

**a) A partir del gen o proteína de interés para ustedes dar su link a NCBI-Gene como una entrada de Entrez, por ej.: <http://www.ncbi.nlm.nih.gov/gene/3630> Expliquen brevemente lo que hace el gen y por qué la eligieron.**

El gen que seleccionamos es conocido como HLA-DQB1, y su link a NCBI-Gene es: <https://www.ncbi.nlm.nih.gov/gene/3119>. Este gen es de tipo codificador de proteínas, fue elegido dado que es uno de los genes que interviene en el desarrollo de la celiaquía, la enfermedad que elegimos para el desarrollo de este trabajo práctico.

El riesgo de desarrollar la enfermedad se incrementa cuando ocurren ciertas variantes en el gen HLA-DQA1 y HLA-DQA2, ambos genes se encuentran en el brazo corto del cromosoma 6 (6p21.3). Estos genes codifican proteínas que se encuentran en las células inmunitarias. Pertenecen a una familia de genes llamados complejo de antígeno leucocitario humano. El complejo HLA ayuda al sistema inmunitario a distinguir los antígenos propios de los antígenos exógenos como virus y bacterias.

La celiaquía es una enfermedad autoinmune que causa intolerancia permanente al gluten de alimentos como el trigo, la cebada, el centeno y la avena. Al ingerir alimentos que contienen gluten se genera inflamación que afecta directamente al tracto gastrointestinal, dañando las vellosidades que tienden a acortarse y aplanarse, provocando pérdida de nutrientes, anemia, deficiencias vitamínicas, y tendencia a padecer algunos cánceres o enfermedades neurológicas.

Esta enfermedad se asocia con una respuesta inmunitaria inapropiada a un segmento de la proteína del gluten llamada gliadina. Esta activación inapropiada del sistema inmunitario desencadena la inflamación que daña a los órganos, dando lugar a los síntomas comentados anteriormente. Es importante destacar que aproximadamente el 30% de las personas poseen las variantes que desencadenan celiaquía en estos genes, pero sin embargo solo el 3% de la población padece la enfermedad.

Otros genes asociados:

<https://www.ncbi.nlm.nih.gov/gene/?term=CELIAC3>

Más información en:

<https://www.ivami.com/es/pruebas-geneticas-mutaciones-de-genes-humanos-enfermedades-neoplasias-y-farmacogenetica/1796-pruebas-geneticas-celiaca-enfermedad-celiac-disease-genes-hla-dqa1-y-hla-dqb1>

**b) ¿Cuántos genes / proteínas homólogas se conocen en otros organismos? Utilicen la información que está en la base de datos de HomoloGene y en la bases de datos Ensembl. Describan los resultados en ambas bases de datos, y en qué se diferencian. Mencionen sobre qué tan común creen que son estos genes o proteínas y a qué grupos taxonómicos pertenecen (sólo en las bacterias, en los vertebrados, etc.)**

En la base de datos HomoloGene, encontramos 702 genes ortólogos a HLA-DBQ1, pertenecientes únicamente a la familia taxonómica de los vertebrados (tetrápodos). El árbol generado por HomoloGene es el siguiente:

- ▾ tetrapods
  - birds
  - turtles
  - alligators and others
  - lizards & snakes
  - mammals
  - amphibians

Dentro de los vertebrados, la mayoría de las coincidencias se encuentran en los mamíferos, con 624 resultados, donde los seres vivos con mayor éxito (además del homo sapiens) son las especies de monos, ratas y cerdos.

En la base de datos de Ensembl encontramos la siguiente información:

**Gene: HLA-DQB1** ENSG00000179344

Description	major histocompatibility complex, class II, DQ beta 1 [Source:HGNC Symbol;Acc:HGNC:4944]
Gene Synonyms	CELIAC1, HLA-DQB, IDDM1
Location	<a href="#">Chromosome 6: 32 659 467-32 668 383</a> reverse strand. GRCh38:CM000668.2 <a href="#">View alleles</a> of this gene on alternative sequences
About this gene	This gene has 9 transcripts ( <a href="#">splice variants</a> ), 1 gene allele, <a href="#">128 orthologues</a> , <a href="#">13 paralogues</a> and is associated with <a href="#">11 phenotypes</a> .
Transcripts	<a href="#">Show transcript table</a>

