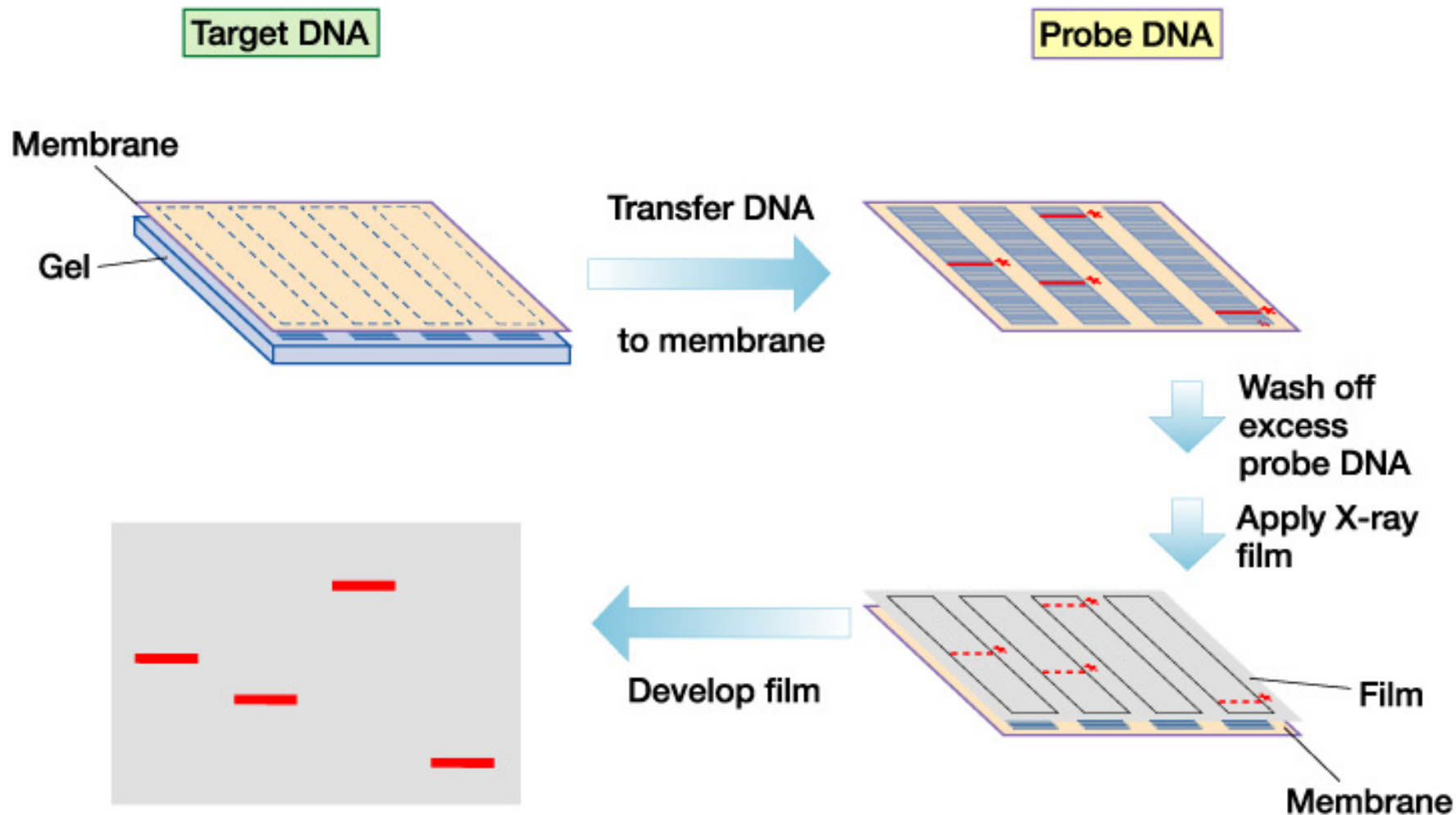


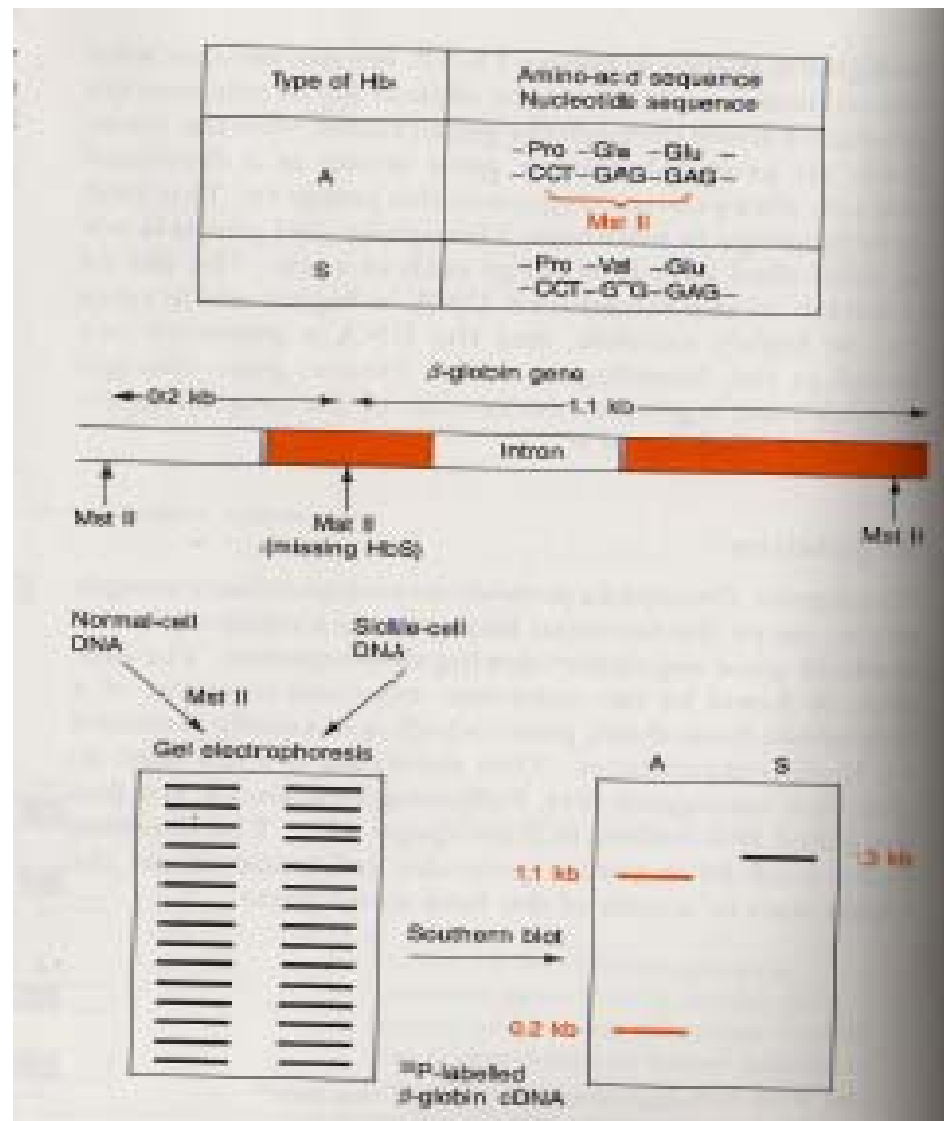
Figure 10-20
 Anthony J. F. Griffiths et al. MODERN GENETIC ANALYSIS
 Copyright © 1999 by W. H. Freeman and Company

