



BIOINFORMÁTICA

Objetivo



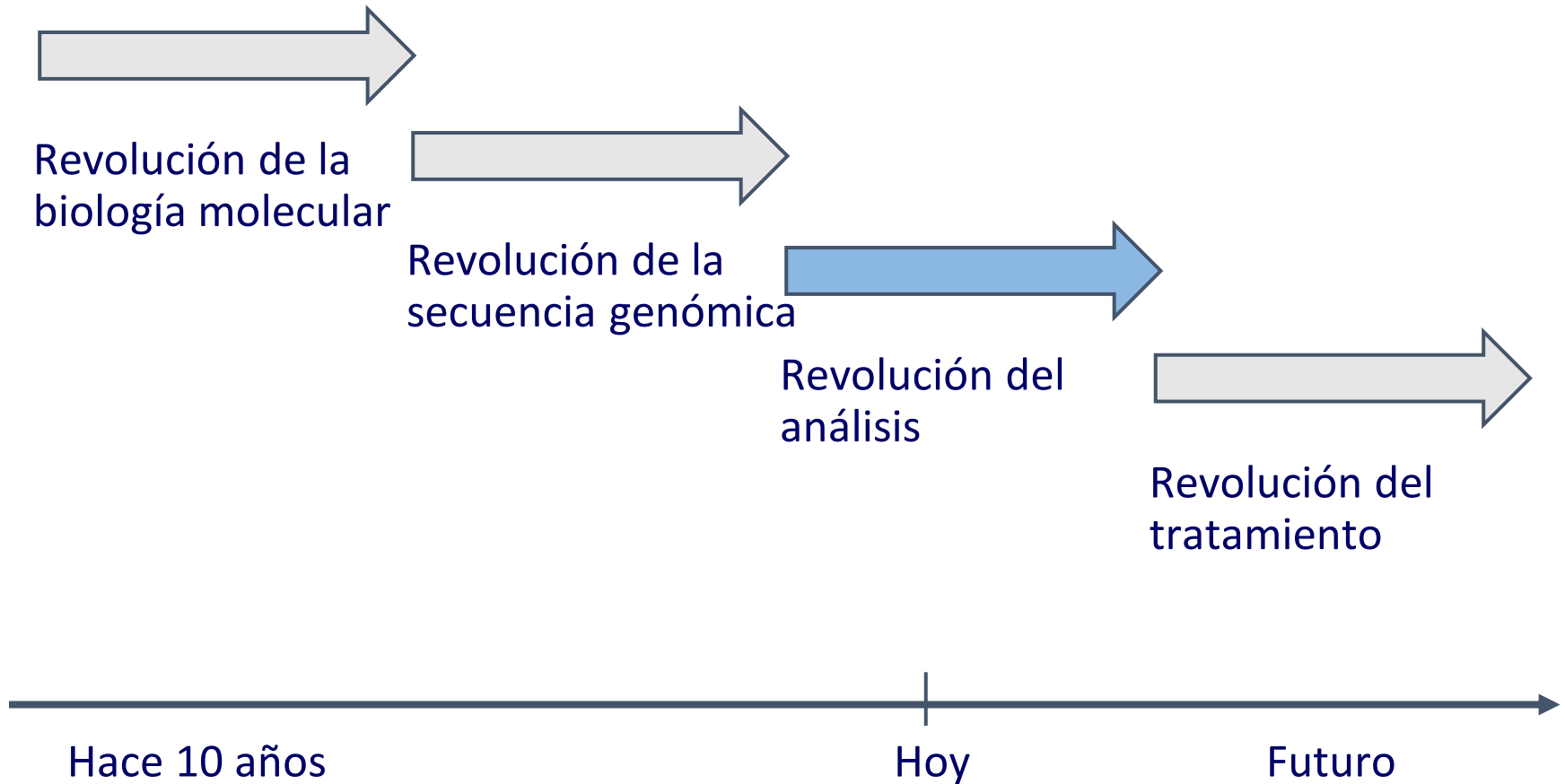
- Conocer los conceptos de la bioinformática y su relevancia para la práctica de la medicina.

