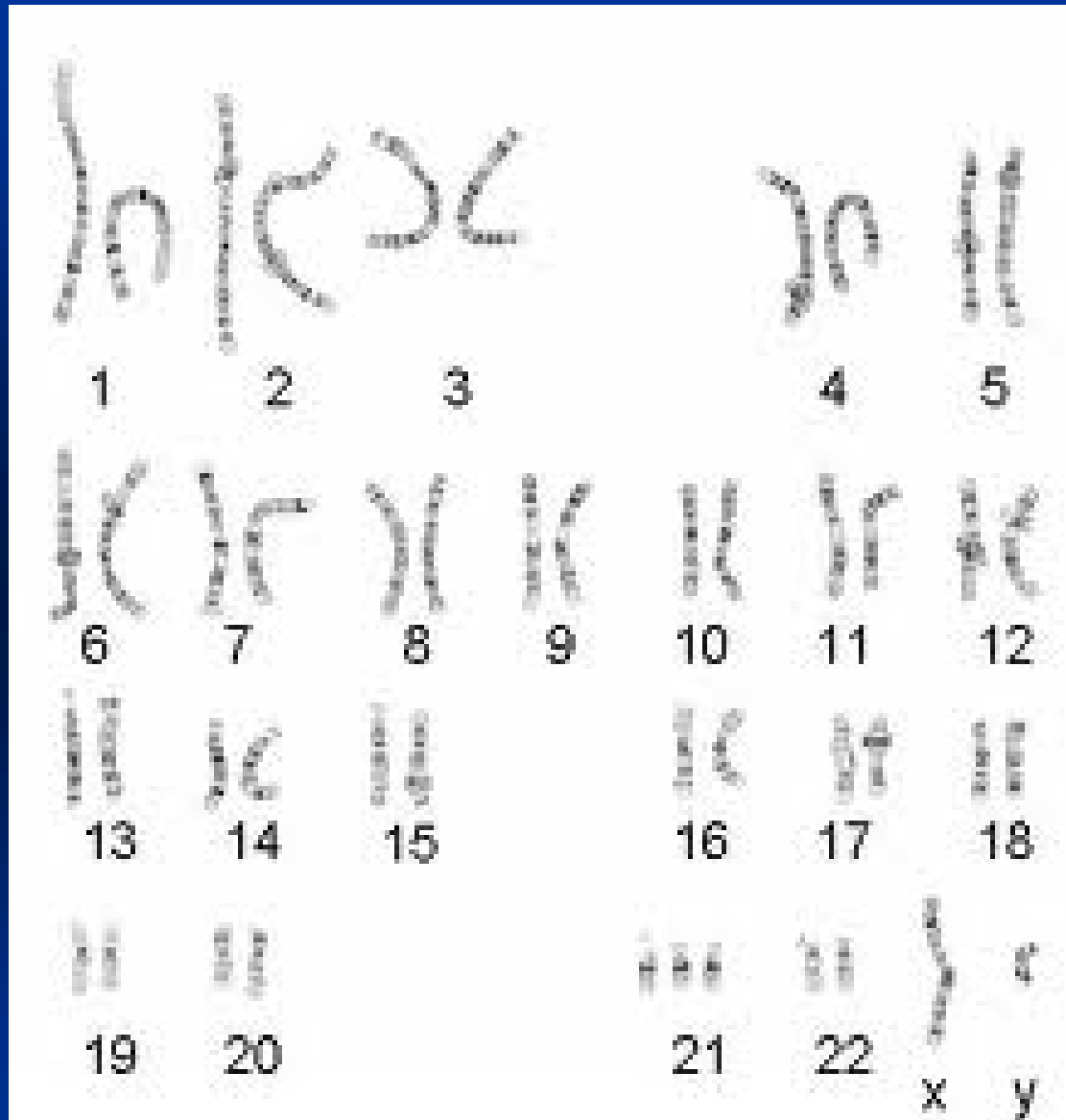


PROYECTO GENOMA HUMANO

*Medicina Molecular 2011
Maestría en Biología Molecular
Médica- UBA*

GENOMA HUMANO



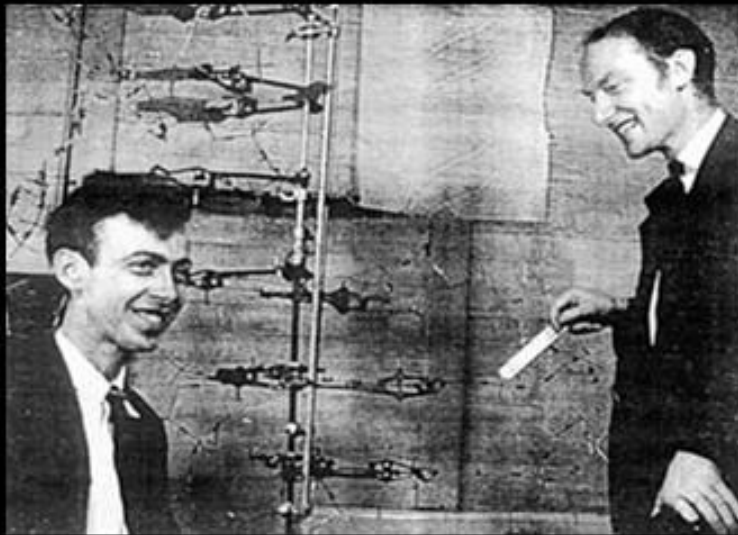
Medicina Molecular 2011

Maestría en Biología Molecular Médica- UBA

Some landmarks on the way to the genome sequence

- 1940s Recognition that DNA is the hereditary material
- 1953 Double-helix structure described
- 1966 Genetic code cracked
- 1972 Recombinant DNA technology developed
- 1975-77 DNA sequencing technology developed

James Watson -nacido en Chicago, Illinois, el 6 de abril de 1928, poseedor de un doctorado en Zoología-, y **Francis Crick** - doctor en física, de nacionalidad británica, nacido en Northampton, Inglaterra en 1916 – trabajando juntos en un laboratorio de la Universidad de Cambridge, Inglaterra, en 1951, descubrieron la estructura secundaria del ADN. Por ello recibieron el premio Nobel de medicina en 1962.

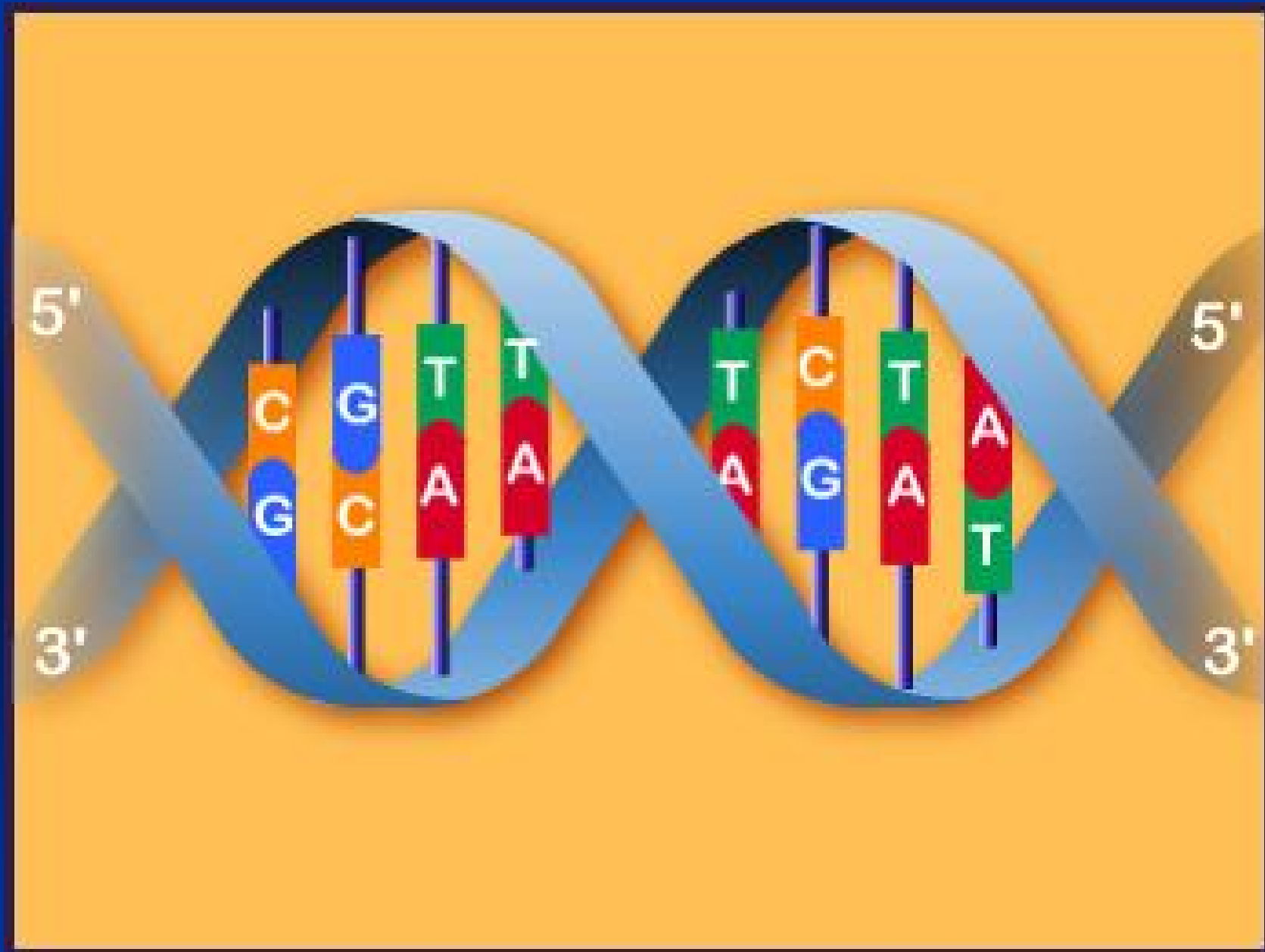


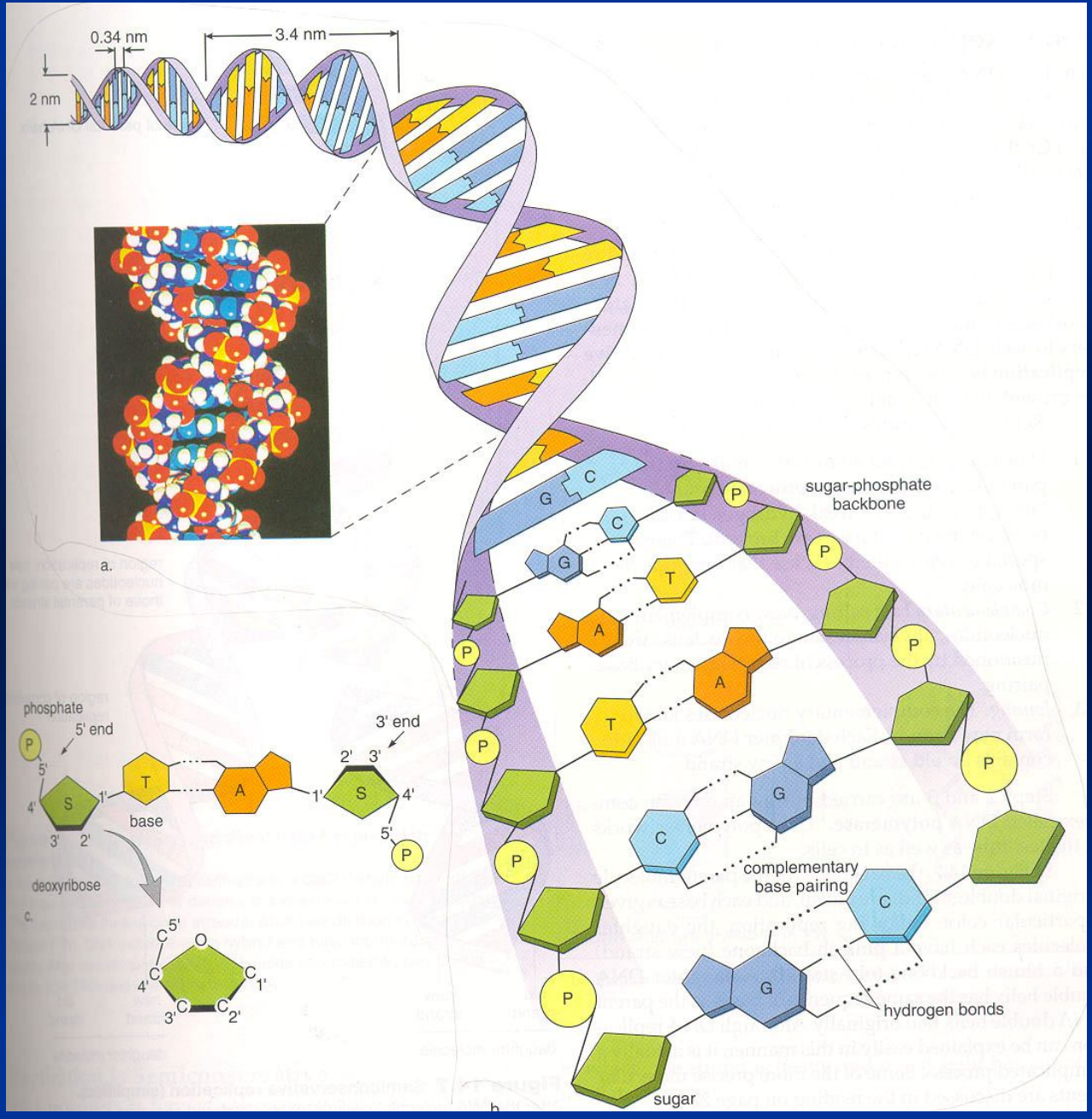
J. Watson



F. Crick

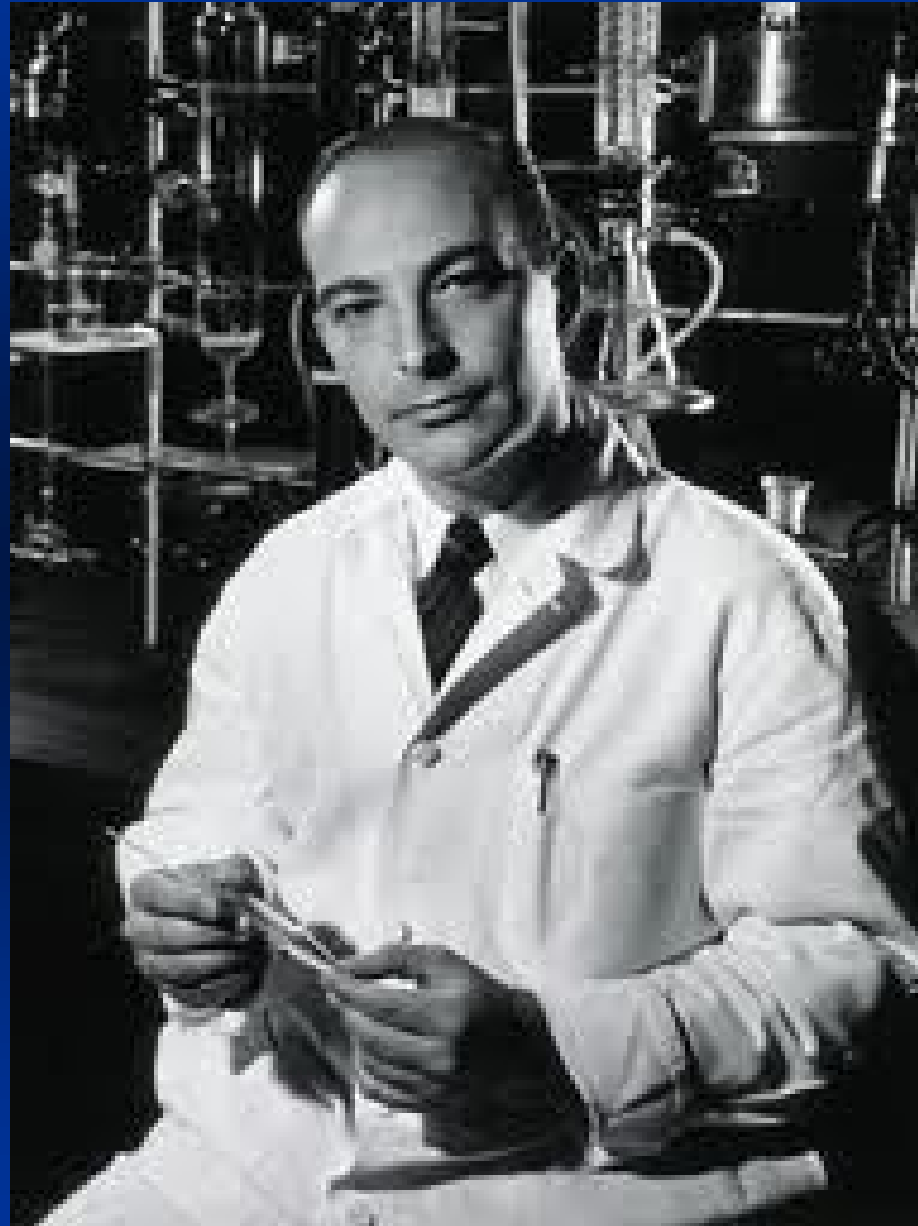




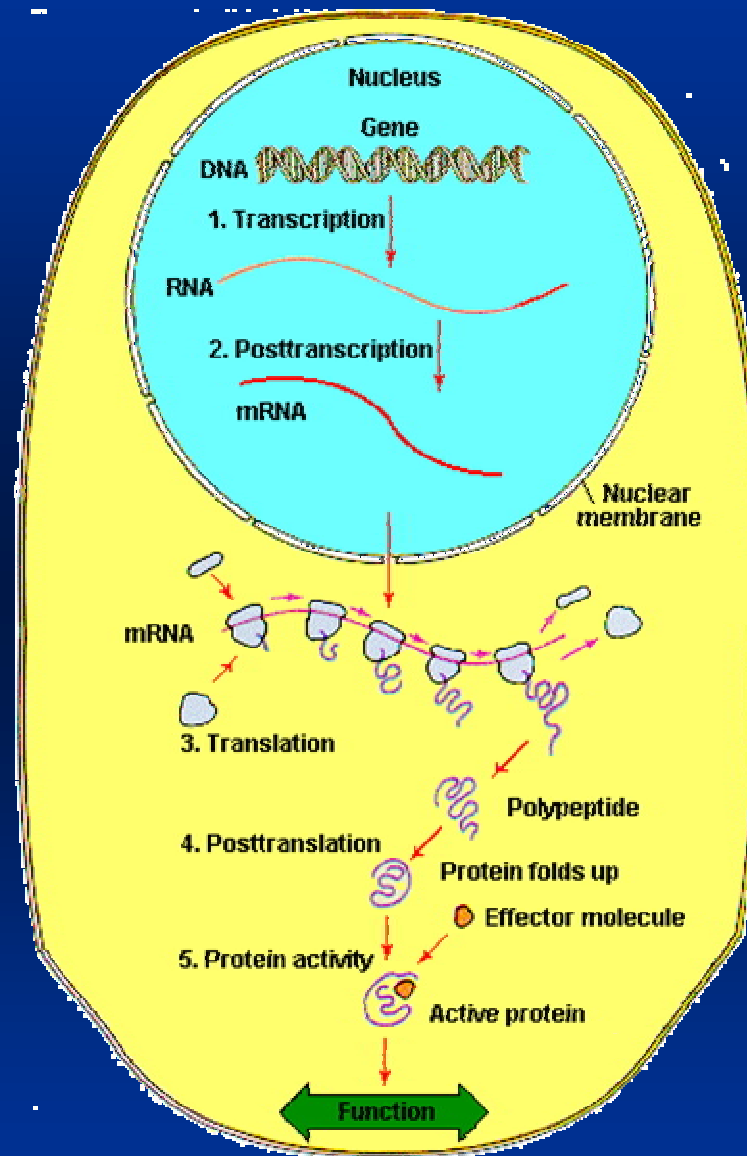


- **SI DESPLEGAMOS TODO EL DNA CONTENIDO EN LOS 46 CROMOSOMAS EL DNA MEDIRIA 2 METROS DE LONGITUD.**

Arthur Kornberg, Premio Nobel de Medicina 1959



TRANSCRIPCION



Severo Ochoa- Premio Nobel de Medicina 1959



Protein Synthesis - Transcription

- Each gene codes for a protein
- DNA sense strand acts as template and is 'transcribed' into messenger RNA (mirror image of the DNA but Uracil instead of Thymine)