Principles of Southern blot





Southern applications- example



DNA polymorphisms

- DNA polymorphisms are specific sites in the genome (locus) where the precise sequence of DNA tends to differ in unrelated individuals
- These polymorphisms when found in genes, accounting for the differences in phenotype, are usually referred as mutations or variants (alleles)

DNA polymorphisms in intergenic regions

- Large number polymorphisms have accumulated in the **intergenic regions** of most eucaryotic organisms (no selective pressure)
- These polymorphisms have turned out to be valuable tools for **genetic mapping** and for **forensic identification**

DNA Markers (or DNA polymorphisms) present in genomic DNA

- Single-nucleotide polymorphisms (**SNPs**)
 - DNA sequence analysis
- Restriction fragment length polymorphisms (**RFLPs**)
 - Southern blot
- Tandem repeat polymorphisms or **SSLPs**
(simple sequence length polymorphisms)
 - ex. VNTR (variable number of tandem repeats, 9-80 bp)
 - PCR, Southern blot

RFLPs

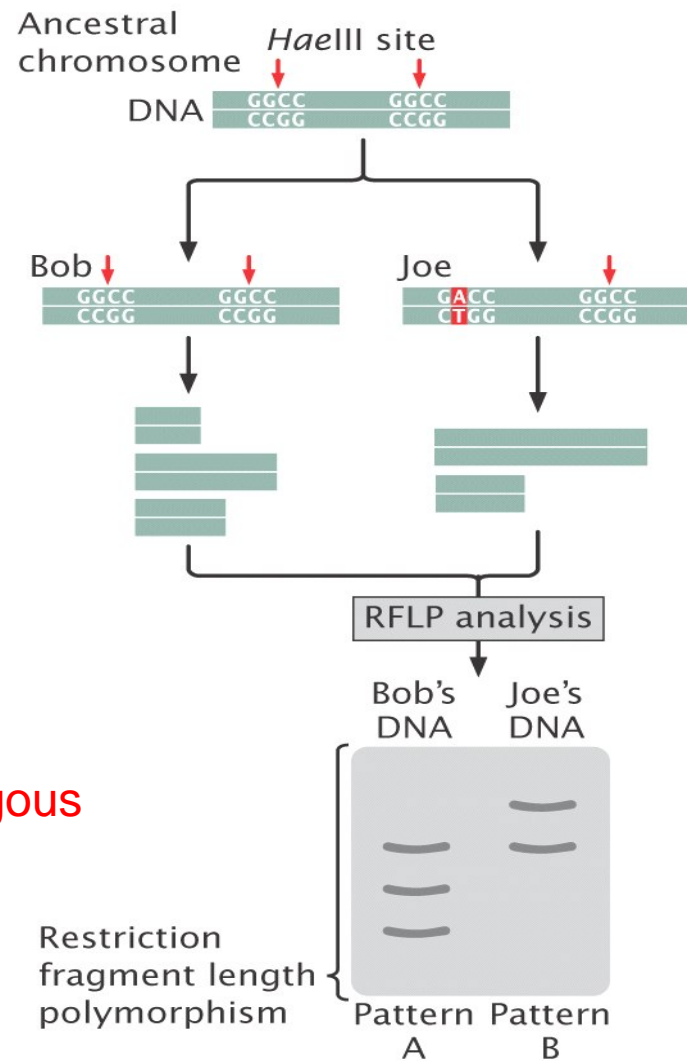
Restriction fragment length polymorphisms

- RFLPs are variations (polymorphisms) in the patterns of fragments produced when DNA molecules are cut with the same restriction enzyme

Different patterns of fragments  Differences in DNA sequences

- Inherited differences used in mapping (genetic markers)

