
Trabajo Práctico Bioinformática

CFTR (*C*ystic *F*ibrosis *T*ransmembrane Conductance *R*egulator)

2do Cuatrimestre 2015

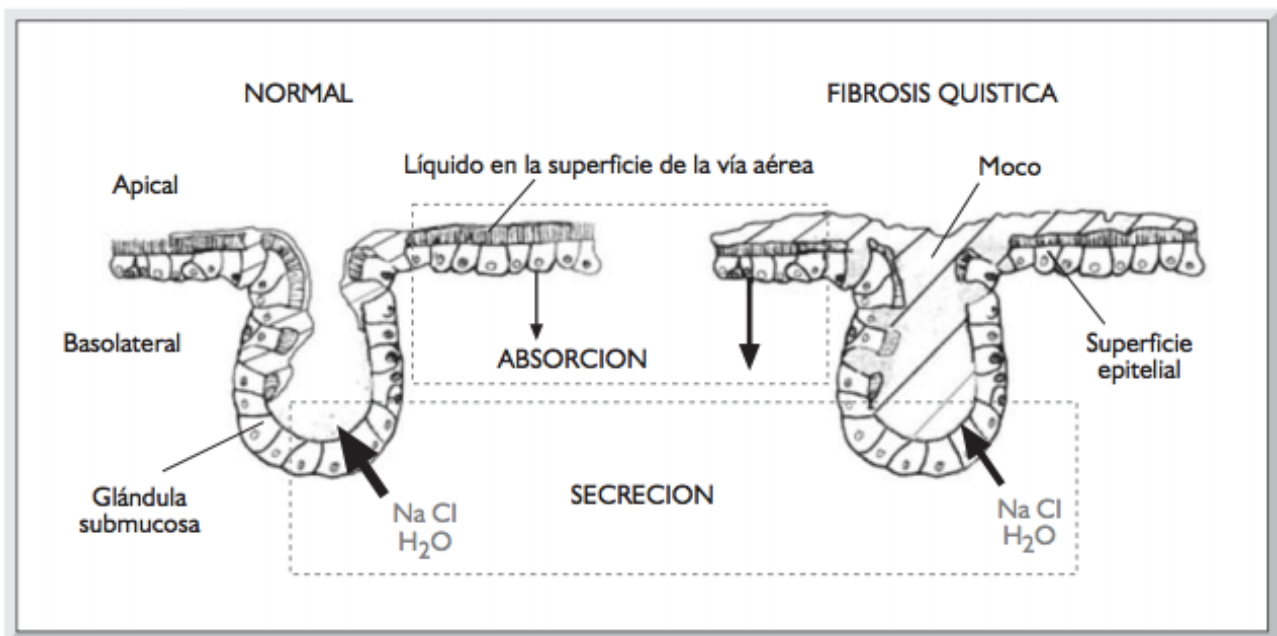
Integrantes

- Cavo, Maria Victoria - 53202 -
- Rossi, Melisa Anabella - 54265 -
- Sakuda, María Eugenia - 53191 -
- Vera, Juan Sebastián - 53852 -

Ejercicio 5

a) El gen elegido es el CFTR (Cystic Fibrosis Transmembrane Conductance Regulator). Este es el encargado de codificar la proteína cuya principal función esta relacionada con el transporte de iones de cloro hacia el exterior de la membrana celular.

Se decidió estudiar este gen debido a su influencia en la enfermedad conocida como fibrosis quística. Esta es la enfermedad autosómica recesiva letal más frecuente. Se produce debido a una mutación en el gen, provocando que haya un aumento en la concentración de sodio y cloro en las secreciones corporales formando una capa de liquido viscoso (moco).



Link : <http://www.ncbi.nlm.nih.gov/gene/1080>

b) Al ingresar a la base de datos **Ensembl** podemos encontrar el “Gene Tree” que se muestra en la Figura 1.