DNA

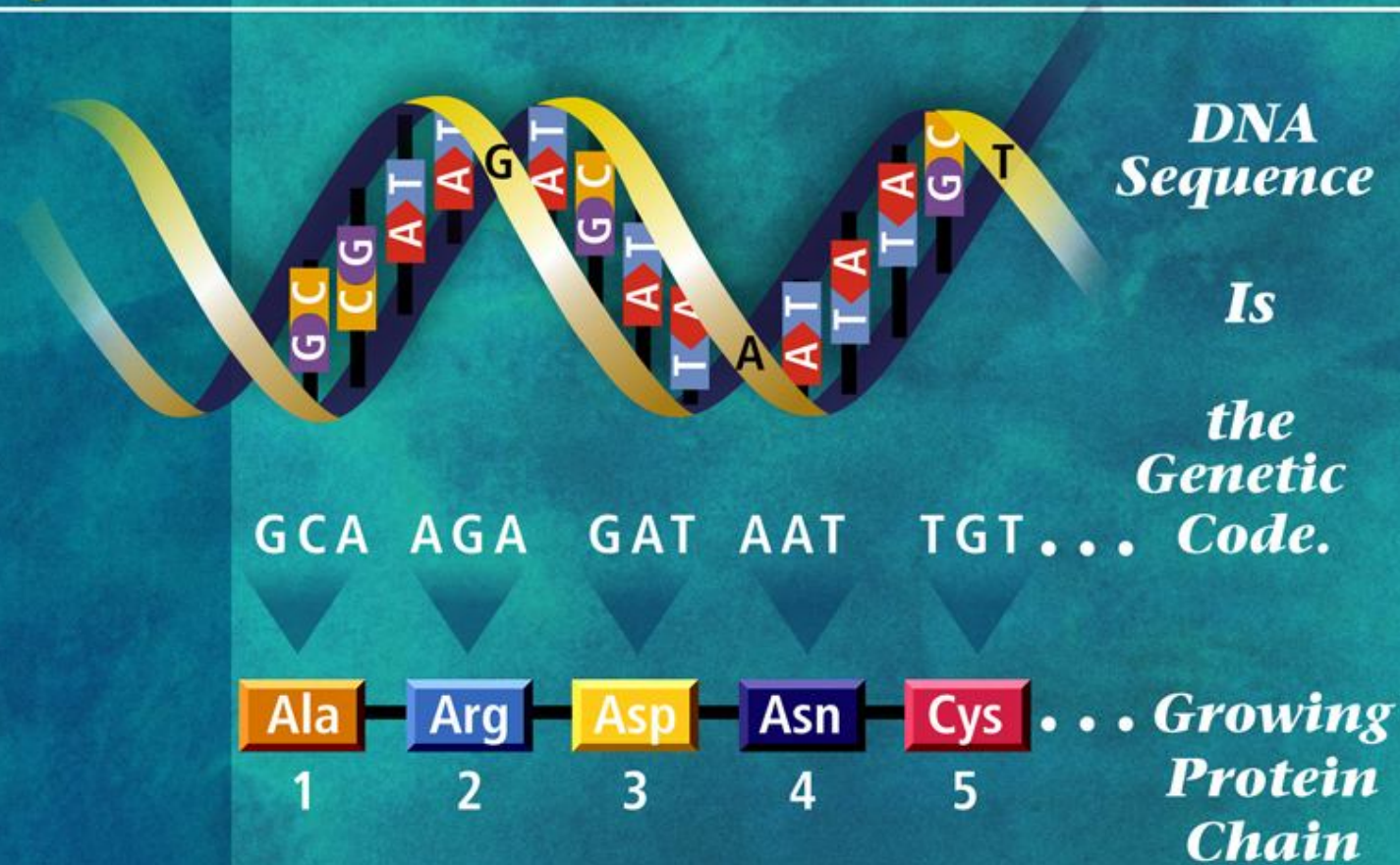
ATCGG

mRNA

UAGCC

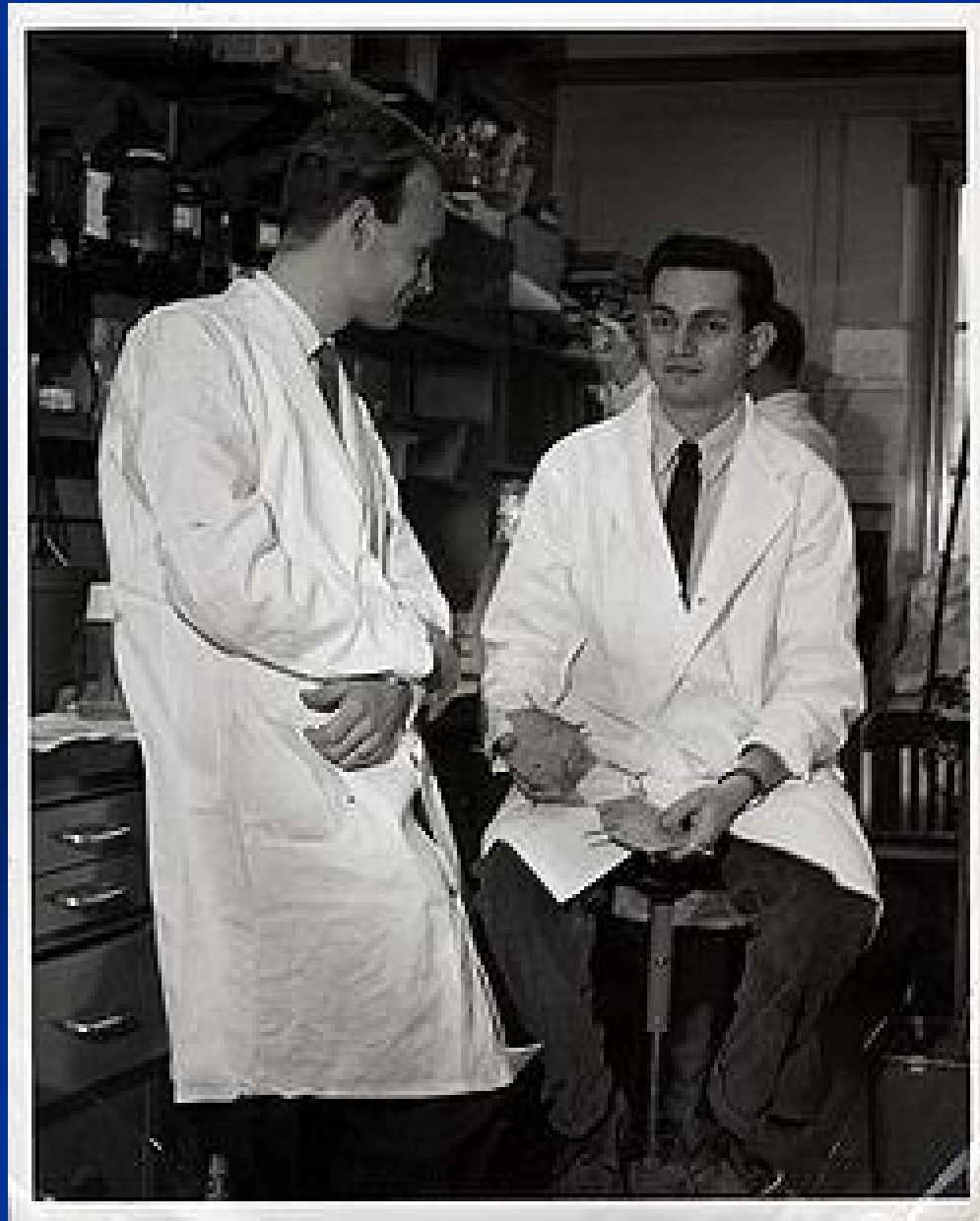
DILUCIDACION DEL CÓDIGO GENÉTICO

DNA Genetic Code Dictates Amino Acid Identity and Order



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

Marshall Nirenberg. Premio Nobel de Medicina 1968

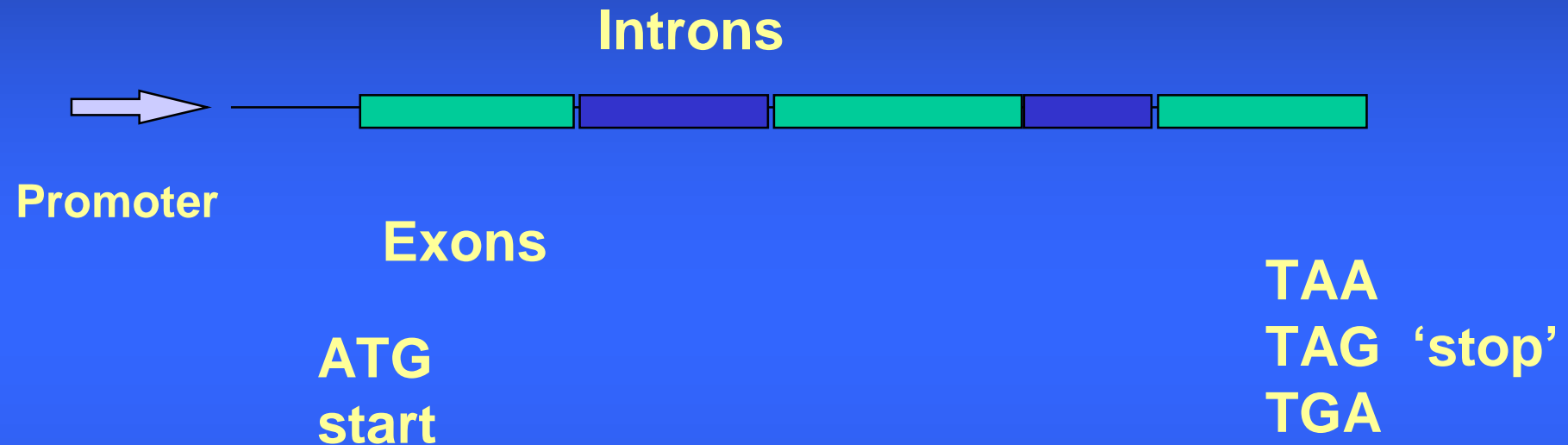


Severo Ochoa- Premio Nobel de Medicina 1959



EL DNA TIENE INTRONES Y EXONES

Gene Structure



Exon = coding sequence

Intron = intervening

sequence (non-coding)

SPLICING

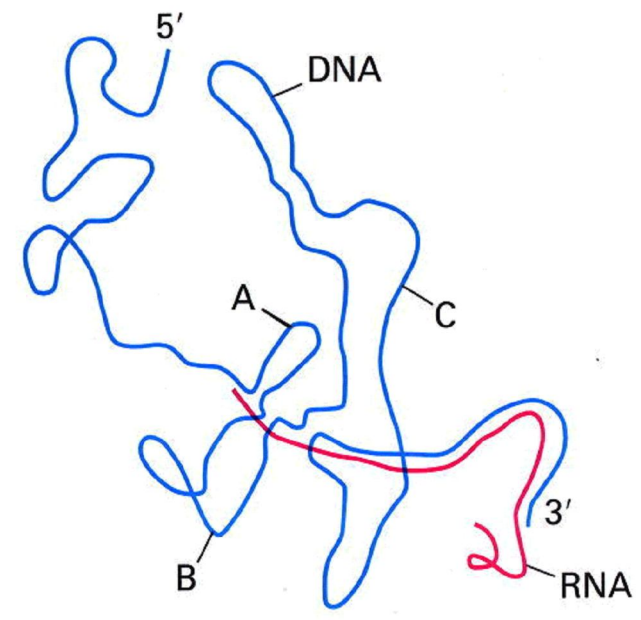
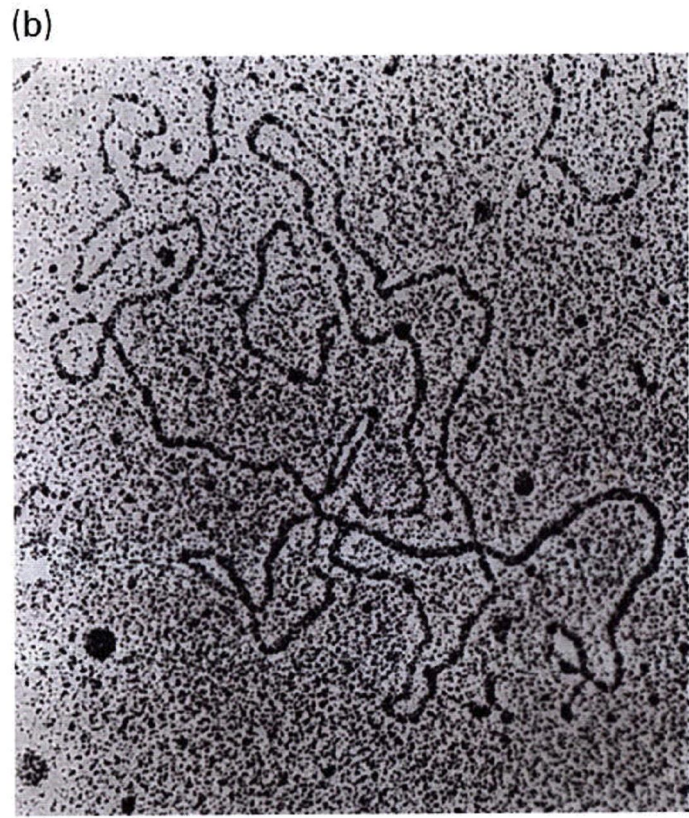
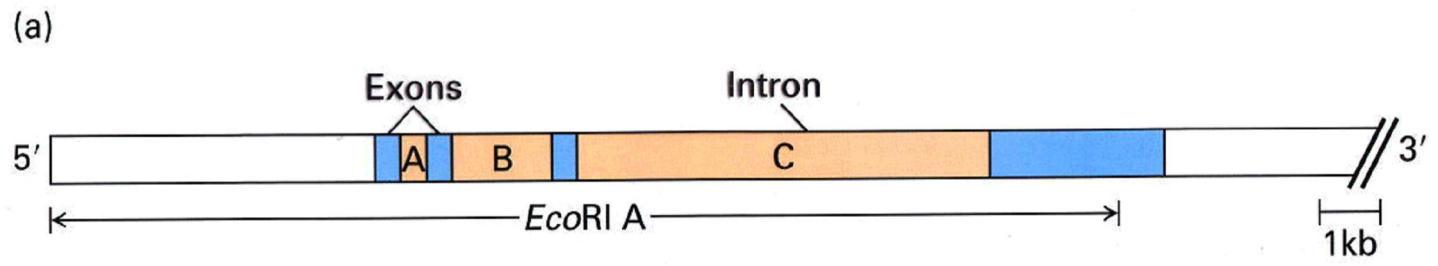


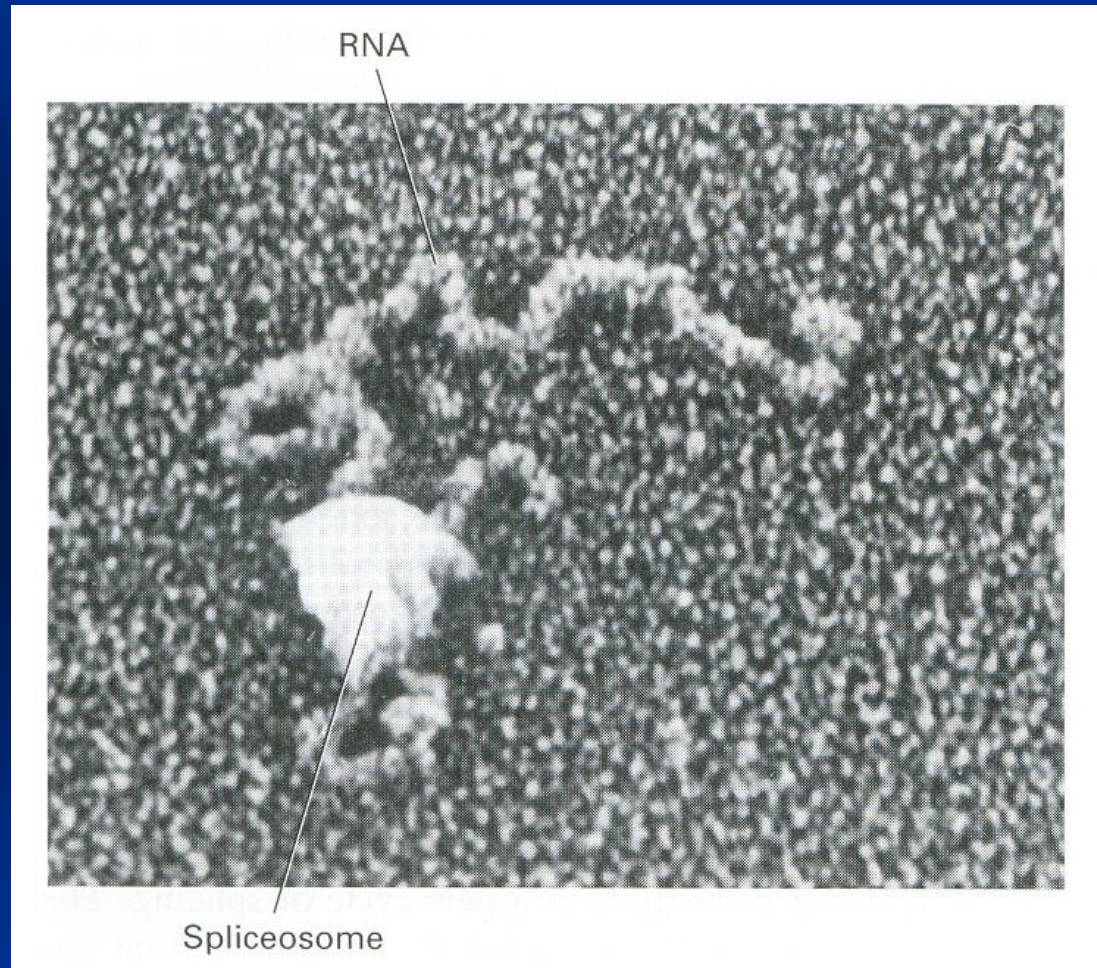
Figure 11-13
 Lodish et al. MOLECULAR CELL BIOLOGY, Fourth Edition
 Copyright © by W. H. Freeman and Company