RFLP- a genetic marker that can be used in mapping



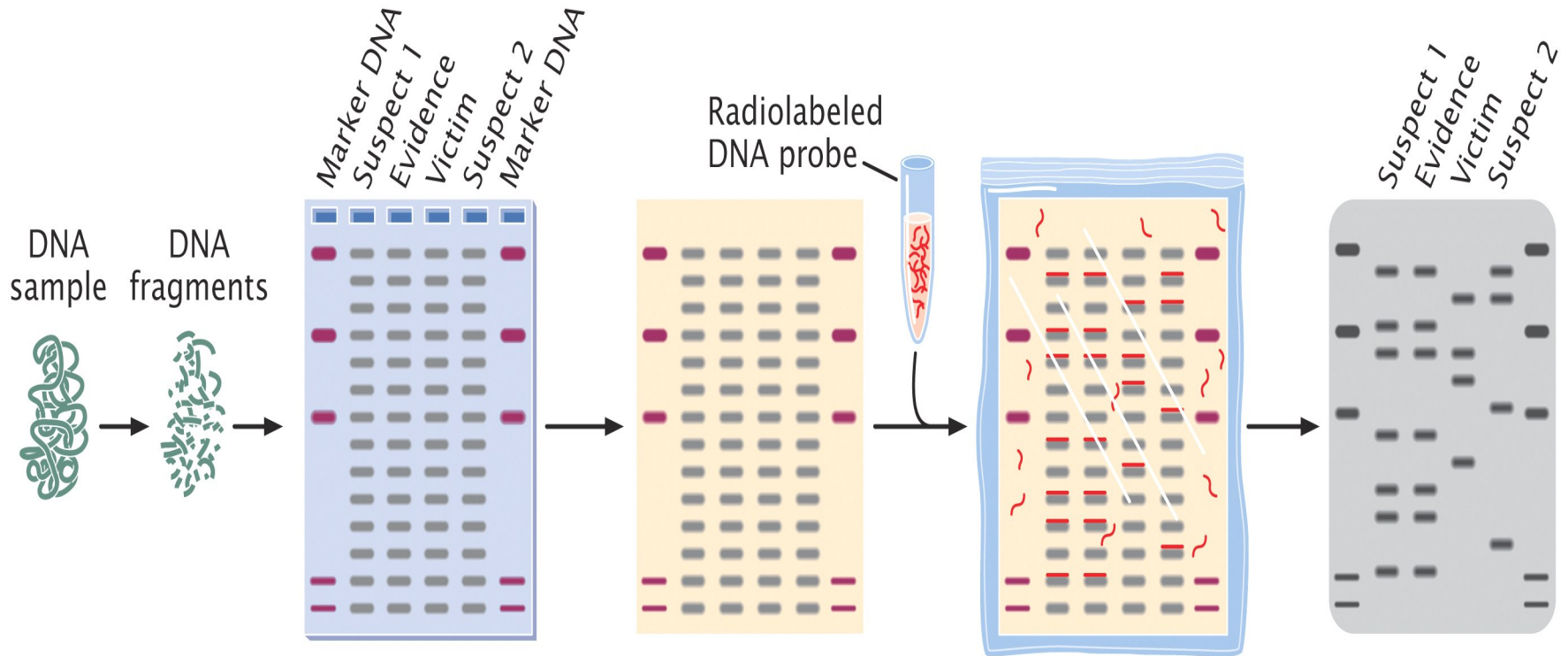
Bob and Joe are homozygous

RFLPs are often found in noncoding regions of DNA and are therefore frequently quite variable in humans.

DNA fingerprinting- is the method in which DNA sequences are used to identify a person.

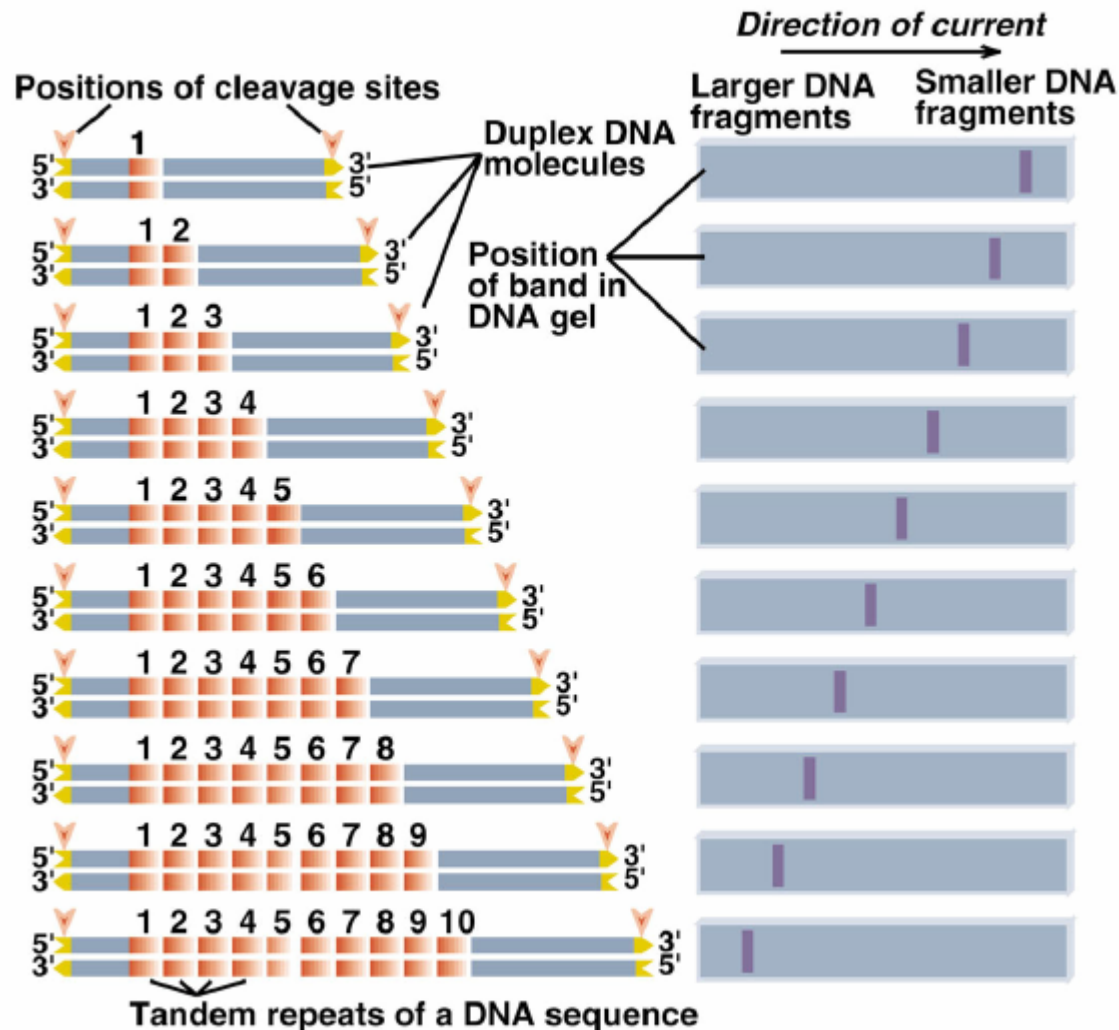
DNA fingerprinting is a powerful tool for criminal investigations and other forensic applications

DNA fingerprinting

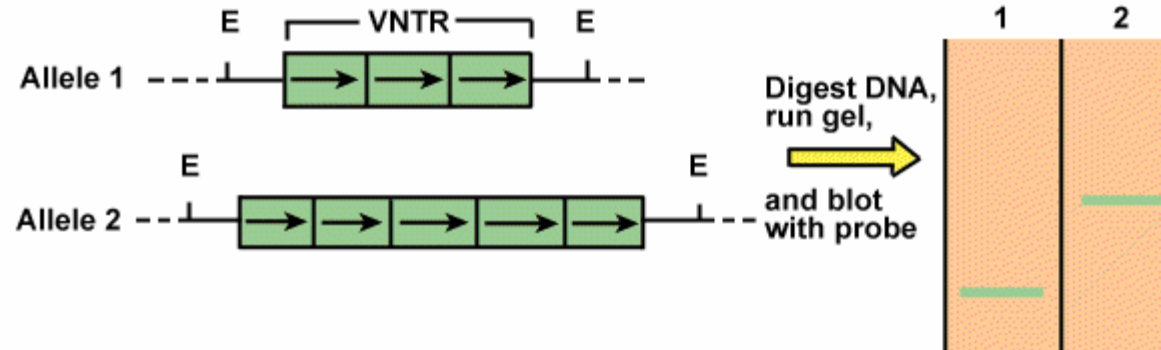


Note: **probes** in DNA fingerprinting DNA fragments of a specific chromosome region (associated to a specific **RFLP** or a **VNTR**)