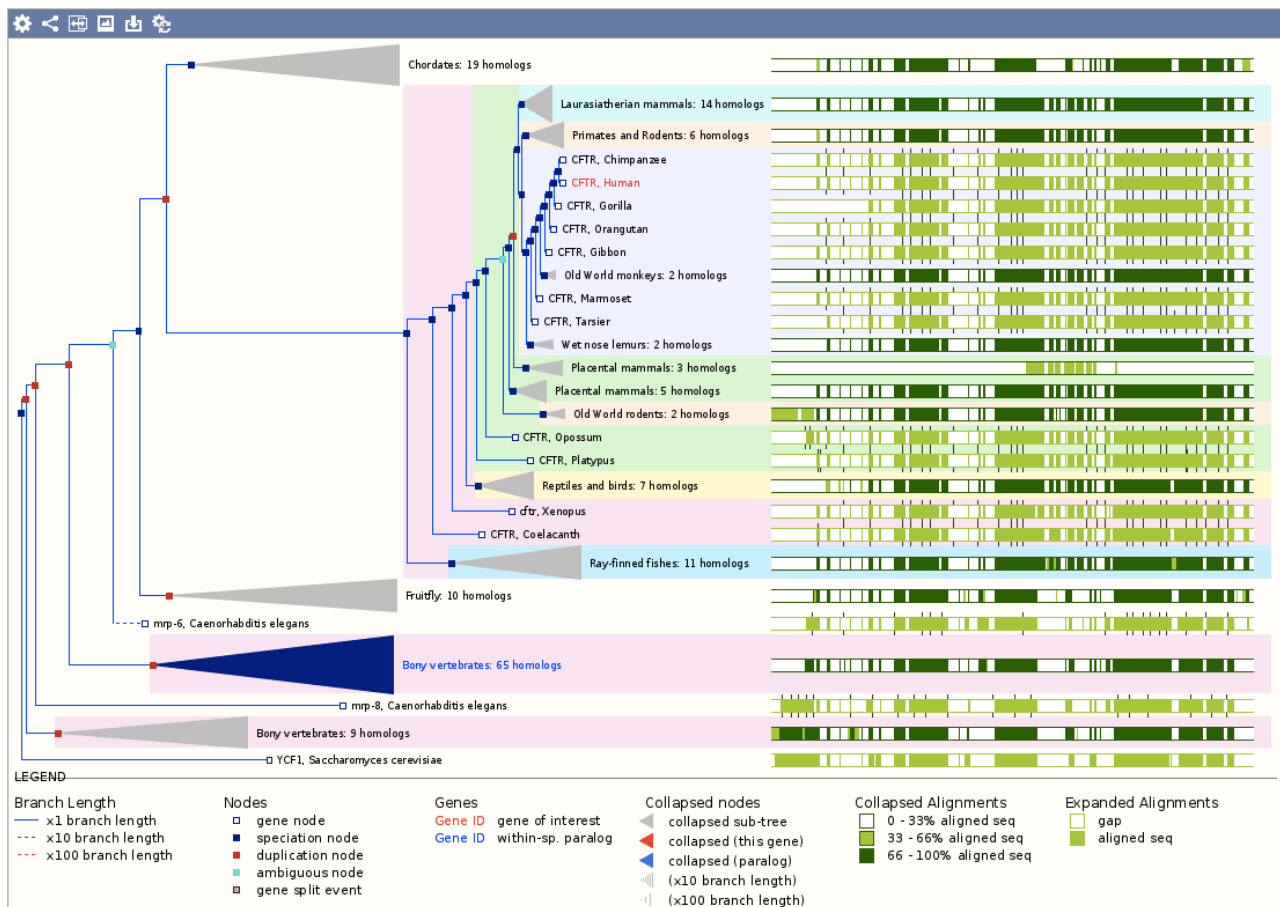


Figura 1 . “Gene Tree” a partir de la base de datos de Ensembl.











En el árbol encontramos los genes homólogos al CFTR. Se encuentra expandido solo el sub-arbol que contiene el gen humano (en rojo), pero podemos observar la gran cantidad de homólogos que existen.

Si analizamos la raíz del árbol encontramos que se encuentra en la categoría taxonómica de los animales y hongos.

Ahora bien, al analizar al base de datos de **HomoloGene** pudimos encontrar los genes homólogos que se muestran en la Figura 2.

Genes

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

-  [ABCC1, *H.sapiens*](#)
ATP-binding cassette, sub-family C (CFTR/MRP), member 1
-  [ABCC1, *P.troglodytes*](#)
ATP-binding cassette, sub-family C (CFTR/MRP), member 1
-  [ABCC1, *C.lupus*](#)
ATP-binding cassette, sub-family C (CFTR/MRP), member 1
-  [ABCC1, *B.taurus*](#)
ATP-binding cassette, sub-family C (CFTR/MRP), member 1
-  [Abcc1, *M.musculus*](#)
ATP-binding cassette, sub-family C (CFTR/MRP), member 1
-  [Abcc1, *R.norvegicus*](#)
ATP-binding cassette, subfamily C (CFTR/MRP), member 1
-  [ABCC1, *G.gallus*](#)
ATP-binding cassette, sub-family C (CFTR/MRP), member 1
-  [abcc1, *D.rerio*](#)
ATP-binding cassette, sub-family C, member 1
-  [mrp-1, *C.elegans*](#)
mrp-1
-  [mrp-2, *C.elegans*](#)
mrp-2
-  [MRP5, *A.thaliana*](#)
MRP5
-  [Os03g0142800, *O.sativa*](#)
Os03g0142800

Proteins

Proteins used in sequence comparisons and their conserved domain architectures.

























-  [NP_004987.2](#) 
1531 aa
-  [XP_001145351.1](#) 
1247 aa
-  [NP_001002971.1](#) 
1531 aa
-  [NP_776648.1](#) 
1530 aa
-  [NP_032602.1](#) 
1528 aa
-  [NP_071617.2](#) 
1532 aa
-  [NP_001012540.1](#) 
1525 aa
-  [XP_001341895.3](#) 
1518 aa
-  [NP_001033553.1](#) 
1534 aa
-  [NP_508121.1](#) 
1525 aa
-  [NP_171908.1](#) 
1514 aa
-  [NP_001048934.1](#) 
1505 aa

Figura 2 . Genes Homologos de HomoloGene

La información dada por HomoloGene es mucho mas incompleta y difícil de comprender que la que se pudo extraer de Ensembl. Allí vemos que encontró solo algunos de los homólogos encontrados en Ensembl, principalmente de la linea de los mamíferos.

A partir de HomoloGene pudimos encontrar también otra fuente que nos brindo información un poco mas detallada que se muestra a continuación.

HomoloGene Vertebrate Homology						
Source		HomoloGene (Release 68, 06-may-2014)				
Class ID		55465				
Comparative GO Graph (mouse, human, rat)		HomoloGene:55465 Multiple Sequence Alignment				
		MGI HomoloGene Information				
Species	Symbol	Gene Links	Genetic Location	Genome Coordinates (mouse and human only)	Associated Human Diseases	Sequences select all deselect all get FASTA Go
human	CFTR	HGNC:1894 (HGNC) 1080 (Entrez Gene) 602421 (OMIM) HGNC homology	Chr7 q31.2	Chr7:117478340-117668665 (+) GRCh38.p2	Bronchiectasis with or without Elevated Sweat Chloride 1; BESC1 Cystic Fibrosis; CF Pancreatitis, Hereditary; PCTT Vas Deferens, Congenital Bilateral Aplasia Of; CBAVD	<input type="checkbox"/> P13569 (UniProt EBI) <input type="checkbox"/> NM_000492 (RefSeq)
mouse	Cfr	MGI:88388 (MGI) 12638 (Entrez Gene) Gene Tree VISTA-Point HGNC homology	Chr6 8.1 cM	Chr6:18170687-18322768 (+) GRCm38	Cystic Fibrosis; CF	<input type="checkbox"/> OTTMUSG00000024173 (VEGA) <input type="checkbox"/> OTTMUSP00000028741 (VEGA) <input type="checkbox"/> OTTMUST00000059141 (VEGA)
rat	Cfr	24255 (Entrez Gene)	Chr4 q21			<input type="checkbox"/> P34158 (UniProt EBI) <input type="checkbox"/> NM_031506 (RefSeq)
chimpanzee	CFTR	463674 (Entrez Gene)	Chr7			<input type="checkbox"/> Q2QLE5 (UniProt EBI) <input type="checkbox"/> NM_001079917 (RefSeq)
rhesus macaque	CFTR	574346 (Entrez Gene)	Chr3			<input type="checkbox"/> Q00553 (UniProt EBI) <input type="checkbox"/> NM_001032938 (RefSeq)
cattle	CFTR	281067 (Entrez Gene)	Chr4 q23-q25			<input type="checkbox"/> P35071 (UniProt EBI) <input type="checkbox"/> NM_174018 (RefSeq)
dog	CFTR	492302 (Entrez Gene)	Chr14			<input type="checkbox"/> Q5U820 (UniProt EBI) <input type="checkbox"/> NM_001007143 (RefSeq)
chicken	CFTR	100049619 (Entrez Gene)	Chr1			<input type="checkbox"/> NP_001099136 (RefSeq) <input type="checkbox"/> NM_001105666 (RefSeq)
western clawed frog	cfr	100486085 (Entrez Gene)	ChrUN			<input type="checkbox"/> XP_012815406 (RefSeq)
zebrafish	cfr	559080 (Entrez Gene)	Chr18			<input type="checkbox"/> NP_001038348 (RefSeq) <input type="checkbox"/> NM_001044883 (RefSeq)

