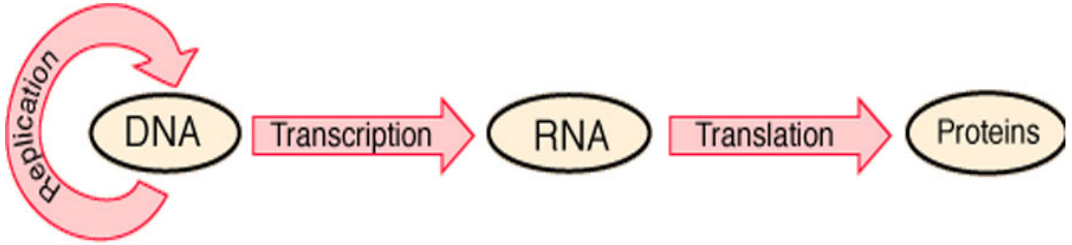
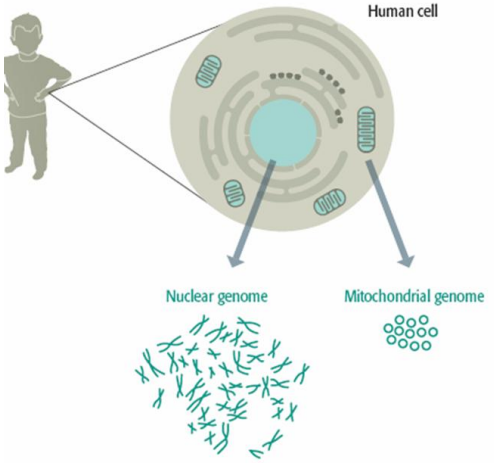
Dogma Central de la Biología

desde el DNA a la proteína.



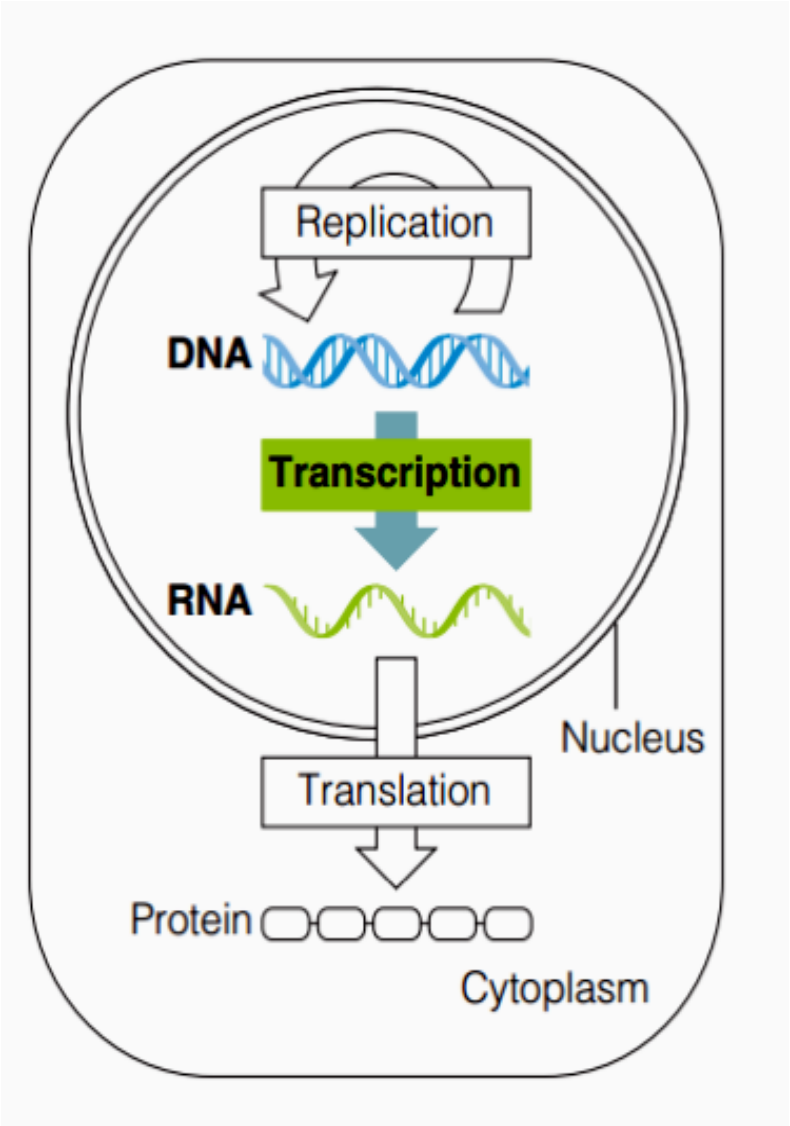
Genoma Humano.

Características y organización del genoma. Variabilidad en poblaciones

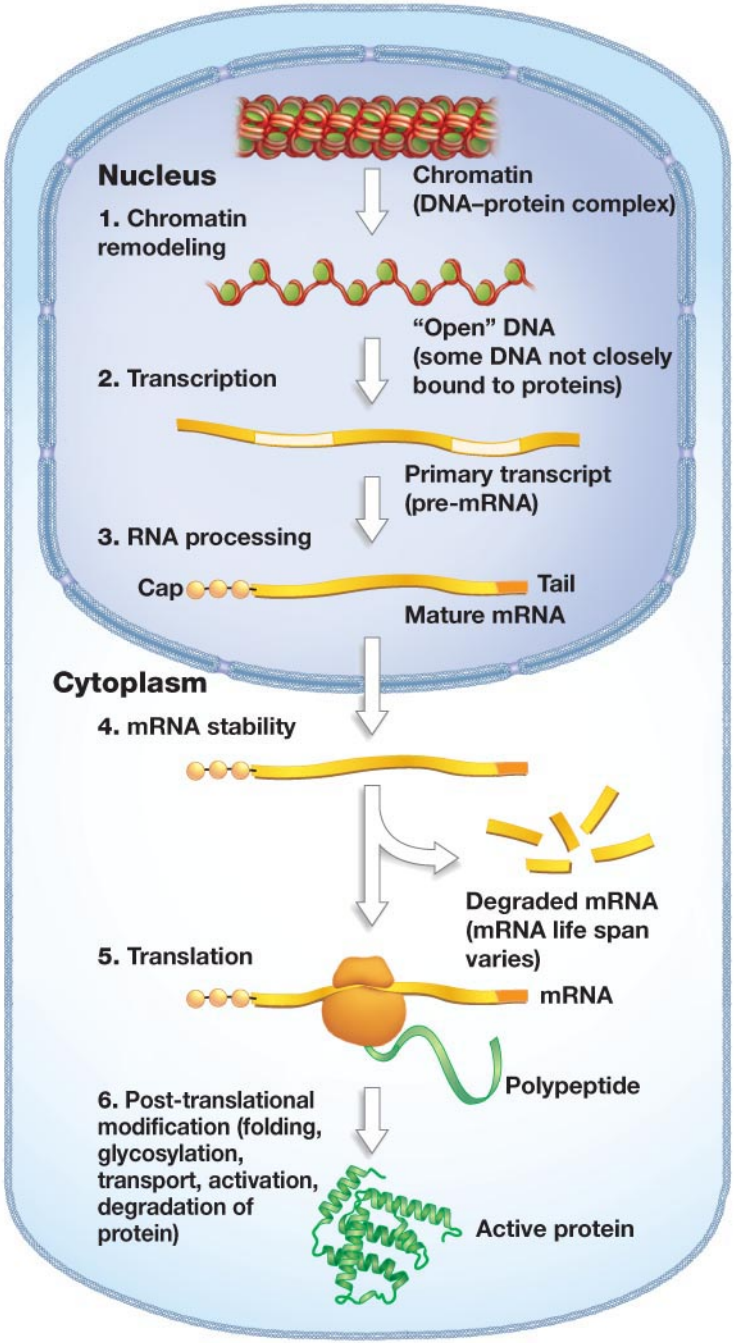


Luisa Herrera
Programa de Genética Humana
ICBM Facultad de Medicina

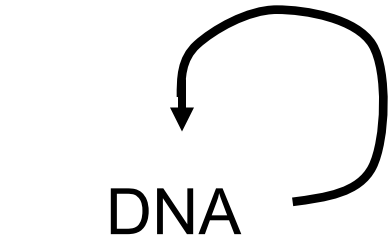
INSTITUTO DE CIENCIAS BIOMÉDICAS



Dogma central de la Biología



© 2011 Pearson Education, Inc.



Replicación

DNA

Transcripción y procesamiento

RNAm (tRNA y rRNA)

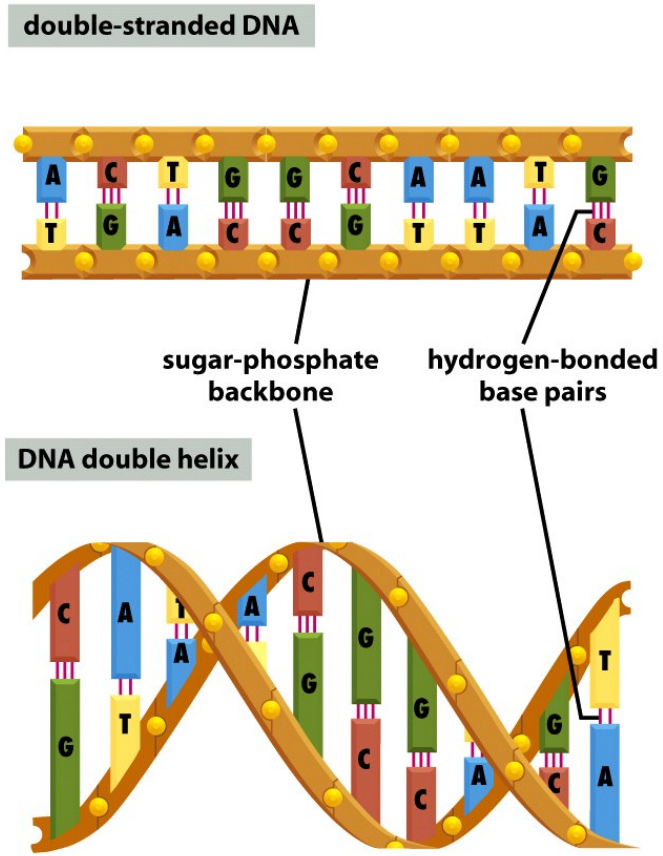
Traducción

PROTEÍNAS

Núcleo

Citoplasma

Flujo de la Información Génica

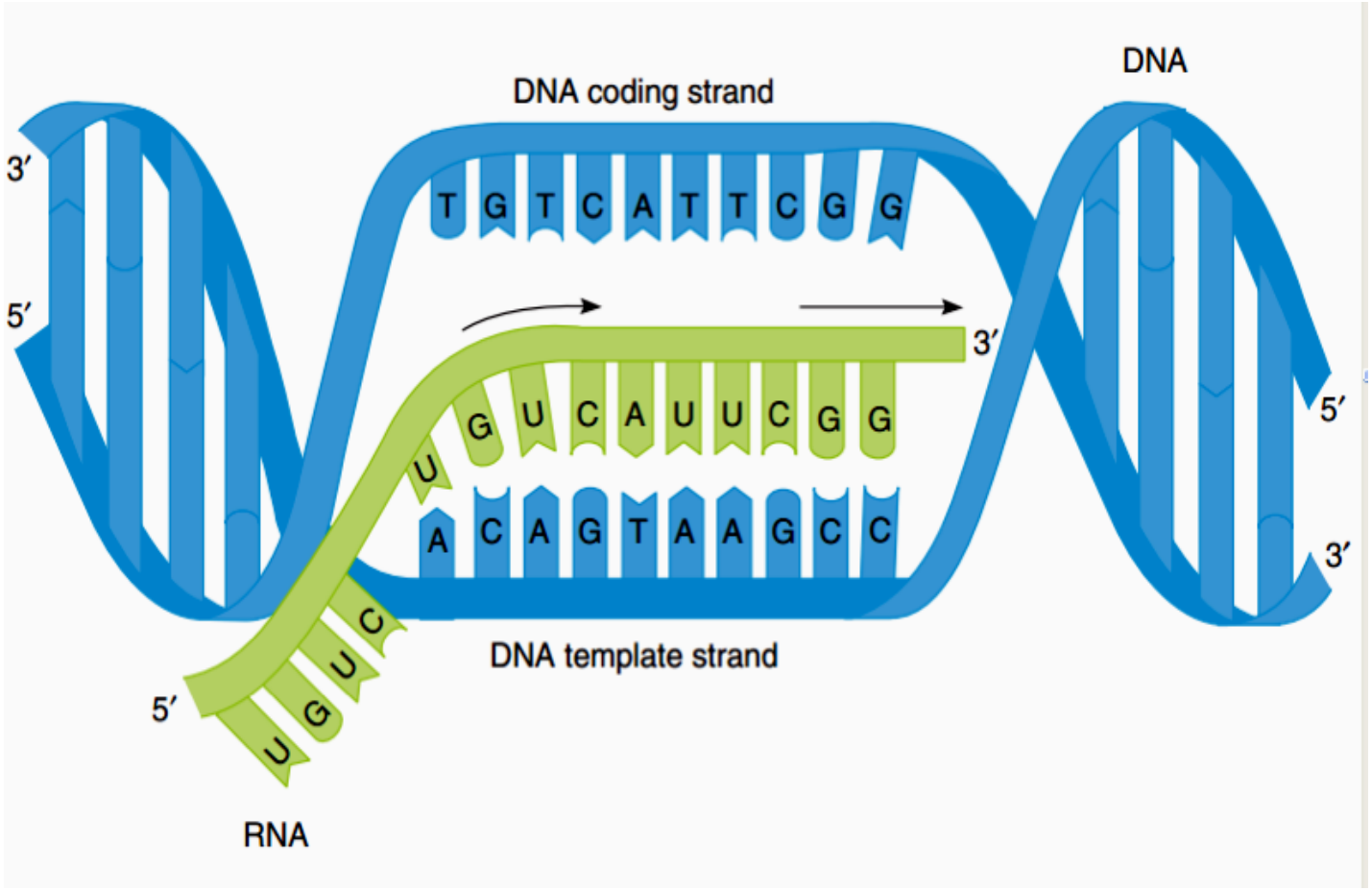


http://www.youtube.com/watch?v=C22BtSCH4aM

http://www.youtube.com/watch?v=qOA25GbUkdA

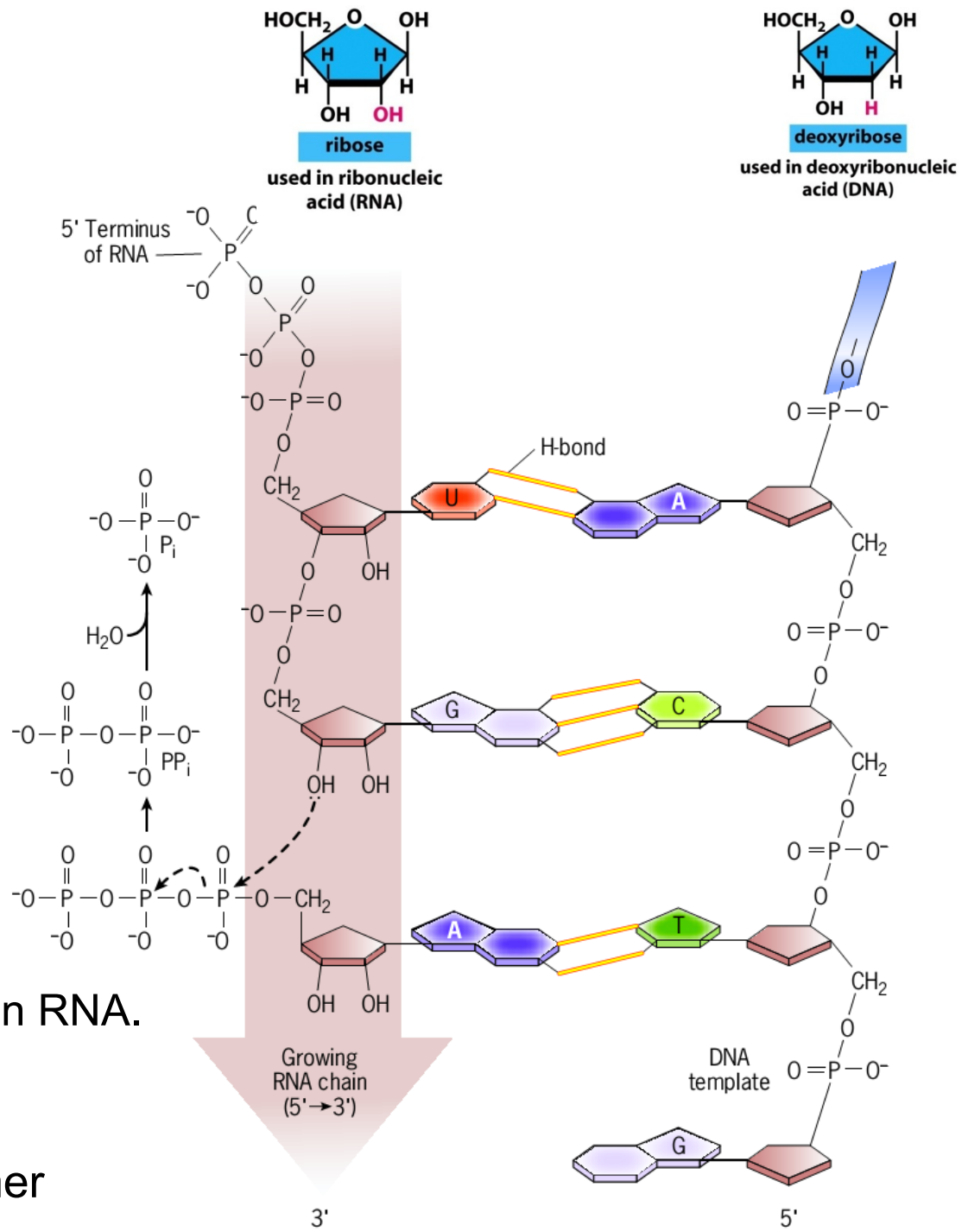
http://www.youtube.com/watch?v=CWLgALHiIvI

Transcripción



(T---› U)

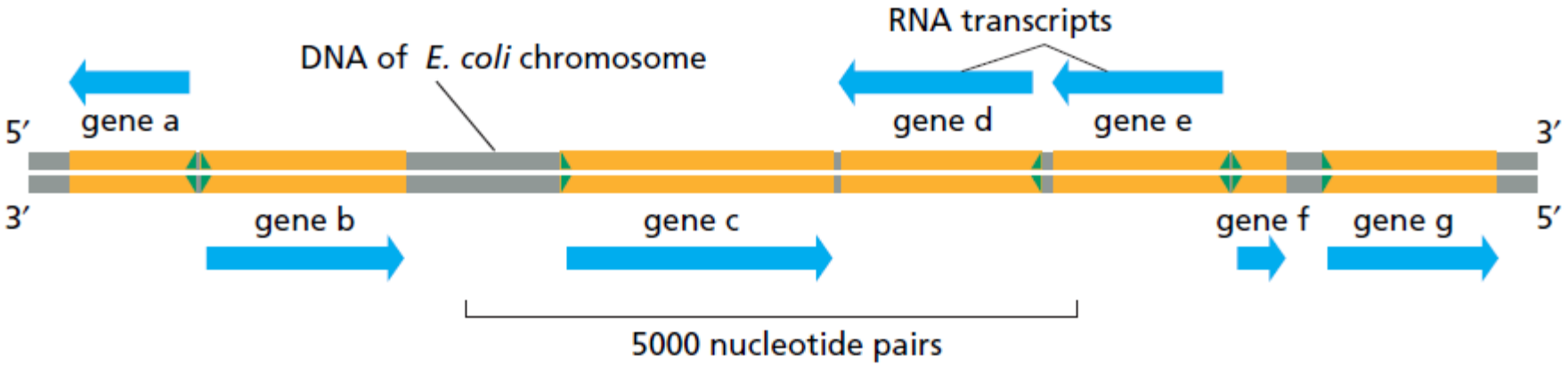
Trascripción



- RNA polimerasa copia el DNA en RNA.
- Síntesis en sentido 5'-3'.

Transcripcion no requiere primer

Algunos genes transcriben usando una hebra de DNA como molde mientras que otros transcriben utilizando la otra hebra.

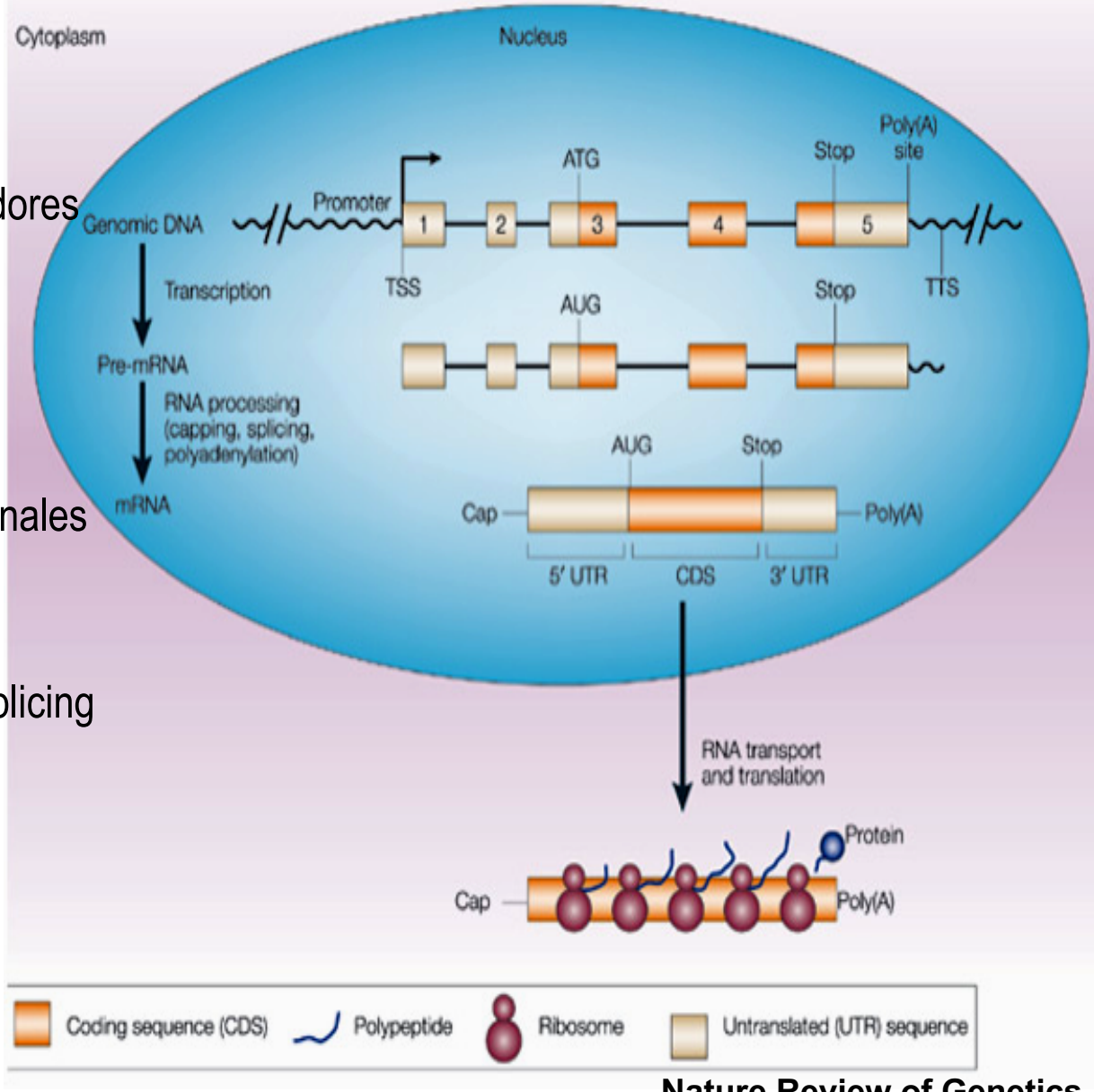


Molecular Biology of the cell. Fifth Ed. , 2008
Alberts, Johnson, Lewis, Raff, Roberts, Walter

Transcripción y procesamiento de RNAs eucariontes codificantes para proteinas

1. Transcripción
   - Promotor, enhancers y silenciadores
   - Factores de transcripción
   - RNA polimerasa II
   - Inicio, elongación y término

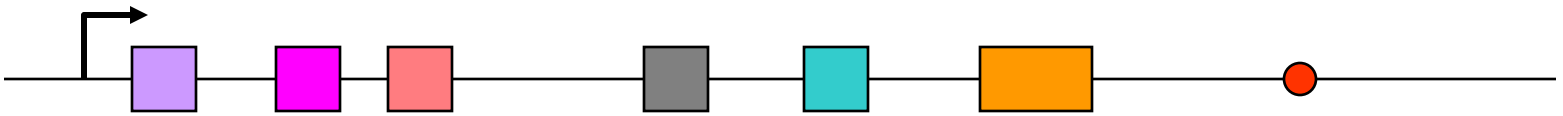
2. Modificaciones post-transcripcionales
   - Adición de Cap en 5´
   - Adición de cola de PoliA en 3´
   - Corte y empalme de exones. Splicing y splicing alternativo
   - Edición de los RNAs



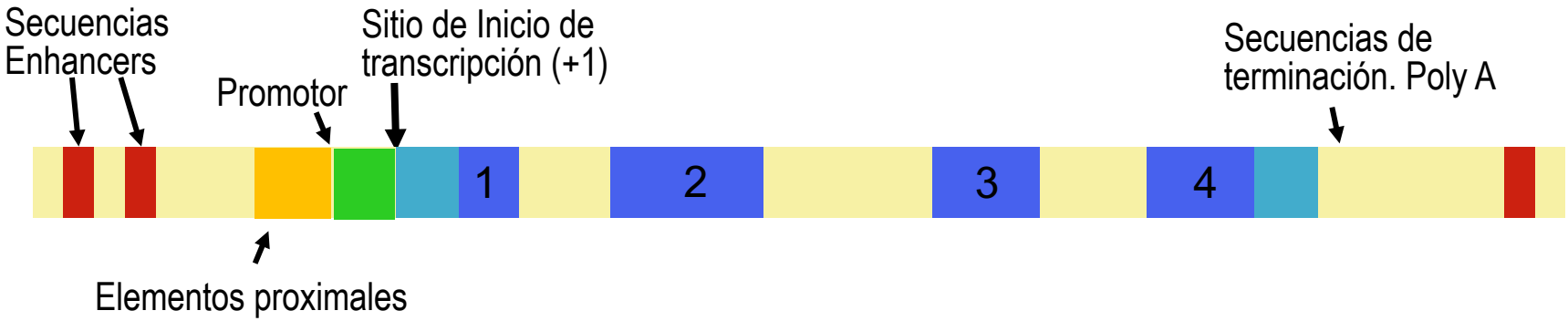
Nature Review of Genetics

Regulación transcripcional

Promotor

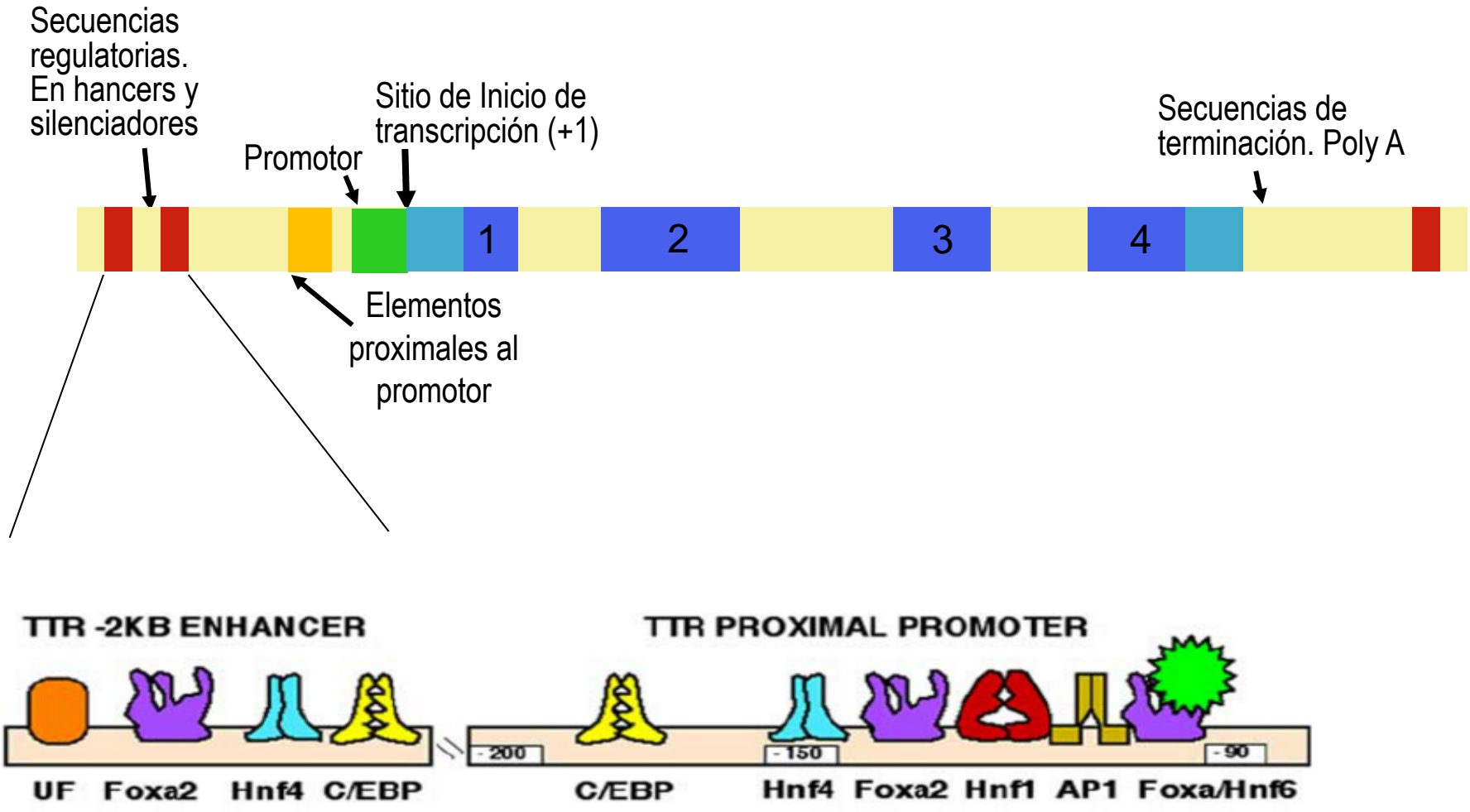


Estructura de un gen eucarionte codificante para proteínas

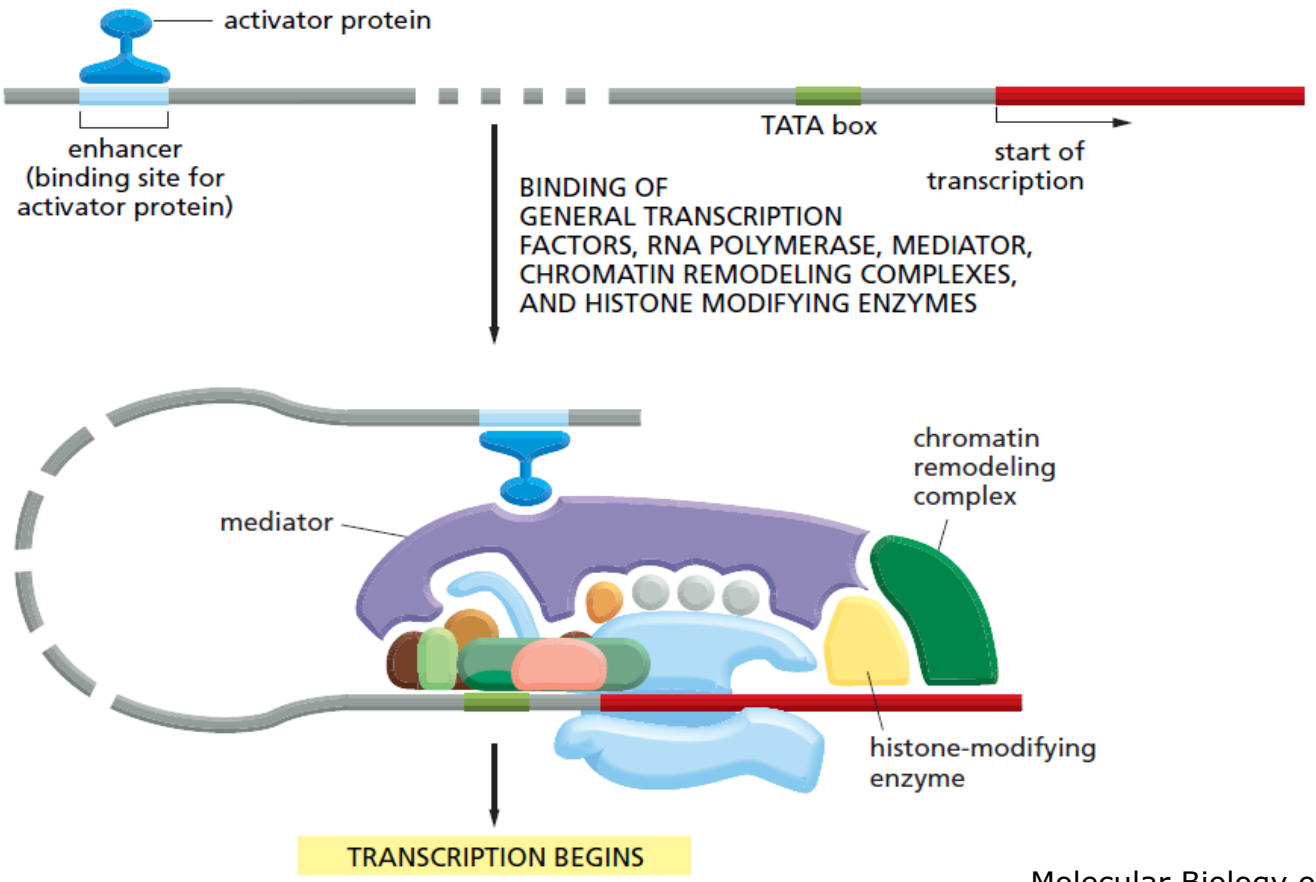


- Promotor: secuencia donde se ensamblan la RNA polimerasa y los factores de transcripcion generales para RNA pol II (TF II) para el inicio de transcripción. Caja TATA, elementos río arriba, DPE (Downstream promoter element), reconocido por TFIID.
- Potenciadores/Enhancers: secuencias reconocidas por activadores transcripcionales
- Silenciadores: secuencias reconocidas por inhibidores transcripcionales

Estructura de un gen eucarionte codificante para proteínas



Acción de factores de transcripción específicos

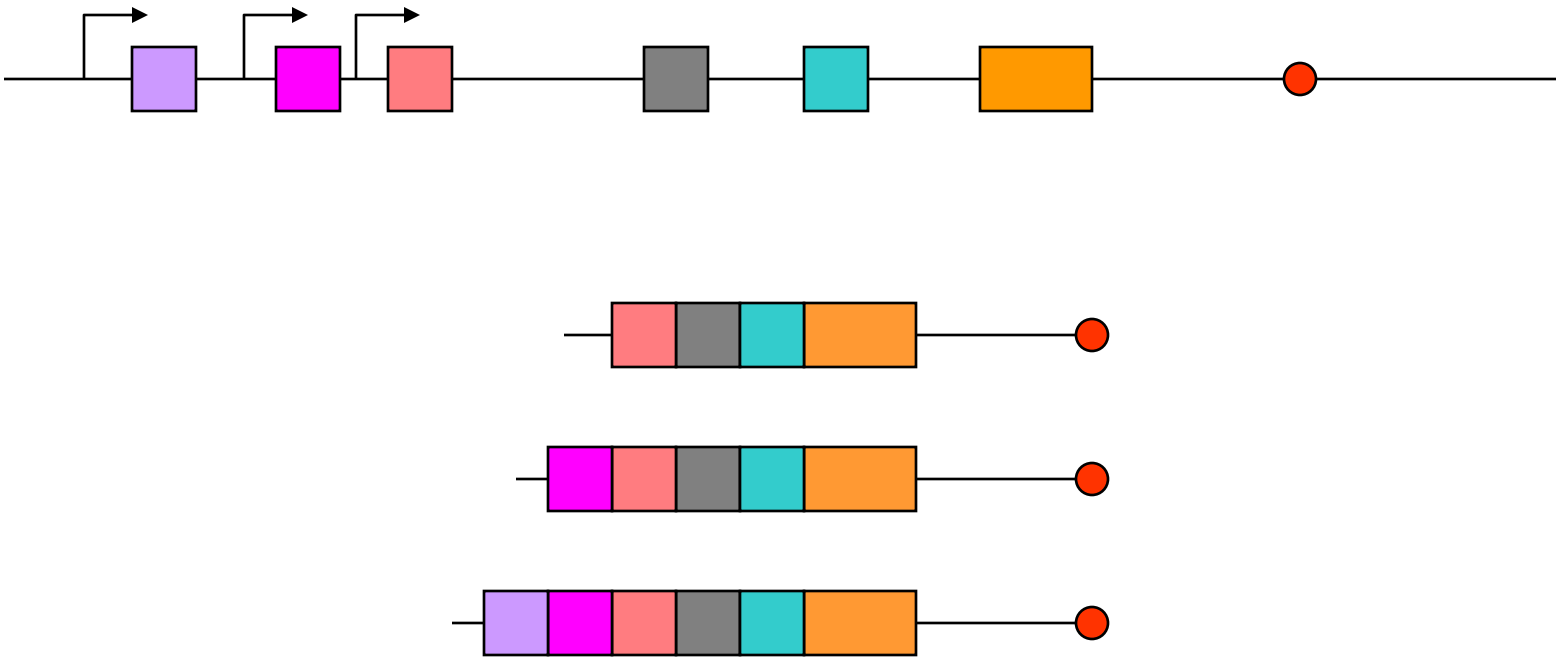


Molecular Biology of the cell. Fifth Ed. , 2008 Alberts, Johnson, Lewis, Raff, Roberts, Walter

- Unión a secuencias específicas.
- Utilizan mediadores o co-activadores para transmitir la señal de activación para que los FT generales y la RNA pol se ensamblen en el promotor.
- Unión de complejos de remodelamiento de cromatina

Muchos genes poseen mas de un promotor.

Uso de promotores alternativos.



