UNIVERSIDAD DE GRANADA

Práctica 2

Herramientas bioinformáticas en Genómica

Biología Molecular - Grado en Medicina
Curso 2020/21

Profesora: Marisol Benítez Cantos

¿Qué vamos a hacer?

Vamos a usar bases de datos y herramientas bioinformáticas para conocer mejor un gen que origina una enfermedad.

Durante la práctica trabajamos con un gen de ejemplo.

En casa elegid otra enfermedad monogénica y repetid la práctica para rellenar la tabla de la derecha.

En el examen habrá preguntas **sencillas** tipo test de opción múltiple.

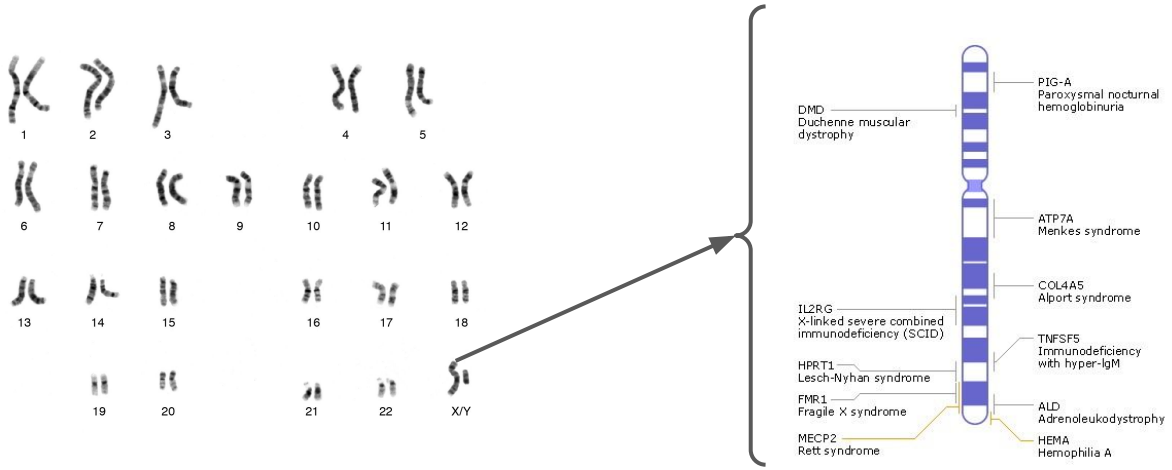
PRÁCTICA II: BIOINFORMÁTICA

Nombre y apellidos: María Soledad Benítez Cantos

Enfermedad:	
Neoplasia endocrina múltiple de tipo 1	
Nombre oficial del gen responsable	*MEN1*
Alias	menin
Evidencia de la asociación gen-enfermedad	Frase: “We conclude that many "hyperplastic" parathyroid tumors in familial MEN-1 are in fact monoclonal and may progress or even begin to develop by inactivation of the MEN-1 gene (at 11q13) in a precursor cell.” Enlace a publicación: https://pubmed.ncbi.nlm.nih.gov/2568586/
Ensembl ID del gen	ENSG00000133895
Localización cromosómica (cromosoma, coordenadas y hebra)	Chromosome 11: 64,803,510-64,811,294 reverse strand
Genes adyacentes	MAP4K2, SF1, RASGRP2
dbSNP ID de una variante patogénica	rs104894266
Consecuencia de la variante	Codón de parada prematuro
Número de transcritos	19
Ensembl ID del transcrito seleccionado	ENST00000450708.7
Longitud del transcrito (pb)	3162
Número de exones	10
Longitud de la proteína (aa)	615
Secuencia del transcrito en formato FASTA	>ENST00000450708.7 MEN1-214 cdna:protein_coding GTGCGCCGCGGTGCCTAGTGTGGGATGTAAGCGCG GAGGCCGCCGCCCACCGCCCGCCGC (...)
Primers diseñados	Forward: CGTGAGCTGGTGAAGAAGGT Reverse: GCTGTCCAATTTGGTGCCTG
Identificador del miRNA que lo regula	hsa-miR-6877-5p
Secuencia del miRNA maduro	agggccgaaggguggaagcugc
Laboratorio que realiza el diagnóstico genético de la enfermedad (preferiblemente en España)	Laboratorio de Genética Clínica SL, LabGenetics (Madrid)

¿Qué es la Genómica?

