

TRABAJO PRÁCTICO ESPECIAL

INTRODUCCIÓN A LA BIOINFORMÁTICA

de la Puerta Echeverría, María (50009)
Romarión, Germán Rodrigo (51296)

Ejercicio 5

- a. NPC1 - http://www.ncbi.nlm.nih.gov/nuccore/NM_000271.4

El transportador de colesterol humano Niemann-Pick C1 (NPC1) parece ser esencial para que el virus del Ébola pueda infectar: una serie de estudios independientes presentaron evidencia de que el virus del Ébola entra a las células humanas luego de unirse con el NPC1.

Elegimos este gen por ser fundamental para el Ébola, una enfermedad que resulta altamente contagiosa, y que últimamente está afectando a varios individuos en Europa y África.

- b.

NPC1L1 (Human) - ENSEMBL:

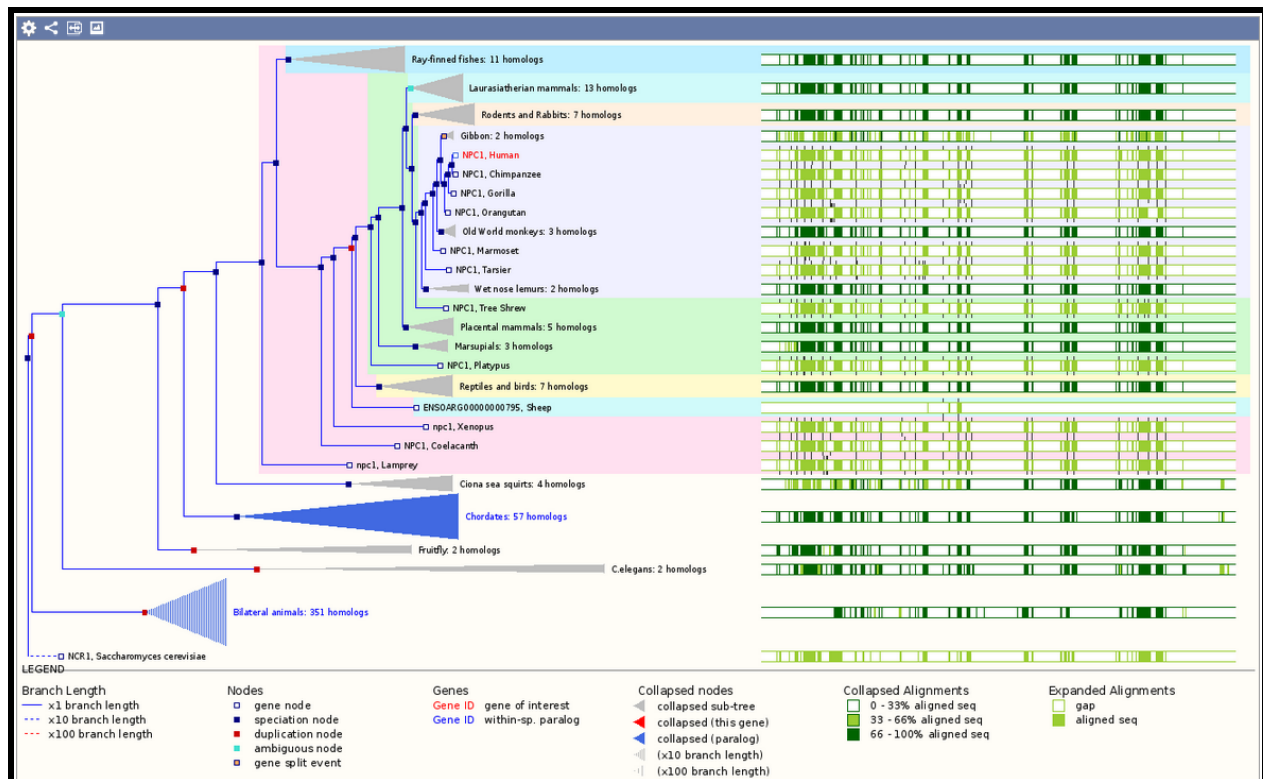


Imagen 1: Árbol de genes homólogos al NPC1 obtenido del sitio ENSEMBL. Se muestra en rojo al gen humano.