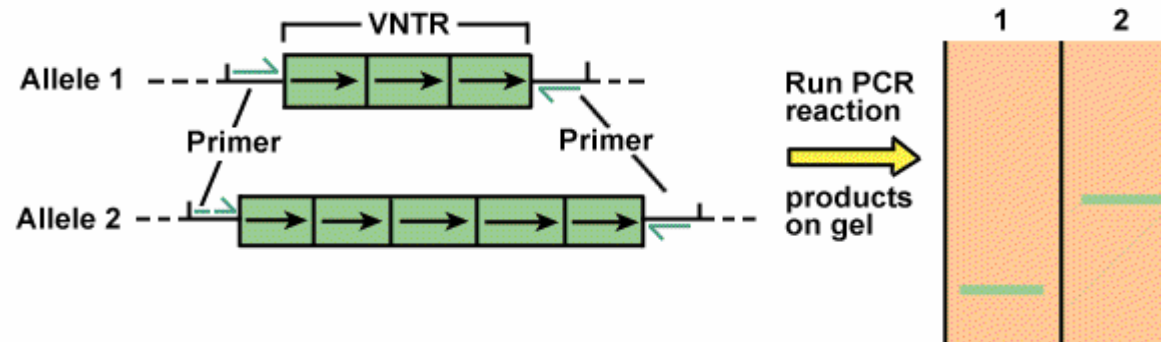
Tandem Repeat Polymorphism or **SSLPs** (simple sequence length polymorphisms)



VNTR detection by Southern blot



VNTR detection by PCR



Genomics

A genome sequence is not an end in itself

Structural genomics
Functional genomics
Comparative genomics

Genomics attempts to understand the:

- Content
- Organization
- Function
- Evolution

of genetic information contained in whole genomes

EX.

Table 19.1

Comparison of the genomes of *Mycobacterium leprae*, which causes leprosy, and *Mycobacterium tuberculosis*, which causes tuberculosis

Characteristics	<i>M. leprae</i>	<i>M. tuberculosis</i>
Genome size (bp)	↓ 3,268,203	4,411,532
Percentage of genome that encodes proteins	↓ 49.5%	90.8%
Protein-encoding genes (bp)	1604	3959
Pseudogenes (bp)	↑ 1116	6
Gene density (bp/gene)	2037	1114
Average length of gene (bp)	1011	1012

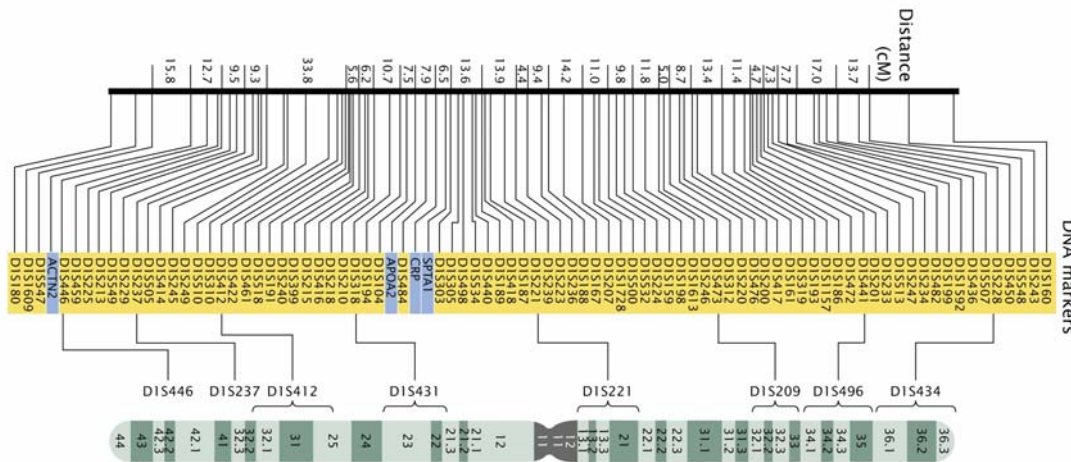
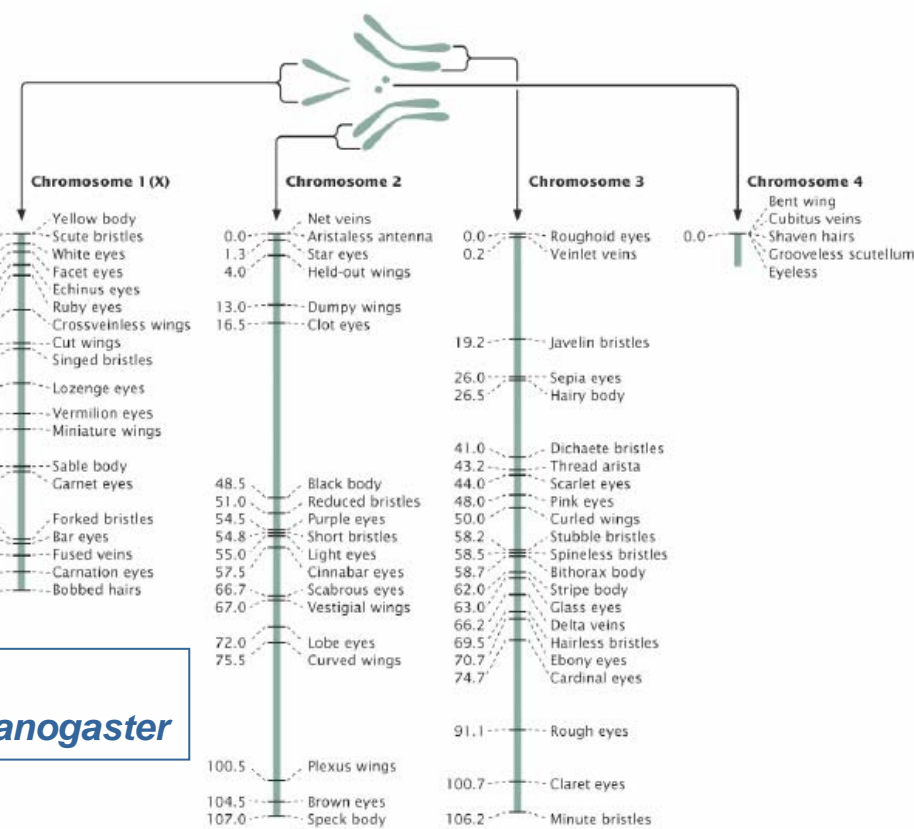
Source: S. T. Cole et al., Massive gene decay in the leprosy bacillus, *Nature* 409 (2001), p. 1007.

