

Bioinformatica

Trabajo Práctico Final

Alumnos

Tylson, Emilio - 54022
Gutierrez, Ignacio - 54293

Profesor

Yankilevich, Patricio

Indice

[Ejercicio 1](#)

[Ejercicio 2](#)

[Ejercicio 3](#)

[Ejercicio 4](#)

[Ejercicio 5](#)

[Ejercicio 6](#)

[Pregunta A](#)

[Pregunta B](#)

[Pregunta C](#)

[Pregunta D](#)

[Pregunta E](#)

[Pregunta F](#)

[Pregunta G](#)

Ejercicio 1

El trabajo se decide en enfocar en la Apolipoproteína E, proteína relacionada con el Alzheimer (https://www.ncbi.nlm.nih.gov/nuccore/NM_001302688.1?report=genbank).

El código toma el archivo genbank del ARN mensajero en formato genbank, y realiza la transcripción de la proteína mediante los 6 posibles marcos de lectura (orf).

De todos los marcos de lecturas realizados, el orf 0 es el que genera la proteína con menos caracteres de finalización intermedios (sólo posee estos al final de la secuencia). Por lo que se estima que este marco de lectura sería el correcto. No obstante en el siguiente ejercicio donde se realizará el BLAST frente a una base de datos de proteínas, se podrá estimar de mejor manera cuál es el marco de lectura correcto.

Ejecución:

```
$ perl Ex1.pl [path archivo Genbank]
```

Ejercicio 2

El código realiza BLAST sobre Apolipoprotein E de forma local o remota. Básicamente entra como input el archivo en formato fasta de la proteína, y genera un blast report en el a carpeta ../data/exercise2_out.

Como conclusión se puede observar que de todos los BLAST hechos, el del marco de lectura 0 es el genera más matches. De hecho todas las proteínas con las que se está alineando son distintas variaciones de la Apolipoproteína E en distintas especies. Por ejemplo la segunda proteína con más score de coincidencia es la Q9GLM7 que pertenece a la especie del orangután, o la tercera (Q9GLM8) es la Apolipoproteína E de los gorilas.

Ejecución:

```
$ perl Ex2.pl [path al fasta input] [path al reporte BLAST] [1 para remoto | 0 para local ]
```

Ejercicio 3

El código toma como input el archivo blast generado en el ejercicio 2. Luego, toma el nombre de las primeras 10 secuencias con mayor score en el alineamiento.

Además busca la correspondiente secuencias en la base de datos swissprot y las escribe en un archivo fasta, en el que también está presente la Apolipoproteína E, nuestra proteína de referencia.

Ejecución:

```
$ perl Ex3.pl [path al blast report] [path al archivo fasta de la proteína referencia]
```

De la página <https://www.ebi.ac.uk/Tools/msa/> que ofrece diferentes alternativas para realizar el secuenciamiento, usamos T-Coffe. En dicha aplicación subimos el archivo generado por el código anterior, el cual genera un archivo fasta con las 10 secuencia con más similitud obtenidas por BLAST sobre Apolipoproteína E del homossapiens. La alineación genera un archivo que está subido en ../data/exercise3_out/msa.out.

The screenshot displays the T-Coffe web interface, which is organized into three main steps:

- STEP 1 - Enter your input sequences:** This section includes a text area for pasting sequences and a file upload option. The file 'top10_matches.fasta' is selected, and the 'Choose File' button is highlighted with a red box.
- STEP 2 - Set your Parameters:** This section contains a 'More options...' button and a note about default settings.
- STEP 3 - Submit your job:** This section features a checkbox for email notifications and a 'Submit' button, which is also highlighted with a red box.

Herramienta online de T-Coffe para hacer el alineamiento múltiple. Primero se carga el archivo fasta con el top 10 de matches generado por el código. Luego apretando "submit" realiza el alineamiento.