c) En Ensembl se pudieron encontrar 11 formas alternativas de splicing.

Name	Transcript ID	bp	Protein	Biotype	CCDS	UniProt	RefSeq	Flags
CFTR-001	ENST00000003084	6132	1480aa	Protein coding	CCDS5773	A0A024R730 P13569	NM_000492 NP_000483	TSL:1 GENCODE basic APPRIS P1
CFTR-005	ENST00000426809	4316	1438aa	Protein coding	-	E7EPB6	-	CDS 3' incomplete TSL:5
CFTR-002	ENST00000468795	682	190aa	Protein coding	-	H0Y8A9	-	CDS 5' incomplete TSL:5
CFTR-004	ENST00000446805	575	37aa	Protein coding	-	C9J6L5	-	CDS 3' incomplete TSL:4
CFTR-008	ENST00000600166	559	156aa	Protein coding	-	M0QYZ3	-	CDS 5' incomplete TSL:5
CFTR-009	ENST00000608965	896	No protein	Processed transcript	-	-	-	TSL:5
CFTR-010	ENST00000610149	519	No protein	Processed transcript	-	-	-	TSL:5
CFTR-007	ENST00000429014	423	No protein	Processed transcript	-	-	-	TSL:5
CFTR-003	ENST00000546407	222	No protein	Processed transcript	-	-	-	TSL:1
CFTR-006	ENST00000472848	148	No protein	Processed transcript	-	-	-	TSL:5
CFTR-011	ENST00000621535	657	No protein	Retained intron	-	-	-	TSL:5

En cambio en NCBI pudimos encontrar 293 formas alternativas.

NCBI
Resources
How To

Gene

Create alert
Advanced

Gene sources
Genomic
Categories
Alternatively spliced
Annotated genes
Protein-coding
Pseudogene
Sequence content
CCDS
Ensembl
RefSeq
RefSeqGene
Status
Current
Chromosome locations
more...
Clear all
Show additional filters

Tabular
20 per page
Sort by Relevance
Send to:

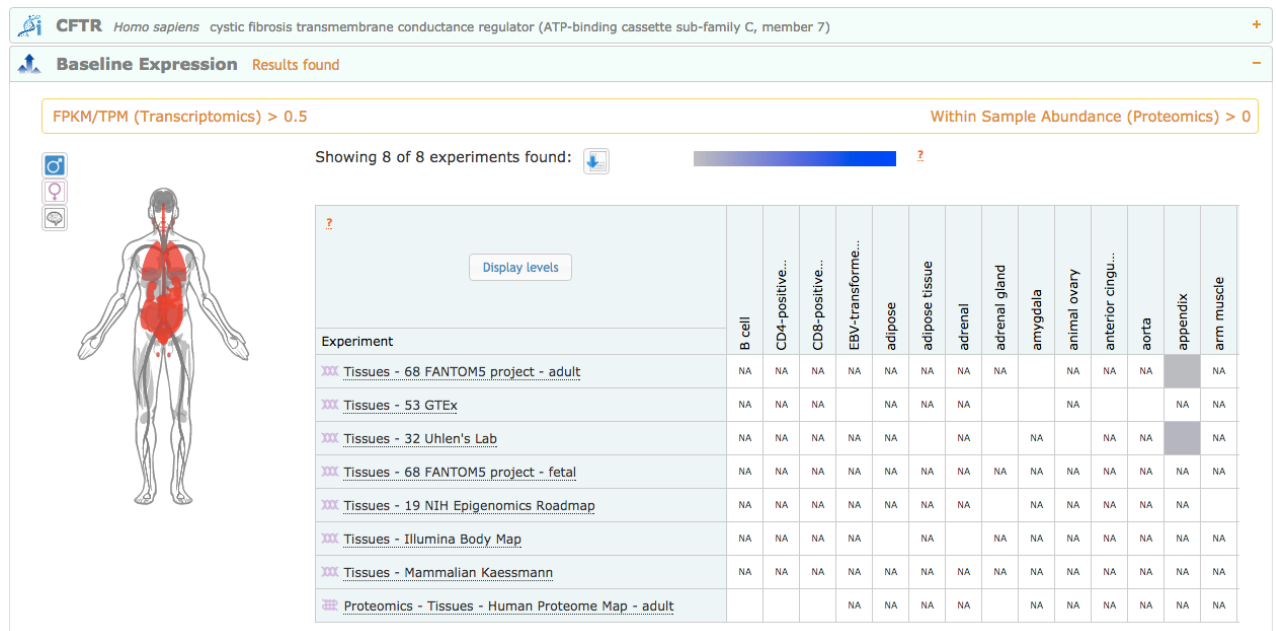
Search results
Items: 1 to 20 of 293
Filters activated: Alternatively spliced. Clear all to show 382 items.
Showing items.