Una frontera digital de la ciencia



- Campo interdisciplinario
 - Medicina
 - Desarrolla métodos
 - Aplica *software*
 - Estadística e ingeniería
 - Genética, genómica

¿Evolución en la medicina?





Bioinformática

Conceptos y
aplicaciones
biológicas en
moléculas

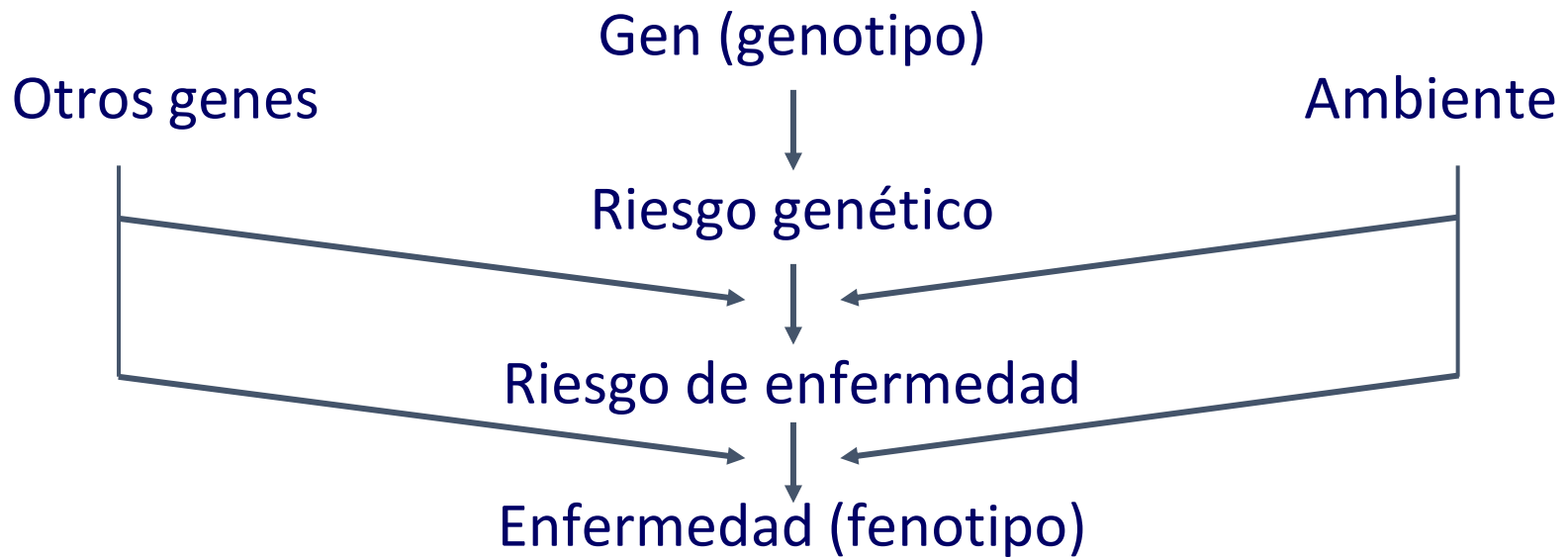
Técnicas informáticas

- Matemáticas
- Estadística
- Ciencias computacionales
 - Big data
 - Cómputo distribuido

Medicina genómica



- ¿Cuál es el riesgo de enfermar si existe alguna mutación genética específica?
- ¿Puede ayudar la medicina?