Se puede observar en el alineamiento, que los primeros aminoácidos no matchean pero luego en la cadena principal de aminoácidos existe una coincidencia de uno a uno en todas las especies, por lo que se podría concluir que sería la estructura principal de la proteína, que le da la funcionalidad principal.

```

A0A0D9S1R0.1 -----
NM_001302688 GTGGALLDKSGILESYSAPAEVKDVLPOEPVRSAVGGTGMSGASRKSWD
P02649.1 -----
P0DKU9.1 -----
P0DKY2.1 -----
P0D094.1 -----
P0D095.1 -----
Q9GJU3.1 -----
Q9GLM6.1 -----
Q9GLM7.1 -----
Q9GLM8.1 -----

A0A0D9S1R0.1 -----MKVLWAALLVTFLAGCQAKVEQVPEPEPEL RQ
NM_001302688 PGNPWPPDWPITGRMKVLWAALLVTFLAGCQAKVEQAVETEPEPELRQ
P02649.1 -----MKVLWAALLVTFLAGCQAKVEQAVETEPEPELRQ
P0DKU9.1 -----MKVLWAALLVTFLAGCQAKVEQVPEPEPELRQ
P0DKY2.1 -----MKVLWAALLVTFLAGCQAKVEQVPEPEPELRQ
P0D094.1 -----MKVLWAALLVTFLAGCQAKVEQVPEPEPELRQ
P0D095.1 -----MKVLWAALLVTFLAGCQAKVEQVPEPEPELRQ
Q9GJU3.1 -----MKVLWAALLVTFLAGCQAKVEQVPEPEPELRQ
Q9GLM6.1 -----MKVLWAALLVTFLAGCQAKVEQVPEPEPELRQ
Q9GLM7.1 -----MKVLWAALLVTFLAGCQAKVEQVPEPEPELRQ
Q9GLM8.1 -----MKVLWAALLVTFLAGCQAKVEQVPEPEPELRQ
***** **.*.***: *

A0A0D9S1R0.1 TEWQSGQPWELALGRFDYLRWVQTLSEQVQEELLSPQVTQELTTLMDET
NM_001302688 TEWQSGQRWELALGRFDYLRWVQTLSEQVQEELLSSQVTQELRALMDET
P02649.1 TEWQSGQRWELALGRFDYLRWVQTLSEQVQEELLSSQVTQELRALMDET
P0DKU9.1 TEWQSGQPWELALGRFDYLRWVQTLSEQVQEELLSSQVTQELTTLMDET
P0DKY2.1 AEWQSGQPWELALGRFDYLRWVQTLSEQVQEELLSPQVTQELTTLMDET
P0D094.1 AEGQSGQPWELALGRFDYLRWVQTLSEQVQEELLSPQVTQELTTLMDET
P0D095.1 TEWQSGQPWELALGRFDYLRWVQTLSEQVQEELLSSQVTQELTTLMDET
Q9GJU3.1 AEWQSGQWELALGRFDYLRWVQTLSEQVQEELLSSQVTQELTALMDET
Q9GLM6.1 AEWQSGQPWELALGRFDYLRWVQTLSEQVQEELLSSQVTQELTALMDET
Q9GLM7.1 AEWQSGQWELALGRFDYLRWVQTLSEQVQEELLSSQVTQELTALMDET
Q9GLM8.1 AEWQSGQWELALGRFDYLRWVQTLSEQVQEELLSSQVTQELTALMDET
:* *** *****:*****.***** :*****

```

Captura del archivo de salida de la alineación múltiple

Ejercicio 4

El código busca el nombre de cada proteína incluida en el BLAST report, cuya descripción contenga el pattern ingresado como input. Luego por cada proteína se busca la secuencia en la base de datos swissprot. Luego cada secuencia es escrita en un archivo fasta con el nombre que es ingresado por parámetro. A su vez por cada match del pattern se escribe en un archivo paralelo el ACCESSION que indica la proteína, su descripción y el índice de la descripción donde hubo match del pattern. Ambos archivos de salida se crean en ./data/exercise4_output .

Ejecución:

```
$ perl Ex4.pl [path al BLAST report] [pattern] [nombre del fasta de salida]
```

Ejercicio 5

El siguiente ejercicio se realizó en python debido a que los wrappers de Bioperl tenían problemas para ser instalados. El primer paso que realiza el código es utilizar la función **getorf** de EMBOSS el cual escribe la salida en `./data/exercise5_out/out.orf` . El programa **getorf** encuentra y extrae los posibles marcos de lectura (orf) de la secuencia de aminoácidos ingresada, en este caso el de la Apolipoproteína E, marcando los codones de START y de STOP.

Luego el código ejecuta **patmatmotifs** de EMBOSS y escribe su salida en `./data/exercise5_out/out.patmatmotifs` . El program lee una secuencia de proteínas y la busca en la base de datos prosite. Escribe un reporte EMBOSS donde detalla la locación y el score de los matches.

Ejecución:

```
$ python Ex5.py
```

Ejercicio 6

Pregunta A

A partir del gen o proteína de interés para ustedes dar su link a NCBI-Gene como una entrada de Entrez. Expliquen brevemente lo que hace la proteína y porque la eligieron.

La apolipoproteína E (APOE) es una proteína con 299 aminoácidos de largo y transporta lipoproteínas, vitaminas liposolubles y colesterol hacia el sistema linfático y luego a la sangre. Es sintetizada principalmente en el hígado, aunque se ha encontrado también en otros tejidos, como el cerebro, riñón y el bazo. El gen que la sintetiza se encuentra en el cromosoma 19 y el mismo tiene 3 alelos principales. La presencia de uno de estos 3 alelos está asociado a una mayor probabilidad de contraer la enfermedad de Alzheimer. Es una proteína de baja densidad, responsable de remover el colesterol del flujo sanguíneo y transportar triglicéridos a los tejidos. Variaciones en APOE está relacionada con el metabolismo del colesterol, que incrementa las chances de ataques cardíacos.