SPLICING ALTERNATIVO

J. Darnell. Splicing alternativo

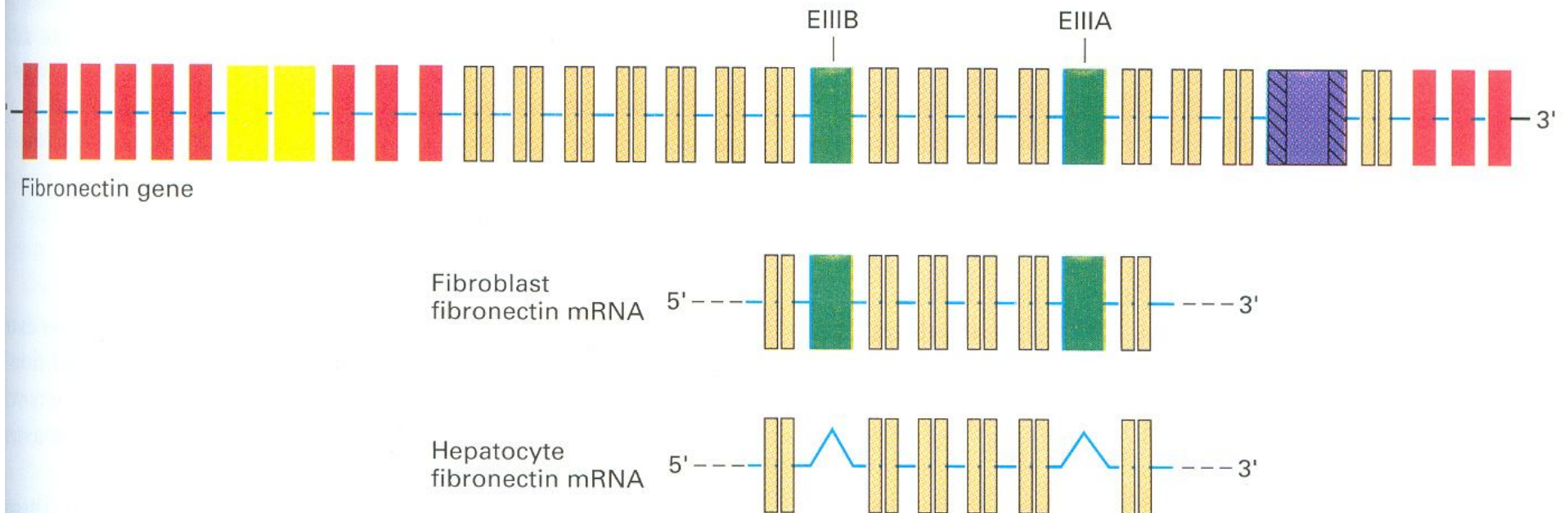


SPLICEOSOMA



Spliceosoma de HeLa + pre-mRNA de beta globina

SPLICING DIFERENCIAL EN DISTINTOS TEJIDOS



Some landmarks on the way to the genome sequence

- 1940s Recognition that DNA is the hereditary material
- 1953 Double-helix structure described
- 1966 Genetic code cracked
- 1972 Recombinant DNA technology developed
- 1975-77 DNA sequencing technology developed

TRADUCCION

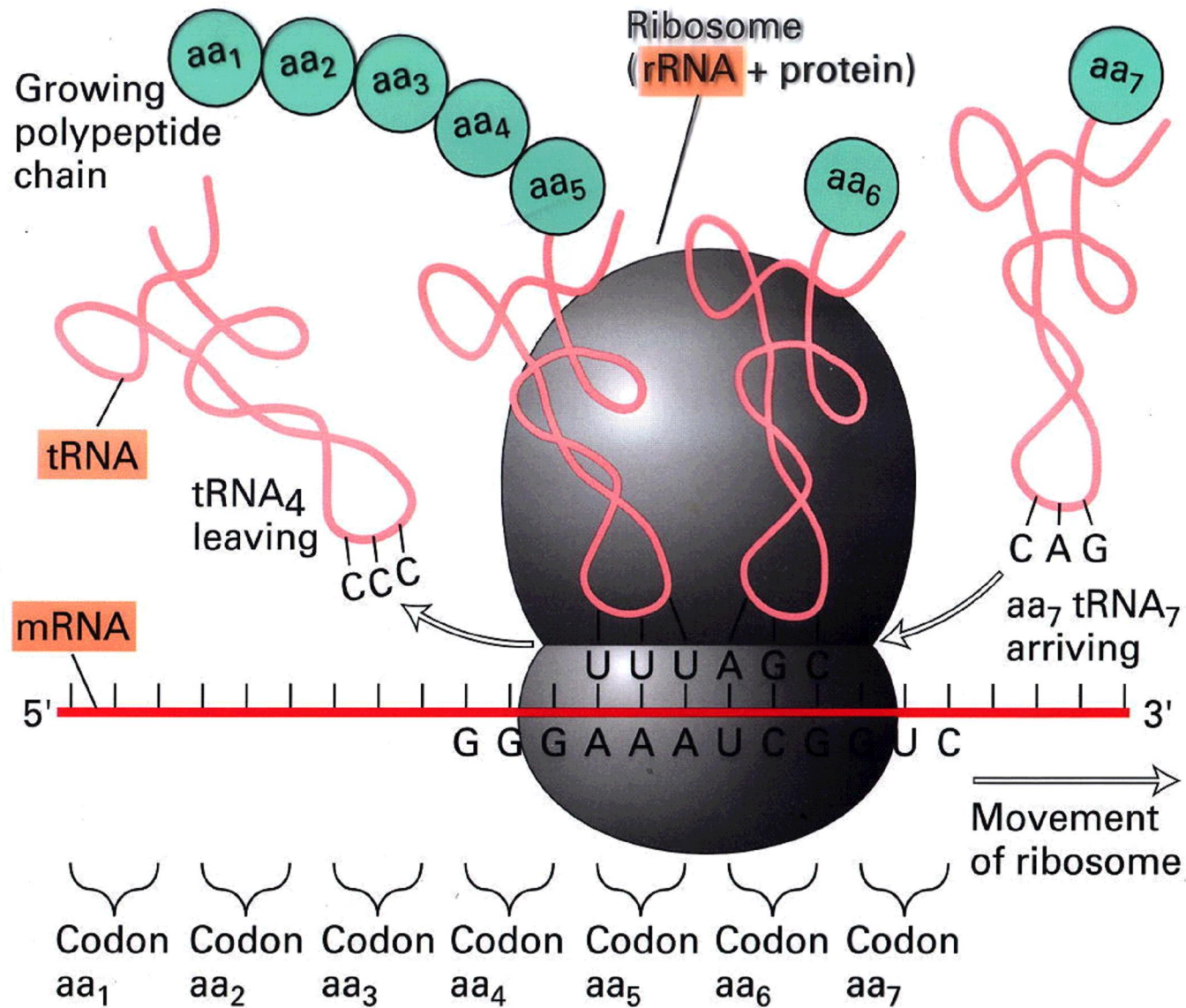


Figure 4-20
 Lodish et al. MOLECULAR CELL BIOLOGY, Fourth Edition
 Copyright © by W. H. Freeman and Company

Some landmarks on the way to the genome sequence

- 1940s Recognition that DNA is the hereditary material
- 1953 Double-helix structure described
- 1966 Genetic code cracked
- 1972 Recombinant DNA technology developed
- 1975-77 DNA sequencing technology developed

ENZIMAS DE RESTRICCIÓN

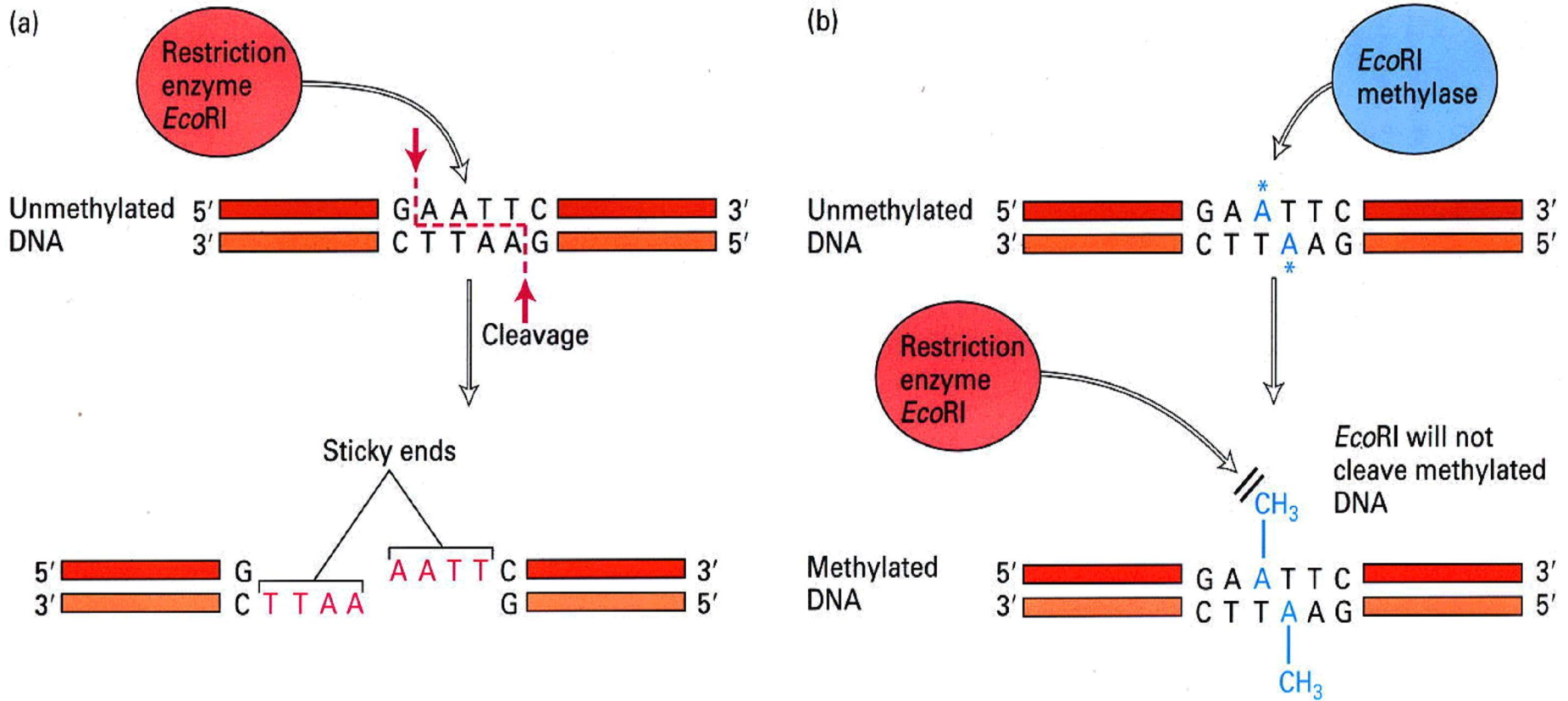
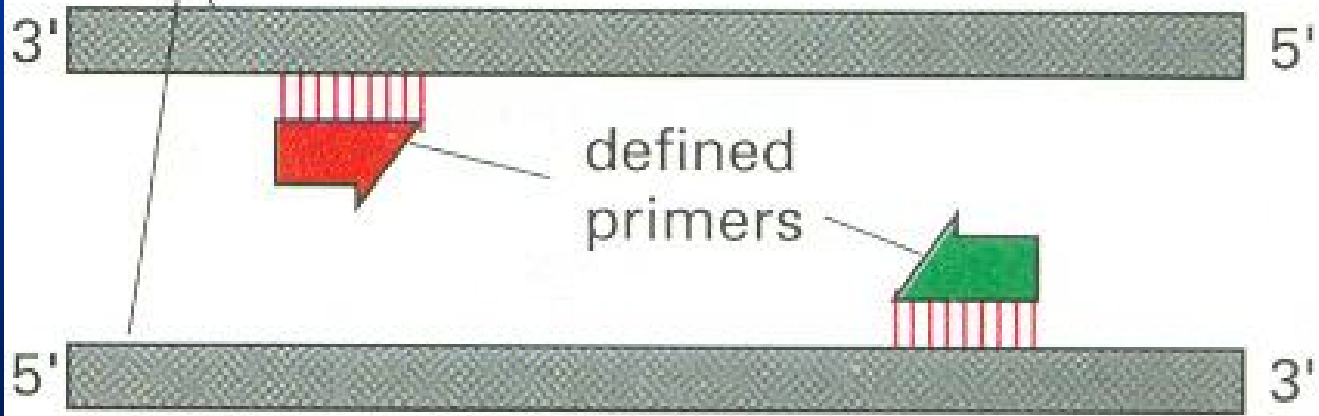


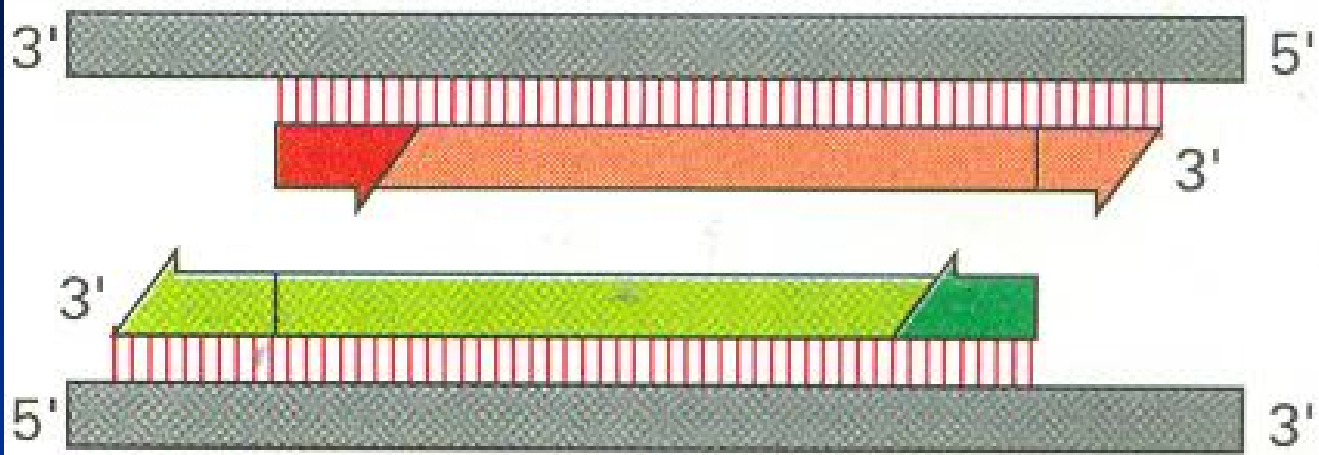
Figure 7-5
 Lodish et al. MOLECULAR CELL BIOLOGY, Fourth Edition
 Copyright © by W. H. Freeman and Company