- Mas del 70% de los genes codificados en el genoma humano usan promotores alternativos
- Permite regulación diferencial en distintos tejidos
- Generan productos proteicos diferentes

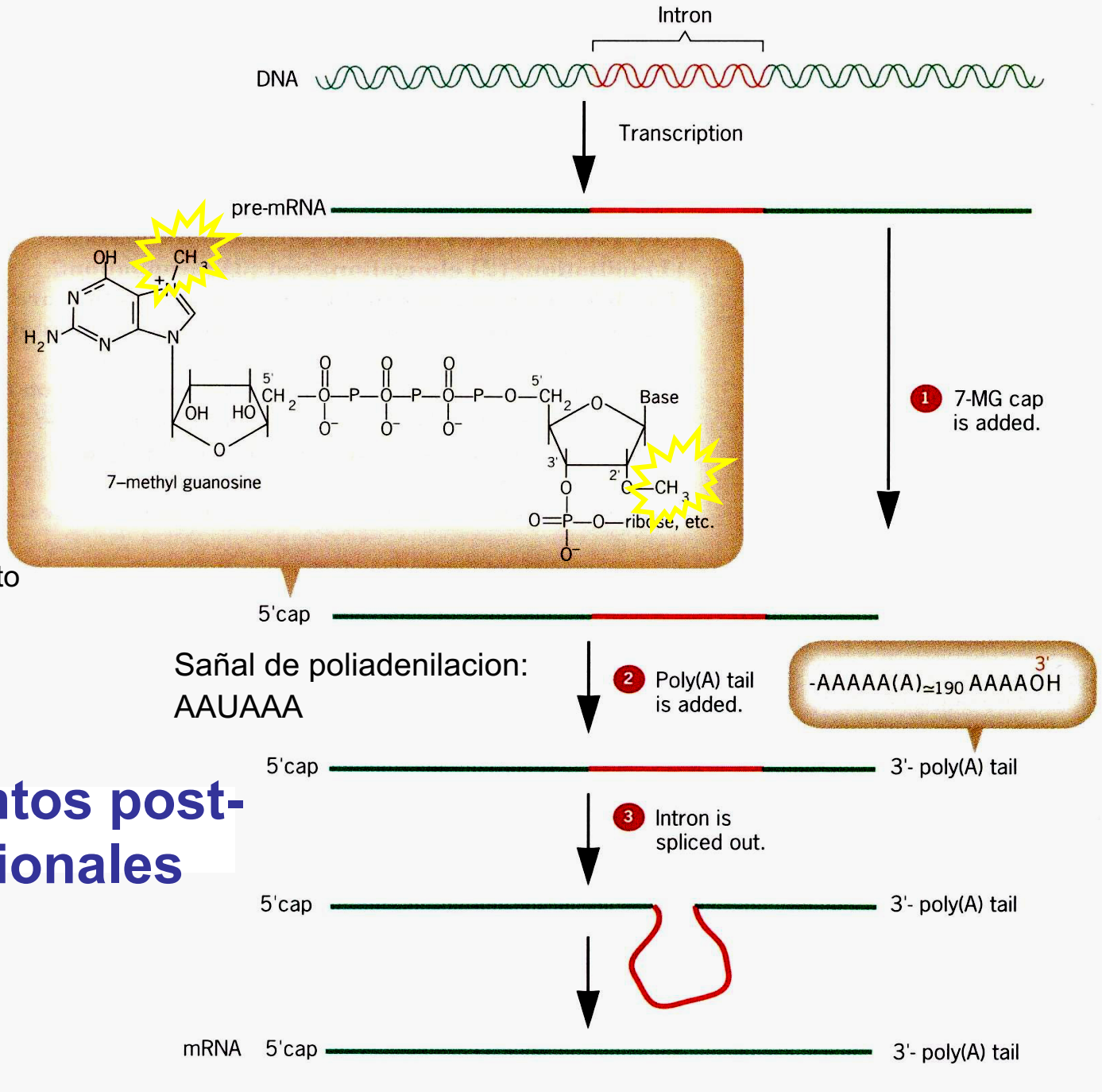
Regulación Post-transcripcional

Procesamiento de los mRNAs

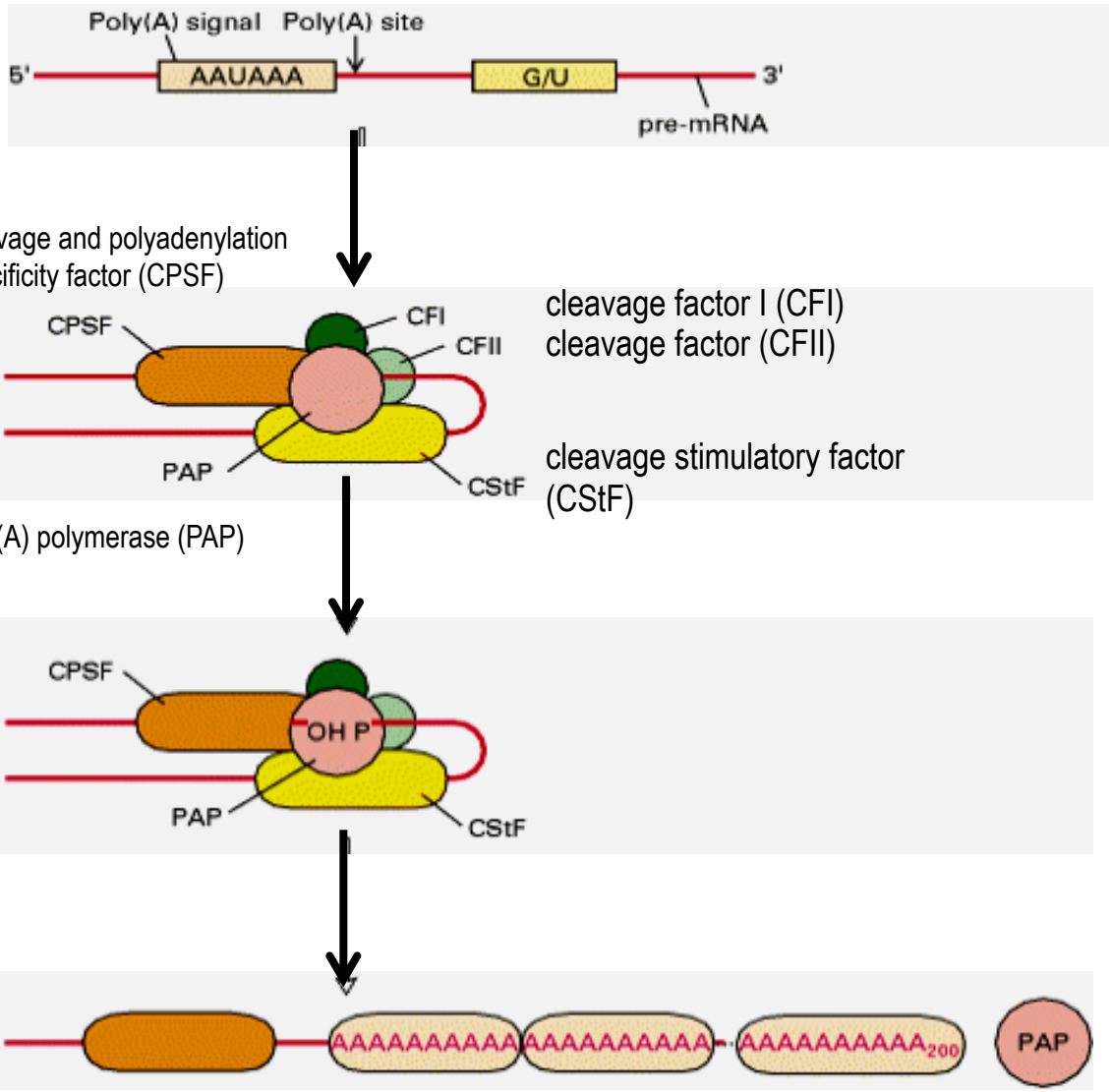
- Adición de capuchón (cap) en extremo 5'. (7-metilguanosina)
- Remoción de intrones. Splicing
- Edición de RNAs
- Poliadenilación (cola Poli A) en extremo 3'

Procesamientos post-transcripcionales

guanosina 5’ -5 -trifosfato

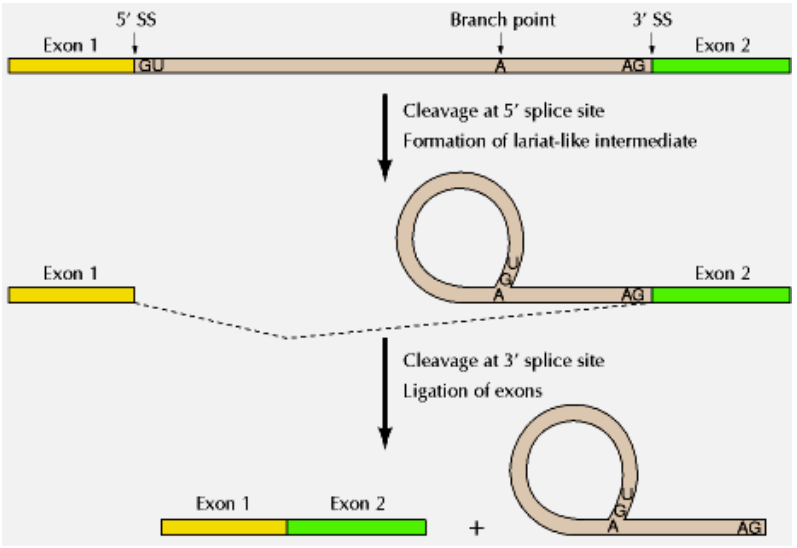


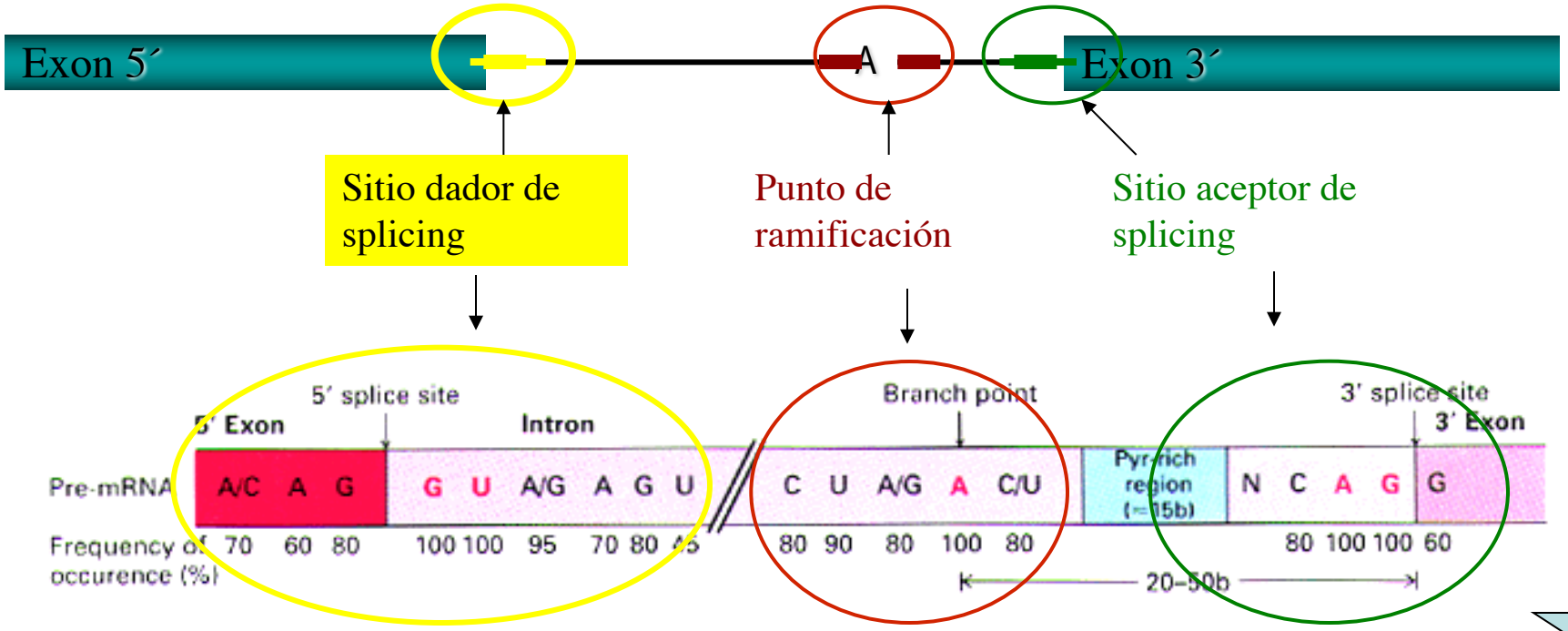
Poliadenilación



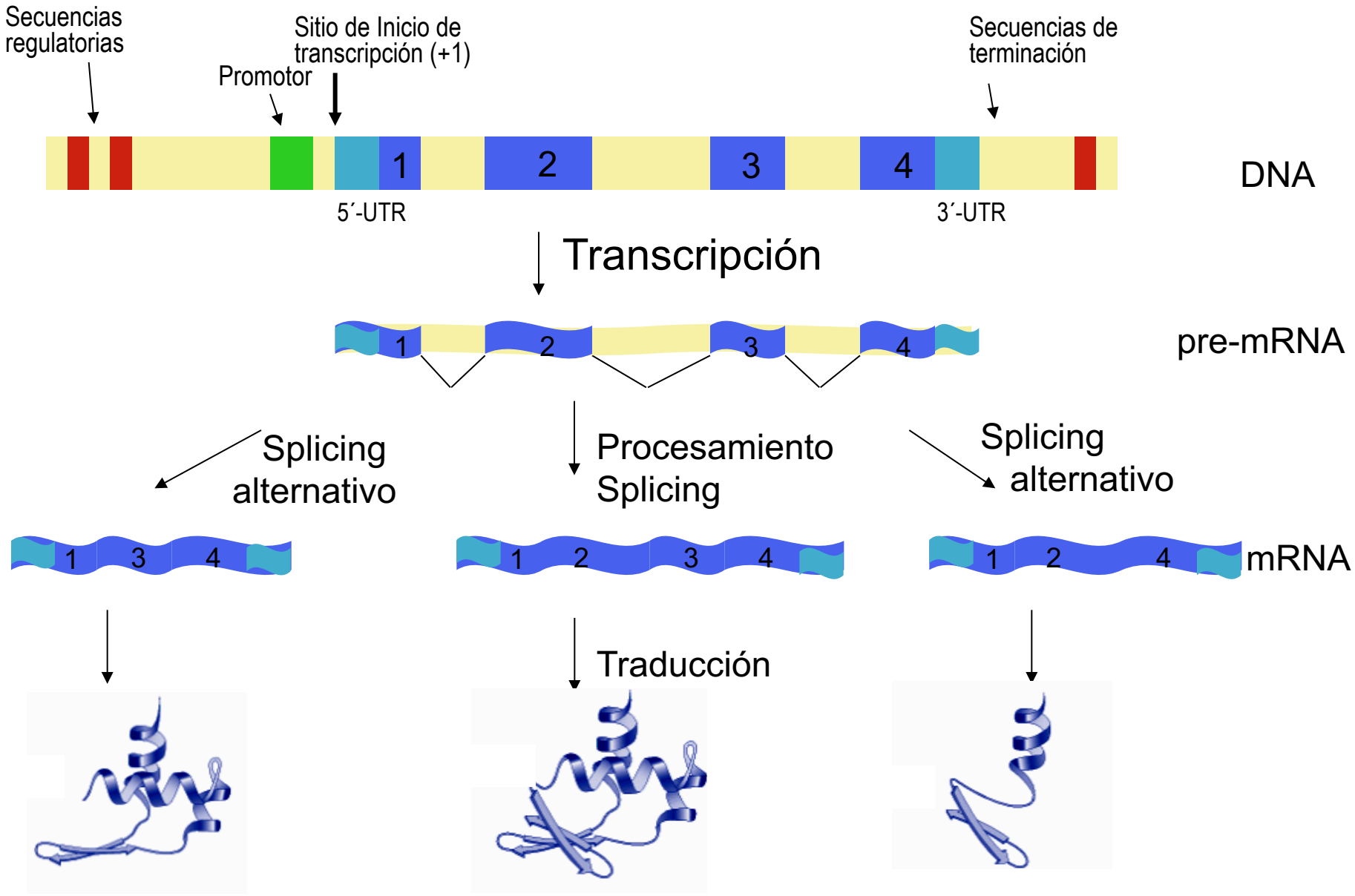
Secuencias de Splicing

Regla GT-AG: los intrones casi siempre comienzan con GT (o realmente GU y terminan con AG.

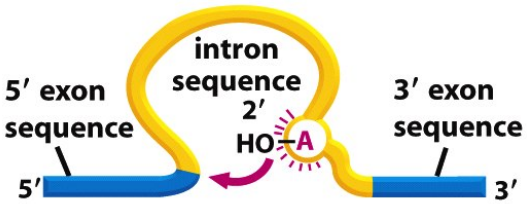




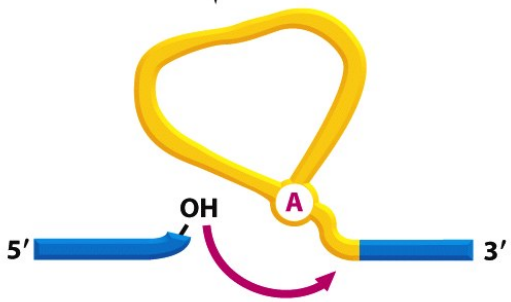
Regulación - variedad de productos por splicing alternativo



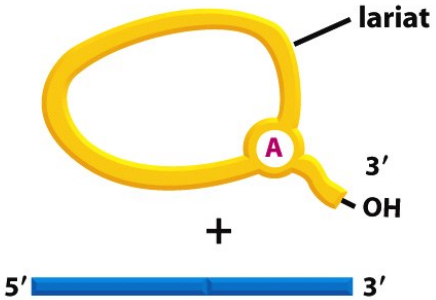
Mecanismo químico del splicing

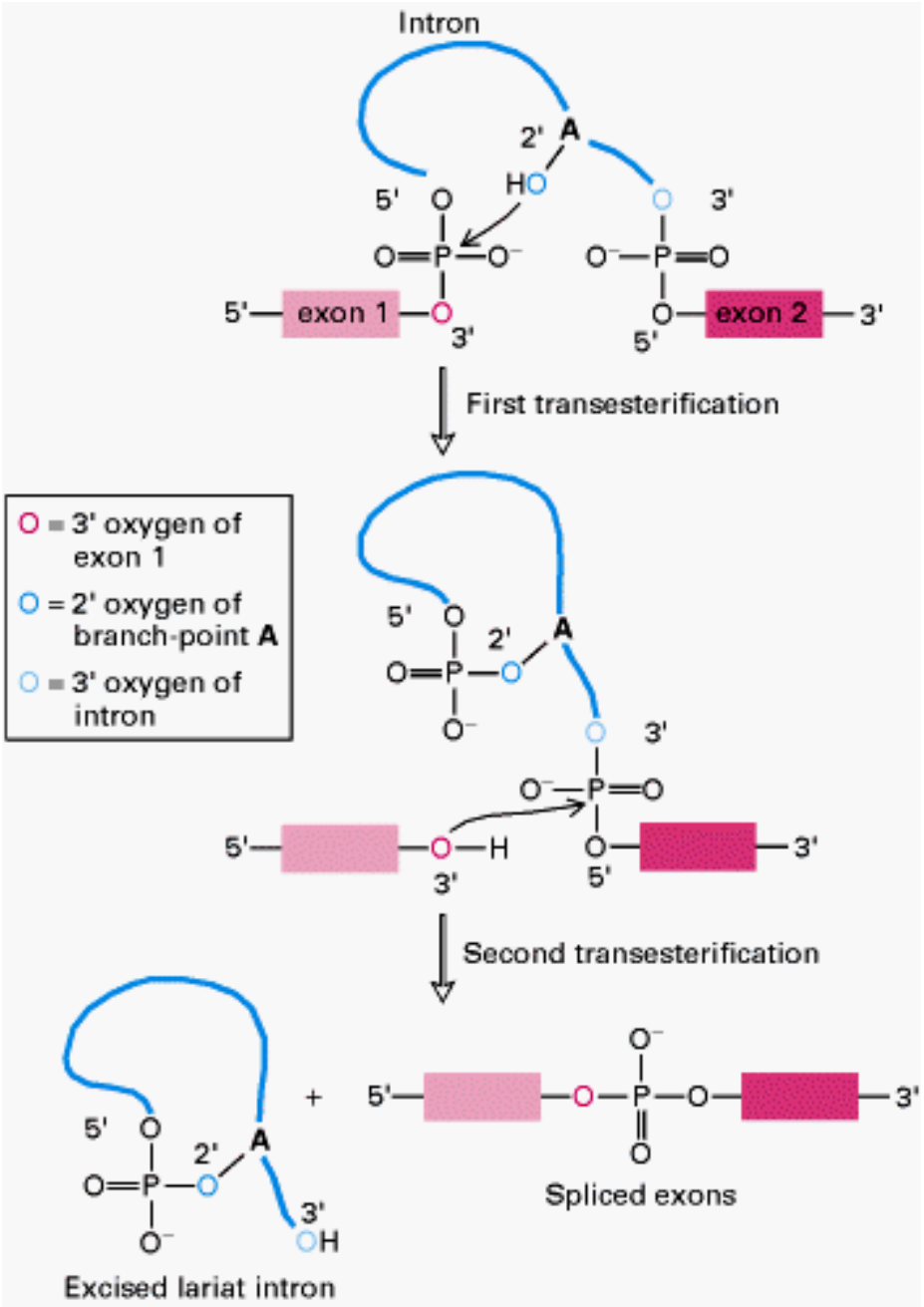


1. Corte en el sitio 5’ o dador de splicing



2. Corte en el sitio 3’ o aceptor de splicing

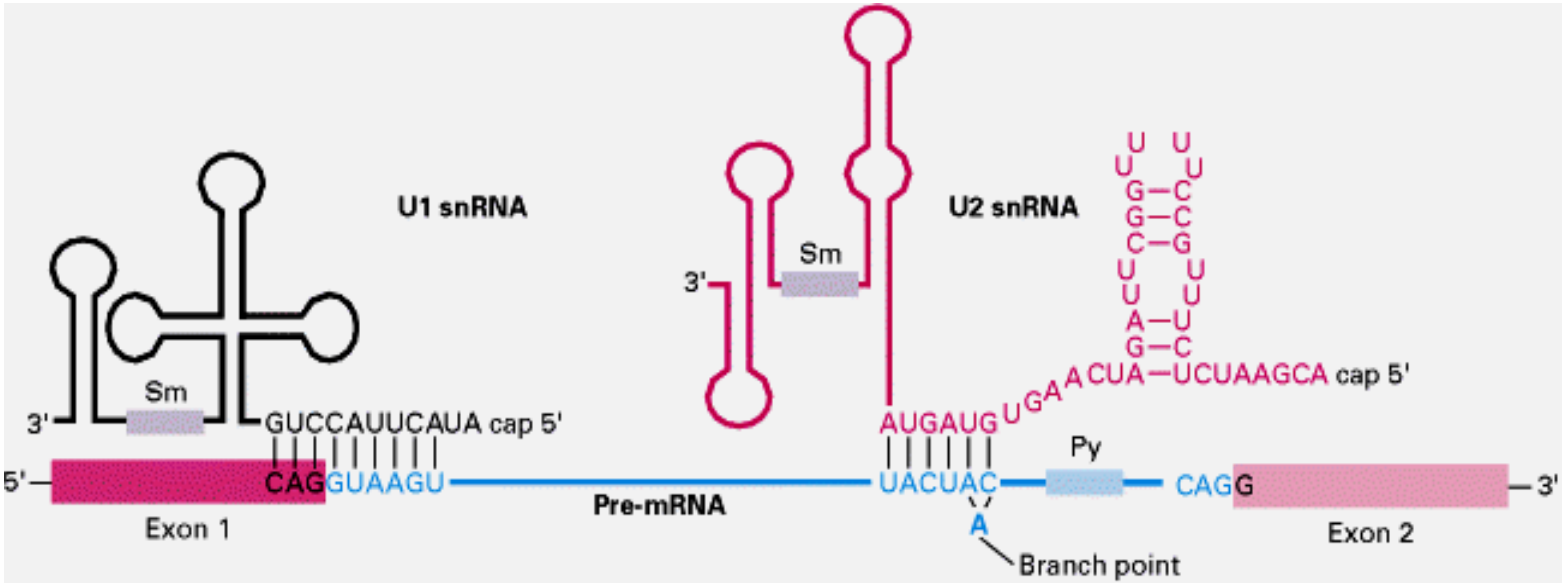




Dos reacciones de transesterificación

Mecanismo químico del splicing...

small nuclear RNAs (snRNAs)



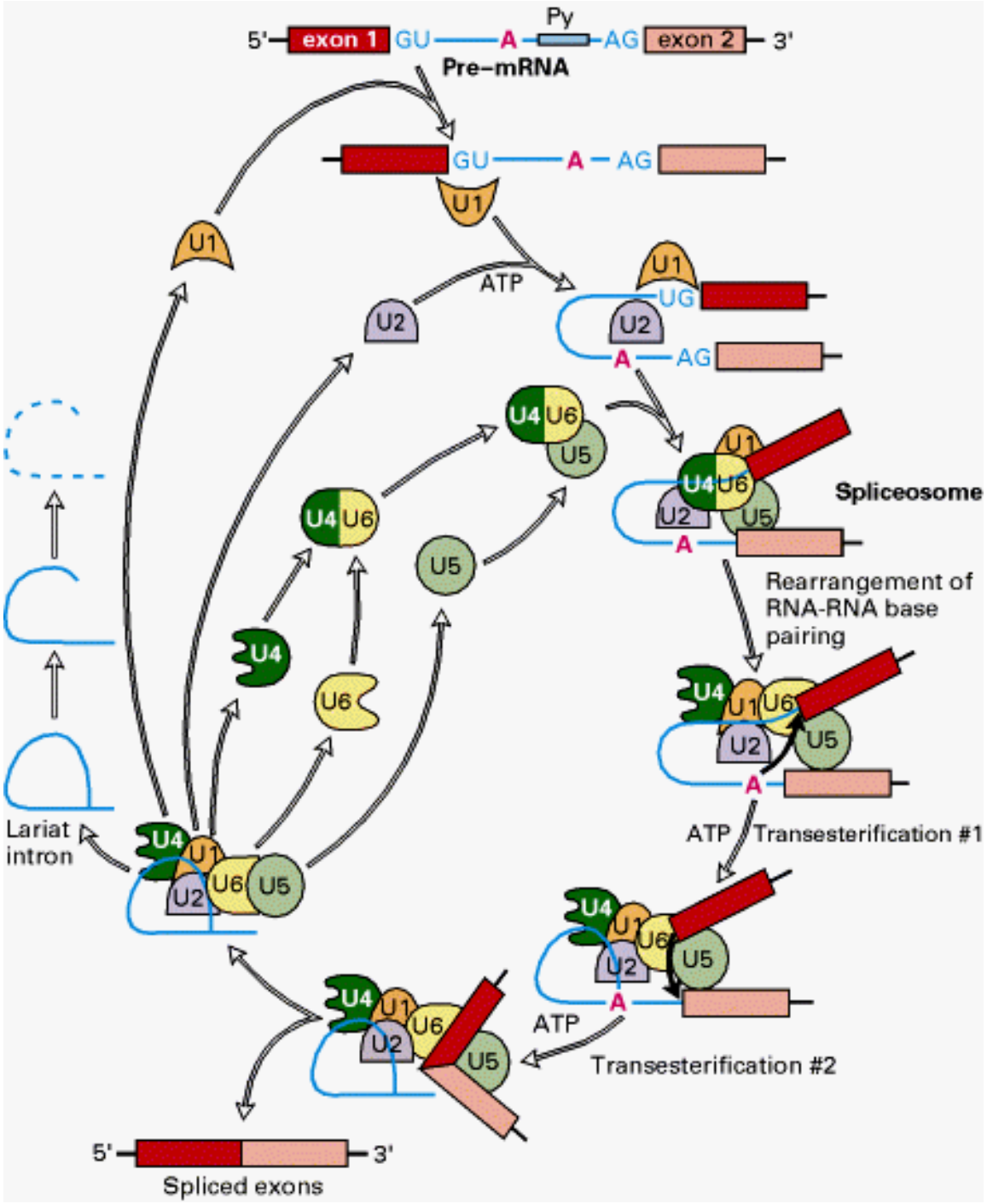
Spliceosomas

Interacciones RNA-RNA

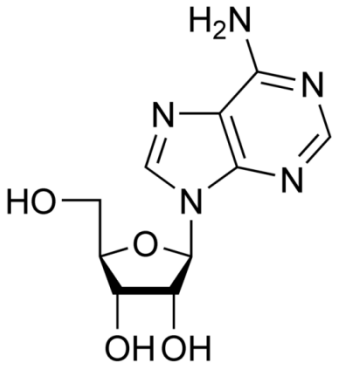
U6 snRNA / U4 snRNA.
U6 snRNA / U2 snRNA,

Etc...

Spliceosoma también interactúa con CTD de RNA pol II



Edición del RNA

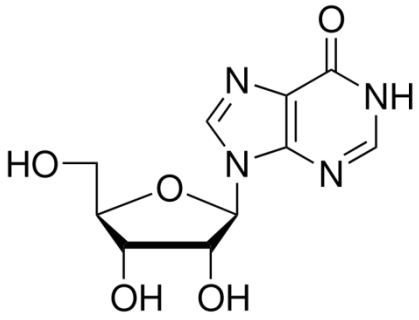


Adenosina

ADAR (1-3)

Adenosine deaminases that act on double- RNA

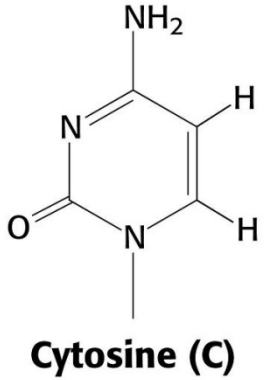
No requiere cofactores adicionales



Inosina

Pre-mRNA que generan sustituciones:

- receptor de glutamato
- receptor de serotonina 2C
- Autoedición de ADAR2

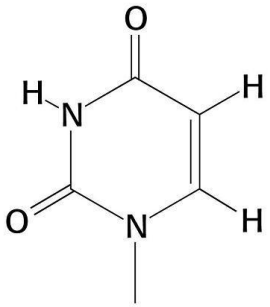


Cytosine (C)

CDAR

Cistosine deaminases that act on RNA

requiere cofactores adicionales

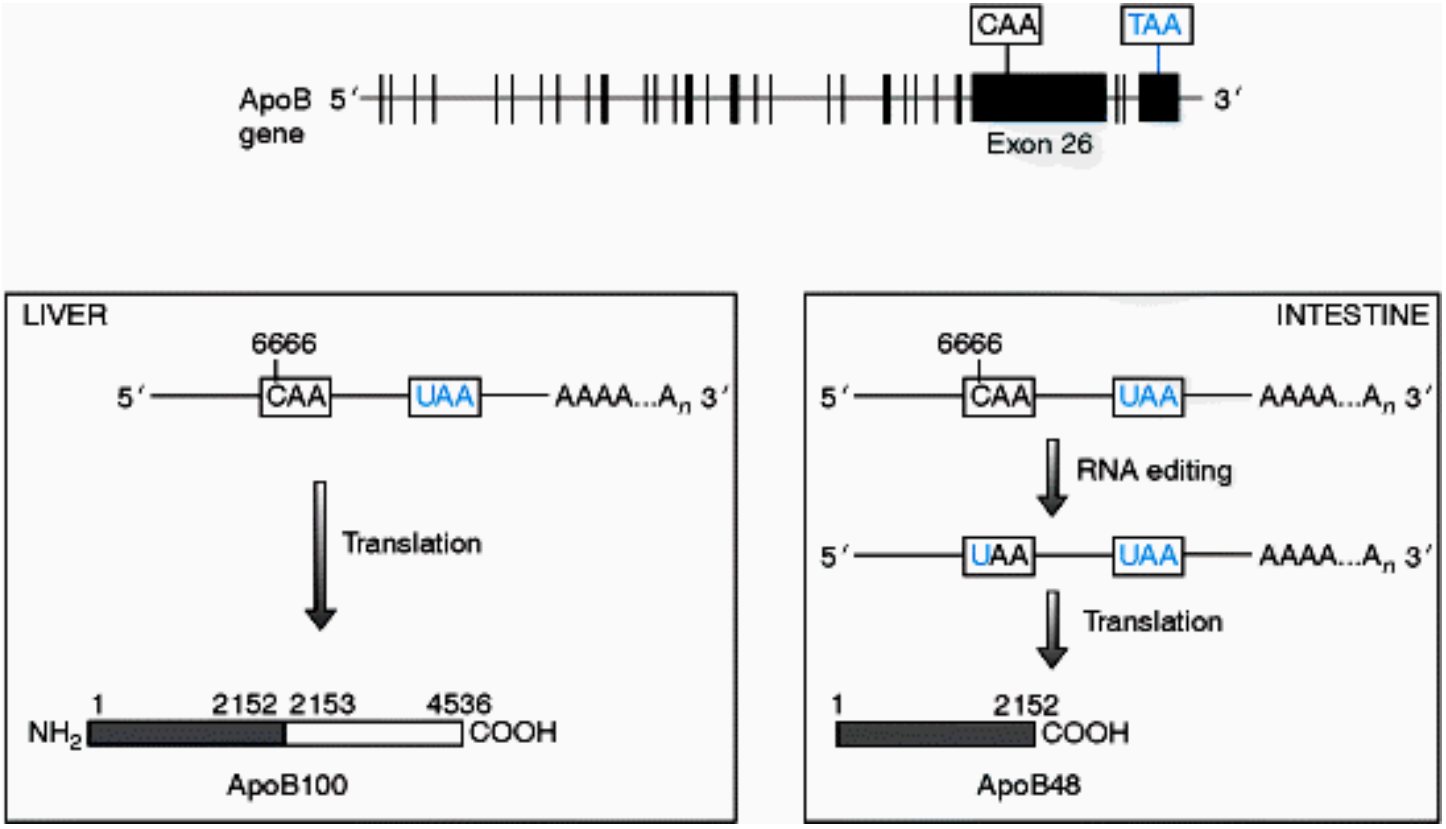


Uracil (C)

mRNA que generan codones de término:

- ApoB
- neurofibromatosis type1 (NF1)

Regulación post transcripcional. Edición del RNA



Producida en hígado.

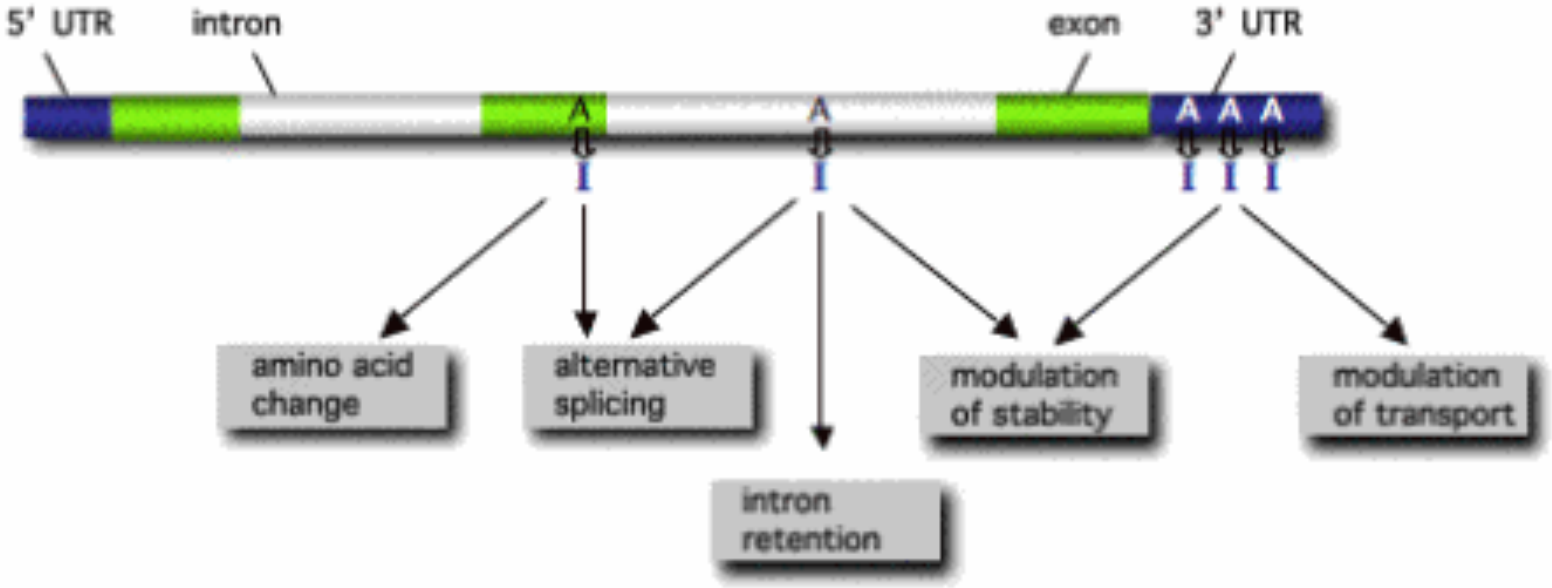
LDL presentes en todas las células.

Transporta lípidos y colesterol endógenos

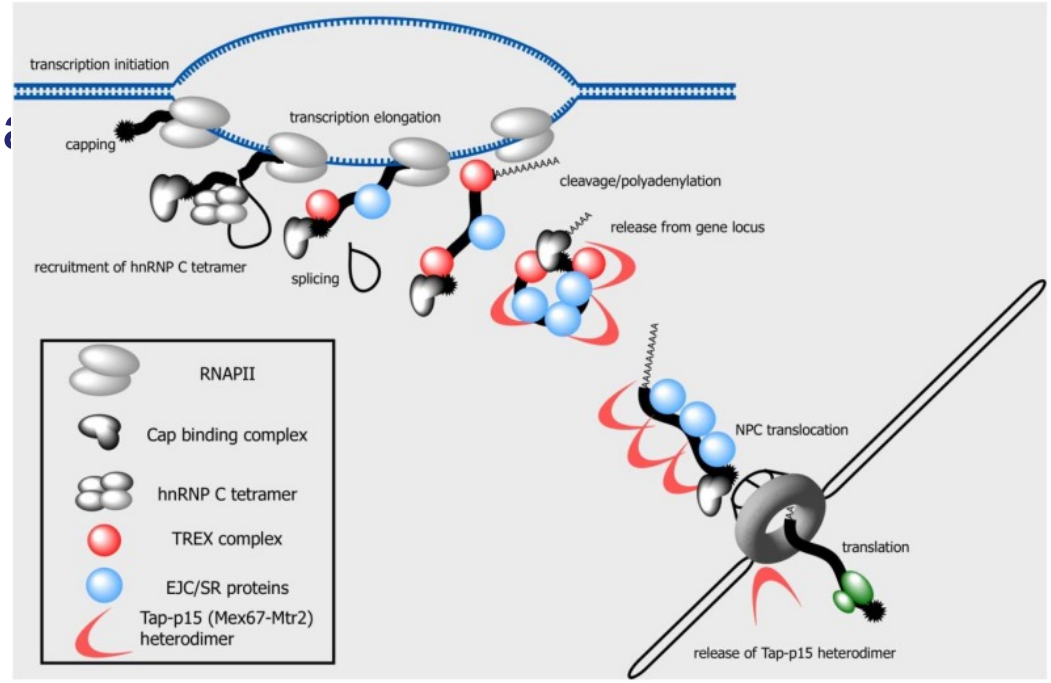
Producida en intestino delgado.

Transporta lípidos dietarios (QM), desde el intestino al hígado.

Regulación post transcripcional. Edición del RNA



Transporte de mRNAs a citoplasma



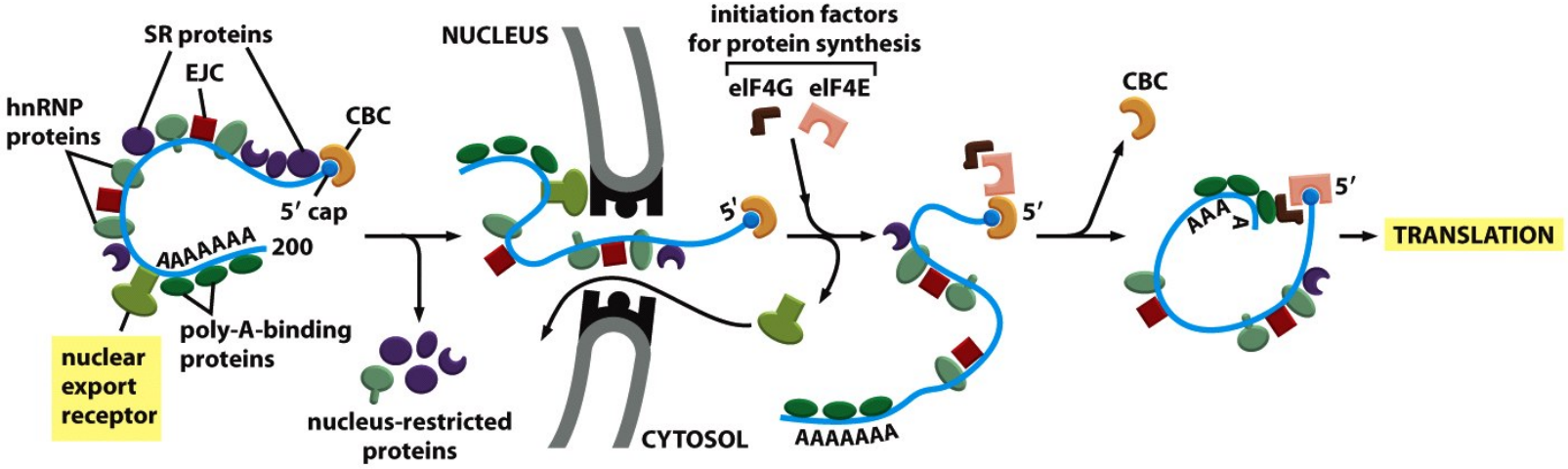
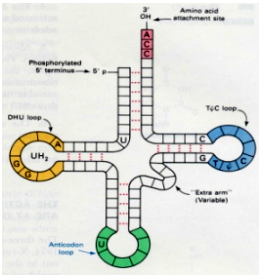


Figure 6-40 *Molecular Biology of the Cell* (© Garland Science 2008)

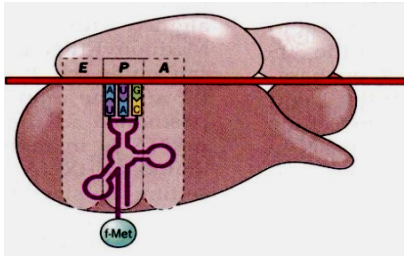
Maquinaria de traducción de proteínas.