Pregunta B

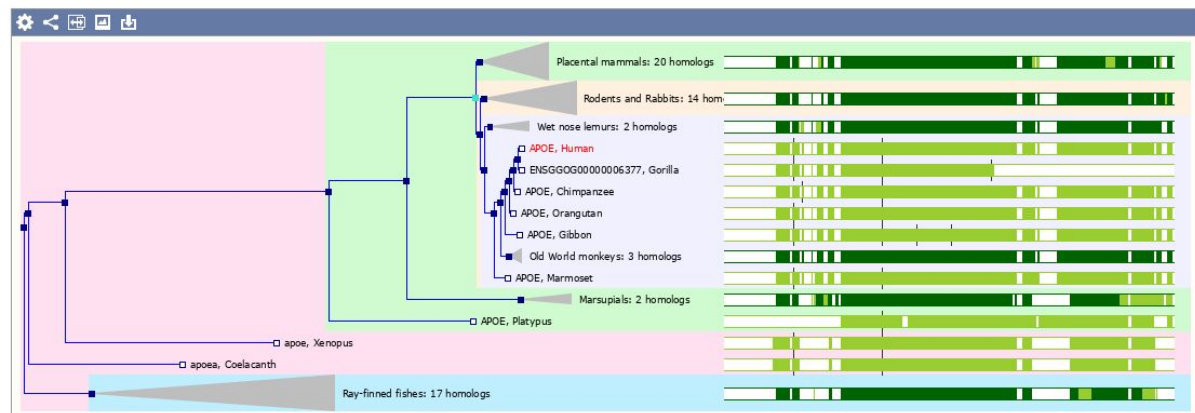
Cuantos genes/proteínas homólogas se conocen en otros organismos? Utilicen la información que está en la base de datos de HomoloGene y en la base de datos Ensembl. Describan los resultados en ambas bases de datos y en qué se diferencian.

Mencionen sobre que tan comun creen son estos genes o proteínas y a que grupos taxonómicos pertenecen (solo en las bacterias, en los vertebrados, etc).

Gene tree

GeneTree [ENSGT00730000111315](#)

Number of genes	67
Number of speciation nodes	64
Number of duplication	1
Number of ambiguous	1
Number of gene split events	0



En el árbol genético del gen APOE (en Ensembl) podemos apreciar en rojo el gen humano y luego otros genes homólogos (67 en total) presentes en otras especies, todas vertebrados.

HomoloGene:30951. Gene conserved in Tetrapoda

Genes

Genes identified as putative homologs of one another during the construction of HomoloGene.

APOE, *H.sapiens*
 apolipoprotein E
 APOE, *P.troglodytes*
 apolipoprotein E
 APOE, *M.mulatta*
 apolipoprotein E
 APOE, *C.lupus*
 apolipoprotein E
 APOE, *B.taurus*
 apolipoprotein E
 Apoe, *M.musculus*
 apolipoprotein E
 Apoe, *R.norvegicus*
 apolipoprotein E
 apoe, *X.tropicalis*
 apolipoprotein E

Proteins

Proteins used in sequence comparisons and their conserved domain architectures.

NP_000032.1 —————
 317 aa
 NP_001009007.1 —————
 317 aa
 XP_001104482.2 —————
 263 aa
 XP_866043.1 —————
 323 aa
 NP_776416.1 —————
 316 aa
 NP_033826.2 —————
 311 aa
 NP_620183.2 —————
 312 aa
 XP_004912063.1 —————
 275 aa

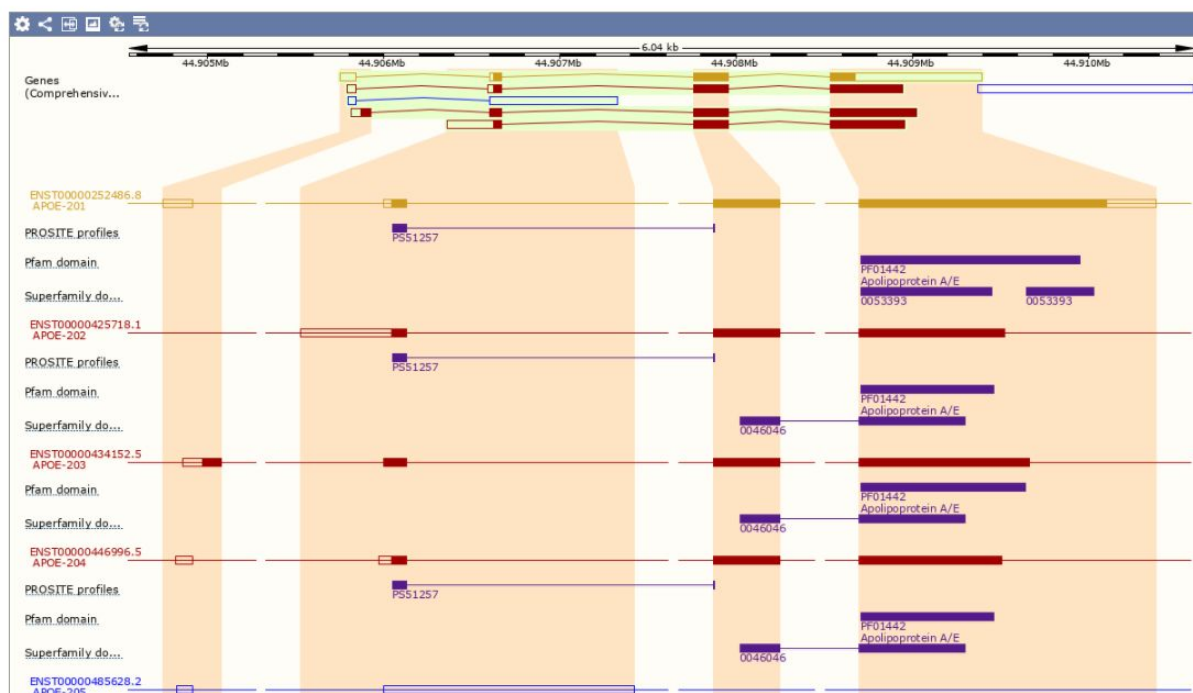
En HomoloGene podemos apreciar las especies en las que se encuentran los homólogos, aunque no la cantidad exacta de genes.