- 45411789

5'pad=0 3'pad=0 strand=+ repeatMasking=none

gtgagaagcgcgatcgggggcacggggatgagctcaggggcctctagaaa
gagctgggaccctgggaacccctggcctccaggtagtctcaggagagcta
ctcggggctcgggcttggggagaggaggagcggggggtgaggcaagcagcag
gggactggacctgggaagggtgggcagcagagacgacccgacccgctag
aaggtgggggtggggagagcagctggactgggatgtaagccatagcaggac
tccacgagttgtcactatcattatcgagcacctactgggtgtccccagt
gtcctcagatctccataactggggagccaggggcagcgacacggtagcta
gccgtcgattggagaactttaaatgaggactgaattagctcataaatgg
aacacggcgcttaactgtgaggttggagcttagaatgtgaagggagaatg
aggaatgcgagactgggactgagatggaaccggcggtggggaggggggtgg
ggggatggaatttgaacccgggagaggaagatggaattttctatggagg
ccgacctggggatggggagataagagaagaccaggaggaggattaaatagg
gaatgggttgggggcggccttggtaaatgtgctgggattaggctgttcag
ataatgcaacaaggcttgaaggctaacctggggtgaggccgggttgggg
ccgggctgggggtgggaggagtccctactggcggttgattgacagtttct
ccttcccaggatatgggggcggggccttgctcggttcccccgctcctccc
cctctatcctcacctcaacctcctggccccattcaggcagacctgggc
ccccttctgaggcttctgtgctgcttctggctctgaacagcgatttg
acgctctctgggcctcggttcccccatccttgagataggagttagaagt
tgttttgttgttgttgttgttgttgttgttgttgttgttgttgttgttgt
tctcgctctgtcgccaggctggagtgcaatggcggtatctcggtcact

ADN codificante y no codificante



- Codificante: secuencias que dictan proteínas.
 - 5% del ADN en vertebrados.
 - Se conoce el código y secuencias.
- No codificante: secuencias reguladoras que controlan cuándo, cuánto y dónde un gen se va a transcribir.
 - Se conoce el alfabeto base, pero el lenguaje es en gran medida desconocido.

Los genes humanos



- 30,000 genes.
- Regiones del genoma que codifican proteínas
 - se transcriben para generar ARN mensajeros
 - proteínas en ribosomas
- 3000 nucleótidos o bases, media de gen
- 3,164.7 millones de bases en genoma
 - todos los genes suponen el 2% del genoma

La transcripción de un gen

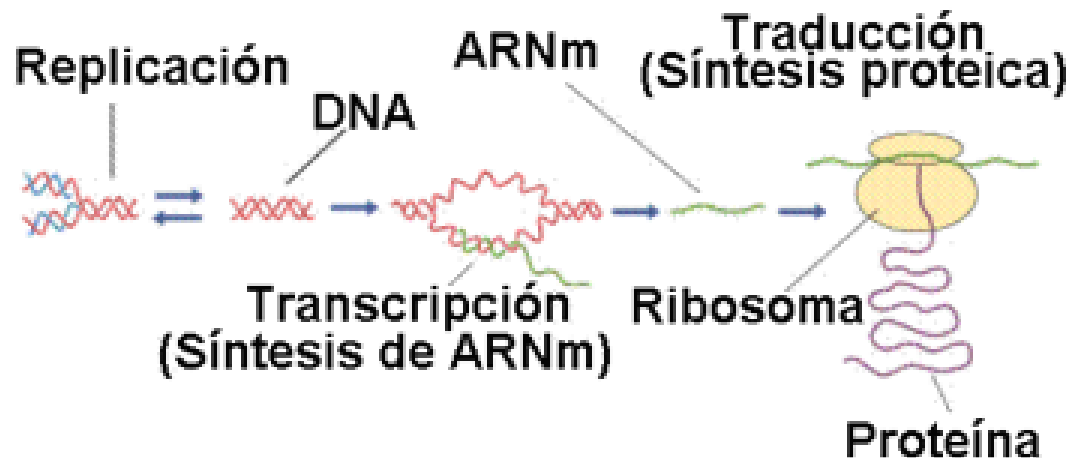


- La expresión génica o transcripción hace posible la diferenciación celular.
 - Según el tipo de célula se activa un subconjunto.
- En las regiones promotoras se controla el inicio de la transcripción, allí se une la polimerasa ARN para iniciar la transcripción
 - los *enhancers* y los *silencers* potencian o evitan el reclutamiento de polimerasa ARN al promotor de un gen, favoreciendo o bloqueando la transcripción génica.