•RNA mensajero (RNAm)

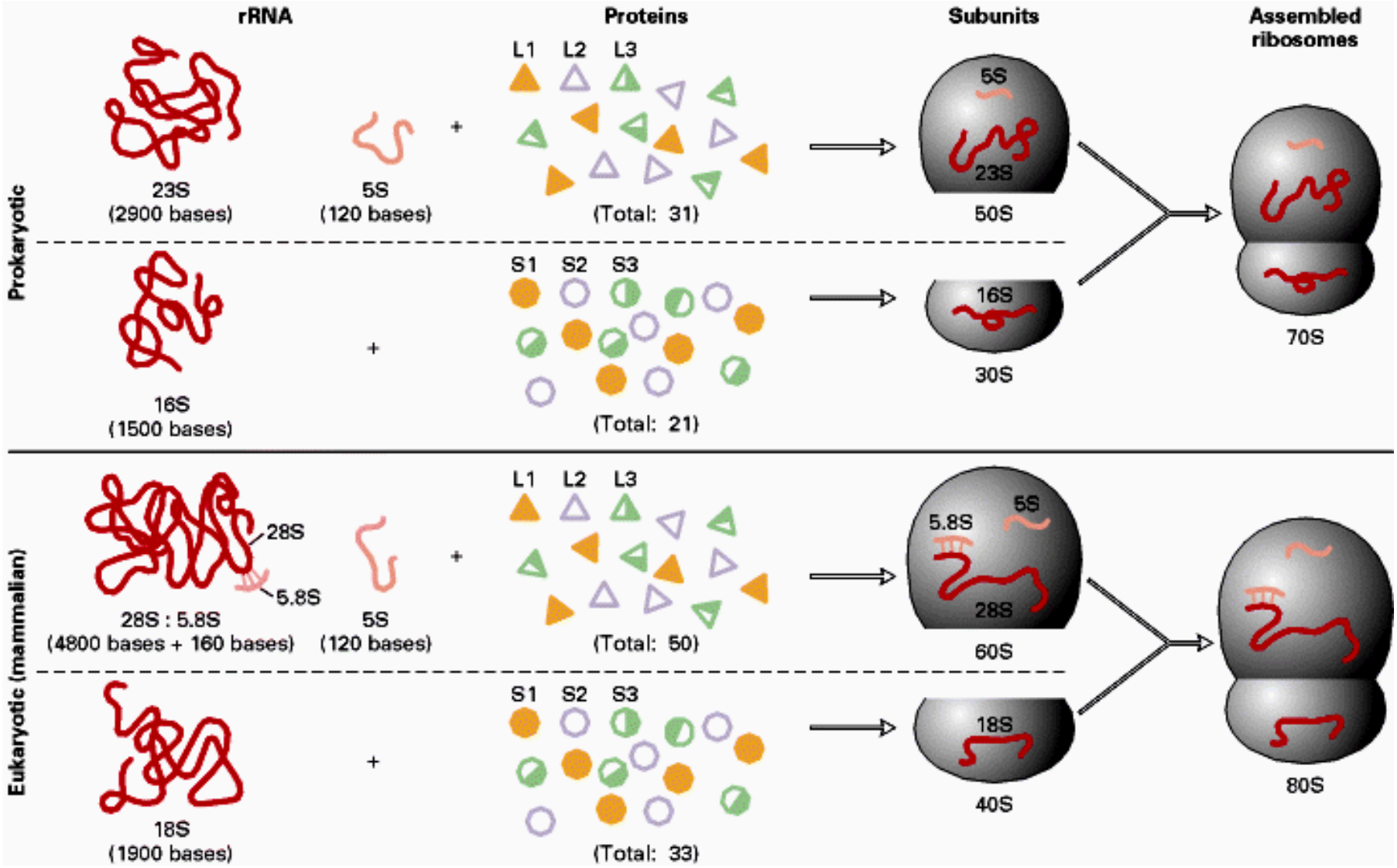
•RNA de transferencia (RNAt), unido a cada uno de los 20 aminoácidos.



•Ribosomas: RNAs y proteínas ribosomales.



Estructura general de los ribosomas de procariontes y eucariontes.



*El código genético: c*orrespondencia entre los codones (grupo de tres nucleótidos) en el DNA (o su transcrito) y la secuencia de aminoácidos *en una proteína.*

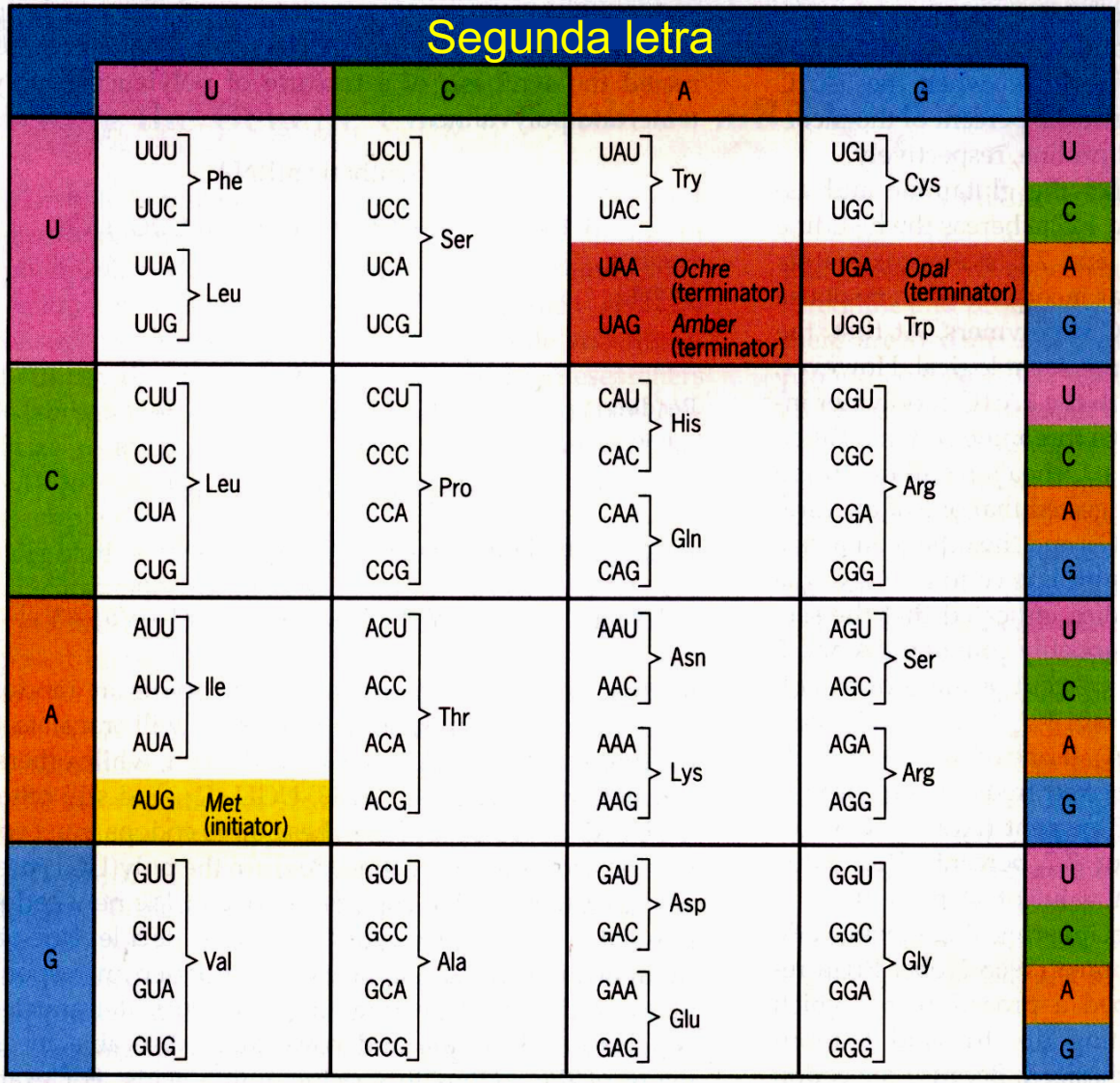
Codones posibles: $4^3 = 64$

Código genético es degenerado

Código es Universal……

5´- UAU | CUA | UCU | AUC | UAU | CUA | UCU | AUC - 3´

Tyr - Leu - Ser - Ile - Tyr - Leu - Ser - Ile



Código genético es:

- Universal
- redundante

Codón de inicio

Codón de término

GCA GCC GCG GCU	AGA AGG CGA CGC CGG CGU	GAC GAU	AAC AAU	UGC UGU	GAA GAG	CAA CAG	GGA GGC GGG GGU	CAC CAU	AUA AUC AUU	UUA UUG CUA CUC CUG CUU	AAA AAG	AUG	UUC UUU	CCA CCC CCG CCU	AGC AGU UCA UCC UCG UCU	ACA ACC ACG ACU	UGG	UAC UAU	GUA GUC GUG GUU	UAA UAG UGA
Ala	Arg	Asp	Asn	Cys	Glu	Gln	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val	stop
A	R	D	N	C	E	Q	G	H	I	L	K	M	F	P	S	T	W	Y	V	

Marco de Lectura

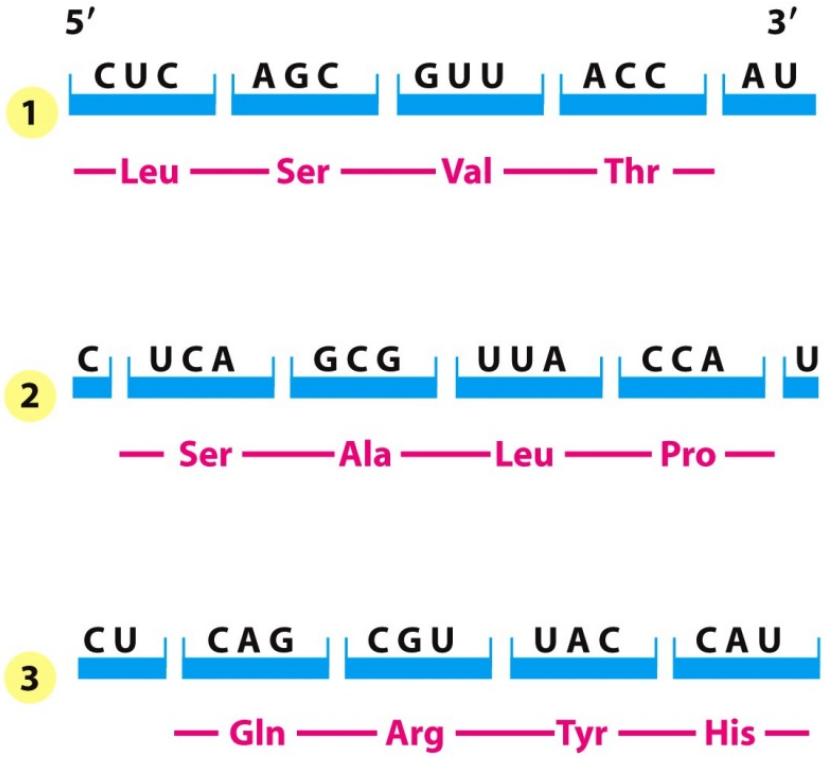
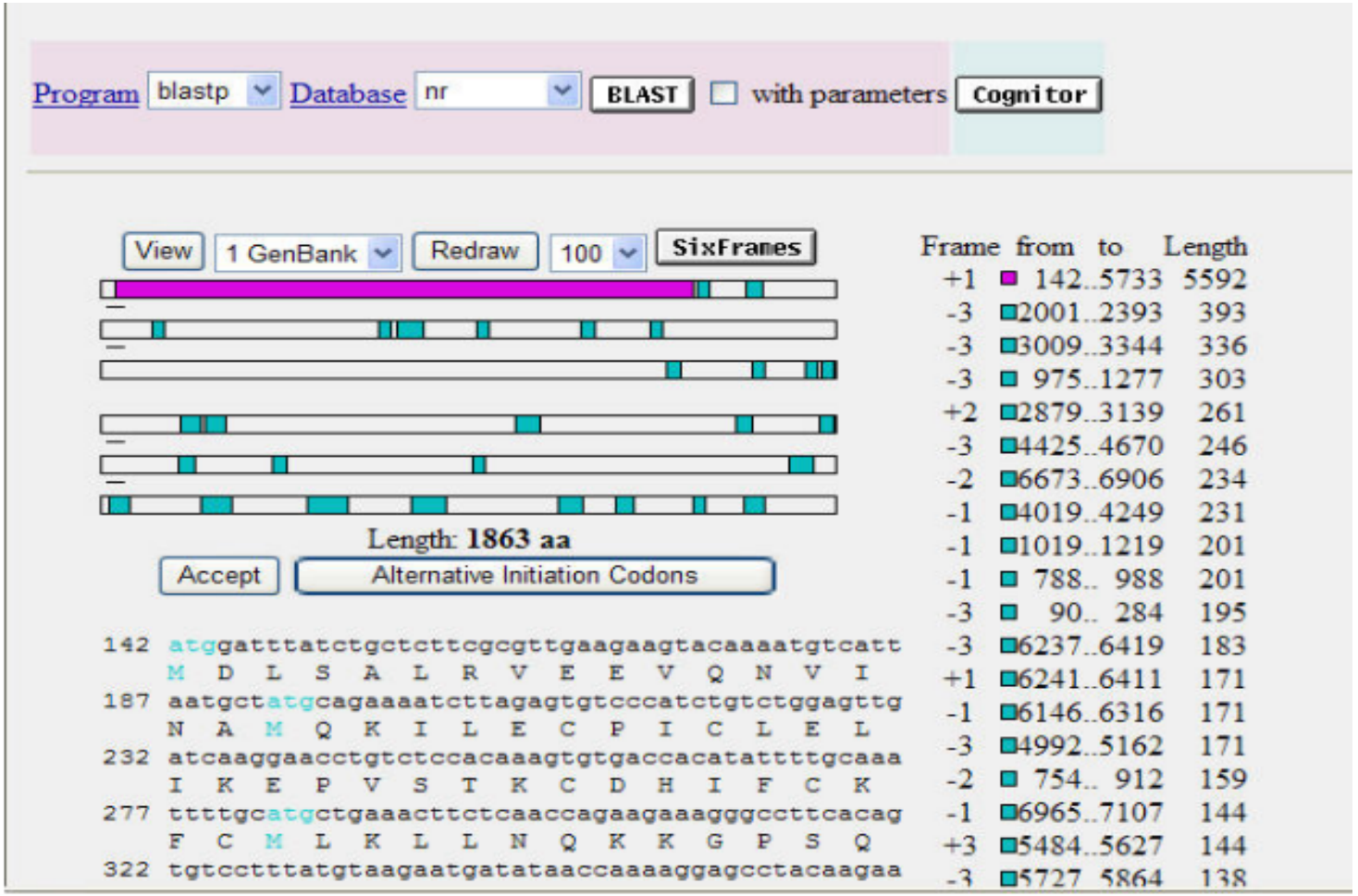


Figure 6-51 *Molecular Biology of the Cell* (© Garland Science 2008)

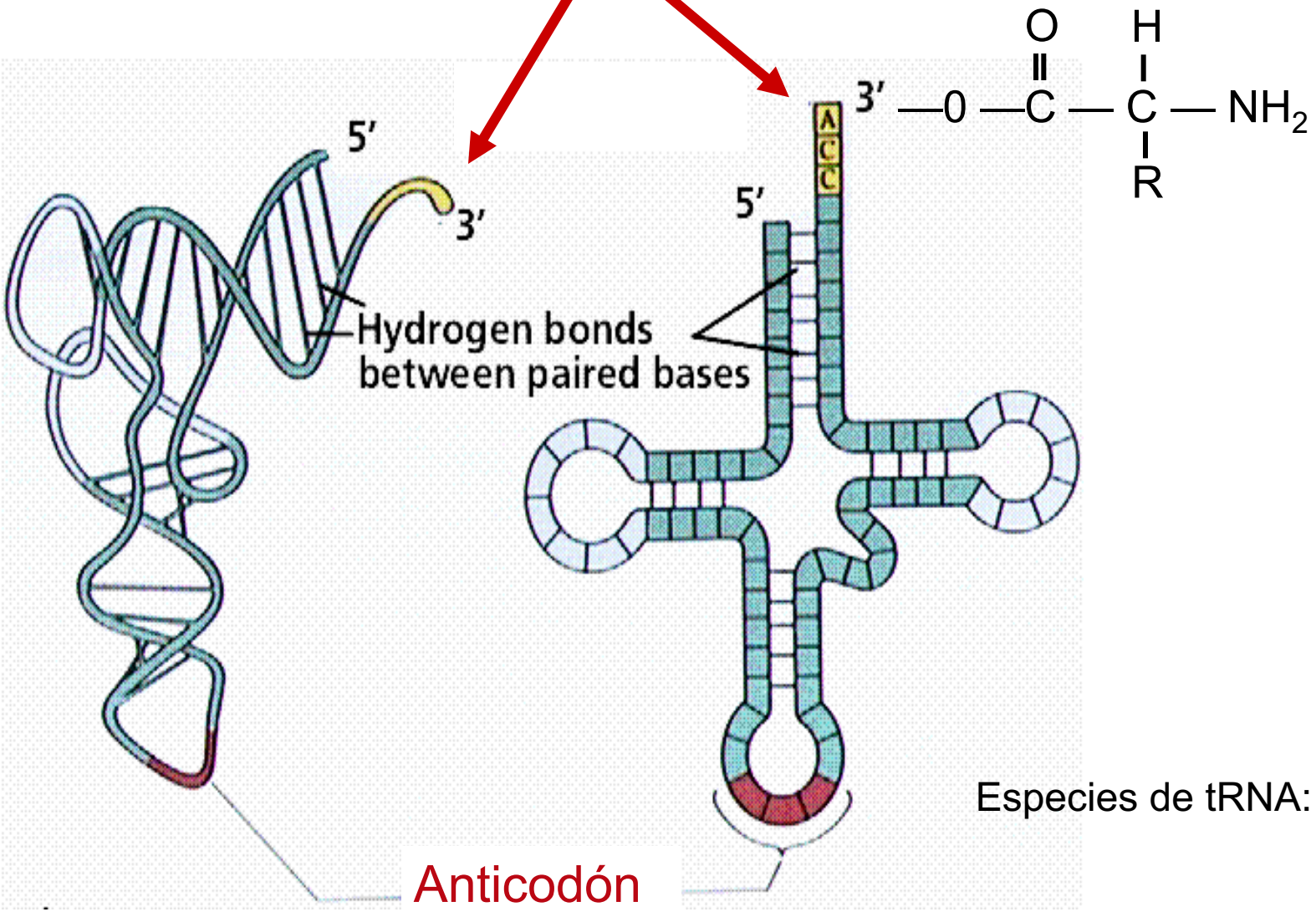
http://www.ncbi.nlm.nih.gov/orffinder/

Marco de Lectura abierto, open reading frame



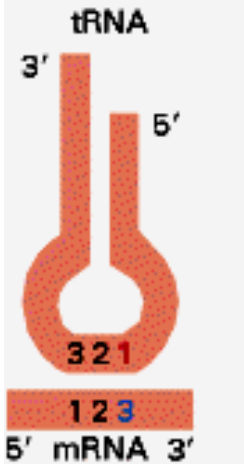
Un mRNA podria ser leído en cualquiera de los tres marcos de lectura, sin embargo sólo uno permite generar la proteína codificada.

Sitio de unión del aminoácido



Especies de tRNA: 41 to 55

Balanceo



anticodón

3’—A—A—G—5’

5’—U—U—U—3’

codón

anticodón

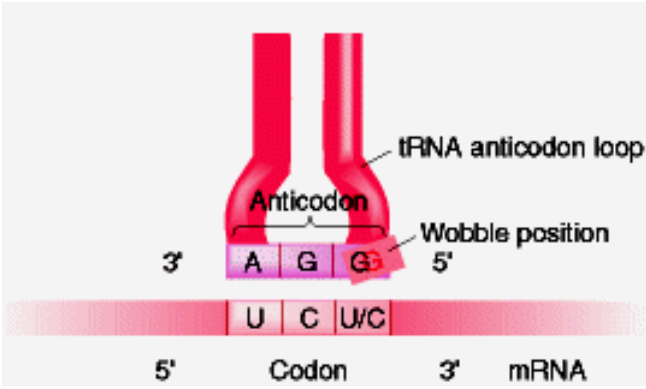
3’—A—A—G—5’

5’—U—U—C—3’

codón

Codigo genetico

Base en la primera posición del anticodón	Base(s) reconocida(s) en la tercera posición del codón
U	A o G
C	Solo G
A	Solo U
G	C o U
I	U, C o A



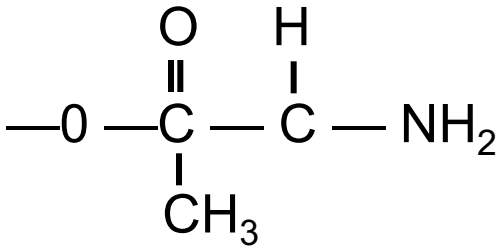
Reglas de la teoría del Balanceo.

Una inosina en la posición de balanceo del tRNA puede leer tres codones diferentes

G o U en la posición de balanceo en el tRNA pueden leer dos codones.

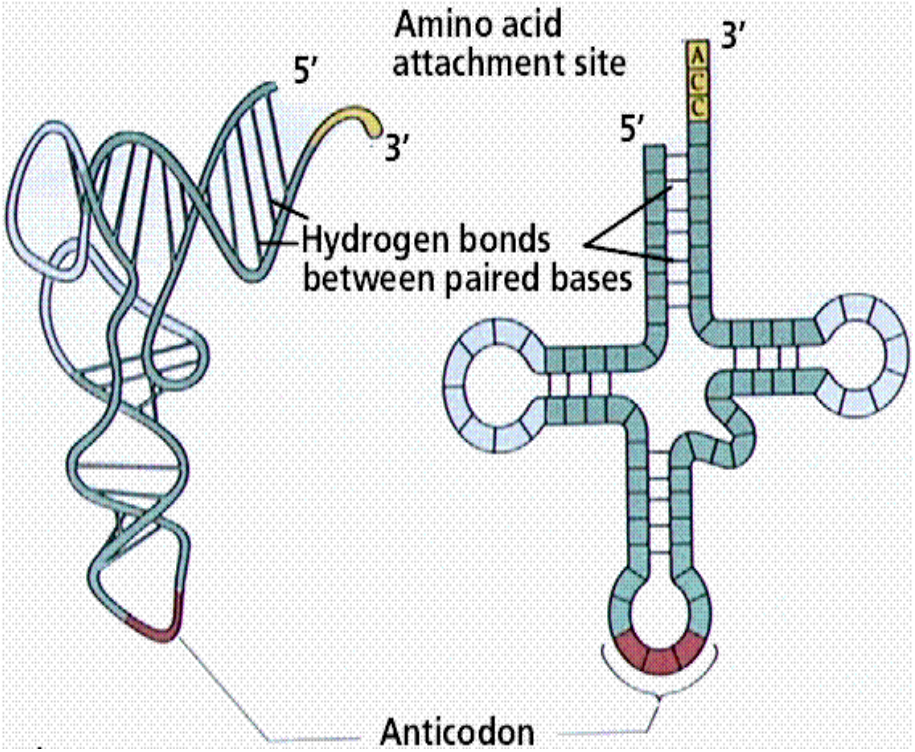
Aminoacil tRNA sintetasa

aminoácido + ATP + tRNA + H_20

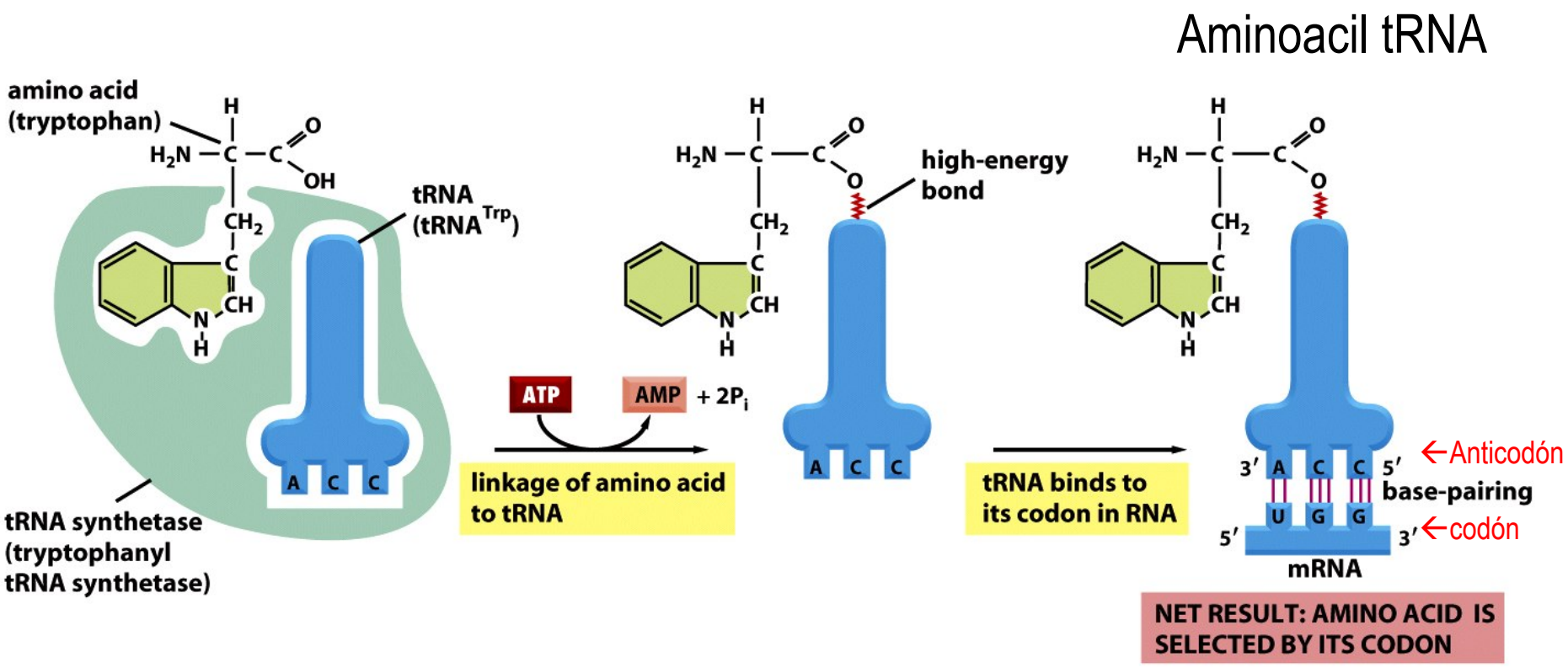


aminoacil–tRNA + AMP + 2Pi

Aminoacil tRNA sintetasas.
Una por cada aminoácido



Aminoacil tRNA sintetasa



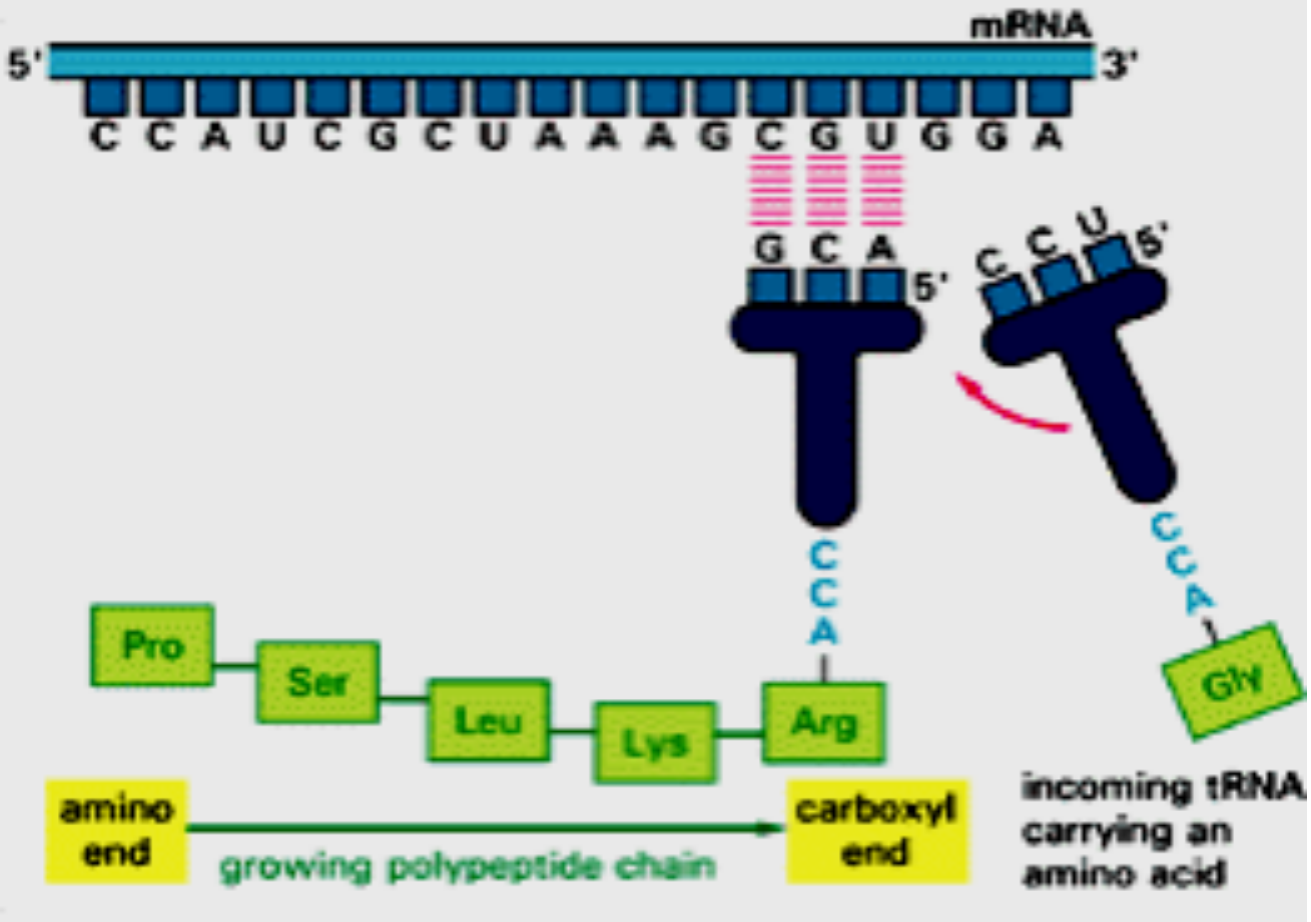
El código genético se logra traducir correctamente mediante la participación de "adaptadores" en el proceso de traducción:

1- Enzima aminoacil-tRNA sintetasa

2- Aminoacil- tRNA

Un error puede resultar en la incorporación de un aminoácido equivocado en la síntesis de la proteína

Codones, tripletes y anticodones

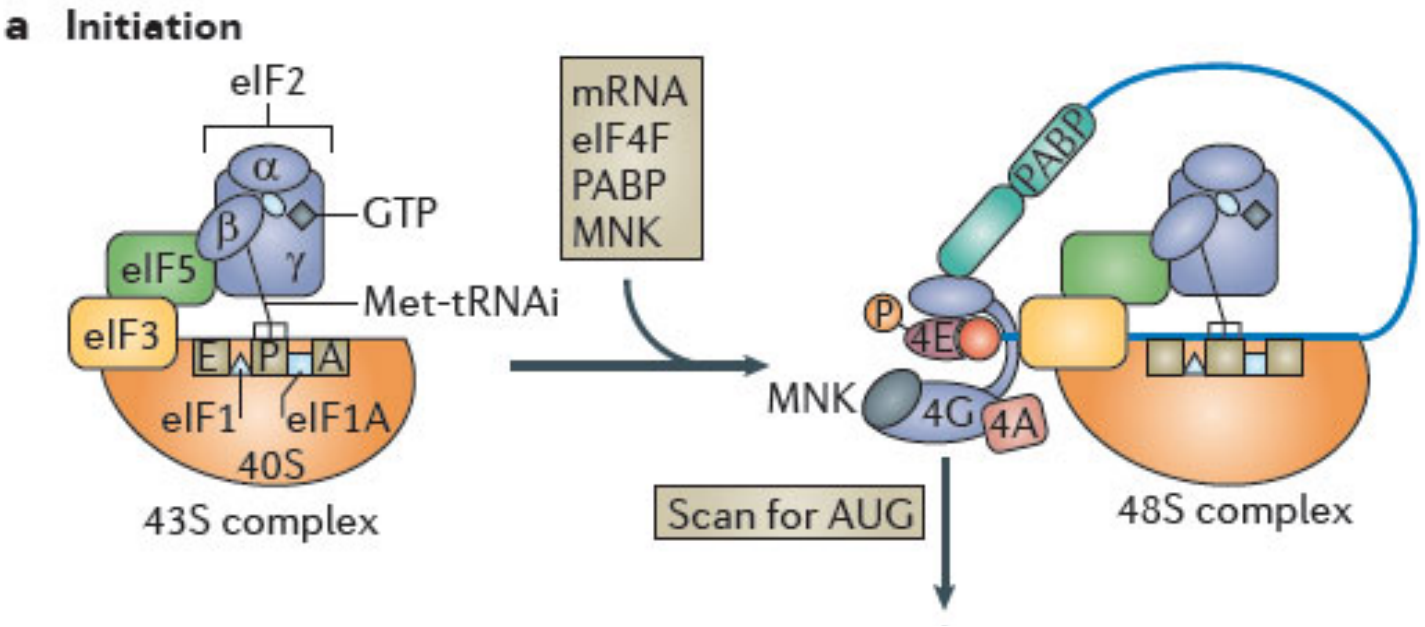


El codón es reconocido por el anticodón del tRNA

La traducción procede en sentido 5’ - 3’ del mRNA

La traducción procede desde el extremo amino al carboxilo terminal del polipéptido

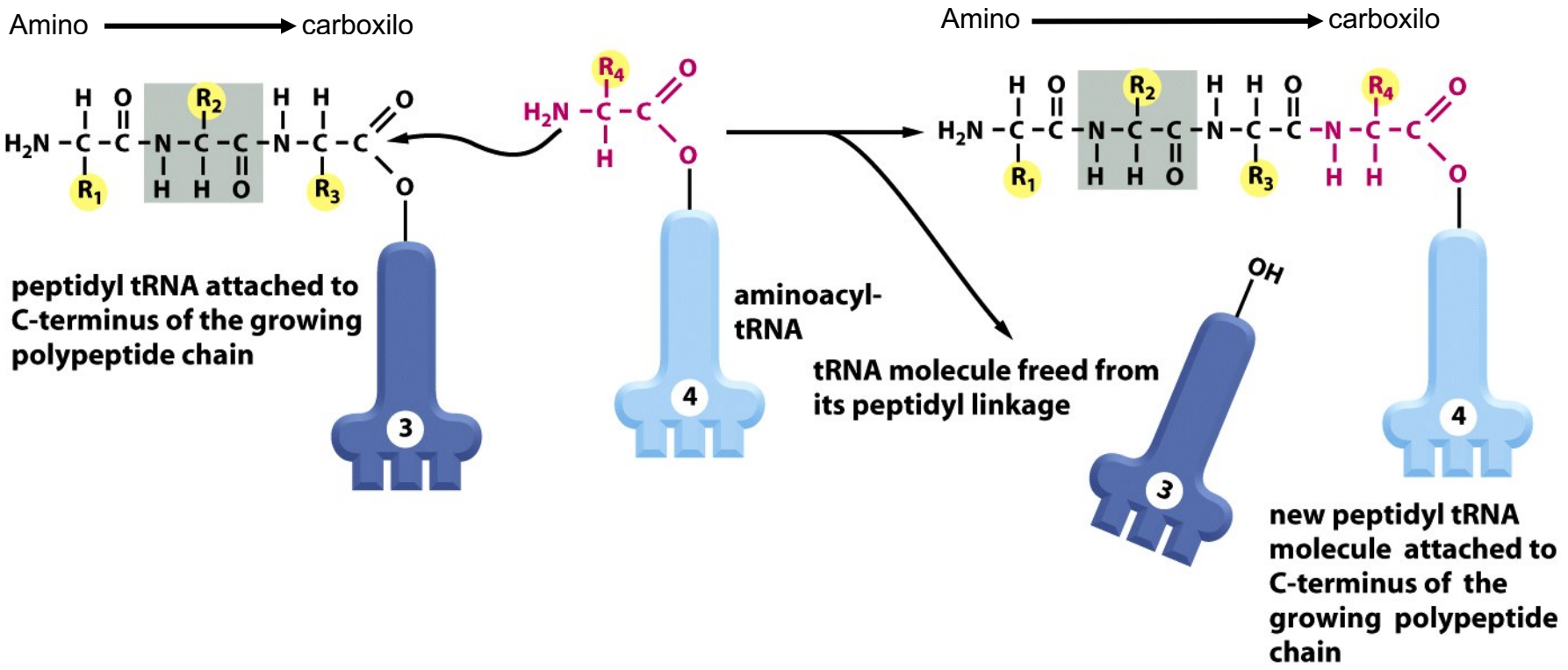
La síntesis proteica se realiza en tres etapas.



- Reconocimiento entre mRNA y ribosoma: subunidad pequeña 40S (eIF4) interactúa con 5´CAP metilado.
- Reconocimiento de secuencia AUG (KOZAC: 5'-ACCAUGG-3')
- Unión de met–$tRNA_i^{met}$ en AUG
- Unión de subunidad 60S con la formación del complejo de inicio *80S*

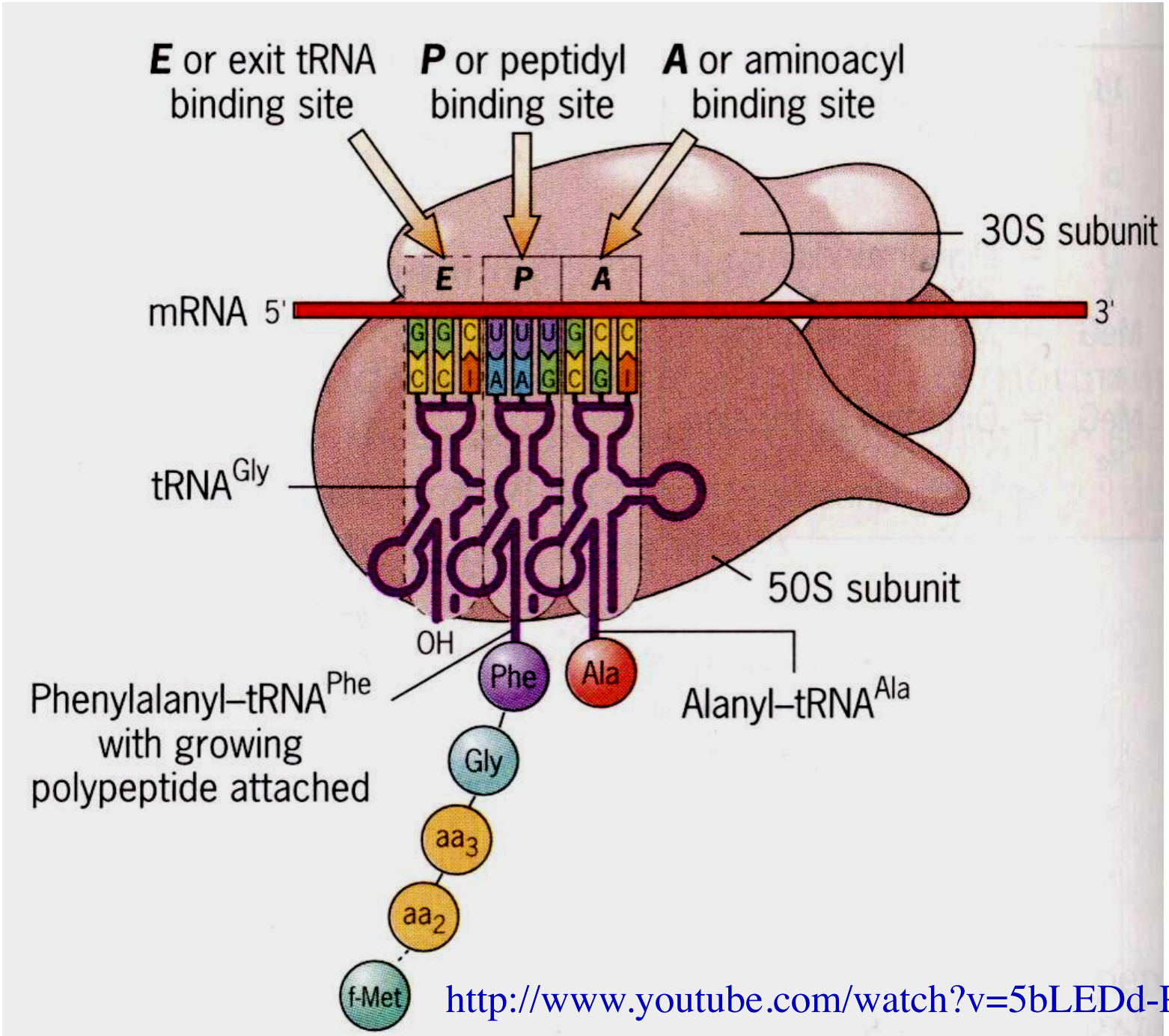
Nature Reviews Microbiology 9, 860-875 (December 2011)

Enlace peptídico

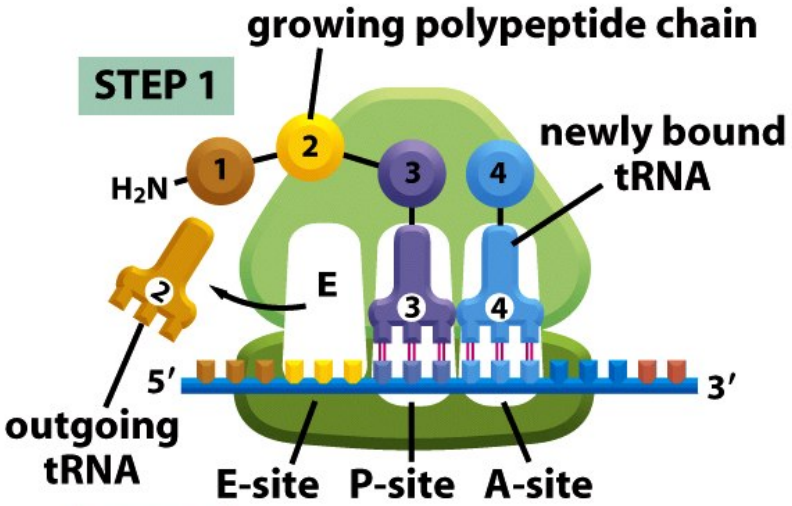


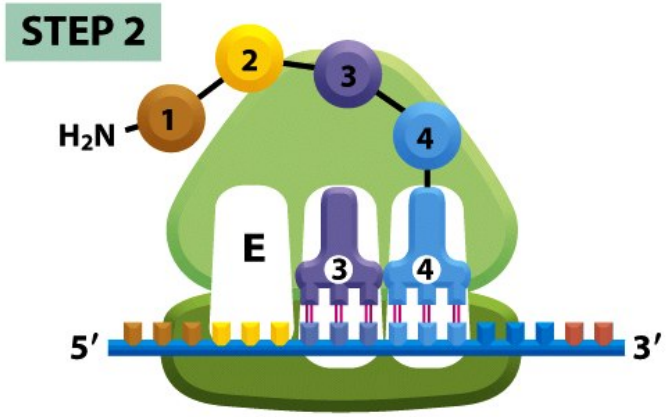
El crecimiento de una cadena polipeptídica ocurre por adición de un aminoácido en el extremo carboxilo mediante la formación de un enlace peptídico.

