Ejercicio 6. Trabajo con Bases de Datos Biológicas

**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 por qué la eligieron.**

El gen elegido forma parte del complejo mayor de histocompatibilidad o CMH, una familia de

genes hallados en todos los vertebrados y ubicados en el brazo corto del cromosoma 6 en

humanos, cuya función es la codificación de moléculas (glucoproteínas) denominadas

antígenos leucocitarios humanos o antígenos de histocompatibilidad, que participan en la

presentación de antígenos a los linfocitos T permitiendo la activación de procesos críticos

en la generación de la respuesta inmunitaria. En general, el CMH permite distinguir lo propio

de lo extraño. En humanos, el CMH se denominó primeramente «complejo HLA» (del

acrónimo inglés para human leucocytic antigen), porque las moléculas del CMH se

descubrieron como antígenos que diferencian los leucocitos de distintas personas y que

producen la respuesta inmunitaria del receptor en trasplantes, produciendo el rechazo.

<https://www.ncbi.nlm.nih.gov/nuccore/NM_002122.5>

<https://www.omim.org/entry/212750>

<https://pubmed.ncbi.nlm.nih.gov/26303674/>

**b) ¿Cuántos genes / proteínas homólogas se conocen en otros organismos?**

En la base de datos Gene del NCBI se pueden encontrar 10 genes ortólogos, todos dentro del magnorden Boreoeutheria (Mamíferos Placentarios). Las especies listadas son:

* Boreoeutheria
* Roedores
  + Mus musculus
* Murciélagos
  + Pipistrellus kuhlii
* Primates
  + Colobus angolensis palliatus
  + Chlorocebus sabaeus
  + Macaca nemestrina
  + Theropithecus gelada
  + Trachypithecus francoisi
  + Rhinopithecus roxellana
  + Homo sapiens
  + Saimiri boliviensis

Fuente:  
<https://www.ncbi.nlm.nih.gov/gene/3117/ortholog/?scope=1437010>

En cuanto a la base de datos Ensembl, vemos que existen 11 genes ortólogos 1:1, es decir que solo una copia del gen se encuentra en cada especie. Estas especies son:

* Black snub-nosed monkey (Rhinopithecus bieti)
* Bonobo (Pan paniscus)
* Chimpanzee (Pan troglodytes)
* Crab-eating macaque (Macaca fascicularis)
* Gibbon (Nomascus leucogenys)
* Golden snub-nosed monkey (Rhinopithecus roxellana)
* Gorilla (Gorilla gorilla gorilla)
* Pig-tailed macaque (Macaca nemestrina)
* Sooty mangabey (Cercocebus atys)
* Vervet-AGM (Chlorocebus sabaeus)
* White-tufted-ear marmoset (Callithrix jacchus)

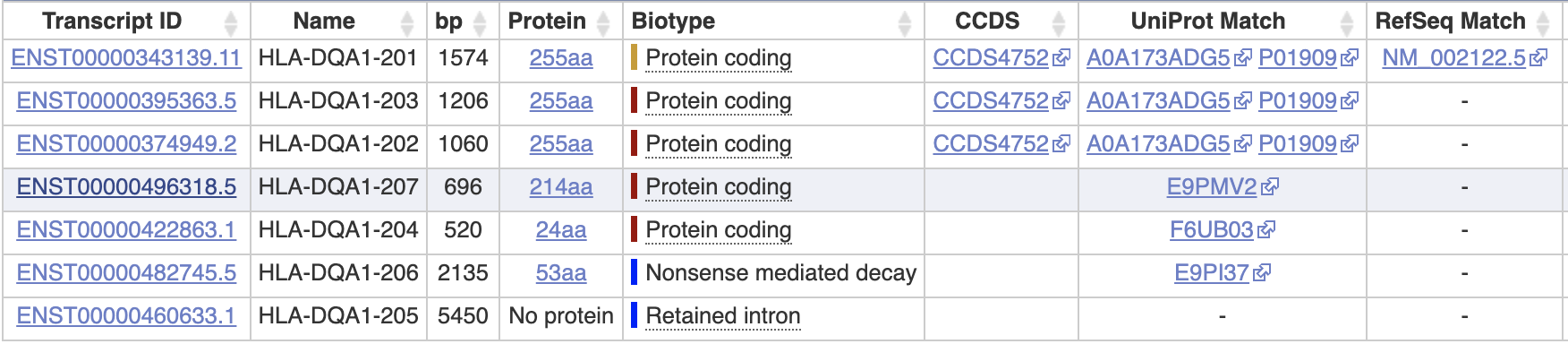
Fuente:

<https://www.ensembl.org/Homo_sapiens/Gene/Compara_Ortholog?db=core;g=ENSG00000196735;r=6:32637406-32642981;t=ENST00000496318>

Se puede observar que, si bien en la base del NCBI aparecen dos resultados que no son Primates, este gen está fuertemente asociado al orden de los Mamíferos Placentarios Primates, existiendo gran cantidad de primates que cuentan con genes ortólogos al mismo.

**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 transcriptos 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?**

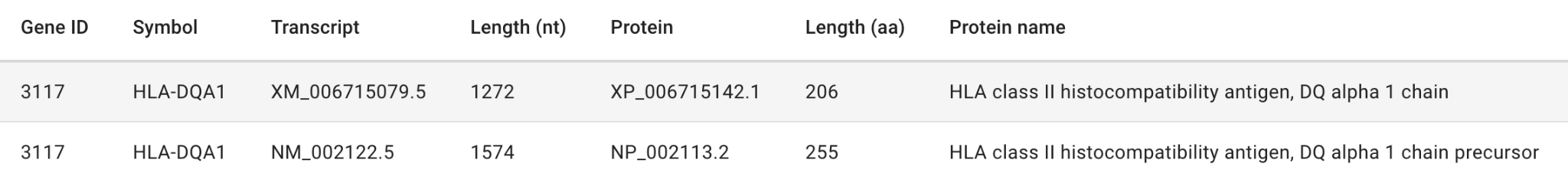
El Gen cuenta con 7 transcriptos en la base de datos de Ensambl, listados a continuación:



<https://www.ensembl.org/Homo_sapiens/Gene/Splice?db=core;g=ENSG00000196735;r=6:32637406-32642981;t=ENST00000496318>

Podemos observar que 5 de ellos cumplen con la función de codeo de proteínas mientras que uno de ellos cumple con la función de **NMD** (*Nonsense Mediated Decay, un mecanismo celular de vigilancia del ARN mensajero para detectar mutaciones terminadoras*) y el restante ‘Retained Intron’.