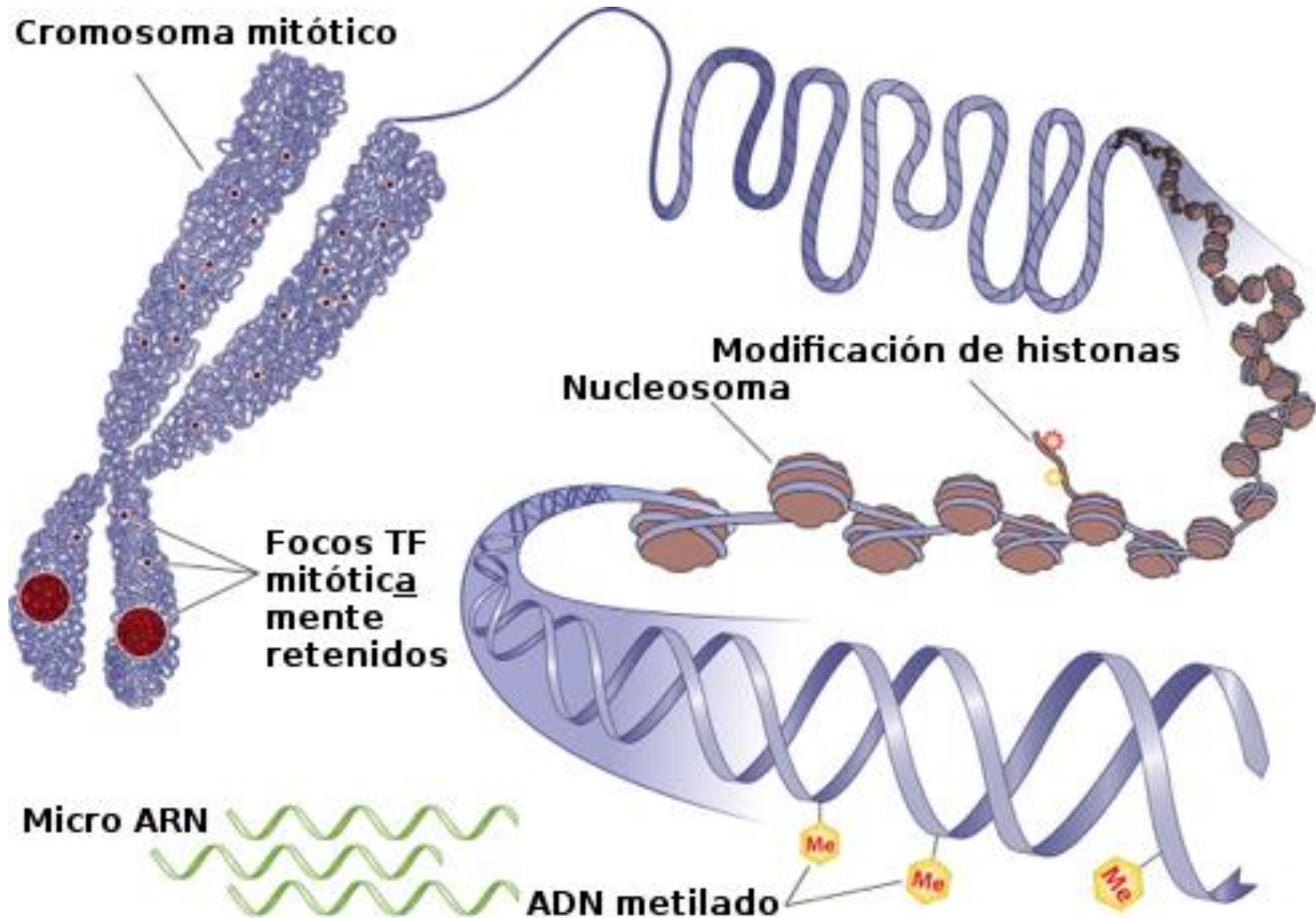
Cantidad de información



- Información de la secuencia genómica
- ARNm (expresión específica de tejido)
- Proteínas
 - Secuencia
 - Estructura
 - Función