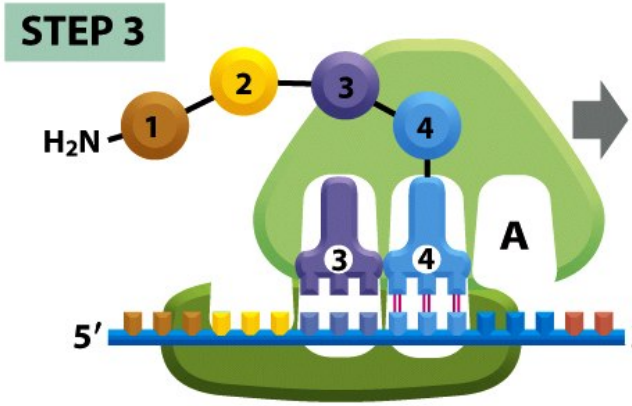
Este enlace es energéticamente favorable solo cuando el extremo carboxilo terminal (COO^-) se encuentra unido covalentemente a un tRNA.

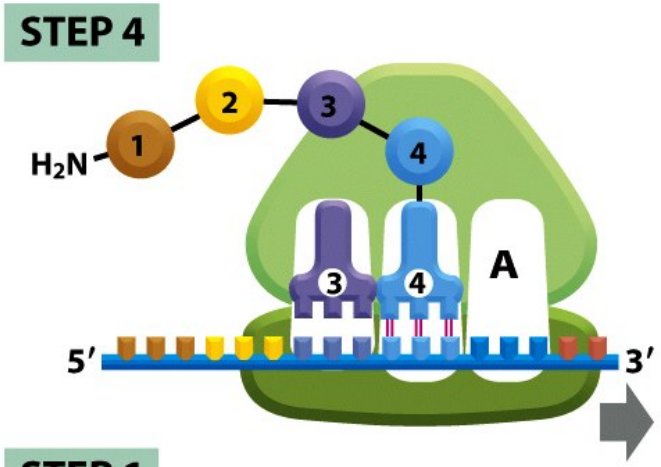


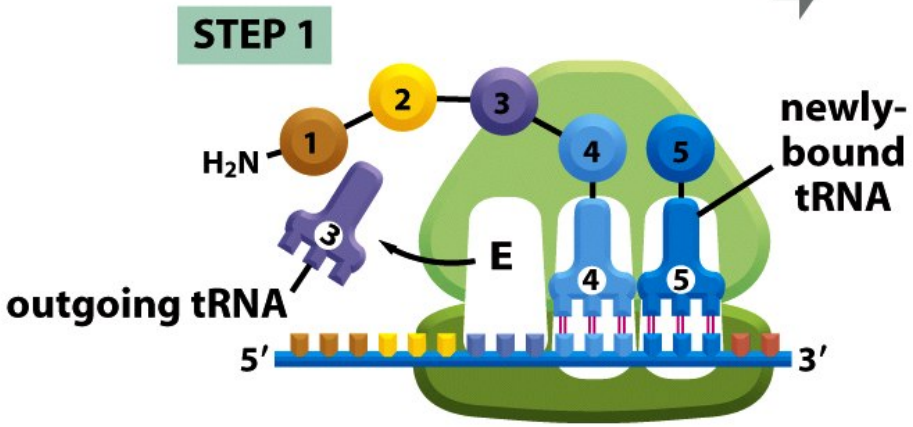
http://www.youtube.com/watch?v=5bLEDd-PSTQ









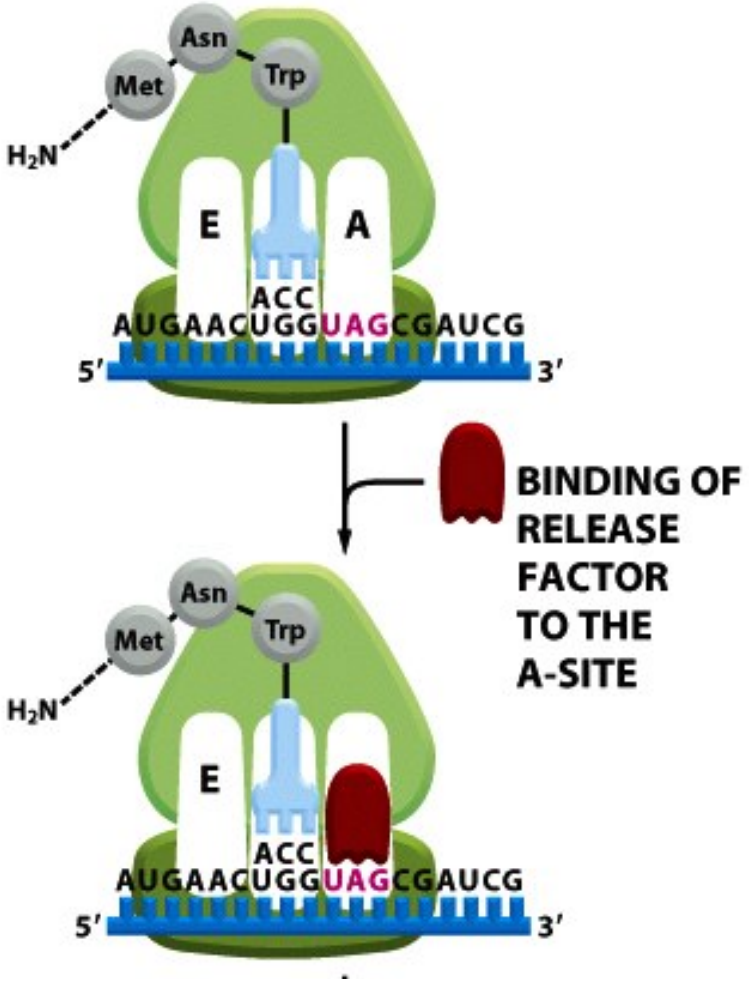


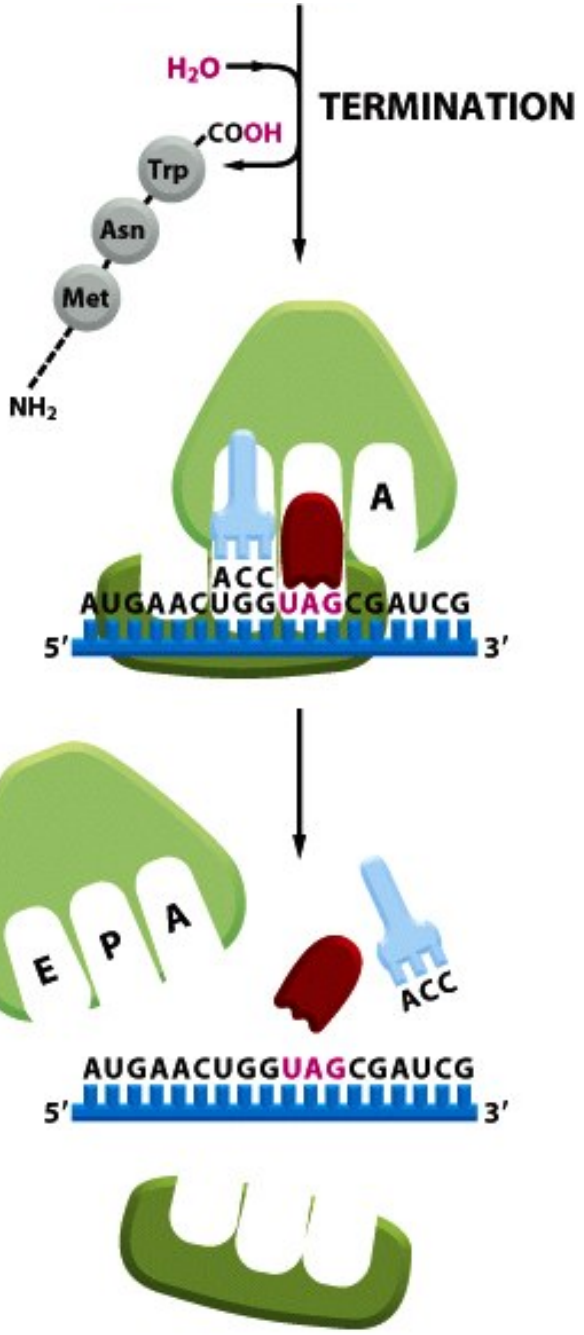
•Cada aminoacil tRNA ingresa al sitio A por complementaridad de bases entre el codón del mRNA y el anticodón del tRNA.

•Transferencia del polipeptido desde el peptidil t-RNA al aminoacil tRNA en el sitio A (Peptidil transferasa que reside en la subunidad grande).

•El polipéptido en crecimiento permanece siempre unido al complejo en el sitio P de la subunidad grande del ribosoma

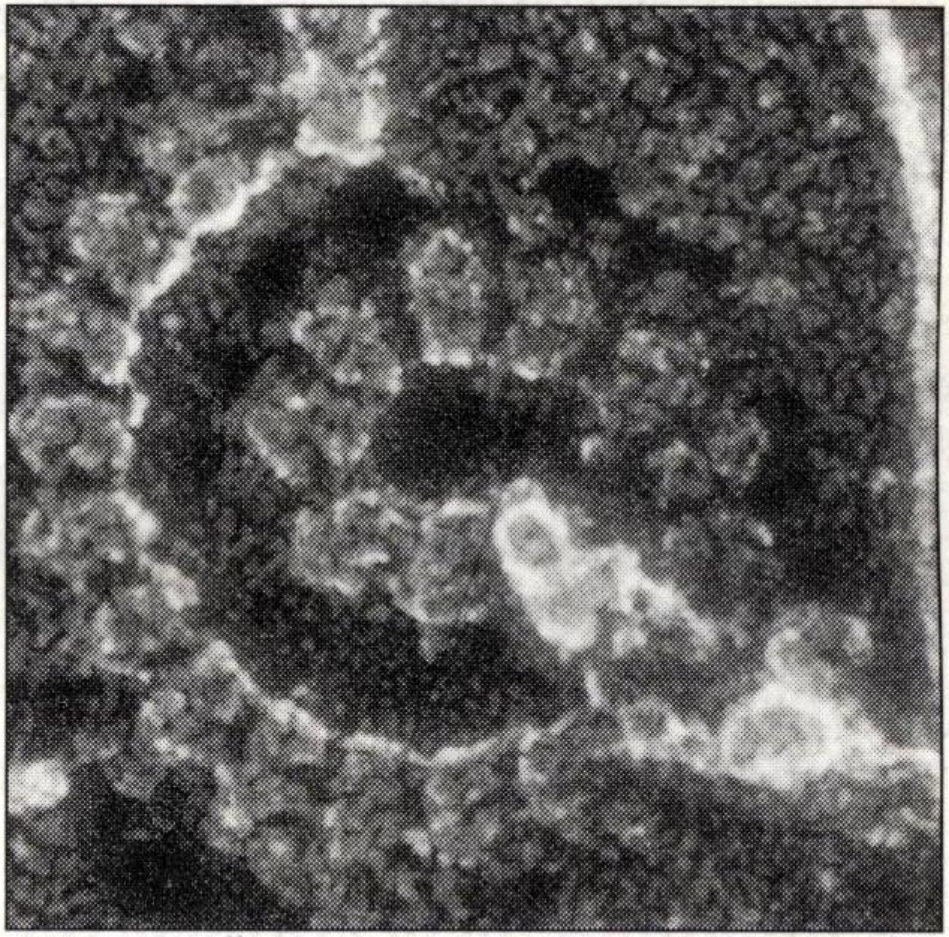
Terminación de traducción

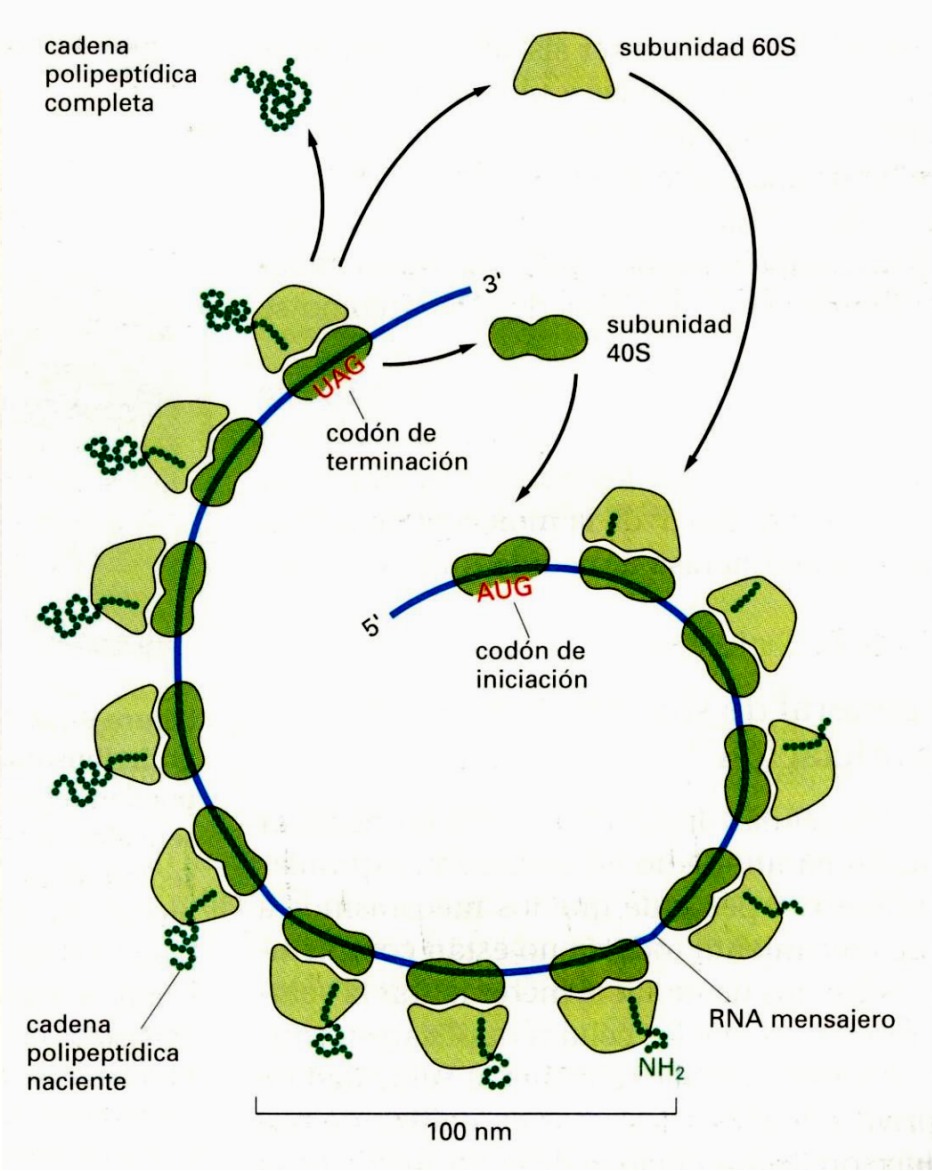




•La traducción termina cuando se une al sitio A un "factor de liberación". La proteína se suelta y se desarma el complejo.

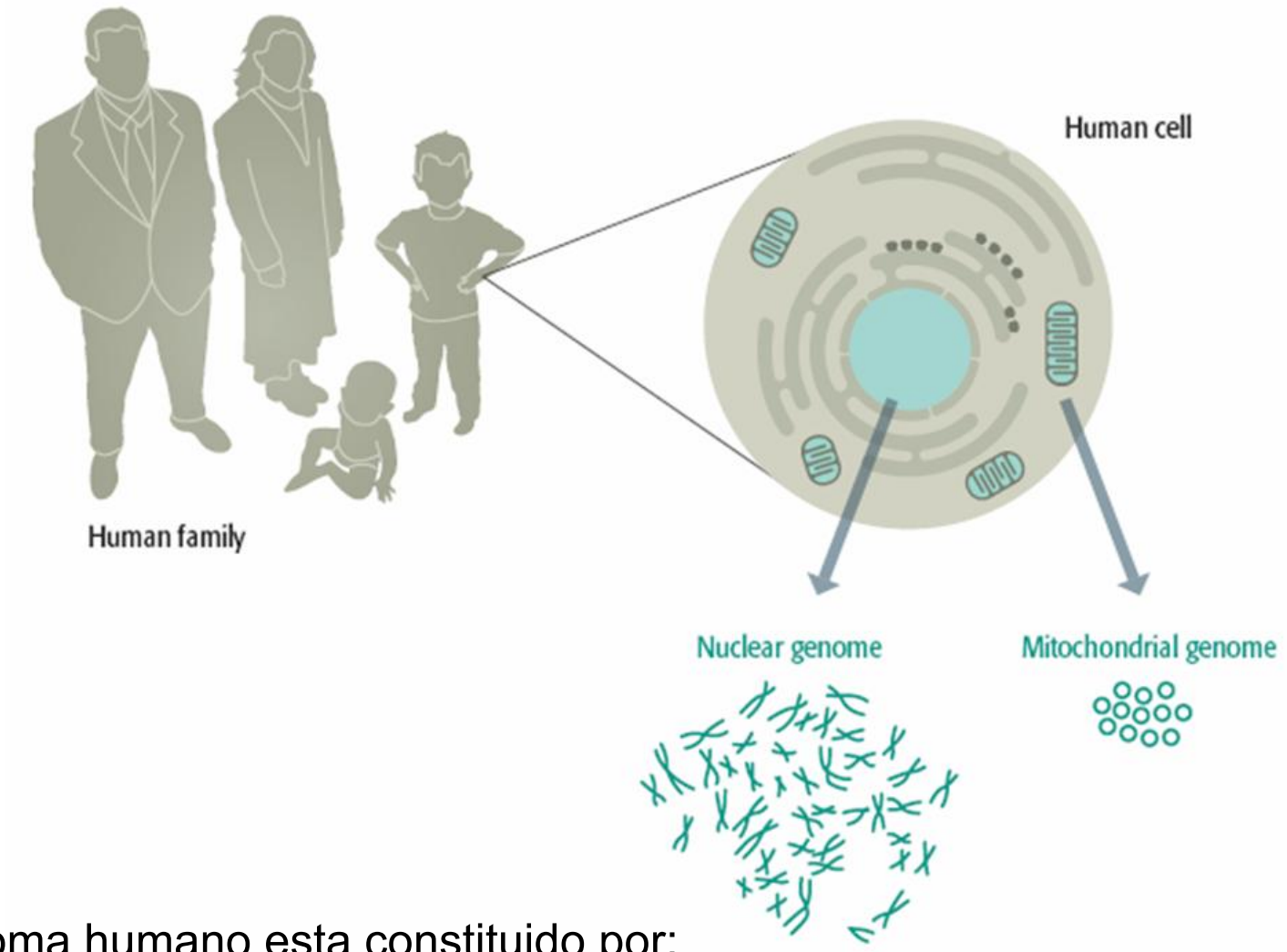
Polirribosomas. Varios ribosomas pueden traducir a la vez una molécula de mRNA





PABI: polyA binding protein I, proteina de union a poliA

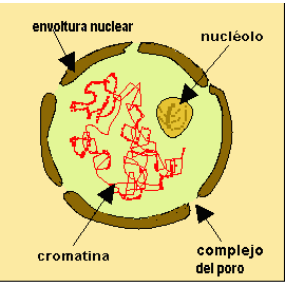
Genoma Humano



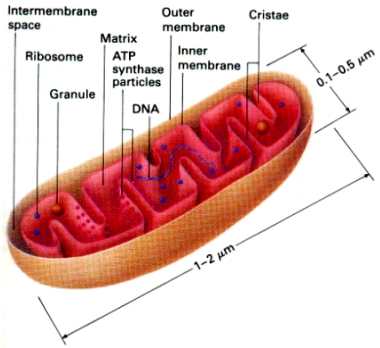
El genoma humano esta constituido por:

- genoma nuclear (23 pares de cromosomas; 3,1Gb o 3100Mb o 3.100.000.000bp)
- mitocondrial (circular; 16,6 kp).

Proporciones de los tipos de secuencias de DNA del genoma humano.



Corte de núcleo



GENOMA HUMANO

Genoma Nuclear
3000 Mb (25000 genes)

Genoma Mitocondrial
16,6 kb (37 genes)

30% **70%**

Genes y secuencias relacionadas

DNA extragénico

Dos genes de rRNAs

22 genes de tRNAs

13 proteínas

5% **95%**

DNA Codificante

DNA No Codificante

25% **75%**

Únicas o bajo número de copias

Moderada a altamente repetitiva

Seudogenes

Fragmentos de genes

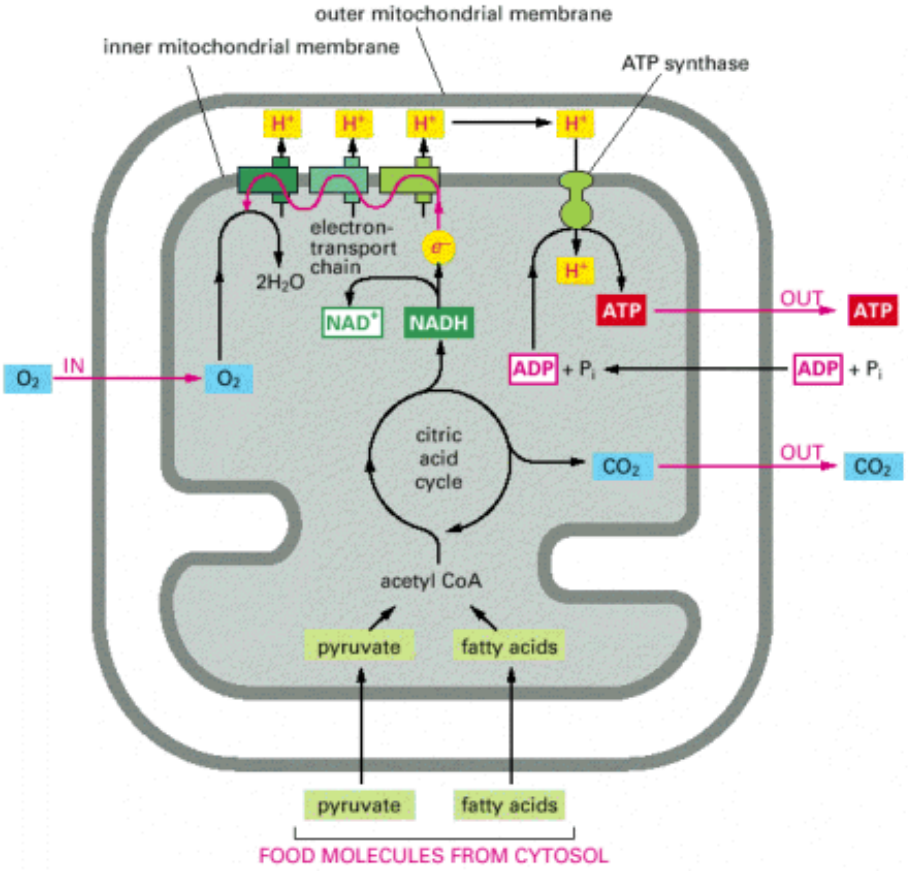
Intrones

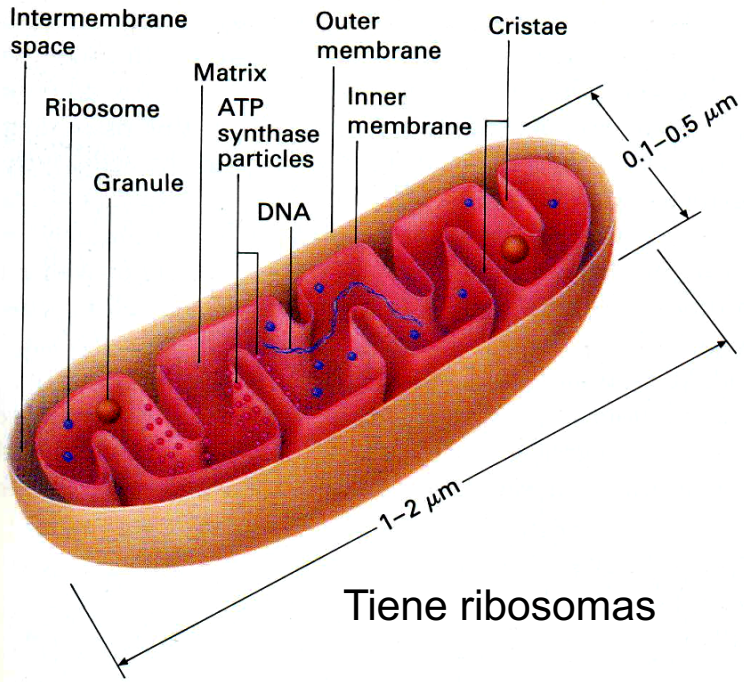
Repetidos agrupados o en tandem

Repetidos dispersos

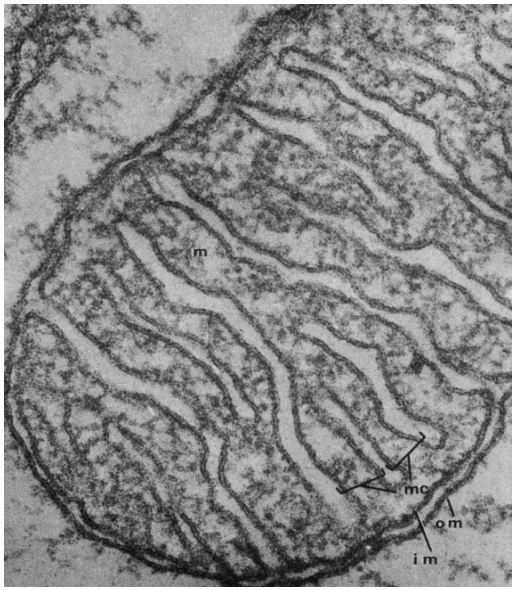
La Mitocondria

La energía se obtiene de la respiración celular





Tiene ribosomas



El tamaño del genoma mitocondrial y el número y tipo de genes presentes difiere en distintos organismos vivos

Feature	*Chlamydomonas reinhardtii*	*Homo sapiens*	*Saccharomyces cerevisiae*	*Arabidopsis thaliana*	*Reclinomonas americana*
Total number of genes	12	37	35	52	92
Types of genes					
Protein-coding genes	7	13	8	27	62
Respiratory complex	7	13	7	17	24
Ribosomal proteins	0	0	1	7	27
Transport proteins	0	0	0	3	6
RNA polymerase	0	0	0	0	4
Translation factor	0	0	0	0	1
Ribosomal RNA genes	2	2	2	3	3
Transfer RNA genes	3	22	24	22	26
Other RNA genes	0	0	1	0	1
Number of introns	1	0	8	23	1
Genome size (kb)	16	17	75	367	69

Based on Palmer (1997a).

Los genes mitocondriales no contienen intrones

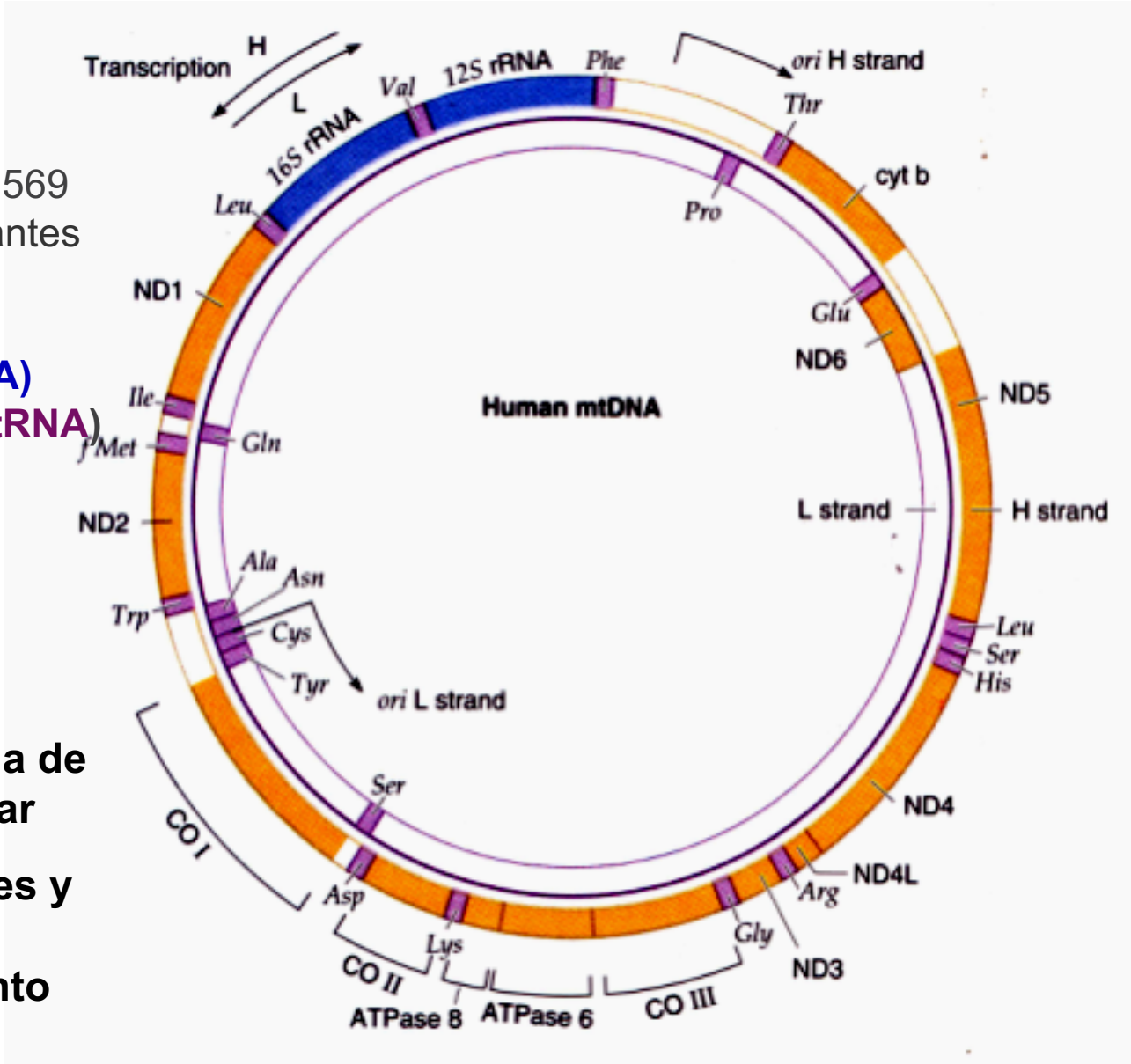
Organización del DNA mitocondrial humano

El genoma mitocondrial (16.569 bp) posee 37 genes codificantes para:

13 proteínas

2 **RNAs ribosomales (rRNA)**

22 **RNA de transferencia (tRNA)**



- **El mtDNA es una molécula de doble hebra de DNA circular**
- **El mtDNA no tiene intrones y los tRNAs sirven como elementos de procesamiento**

Cantidad relativa de DNA organelar en algunas células y tejidos

Table 14-3 Relative Amounts of Organelle DNA in Some Cells and Tissues

Organism	Tissue or Cell Type	DNA Molecules per Organelle	Organelles per Cell	Organelle DNA as Percent of Total Cellular DNA
Mitochondrial DNA				
Rat	liver	5–10	1000	1
Yeast*	vegetative	2–50	1–50	15
Frog	egg	5–10	10^7	99
Chloroplast DNA				
Chlamydomonas	vegetative	80	1	7
Maize	leaves	20–40	20–40	15

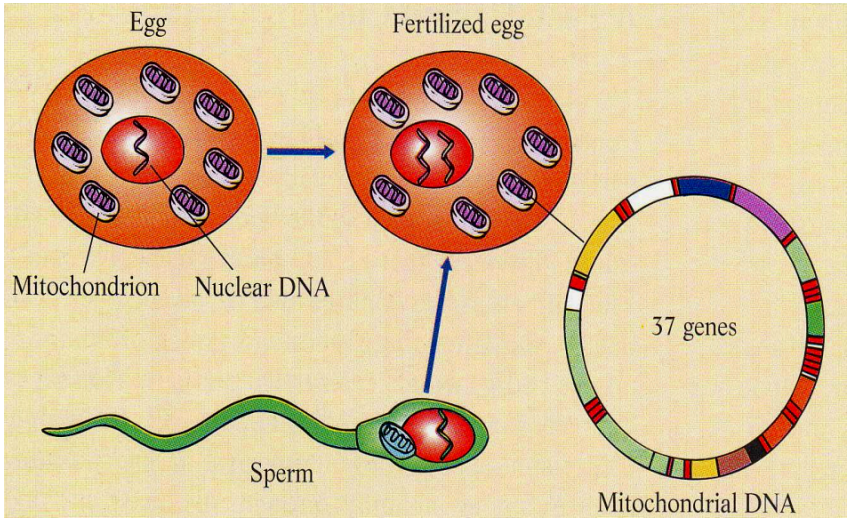
*The large variation in the number and size of mitochondria per cell in yeast is due to mitochondrial fusion and fragmentation.

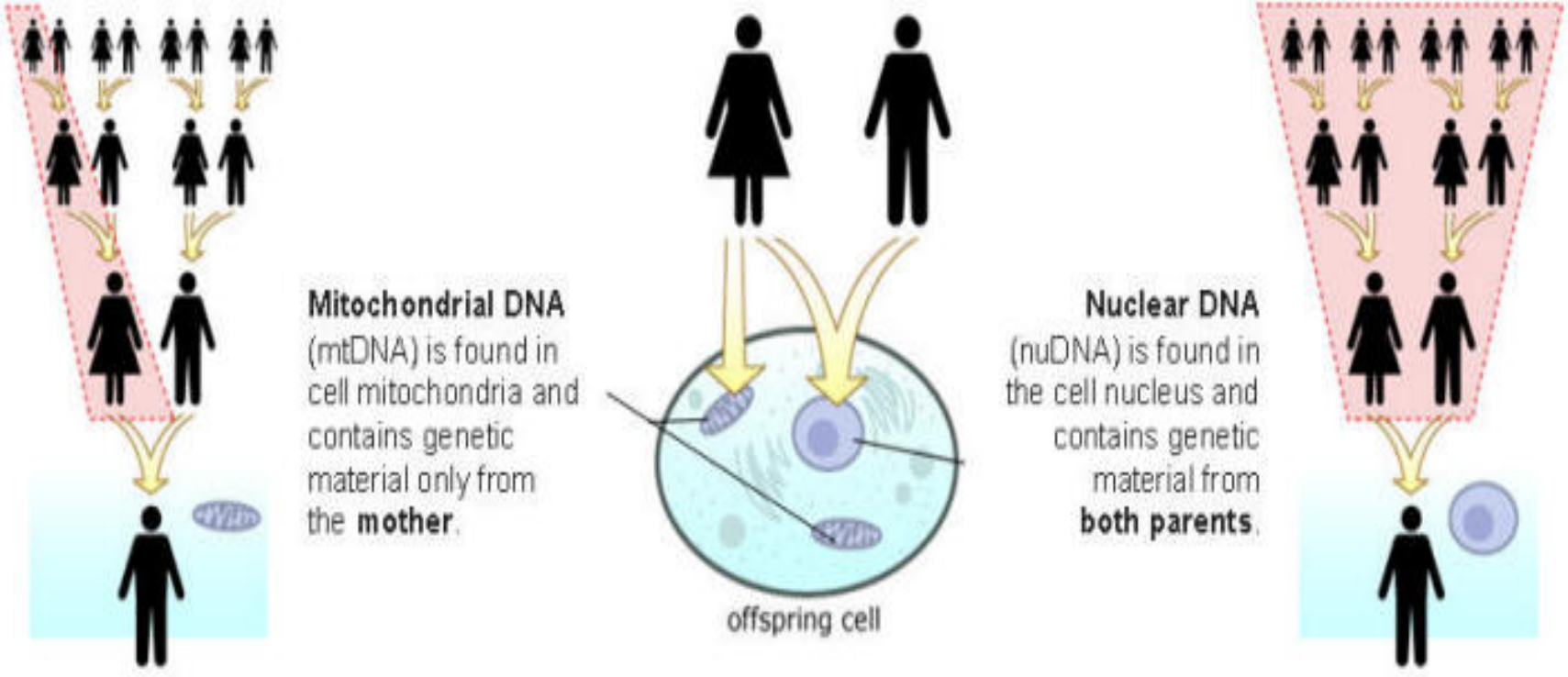
El DNA mitocondrial presenta diferencias con el Código Genético "Universal"

Codon	"Universal" Code	Mitochondrial Codes			
		Mammals	*Drosophila*	Yeasts	Plants
UGA	STOP	*Trp*	*Trp*	*Trp*	STOP
AUA	Ile	*Met*	*Met*	*Met*	Ile
CUA	Leu	Leu	Leu	*Thr*	Leu
AGA, AGG	Arg	*STOP*	*Ser*	Arg	Arg

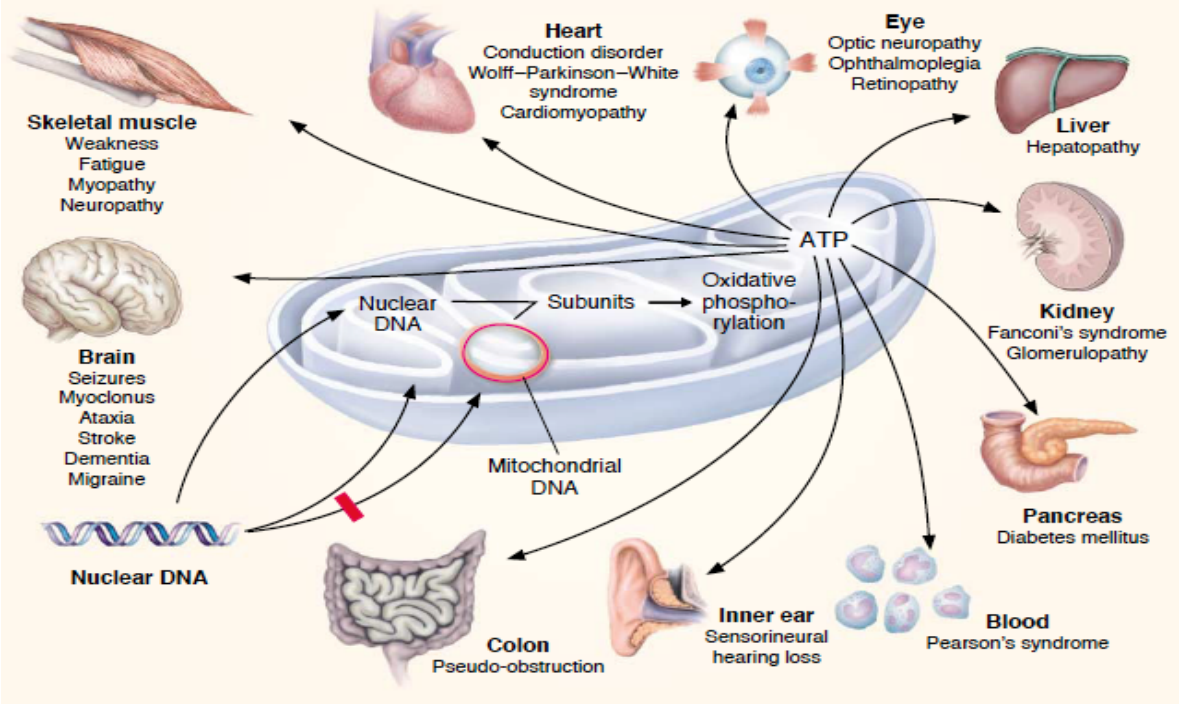
Herencia materna.