Vemos que la búsqueda devuelve información sobre 128 ortólogos

Species set	Show details	<a href="#">With 1:1 orthologues</a>	<a href="#">With 1:many orthologues</a>	<a href="#">With many:many orthologues</a>	<a href="#">Without orthologues</a>
<b>Primates (22 species)</b> Humans and other primates	<input type="checkbox"/>	12	5	1	<a href="#">4</a>
<b>Rodents and related species (24 species)</b> Rodents, lagomorphs and tree shrews	<input type="checkbox"/>	0	19	3	<a href="#">2</a>
<b>Laurasiatheria (38 species)</b> Carnivores, ungulates and insectivores	<input type="checkbox"/>	0	18	14	<a href="#">6</a>
<b>Placental Mammals (89 species)</b> All placental mammals	<input type="checkbox"/>	12	44	19	<a href="#">14</a>
<b>Sauropsida (27 species)</b> Birds and Reptiles	<input type="checkbox"/>	0	5	2	<a href="#">20</a>
<b>Fish (65 species)</b> Ray-finned fishes	<input type="checkbox"/>	0	0	0	<a href="#">65</a>
<b>All (200 species)</b> All species, including invertebrates	<input checked="" type="checkbox"/>	12	53	21	<a href="#">114</a>

Dentro de los cuales encontramos nuevamente, solo vertebrados.

En conclusión, este gen está ampliamente distribuido en los animales vertebrados, con una mayor aparición en mamíferos.

Fuente: <https://www.ncbi.nlm.nih.gov/gene/3119/ortholog/similargenes/?scope=40674>

c) ¿Cuántos transcritos y cuántas formas alternativas de splicing son conocidos para este gen / proteína? ¿Cuáles de estos splicing alternativos se expresan? ¿Tienen funciones alternativas? Buscar evidencia de esto en las base de datos de NCBI y en los transcritos de Ensembl ¿Cómo el número de splicings alternativos diferente entre las dos bases de datos y cuál piensan que es más precisa y por qué?

Según Ensembl, filtrando por el nombre del gen “HLA-DQB1” vemos que existen 9 transcritos, listados a continuación:

Transcript ID	Name	bp	Protein	Biotype	CCDS	UniProt Match
<a href="#">ENST00000434651.7</a>	HLA-DQB1-205	1605	<a href="#">261aa</a>	Protein coding	<a href="#">CCDS43451</a>	<a href="#">Q5Y7D6</a>
<a href="#">ENST00000399084.5</a>	HLA-DQB1-204	1734	<a href="#">261aa</a>	Protein coding	<a href="#">CCDS43451</a>	<a href="#">Q5Y7D6</a>
<a href="#">ENST00000374943.8</a>	HLA-DQB1-201	1656	<a href="#">269aa</a>	Protein coding	<a href="#">CCDS59006</a>	<a href="#">Q5SU54</a>
<a href="#">ENST00000399079.7</a>	HLA-DQB1-202	1501	<a href="#">224aa</a>	Protein coding		<a href="#">A2AAZ0</a>
<a href="#">ENST00000399082.7</a>	HLA-DQB1-203	1219	<a href="#">134aa</a>	Protein coding		<a href="#">A2AAY8</a>
<a href="#">ENST00000484729.2</a>	HLA-DQB1-208	783	<a href="#">136aa</a>	Nonsense mediated decay		<a href="#">E9PIB1</a>
<a href="#">ENST00000460185.1</a>	HLA-DQB1-207	428	No protein	Protein coding CDS not defined		-
<a href="#">ENST00000443574.1</a>	HLA-DQB1-206	309	No protein	Protein coding CDS not defined		-
<a href="#">ENST00000487676.1</a>	HLA-DQB1-209	4644	No protein	Retained intron		-

Según la base de datos de NCBI, solo contamos con 3 transcritos

## HLA-DQB1 Transcripts BETA

HLA-DQB1 — major histocompatibility complex, class II, DQ beta 1 (*Homo sapiens*)