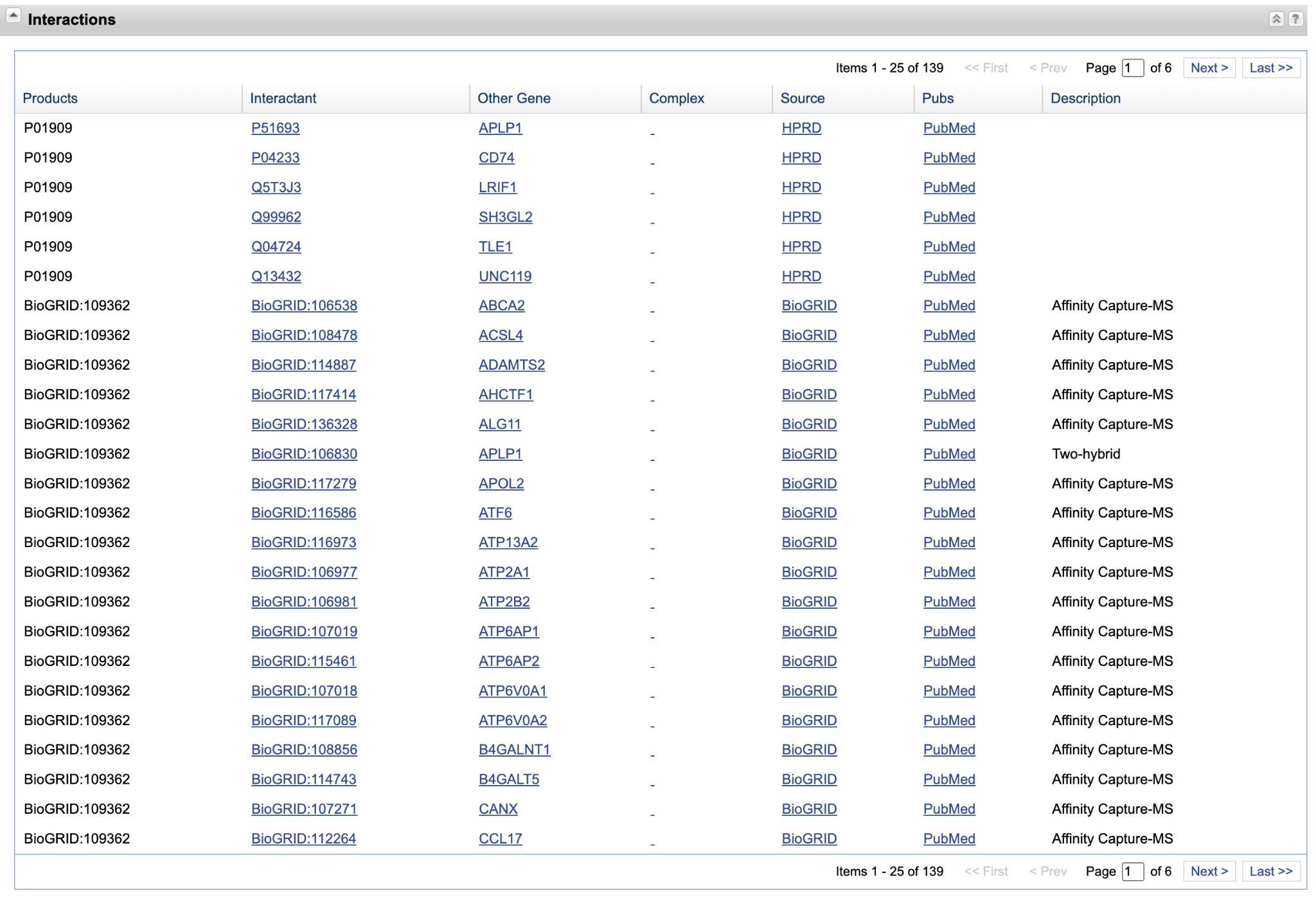
En la base de datos del NCBI, podemos encontrar solamente 2 transcriptos para este gen:



<https://www.ncbi.nlm.nih.gov/datasets/tables/genes/?table_type=transcripts&ids=3117&utm_source=gene&utm_medium=referral&utm_campaign=summary>