Las mitocondrias se heredan exclusivamente de la madre



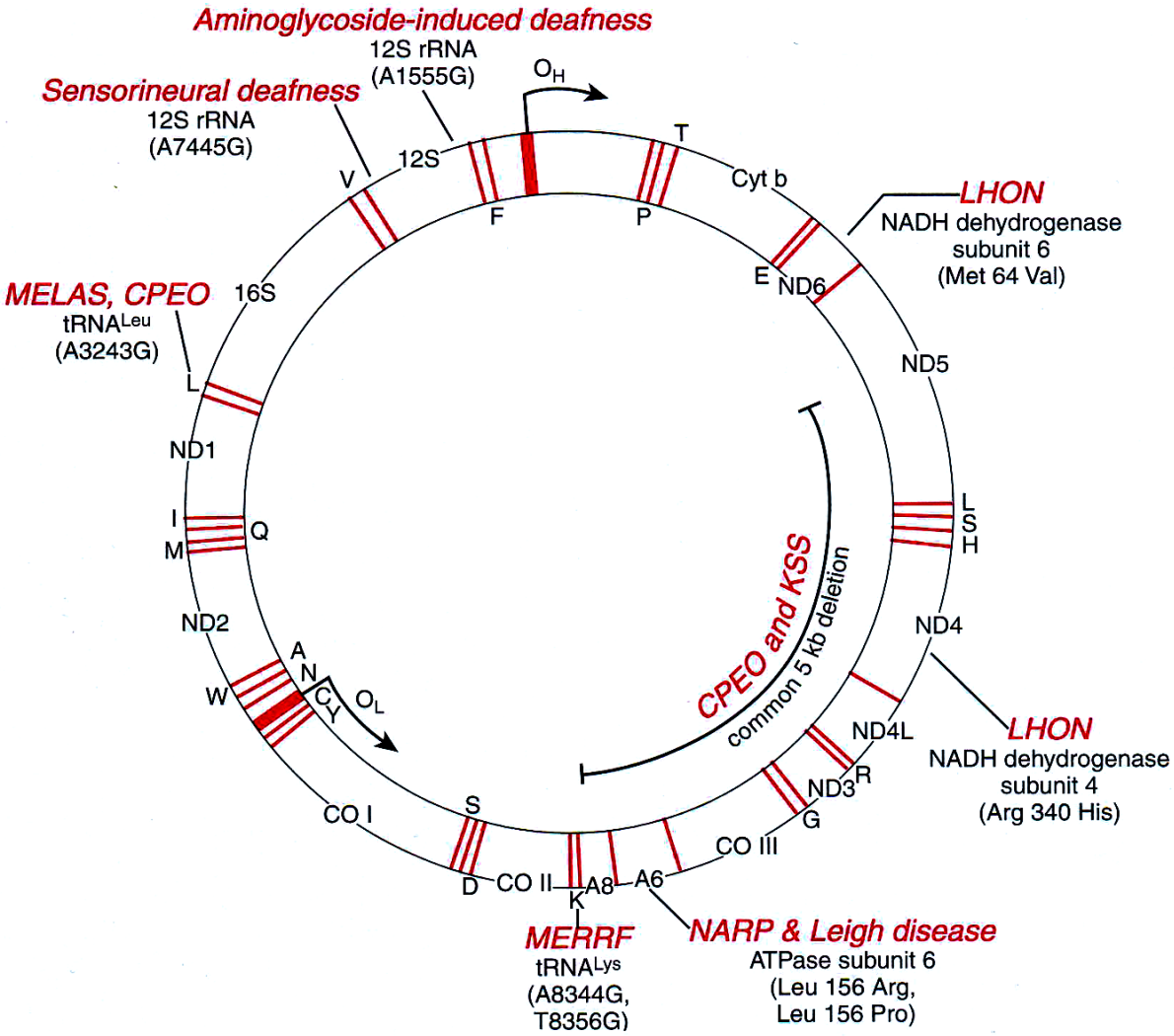


Existen mutaciones en mtDNA asociadas a enfermedades



- Hay más de 150 enfermedades mitocondriales
- Se han detectado ~275 genes que dan lugar a enfermedades mitocondriales
- La edad de inicio y el rango de la expresión fenotípica de estas enfermedades son diversos, con un inicio desde la vida neonatal hasta la séptima década de la vida.
- El rango de disfunción es heterogéneo, y abarca desde la participación de un solo órgano hasta multisistemico.
- Las mutaciones del mtDNA se encuentran en no más del 15% - 30% de los niños con trastornos mitocondriales

Algunas mutaciones puntuales en el mtDNA asociadas a enfermedades



Segregación replicativa de una mutación mitocondrial heteroplásmica

Poliplasmia: Entre 100 y 10.000 copias de mtDNA/célula dependiendo del tejido

- Homoplasmia
- Heteroplasmia

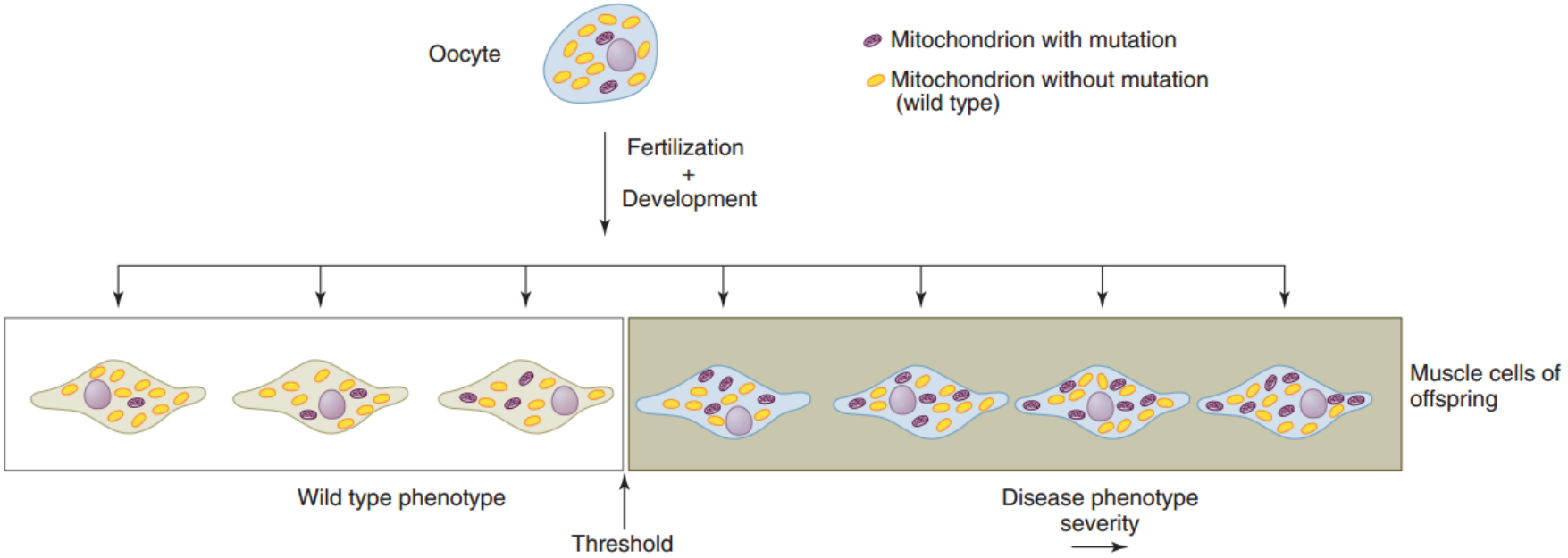
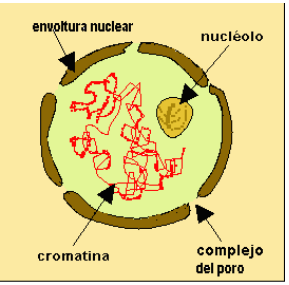
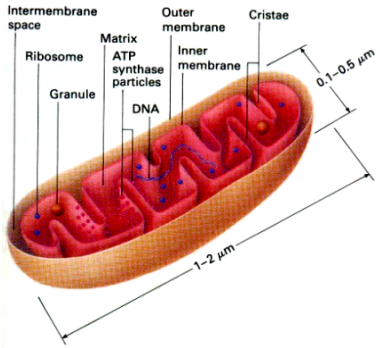


Figure 5.9 **Mitochondrial inheritance.** Mitochondria and their genes are only passed from the mother. Cells have many mitochondria. If an oocyte is heteroplasmic, differing numbers of copies of a mitochondrial mutation may be transmitted. The phenotype reflects the proportion of mitochondria bearing the mutation.

Proporciones de los tipos de secuencias de DNA del genoma humano.



Corte de núcleo



GENOMA HUMANO

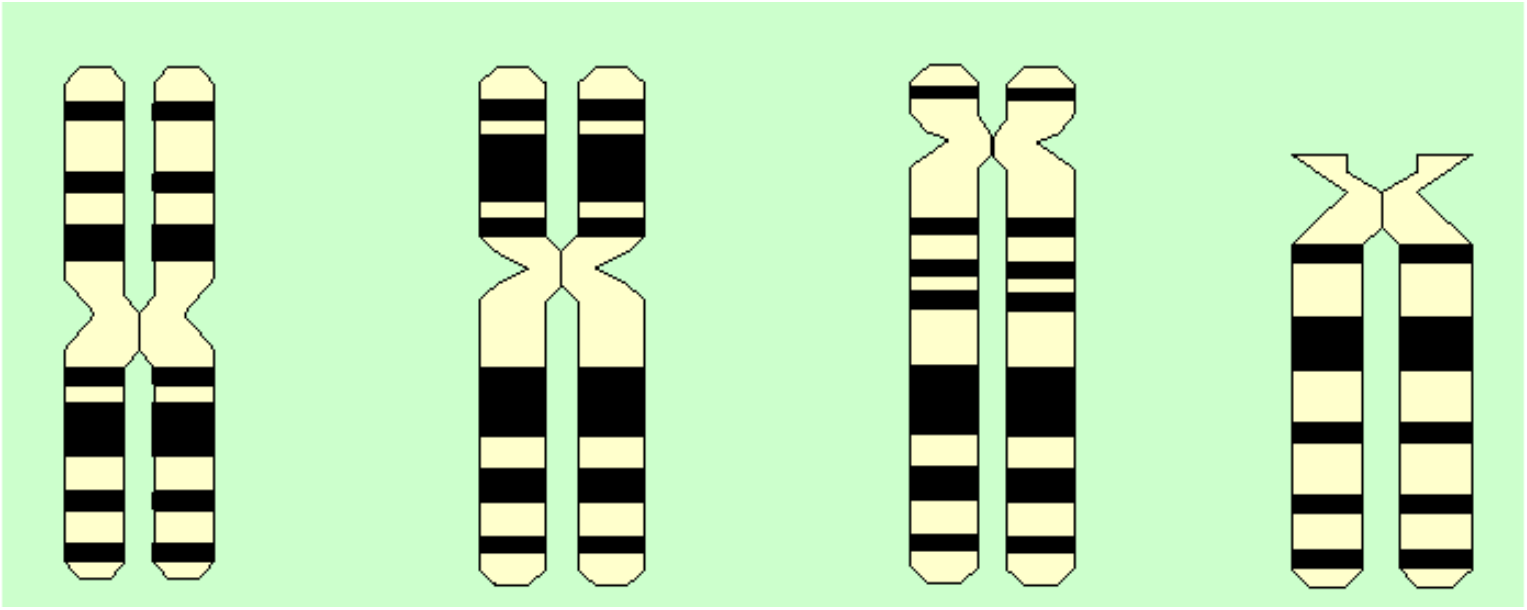
- **Genoma Nuclear** 3000 Mb (25000 genes)
  - 30%
    - **Genes y secuencias relacionadas**
      - 5% → **DNA Codificante**
      - 95% → **DNA No Codificante**
        - ***Seudogenes***
        - ***Fragmentos de genes***
        - ***Intrones***
  - 70% → **DNA extragénico**
    - 25% → **Únicas o bajo número de copias**
    - 75% → **Moderada a altamente repetitiva**
      - **Repetidos agrupados o en tandem**
      - **Repetidos dispersos**
- **Genoma Mitocondrial** 16,6 kb (37 genes)
  - **Dos genes de rRNAs**
  - **22 genes de tRNAs**
  - **13 proteínas**

Placa metafásica y cariotipo humano con bandeo G.

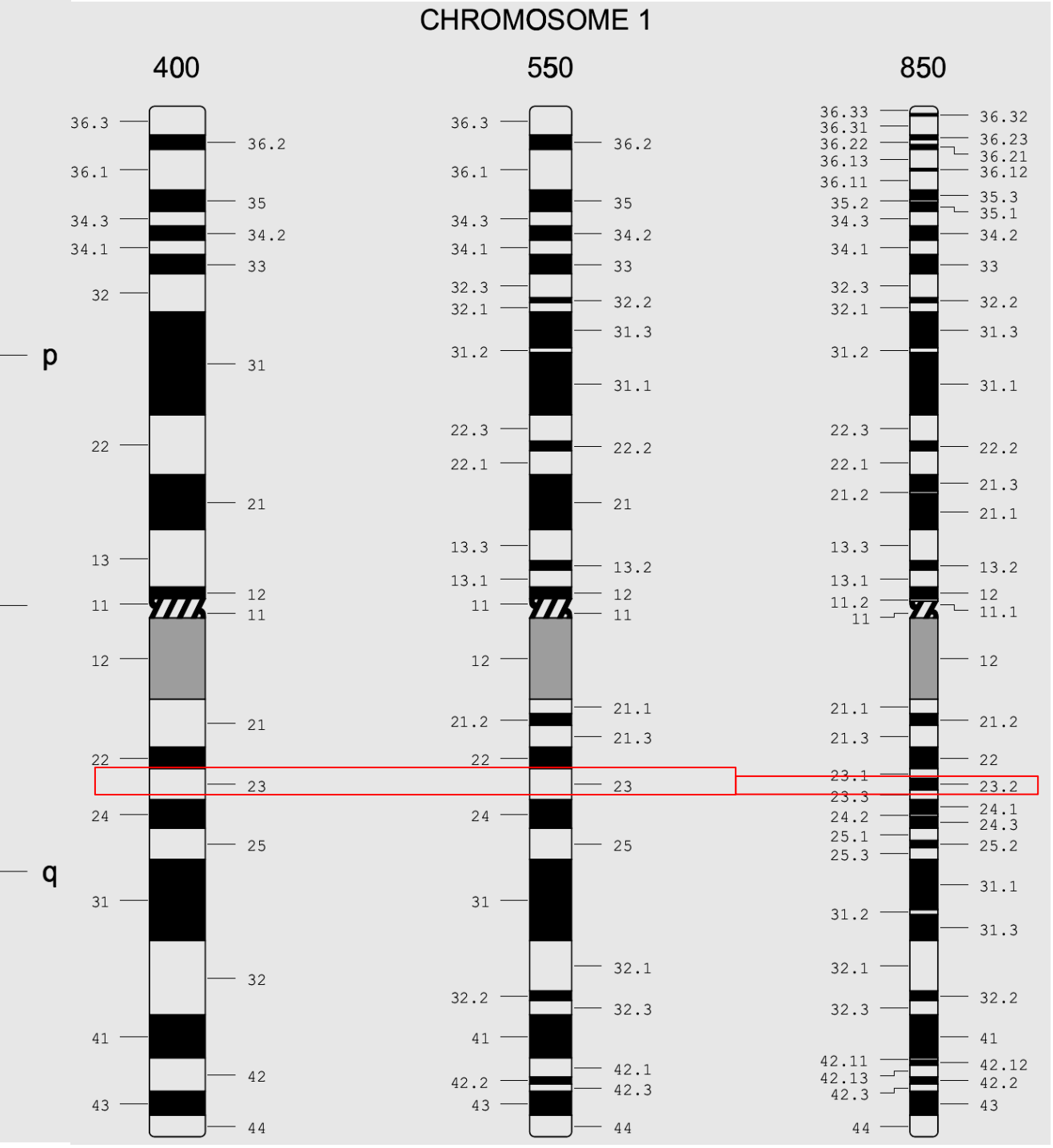
Tratamiento controlado con tripsina que degrada las proteínas y tinción con Giemsa produce bandas claras y oscuras. A las oscuras se las llama G+ y a las claras G-.

Los miembros de los pares homólogos de:
- autosomas presentan igual morfología, tamaño y patrón de bandas G.
- cromosomas sexuales (X e Y) difieren entre sí en todas estas características.

Clases de cromosomas según la posición del centrómero



Metacéntrico Submetacéntrico Subtelocéntrico Telocéntrico



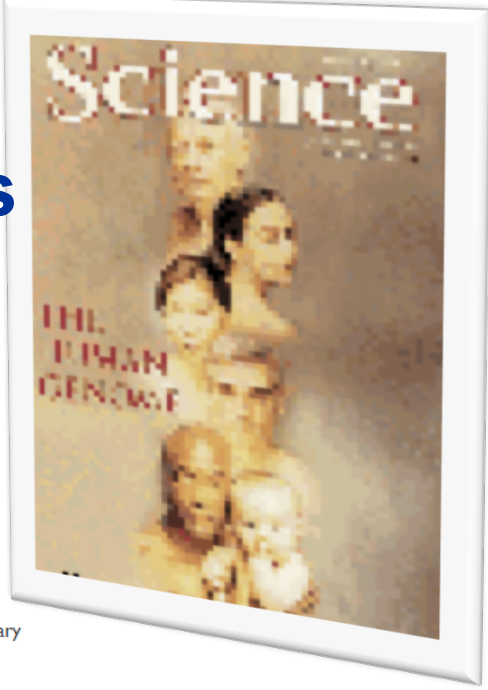
ISCN. International System for Human Cytogenetic Nomenclature

- Brazo corto (p)
- Brazo largo (q)
- Se asignó un número a cada región de cada cromosoma.

1q23.2

- Sub-banda
- Banda
- Región
- Brazo
- cromosoma

El Proyecto del Genoma Humano + otros Genomas



New Results

Deep Sequencing of 10,000 Human Genomes

Amalio Telenti, Levi T Pierce, William H Biggs, Julia di Iulio, Emily H.M. Wong, Martin M Fabani, Ewen F Kirkness, Ahmed Moustafa, Naisha Shah, Chao Xie, Suzanne C Brewerton, Nadeem Bulsara, Chad Garner, Gary Metzker, Efren Sandoval, Brad A Perkins, Franz J Och, Yaron Turpaz, J. Craig Venter

doi: http://dx.doi.org/10.1101/061663

This article is a preprint and has not been peer-reviewed [what does this mean?].

Abstract | Info/History | Metrics | Data Supplements | Preview PDF

Abstract

We report on the sequencing of 10,545 human genomes at 30-40x coverage with an emphasis on quality metrics and novel variant and sequence discovery. We find that 84% of an individual human genome can be sequenced confidently. This high confidence region includes 91.5% of exon sequence and 95.2% of known pathogenic variant positions. We present the distribution of over 150 million single nucleotide variants in the coding and non-coding genome. Each newly sequenced genome contributes an average of 8,579 novel variants. In addition, each genome carries in average 0.7 Mb of sequence that is not found in the main build of the hg38 reference genome. The density of this catalog of variation allowed us to construct high resolution profiles that define genomic sites that are highly intolerant of genetic variation. These results indicate that the data generated by deep genome sequencing is of the quality necessary for clinical use.

Copyright The copyright holder for this preprint is the author/funder. It is made available under a CC-BY-ND 4.0 International license.

Tamaño de los Genomas Procariontes

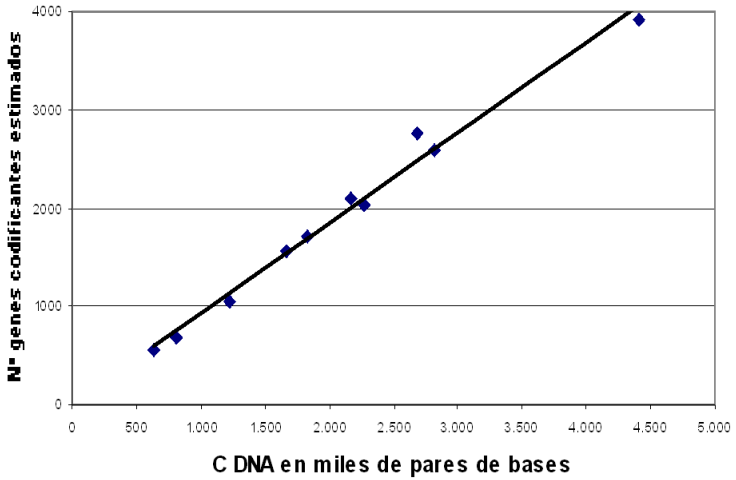
Su relación con el número de genes

Codificantes para proteínas

Procariontes	C DNA *	N° de genes (estimados)
Buchnera species	640.681	564
Mycoplasma pneumoniae	816.394	677
Chlamydia pneumoniae	1.230.230	1.052
Helicobacter pylori	1.667.867	1.566
Haemophilus influenzae	1.830.138	1.709
Streptococcus pneumoniae	2.160.837	2.094
Neisseria meningitidis	2.272.351	2.025
Xilella fastidiosa	2.679.306	2.766
Staphylococcus aureus	2.813.641	
Mycobacterium tuberculosis	4.411.529	

En procariontes existe una buena correlación e

- la “complejidad” morfológica
- el número de genes codificantes para prote

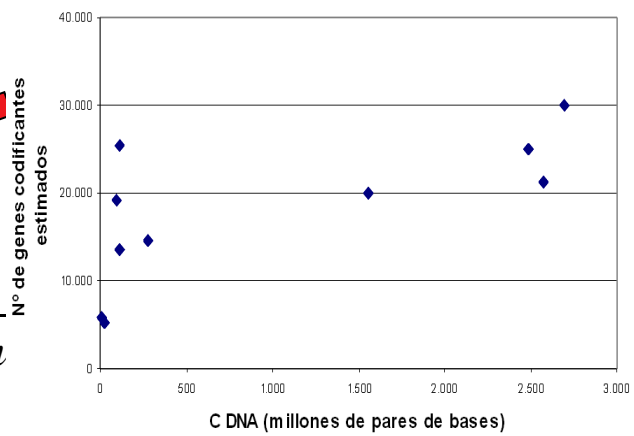


En eucariontes no existe correlación entre tamaño genómico y número de genes codificantes para proteínas

Codificantes para proteínas

Eucariontes	C DNA *	N° de genes (estimados)
Saccharomyces cerevisiae	12.000.000	5.800
Plasmodium falciparum	23.000.000	5.300
Caenorhabditis elegans	97.000.000	19.099
Arabidopsis thaliana	115.000.000	25.498
Drosophila melanogaster	116.000.000	13.601
Anopheles gambiae	278.000.000	14.653
Danio rerio (zebrafish)	1.560.000.000	20.000
Mus musculus	2.490.000.000	24.048
Rattus norvegicus	2.570.000.000	
Homo sapiens	3.000.000.000	
sapo	6.900.000.000	
Amoeba dubia	670.000.000.000	

Valor C - *cantidad de DNA presente en el genoma haploide de u*

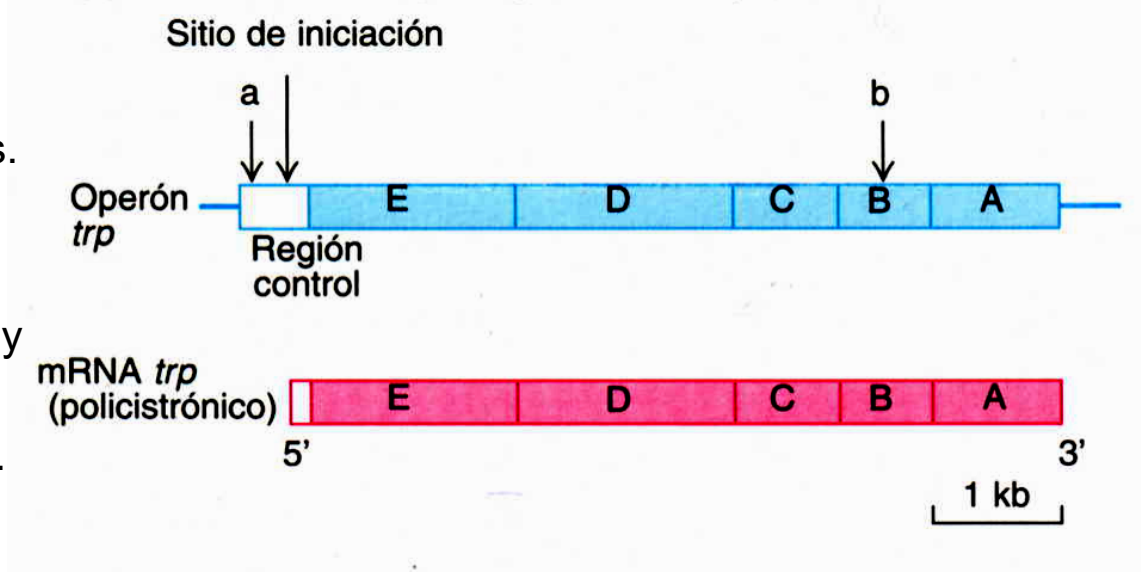


Organización de los Genomas

Factores que contribuyen a un mayor tamaño genómico

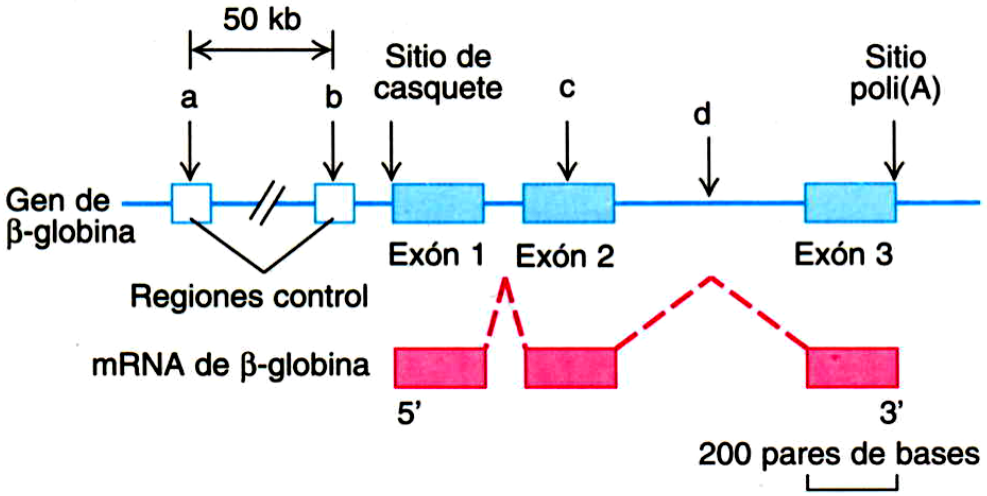
Procariontes:

- No presentan intrones, son continuos.
- Muchos genes están organizado en operones.
- Comparten las regiones reguladoras y promotoras.
- Las regiones intergénicas son cortas.

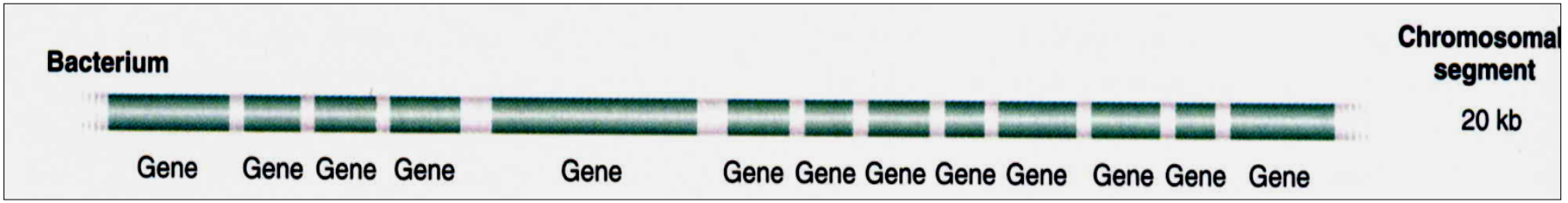


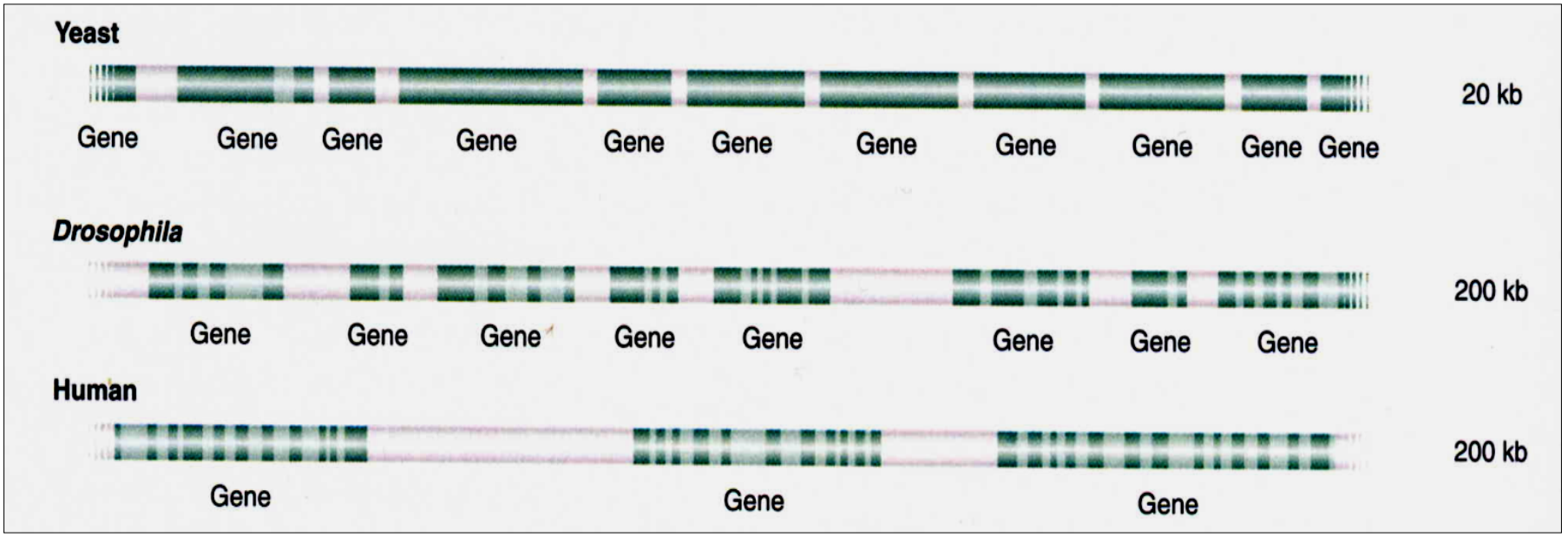
Eucariontes

- Muchos genes presentan **intrones**.
- Tienen extensas regiones reguladoras.
- Los mRNAs tienen extensas regiones no traducibles 5´ y 3´.



Los genomas eucariontes presentan mas DNA intergénico y sus genes tienen intrones. Estas condiciones aumentan al pasar de unicelulares a pluricelulares

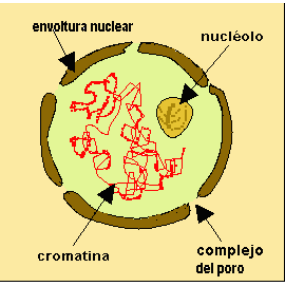




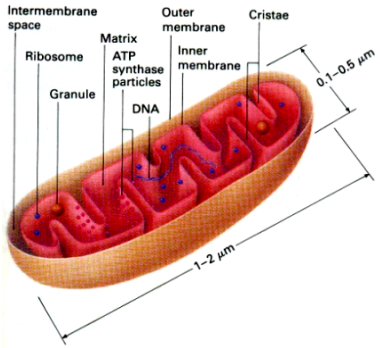
Principales hallazgos

- el genoma nuclear contiene 3.100 **millones de pares de bases** (Mb).
- el número total de genes **codificantes para proteínas** se estima en 25.000 aproximadamente
- el número total de pseudogenes se estima en 12.000 aprox.
- Aprox. 1.4% del genoma codifica para proteínas
- Nº de exones por gen: 1 - 363 (titina) (promedio 9,8)
- genes para RNA : MUCHOS *noticia en desarrollo*
- 50% del genoma son secuencias repetidas
- el 99,9% de la secuencia del DNA es exactamente la misma en todos los individuos

Proporciones de los tipos de secuencias de DNA del genoma humano.



Corte de núcleo



GENOMA HUMANO

Genoma Nuclear
3000 Mb (25000 genes)

Genoma Mitocondrial
16,6 kb (37 genes)

30% **70%**

Genes y secuencias relacionadas

DNA extragénico

Dos genes de rRNAs

22 genes de tRNAs

13 proteínas

5% **95%**

DNA Codificante

DNA No Codificante

25% **75%**

Únicas o bajo número de copias

Moderada a altamente repetitiva

Seudogenes

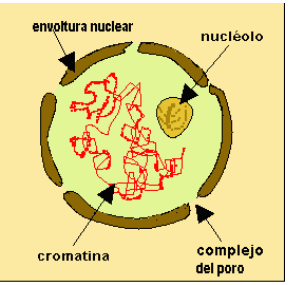
Fragmentos de genes

Intrones

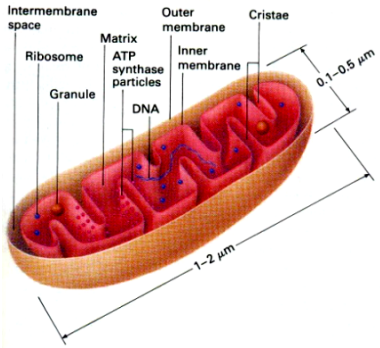
Repetidos agrupados o en tandem

Repetidos dispersos

Proporciones de los tipos de secuencias de DNA del genoma humano.