3 transcripts

[Switch to gene view](#)

Edit

Download

Select columns

<input type="checkbox"/>	Gene ID	Symbol	Transcript	Length (nt)	Protein	Length (aa)	Protein name	Isoform
<input type="checkbox"/>	3119	HLA-DQB1	NM_002123.5	1605	NP_002114.3	261	HLA class II histocompatibility antigen, DQ beta 1 chain	1 precursor
<input type="checkbox"/>	3119	HLA-DQB1	NM_001243961.2	1629	NP_001230890.1	269	HLA class II histocompatibility antigen, DQ beta 1 chain	2 precursor
<input type="checkbox"/>	3119	HLA-DQB1	NM_001243962.1	1638	NP_001230891.1	261	HLA class II histocompatibility antigen, DQ beta 1 chain	1 precursor
Rows per page					25	Showing transcripts for 1-1 of 1 Gene		

Genes transcritos según Ensembl:

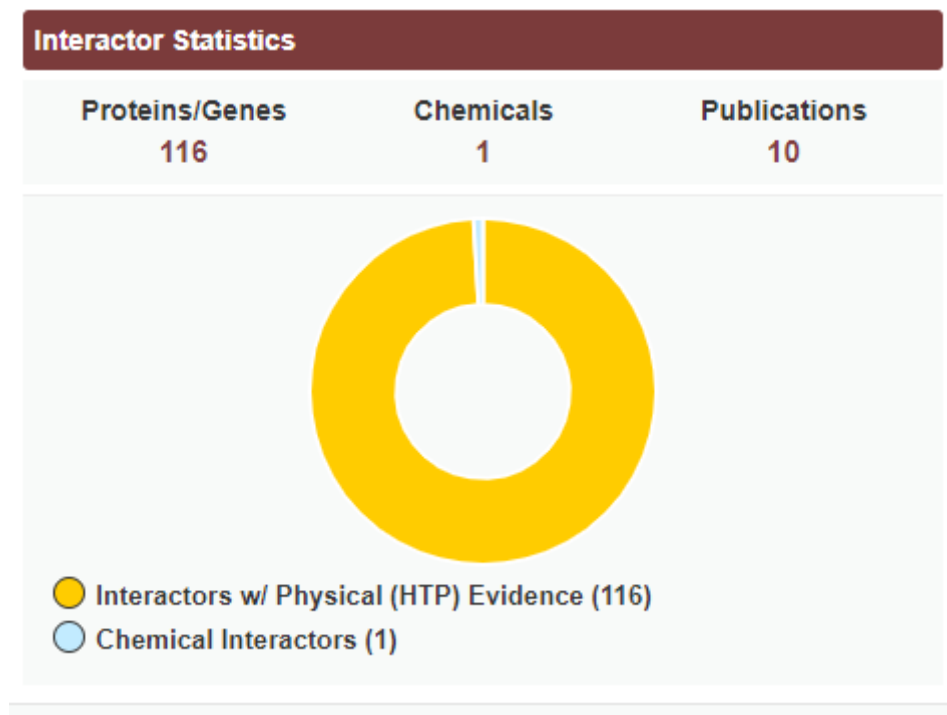
[http://www.ensembl.org/Homo\\_sapiens/Gene/Summary?q=ENSG00000179344;r=6:32659467-32668383](http://www.ensembl.org/Homo_sapiens/Gene/Summary?q=ENSG00000179344;r=6:32659467-32668383)

Genes transcritos según NCBI HLA-DBQ1:

[https://www.ncbi.nlm.nih.gov/datasets/tables/genes/?table\\_type=transcripts&key=375a0f9b3d2455fbc31cc4969612b05f](https://www.ncbi.nlm.nih.gov/datasets/tables/genes/?table_type=transcripts&key=375a0f9b3d2455fbc31cc4969612b05f)

d) ¿Con cuántas otras proteínas interactúa 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 gene/proteína tanto en la base de datos NCBI Gene como en la base de datos UniProt. Compare las dos tablas entre sí. ¿Hay proteínas que interactúan únicas para cada tabla?