Pregunta C

Cuantos transcritos y cuantas formas alternativas de splicing son conocidos para este gen/proteína? Cuales de estos splicing alternativos se expresan? Tienen funciones alternativas? Buscar evidencia de esto en las bases de datos de NCBI y en los transcritos de Ensembl. Como es el número de splicings alternativos diferente entre las dos bases de datos y cual piensan que es más precisa y porque?

En Ensembl encontramos 5 formas alternativas de splicing, aunque no tenemos mucha información de cada una.

Splice variants



En NCBI encontramos la misma cantidad de formas alternativas de splicing, pero nos da más datos. Nos dice que existen 2 isoformas de la Apolipoproteína E (isoforma A e isoforma B) y que 4 de las variantes de splicing codifican la isoforma B y la otra la isoforma A.

mRNA and Protein(s)	
1. NM_000041.3 → NP_000032.1 apolipoprotein E isoform b precursor	
See identical proteins and their annotated locations for NP_000032.1	
Status: REVIEWED	
Description	Transcript Variant: This variant (2) contains an alternate 5' terminal exon, and it thus differs in the 5' UTR and initiates translation at a downstream in-frame start codon, compared to variant 1. The encoded isoform (b) is shorter at the N-terminus, compared to isoform a. Variants 2, 3, 4 and 5 all encode isoform b.
Source sequence(s)	AL037046 , BC003557 , HY056112
Consensus CDS	CCDS12647.1
UniProtKB/Swiss-Prot	P02649
UniProtKB/TrEMBL	A0A0S2Z3D5
Related	ENSP00000252486.3 , OTTHUMP00000159143 , ENST00000252486.8 , OTTHUMT00000250865
Conserved Domains (1) summary	
	pfam01442 Apolipoprotein; Apolipoprotein A1/A4/E domain Location:86 → 273
2. NM_001302688.1 → NP_001289617.1 apolipoprotein E isoform a precursor	
Status: REVIEWED	
Description	Transcript Variant: This variant (1) represents the longest transcript and encodes the longer isoform (a).
Source sequence(s)	AL037046 , BC003557 , BM042032 , HY056112
UniProtKB/Swiss-Prot	P02649
Conserved Domains (1) summary	
	pfam01442 Apolipoprotein; Apolipoprotein A1/A4/E domain Location:112 → 292
3. NM_001302689.1 → NP_001289618.1 apolipoprotein E isoform b precursor	
See identical proteins and their annotated locations for NP_001289618.1	
Status: REVIEWED	
Description	Transcript Variant: This variant (3) contains an alternate 5' terminal exon, and it thus differs in the 5' UTR and initiates translation at a downstream in-frame start codon, compared to variant 1. The encoded isoform (b) is shorter at the N-terminus, compared to isoform a. Variants 2, 3, 4 and 5 all encode isoform b.

Con cuantas otras proteínas interactúan el producto génico de su gen? Existe un patrón o relación entre las interacciones? Mencione las interacciones interesantes o inusuales. Usted encontrará las interacciones de su gen/proteína tanto en la base de datos NCBI Gene como en la base de datos UniProt. Compare las dos tablas entre si. Hay proteínas que interactúan unicas para cada tabla?

Interaction¹

P02649 has binary interactions with 10 proteins

Diseases ▼



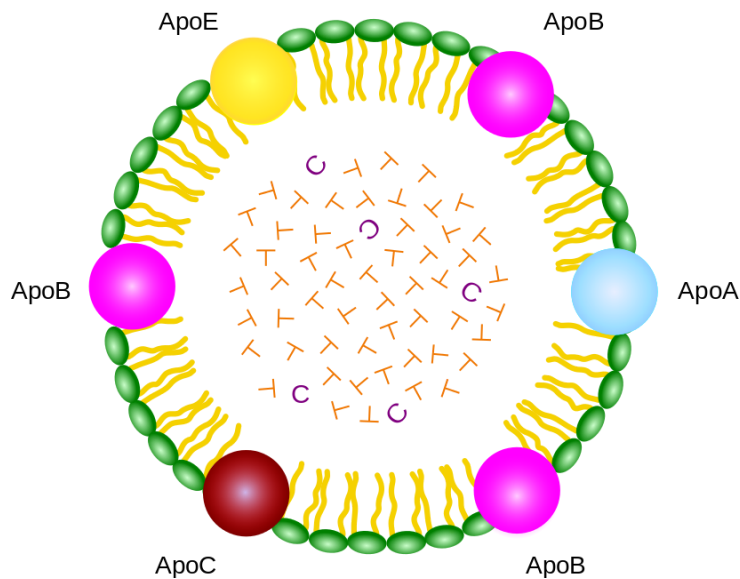
En NCBI vemos que existen 78 interacciones con el gen. En particular menciona una interacción con la proteína NS5A, proteína que juega un rol esencial en el desarrollo de Hepatitis C.