Es el campo de la Biología Molecular dedicado al estudio de **todo el material genético** de un organismo. Con técnicas de Biología Molecular y Bioinformática, se estudia la **estructura**, **mapeo**, **función**, evolución y edición de los genomas.

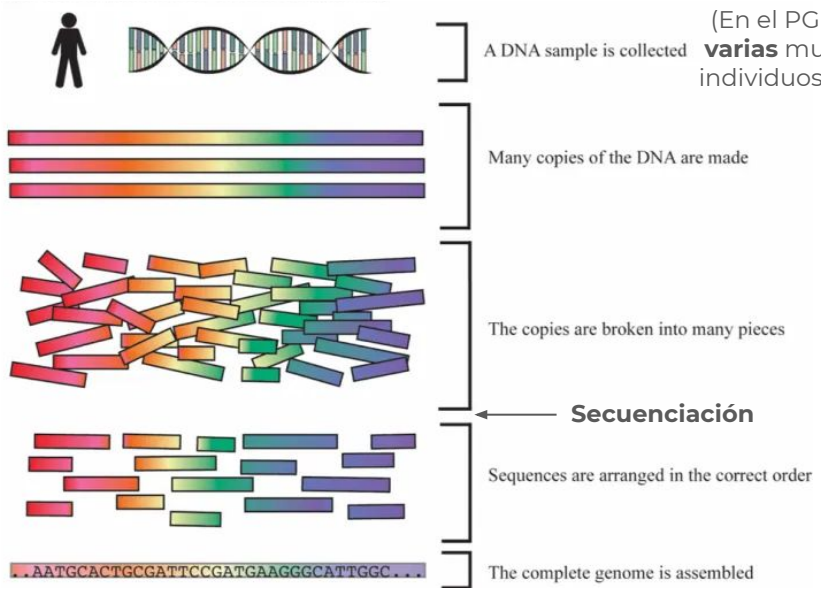


Proyecto Genoma Humano (1999-2003)

Secuenciar el genoma humano (3 millones de pares de bases) y **publicar** los datos:

- Identificar genes y regiones regulatorias importantes
- Entender su papel en las enfermedades
- Investigar nuestros orígenes

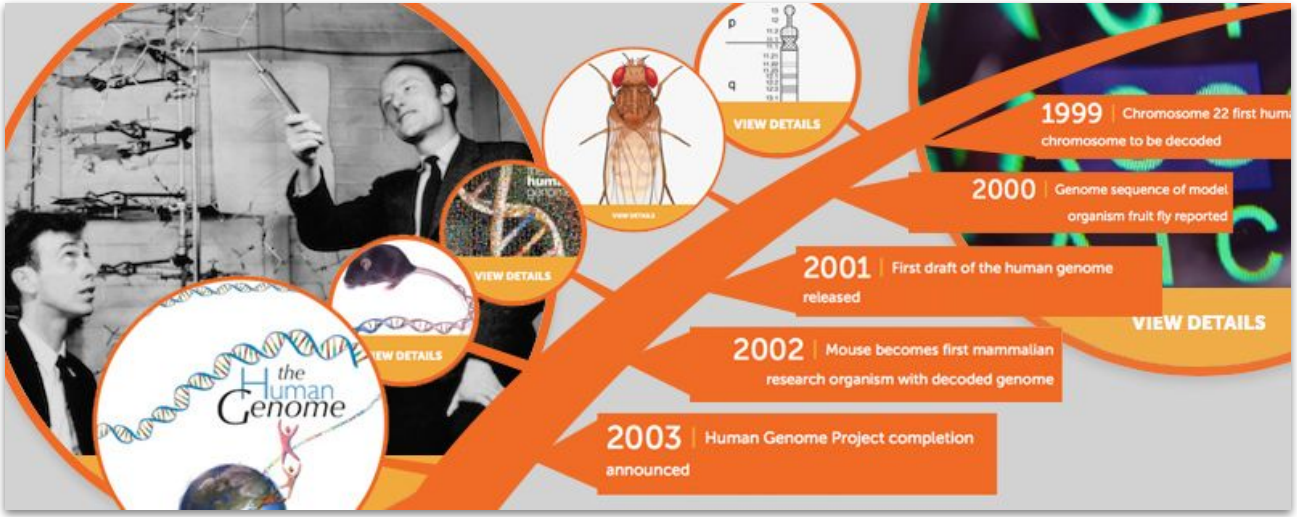
Proceso de secuenciación de un genoma



(En el PGM fueron **varias** muestras de individuos distintos)