Al buscar el gen HLA-DBQ1 en UniProt, encontramos 116 coincidencias, que se detallan en profundidad en la base de datos de bioGrid



A continuación, listamos las primeras 5 interacciones que se muestran en la tabla

Interactor	Organism / Chemical Type	Aliases	Description	Evidence
Insulin Pork	Biologic	Insulin (pork), Porcine insulin, Insulin porcine, Insulin, porcine, Insulin purified ... <a href="#">more</a>	Insulin isolated from pig pancreas. Composed of alpha and beta chains, processed from pro-insulin. Forms ... <a href="#">more</a>	5 <a href="#">View</a>
HLA-DQA1	H. sapiens	CD, GSE, DQ-A1, CELIAC1, HLA-DQA, DAQB-109B10.1	major histocompatibility complex, class II, DQ alpha 1	3 <a href="#">View</a>
APOM	H. sapiens	G3a, NG20, apo-M, HSPC336, DADB-127H9.5	apolipoprotein M	1 <a href="#">View</a>
ASIC4	H. sapiens	BNAC4, ACCN4	acid-sensing (proton-gated) ion channel family member 4	1 <a href="#">View</a>
ATP2A1	H. sapiens	ATP2A, SERCA1	ATPase, Ca++ transporting, cardiac muscle, fast twitch 1	1 <a href="#">View</a>

Es interesante destacar que este gen interactúa con 5 químicos, todos relacionados directamente con la insulina porcina

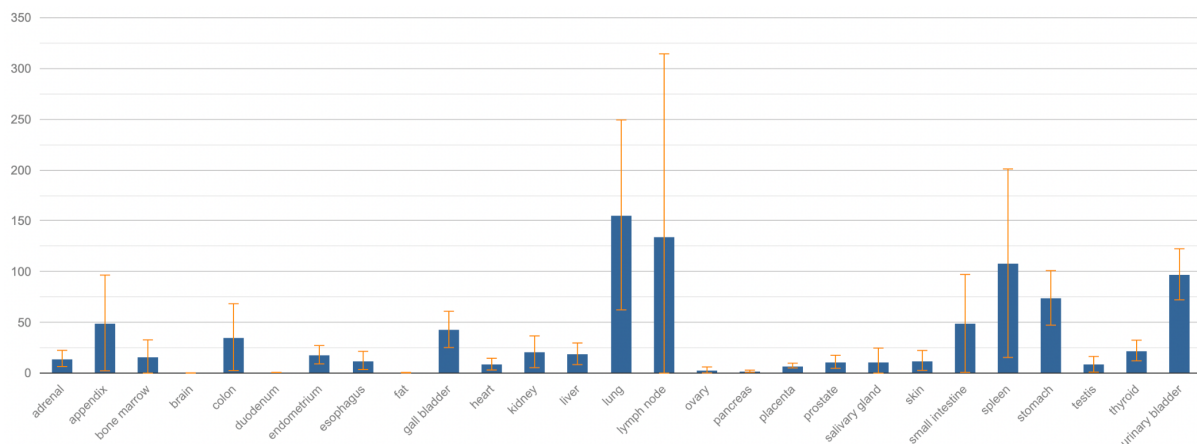
Chemical ▲	Chemical Type ◆	Action ◆	Dataset ◆
Insulin Pork	biologic	Unknown	Imming P (2006)
Insulin Pork	biologic	Unknown	Levy-Marchal C (1992)
Insulin Pork	biologic	Unknown	Overington JP (2006)
Insulin Pork	biologic	Unknown	Mahrn MZ (1992)
Insulin Pork	biologic	Unknown	Berman HM (2000)

BioGrid: <https://thebiogrid.org/109364>

En la base NCBI encontramos 79 interacciones:  
<https://www.ncbi.nlm.nih.gov/gene/3119#interactions>

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

En el siguiente diagrama obtenido en NCBI Gene, podemos apreciar en qué tejidos se expresa más este gen



Dentro de sus procesos biológicos podemos encontrar:

- Este gen se ve involucrado en la señalización de la ruta metabólica de las células T
- La respuesta adaptativa inmune, lo cual hace que reaccione de forma errónea al gluten, entre otros químicos
- La producción de inmunoglobulina mediante la respuesta inmune
- La respuesta positiva en la regulación de la activación de las células T

- Este gen se ve involucrado en el procesamiento y la presentación de péptidos exógenos vía MHC clase 2.