Name/Gene ID	Description	Location	Aliases	MIM
<input type="checkbox"/> CFTR ID: 1080	cystic fibrosis transmembrane conductance regulator [<i>Homo sapiens</i> (human)]	Chromosome 7, NC_000007.14 (117478340..117668665)	ABC35, ABCG7, CF/MRP, MRP7, TNR-CFTR, dJ760C5.1, CFTR	602421
<input type="checkbox"/> ABCC8 ID: 6833	ATP binding cassette subfamily C member 8 [<i>Homo sapiens</i> (human)]	Chromosome 11, NC_000011.10 (17392885..17476845, complement)	ABC36, HHF1, HI, HRINS, MRP8, PHH1, SUR, SUR1, SUR1delta2, TNDM2	600509
<input type="checkbox"/> ABCC2 ID: 1244	ATP binding cassette subfamily C member 2 [<i>Homo sapiens</i> (human)]	Chromosome 10, NC_000010.11 (99782598..99852576)	ABC30, CMOAT, DJS, MRP2, cMRP	601107
<input type="checkbox"/> ABCC4 ID: 10257	ATP binding cassette subfamily C member 4 [<i>Homo sapiens</i> (human)]	Chromosome 13, NC_000013.11 (95010820..95301446)	MOAT-B, MOATB, MRP4	605250

Al buscar razones por esta gran diferencia, pudimos descubrir que el sitio principal de Ensembl, en el que se busco los splicing, realiza las búsquedas solo en base de datos de vertebrados. Es decir, la búsqueda se reduce muchísimo en comparación con NCBI. Ensembl cuenta con páginas que denomina "hermanas" que localizan en otras especies como

puede ser invertebrados, plantas, bacterias, etc. Por eso, se considera a NCBI como la más precisa de ambas.

Al analizar el CFTR - 2 en <http://www.ebi.ac.uk/gxa/genes/ENSG00000001626>, así como también otros de los indicados por Ensembl, pudimos ver que este se expresan en el apéndice humano.



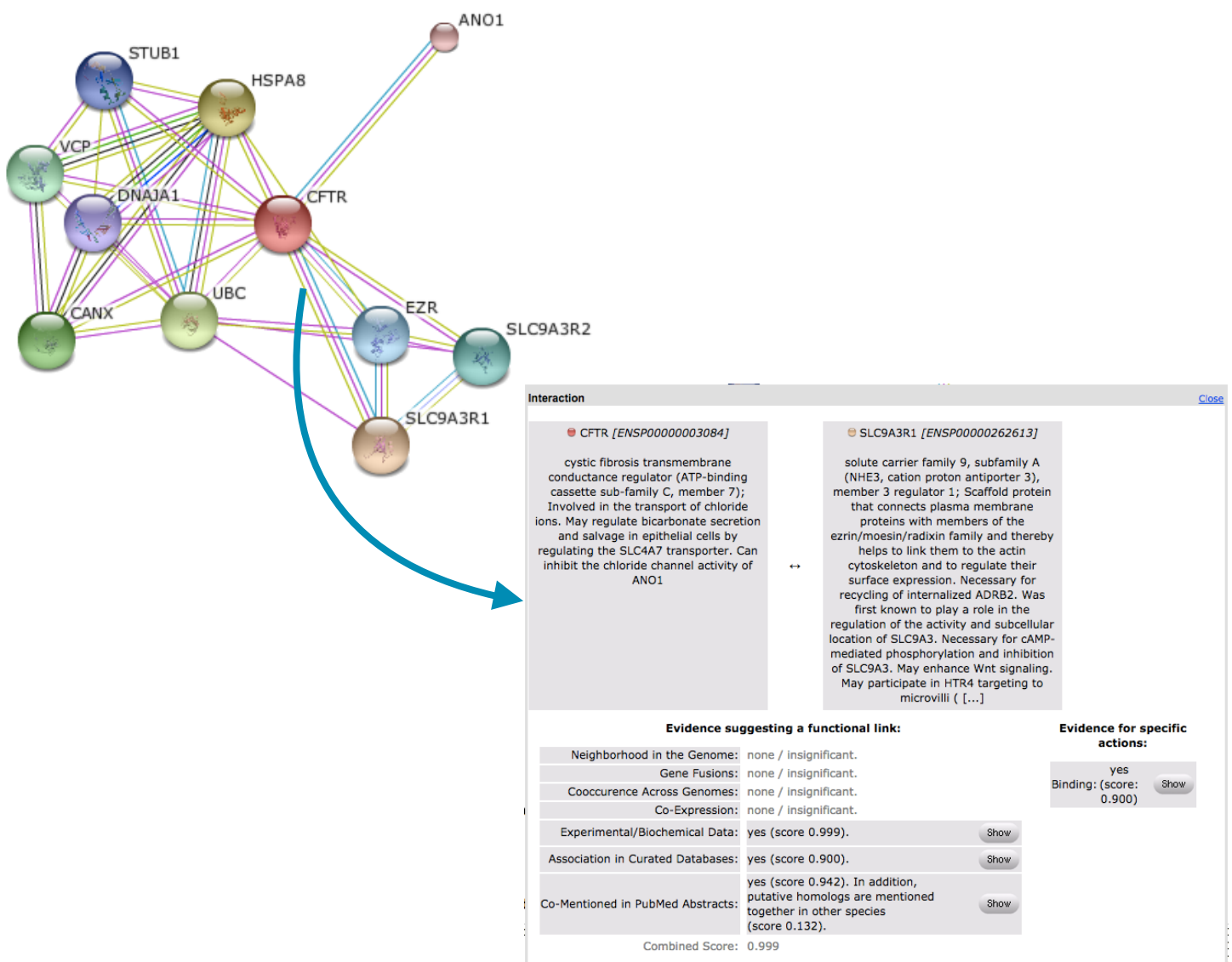
La celdas subrayadas, significa que se pueden ver su expresión mediante experimentos. Aquellas celdas marcadas con NA indican que esas zonas no fueron analizadas.

d) En UniPort pudimos ver que la cantidad de interacciones proteína-proteína varía según la base de datos tomada.

Protein-protein interaction databases

BioGrid ⁱ	107506. 270 interactions.
DIP ⁱ	DIP-32788N.
IntAct ⁱ	P13569. 138 interactions.
MINT ⁱ	MINT-148539.
STRING ⁱ	9606.ENSPP0000003084.

Al entrar a analizar estas bases vimos que rondan las 270 interacciones. El siguiente gráfico muestra los datos obtenidos en STRING, donde cada línea representa una interacción proteína-proteína con CFTR.



En ncbi encontramos, en cambio, que CFTR interactua con 304 proteínas.