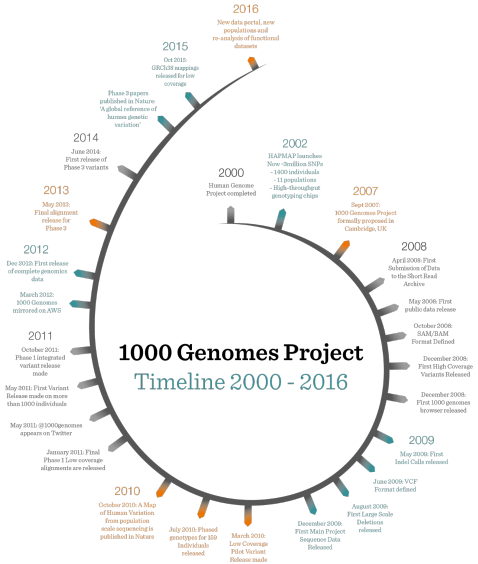
Proyecto Genoma Humano (1999-2003)

- El ADN secuenciado procedía de **varios individuos** donantes anónimos.
- La secuenciación se llevó a cabo entre **20 laboratorios colaboradores** de Estados Unidos, Reino Unido, Francia, Alemania, Japón y China y se completó en **4 años**.

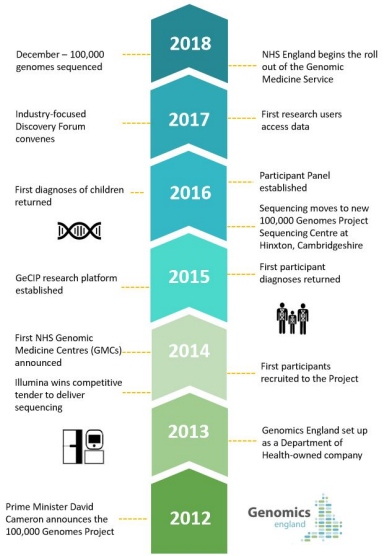


Nuevos proyectos genómicos

Secuenciar a humanos **individualmente** para conocer la **variación genética** humana.