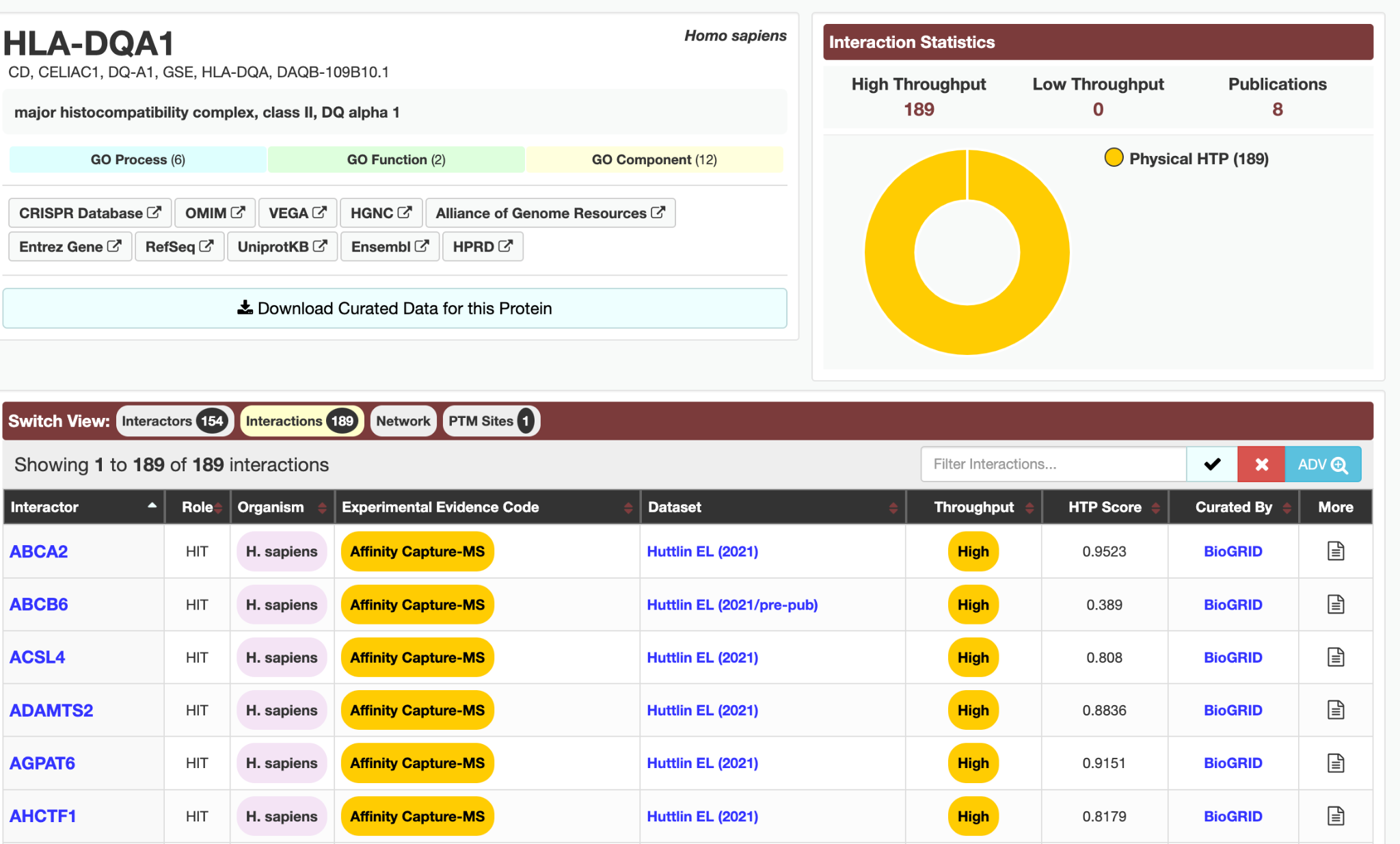
En este caso, la base de datos de Ensembl parece ser la más precisa, por contar con mayor número de transcripciones alternativas de este gen.

**d) ¿Con cuántas 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 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?**

En la base de datos del NCBI podemos encontrar 139 interacciones con otras proteínas. A modo de ejemplo, listamos algunas de ellas:  


<https://www.ncbi.nlm.nih.gov/gene/3117#interactions>

En cambio, dentro de Uniprot / BioGrid se encuentran listadas 189 interacciones del gen con otras proteinas.



<https://thebiogrid.org/109362>

**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)? 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.**

**Componentes:**

* Vesícula endocítica
* Endosoma
* Membrana lisosomal
* Membrana plasmática
* Membrana de Golgi

**Procesos:**

* Respuesta inmune adaptativa
* Procesamiento de antígenos y la presentación de antígenos peptídicos exógenos a través del MHC de clase II
* Respuesta inmune
* Producción de inmunoglobulina implicado en la respuesta inmunitaria mediada por inmunoglobulina
* Ensamblaje del antígeno peptídico con el complejo proteico MHC clase II
* Regulación positiva de la activación de células T

**Funciones:**

* Unión del complejo proteico MHC clase II
* Actividad del receptor MHC clase II
* Actividad del receptor MHC clase II
* Unión al antígeno peptídico
* Unión a proteínas