Regulación del ADN



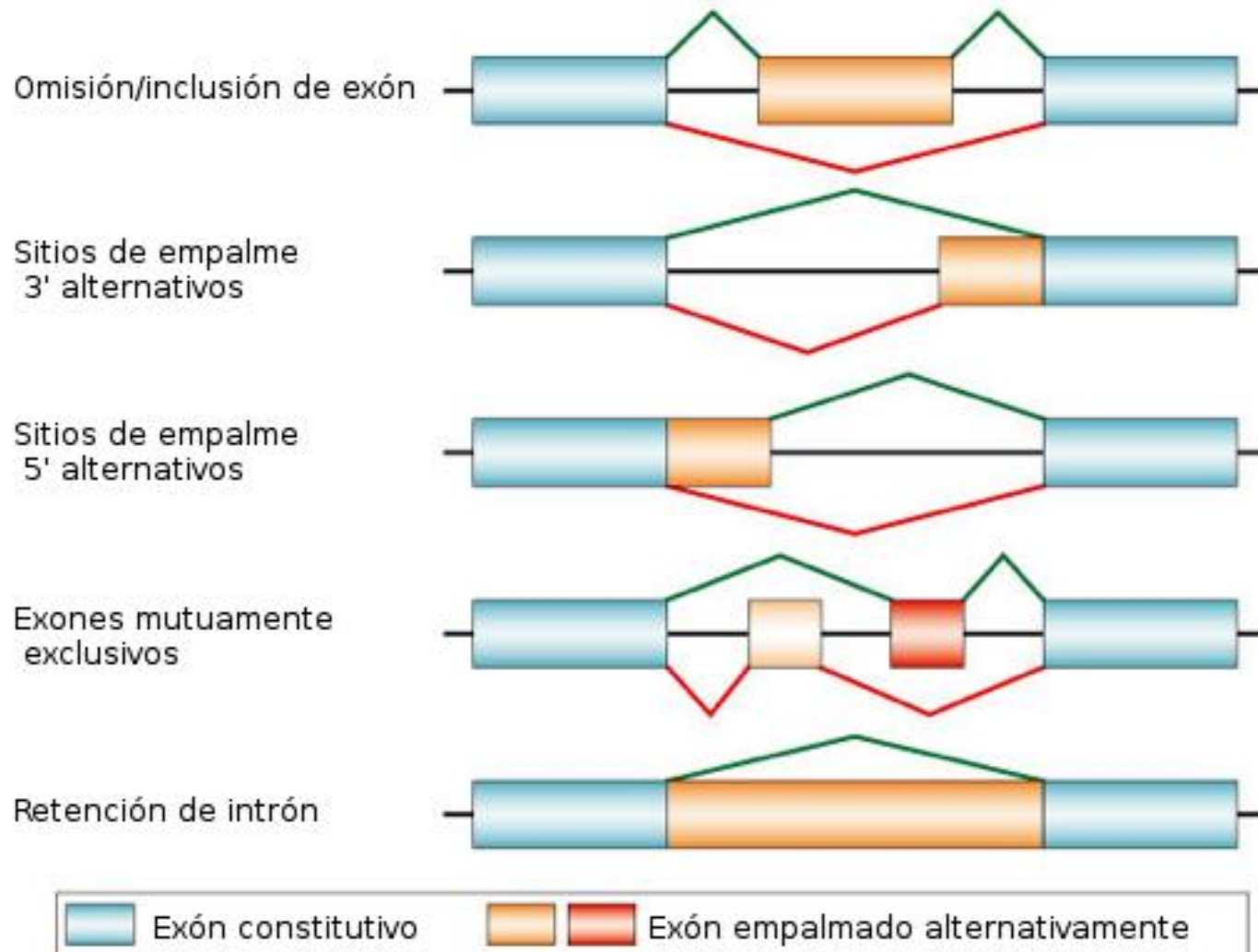
ARN Mensajero



- Secuencia de ARNm
 - ¿Procesamiento alternativo?

```
>uc002pab.3      (APOE)      length=1223
gggatccttgagtcctactcagccccagcggaggtgaaggacgtccttcccaggagccg
actggccaatcacaggcaggaagatgaaggttctgtgggctgcgttgctggtcacattcc
tggcaggatgccaggccaaggtggagcaagcggaggagacagagccggagcccagctgc
gccagcagaccgagtggcagagcggccagcgtgggaactggcactgggtcgcttttggg
attacctgcgctgggtgcagacactgtctgagcaggtgcaggaggagctgctcagctccc
aggtcaccaggaactgagggcgctgatggacgagaccatgaaggagtgaaggctaca
aatcggaactggaggaacaactgaccccggtggcggaggagacgcgggcacggctgtcca
aggagctgcaggcggcgaggcccggtggcgcgacatggaggacgtgtgcgccgccc
tggtgcagtaccgcggcgaggtgcaggccatgctcgccagagcacaggaggagctgcggg
tgcgcctcgctcccacctgcgcaagctgcgtaagcggctcctccgcgatgccgatgacc
tgcagaagcgcctggcagtgtagcaggccggggccgcgagggcgccgagcgcggcctca
gcgcatccgcgagcgctggggcccctggtggaacagggccgctgcgggcccgcactg
tgggtcctctggccggccagccgctacaggagcggggccaggcctggggcgagcggctgc
gcgcgaggatggaggagatgggcagccggacccgcgaccgctggacgaggtgaaggagc