POLYMERASE CHAIN REACTION

separated complementary DNA strands

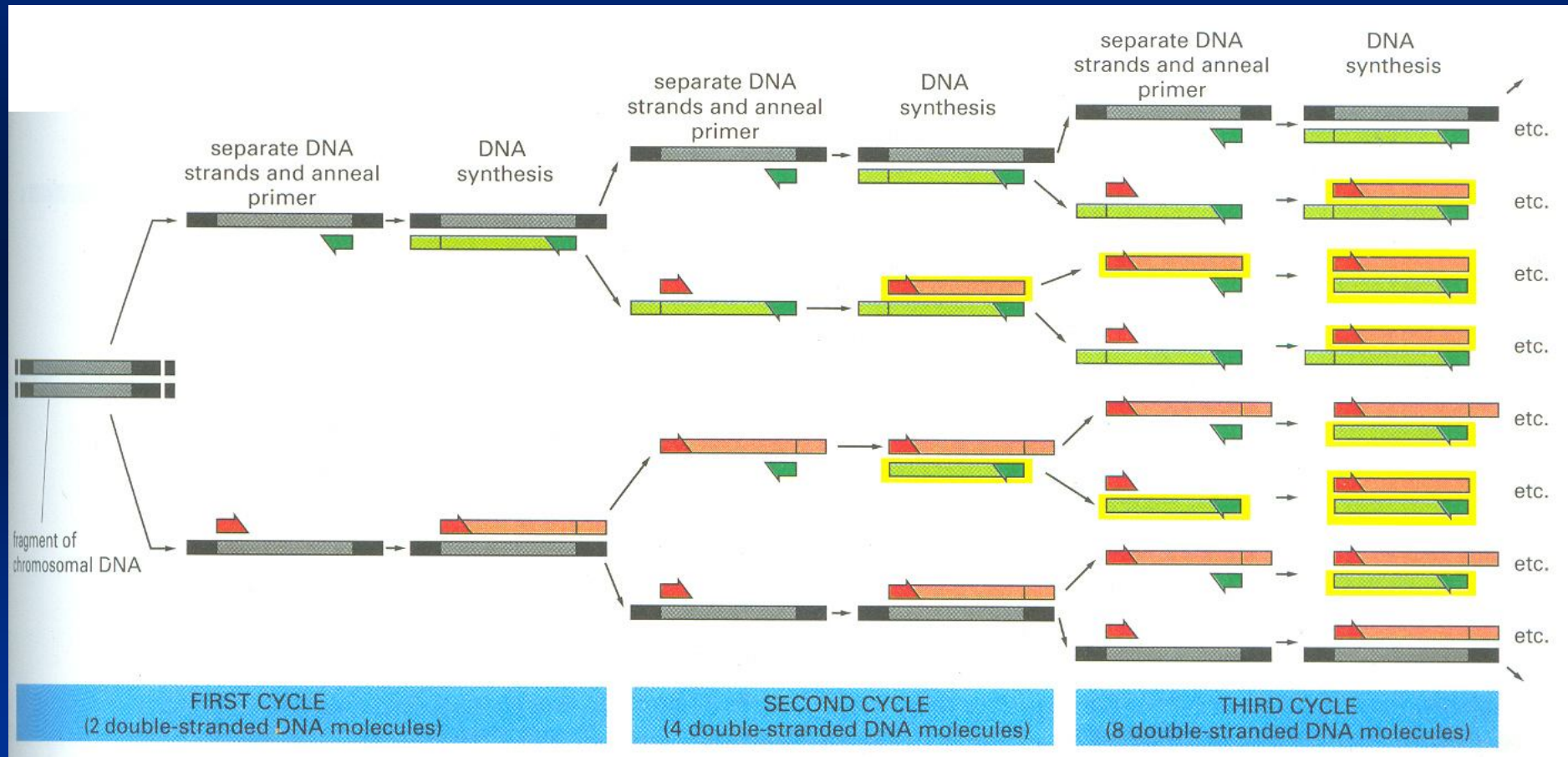


DNA synthesis



X nucleotide pairs

PCR



PLASMIDOS

Bacterial DNA

Plasmids



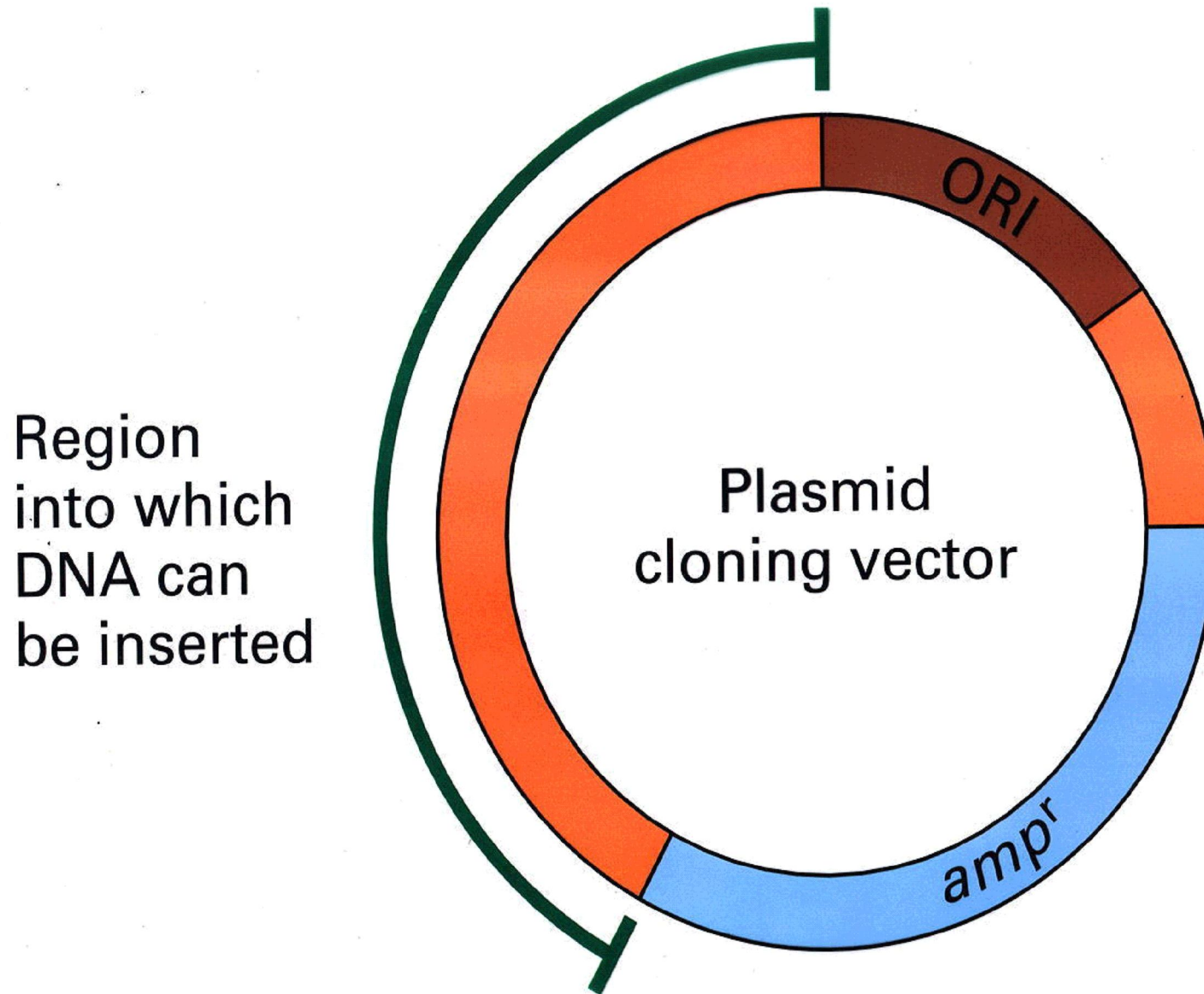
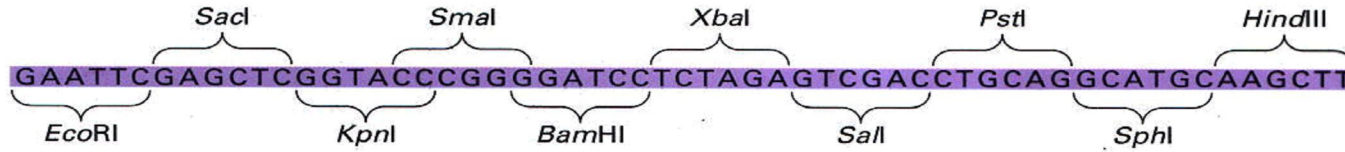
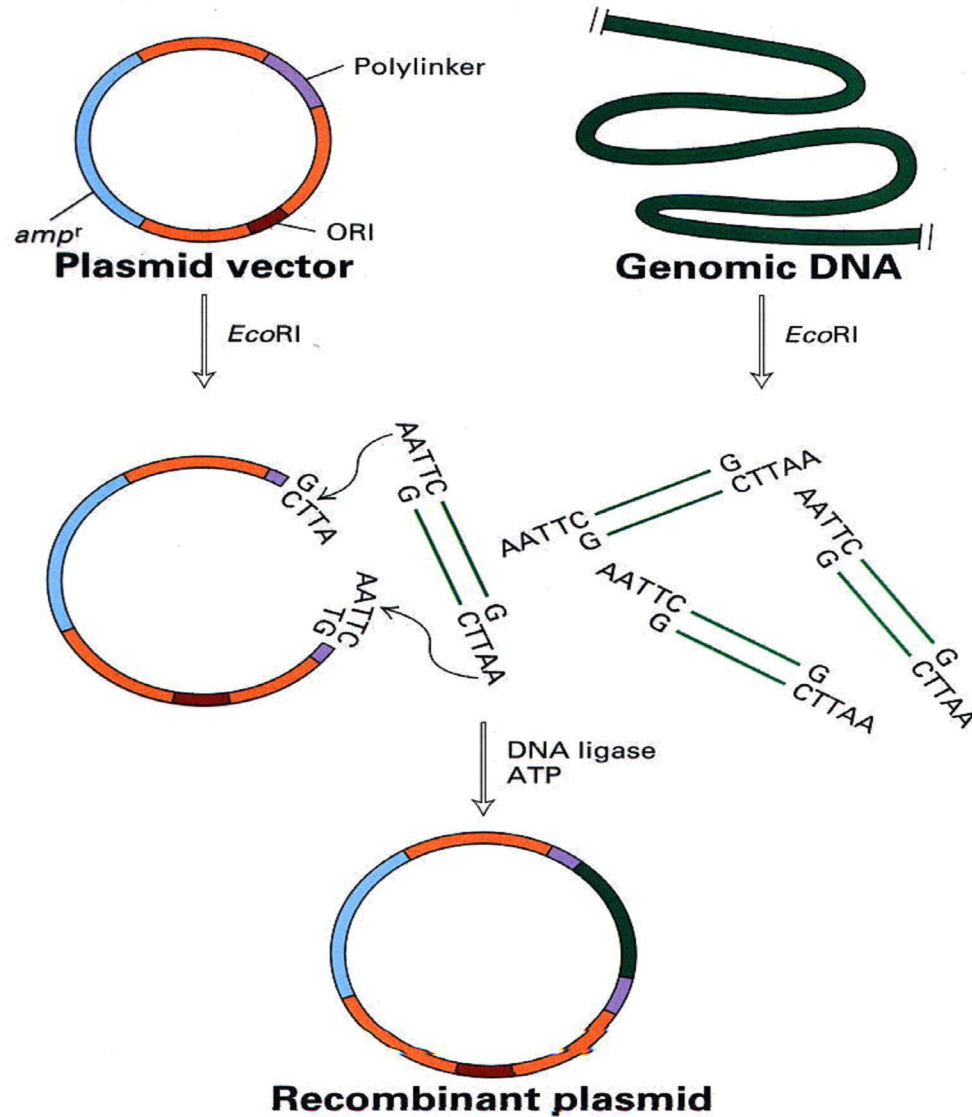


Figure 7-1
Lodish et al. MOLECULAR CELL BIOLOGY, Fourth Edition
Copyright © by W. H. Freeman and Company

(a) Sequence of polylinker



(b) Insertion of *EcoRI* restriction fragments



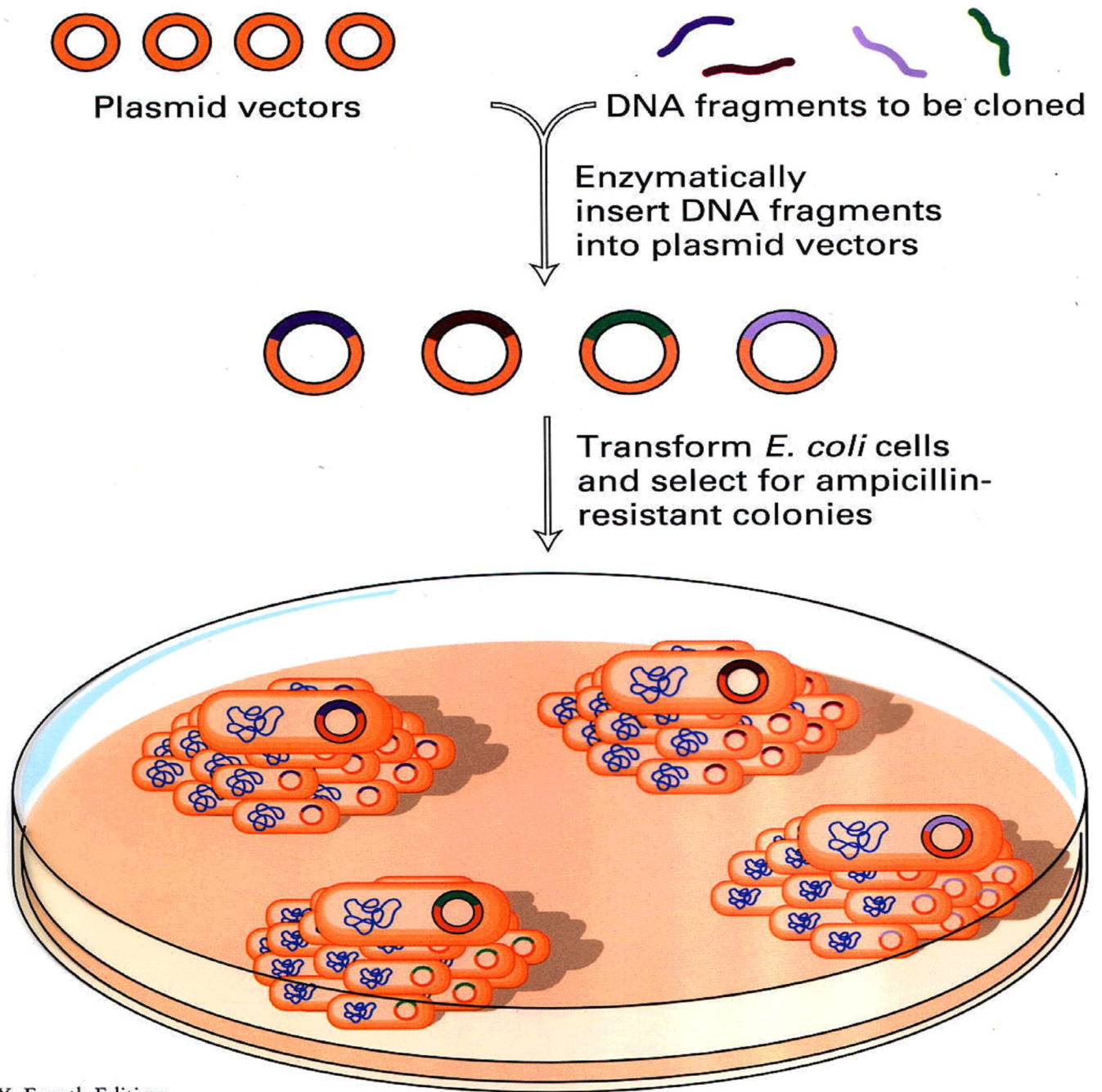


Figure 7-4
Lodish et al. MOLECULAR CELL BIOLOGY, Fourth Edition
Copyright © by W. H. Freeman and Company

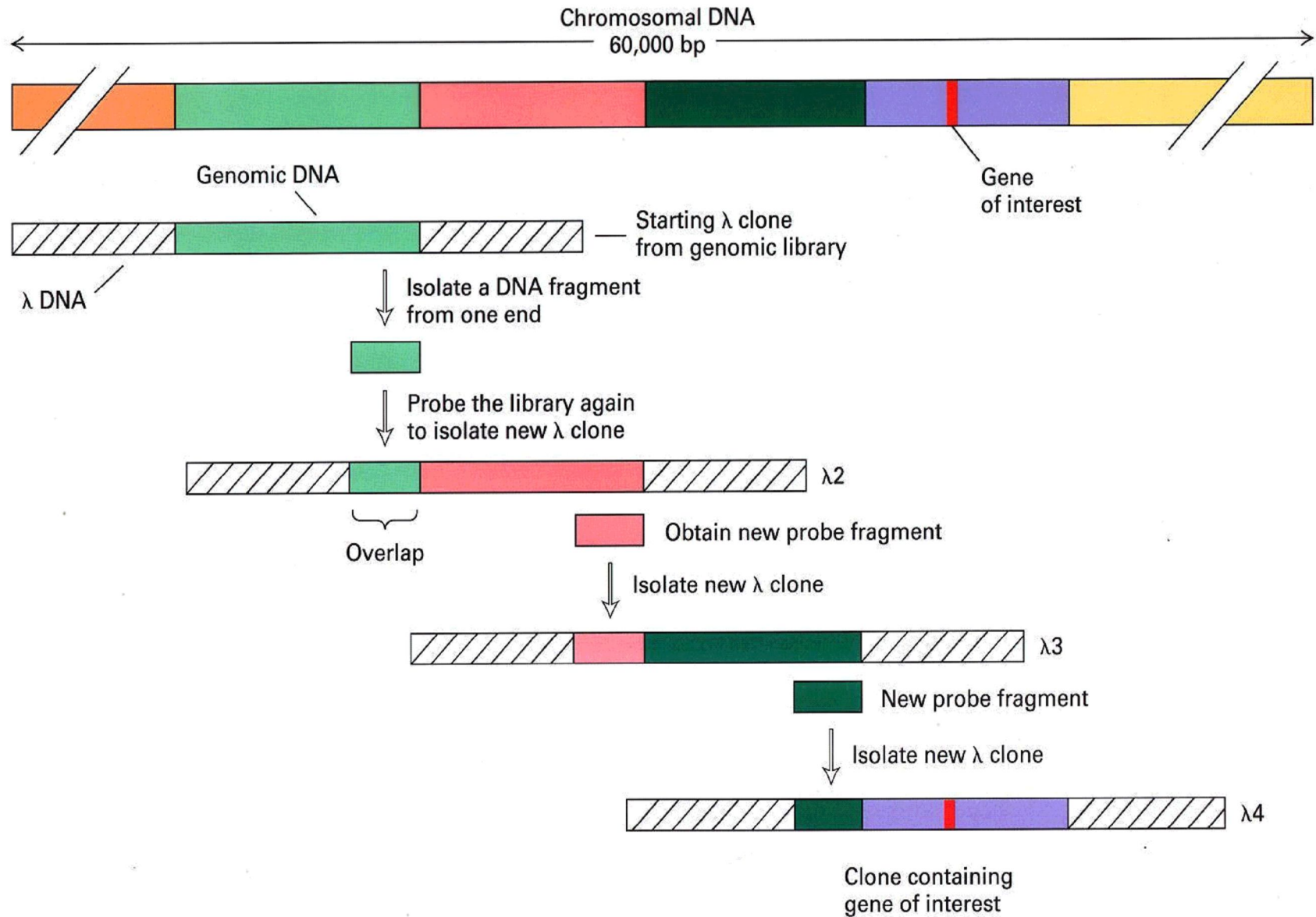
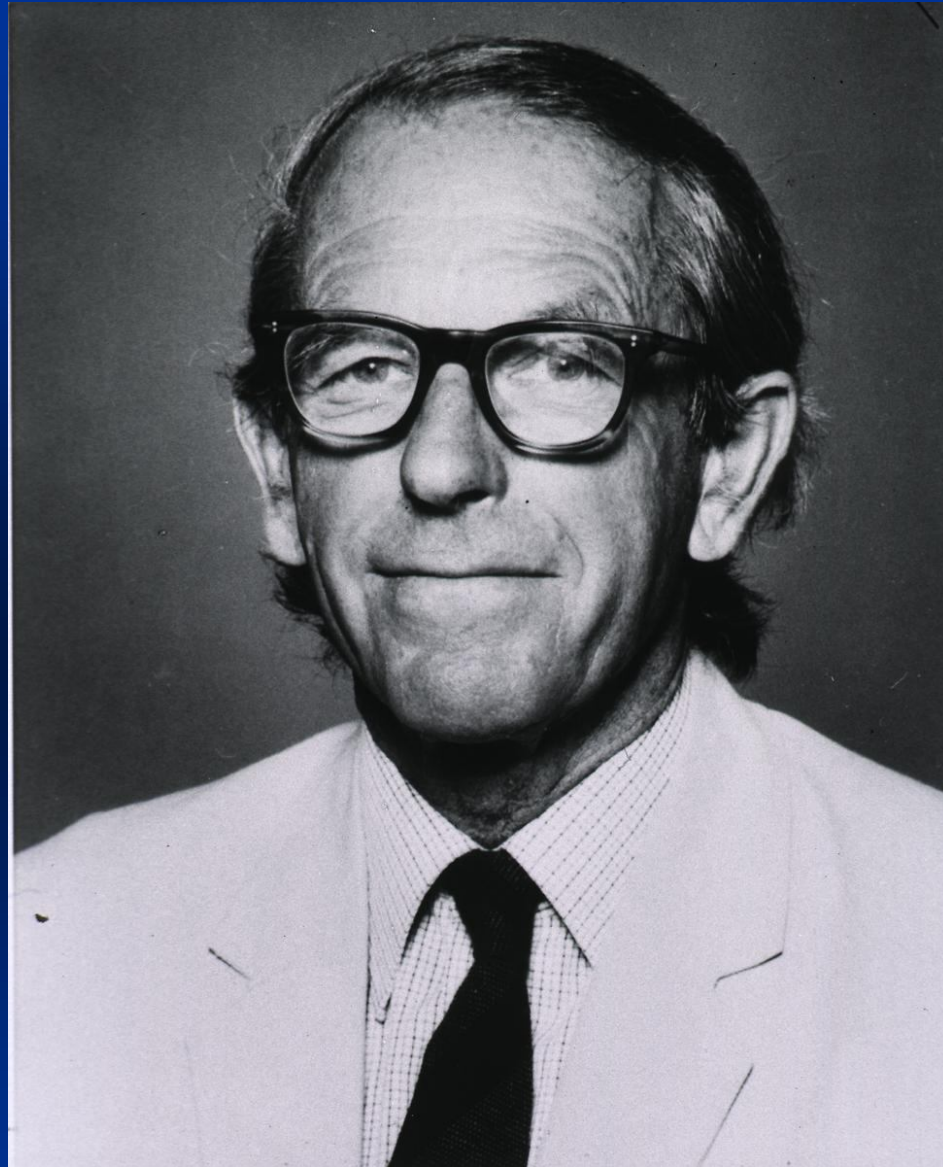