Corte de núcleo



GENOMA HUMANO

Genoma Nuclear
3000 Mb (25000 genes)

Genoma Mitocondrial
16,6 kb (37 genes)

30% | **70%**

Genes y secuencias relacionadas

DNA extragénico

Dos genes de rRNAs

22 genes de tRNAs

13 proteínas

5% | **95%**

DNA Codificante 48Mb

DNA No Codificante

25% | **75%**

Unicas o bajo número de copias

Moderada a altamente repetitiva

Seudogenes

Fragmentos de genes

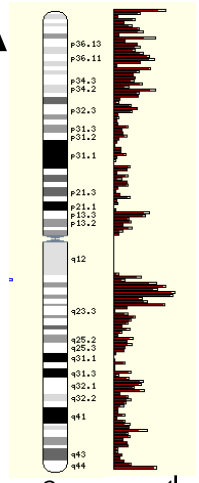
Intrones

Repetidos agrupados o en tandem

Repetidos dispersos

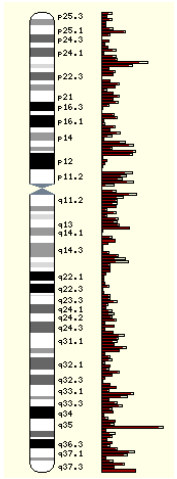
La densidad génica es diferente en distintos cromosomas y NO es homogénea a lo largo de los cromosomas

A



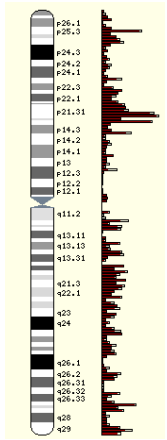
Cromosoma 1

Tamaño	245MB
genes conocidos	1.945
genes nuevos	316
SNPs	309,464



Cromosoma 2

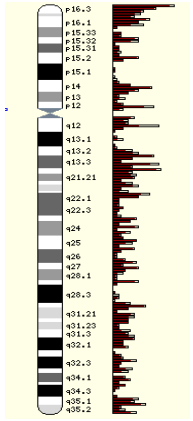
Tamaño	243MB
genes conocidos	1.289
genes nuevos	313
SNPs	259,706



Cromosoma 3

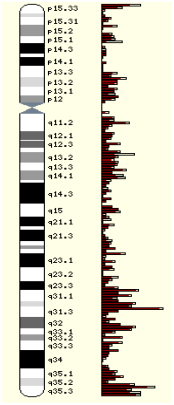
Tamaño	199MB
genes conocidos	1.049
genes nuevos	200
SNPs	199,756

B



Cromosoma 4

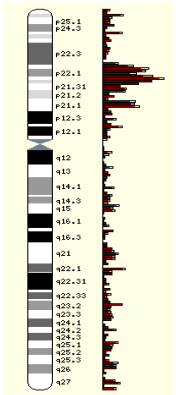
Tamaño	192MB
genes conocidos	765
genes nuevos	177
SNPs	198,494



Cromosoma 5

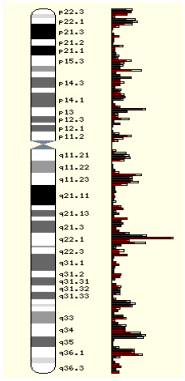
Tamaño	180MB
genes conocidos	879
genes nuevos	242
SNPs	223,799

C



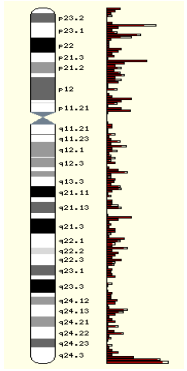
Cromosoma 6

Tamaño	170MB
genes conocidos	1.053
genes nuevos	214
SNPs	207,927



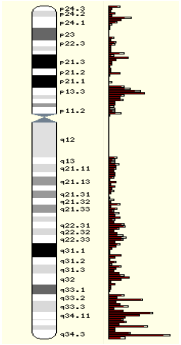
Cromosoma 7

Tamaño	158MB
genes conocidos	952
genes nuevos	251
SNPs	191,878



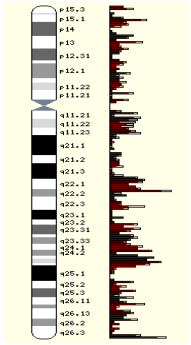
Cromosoma 8

Tamaño	145MB
genes conocidos	717
genes nuevos	134
SNPs	162.720



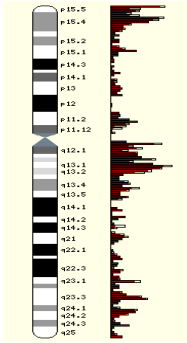
Cromosoma 9

Tamaño	134MB
genes conocidos	755
genes nuevos	157
SNPs	187.861



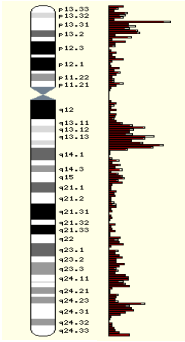
Cromosoma 10

Tamaño	135MB
genes conocidos	756
genes nuevos	164
SNPs	170.152



Cromosoma 11

Tamaño	135MB
genes conocidos	1.294
genes nuevos	212
SNPs	172.628



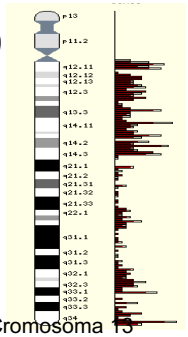
Cromosoma 12

Tamaño	133MB
genes conocidos	1.006
genes nuevos	147
SNPs	137.582

http://www.ensembl.org/

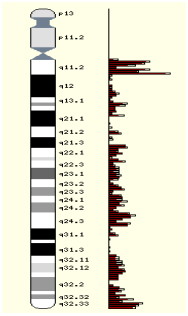
La densidad génica es diferente en distintos cromosomas y NO es homogénea a lo largo de los cromosomas

D



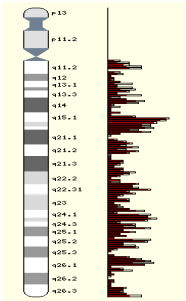
Cromosoma 13

Tamaño	114MB
genes conocidos	341
genes nuevos	94
SNPs	110.733



Cromosoma 14

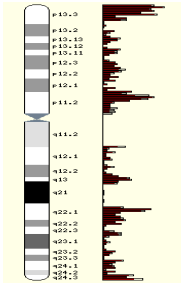
Tamaño	105MB
genes conocidos	647
genes nuevos	139
SNPs	96.837



Cromosoma 15

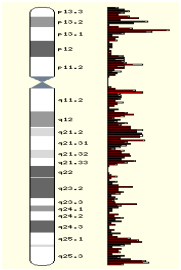
Tamaño	100MB
genes conocidos	592
genes nuevos	167
SNPs	112.606

E



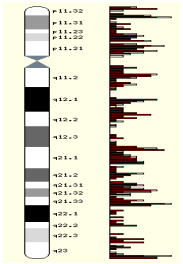
Cromosoma 16

Tamaño	90MB
genes conocidos	900
genes nuevos	136
SNPs	127.253



Cromosoma 17

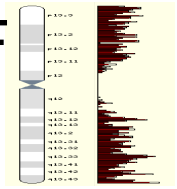
Tamaño	82MB
genes conocidos	1.121
genes nuevos	153
SNPs	96.210



Cromosoma 18

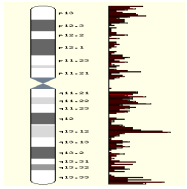
Tamaño	78MB
genes conocidos	267
genes nuevos	65
SNPs	92.281

F



Cromosoma 19

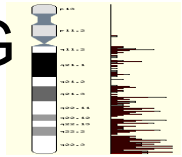
Tamaño	64MB
genes conocidos	1.303
genes nuevos	130
SNPs	73.282



Cromosoma 20

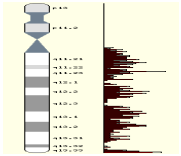
Tamaño	64MB
genes conocidos	631
genes nuevos	54
SNPs	185.391

G



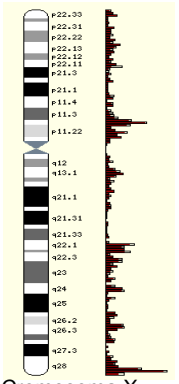
Cromosoma 21

Tamaño	47MB
genes conocidos	231
genes nuevos	42
SNPs	66.678



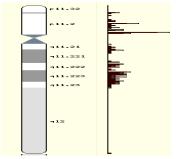
Cromosoma 22

Tamaño	49MB
genes conocidos	485
genes nuevos	91
SNPs	99.388



Cromosoma X

Tamaño	153MB
genes conocidos	750
genes nuevos	188
SNPs	145.535



Cromosoma Y

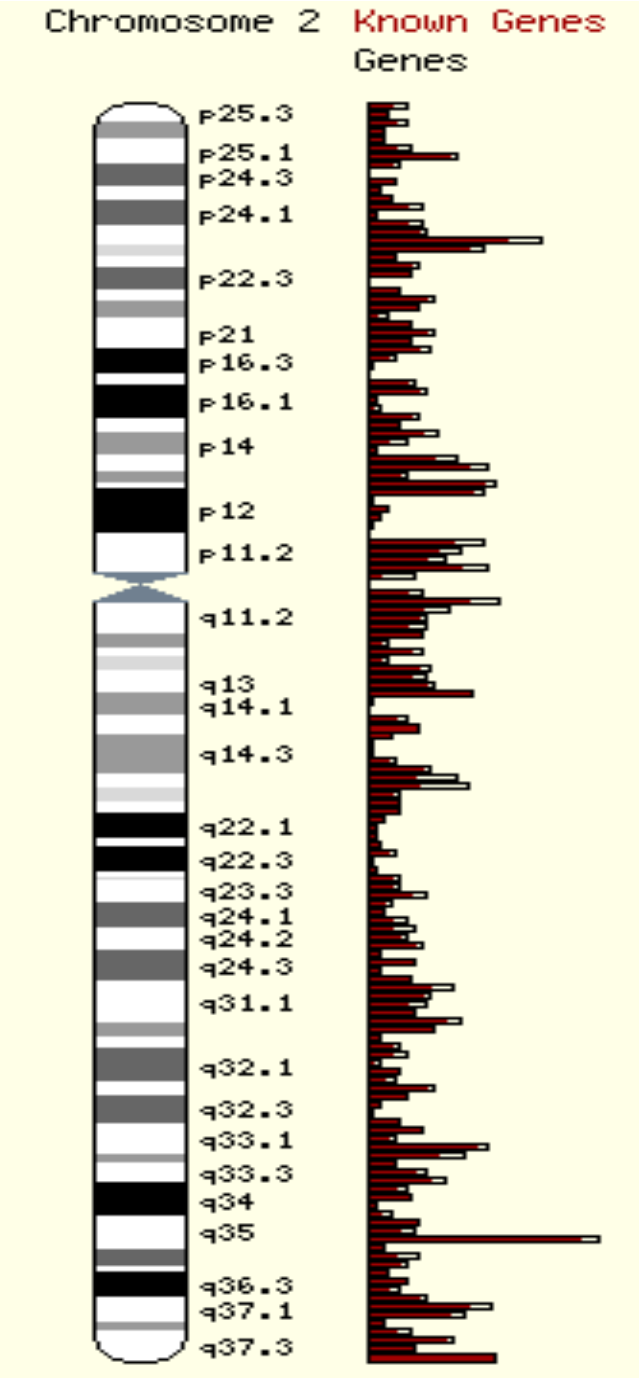
Tamaño	50MB
genes conocidos	94
genes nuevos	35
SNPs	28.864

http://www.ensembl.org/

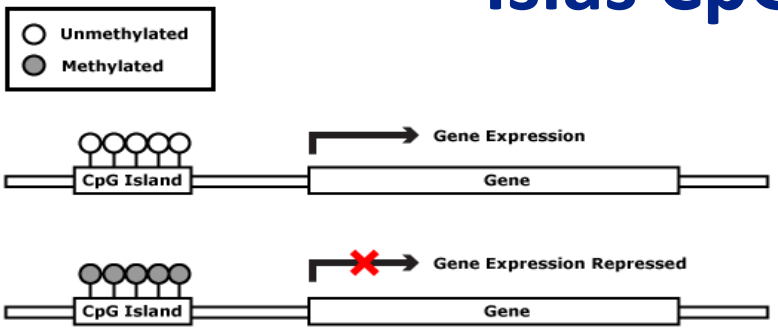
La densidad génica correlaciona con el bandeo cromosómico.

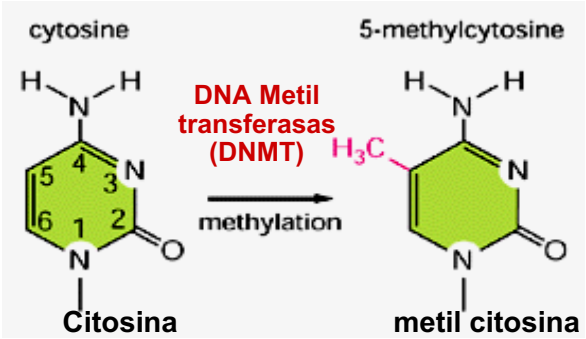
B*andas G*

Banda **G+**	Pobres en genes y ricas en secuencias Line
Banda **G-**	Ricas en genes y ricas en secuencias Alu



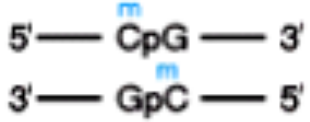
Islas CpG

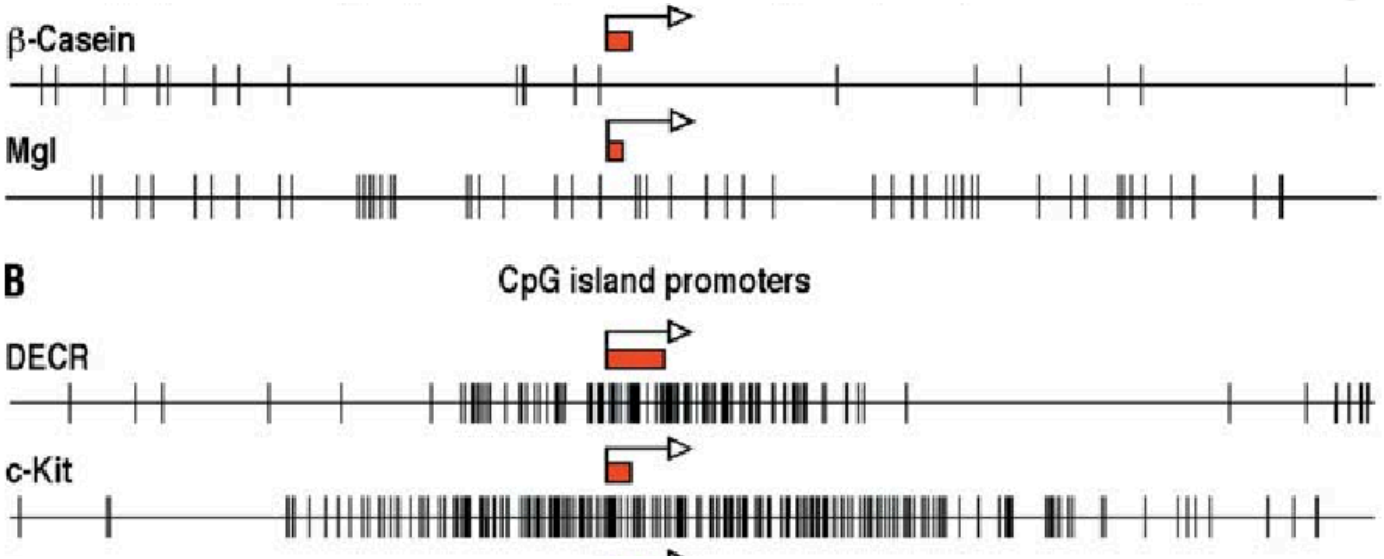




SAM: S-Adenosilmetionina (Donador del grupo metilo). SAH: S-AdenosilHomocisteína.

- Islas CpG: segmentos de DNA 500-1000bp con alta concentración de dinucleótidos CpG
- Aprox el 60% de los promotores contienen islas CpG





Antequera et al, Cel Mol Life Sci, 2003, 60: 1647-58.

Densidad génica por cromosoma

Cromosoma	Nº de Genes Conocidos	Tamaño (MB)	Densidad génica (genes/MB)
1	1945	245	**7,94**
2	1289	243	**5,30**
3	1049	199	**5,27**
4	765	192	**3,98**
5	879	180	**4,88**
6	1053	170	**6,19**
7	952	158	**6,03**
8	717	145	**4,94**
9	755	134	**5,63**
10	756	135	**5,60**
11	1294	135	**9,59**
12	1006	133	**7,56**
13	341	114	**2,99**
14	647	105	**6,16**
15	592	100	**5,92**
16	900	90	**10,00**
17	1121	82	**13,67**
18	267	78	**3,42**
19	1303	64	**20,36**
20	631	64	**9,86**
21	231	47	**4,91**
22	485	49	**9,90**
X	750	153	**4,90**
Y	94	50	**1,88**

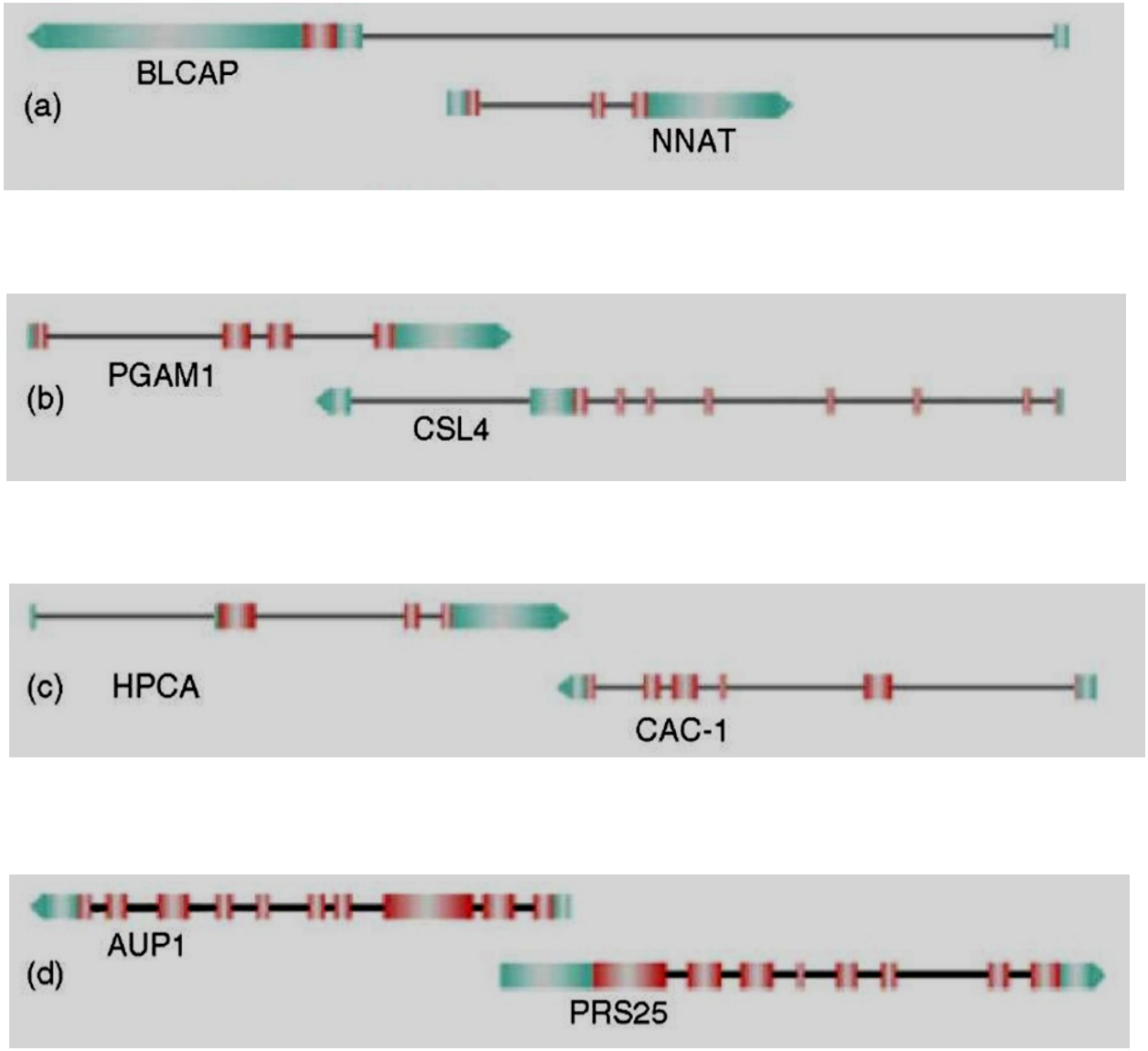
Diversidad en tamaños y estructura de genes codificantes para proteínas

TABLE 9.4 STRUCTURAL VARIATION IN SIZE AND ORGANIZATION OF HUMAN PROTEIN-CODING GENES

Human protein	Size of protein (no. of amino acids)	Size of gene (kb)	No. of exons	Coding DNA (%)	Average size of exon (bp)	Average size of intron (bp)
SRY	204	0.9	1	94	850	–
β-Globin	146	1.6	3	38	150	490
p16	156	7.4	3	17	406	3064
Serum albumin	609	18	14	12	137	1100
Type VII collagen	2928	31	118	29	77	190
p53	393	39	10	6.0	236	3076
Complement C3	1641	41	29	8.6	122	900
Apolipoprotein B	4563	45	29	31	487	1103
Phenylalanine hydroxylase	452	90	26	3	96	3500
Factor VIII	2351	186	26	3	375	7100
Huntingtin	3144	189	67	8.0	201	2361
RB1 retinoblastoma protein	928	198	27	2.4	179	6668
CFTR (cystic fibrosis transmembrane receptor)	1480	250	27	2.4	227	9100
Titin	34,350	283	363	40	315	466
Utrophin	3433	567	74	2.2	168	7464
Dystrophin	3685	2400	79	0.6	180	30,770

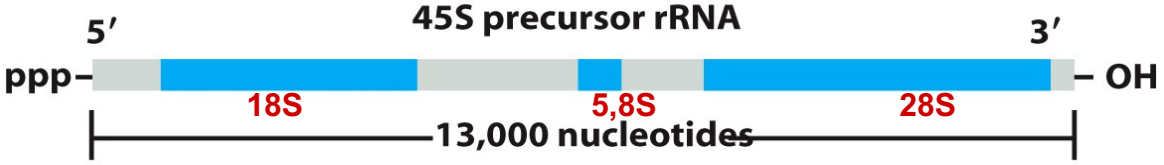
- No hay relación entre el tamaño de los genes y el tamaño de las proteínas
- Los intrones son mucho mas grandes que los exones.

Los genes humanos con frecuencia se encuentran superpuestos

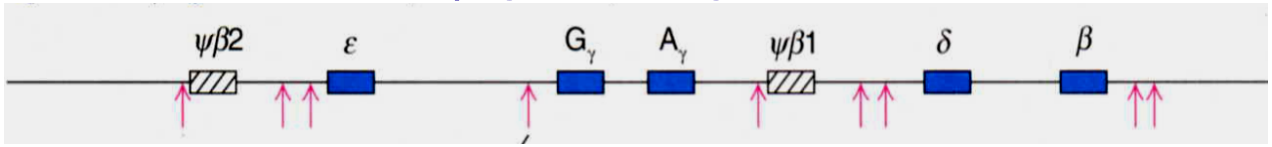


Familias multigénicas. Grupo de genes de un genoma que derivan de una secuencia ancestral común.

1. Genes agrupados (en tandem): Ej.
   - rRNA 5 S con 2000 genes agrupados en el cromosoma 1.
   - rRNA con una única unidad transcripcional repetida unas 280 veces agrupados en 5 clusters ubicados en los cromosomas 13. 14, 15, 21 y 22;

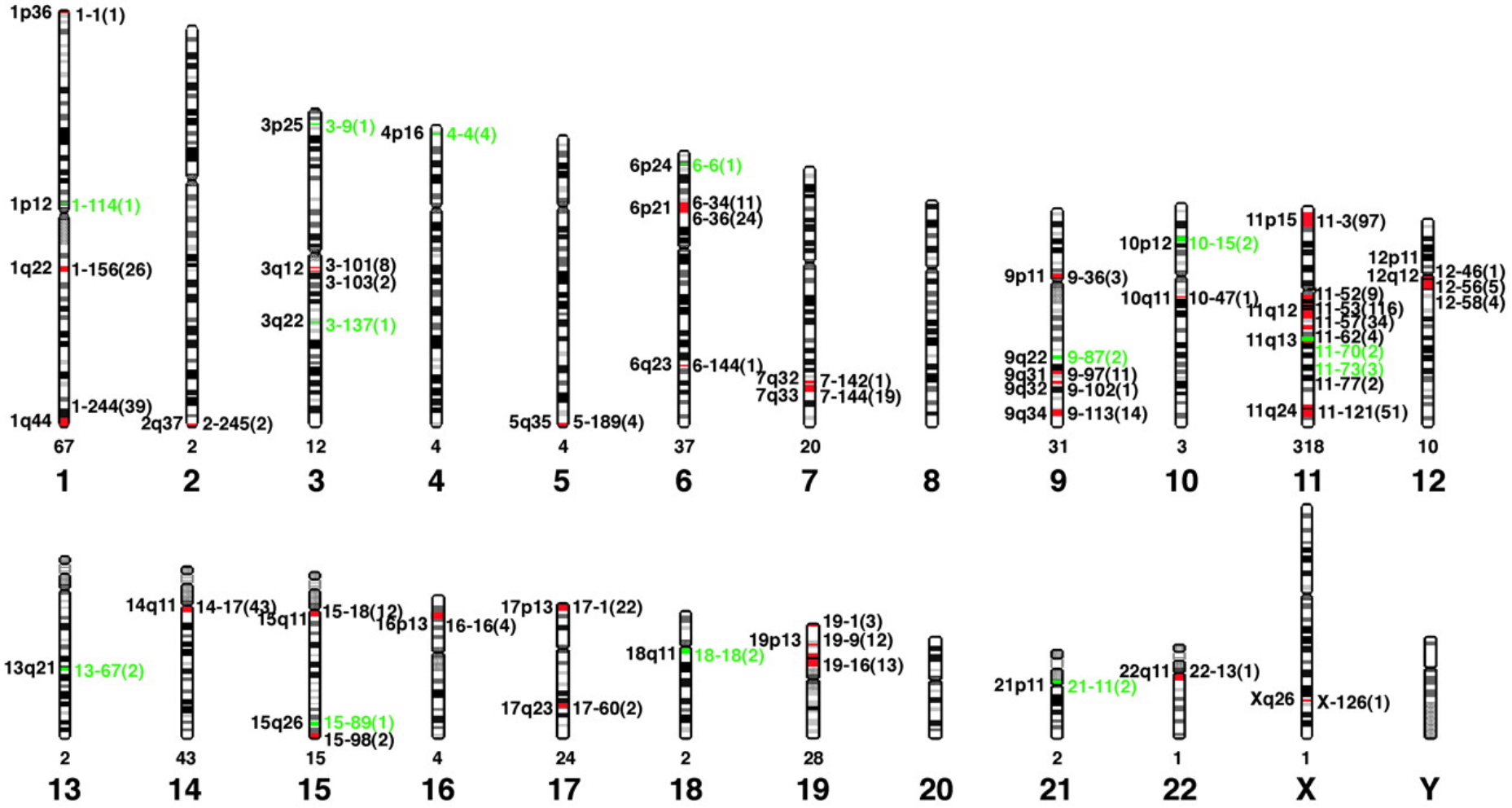


   - tRNAs 50 sitios cromosómicos con 10 a 100 copias por sitio.
   - Genes codificantes para β-globina agrupados en el cromosoma 11.



   - clusters de genes homeóticos agrupados en 4 cromosomas diferentes.

2. Genes dispersos: Ej. Aldolasa en cromosomas 3,9,10,16 y 17; Actina (5 a 30 genes), tubulina (3 a 15 genes), etc



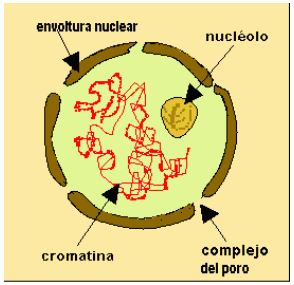
Malnic B et al. PNAS 2004;101:2584-2589

Los 630 genes codificantes para receptores olfatorios (OR) se localizan en 51 loci diferentes distribuidos en 21 cromosomas. Rojo Loci de OR con uno o mas genes intactos, Verde loci que contienen solo pseudogenes. El número de genes de OR y el número por cromosoma se muestra abajo.

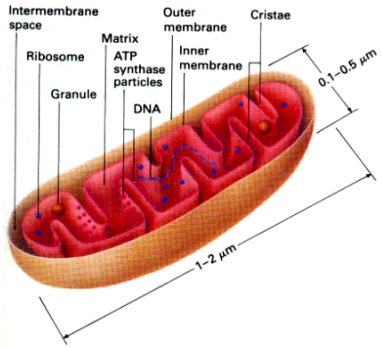
Genes de RNA no codificantes son tremendamente diversos.

Clase de RNA	Familia o subclase	N° de diferentes tipos	Función	Tamaño (nt)
RNA ribosomal (rRNA)	12S rRNA, 16S rRNA, 5S rRNA, 5,8S rRNA, 18S rRNA, 28S rRNA	1 de cada uno, miles de copias	Componentes de ribosomas mitocondriales y citoplasmáticos, Traducción	120-5000
RNA de transferencia (tRNA)	Mitocondriales y citoplasmáticos	22 y 49	Traducción	70-100
RNA nuclear pequeño (snRNA)	Familia spliceosomal con subclase Sm y Lsm (U1, U2, U4, U5 Y U6 snRNAs)	9	Splicing	60-360
RNA nucleolar pequeño (snoRNA)	Clase C/D box	246	Modificación de RNAs y procesamiento de rRNA.	70-300
Micro RNA (miRNA)	>70 familias de miRNAs	Mas de 1000	Regulación de expresión génica (traducción)	22
RNA de unión a piwi (piRNA)	89 clusters individuales	>15.000	Generalmente en secuencias repetidas, expresados en células germinales donde limitan actividad de transposones	24-31
Short interfering RNAs (siRNA)	muchos	Mas de 10.000	Regula expresión génica, degradación de RNAs	21-22
Long non coding RNA (lncRNA)	muchos	> 30.000	Regula expresión, estructura de cromatina	Generalmente > 1kb

Proporciones de los tipos de secuencias de DNA del genoma humano.



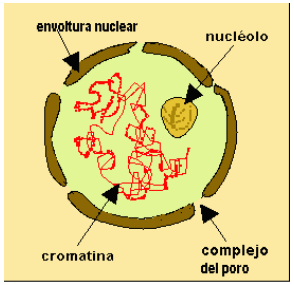
Corte de núcleo



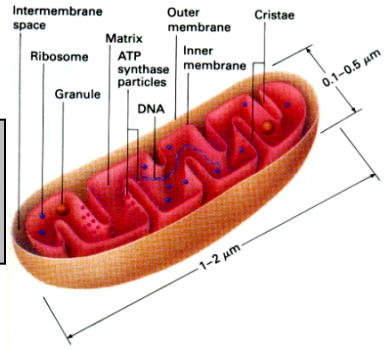
GENOMA HUMANO

- **Genoma Nuclear** 3300 Mb (25000 genes)
  - 30%
    - **Genes y secuencias relacionadas**
      - 5%
        - **DNA Codificante**
      - 95%
        - **DNA No Codificante**
          - **Pseudogenes**
          - **Fragmentos de genes**
          - **Intrones**
  - 70%
    - **DNA extragénico**
      - 25%
        - **Unicas o bajo número de copias**
      - 75%
        - **Moderada a altamente repetitiva**
          - **Repetidos agrupados o en tandem**
          - **Repetidos dispersos**
- **Genoma Mitocondrial** 16,6 kb (37 genes)
  - **Dos genes de rRNAs**
  - **22 genes de tRNAs**
  - **13 proteínas**

Proporciones de los tipos de secuencias de DNA del genoma humano.



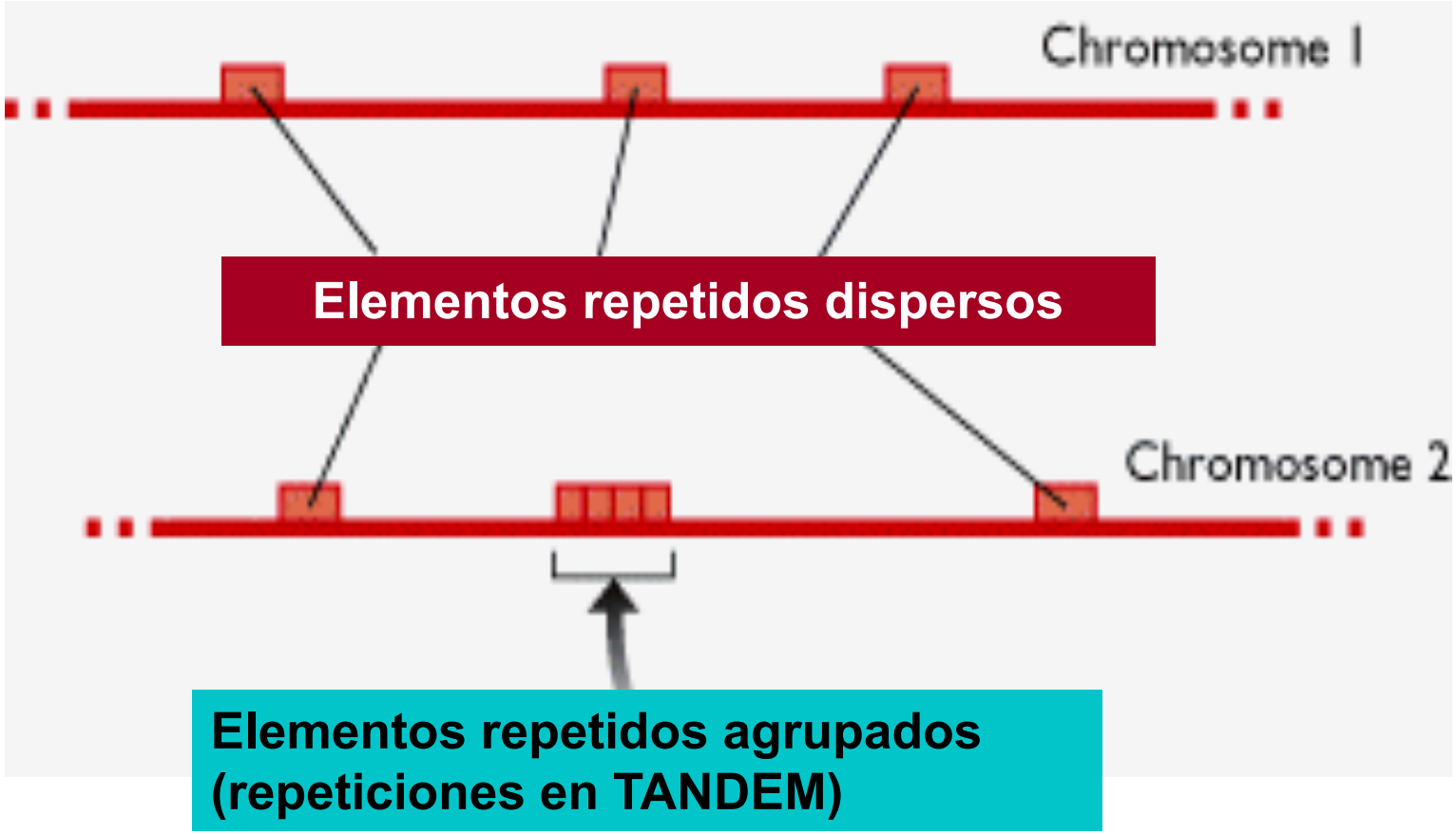
Corte de núcleo



GENOMA HUMANO

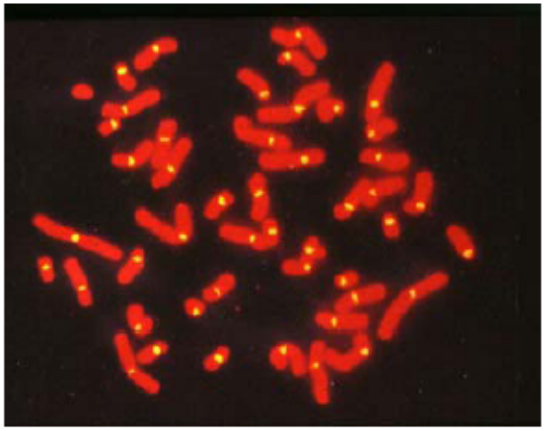
- **Genoma Nuclear** 3300 Mb (30000 genes)
  - 30%
    - **Genes y secuencias relacionadas**
      - 5%
        - **DNA Codificante**
      - 95%
        - **DNA No Codificante**
          - **Seudo genes**
          - **Fragmentos de genes**
          - **Intrones**
  - 70%
    - **DNA extragénico**
      - 25%
        - **Únicas o bajo número de copias**
      - 75%
        - **Moderada a altamente repetitiva**
          - **Repetidos agrupados o en tandem**
          - **Repetidos dispersos**
- **Genoma Mitocondrial** 16,6 kb (37 genes)
  - **Dos genes de rRNAs**
  - **22 genes de tRNAs**
  - **13 proteínas**

Secuencias repetidas: Agrupadas y dispersas



DNA repetido agrupado

DNA repetitivo centromérico (Satélite)
Tamaño de la repetición: 171 pb.
Tamaño del conglomerado: 100 a 5.000 Kb.
Localización: Centrómeros



Detección del DNA satélite clásico humano, mediante FISH con una sonda para α DNA centromérico.

•**Minisatélites,VNTRs** (variable number of tandem repeats, subtelomericos): unidad de repetición es de hasta 7-64 bp
•**Microsatélites, STRs** (short tandem repeats): unidad de repetición es de 1-6 bp. Un STR típico consiste de 10–30 copias .

DNA repetitivo telomérico (Minisatélites)
Tamaño de la repetición: 6 pb (TTAGGG).
Tamaño del conglomerado: 100pb a 20 Kb.
Localización: Telómeros.

Detección de DNA minisatélite telomérico humano, mediante FISH con una sonda marcada con un fluorocromo.

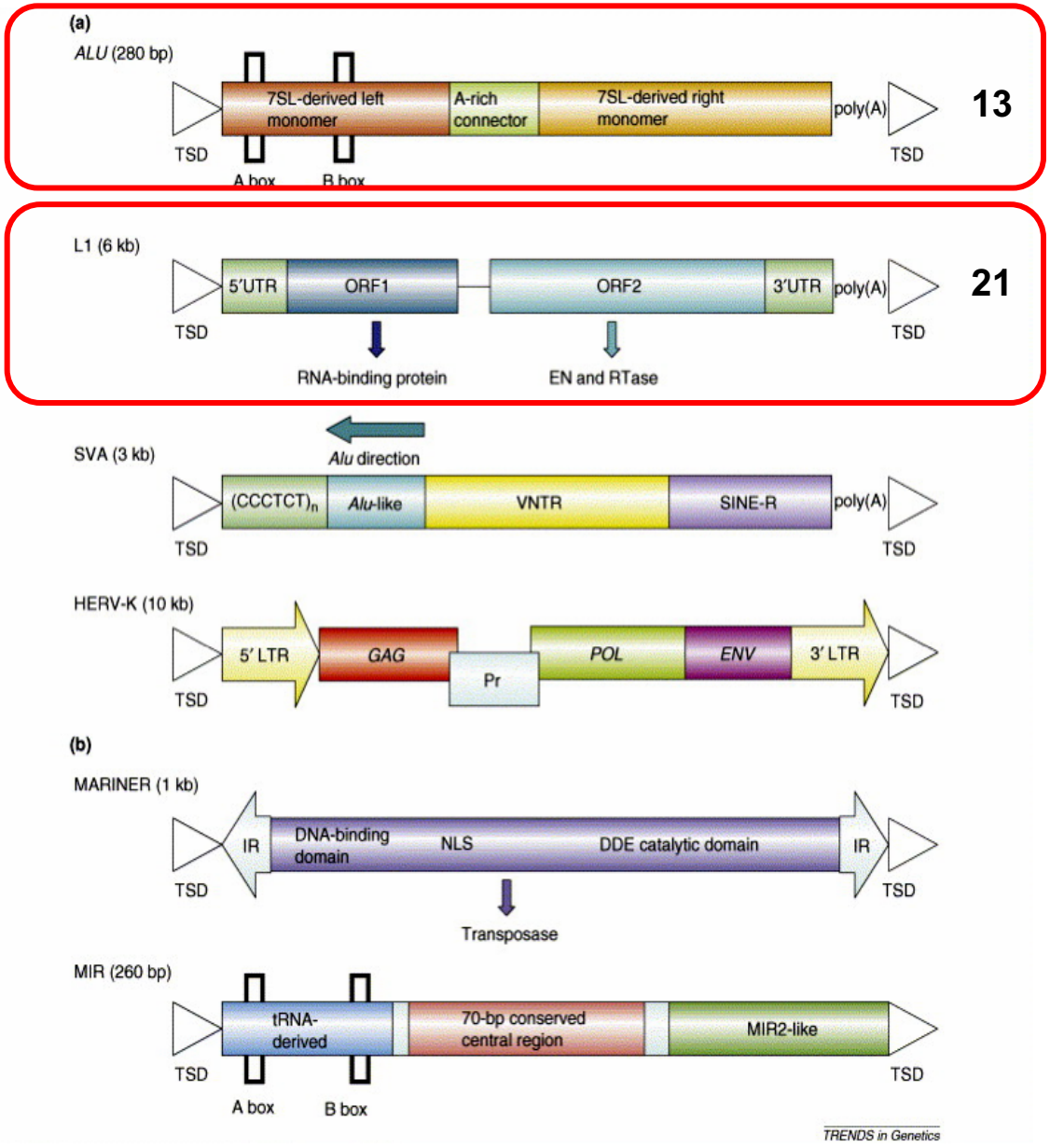
Los elementos repetidos dispersos

%

No autónomo

Autónomo

competentes=147



Mills et al., Trends in Genetics (2007) 23, 183

Elementos repetidos dispersos

	Tipos de DNA repetido disperso	
	SINEs (13%)	**LINEs (21%)**
Familia más representativa	Alu (primate)	Line-1
Longitud del fragmento de repetición	100 a 300 pb	6 a 8 Kbp
Cantidad de copias en el genoma (Mpb)	1.500.0000	850.000 147 competentes

Variabilidad Genética en las poblaciones

Generación de secuencias variables

Tasa de mutaciones.

~ 3000 mutaciones en cada célula a lo largo de la vida.

~ 0.1% variación entre individuos

~ 0.01% variación entre padres e hijos

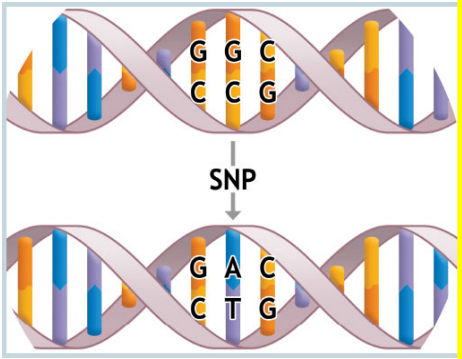
100.000 mutaciones están asociadas a enfermedades (HGMD), 68% sustituciones de un nucleótido, 31% pequeñas indels y 300 mutaciones relacionadas a elementos repetidos

Un locus es polimórfico, cuando presenta al menos dos alternativas alélicas y la frecuencia del alelo menos frecuente es superior al 1% en la población .

Tipos de secuencias variables en las poblaciones

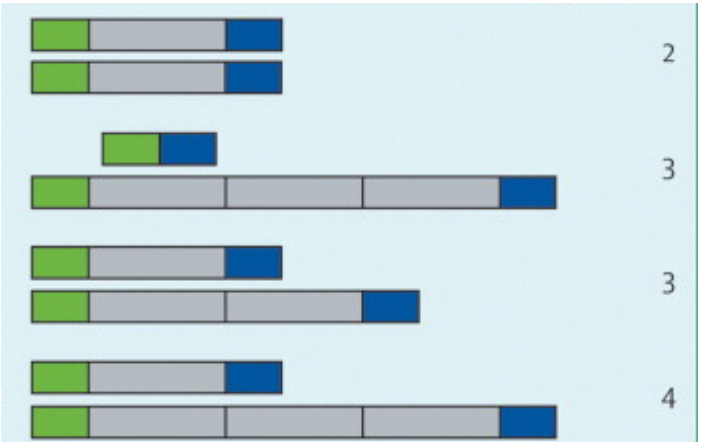
- Single Nucleotide Polymorphisms, SNPs

  Polimorfismos de un nucleótido. Se estima que los SNPs explican aproximadamente el 90% de la variación existente en poblaciones humanas. Uno cada 300bp en promedio, con 2 (mas frecuentes), 3 o 4 alelos. (SNV, single nucleotide variant si la frecuencia es <1%).



- Copy Number Variations, CNV.