Interactions						
Products	Interactant	Other Gene	Complex	Source	Pubs	Description
NP_000032.1	NP_000605.1	CNTF		BIND	PubMed	Apolipoprotein E binds to ciliary neurotrophic factor
NP_000032.1	NP_671491.1	HCVgp1		BIND	PubMed	ApoE interacts with NS5A. This interaction was modeled on a demonstrated interaction between ApoE from an unspecified species and NS5A from Hepatitis C virus type 1a isolate H77.
NP_000032.1	CAB46677.1			BIND	PubMed	ApoE interacts with NS5A. This interaction was modeled on a demonstrated interaction between ApoE from an unspecified species and NS5A from Hepatitis C virus type 1b isolate Con1.
P02649	P01023	A2M		HPRD	PubMed	
P02649	P02768	ALB		HPRD	PubMed	
P02649	P02649	APOE		HPRD	PubMed	
P02649	P05067	APP		HPRD	PubMed	
P02649	P26441	CNTF		HPRD	PubMed	
P02649	P07858	CTSB		HPRD	PubMed	
P02649	P01130	LDLR		HPRD	PubMed	
P02649	Q07954	LRP1		HPRD	PubMed	
P02649	P98164	LRP2		HPRD	PubMed	
P02649	Q14114	LRP8		HPRD	PubMed	
P02649	P11137	MAP2		HPRD	PubMed	
P02649	P10636	MAPT		HPRD	PubMed	
P02649	P07197	NEFM		HPRD	PubMed	
P02649	P55058	PLTP		HPRD	PubMed	
P02649	P04156	PRNP		HPRD	PubMed	
P02649	Q8WTV0	SCARB1		HPRD	PubMed	
P02649	P08155	VLDLR		HPRD	PubMed	

Pregunta E

Expliquen brevemente de qué componente celular forma parte su proteína (pista: se puede estudiar la información de Gene Ontology - GO). A qué procesos biológicos pertenece (pista idem)? En qué función molecular trabaja esta proteína? Los términos ontológicos de genes los pueden encontrar en NCBI Gene, en la base de datos UniProt y haciendo una búsqueda en AmiGO.

La Apolipoproteína E (APOE) es un componente de la proteína de los Quilomicrones, lipoproteína más compleja localizada en el epitelio del intestino y se encarga de transportar y catabolizar por el sistema linfático los triglicéridos, los fosfolípidos y el colesterol ingeridos. En este caso APOE es el receptor encargado de metabolizar y transportar las lipoproteínas ricas en triglicéridos que son de baja densidad.



APOE es una molécula que compone a los Quilomicrones, una proteína de mayor complejidad

APOE que forma parte de la estructura más compleja de lo Quilomicrones, forma parte de la ruta exógena de transporte de lípidos al hígado, cerebro o riñón:

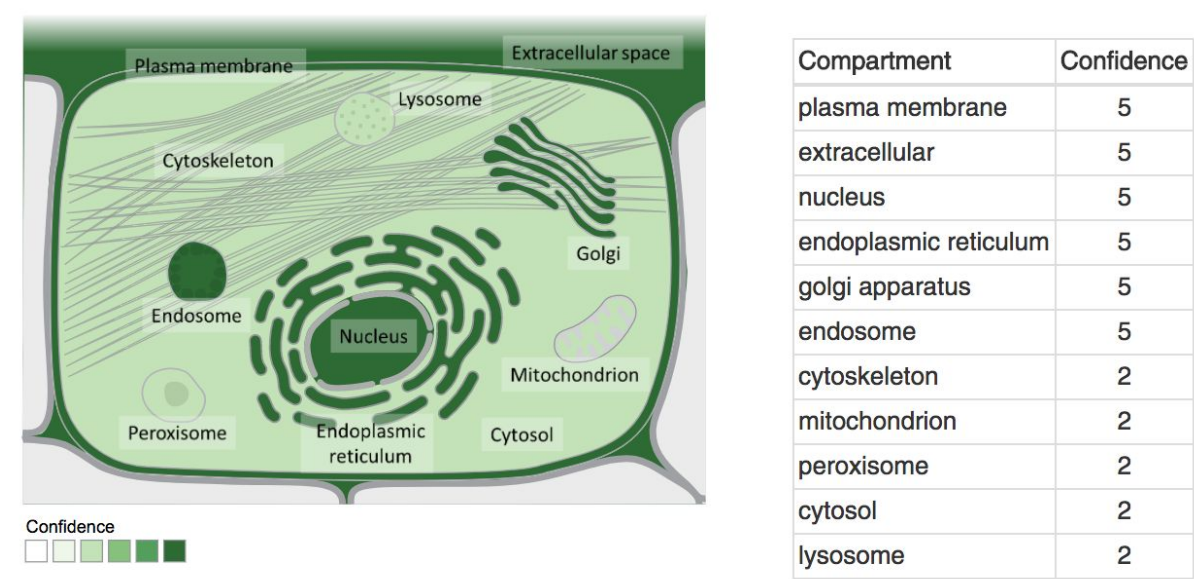
- 1 - Los triglicéridos son captados en el intestino delgado.
- 2- Luego desplazarse por el sistema linfático las proteínas APOE y APO CII son agregadas, formando la estructura del Quilomión.
- 3- Luego, esta molécula se transporta por sangre hasta los tejidos periféricos como los músculos y los tejidos adiposos, donde el Quilomión se degrada, y los ácidos grasos se usan como fuente de energía de las células.