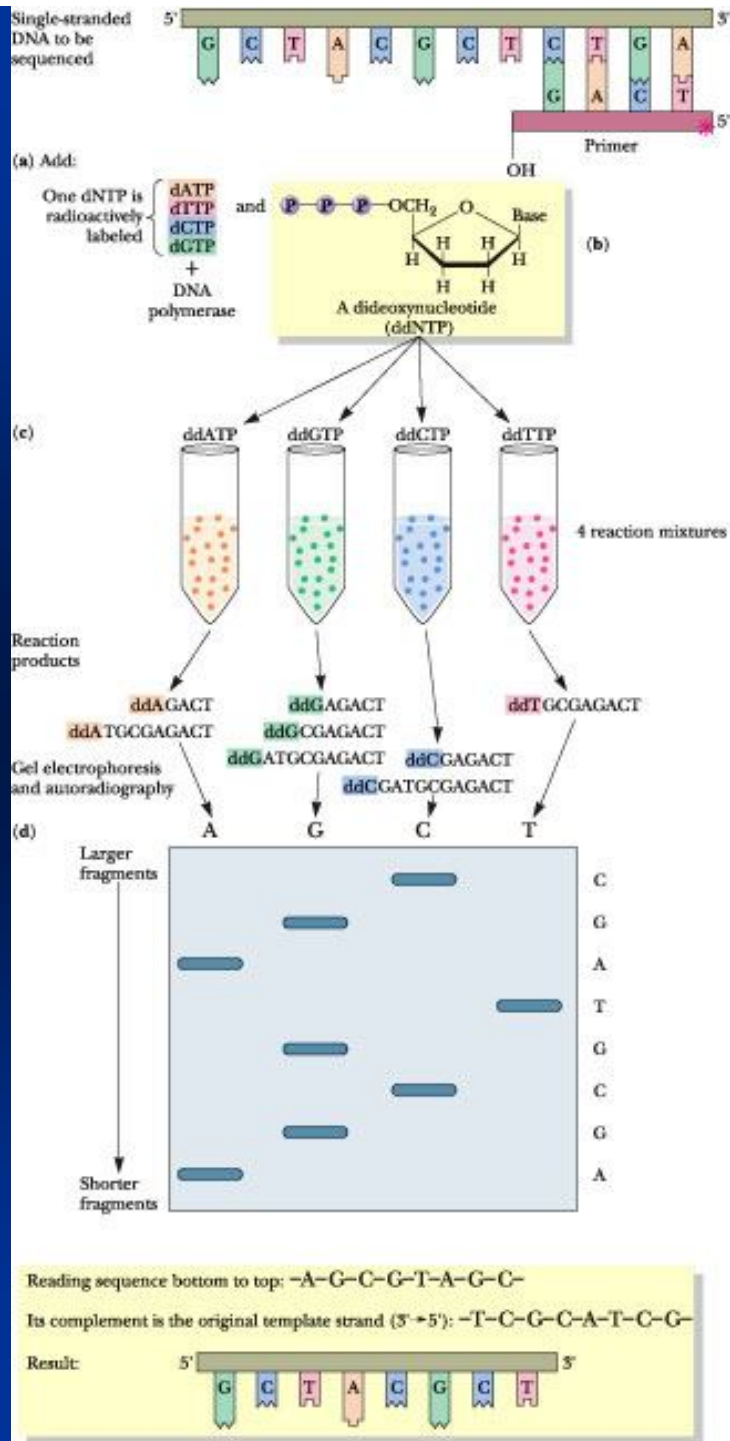
Figure 8-24
Lodish et al. MOLECULAR CELL BIOLOGY, Fourth Edition
Copyright © by W. H. Freeman and Company

Some landmarks on the way to the genome sequence

- 1940s Recognition that DNA is the hereditary material
- 1953 Double-helix structure described
- 1966 Genetic code cracked
- 1972 Recombinant DNA technology developed
- 1975-77 DNA sequencing technology developed



Frederick Sanger, Nobel Prize 1958 y 1980



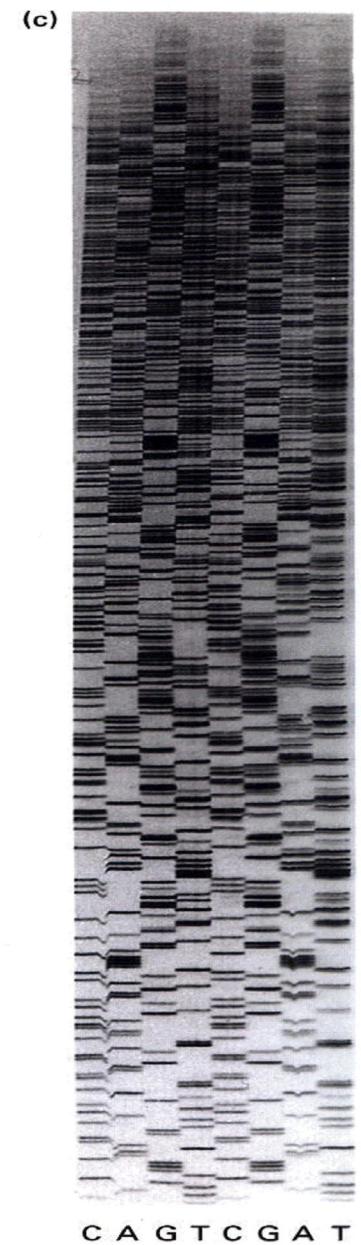
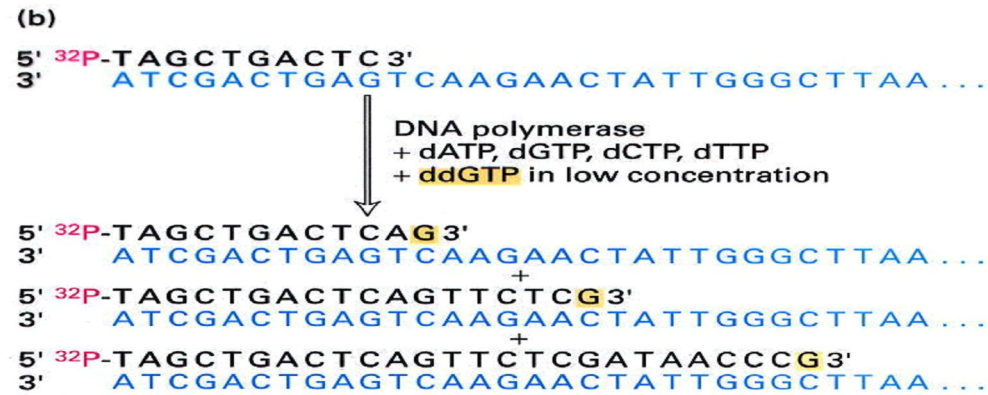
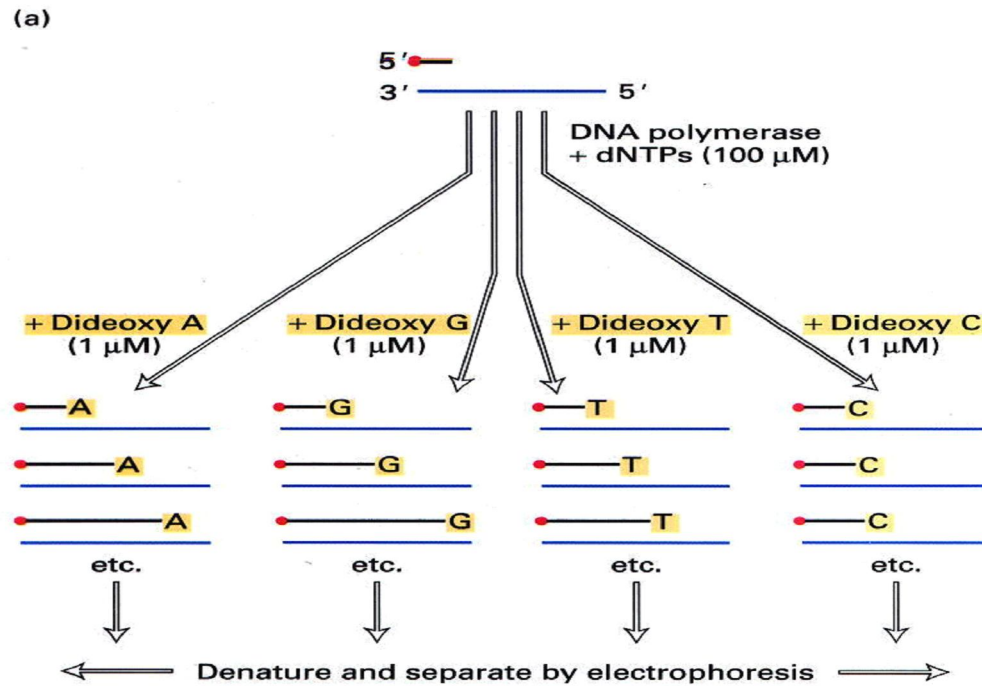
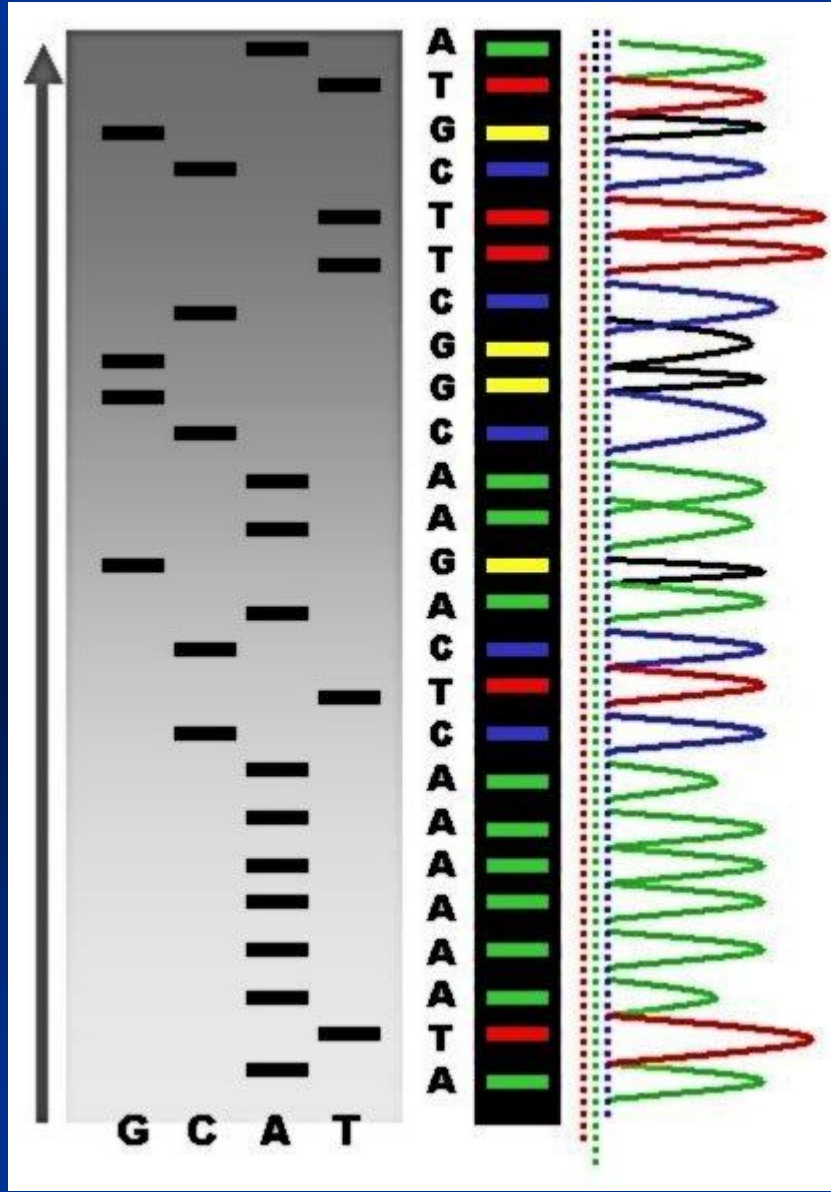


Figure 7-29
Lodish et al. MOLECULAR CELL BIOLOGY, Fourth Edition
Copyright © by W. H. Freeman and Company



Landmarks

- 1983 First human disease gene mapped (Huntington's disease). Gene finally isolated in 1993
- 1990 Human Genome Project launched
- 1995 Human Physical map completed
- 1996 Sequencing begins
- 1999 Full-scale sequencing begins
- 2001 'Working draft' produced
- 2003 Final sequence published





Meeting en Alta, Utah, 1984

Human Genome Project Goals

- identify all the genes in human DNA,
- determine the sequences of the 3 billion chemical base pairs that make up human DNA
- store this information in databases,
- improve tools for data analysis,
- transfer related technologies to the private sector, and
- address the ethical, legal, and social issues (ELSI) that may arise from the project.

Why?

- virtually all disease states arise through complex interplay between genes & environment
- almost all progress to date is on single-gene disorders
- Cancers, heart disease, hypertension etc, all have genetic component,
 - many genes involved
 - hard to get at by traditional techniques

Proyecto Genoma Humano

Cinco instancias de análisis del GH y sus productos:

- **GENOMA HUMANO:** análisis de la secuencia completa
- **TRANSCRIPTOMA HUMANO:** análisis de las secuencias transcritas en mRNA (genes)
- **PROTEOMA HUMANO:** análisis de las proteínas presentes en las células humanas
- **GLICOMA HUMANO:** análisis de las glicoproteínas presentes en las células humanas
- **LIPIDOMA HUMANO:** análisis de los lípidos presentes en las células humanas

Proyecto Genoma Humano

Tabla 1. Tamaño comparativo de los genomas de diversas especies

Especie	Genoma Haploide (pares de Bases, bp)	Genes (n)
Homo sapiens	3.000.000.000.	30.000 - 35.000
Mus musculus	3.000.000.000.	30.000-35.000
Rattus norvegicus	3.000.000.000.	30.000-35.000
Drosophila (mosca de la fruta)	165.000.000.	15.000 - 25.000
C. elegans	100.000.000.	19.000
Levadura y hongos	14.000.000.	8.355 - 8.947
E. Coli	4.670.000.	3.237
H.Influenzae	1.800.000.	
M. Genitalium	580.000.	

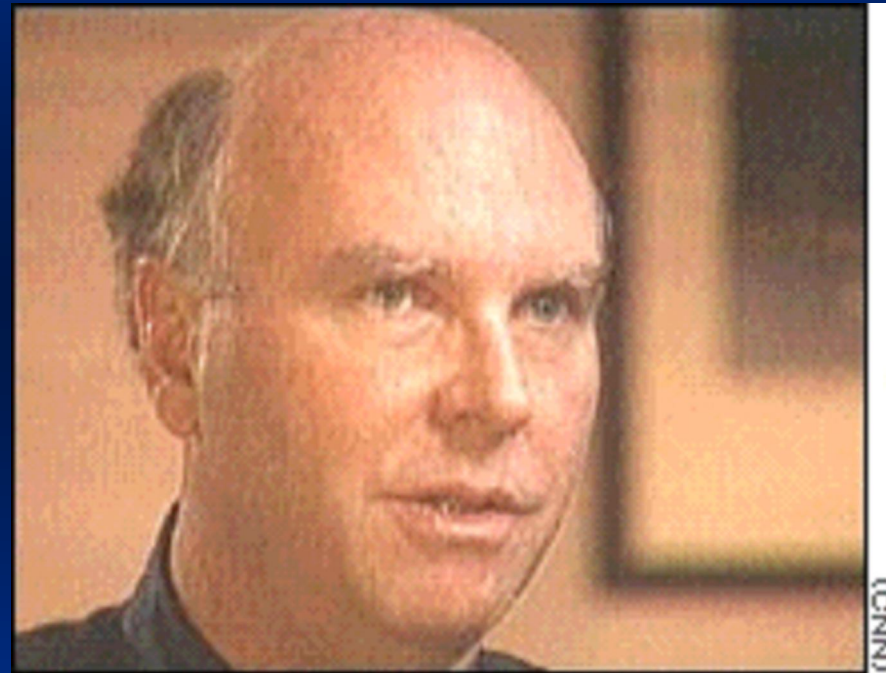


The Human Genome Race

Collins vs. Venter



Francis Collins



Craig Venter



2.000

Collins

- Francis Collins, a physician, is director of the National Human Genome Research Institute.
- His research laboratory was responsible for identifying the genes responsible for **Cystic Fibrosis, Neurofibromatosis, and Huntington's disease.**

Venter

Venter founded the nonprofit Institute for Genomic Research in 1992. Before that he was section chief and a laboratory chief at the National Institutes of Neurological Disorders and the National Institutes of Health. Celera Genomics is part of the PE Corporation.

Venter

**Celera is a for-profit
organization whose motto is
“Discovery Can’t Wait”**



Intro to Sequencing

- To read the DNA, the chromosomes are cut into tiny pieces, each of which is read individually.
- When all the segments have been read they are assembled in the correct order. Link these fragments to self-replicating forms of DNA = vectors.

Intro to Sequencing

- Two approaches have been used to sequence the genome.
- They differ in the methods they use to cut up the DNA, assemble it in the correct order, and **whether** they map the chromosomes before decoding the sequence.

Intro to Sequencing: BAC to BAC

- The BAC-to-BAC method:
 - the first to be employed in human genome studies
 - slow but sure
 - also called the “map based method”

BAC to BAC - 1

- First create a rough physical map of the whole genome before sequencing the DNA
 - requires cutting the chromosomes into large pieces and then figuring out the order of these big chunks of DNA before taking a closer look and sequencing all the fragments.

BAC to BAC - 2

II. Several copies of the genome are randomly cut into pieces that are about 150,000 base pairs (bp) long.

BAC to BAC - 2

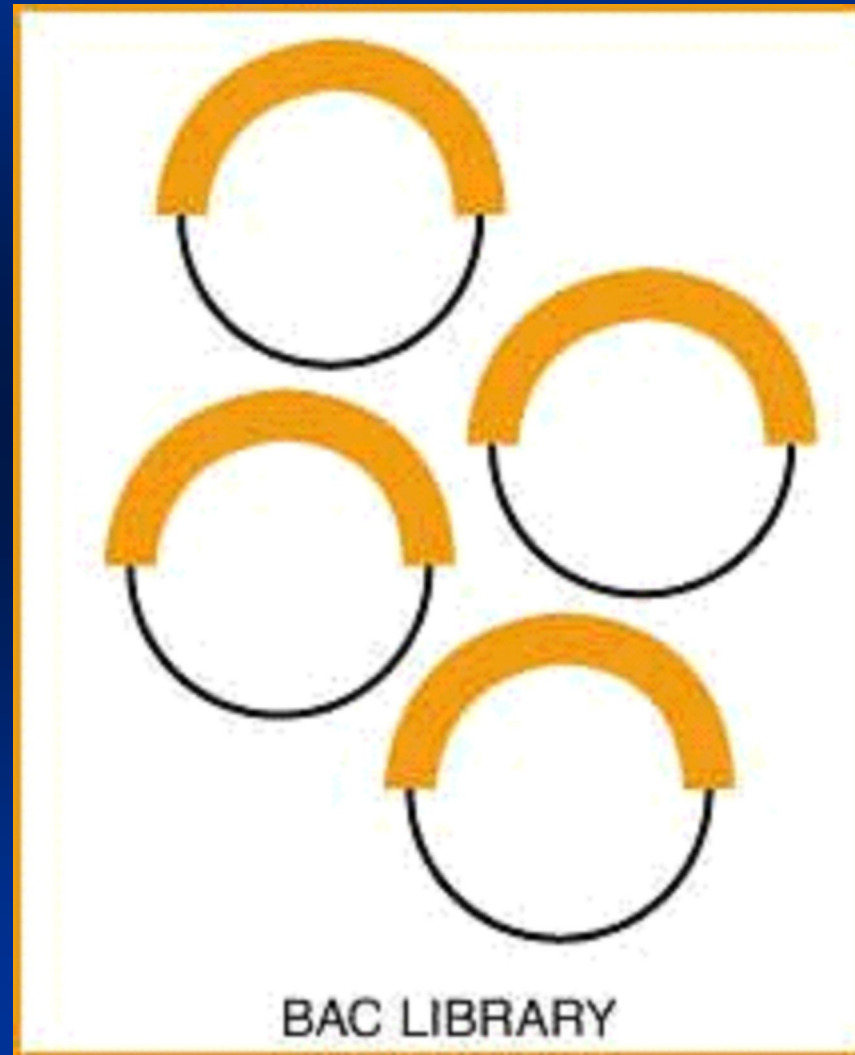


BAC to BAC - 3

Each of these 150,000 bp fragment is inserted into a BAC

- ◆ A BAC is a man made piece of DNA that can replicate inside a bacterial cell.
- ◆ The collection of BACs containing the entire human genome is called “a **BAC library**”.