Structural genomics

Determines the **organization** and **sequence** of the genetic information

Genetic maps
are based
on rates of
recombination

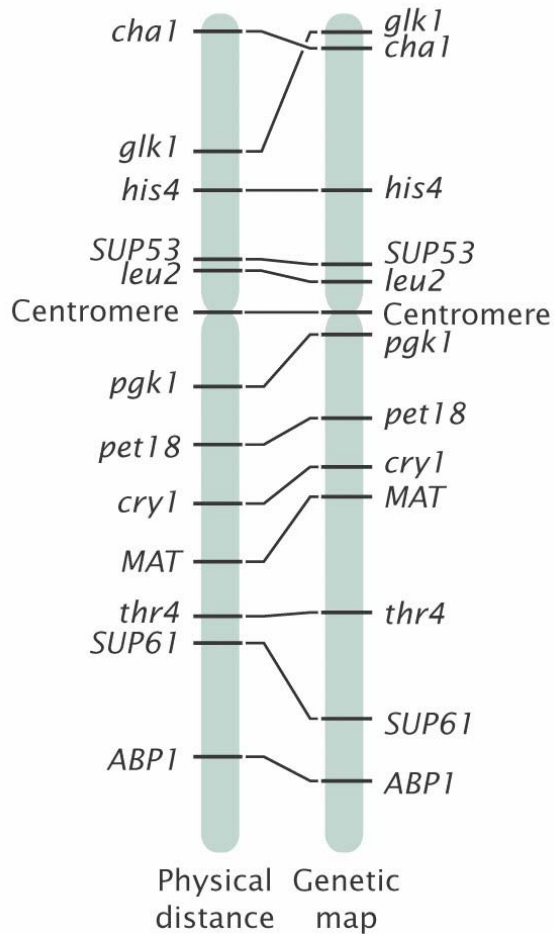
Genetic map of
Drosophila melanogaster



Physical map of
human chromosome 1

Techniques for creating
physical maps

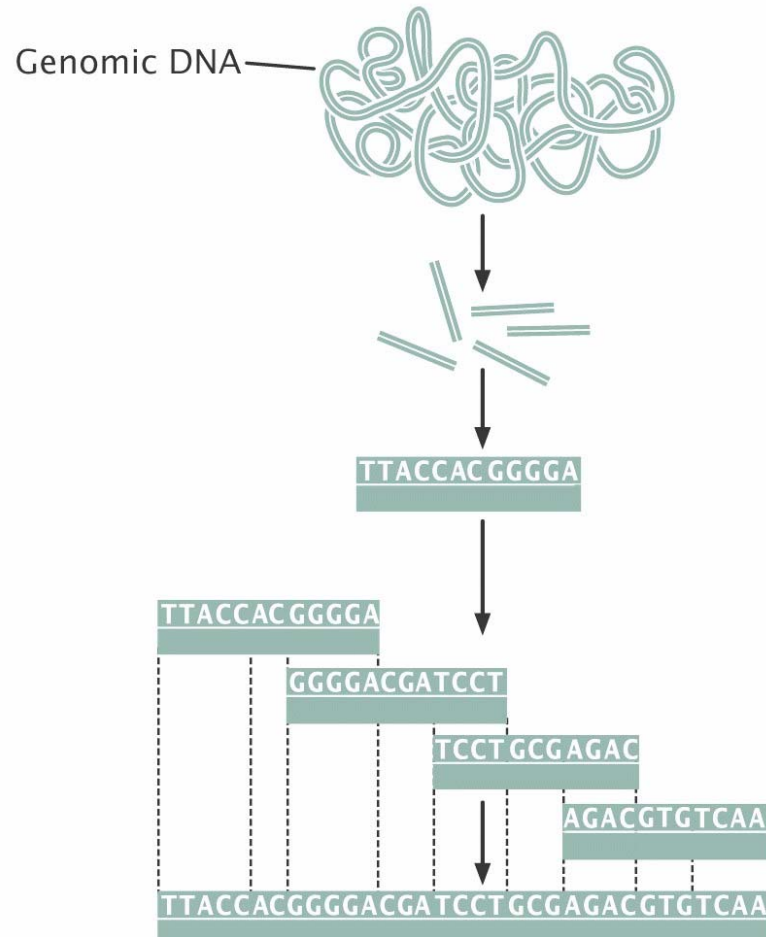
- Restriction mapping
- Sequence-tagged site (STS) mapping
- DNA sequencing



Data from genetic and physical maps may differ in relative distances and even in the position of genes on a chromosome

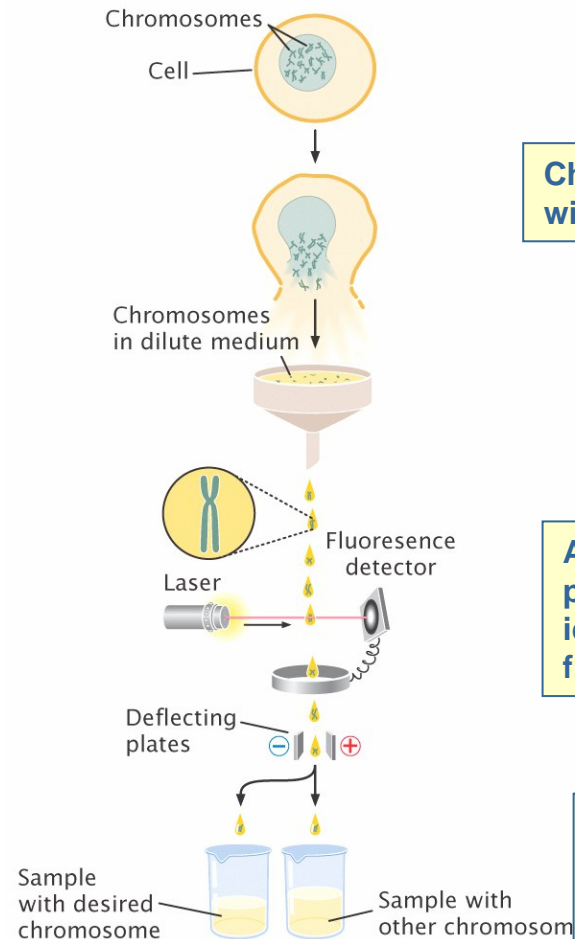
Yeast chromosome III

**Whole-genome sequencing
utilizes sequence overlap to
align sequenced fragments**



**Genomic sequence assembled
by powerful computer programs**

**Flow cytometry can be used to
separate individual chromosomes**



**Chromosomes are stained
with fluorescent dye**

**The dye taken up is
proportional to
chromosome size**

**A detector determines a
particular chromosome's
identity from its unique
fluorescence**

**And signals a charge ring
to apply a charge
to the designated drops,
which are deflected into
a separate receptacle**

Functional genomics

Characterizes the **function** elucidated
by structural genomics

Goals of functional genomics

Identifying genes

Recognizing their organization

- Understanding their function
and
- Identifying all the RNA molecules transcribed from a genome (**Transcriptome**)
- All the proteins encoded by the genome (**Proteome**)
- Computational methods
- Experimental methods