NPC1 (Human) - HomoloGene:

Genes	
<i>Genes identified as putative homologs of one another during the construction of HomoloGene.</i>	
 NPC1, <i>H.sapiens</i>	Niemann-Pick disease, type C1
 NPC1, <i>P.troglodytes</i>	Niemann-Pick disease, type C1
 NPC1, <i>M.mulatta</i>	Niemann-Pick disease, type C1
 NPC1, <i>C.lupus</i>	Niemann-Pick disease, type C1
 NPC1, <i>B.taurus</i>	Niemann-Pick disease, type C1
 Npc1, <i>M.musculus</i>	Niemann Pick type C1
 Npc1, <i>R.norvegicus</i>	Niemann-Pick disease, type C1
 NPC1, <i>G.gallus</i>	Niemann-Pick disease, type C1
 npc1, <i>X.tropicalis</i>	Niemann-Pick disease, type C1
 npc1, <i>D.rerio</i>	Niemann-Pick disease, type C1
 Npc1a, <i>D.melanogaster</i>	Niemann-Pick type C-1a
 AgaP_AGAP008137, <i>A.gambiae</i>	AgaP_AGAP008137
 NCR1, <i>S.cerevisiae</i>	NCR1
 KLLA0F27137g, <i>K.lactis</i>	KLLA0F27137g
 AGOS_AGL008W, <i>E.gossypii</i>	AGOS_AGL008W
 MGG_00819, <i>M.oryzae</i>	MGG_00819
 NCU04090, <i>N.crassa</i>	NCU04090
 AT4G38350, <i>A.thaliana</i>	AT4G38350
 AT1G42470, <i>A.thaliana</i>	AT1G42470

Imagen 2: Genes homólogos al NPC1 obtenido del sitio HomoloGene.

Tras haber expandido el árbol que obtuvimos en el sitio de ENSEMBL, pudimos ver que la raíz del mismo indicaba que el gen NPC1 pertenece a los grupos taxonómicos de animales y hongos:

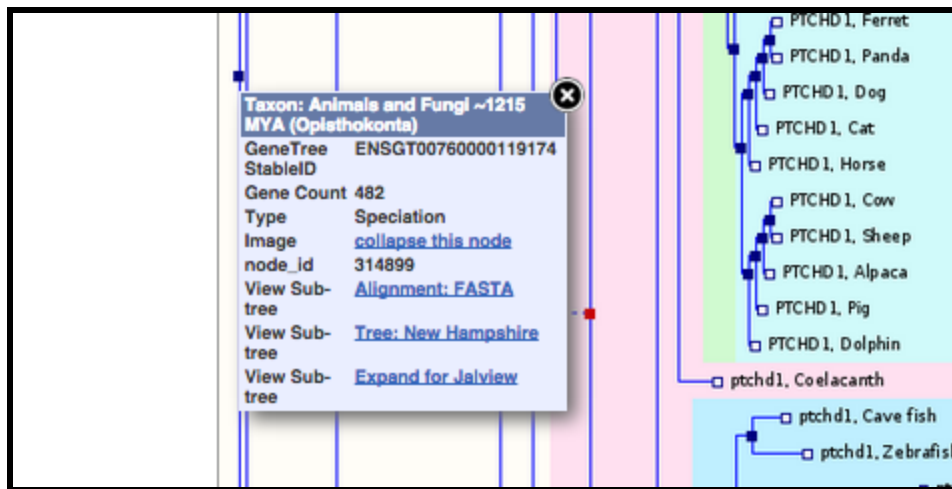


Imagen 3: Raíz del árbol de genes homólogos a NPC1 expandido de ENSEMBL.

La principal diferencia que pudimos ver es en el formato de la respuesta: en ENSEMBL se muestra el árbol con todos los genes homólogos, ofreciendo la posibilidad de expandirlo para mostrar mayor cantidad de información. En HomoloGene en cambio muestra la información de manera más reducida, y no muestra la información en formato de árbol, si no que uno tiene que hacer links en los genes para ver sus homólogos correspondientes, lo cual resulta más incómodo a nuestro modo de ver.

A primera vista ambos muestran un denominador común de especies homólogas para el gen NPC1 en los resultados de ENSEMBL y los de HomoloGene: Homo Sapiens, Lobos, Toros y Gallos.

Aparentemente, por la amplia cantidad de resultados obtenidos en ENSEMBL, opinamos que la presencia del gen NPC1 es altamente común en varios seres vivos.

C.