Interactions						
Products	Interactant	Other Gene	Complex	Source	Pubs	Description
NP_000483.2	NP_001737.1	CANX		BIND	PubMed	CFTR interacts with CANX.
NP_000483.2	NP_001737.1	CANX		BIND	PubMed	CFTR interacts with CANX. This interaction was modeled on a demonstrated interaction between human CFTR and hamster CANX.
NP_000483.2	NP_000483.2	CFTR		BIND	PubMed	CFTR forms a homodimer
NP_000483.2	CAC15494.1	DNAJC5		BIND	PubMed	Csp2 interacts with CFTR.
NP_000483.2	CAC15495.1	DNAJC5		BIND	PubMed	Csp1 interacts with CFTR.
NP_000483.2	NP_004148.1	PRKAR2A		BIND	PubMed	PKA II interacts with and phosphorylates CFTR.
NP_000483.2	NP_004243.1	SLC9A3R1		BIND	PubMed	First PDZ domain of EBP50 binds to the C-terminus of CFTR.
NP_000483.2	AAC63061.1	SLC9A3R2		BIND	PubMed	The second PDZ domain of E3KARP interacts with the C-terminal PDZ binding motif of CFTR.
NP_000483.2	NP_004594.1	STX1A		BIND	PubMed	Syntaxin 1A interacts with CFTR.
NP_000483.2				BIND	PubMed	CFTR interacts with ATP.
P13569	P27824	CANX		HPRD	PubMed	
P13569	P13569	CFTR		HPRD	PubMed	
P13569	Q9BUN8	DERL1		HPRD	PubMed	
P13569	P31689	DNAJA1		HPRD	PubMed	
P13569	Q9H3Z4	DNAJC5		HPRD	PubMed	
P13569	P15311	EZR		HPRD	PubMed	
P13569	Q9HD26	GOPC		HPRD	PubMed	
P13569	Q86UT5	PDZD3		HPRD	PubMed	
P13569	Q5T2W1	PDZK1		HPRD	PubMed	
P13569	P30154	PPP2R1B		HPRD	PubMed	

e) Para encontrar información a cerca de las funciones moleculares en las cuales interviene la proteína se entro a <http://www.genecards.org/cgi-bin/carddisp.pl?gene=CFTR>. Allí se pudo encontrar que las principales funciones que realiza son las siguientes:

- ATP-binding and phosphorylation-dependent chloride channel activity : Permite la transferencia de ion cloruro a traves de la membrana.
- Chloride-specific channel activity : Permite la transferencia selectiva y difusión de cloruro mediante un proceso independiente de energía sin evidencias de un mecanismo de transporte intermedio.

GO:0005737	cytoplasm	IDA	18570918
GO:0005765	lysosomal membrane	TAS	
GO:0005769	early endosome	IDA	19398555
GO:0005886	plasma membrane	TAS	
GO:0005902	microvillus	IEA	
GO:0009986	cell surface	IDA	20658517
GO:0016021	integral component of membrane	undefined	
GO:0016323	basolateral plasma membrane	NAS	11707463
GO:0016324	apical plasma membrane	IDA	12801959
GO:0030659	cytoplasmic vesicle membrane	undefined	
GO:0030660	Golgi-associated vesicle membrane	TAS	
GO:0031205	colocalizes_with endoplasmic reticulum Sec complex	IDA	9792704
GO:0031901	early endosome membrane	IEA	
GO:0034707	chloride channel complex	IEA	
GO:0043234	protein complex	IDA	17462998
GO:0055037	recycling endosome	IDA	18570918
GO:0070062	extracellular exosome	IDA	23376485

- Channel-conductance-controlling: Funciona como un catalizador que transfiere una solución desde un lado de una membrana al otro. La falta de esto causa la fibrosis quística.
- Proteing binding : interactua selectivamente con proteínas.
- ATP binding : interactua selectivamente con ATP.
- bicarbonate transmembrane transporter activity : Permite el traslado de bicarbonato de un lado de la membrana hacia el otro.
- chloride channel inhibitor activity: Frena, previene y reduce la actividad de un canal de cloruro.

En cuanto a componentes celulares

Por último, en cuanto a los procesos biológicos en los que influye se encontraron los siguientes:

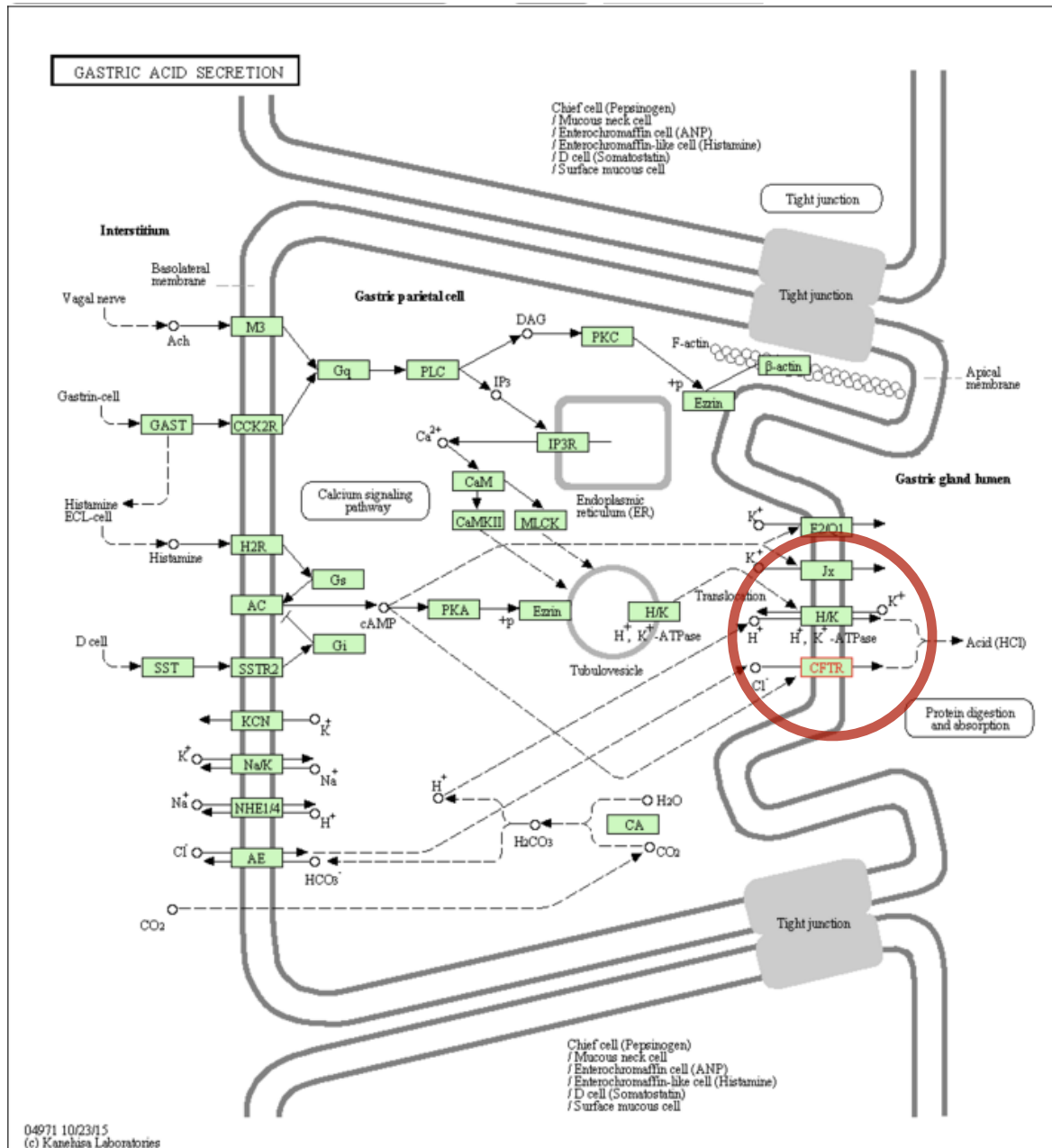
- Transporte de bicarbonato.
- Transporte de cloruro a través de la membrana.
- Transporte de colesterol.
- Elevación intracelular de ph.
- Respuesta celular a cAMP
- Sintetización de colesterol.
- Hiperpolarización de la membrana.
- Regulación positiva de segregación de insulina en la respuesta celular a estímulos de glucosa.