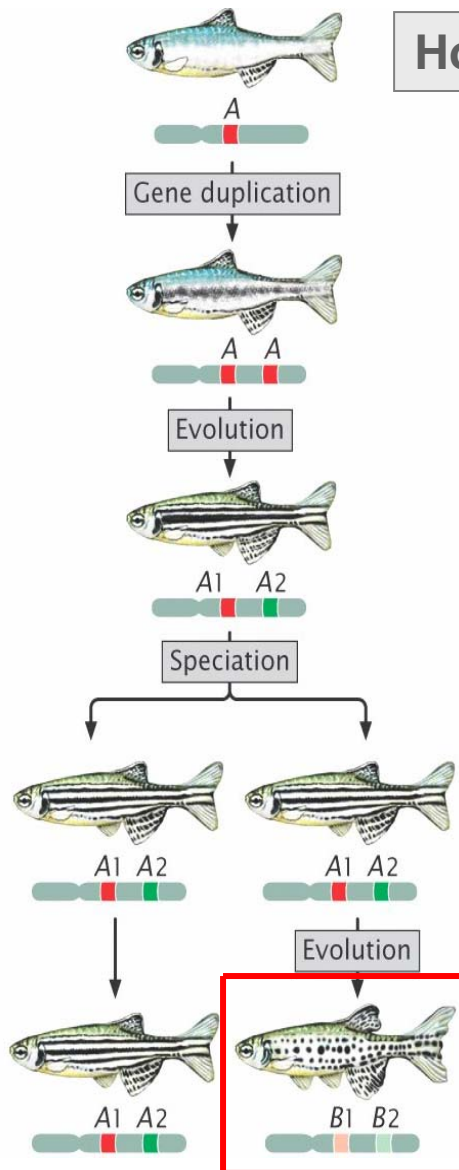
Computational methods

- Develop computational methods bypass the isolation and chracterization of individual genes
 - Homology searches
 - DNA or protein sequences (ex. protein domains)
 - Same or different organisms

Genes evolutionary related- **Homologous**

- **paralogs** (homologous gene in the same organism from duplication of a single gene- alfa and beta subunit of hemoglobin)
- **orthologs** (same gene in different species evolved from a common ancestor- alfa subunit of mice and humans)

Paralogs often involved in new functions



Homologous sequences are evolutionarily related

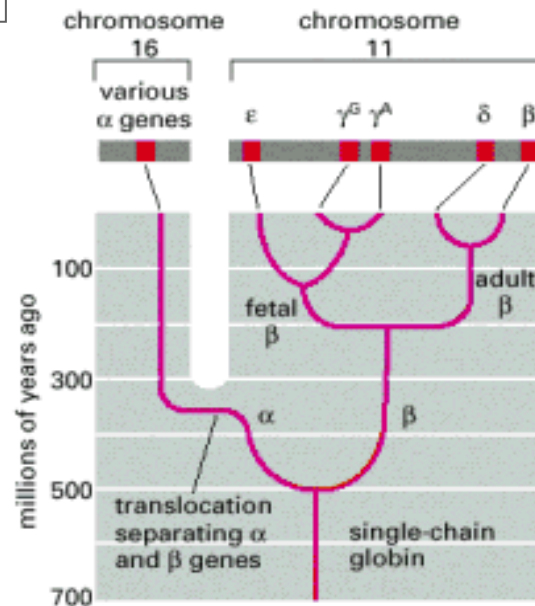
Paralogous genes – homologous genes in the **same species** that arose through the duplication of a single ancestral gene

Orthologous genes – homologous genes in the different species, because the **two species** have a common ancestor that also possessed the gene

Genes A1 and A2 are paralogs
 Genes B1 and B2 are paralogs
 Genes A1 and B1 are orthologs
 Genes A2 and B2 are orthologs

An evolutionary scheme for the globin chains that carry oxygen in the blood of animals

β -like globin gene family



A relatively recent gene duplication of the γ -chain gene produced γ^G and γ^A , which are fetal β -like chains of identical function.

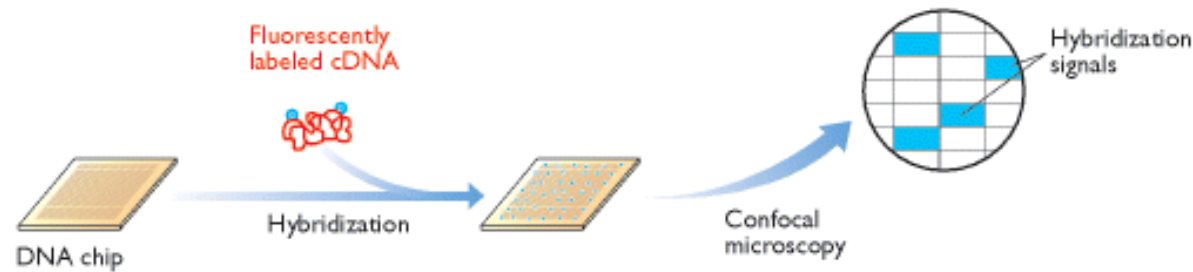
The location of the globin genes in the human genome is shown at the top of the figure

Gene expression and microarrays (or gene chips)

- Rely on nucleic acids hybridization
- Monitors expression of thousands of genes simultaneously- which genes are active in a particular tissue or moment of a biological process such as development or disease progression.