aggtggcggaggtgcgcgccaagctggaggagcaggcccagcagatacgctgcaggccg
aggccttcaggcccgcctcaagagctggttcgagcccctggtggaagacatgcagcgcc
agtgggcccggctggtggagaaggtgcaggctgccgtgggcaccagcgccgcccctgtgc
ccagcgacaatcactgaacgccgaagcctgcagccatgcgacccacgccaccccgctgcc
tcctgcctccgcgagcctgcagcgggagaccctgtccccggccagccgtcctctggg
gtggaccctagtttaataaagattcaccaagtttcacgcaaaaaaaaaaaaaaaaaaaaa
aaaaaaaaaaaaaaaaaaaaaaaa
```

Tipos de procesamiento del ARNm



Secuencia de la proteína

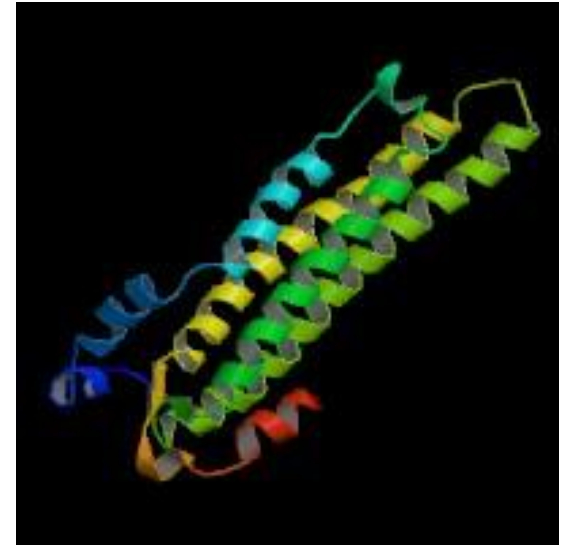


- Alfabeto de 20 letras
- ~300 aminoácidos en una proteína promedio en bacterias
- ~200 mil secuencias de proteínas conocidas.

>>uc002pab.3 (APOE) length=317
 MKVLWAALLVTFLAGCQAKVEQAVETEPEPELRQQTEWQSGQRWEL
 ALGRFWDYLRWVQTLSEQVQEELLSSQVTQELRALMDETMKELKAYK
 SELEEQLTPVAEETRARLSKELQAAQARLGADMEDVCGRLVQYRGEV
 QAMLGQSTEELRVRLASHLRKLRKLLRDADDLQKRLAVYQAGAREG
 AERGLSAIRERLGPLVEQGRVRAATVGSLAGQPLQERAQAWGERLRA
 RMEEMGSRTDRDLDEVKEQVAEVRAKLEEQAQQIRLQAEAFQARLKS
 WFEPLVEDMQRQWAGLVEKVQAAVGTSAAPVPSDNH