<https://www.ncbi.nlm.nih.gov/gene/3117#general-gene-info>

**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).**

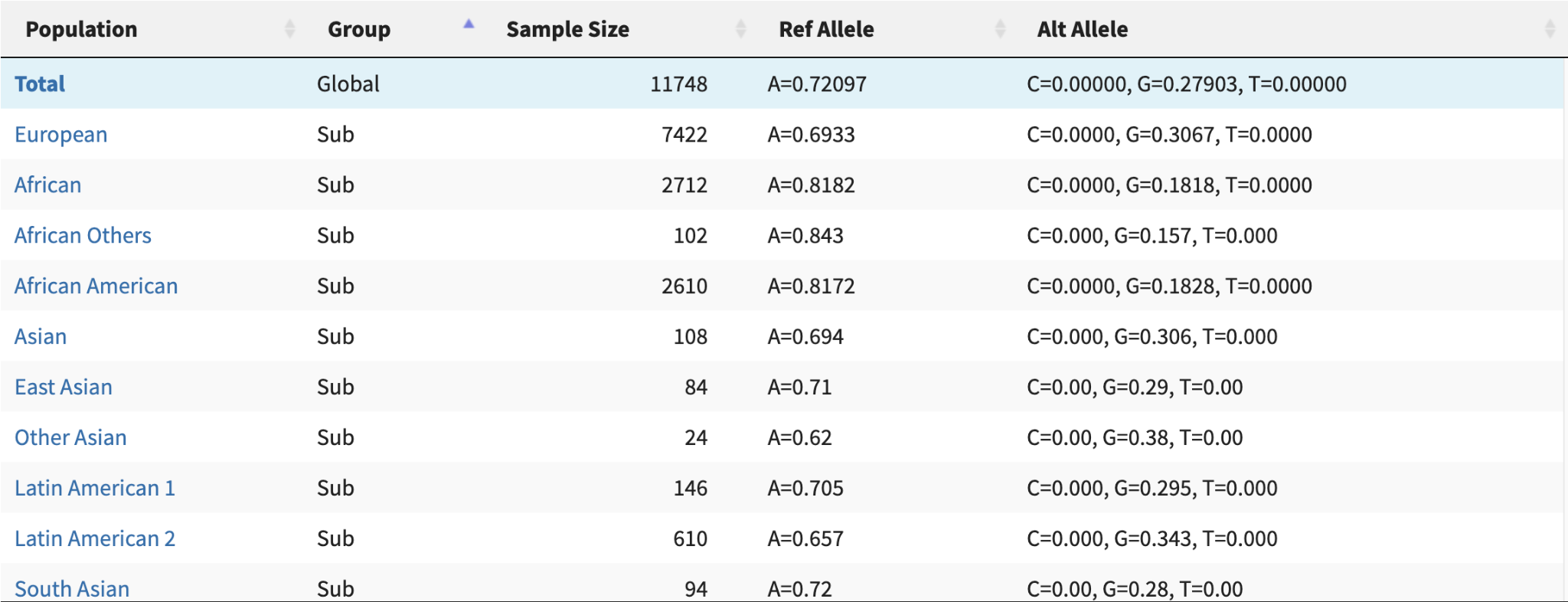
En la base de datos de KEGG, se pueden encontrar 24 pathways en las que participa el producto génico del gen de interés (HLA-DQA1)

* hsa04145 -> fagosoma
* hsa04514 -> moléculas de adhesión celular
* hsa04612 -> Procesamiento y presentación de antígenos
* hsa04640 -> Linaje de células hematopoyéticas
* hsa04658 -> Diferenciación de células Th1 y Th2
* hsa04659 -> diferenciación de células Th17
* hsa04672 -> Red inmune intestinal para la producción de IgA
* hsa04940 -> Diabetes mellitus tipo I
* hsa05140 -> leishmaniosis
* hsa05145 -> toxoplasmosis
* hsa05150 -> Infección por Staphylococcus aureus
* hsa05152 -> Tuberculosis
* hsa05164 -> gripe A
* hsa05166 -> Infección por el virus de la leucemia de células T humanas 1
* hsa05168 -> Infección por el virus del herpes simple 1
* hsa05169 -> Infección por el virus de Epstein-Barr
* hsa05310 -> Asma
* hsa05320 -> Enfermedad tiroidea autoinmune
* hsa05321 -> Enfermedad inflamatoria intestinal
* hsa05322 -> Lupus eritematoso sistémico
* hsa05323 -> Artritis Reumatoide
* hsa05330 -> Rechazo de aloinjerto
* hsa05332 -> Enfermedad de injerto contra huésped
* hsa05416 -> Miocarditis viral

<https://www.genome.jp/entry/hsa:3117>

**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?**

Encontramos la variación rs3667 (variante de intrón). A continuación mostramos en un gráfico la frecuencia que tiene esta variante en la población:



<https://www.ncbi.nlm.nih.gov/snp/rs3667#frequency_tab>