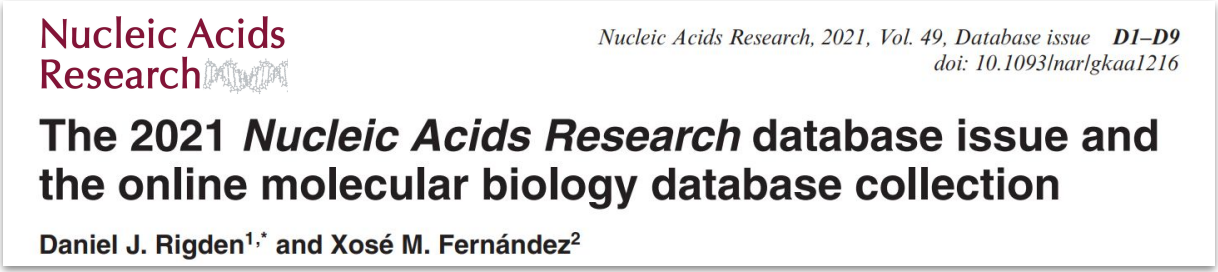


Proyecto **1000** genomas



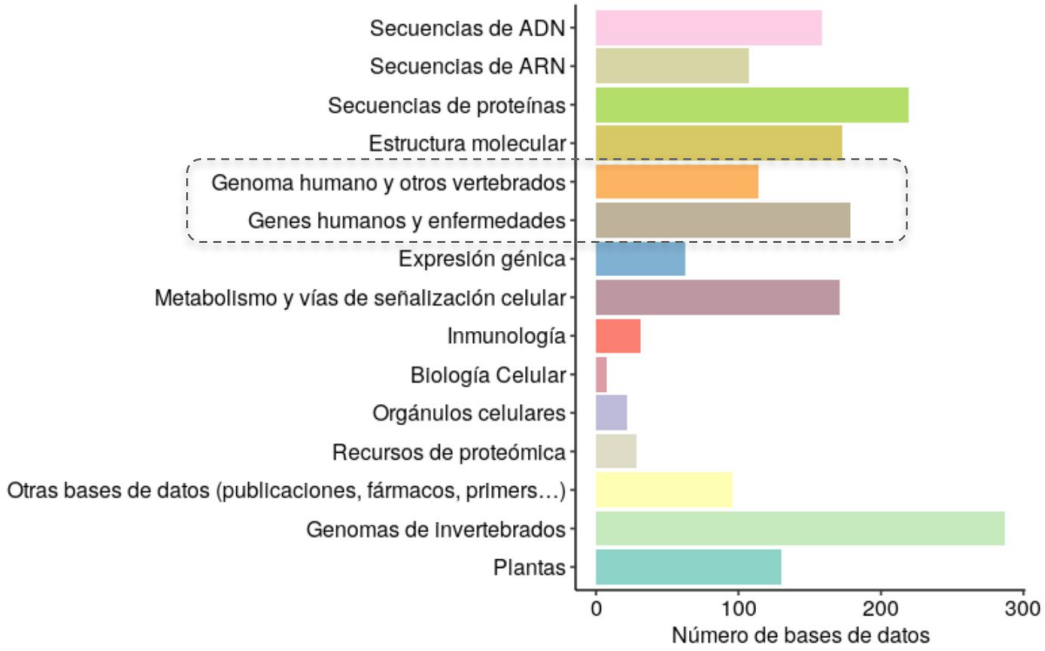
Proyecto **100000** genomas

Bases de datos



- La revista *Nucleic Acids Research* recopila anualmente todas las bases de datos de Biología Molecular que existen (actualmente **1641**)
- En 2021 se han añadido **90** bases de datos nuevas, **7** de ellas sobre COVID-19 y SARS-CoV-2.
- Puedes explorar el catálogo completo **aquí**.

Bases de datos



En esta práctica vamos a usar las siguientes bases de datos y herramientas bioinformáticas:

DisGeNET

Primer-BLAST

Ensembl

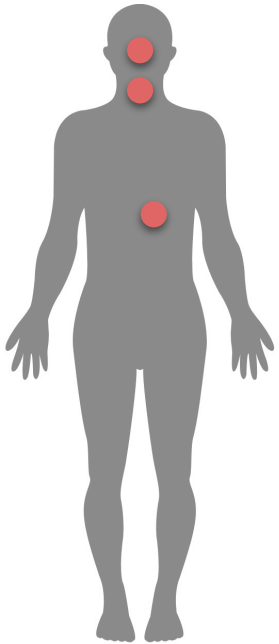
mRWalk

UniProt

miRBase

NCBI GTR Genetic Testing Registry

Neoplasia endocrina múltiple de tipo 1



Síndrome autosómico dominante causado por mutaciones en el gen supresor de tumores ***MEN1***.

Criterios para el diagnóstico

- **Clínico**: 2 o más tumores primarios endocrinos en las glándulas **p**aratiroides, **p**ituitaria anterior y/o islotes **p**ancreáticos (las 3 P's).
- **Familiar**: ocurrencia de un tumor asociado a *MEN1* en un familiar de primer grado de un paciente diagnosticado.
- **Genético**: identificación de una mutación germinal en *MEN1* en un individuo asintomático.

Antes de empezar

Consideraciones para seguir mejor la práctica:

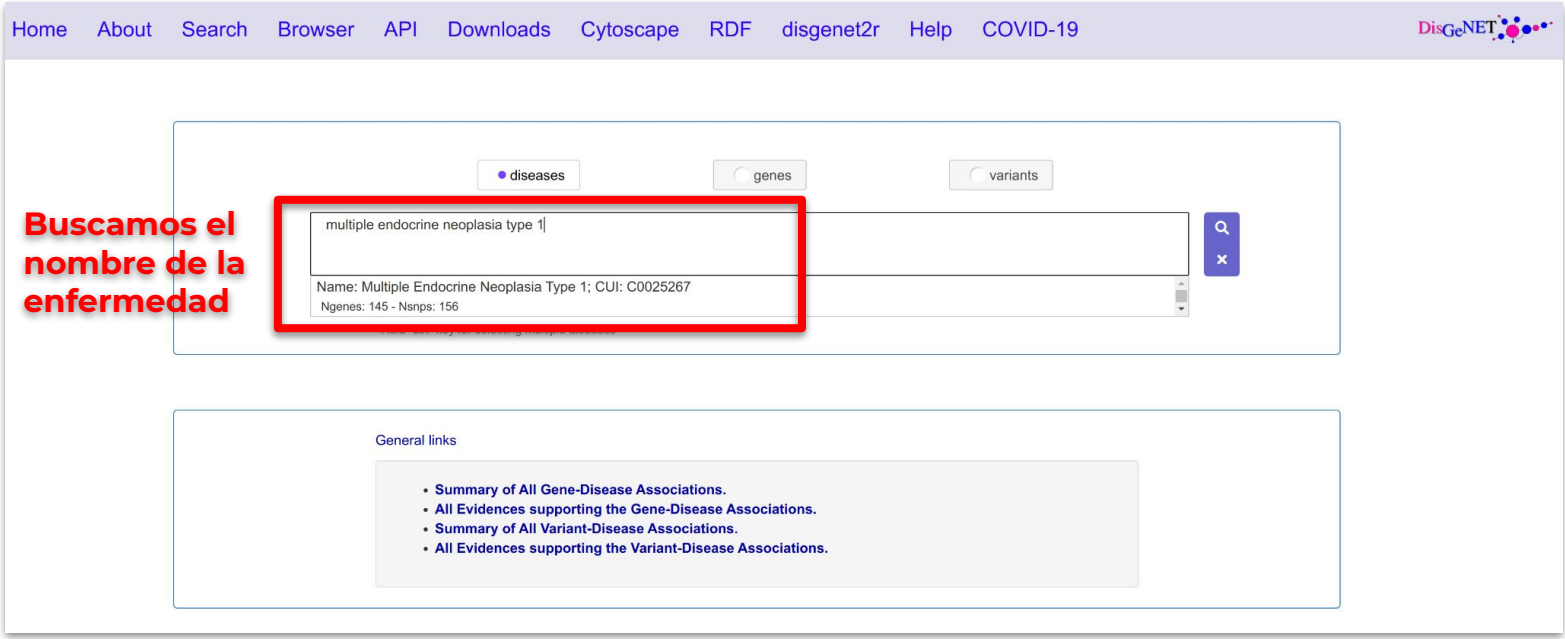
- Los **rectángulos rojos** de las diapositivas indican dónde hay que escribir o hacer click en las páginas web.
- Abre todos los enlaces en **pestañas nuevas**.
- Si en algún momento te pierdes (¡es totalmente normal!), levanta la mano.

¿Hay algún gen asociado a esta enfermedad?