Existen **14 transcripts** para el gen NPC1 en ENSEMBL:

Transcripts

This gene has 14 transcripts (splice variants) [Hide transcript table](#)

Show **All** entries Show/hide columns (1 hidden) Filter

Name	Transcript ID	bp	Protein	Biotype	CCDS	RefSeq	Flags
NPC1-001	ENST00000269228	5157	1278 aa	Protein coding	CCDS11878	NM_000271 NP_000262	TSL:1 GENCODE basic
NPC1-003	ENST00000591051	3171	971 aa	Protein coding	-	-	CDS 5' incomplete TSL:2
NPC1-013	ENST00000593280	1607	42 aa	Protein coding	-	-	CDS 5' incomplete TSL:1
NPC1-002	ENST00000591107	890	179 aa	Protein coding	-	-	CDS 5' incomplete TSL:3
NPC1-014	ENST00000590723	755	59 aa	Protein coding	-	-	CDS 5' incomplete TSL:4
NPC1-012	ENST00000586150	571	175 aa	Protein coding	-	-	CDS 5' incomplete TSL:3
NPC1-004	ENST00000540608	2790	No protein	Processed transcript	-	-	TSL:2
NPC1-005	ENST00000587223	590	No protein	Processed transcript	-	-	TSL:4
NPC1-010	ENST00000588867	2000	No protein	Retained intron	-	-	TSL:5
NPC1-007	ENST00000591075	809	No protein	Retained intron	-	-	TSL:4
NPC1-009	ENST00000586718	574	No protein	Retained intron	-	-	TSL:3
NPC1-008	ENST00000591955	561	No protein	Retained intron	-	-	TSL:4
NPC1-011	ENST00000587163	431	No protein	Retained intron	-	-	TSL:2
NPC1-006	ENST00000590301	410	No protein	Retained intron	-	-	TSL:2

Imagen 4: Transcripts para NPC1 obtenidos de ENSEMBL.

Hay **14 formas alternativas de *splicing*** en ENSEMBL:

Transcripts

This gene has 14 transcripts (splice variants) [Hide transcript table](#)

Name	Transcript ID	bp	Protein	Biotype	CCDS	RefSeq	Flags
NPC1-001	ENST00000269228	5157	1278 aa	Protein coding	CCDS11878	NM_000271 NP_000262	TSL:1 GENCODE basic {PI}
NPC1-003	ENST00000591051	3171	971 aa	Protein coding	-	-	CDS 5' incomplete TSL:2
NPC1-013	ENST00000593280	1607	42 aa	Protein coding	-	-	CDS 5' incomplete TSL:1
NPC1-002	ENST00000591107	890	179 aa	Protein coding	-	-	CDS 5' incomplete TSL:3
NPC1-014	ENST00000590723	755	59 aa	Protein coding	-	-	CDS 5' incomplete TSL:4
NPC1-012	ENST00000586150	571	175 aa	Protein coding	-	-	CDS 5' incomplete TSL:3
NPC1-004	ENST00000540608	2790	No protein	Processed transcript	-	-	TSL:2
NPC1-005	ENST00000587223	590	No protein	Processed transcript	-	-	TSL:4
NPC1-010	ENST00000588867	2000	No protein	Retained intron	-	-	TSL:5
NPC1-007	ENST00000591075	809	No protein	Retained intron	-	-	TSL:4
NPC1-009	ENST00000586718	574	No protein	Retained intron	-	-	TSL:3
NPC1-008	ENST00000591955	561	No protein	Retained intron	-	-	TSL:4
NPC1-011	ENST00000587163	431	No protein	Retained intron	-	-	TSL:2
NPC1-006	ENST00000590301	410	No protein	Retained intron	-	-	TSL:2

Imagen 5: Formas alternativas de splicing para el gen NPC1 en ENSEMBL.

Pudimos determinar en NCBI que **existen 76 formas alternativas de *splicing*** para el gen NPC1:

Categories [clear](#) **Results: 1 to 20 of 76**

☒ Alternatively spliced
☐ Annotated genes
☐ Protein-coding

Sequence content
☐ CCDS
☐ Ensembl

Filters activated: Alternatively spliced, Current only. [Clear all](#) to show 316 items.

Name/Gene ID	Description	Location
<input type="checkbox"/> Npc1 ID: 18145	Niemann-Pick type C1 [<i>Mus musculus</i> (house mouse)]	Chromosome 18, NC_000084.6 (12189693..12236433, complement)

Imagen 6: Formas alternativas de splicing para el gen NPC1 en NCBI.