BAC to BAC - 3



BAC to BAC - 4

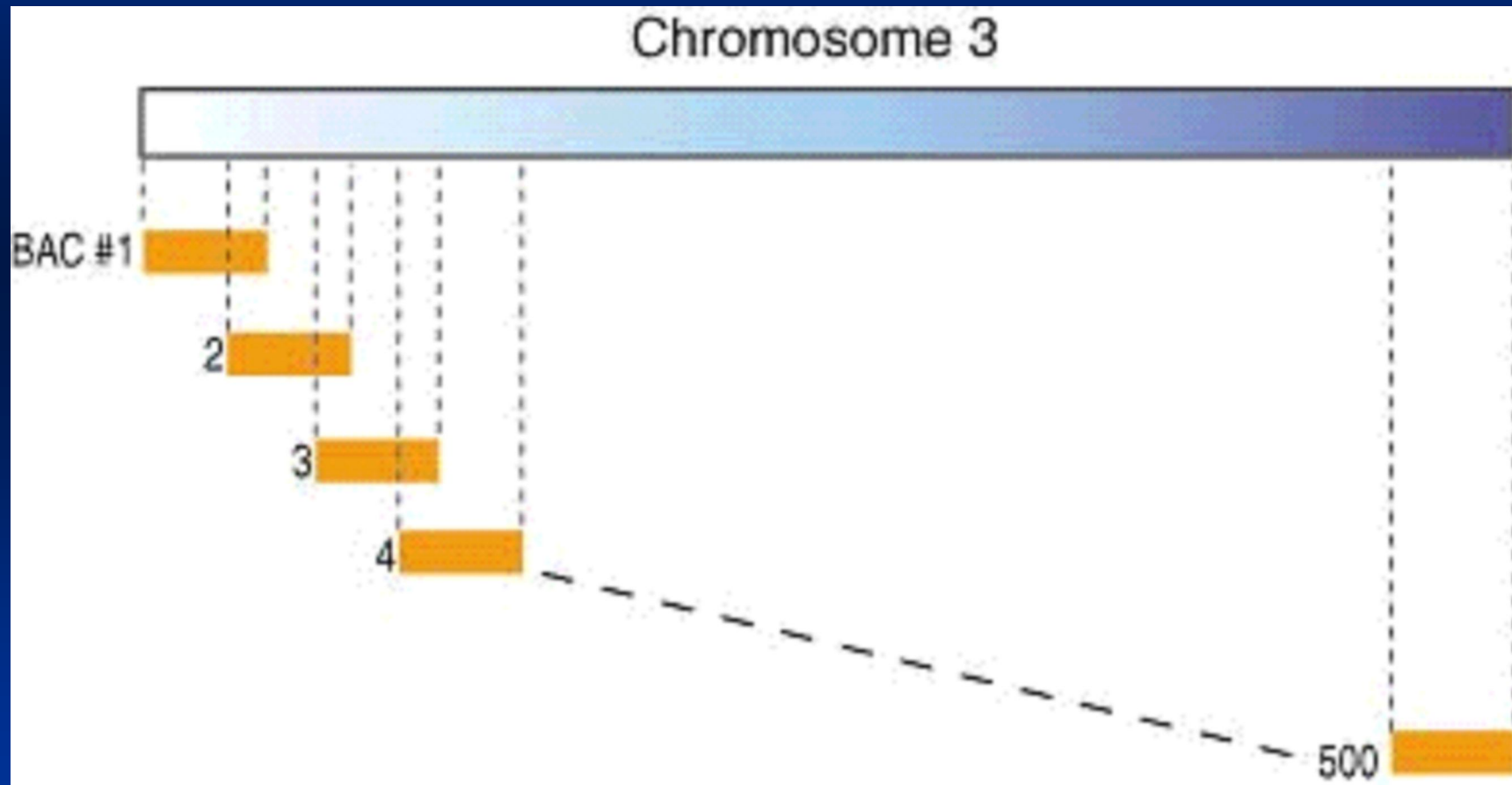
These pieces are fingerprinted to give each piece a unique identification tag that determines the order of the fragments.

Cutting each BAC fragment with a single enzyme and finding common sequence landmarks in overlapping fragments that determine the location of each BAC along the chromosome.

BAC to BAC - 4

Then overlapping BACs with markers every 100,000 bp form a map of each chromosome

BAC to BAC - 4



BAC to BAC - 5

- 1 Each BAC is then broken randomly into 1,500 bp pieces and placed in another artificial piece of DNA called M13. This collection is known as an M13 library.

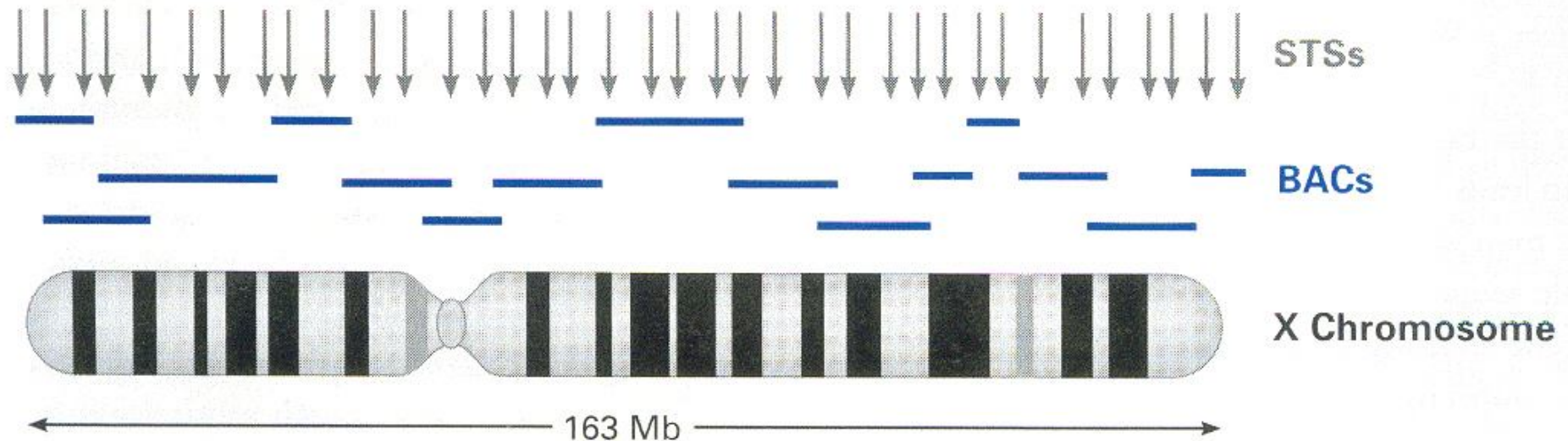
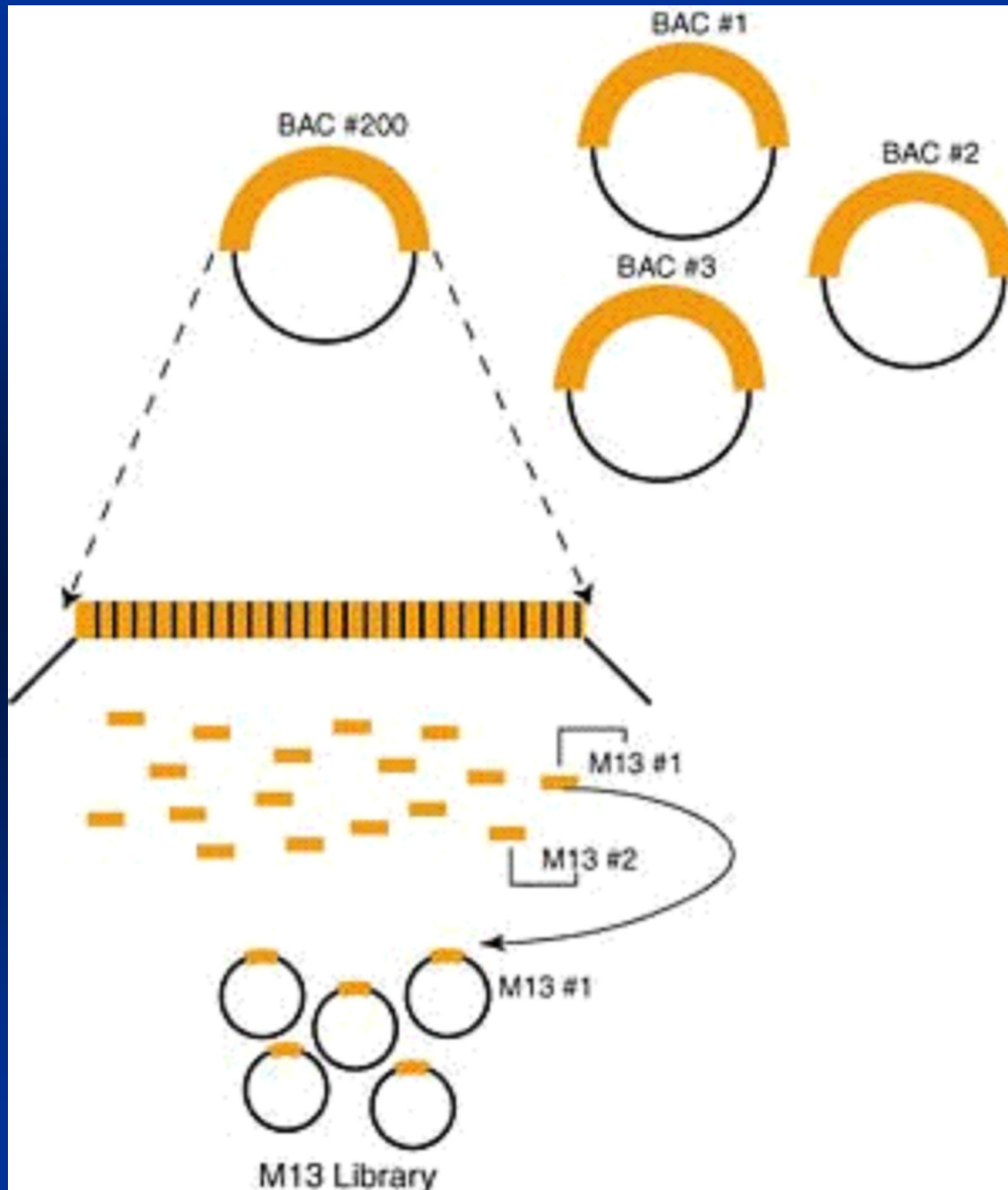


FIGURE 1.3 • Relationships of chromosomes to genome sequencing markers. The X chromosome is about 163 Mb in length. In this diagram, there are 16 overlapping BAC clones that span the entire length. In reality, 1,408 BACs were needed to span the X chromosome. Arrows (top) mark STSs scattered throughout the chromosome and on overlapping BACs.

BAC to BAC - 5



BAC to BAC - 6

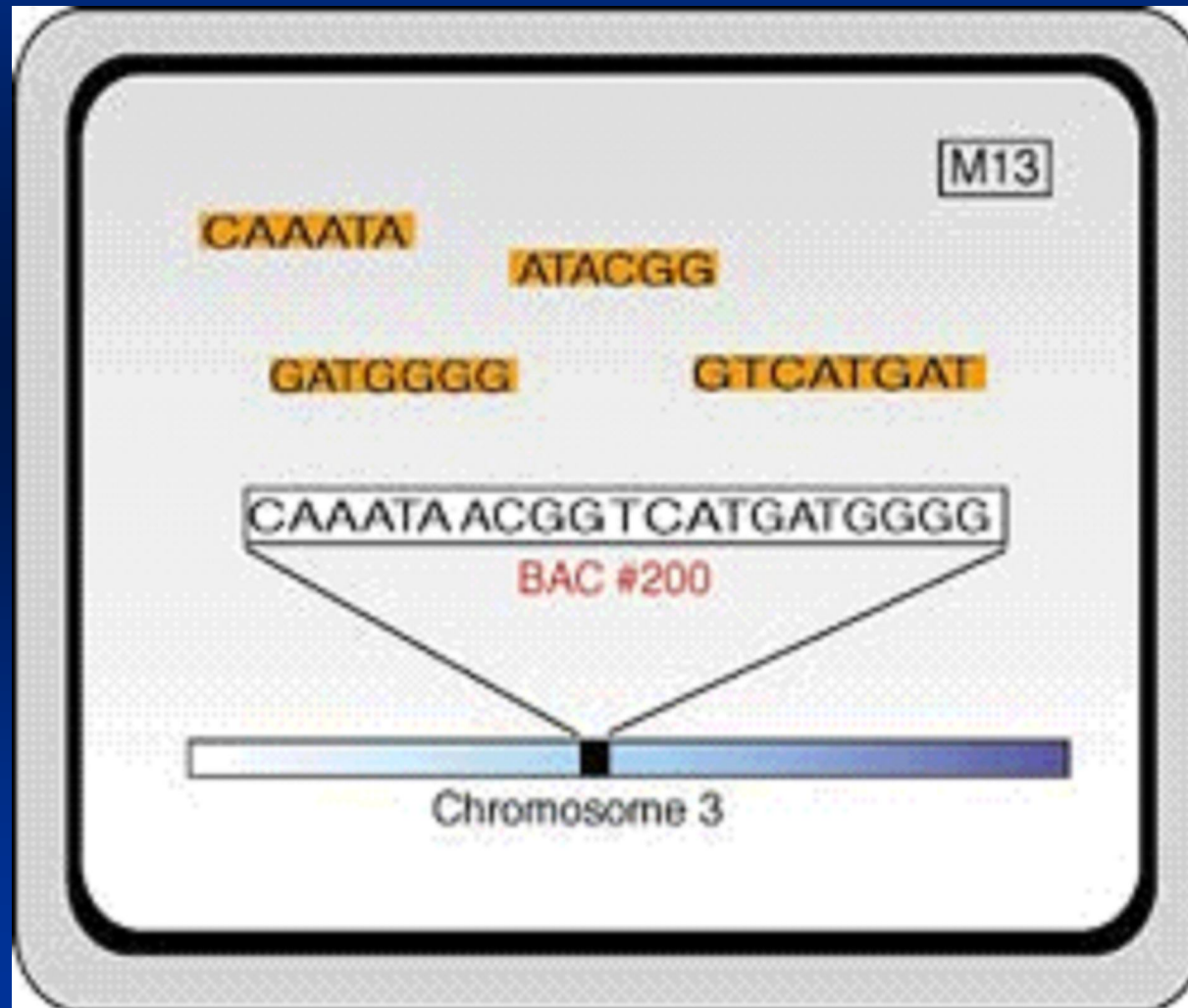
All the M13 libraries are sequenced.

500 bp from one end of the fragment are sequenced generating millions of sequences

BAC to BAC - 7

These sequences are fed into a computer program called PHRAP that looks for common sequences that join two fragments together.

BAC to BAC - 7



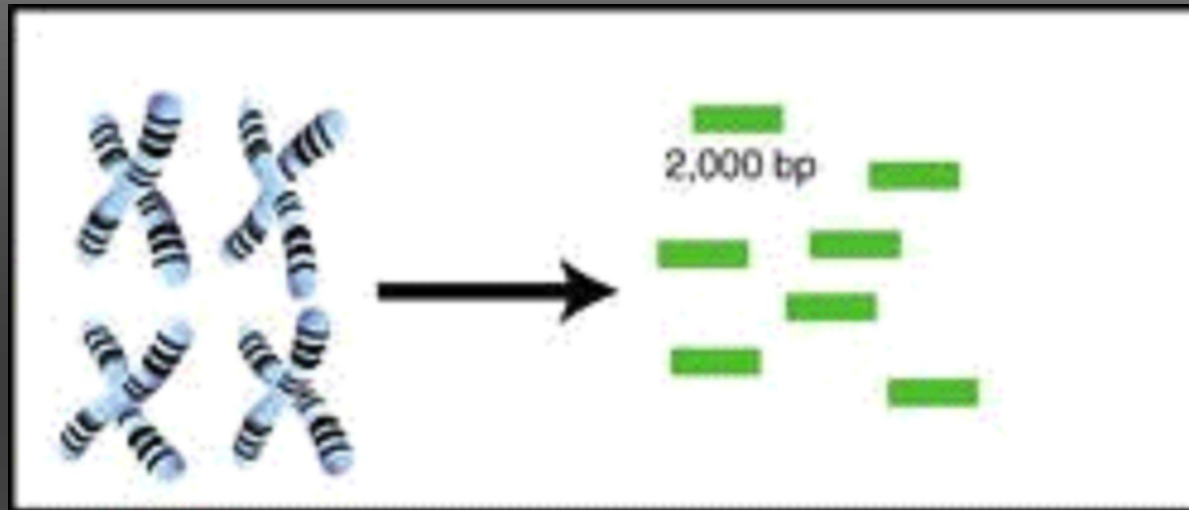
Whole Genome Shotgun - 1

- 1 The shotgun sequencing method goes straight to the job of decoding, bypassing the need for a physical map.

Whole Genome Shotgun - 2

- 1 Multiple copies of the genome are randomly shredded into pieces that are 10,000 bp long by squeezing the DNA through a pressurized syringe. This is done a second time to generate pieces that are 2,000 bp long.