Compartment	Confidence
endosome	5
plasma membrane	5
endoplasmic reticulum	3
cytoskeleton	2
cytosol	2
extracellular	2
nucleus	2
vacuole	2
golgi apparatus	1
lysosome	1
mitochondrion	1

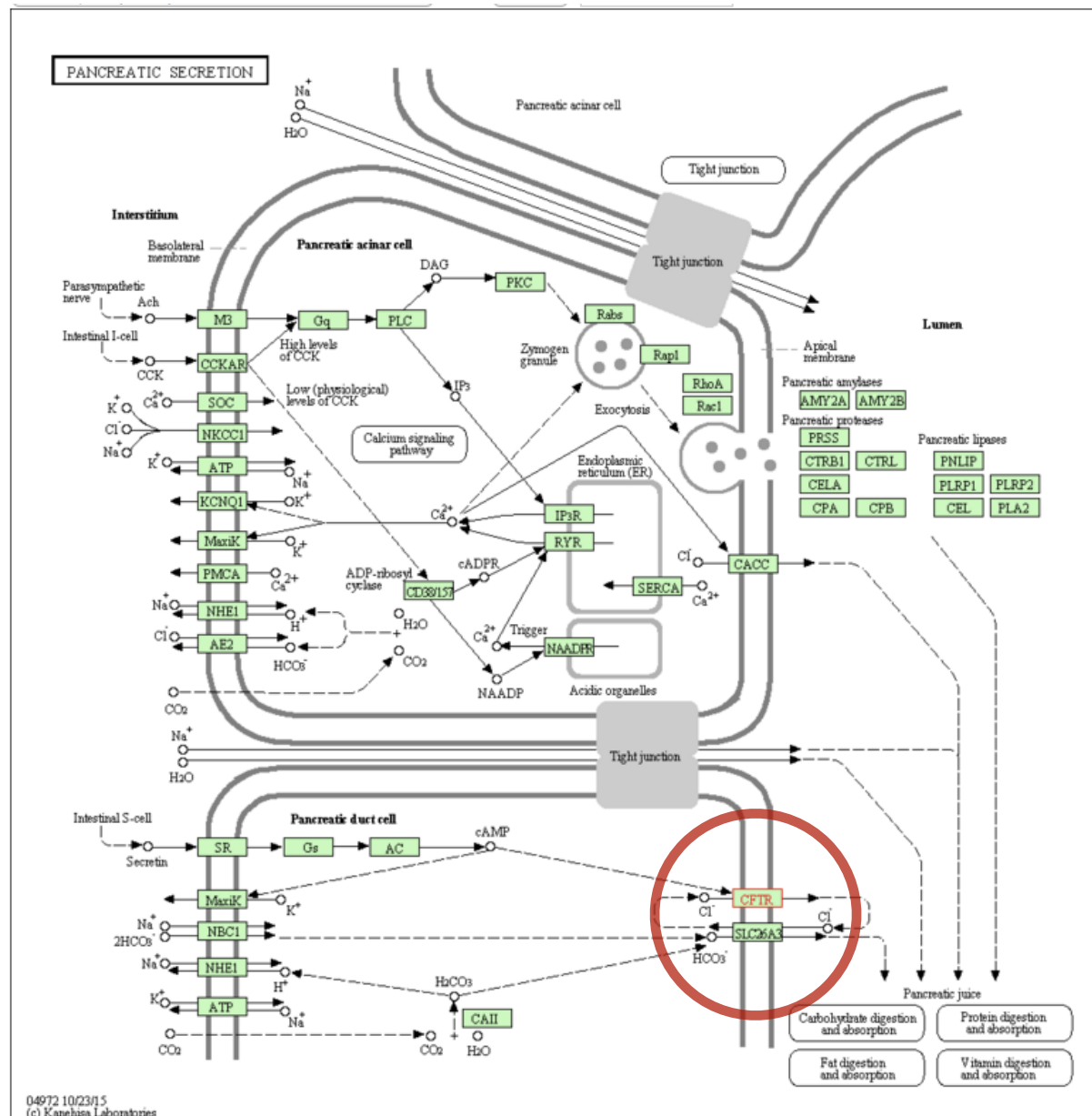
- intercambio gaseoso respiratorio.
- Regulación positiva de la exocitosis.
- Capacitación de espermatozoides (proceso para alcanzar fertilización)
- Regulación positiva de los canales de iones.

f) En NCBI podemos encontrar una gran cantidad de links a distintos pathways. A continuación se detallan algunos de ellos obtenidos de KEGG con una breve descripción de cada uno:

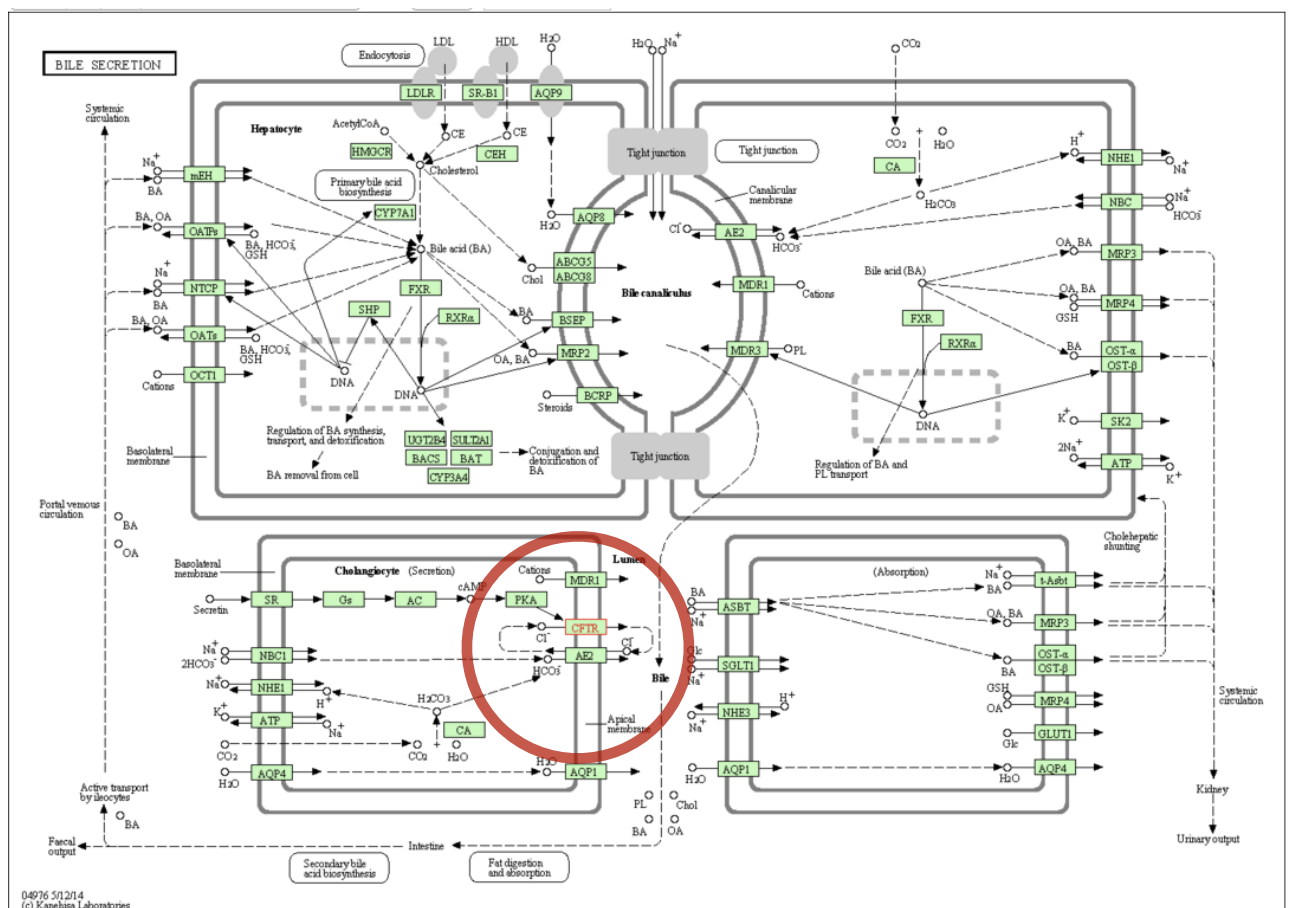


El ácido gástrico es un factor clave dentro de las funciones normales del sistema gastrointestinal superior, que incluyen la digestión proteica y la absorción de calcio y hierro, al igual que la protección contra infecciones bacterianas. Los principales estimulantes de la secreción de dicho ácido al

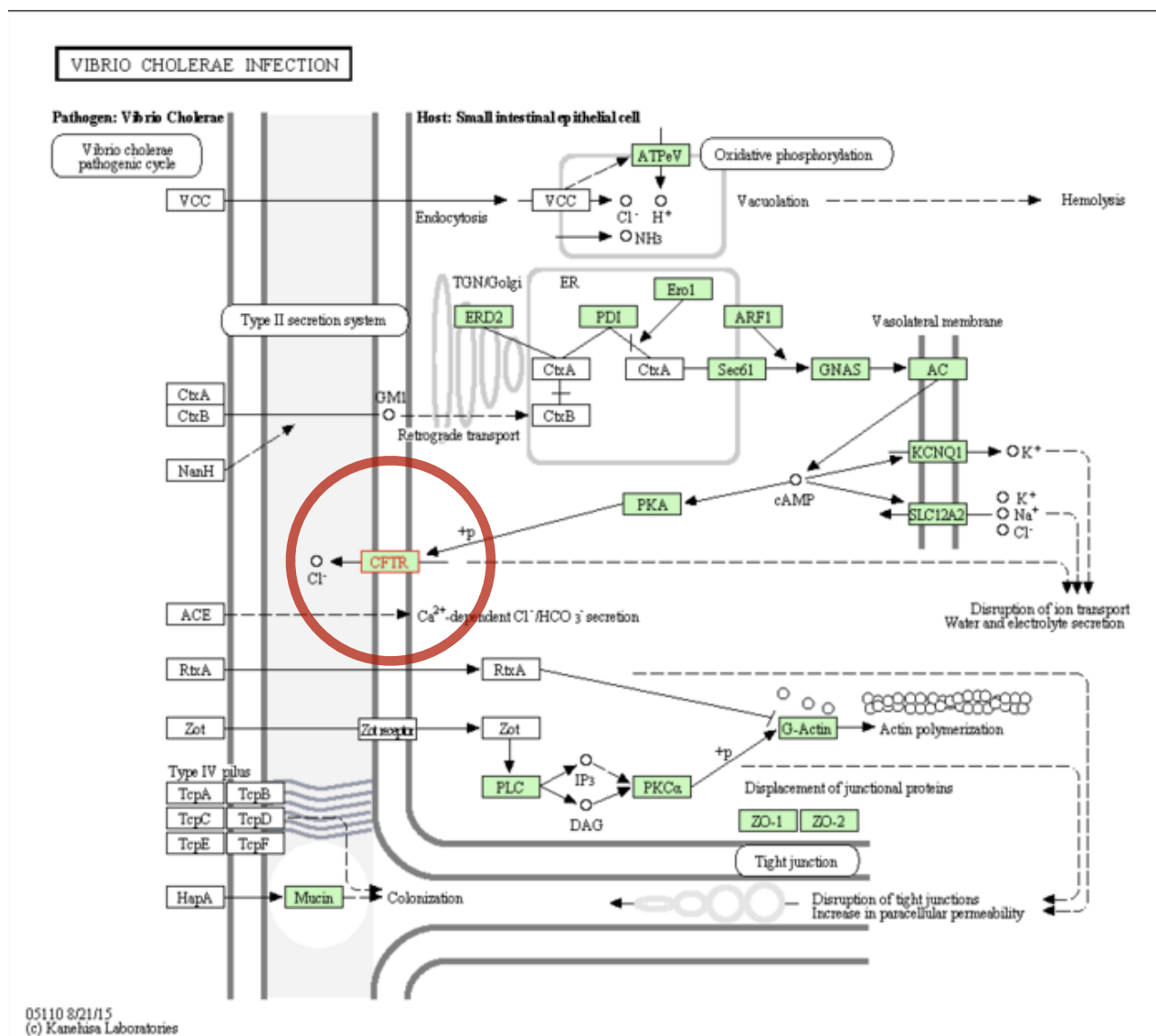
nivel de la célula parietal son la histamínicos (paracrinós), gastrin (hormonal), y acetylcholine (ACh; neurocrinos). La estimulación de la secreción de ácido por lo general implica una elevación inicial de AMP cíclico y calcio intracelular, seguido de la activación de cascadas de quinasa proteica, que activan la traslado de la bomba de protones, H^+ , K^+ -ATPase, desde las túbulo-vesículas citoplasmáticas hasta la membrana plasmática apical y, por lo tanto, de la secreción de H^+ al lumen del estómago.