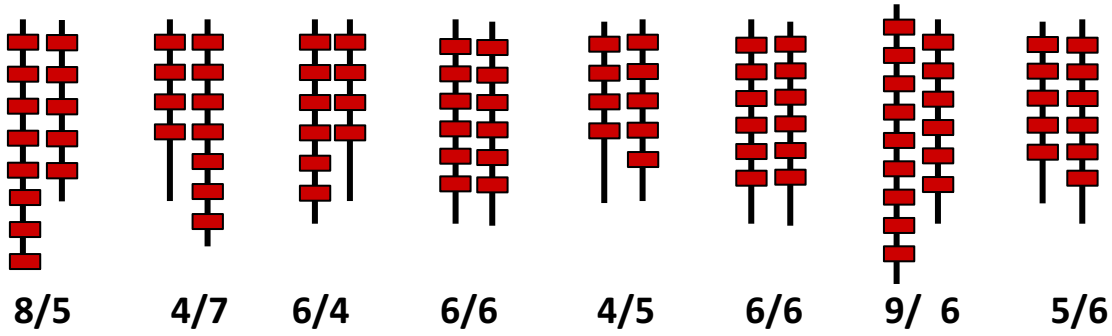
  Variación en el número de copias



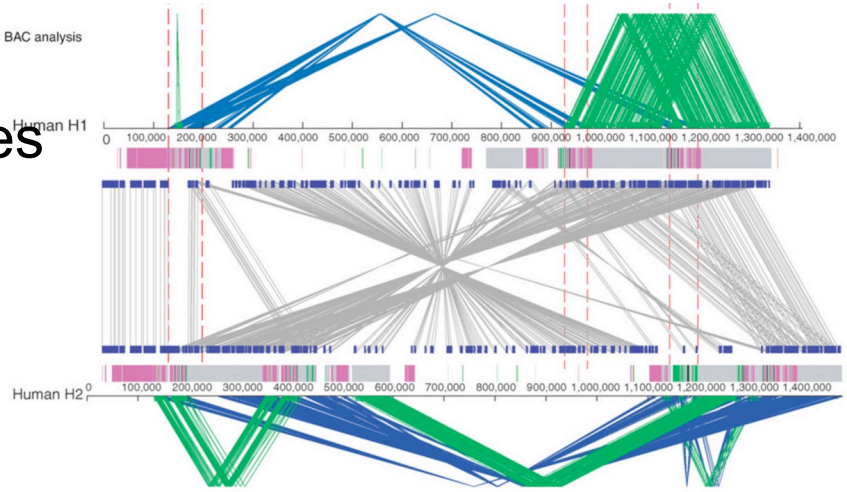
Tipos de secuencias variables en las poblaciones

- **V**ariable **N**umber of **T**andem **R**epeats (**VNTR**).
- **S**hort **T**andem **R**epeats (**STR**). Repetciones de 1-6 bp, distribuidas en todo el genoma. Conglomerado de 150 bp, localizados generalmente fuera de genes.

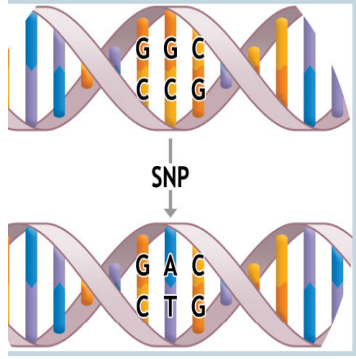
STR analizado en 8 personas (un locus con 6 posibles alelos)



- Alteraciones estructurales genómicas (kb→cientos de kb). Indels, Inversiones



Single Nucleotide Polymorphisms, SNPs



- 10 millones de SNPs en el genoma humano, es decir 1 cada 300bp en promedio.
- Nomenclatura: código rsxxxxxx (reference sequence)

Allele(s):	
IUPAC code	**Meaning**
A	A
C	C
G	G
T	T
M	A or C
R	A or G
W	A or T
S	C or G
Y	C or T
K	G or T
V	A or C or G
H	A or C or T
D	A or G or T
B	C or G or T
N	G or A or T or C

Nomenclatura IUPAC

http://www.ncbi.nlm.nih.gov/snp/

Iniciativas y bases de datos para conocer la variabilidad genética en el mundo

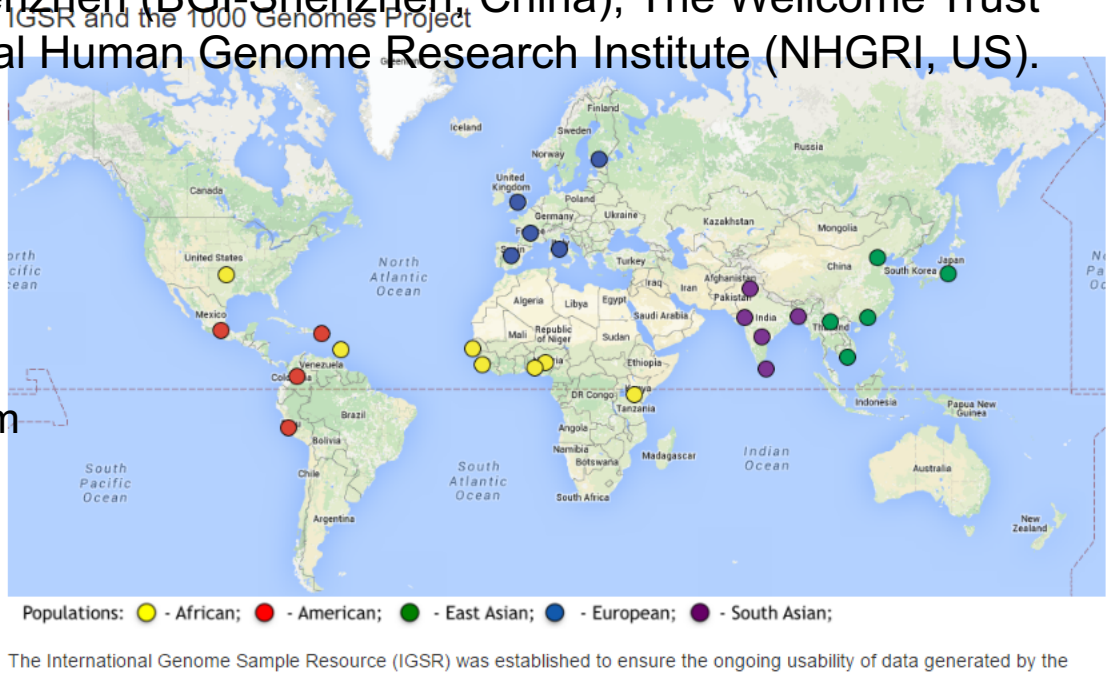
1. HapMap Project: http://www.hapmap.org

- Catálogo de variantes genéticas comunes en seres humanos.
- Describe: que son esas variantes, donde están, como se distribuyen dentro y entre las poblaciones.

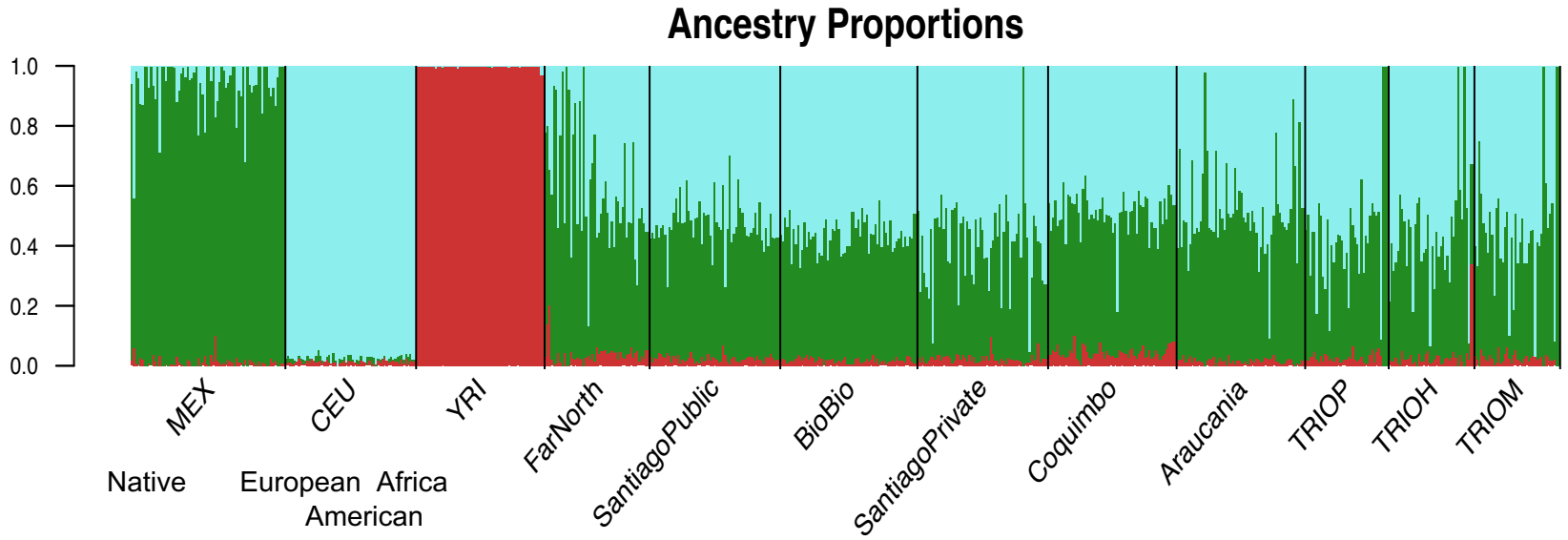
2. The International 1000 Genomes Project.

Enero de 2008. International research consortium. Mapa detallado de variaciones. Beijing Genomics Institute at Shenzhen (BGI-Shenzhen, China), The Wellcome Trust Sanger Institute (UK) and National Human Genome Research Institute (NHGRI, US).

http://www.1000genomes.org/page.php .

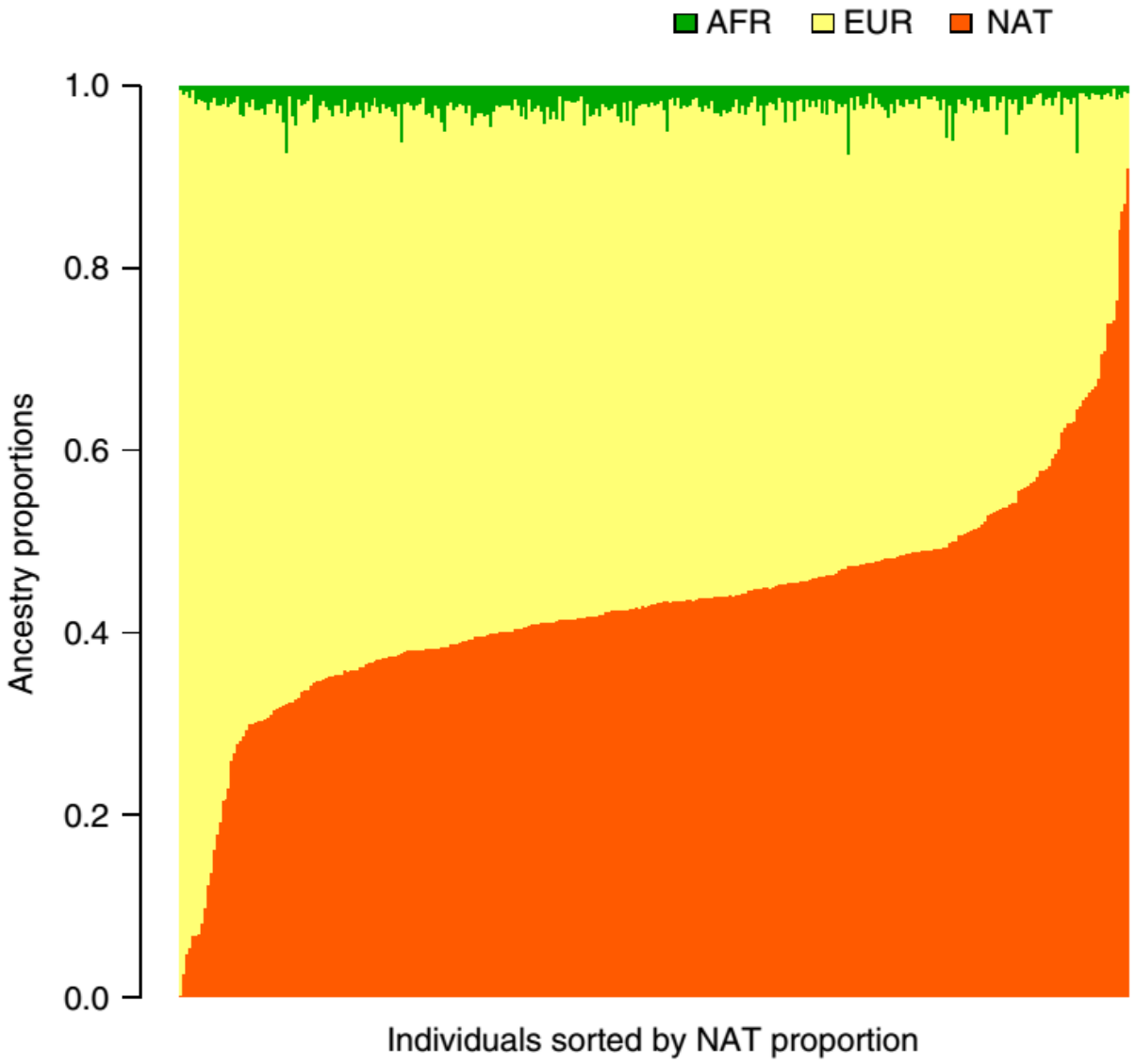


Phases of the project



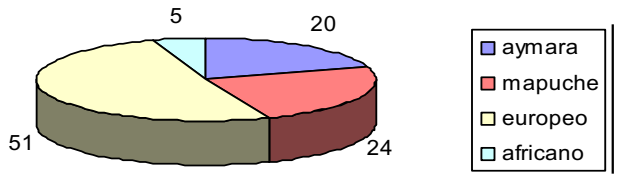
ChileGenomico

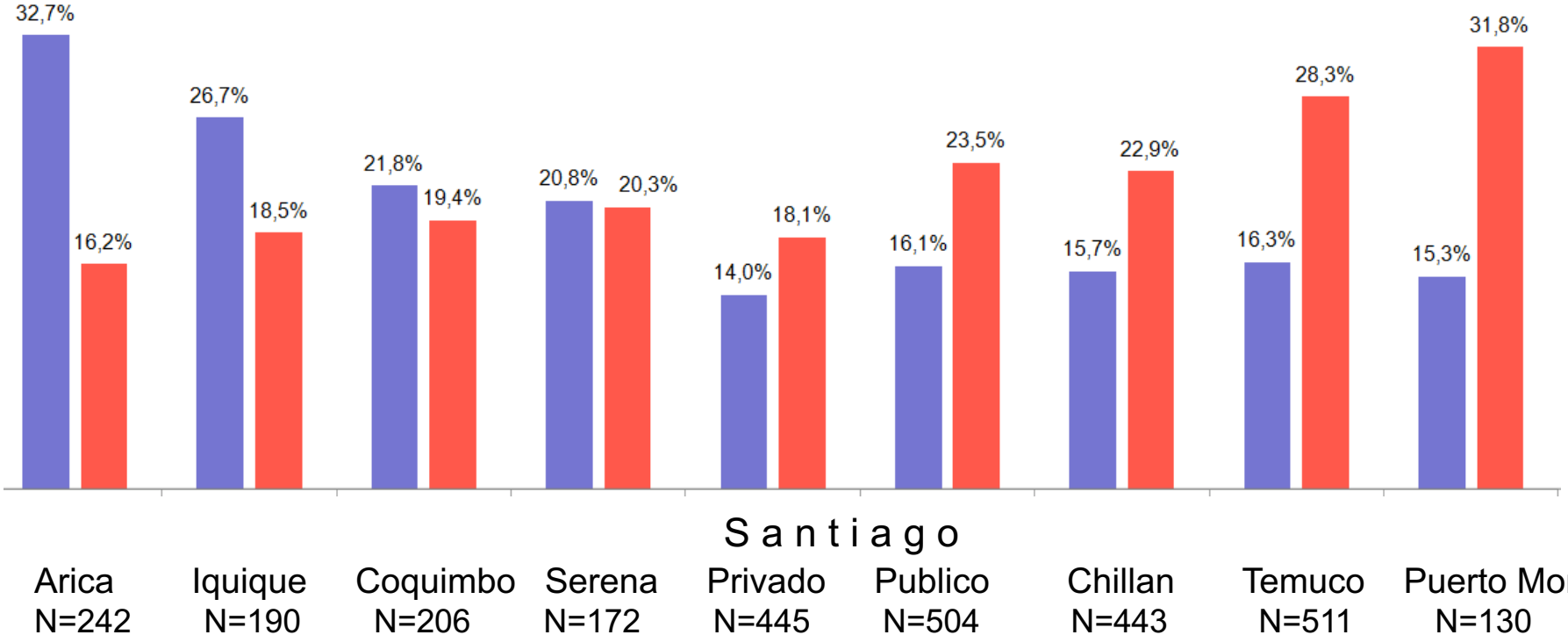
Ancestría de la población chilena



Eyheramendy et al NATURE COMMUNICATIONS| 6:6472

Aymara and Mapuche components along Chile





ChileGenomico

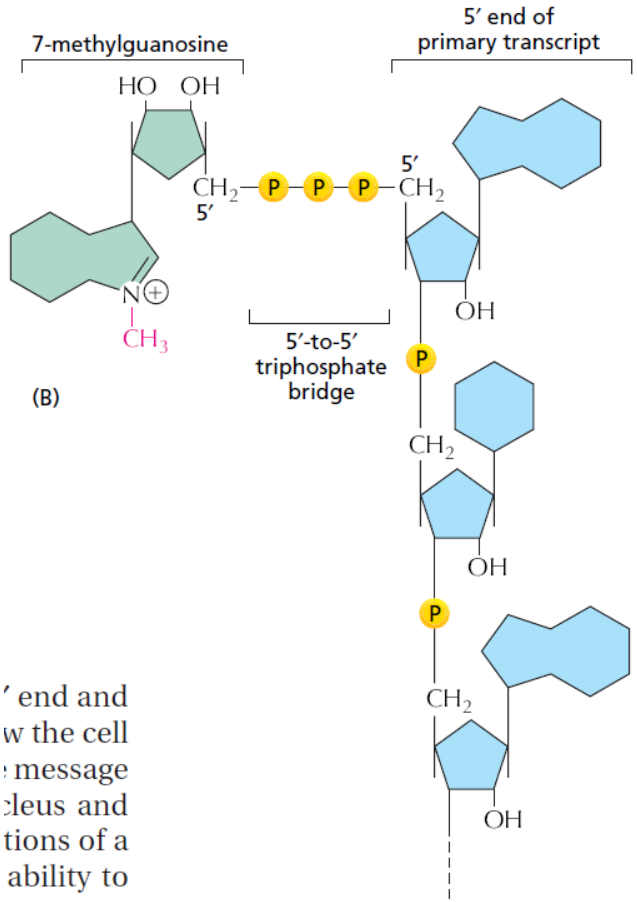
Yuspagara

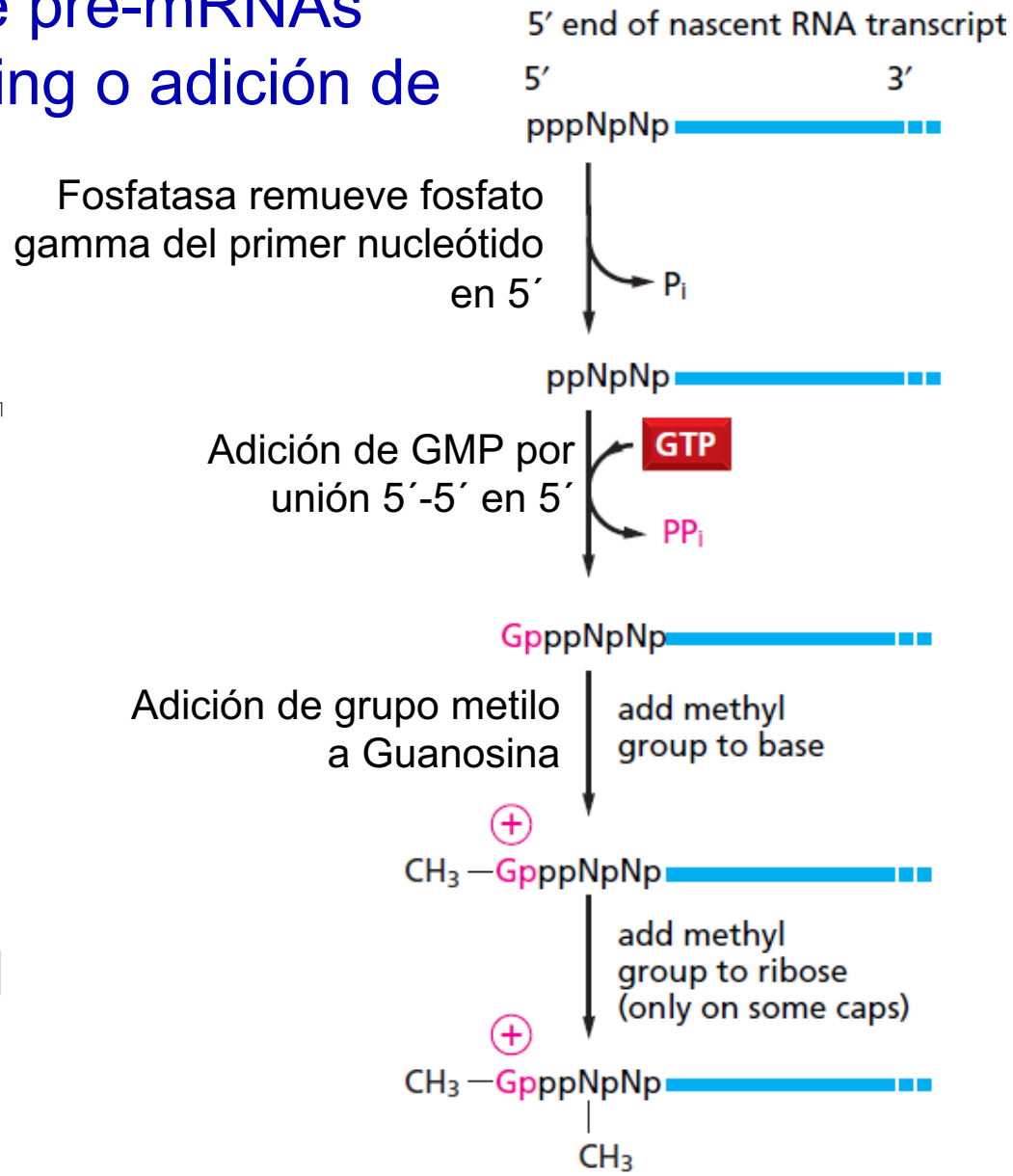
Chaltu

Maururu

Gracias

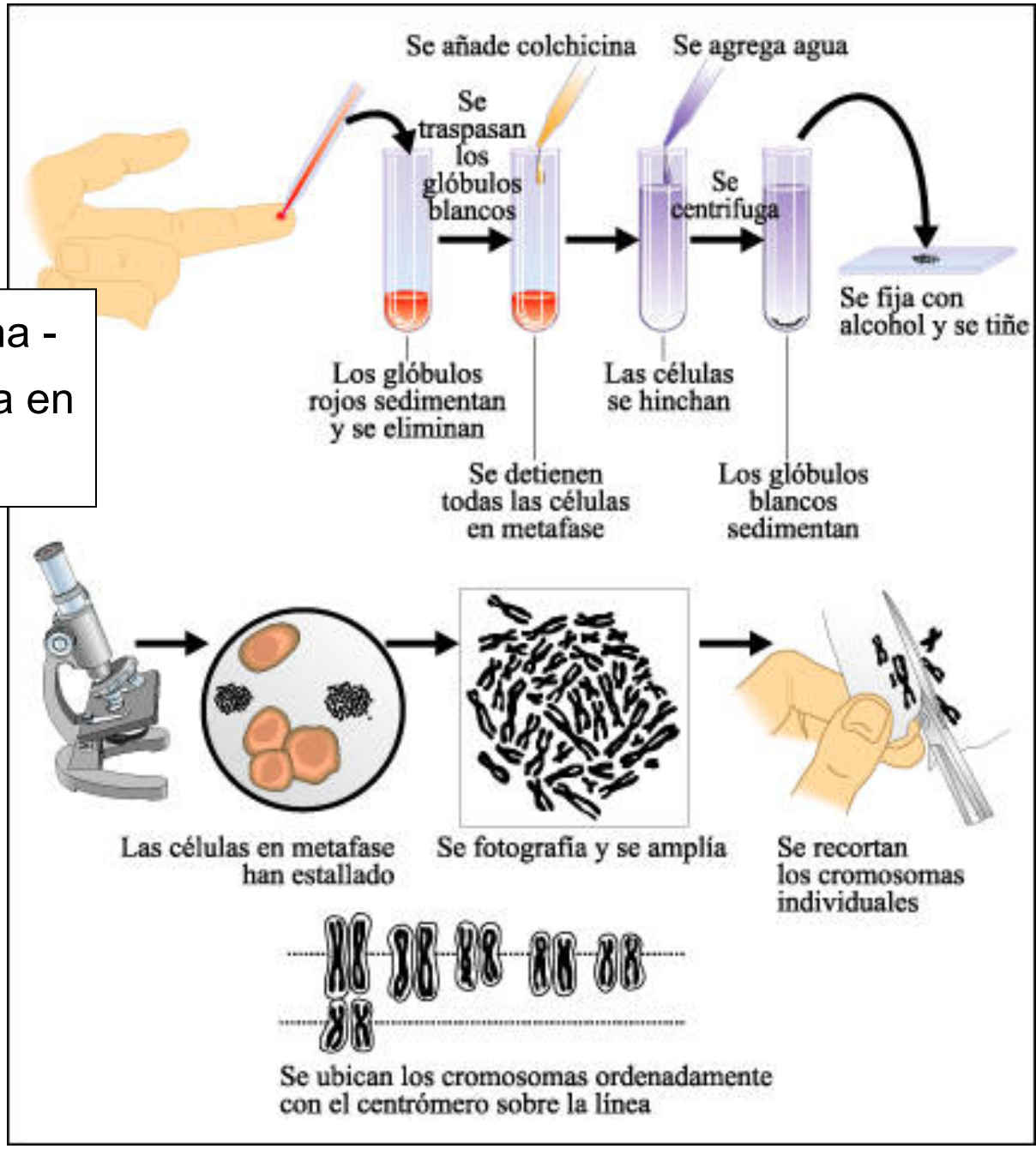
Primera modificación de pre-mRNAs eucariontes: RNA Capping o adición de CAP o capuchón.





Molecular Biology of the cell. Fifth Ed. , 2008
Alberts, Johnson, Lewis, Raff, Roberts, Walter

Cultivo de linfocitos - colchicina - solución hipotónica - se gotea en portaobjeto – tinción giemsa.



Cantidad relativa de DNA organelar en algunas células y tejidos

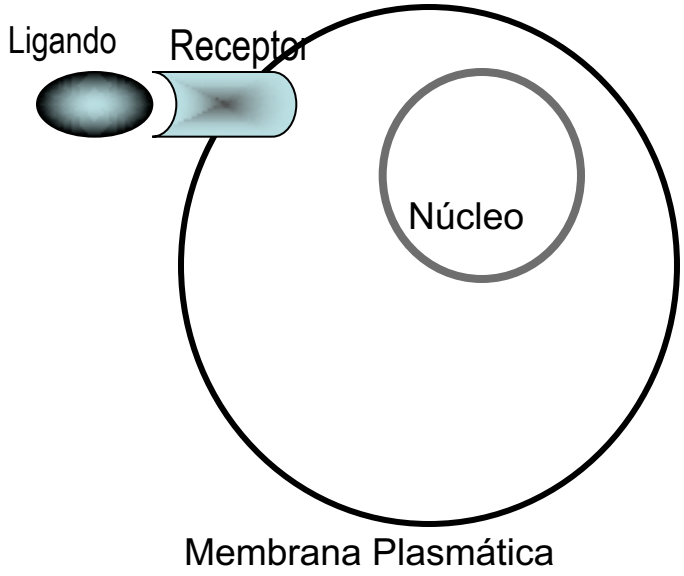
Table 14-3 Relative Amounts of Organelle DNA in Some Cells and Tissues

Organism	Tissue or Cell Type	DNA Molecules per Organelle	Organelles per Cell	Organelle DNA as Percent of Total Cellular DNA
Mitochondrial DNA				
Rat	liver	5–10	1000	1
Yeast*	vegetative	2–50	1–50	15
Frog	egg	5–10	10^7	99
Chloroplast DNA				
Chlamydomonas	vegetative	80	1	7
Maize	leaves	20–40	20–40	15

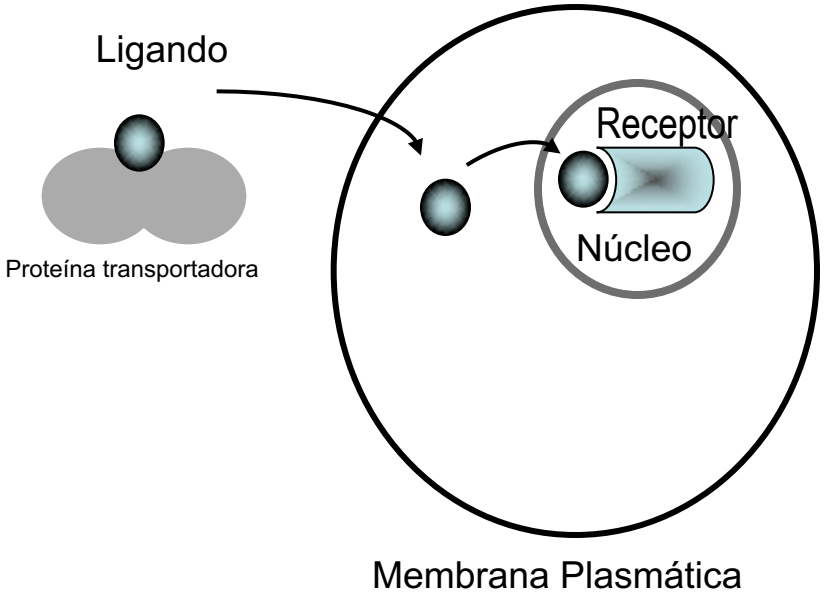
*The large variation in the number and size of mitochondria per cell in yeast is due to mitochondrial fusion and fragmentation.

Control de la función de las células (expresión génica y actividad) en respuesta a señales

Receptor de superficie



Receptor de intracelular



Señales: hormonas, factores de proliferación, o ambientales tales como los nutricinales, ejercicio, etc

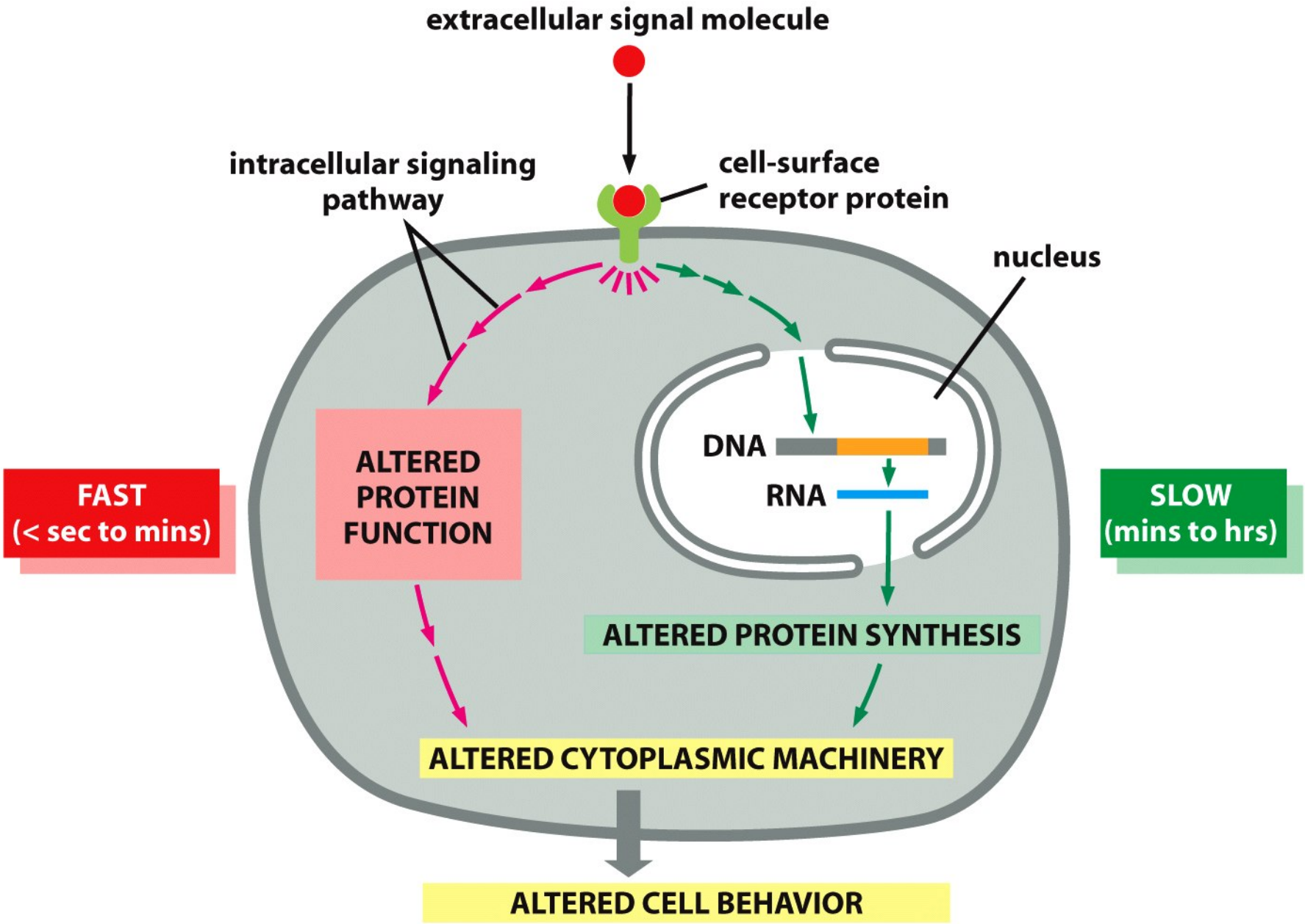
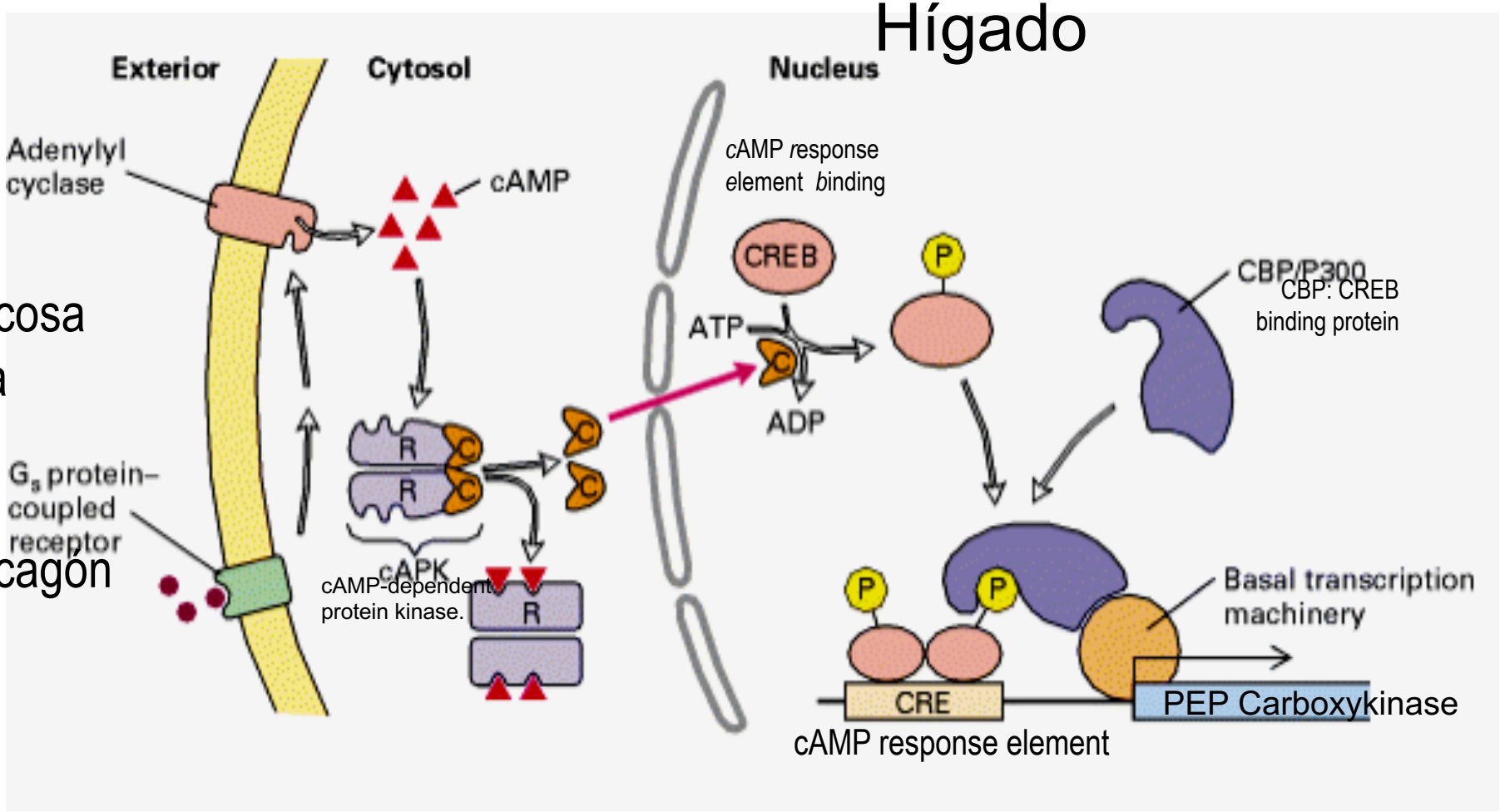


Figure 15-6 Molecular Biology of the Cell 5/e (© Garland Science 2008)

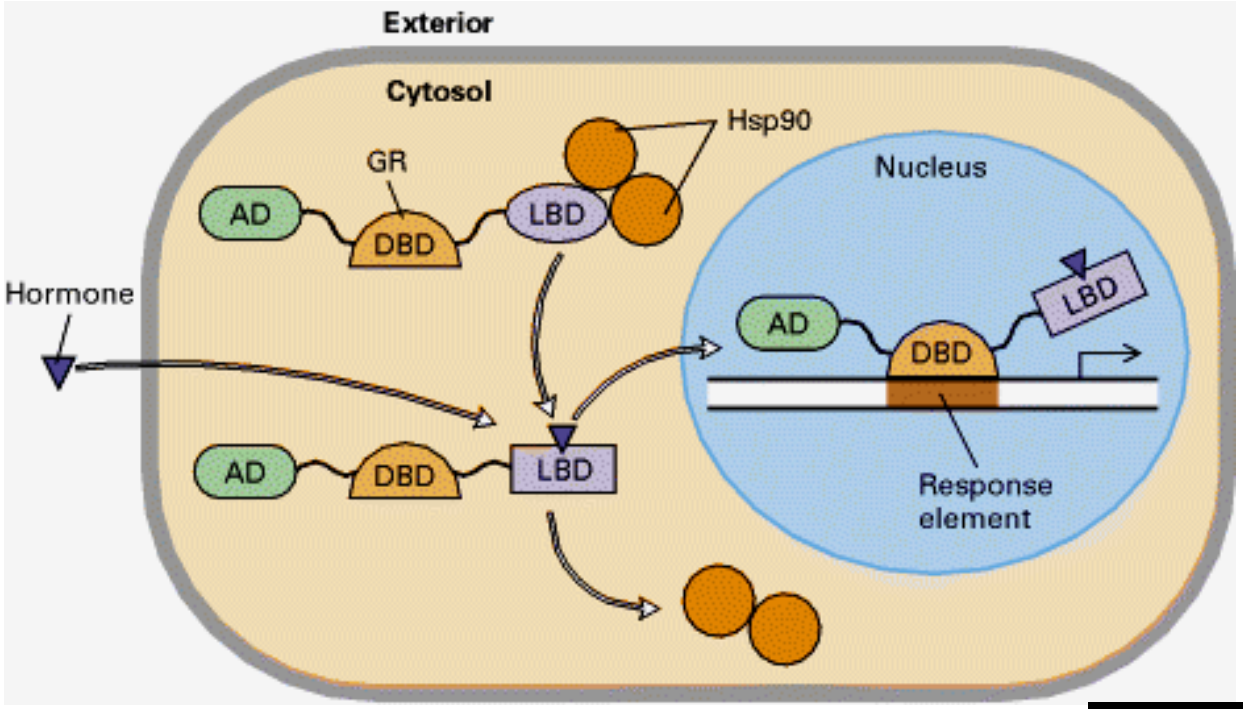
Modelo de activación de expresión génica por activación del receptores de membrana. Receptores acoplados a proteínas |G



CREB regula la expresión de enzimas gluconeogénicas

Nature Reviews Molecular Cell Biology 2, 599 -609 (2001)

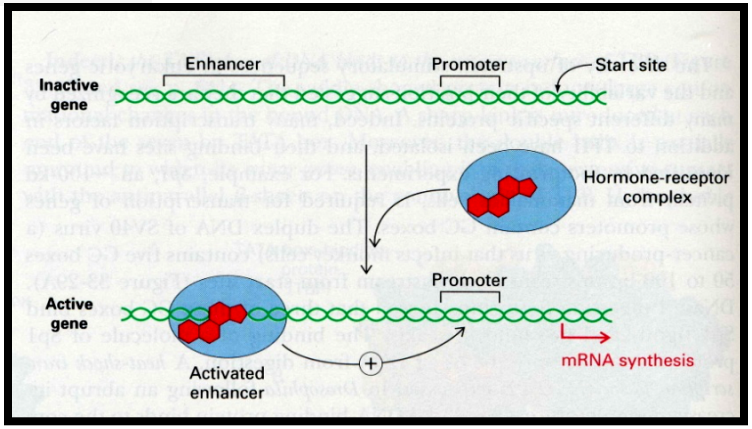
Modelo de activación de expresión génica por activación del receptor esteroidales. Glucocorticoides