Analizando su funcionamiento a un nivel molecular, se pudo encontrar realizando búsquedas en la base de datos AmiGO2 que la APOE se encarga de diversas funciones:

- Unión con Beta-amiloide:
Interacción selectiva y no covalente con el péptido Beta-amiloide y sus precursores. Este péptido de 36 a 43 aminoácidos, está fuertemente relacionada con la enfermedad del Alzheimer a pesar de que se encarga de otras diversas funciones como activación de quinasas, la protección contra estrés oxidativo, regulación del transporte de Colesterol actuando como un factor de transcripción y actividad antimicrobiana (especialmente asociado con su acción pro-inflamatoria).
- Actividad molecular estructural:
Contribuye a la integridad de la estructura de los Quilomicrones
- Transporte de lípidos:
Permite el movimiento directo de los lípidos entre células.
- Unión proteica:
Interaccion selectiva y no covalente con los triglicéridos.

- Unión con fosfolípidos:
 Interacción selectiva y no-covalente con los fosfolípidos, una clase de lípidos que contienen ácido fosfórico y forman parte de la membrana celular

La proteína APOE puede encontrarse en diversas partes de la célula, pero como su función es mayormente de captación de triglicérido, ésta se encuentra mayormente en sectores extracelulares y en su membrana. Otro lugares en donde se puede localizar, una vez que se está transportando el lípido es en ene el núcleo , el aparato de golgi o en el retículo endoplasmático.