En cuanto a su localización, este gen puede encontrarse:

- En la membrana de Golgi
- En el retículo endoplasmático
- En la membrana de plasma

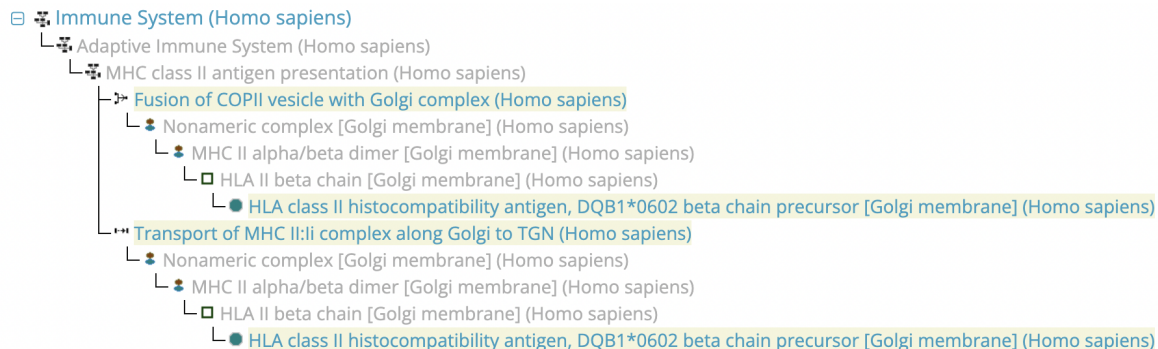
Entre otros, más información sobre los procesos y los componentes en los cuales se encuentra el gen se pueden ver en <https://www.ncbi.nlm.nih.gov/gene/3119/#process>

### Otros enlaces de interés

GeneOntology: <http://amigo.geneontology.org/amigo/search/ontology?q=HLA-DQB1>

**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)**

No pudimos obtener coincidencias buscando en la base de datos de KEGG con las keys "HLA-DQB1" ni "CELIAC1". Por otro lado, en reactome obtuvimos las siguientes coincidencias:



<https://reactome.org/content/detail/R-HSA-2214373>

**g) Entrar en la base de datos de variantes genéticas dbSNP e intentar interpretar o encontrar info sobre alguna variante (reference SNP - rsXXXX) asociada con la patología investigada en su gen de interés. ¿Qué variante es? ¿Hay información sobre la frecuencia que tiene esta variante en la población? ¿Qué grupo étnico parece ser el más afectado? NOTA: Para hacer este ejercicio les pueden servir algunas otras bases de datos como estas (entre otras): <http://www.genecards.org> <https://www.ncbi.nlm.nih.gov/snp/> (para obtener información de la variante en la población) <http://www.ncbi.nlm.nih.gov/clinvar/> (para obtener información clínica del gen y sus variantes) <https://ghr.nlm.nih.gov>**

En la base de SNP de NCBI encontramos 10.248 resultados asociados a variaciones del gen HLA-DQB1 asociado a la celiaquía. Algunas de estas variaciones son: rs1762 [Homo sapiens], rs1049066 [Homo sapiens] y rs1337325077 [Homo sapiens].

Al menos dos combinaciones específicas de variantes del gen HLA-DQB1 se relacionan directamente con un incremento en el riesgo de contraer la enfermedad celíaca. Uno de estos haplotipos (agrupación física de variantes genómicas o polimorfismos) que tienden a heredarse juntas., conocido como DQ2, se compone por la proteína producida por las variantes del gen HLA-DQB1 conocidas como HLA-DQB1\*02:01 or HLA-DQB1\*02:02. El otro haplotipo, se compone por la proteína HLA-DQB1\*03:02, producida por el en en cuestión , combinado con proteínas producidas por otro en asociado a la celiaquía, conocido como HLA-DQA1.