Whole Genome Shotgun - 2



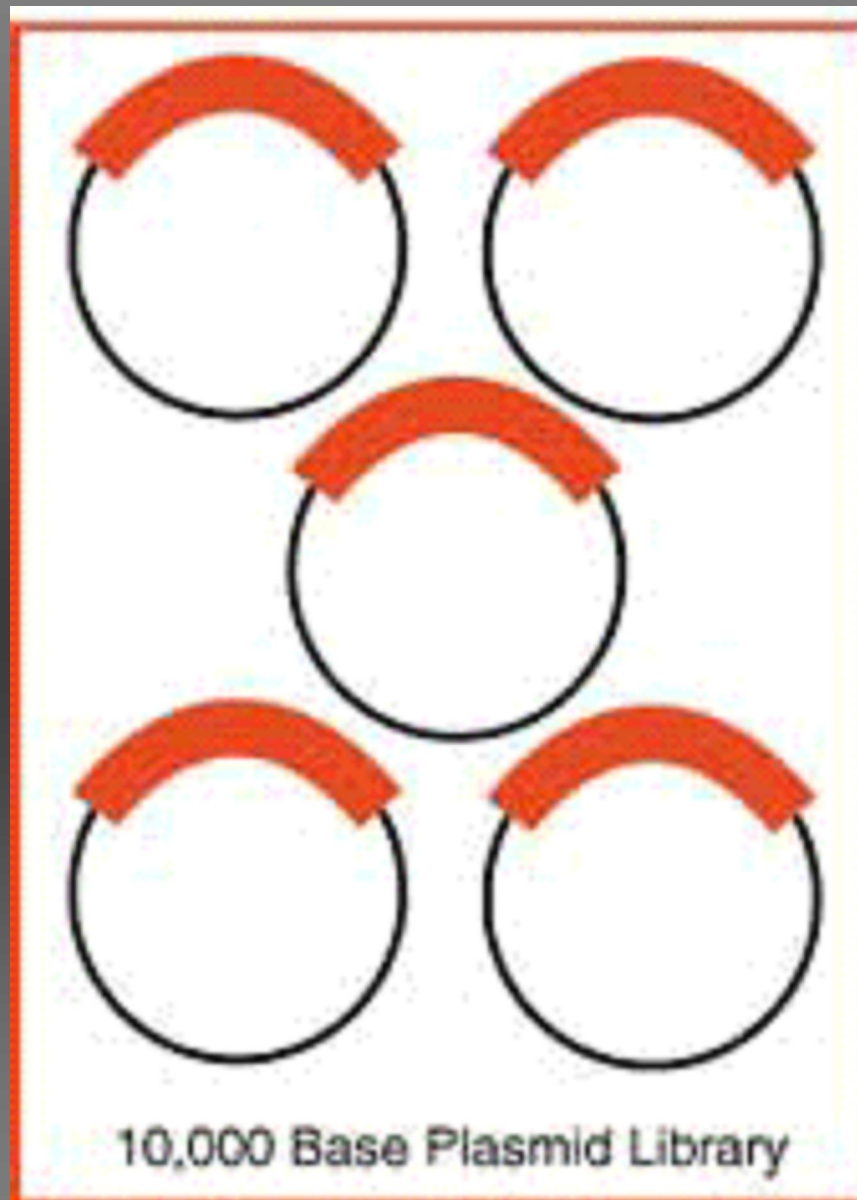
Whole Genome Shotgun - 3

- 1 Each 2,000 and 10,000 bp fragment is inserted into a plasmid, which is a piece of DNA that can replicate in bacteria.
 - 1 The two collections of plasmids containing 2,000 and 10,000 bp chunks of human DNA are known as plasmid libraries.

Intro to Sequencing: Whole Genome Shotgun

- Whole Genome Shotgun Method brings speed into the picture, enabling researchers to do the job in months to a year.
- Developed by Celera president Craig Venter in 1996 when he was at the Institute for Genomic Research.

Whole Genome Shotgun - 3

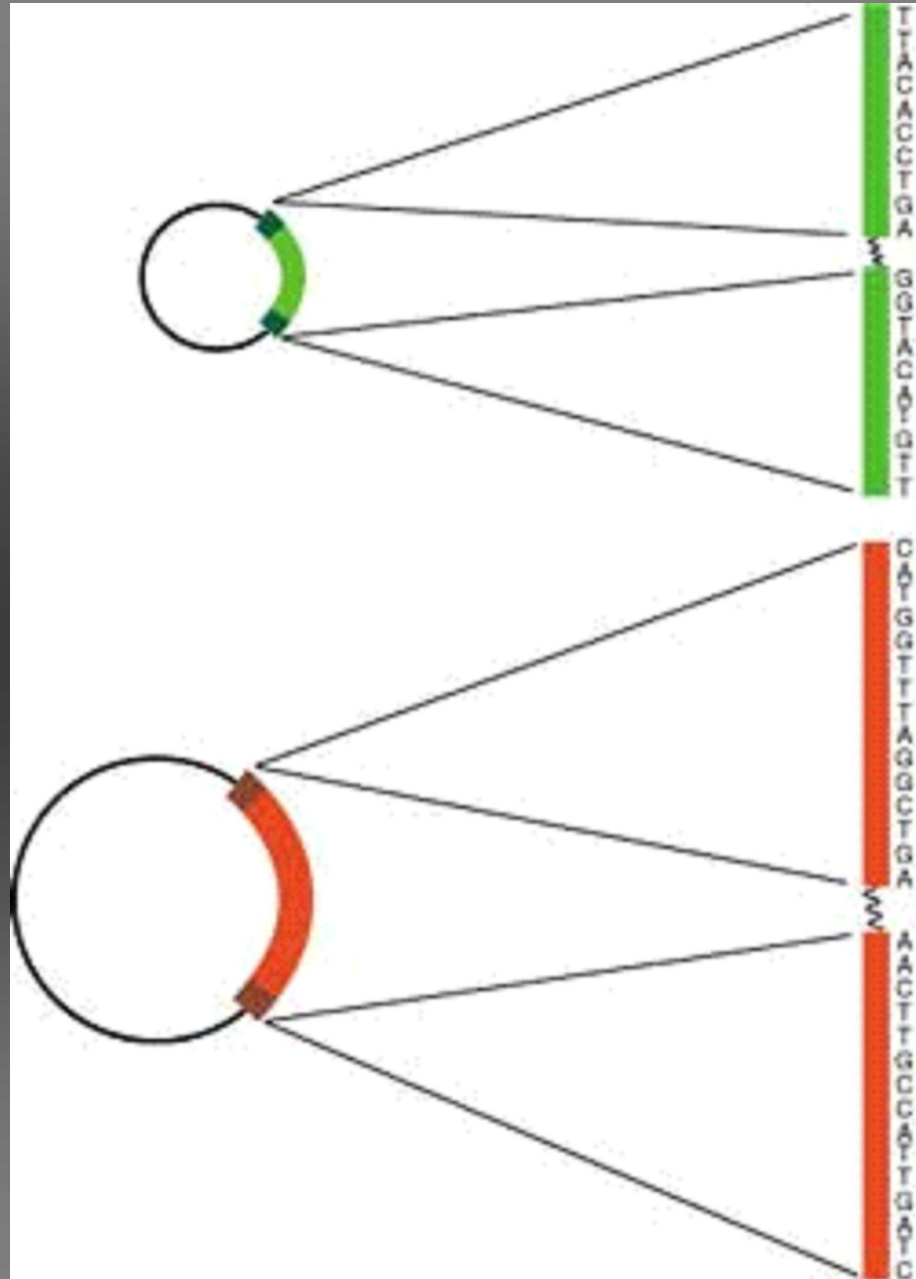


Whole Genome Shotgun - 4

Both plasmid libraries are sequenced.

500 bp from each end of each fragment are decoded generating millions of sequences.

Sequencing both ends of each insert is critical for assembling the entire chromosome.

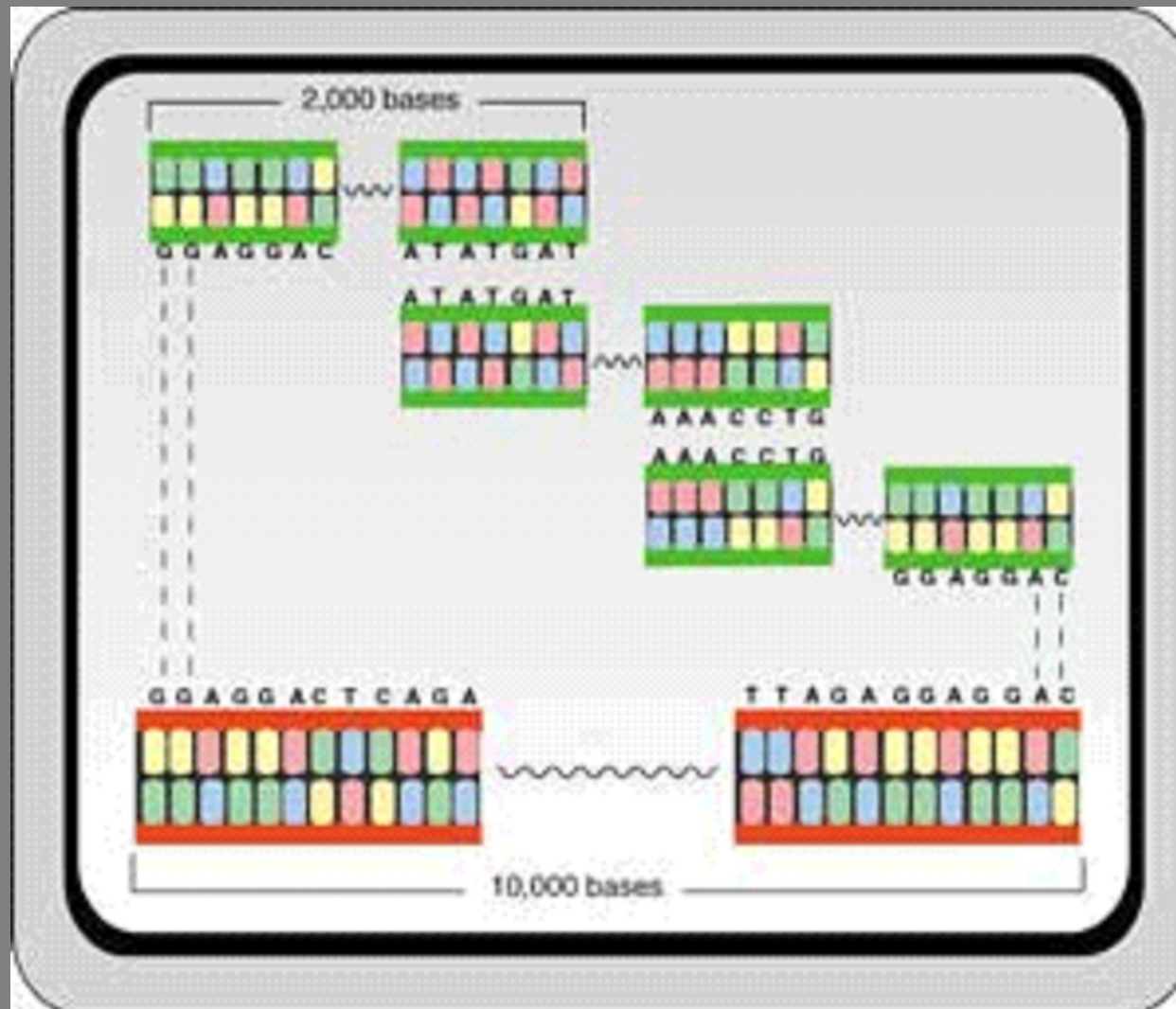


Whole Genome Shotgun - 4

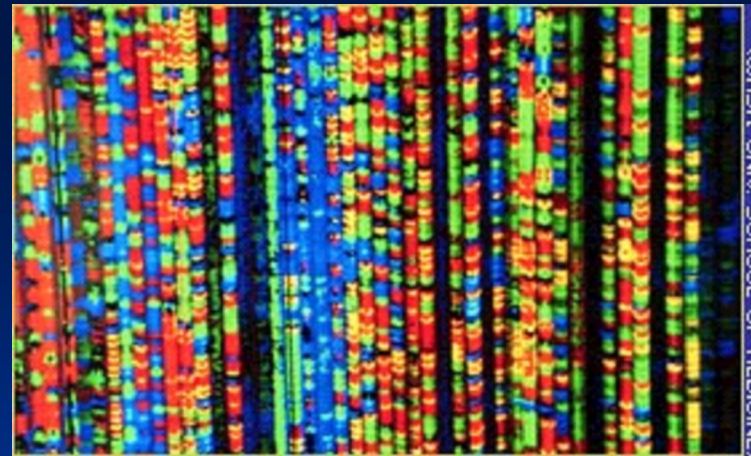
Whole Genome Shotgun - 5

- 1 Computer algorithms assemble the millions of sequenced fragments into a continuous stretch resembling each chromosome.

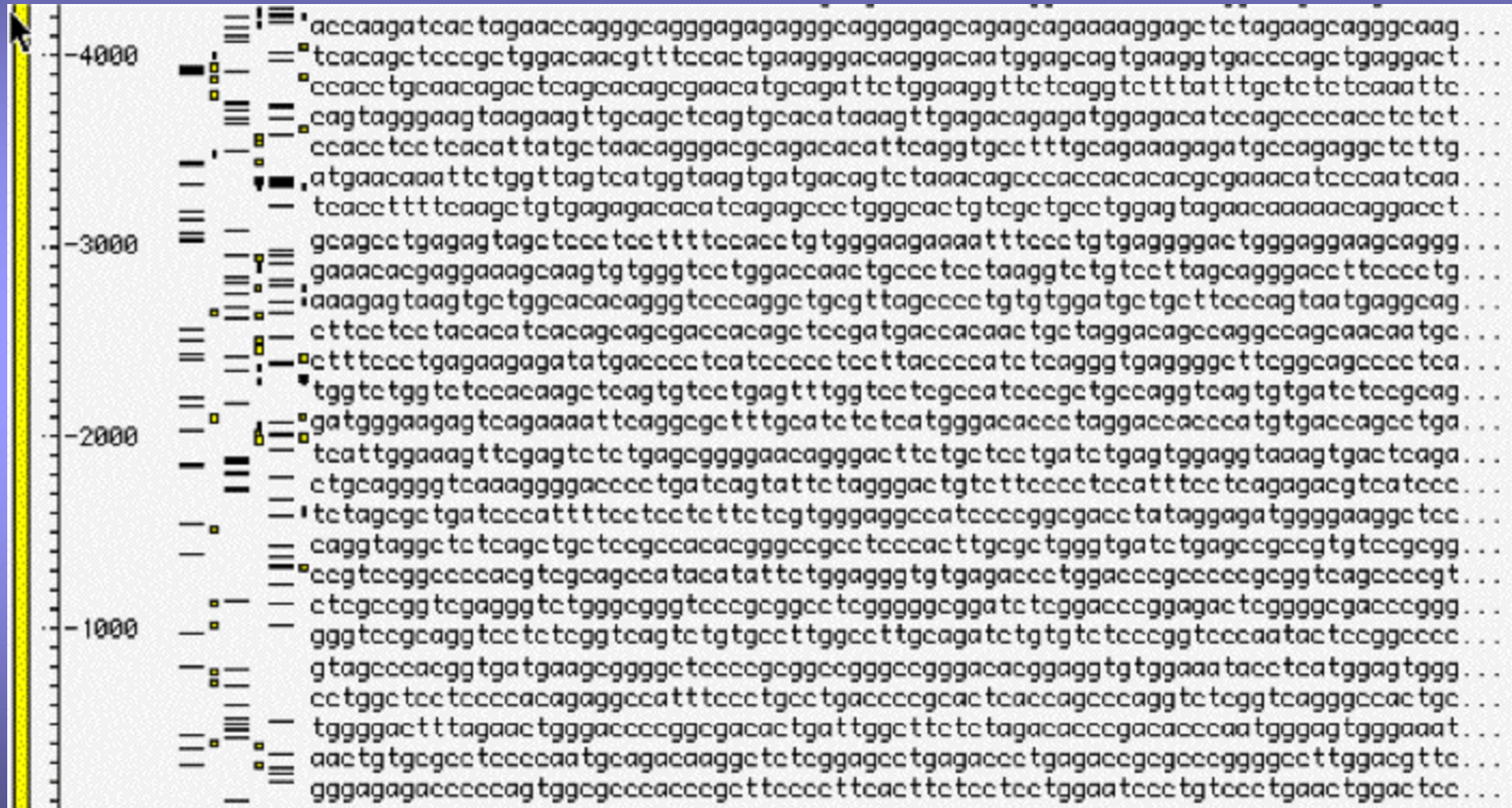
Whole Genome Shotgun - 5



- **Genome sequencing factories churn out raw sequence data at an ever increasing rate**
- **Fewer scientists are involved in generating data and more are involved in data analysis**



Raw Genome Data:



Finding genes in genome sequence is not easy

- About 1.3 % of human genome encodes functional protein.
- 28 % of human genome is transcribed into RNA
- 50 % of human genome are repetitive sequences
- Genes are interspersed among long stretches of non-coding DNA.