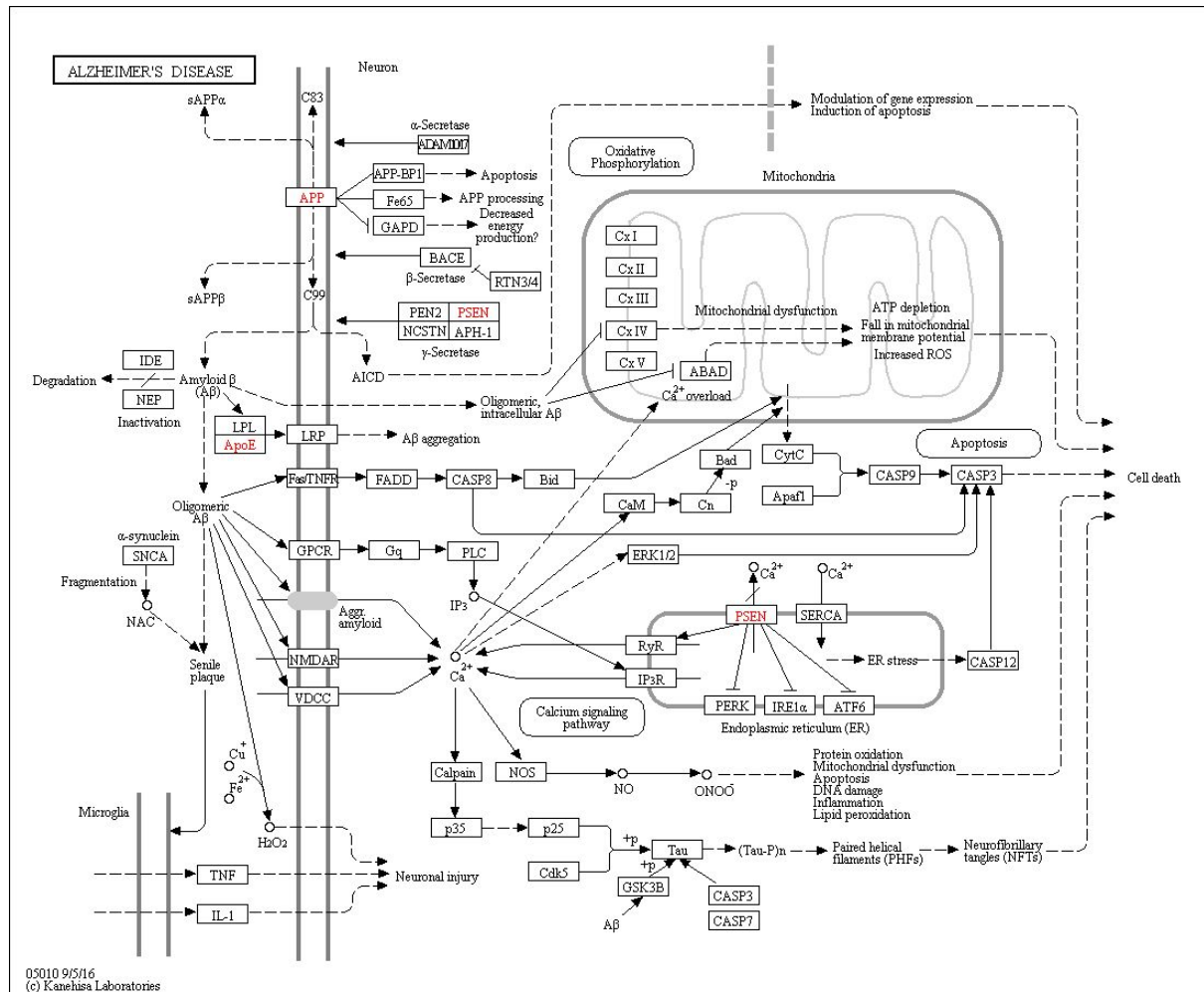


Distintos niveles de confianza en la locación de la proteína APOE

Pregunta F

Discutan brevemente en qué estructura o vías metabólicas específicas (pathways) estaría participando su gen/proteína (Reactome, KEGG son algunas bases de datos de pathways).

En una primera instancia se utilizó la Kyoto Encyclopedia of Genes and Genomes (KEGG) para averiguar el pathway de la Apolipoproteína E. En esta base de datos se muestran datos sobre el funcionamiento a alto nivel de los sistemas biológicos. El pathway que figura se en esta página está relacionado con la enfermedad del Alzheimer. El pathway se presenta con el siguiente mapa donde se resalta en rojo la principales proteínas que interactúan en la enfermedad del Alzheimer.



Esta enfermedad, es un desorden crónico que destruye las neuronas causando serios trastornos cognitivos. Se asocia con las placas seniles y los ovillos neurofibrilares, que son depósitos extracelulares de beta-amiloide en la sustancia gris del cerebro y que se asocian con la degeneración de las estructuras neuronales, así como una abundancia en microglía y de astrocitos. Los beta-amiloide (Abeta), perteneciente a las placas seniles, tiene varios efectos patógenos en células y en las organelas. Los Abeta activan a las enzimas Caspasa que se encargan de la muerte celular programada. Alternativamente, la parte intracelular de Abeta contribuye a la patología ayudando al proceso del hyper-fosforilación (mecanismo de la regulación de la mitosis), terminando el funcionamiento de las mitocondrias y produciendo disfunción de calcio, degradando las estructuras neuronales. Estudios revelaron cuatro genes que están relacionados con esta enfermedad. Estos genes son: proteína amiloidea precursora (APP), presenilina 1 (PS1) y 2 (PS2) y la Apolipoproteína E (APOE). Por lo que mutaciones asociadas a sus genes pueden producir un aumento de la producción de los péptidos Abeta y la interacción con estos

Pregunta G

Entrar en la base de datos de variantes genéticas de SNP e intentar interpretarlo encontrar info sobre alguna variante (reference SNP-rsXXXX) asociada con la patología investigada en su gen de interés. Que variante es? Hay información sobre la frecuencia que tiene esta variante en la población? Que grupo étnico parece ser el mas afectado?

Existen tres variantes comunes, o isoformas de APOE que están definidos por dos SNPs, el rs429358 y el rs7412, nombrados como APOE- ϵ 2, APOE- ϵ 3 y APOE- ϵ 4. Las proteínas generadas por los genes se llaman respectivamente APOE2, APOE3 y APOE4. Siendo APOE3 la variante estándar de la proteína, por lo que la gente hereda APOE- ϵ 3 mas otras variantes de cada padre.

A pesar de que estos son las variantes más comunes, existen otras combinación de nucleótidos por lo que se pueden formar subtipos

rs429358	rs7412	Name
C	T	ϵ 1
T	T	ϵ 2
T	C	ϵ 3
C	C	ϵ 4

Combinación de nucleótidos para formar cada subclase del gen

Se estudió que las personas que poseen la variante APOE- ϵ 4 mas cierto SNPS en otras partes del genoma, tiene mayor riesgo de desarrollar la enfermedad del Alzheimer. Estas otras variantes son:

- Rs2373115 en GAB2
- Heredar el alelo rs1799724 (T) , aparentemente incrementa el riesgo de Alzheimer en APOE- ϵ 4 y está asociado con alteraciones los niveles de concentración Abeta42
- Una variación de POLD1, más la presencia de APOE- ϵ 4, dobla las probabilidades de la enfermedad.

Más en detalle, el alelo rs429358 es un SNP localizado en el cuarto exón de APOE, y afecta al a la posición 130 de la proteína resultante. Como se puede ver en el cuadro anterior, la variación más común del alelo es en nucleótido (T) , y combinado con el SNP rs7412 (C) forman el APOE- ϵ 4. Como se dijo anteriormente, APOE- ϵ 4 está fuertemente relacionado con el riesgo de tener Alzheimer. Sin embargo mucho estudios

estimaron que el nivel de riesgo dependen de la edad, sexo y varios factores. Heredando dos pares (homozygous) de rs429358 (C) o heredando (C;T) incrementan fuertemente las chances de la enfermedad.

☐ **rs429358** [*Homo sapiens*]

1.