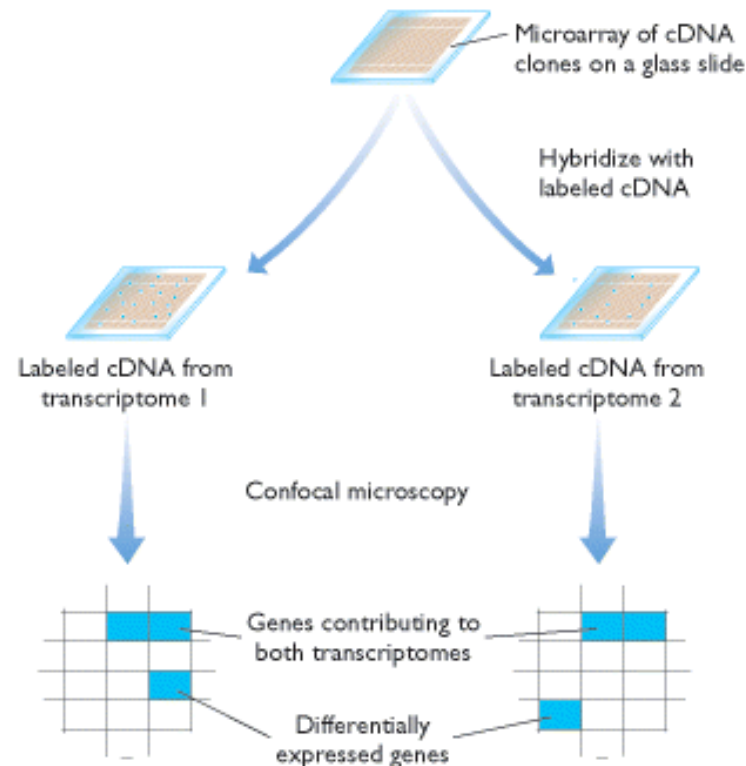
Transcriptome analysis

(A) Transcriptome analysis with a small genome



DNA chip carrying oligonucleotides representing all the genes in a small genome

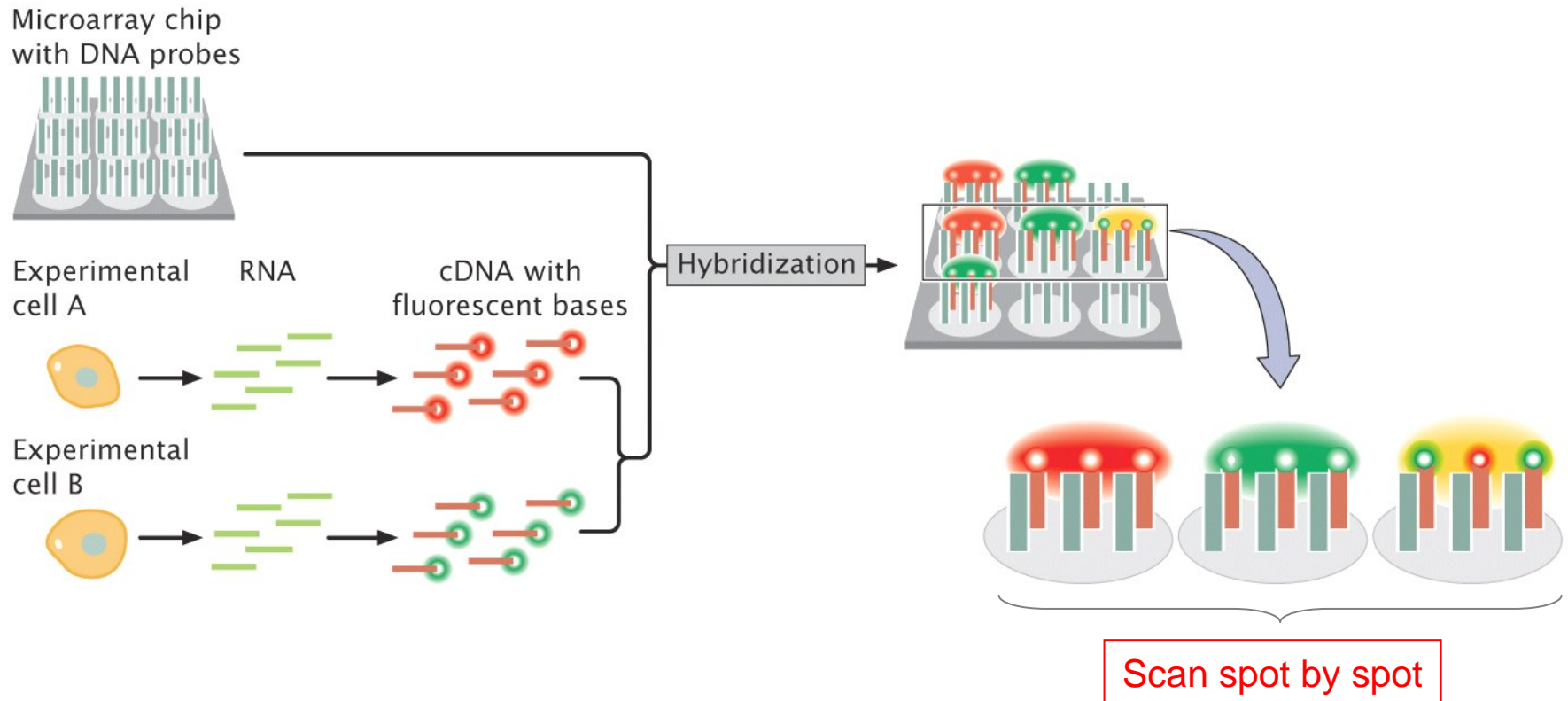
(B) Microarray analysis of a large genome