		Segunda letra									
		U		C		A		G			
1a le tra	U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U	
		UUC		UCC		UAC		UGC		C	
		UUA	Leu	UCA		UAA	Stop	UGA	Stop	A	
		UUG		UCG		UAG	Stop	UGG	Trp	G	
	C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	U	
		CUC		CCC		CAC		CGC		C	
		CUA		CCA		CAA	Gln	CGA		A	
		CUG		CCG		CAG		CGG		G	
	A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U	
		AUC		ACC		AAC		AGC		C	
		AUA		ACA		AAA	Lys	AGA	Arg	A	
		AUG	Met	ACG		AAG		AGG		G	
	G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U	
		GUC		GCC		GAC		GGC		C	
		GUA		GCA		GAA	Glu	GGA		A	
		GUG		GCG		GAG		GGG		G	
										3a le tra	

Estructura de la proteína

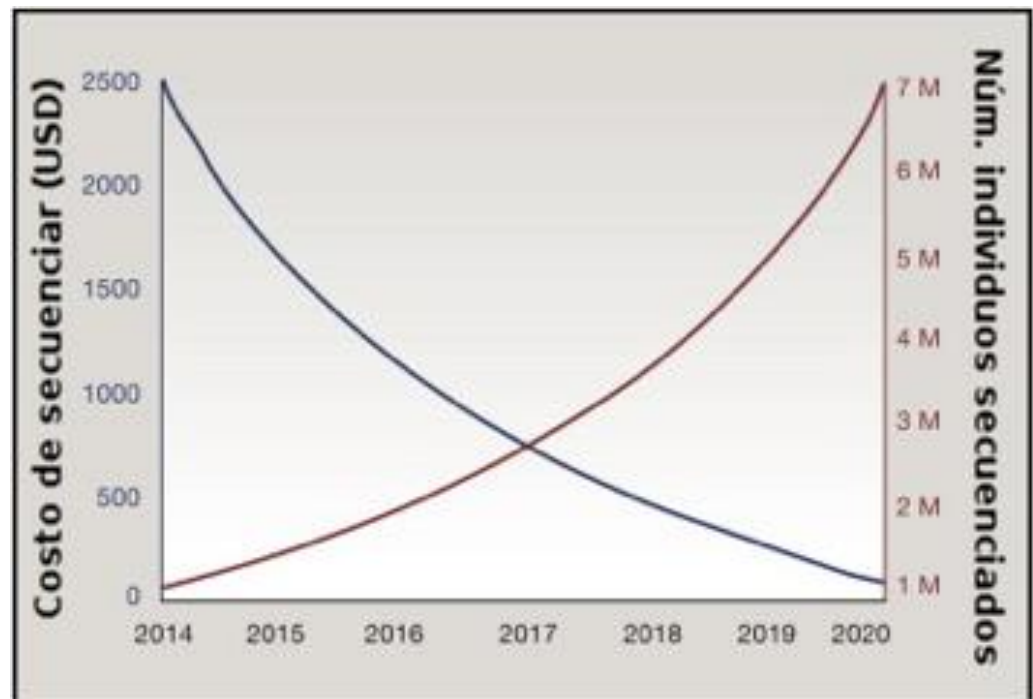


Información de todo el genoma



Genomas

- Bacteria: 1.6 Mb (1995)
- Eucarionte: 13 Mb (1997)
- Animal: 100 Mb (1998)
- Humano: 3 Gb (2001)



Método de secuenciación de disparo a todo el genoma



DNA genoma



Secuenciar cada fragmento
con acercamiento de disparo

GCTTTGAGTTTCCCTGACACCGCTT

CCAGTGGTACTGAGACCCAGAGGCTTA

GCTTATTTGCCATTTATTTATTT

Alinear **secuencias** **contiguas**

GCTTTGAGTTTCCCTGACACCGCTTGGTACTGAGACCCAGAGGCTTATTTGCCATTTATTTATTT

Secuencia generada terminada