GCTGGGCGCGGACATGGAGGACGTG [C/T] GCGGCCGCCTGGTGCAGTACCGCGG

Chromosome: 19:44908684

Gene: APOE (GeneView)

Functional Consequence: missense

Allele Origin: T(germline)/C(germline)

Clinical significance: Pathogenic

Validated: by 1000G, by 2hit Zallele, by cluster, by frequency, by hapmap

Global MAF: C=0.1506/754

HGVs: NC_000019.10:g.44908684T>C, NC_000019.9:g.45411941T>C, NG_007084.2:g.7903T>C, NM_000041.3:c.388T>C, NM_001302688.1:c.466T>C, NM_001302689.1:c.388T>C, NM_001302690.1:c.388T>C, NM_001302691.1:c.388T>C, NP_000032.1:p.Cys130Arg, NP_001289617.1:p.Cys156Arg, NP_001289618.1:p.Cys130Arg, NP_001289619.1:p.Cys130Arg, NP_001289620.1:p.Cys130Arg, XM_005258867.1:c.466T>C, XM_005258868.1:c.388T>C, XP_005258924.1:p.Cys156Arg, XP_005258925.1:p.Cys130Arg

[PubMed](#) [Varview](#)

Reference SNP (refSNP) Cluster Report: rs429358			** With Pathogenic allele **		Links , Linkout
RefSNP	Allele	HGVS Names			
Organism: human (<i>Homo sapiens</i>)	Variation Class: SNV: single nucleotide variation	NC_000019.10:g.44908684T>C	C/T (FWD) C:germline T:germline Ancestral Allele: C Variation Viewer: VarView Clinical Significance: With Pathogenic allele [ClinVar] MAF/MinorAlleleCount: C=0.1843/5332 (ExAC) C=0.1506/754 (1000 Genomes) C=0.1416/1769 (GO-ESP) C=0.1725/5024 (TOPMED)		...more
Molecule Type: Genomic	RefSNP Alleles: C/T (FWD)	NC_000019.9:g.45411941T>C			
Created/Updated in build: 80/150	Allele Origin: C:germline	NG_007084.2:g.7903T>C			
Map to Genome Build: 108/Weight 1	Ancestral Allele: C	NM_000041.3:c.388T>C			
Validation Status:	Variation Viewer: VarView	NM_001302688.1:c.466T>C			
Citation: PubMed	Clinical Significance: With Pathogenic allele [ClinVar]	NM_001302689.1:c.388T>C			
Association: NHGRI GWAS PheGenI	MAF/MinorAlleleCount: C=0.1843/5332 (ExAC)	NM_001302690.1:c.388T>C			
		NM_001302691.1:c.388T>C			
		NP_000032.1:p.Cys130Arg			
		NP_001289617.1:p.Cys156Arg			
		NP_001289618.1:p.Cys130Arg			

SNP Details are organized in the following sections:

[GeneView](#) [Map](#) [Submission](#) [Fasta](#) [Resource](#) [Diversity](#) [Validation](#)

Integrated Maps (Hint: click on 'Chr Pos' to see variant in the new NCBI variation viewer)

Assembly	Annotation Release	Chr	Chr Pos	Contig	Contig Pos	SNP to Chr	Contig allele	Contig to Chr	Neighbor SNP	Map Method
GRCh38.p7	108	19	44908684	NT_011109.17	17667810	Fwd	T	Fwd	view	mapup
GRCh37.p13	105	19	45411941	NT_011109.16	17680159	Fwd	T	Fwd	view	blast

Datos obtenido de la base de datos de NCBI

Como dato adicional, un estudio realizado con más de dos mil personas en Costa Rica, encontró que dietas altas en grasas que producen altos niveles de colesterol incrementa el riesgo de ataques cardiacos en presencia de las variaciones APOE-ε2 y APOE-ε4 comparada con el APOE-ε3. ([abstract del estudio](#)).

Referencias

<https://www.ncbi.nlm.nih.gov/gene/348>
https://www.ncbi.nlm.nih.gov/nuccore/NM_001302688.1
<https://www.omim.org/entry/107741?search=alzheimer&highlight=alzheimer>
<http://amigo.geneontology.org/amigo/search/ontology?q=APOE>
<http://www.uniprot.org/uniprot/P02649>
<http://www.genecards.org/cgi-bin/carddisp.pl?gene=APOE&keywords=apolipoprotein>
<https://reactome.org/content/detail/R-HSA-174646>
http://www.genome.jp/dbget-bin/www_bget?ko:K04524
<https://www.snpedia.com/index.php/APOE>
<https://www.snpedia.com/index.php/APOE-%CE%B54>
<https://www.snpedia.com/index.php/Rs429358>
<https://www.ncbi.nlm.nih.gov/clinvar/?term=APOE%5Bgene%5D>
<https://www.ncbi.nlm.nih.gov/snp/?term=APOE++rs429358>
<https://ghr.nlm.nih.gov/gene/APOE#location>
<https://www.ncbi.nlm.nih.gov/pubmed/18494374?dopt=Abstract>
<https://www.ebi.ac.uk/Tools/msa/>
https://www.ensembl.org/Homo_sapiens/Gene/Splice?db=core:g=ENSG00000130203;r=19:44905754-44909393