Sospechamos que NCBI es el sitio que tiene la respuesta correcta en cuanto a la cantidad de formas alternativas de splicing para el gen NPC1. Nuestra sospecha se basa en la descripción del sitio de ENSEMBL: “The Ensembl project produces genome databases for vertebrates and other eukaryotic species, and makes this information freely available online.” Esto significa que ENSEMBL maneja un grupo acotado de especies en sus bases de datos, lo cual indicaría que NCBI tiene información más acertada en este caso.

d.

En <http://www.uniprot.org/uniprot/O15118>, pudimos ver que el producto génico de nuestro gen interactúa con otras proteínas de la siguiente manera:

Protein-protein interaction databases	
BioGrid ¹	110925. 12 interactions.
IntAct ¹	O15118. 4 interactions.
STRING ¹	9606. ENSP00000269228.

Imagen 7: Interacción del producto génico de NPC1 con otras proteínas en UNIPROT.

En <http://www.ncbi.nlm.nih.gov/gene/4864>, pudimos ver en cambio que el producto génico de nuestro gen interactúa con otras proteínas de la siguiente manera:

Interactions						
Products	Interactant	Other Gene	Complex	Source	Pubs	Description
BioGRID:110925	BioGRID:117654	FBXO6		BioGRID	PubMed	Affinity Capture-MS
BioGRID:110925	BioGRID:120644	LGR4		BioGRID	PubMed	Affinity Capture-MS
BioGRID:110925	BioGRID:110219	LRP6		BioGRID	PubMed	Affinity Capture-Western
BioGRID:110925	BioGRID:118784	MRPS28		BioGRID	PubMed	Co-fractionation
BioGRID:110925	BioGRID:125380	OSBPL5		BioGRID	PubMed	Affinity Capture-Western
BioGRID:110925	BioGRID:130988	TMEM173		BioGRID	PubMed	Affinity Capture-MS
BioGRID:110925	BioGRID:118899	TMOD3		BioGRID	PubMed	Co-fractionation
BioGRID:110925	BioGRID:113164	UBC		BioGRID	PubMed	Affinity Capture-MS; Affinity Capture-Western
BioGRID:110925	BioGRID:123557	UNC93B1		BioGRID	PubMed	Affinity Capture-MS
BioGRID:110925	BioGRID:118059	VPS4A		BioGRID	PubMed	Affinity Capture-Western; Co-localization; Reconstituted Complex

Imagen 8: Interacción del producto génico de NPC1 con otras proteínas en NCBI.

Encontramos que hay algunas proteínas que interactúan de modo único en UNIPROT, que no figuran en NCBI:

LGALS8 RP11-385F5.1, PCTA1, Po66-CBP, PCTA-1, Gal-8 lectin, galactoside-binding, soluble, 8	1 [details]
PTPRK RP3-480J14.1, R-PTP-kappa dJ480J14.2.1 (protein tyrosine phosphatase, receptor type, K (R-PTP-KAPPA, protein tyrosine phosphatase kappa , protein tyrosine phosphatase kappa	1 [details]

Imágenes 9 y 10: Proteínas de UNIPROT que no se visualizan en NCBI.

- e. Aunque no encontramos el gen NPC1 para humanos, sino para ratas, mostramos a continuación los componentes celulares que forman parte de nuestro gen:

<input type="checkbox"/> Acc	Name	Taxon	PANTHER family	Type	Source	Direct annotation	Synonyms
<input type="checkbox"/> Npc1	Niemann-Pick disease, type C1	Rattus norvegicus	patched-related pthr10796	gene	RGD	membrane raft organization perinuclear region of cytoplasm plasma membrane endoplasmic reticulum negative regulation of macroautophagy cholesterol transport nuclear envelope lysosomal membrane lysosome cholesterol binding adult walking behavior negative regulation of cell death extracellular vesicular exosome cholesterol homeostasis cholesterol efflux Golgi apparatus establishment of protein localization to membrane response to drug autophagy extracellular region vesicle endocytosis cellular response to low-density lipoprotein particle stimulus signal transduction membrane integral component of plasma membrane cellular response to steroid hormone stimulus hedgehog receptor activity protein glycosylation membrane raft response to cadmium ion endosome less	

Imagen 10: Listado de componentes celulares del gen NPC1 en ratas.

- f. Pathway para NPC1L1-1 obtenido en <http://www.reactome.org/content/detail/6251834>.