El páncreas cumple funciones tanto endocrinas como exocrinas. El páncreas exocrino consiste de dos partes, las células del acino y de conducto. Las principales funciones de las células pancreáticas del acino sirven para sintetizar y secretar enzimas digestivas. El estímulo de la célula por parte de secretagogos tales como la acetilcolina (ACh) y la colecistoquinina (CCK) genera una señal Ca^{2+} intracelular. Dicha señal, a su vez genera la fusión de los gránulos zimógenos con la membrana plasmática apical, que luego provoca la secreción polarizada de las enzimas. La principal función de las células pancreáticas de conducto es la secreción de fluido e iones de bicarbonato (HCO_3^-), que neutralizan la acidez de los contenidos gástricos que ingresan al duodeno. El aumento del AMP cíclico intracelular por secreción es una de las mayores señales de secreción pancreática de HCO_3^- . La activación del canal CFTR Cl^- y de las actividades de intercambio de $\text{Cl}^-/\text{HCO}_3^-$ dependientes del CFTR- es responsable de la secreción de HCO_3^- inducida por el AMP cíclico.



La bilis es una secreción vital, esencial para la digestión y absorción de grasas y vitaminas liposolubles en el intestino delgado. Asimismo, la bilis es una importante ruta de eliminación del exceso de colesterol y muchos productos residuales, bilirrubina, drogas y compuestos tóxicos. La secreción de bilis depende del correcto funcionamiento de sistemas de membranas de transporte en hepatocitos y colangiocitos, y de la integridad funcional y estructural del árbol biliar. Los hepatocitos generan la llamada bilis primaria en sus canalículos. Los colangiocitos modifican la bilis canalicular mediante procesos de secreción y reabsorción a medida que la bilis pasa a través de los conductos biliares. Los principales solutos en bilis son los ácidos biliares, que estimulan la secreción biliar por ósmosis, y a la vez facilitan la absorción intestinal de lípidos alimentarios gracias a sus propiedades detergentes. Los receptores bilis-nucleares regulan su propia síntesis y transporte.



La toxina del cólera (CTX) es uno de los principales factores virulentos de la *Vibrio cholerae*. Una vez que se secreta, la CTX B-chain (CTXB) se une al gangliósido GM1 en la superficie de la célula del huésped. Luego de dicha unión, la compleja CTX se traslada de la membrana plasmática (PM) al retículo endoplasmático (ER). En el ER, el disulfuro isomerasa de proteínas reconoce y despliega la A-chain (CTXA), y la traslada a la membrana en que la oxidasa-ER asociada con la membrana, Ero1, oxida PDI para liberar la CTXA dentro del canal conductor de proteínas, Sec61. La CTXA es luego retro translocada al citosol e induce la secreción de agua y electrolitos al aumentar los niveles de AMP cíclico mediante la adenilil ciclasa (AC) para ejercer toxicidad. Además de la CTX, *Vibrio cholera* genera diversas toxinas que son riesgosas para las células eucariotas.

g) Para este inciso se buscaron variantes genéticas que producen la fibrosis quística entre las que encontramos $\Delta F508$ (rs113993960).

Según la información recolectada las distintas mutaciones de CFTR se pueden dividir en 6 clases. La mutación a estudiar pertenece a la clase II. Esta clase comprende a las mutaciones que presentan un procesamiento defectuoso de la proteína, lo que impide su correcta localización en la membrana celular.

Al buscar esta variante en http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=113993960, no encontramos información a cerca de la frecuencia y grupos étnicos. Sin embargo, pudimos encontrar en otras fuentes que es una de las mutaciones más frecuentes de CFTR y esta presente en el 70% de los alelos de Fibrosis Quística. En cuanto a la distribución poblacional existe una gran heterogeneidad con un gradiente noroeste/sudeste, por ejemplo en Dinamarca hay 88% de casos $\Delta F508$ pero 50% en Italia. Se dice que los inicios de esta mutación ocurrieron hace 52,000 años en el norte de Europa.

La extensión de esta mutación en la población europea se ha explicado por una ventaja selectiva de los individuos portadores de ella. Según esta hipótesis, estarían más protegidos frente al cólera, la enfermedad letal mas común en esa época.