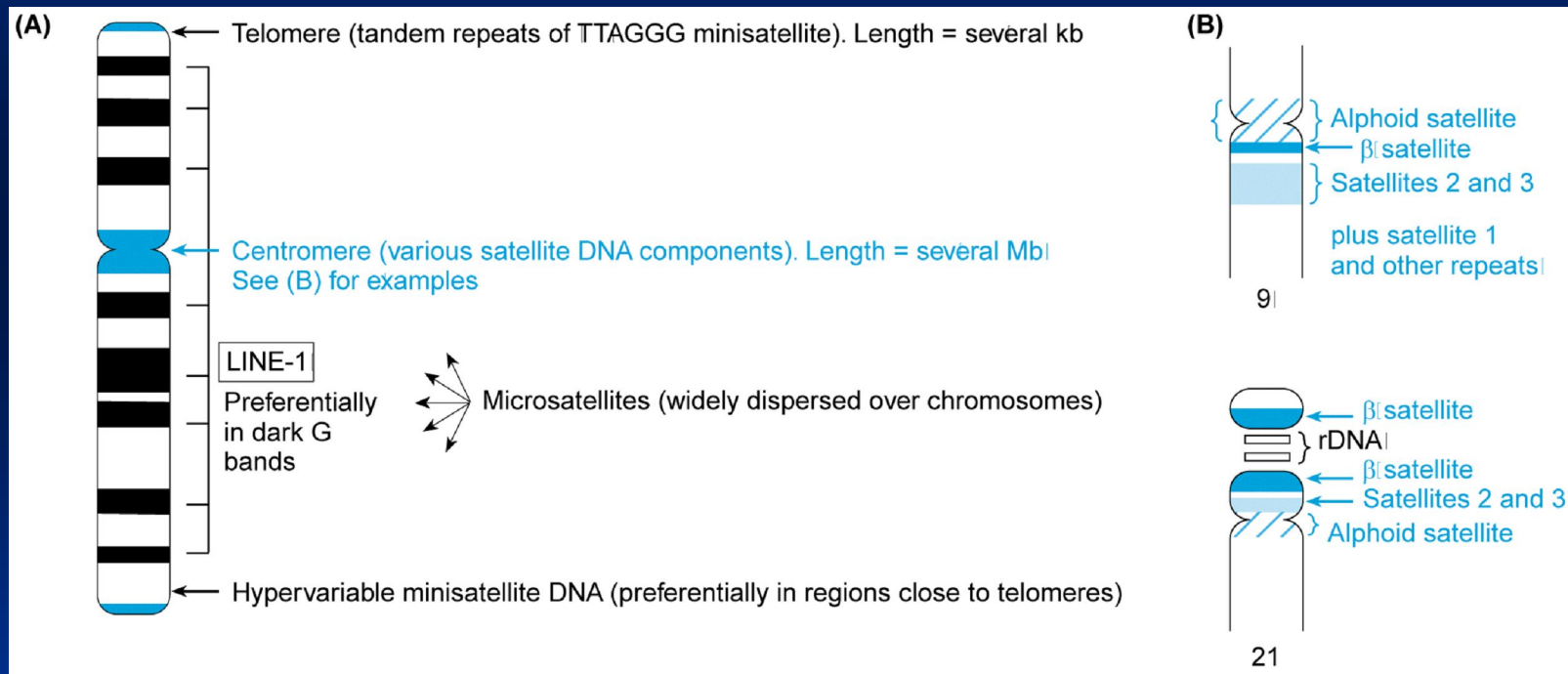
Repeats, pseudo-genes, and introns confound matters



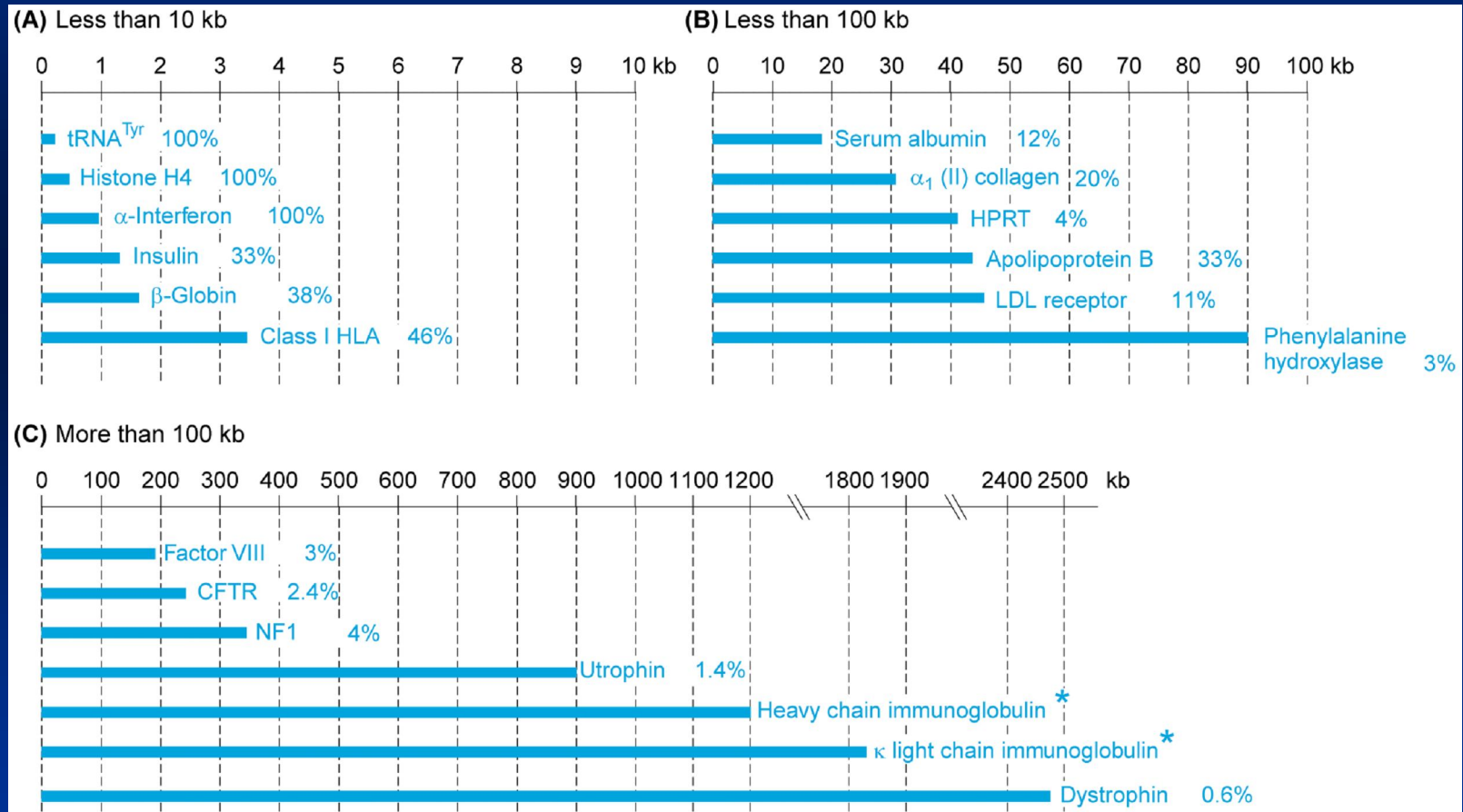
Gene families

- **Members may exhibit high sequence homology**
- **sometimes contain a highly conserved domain**
(e.g. SOX box)
- **sometimes contain a very short conserved “motif”** (e.g. DEAD box, asp-glu-ala-asp RNA helicases)
- **superfamilies** (e.g. Ig superfamily)
- **sometimes clustered** (e.g. globin genes)
- **Often associated with truncated and non processed pseudogenes**

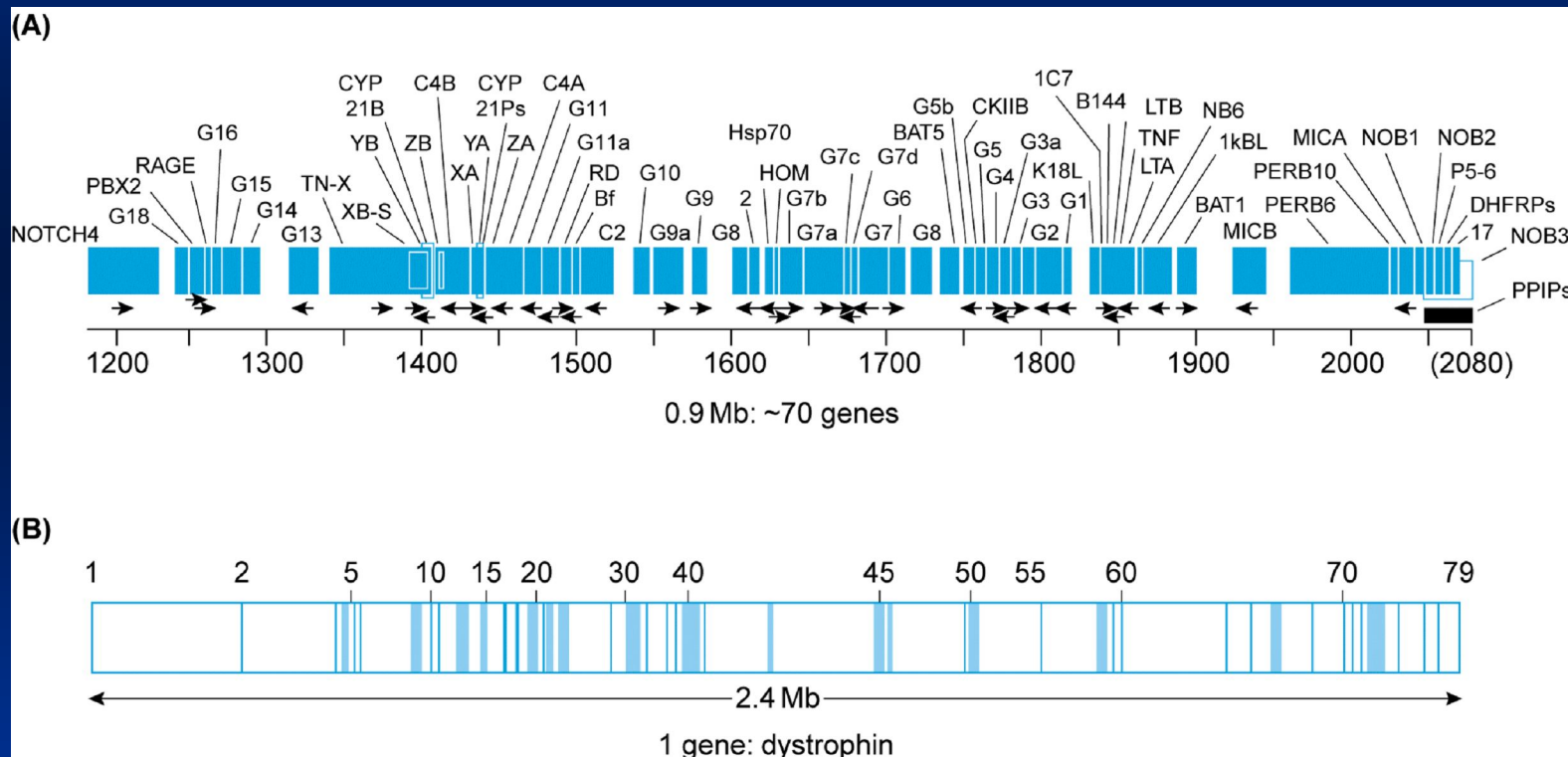
Chromosomal location of repetitive DNA



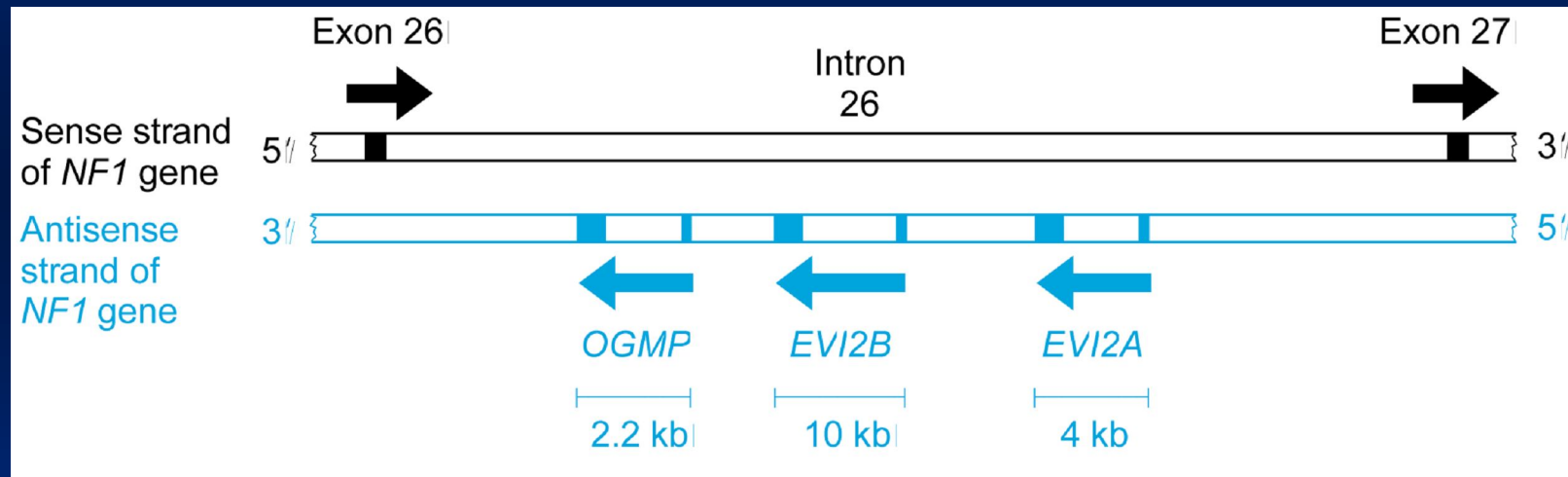
Human genes vary enormously in size and exon content



GENE DENSITIES

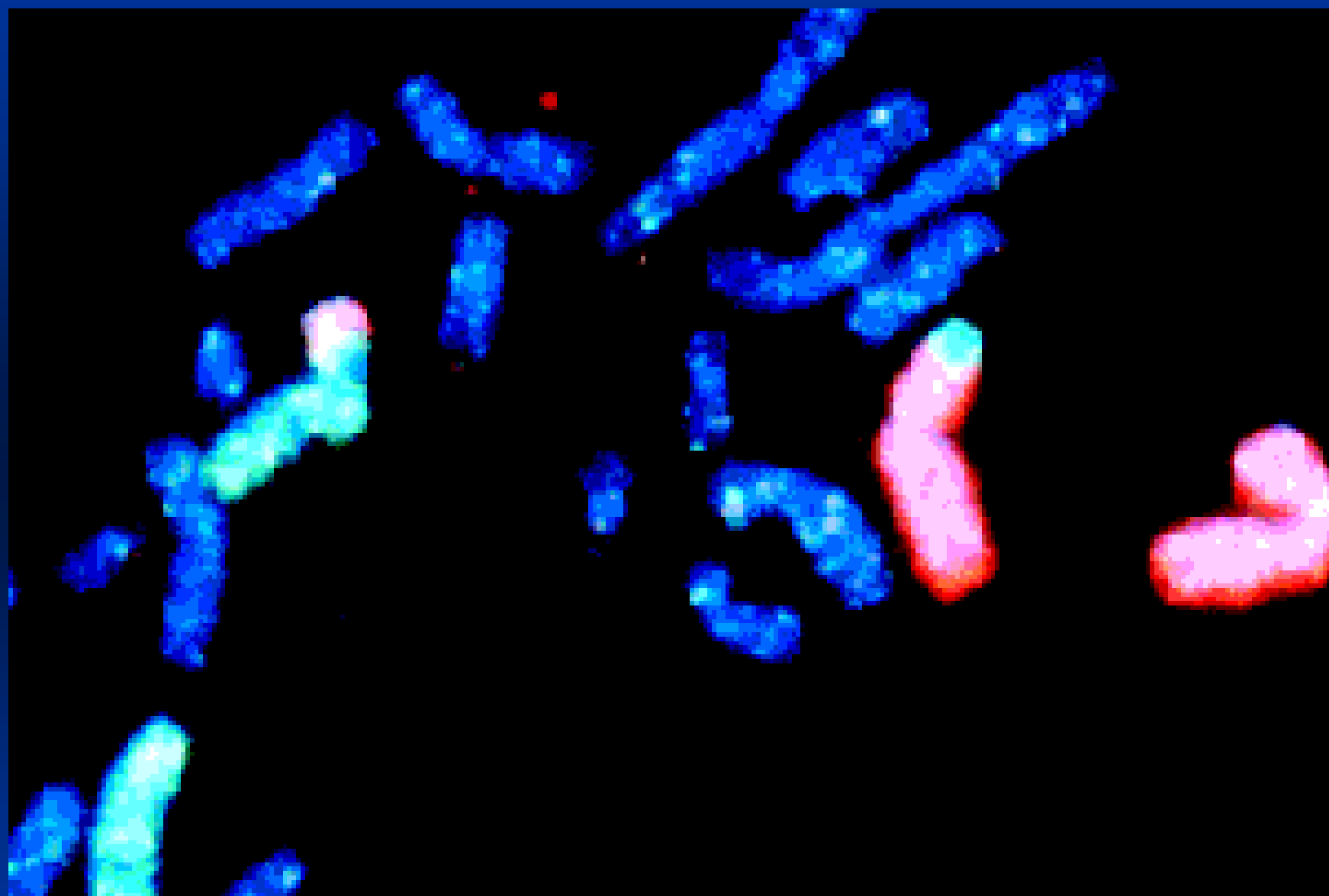


GENES WITHIN GENES



**“Tenemos una enorme cantidad de
información, pero nos falta
conocimiento”**

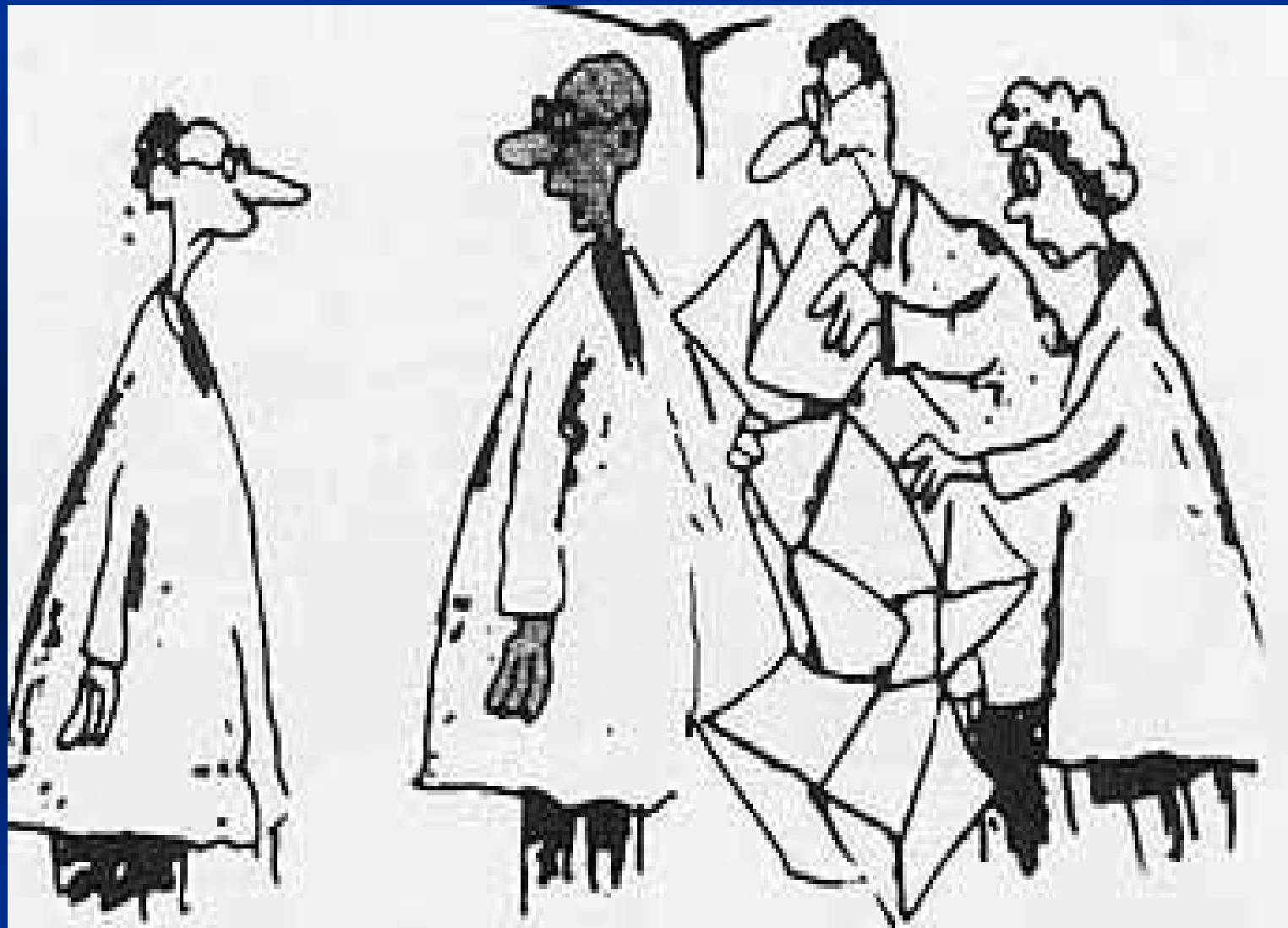
Sidney Brenner



Proteomics

- **Identify all of the proteins in an organism**
 - **Potentially many more than genes due to alternative splicing and post-translational modifications**
- **Quantitate in different cell types and in response to metabolic/environmental factors**
- **Protein-protein interactions**

- **La Medicina Molecular debería ser una especialidad médica.**



"We finished the genomic map, now we can't figure out how to fold it."