Metabolism(Drosophila melanogaster)

- Metabolism(Drosophila melanogaster)
 - Metabolism of lipids and lipoproteins(Drosophila melanogaster)
 - Lipid digestion, mobilization, and transport(Drosophila melanogaster)
 - Trafficking of dietary sterols(Drosophila melanogaster)
 - NPC1L1-mediated phytosterol uptake(Drosophila melanogaster)
 - Homologues of NPC1L1-1 [plasma membrane](Drosophila melanogaster)
 - NPC1L1-1 [plasma membrane](Drosophila melanogaster)
 - NPC1L1 inactivation by ezetimibe(Drosophila melanogaster)
 - Homologues of NPC1L1-1 [plasma membrane](Drosophila melanogaster)

- NPC1L1-1 [plasma membrane](Drosophila melanogaster)
- NPC1L1:ezetimibe complex (name copied from entity in Homo sapiens) [plasma membrane](Drosophila melanogaster)
 - Homologues of NPC1L1-1 [plasma membrane](Drosophila melanogaster)
 - NPC1L1-1 [plasma membrane](Drosophila melanogaster)
- NPC1L1-mediated cholesterol uptake(Drosophila melanogaster)

ANEXO

Ex1.rb

```
#!/usr/bin/env ruby

require 'bio'

if ARGV.length != 1
  puts 'FATAL ERROR: Invalid amount of arguments!'
  puts 'Example: ruby Ex1.rb input_file.gb'
  exit
end

genebank_content = Bio::GenBank.open(ARGV[0])

string_sequence = ''

genebank_content.each_entry do |gc|
  string_sequence << gc.to_biossequence
end

6.times do |frame|
  File.open("fasta_frame_#{frame + 1}.fa", 'w') do |f|
    f.write(Bio::Sequence::NA.new(string_sequence).translate(
      frame + 1,
      1,
      '-',
    ).to_fasta
  )
end

end
```


Ex2.rb

```
#!/usr/bin/env ruby

require 'bio'

if ARGV.length != 1
  puts 'FATAL ERROR: Invalid amount of arguments!'
  puts 'Example: ruby Ex2.rb input_file.fa'
  exit
end

blast = Bio::Blast.remote('blastp', 'swissprot', '-e 0.0001', 'genomenet')

fasta_content = Bio::FlatFile.open(Bio::FastaFormat, ARGV[0])

File.open('out.blast', 'w') do |f|
  fasta_content.each_entry do |fc|
    report = blast.query(fc.seq)
    report.hits.each_with_index do |hit, hit_index|
      f.puts '_____'
      f.puts "Hit number: #{hit_index}"
      f.puts hit.accession
      f.puts hit.definition
      f.puts " * Query length: #{hit.len}"
      f.puts " * Identities number: #{hit.identity}"
      f.puts " * Overlapping: #{hit.overlap}"
      f.puts " * % Overlapping: #{hit.percent_identity}"
      f.puts " * Query sequence: #{hit.query_seq}"
      f.puts " * Target sequence: #{hit.target_seq}"
      hit.hsps.each_with_index do |hsps, hsps_index|
        f.puts " * Bit score: #{hsps.bit_score}"
        f.puts " * Gaps: #{hsps.gaps}"
      end
    end
  end
end
```

Ex3.rb

```
#!/usr/bin/env ruby

require 'bio'

if ARGV.length != 2
  puts 'FATAL ERROR: Invalid amount of arguments!'
  puts 'Example: ruby Ex3.rb input_file.blast INPUT_PATTERN'
  exit
end

pattern = ARGV[1].upcase

Bio::NCBI.default_email = '@'

File.open('Ex3.out', 'w') do |f|
  f.puts "Pattern: #{pattern}"
  Bio::Blast.reports_xml(File.new(ARGV[0])) do |report|
    report.each do |hit|
```

```
if hit.definition.upcase.index(pattern)
  f.puts ' _____'
  f.puts " * Definition: #{hit.definition}"
  f.puts " * Accession: #{hit.accession}"
  f.write " * Fasta sequence: "
  f.puts Bio::NCBI::REST::EFetch.protein(hit.accession, 'fasta')
end
end
end
end
```

Ex4

```
EMBOSS-6.6.0/emboss/getorf -sequence GENE BANK_INPUT_FILE
EMBOSS-6.6.0/emboss/patmatmotifs FASTA_INPUT_FILE
```