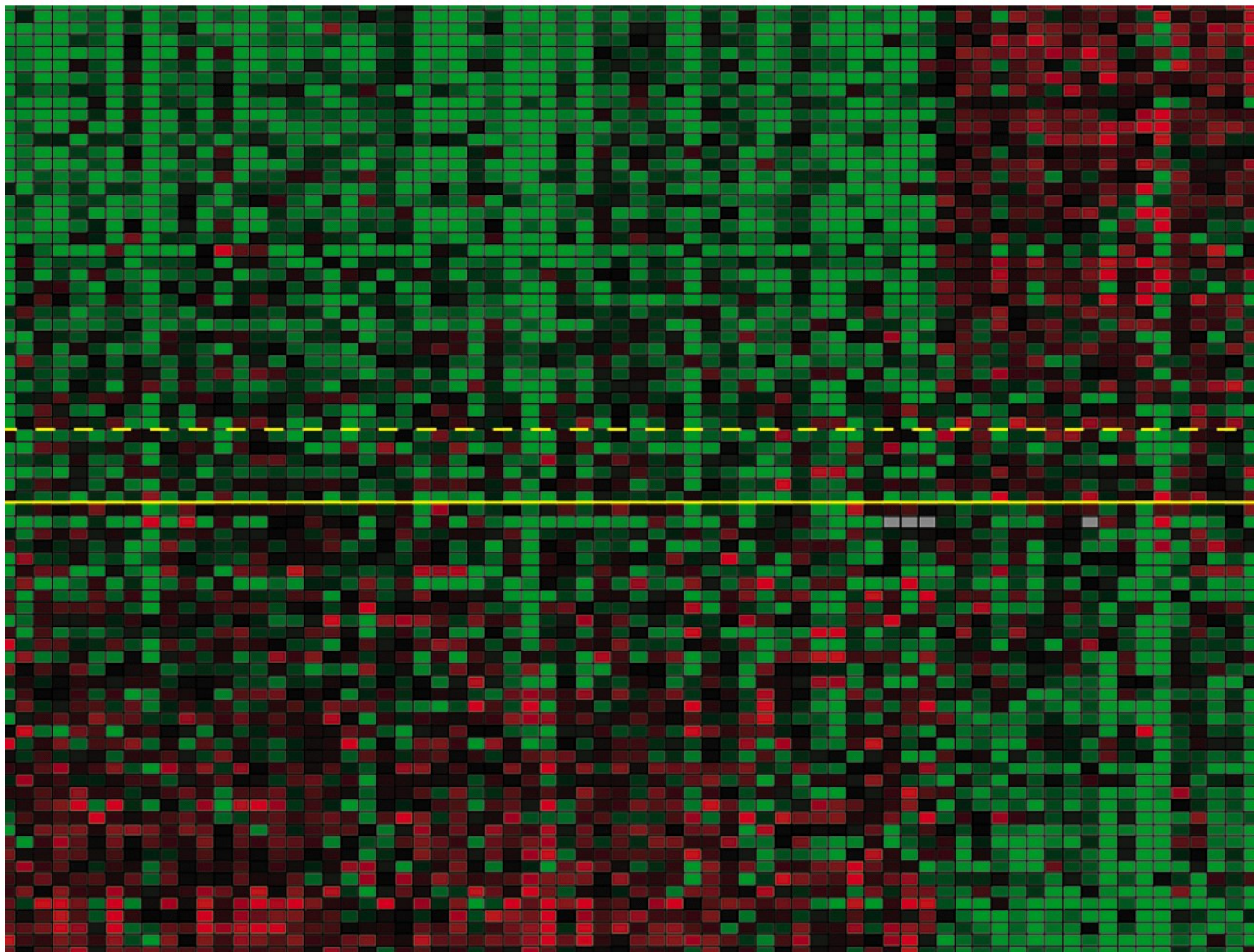
Ex. of one tissue

Microarrays used to compare levels of gene expression in different types of cells

Ex.



Yellow fluorescence- equal expression of the gene in cells A and B
Red fluorescence- more expression in cells A
Green fluorescence- more expression in cells B

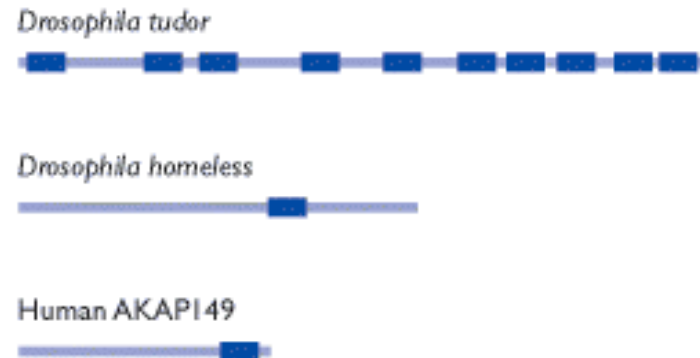


Comparative genomics

Compares the gene
content, function and organization of
genomes of **different organisms**

The tudor domain

Structure of the *Drosophila* tudor protein, which contains ten copies of the tudor domain



The domain is also found in a second *Drosophila* protein, *homeless*, involved in RNA transport during oogenesis and in the human A-kinase anchor protein (AKAP149), which plays a role in RNA metabolism.

The proteins have dissimilar structures other than the presence of the tudor domains.

The activity of each protein involves RNA in one way or another

Examples of human disease genes that have homologs in *Saccharomyces cerevisiae*

Human disease gene	Yeast homolog	Function of the yeast gene
Amyotrophic lateral sclerosis	<i>SOD1</i>	Protein against superoxide (O_2^-)
Ataxia telangiectasia	<i>TEL1</i>	Codes for a protein kinase
Colon cancer	<i>MSH2</i> , <i>MLH1</i>	DNA repair
Cystic fibrosis	<i>YCF1</i>	Metal resistance
Myotonic dystrophy	<i>YPK1</i>	Codes for a protein kinase
Type 1 neurofibromatosis	<i>IRA2</i>	Codes for a regulatory protein
Bloom's syndrome, Werner's syndrome	<i>SGS1</i>	DNA helicase
Wilson's disease	<i>CCC2</i>	Copper transport?

Table 19.2

Characteristics of some completely sequenced representative prokaryotic genomes

Species	Size (Millions of Base Pairs)	Number of Predicted Genes	G + C (%)
Archaea			
<i>Archaeoglobus fulgidus</i>	2.18	2407	49
<i>Methanobacterium thermoautotrophicum</i>	1.75	1869	50
<i>Methanococcus jannaschii</i>	1.66	1715	32
<i>Thermoplasma acidophilum</i>	1.56	1478	46
Eubacteria			
<i>Bacillus subtilis</i>	4.21	4100	44
<i>Bordetella parapertussis</i>	4.75	*	69
<i>Buchnera</i> species	0.64	564	27
<i>Campylobacter jejuni</i>	1.64	1654	31
<i>Escherichia coli</i>	4.64	4289	51
<i>Haemophilus influenzae</i>	1.83	1709	39
<i>Mesorhizobium loti</i>	7.04	6752	63
<i>Mycobacterium tuberculosis</i>	4.41	3918	66
<i>Mycoplasma genitalium</i>	0.58	480	32
<i>Staphylococcus aureus</i>	2.88	2697	33
<i>Treponema pallidum</i>	1.14	1031	53
<i>Ureaplasma urealyticum</i>	0.75	611	26
<i>Vibrio cholerae</i>	4.03	3828	48

Source: Data from the Genome Atlas of the Center for Biological Sequence Analysis,
<http://www.cbs.dtu.dk/services/GenomeAtlas/>

* Data not available.

Density of genes is rather constant across all species; bacteria with larger genomes have more genes