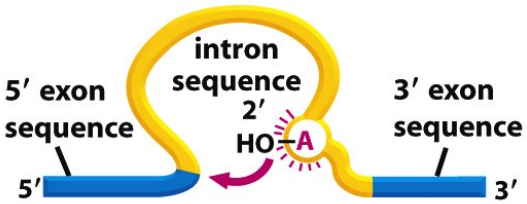
Unión a ligando

unión del receptor al DNA

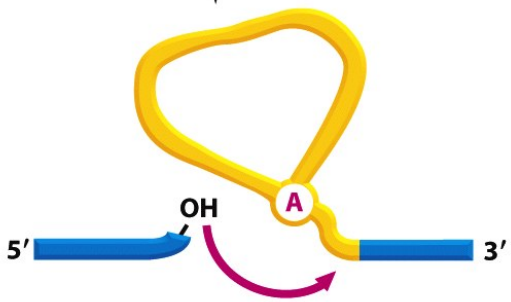
Ensamblaje de complejos de inicio



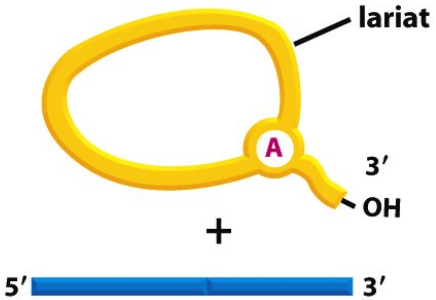
Mecanismo químico del splicing

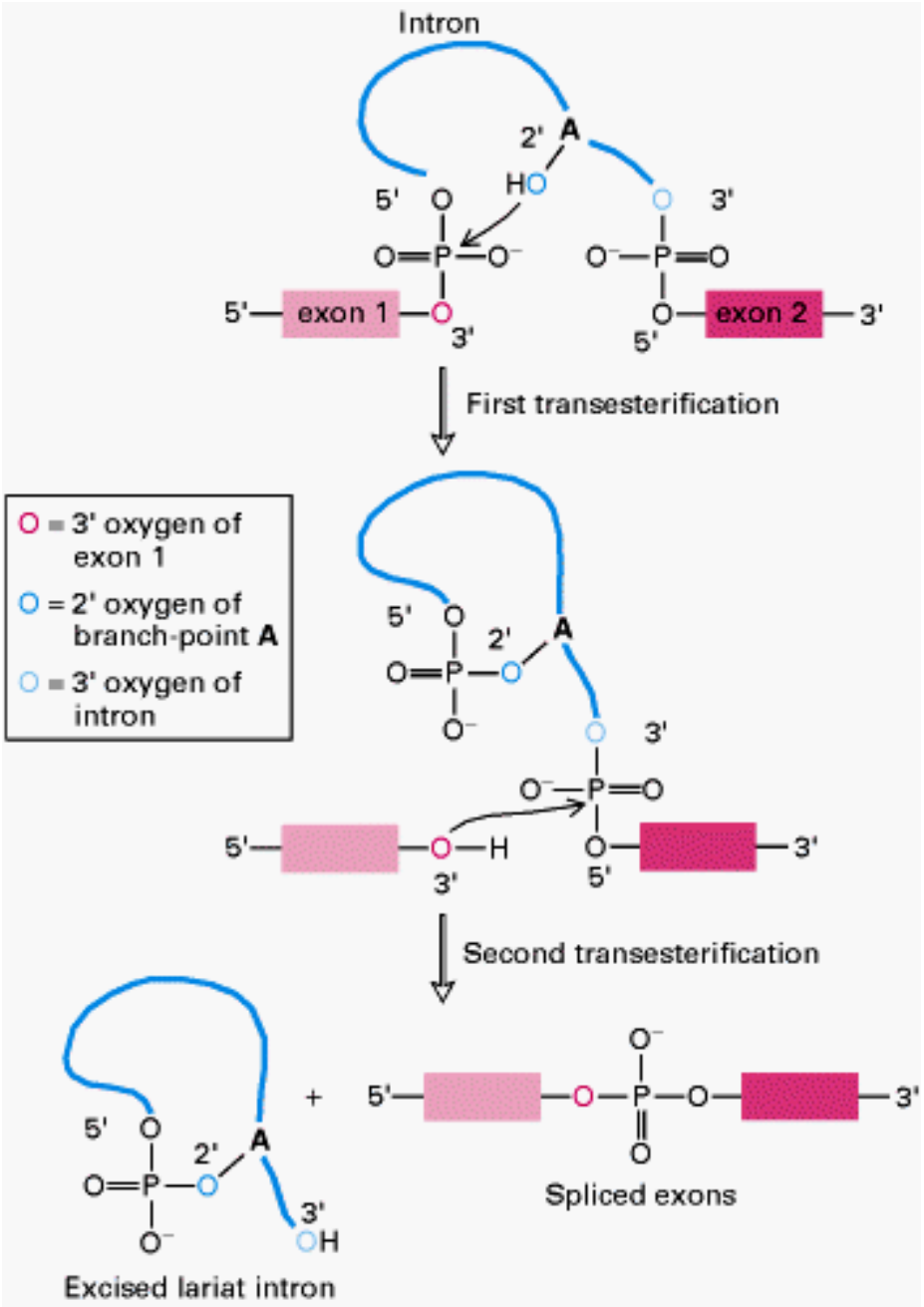


1. Corte en el sitio 5’ o dador de splicing



2. Corte en el sitio 3’ o aceptor de splicing

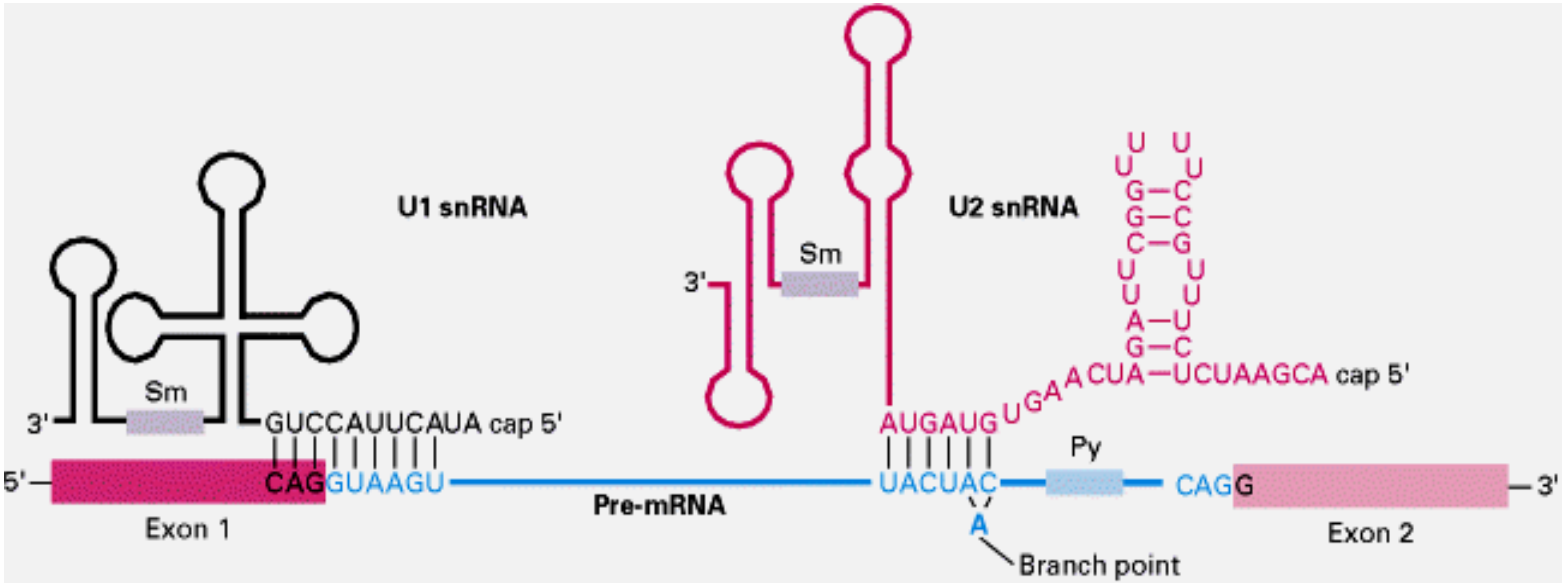




Dos reacciones de transesterificación

Mecanismo químico del splicing...

small nuclear RNAs (snRNAs)



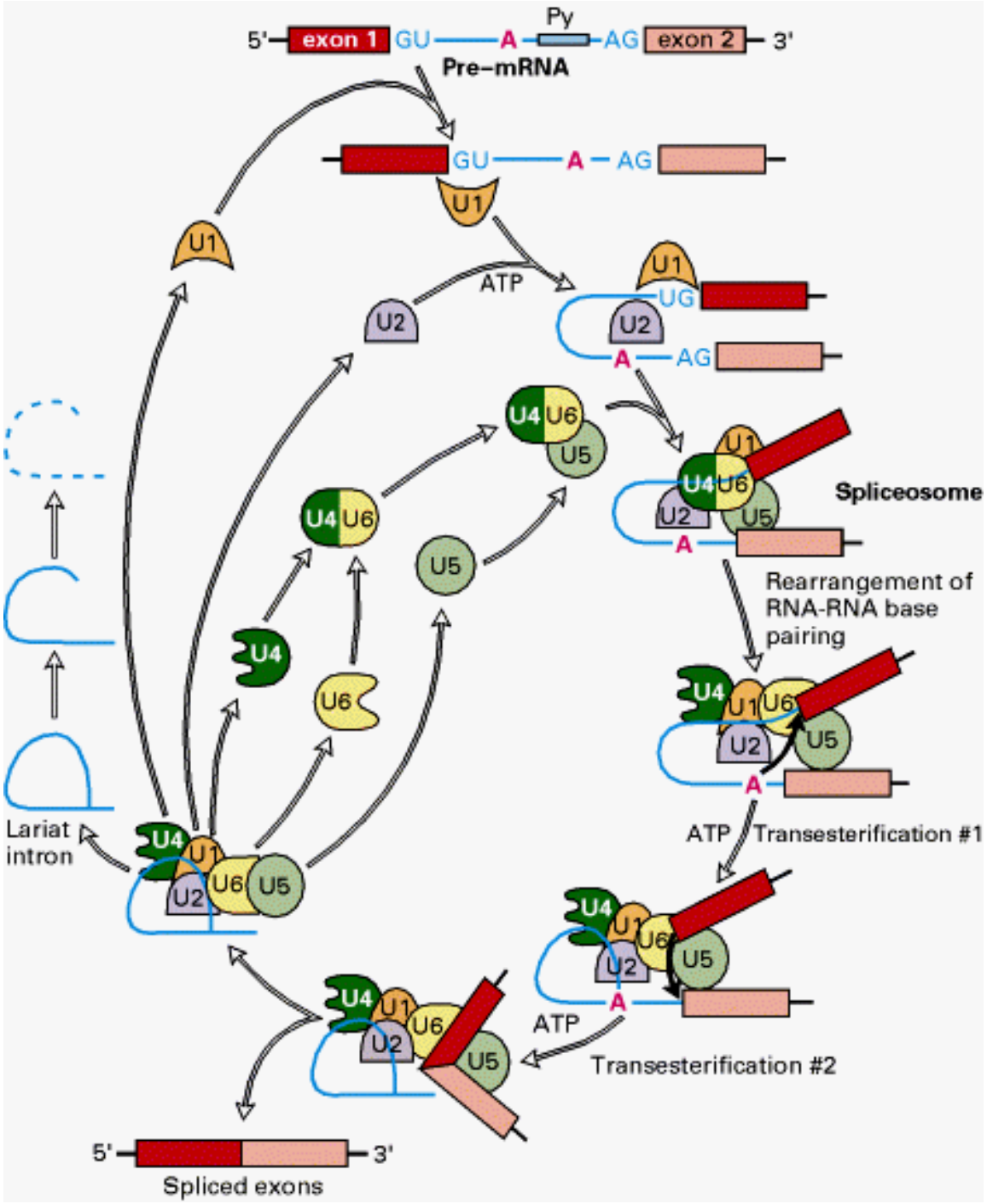
Spliceosomas

Interacciones RNA-RNA

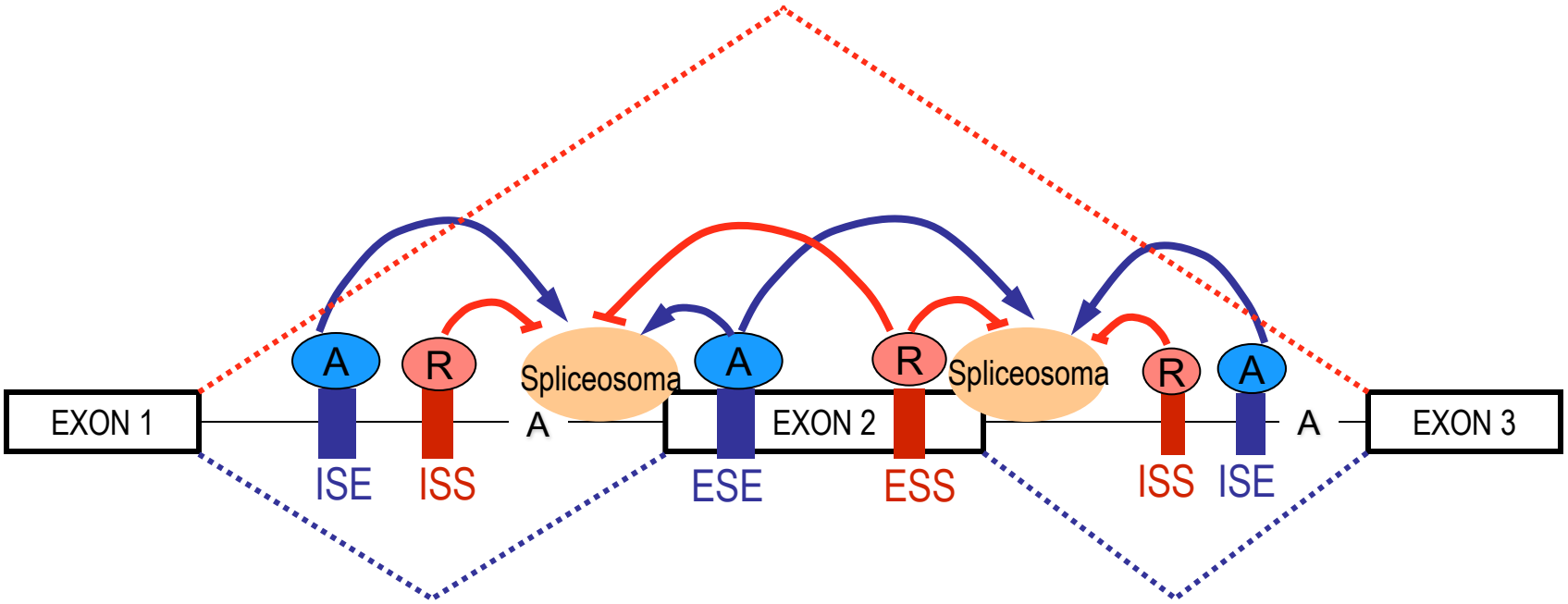
U6 snRNA / U4 snRNA.
U6 snRNA / U2 snRNA,

Etc...

Spliceosoma también interactúa con CTD de RNA pol II



Regulación del splicing alternativo



Enhancers de splicing exónicos (ESE)

Enhancers de splicing intrónicos (ISE)

Silenciadores de splicing exónicos (ESS)

Silenciadores de splicing intrónicos (ISS)

Sitio dador de splicing (sitio 5´ de splicing): límite exon-intron en el extremo 5´ del intron.

splice site: Sitio aceptor de splicing (sitio 3´ de splicing): límite intron-exon en el extremo 3´ del intron.

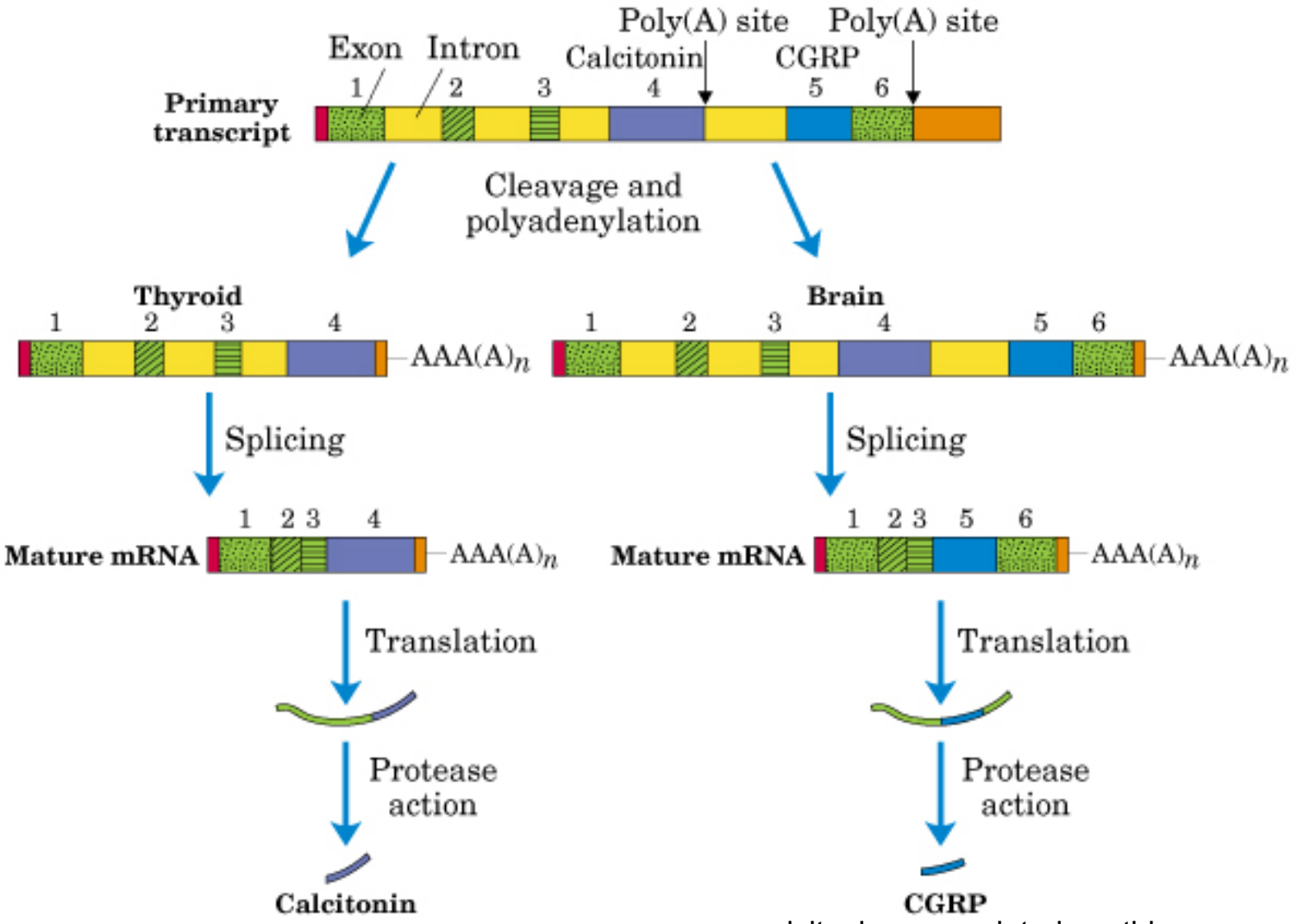
Punto de ramificación: una A cercana al extremo 3´ del intron, la cual es seguida por una secuencia de poli-pirimidina

El splicing es realizado en un gran complejo ribonucleoproteico denominado spliceosoma.

El spliceosoma contiene alrededor de 150 proteínas y 5 small nuclear RNAs (snRNAs, U1, U2, U4, U5, and U6, 100-300 nt).

Muchas funciones del spliceosoma son realizadas por los snRNAs.

Splicing alternativo y poliadenilacion alternativa



calcitonin gene-related peptide

Splicing alternativo posibles (59% de los genes)

Gen promedio:

- 8 exones
- 145 nt/exon
- Intrones al menos diez veces mas grandes

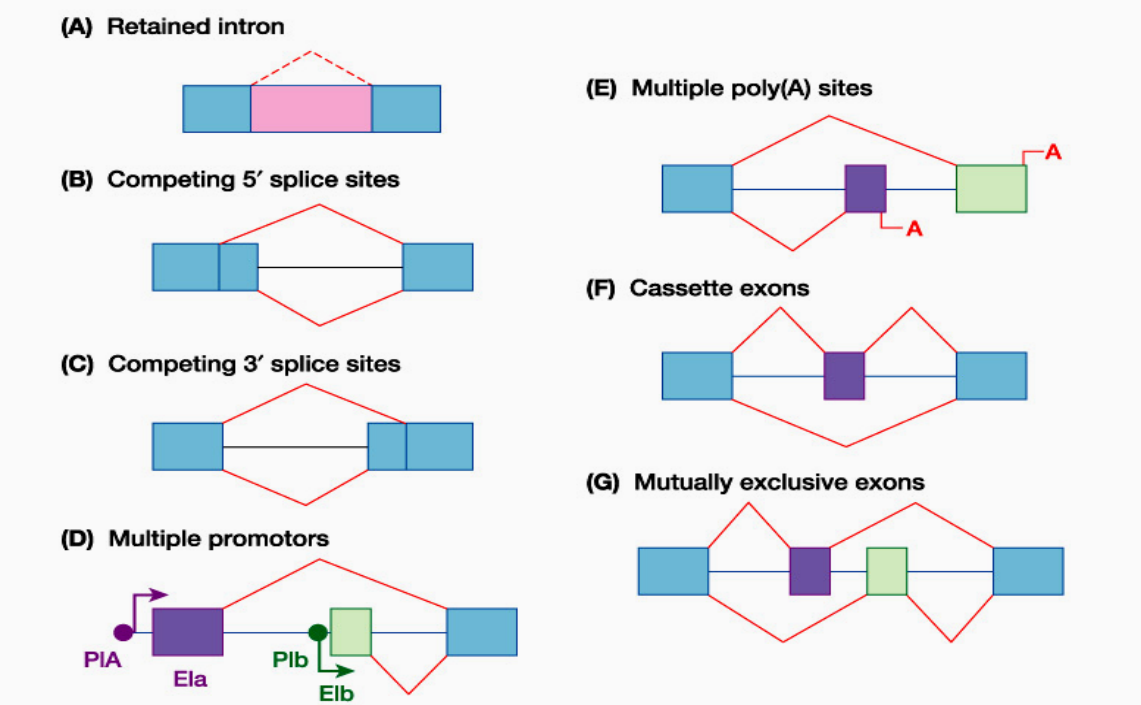


Figure 10-15 Human Molecular Genetics, 3/e. (© Garland Science 2004)

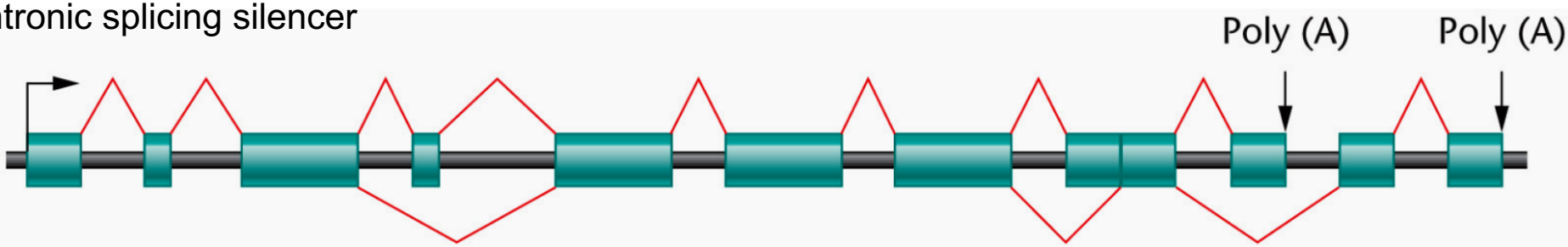
Elementos enhancer o supresores de splicing (6 bp). Poco conservados.

Exonic splicing enhancers (exones) (ESEs)/serine/arginine (SR)-rich splicing factors. SR1,2,3,4…

Intronic splicing

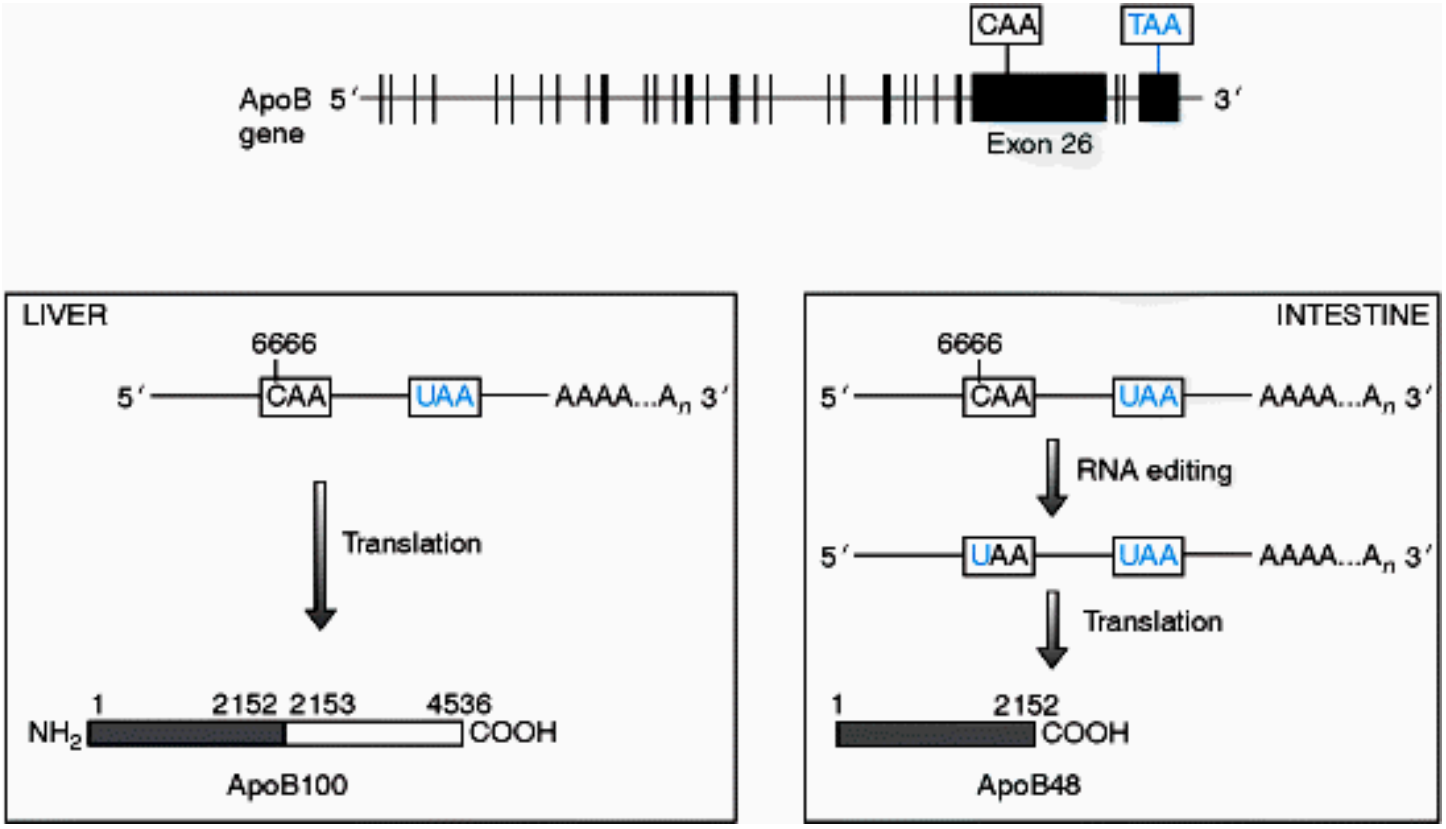
Exonic splicing silencer (ESS)

Intronic splicing silencer



Copyright © 2006 Pearson Prentice Hall, Inc.

Regulación post transcripcional. Edición del RNA



Producida en hígado.

LDL presentes en todas las células.

Transporta lípidos y colesterol endógenos

Producida en intestino delgado.

Transporta lípidos dietarios (QM), desde el intestino al hígado.

Procesamiento y mecanismo de acción de los RNAs pequeños

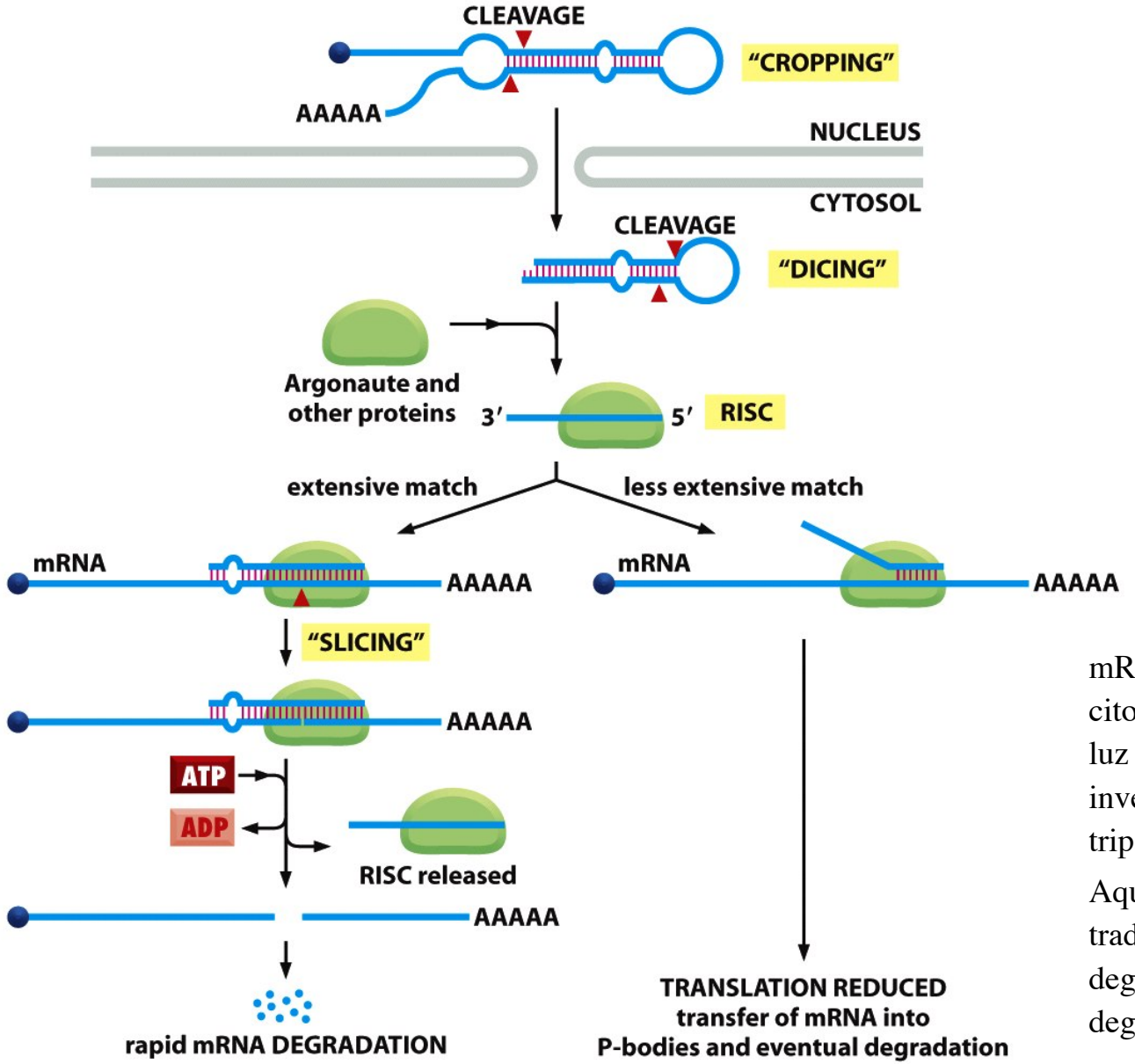


Figure 7-112 Molecular Biology of the Cell 5/e (© Garland Science 2008)

mRNA-processing *(P) bodies:* focos citoplasmaticos visibles al microscopio de luz en celulas somatticas de vertebrados, invertebrados, levaduras, plantas y tripanosomas.

Aquí colocalizan factores de inicio de traducción, deadenilación, decapping degradación exonucleolítica 5´→3´, degradación mediada por miRNA.

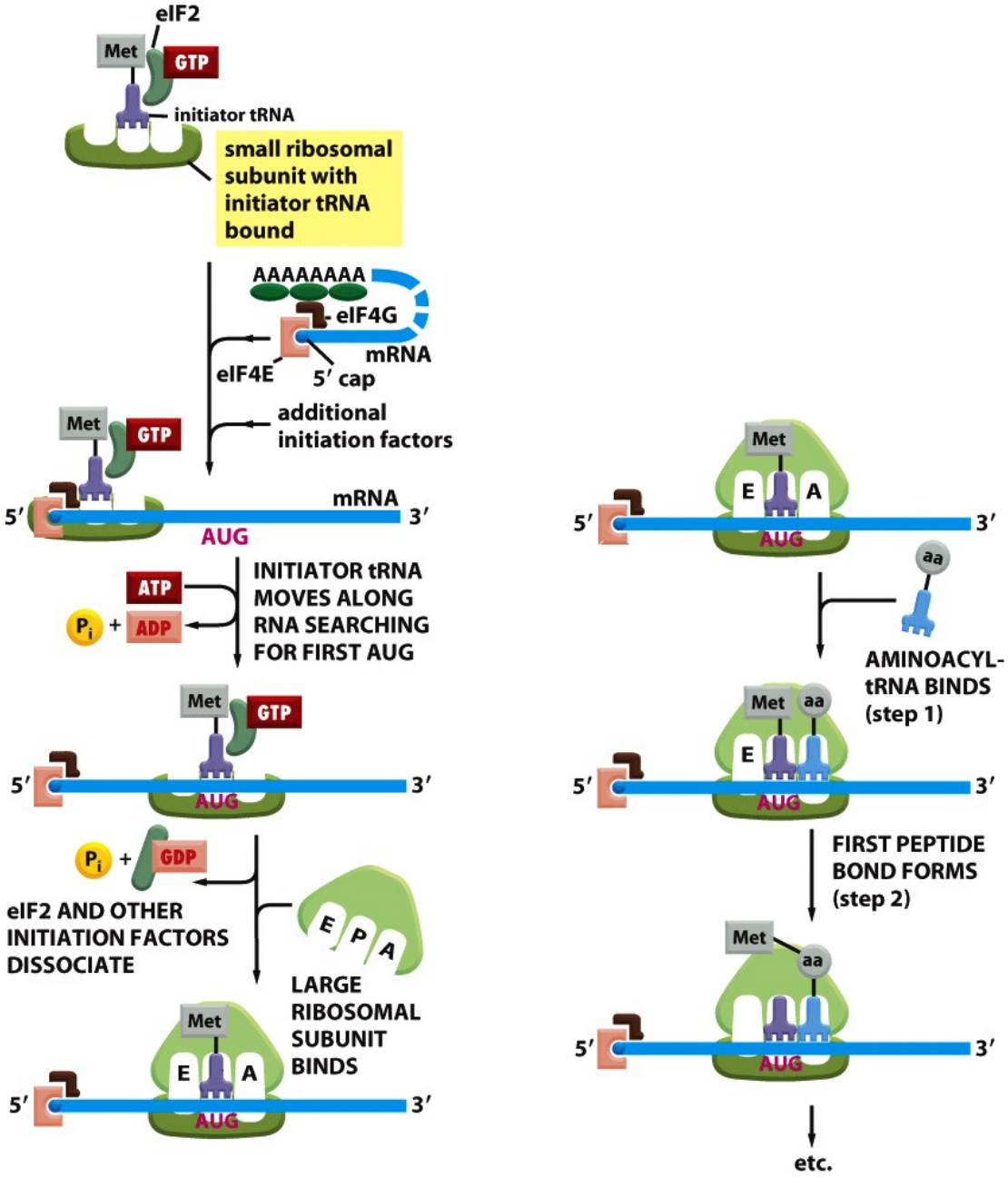


Figure 6-72 *Molecular Biology of the Cell* (© Garland Science 2008)

Inicio de traducción en CAP → AUG o en IRES (internal ribosome entry site)

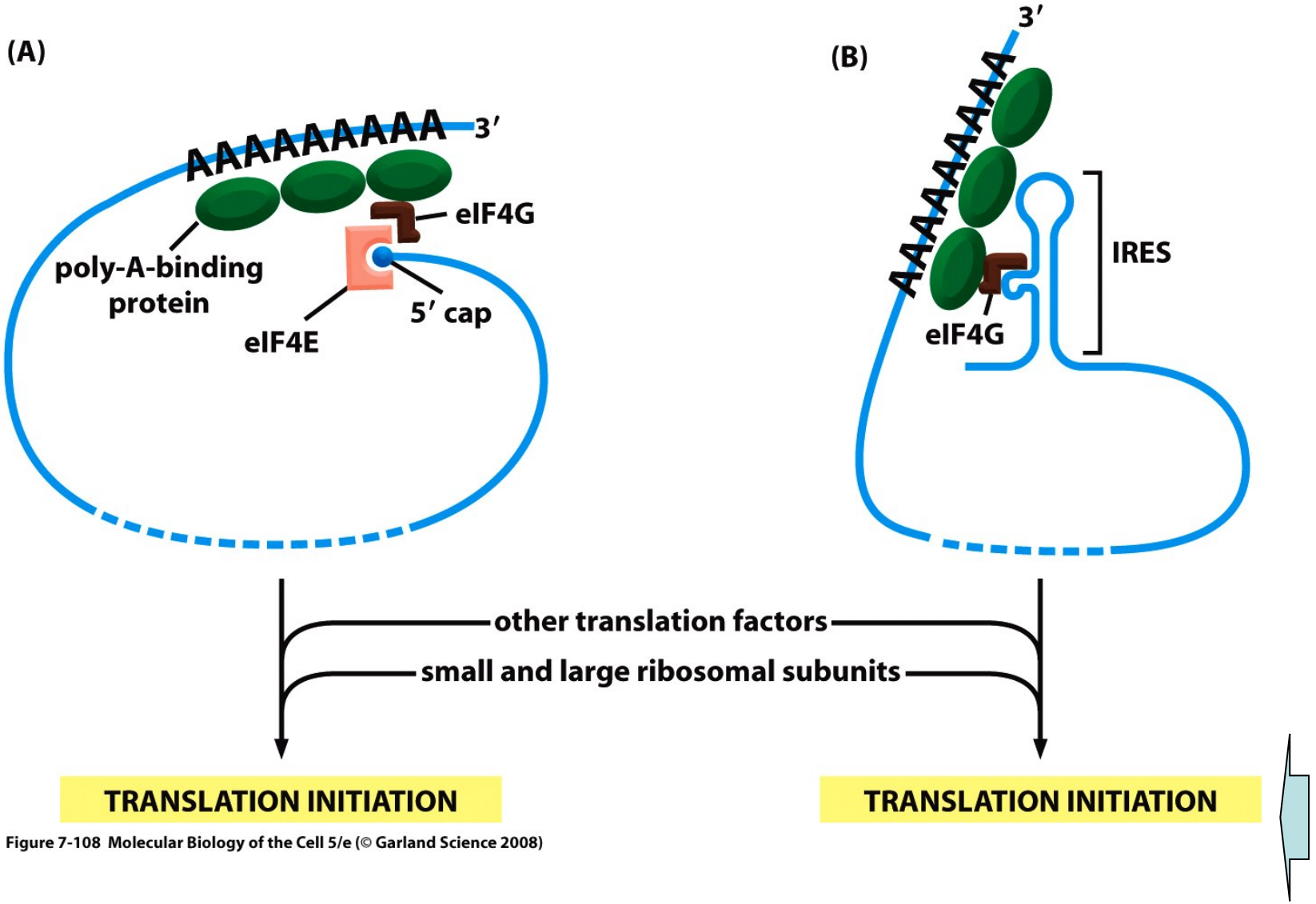


Figure 7-108 Molecular Biology of the Cell 5/e (© Garland Science 2008)