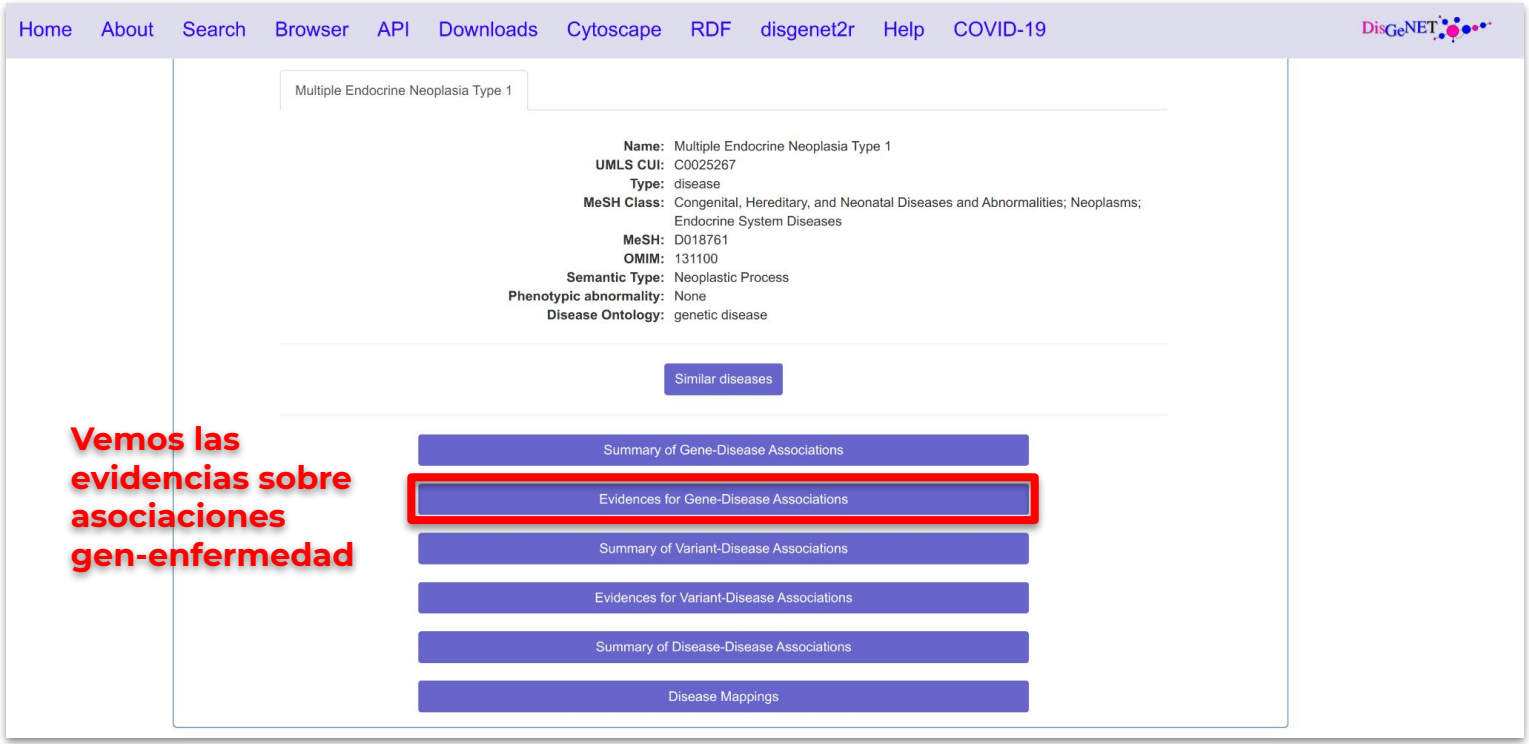
Respuesta: DisGeNET

Enlace

Buscar en DisGeNet



Asociaciones gen-enfermedad



Vemos las evidencias sobre asociaciones gen-enfermedad

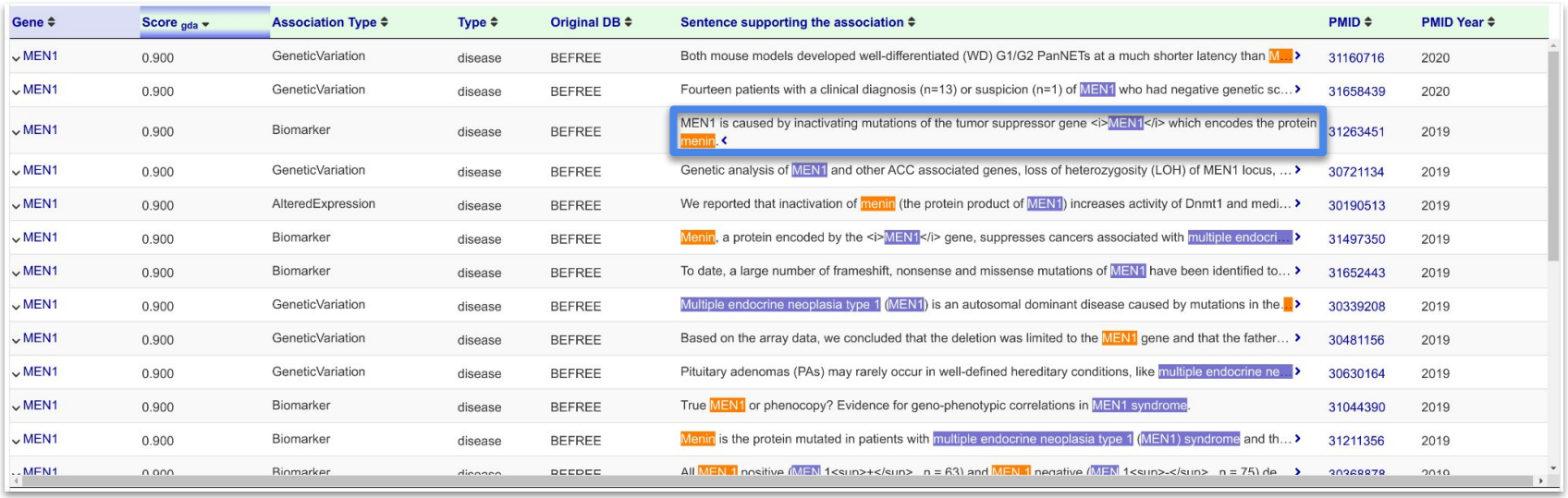
Asociaciones gen-enfermedad

Gen asociado

Puntuación de la asociación (máx. = 1)*

Frase que apoya la asociación gen-enfermedad

Enlace a la publicación (PubMed)



Gene	Score gda	Association Type	Type	Original DB	Sentence supporting the association	PMID	PMID Year
MEN1	0.900	GeneticVariation	disease	BEFREE	Both mouse models developed well-differentiated (WD) G1/G2 PanNETs at a much shorter latency than M…	31160716	2020
MEN1	0.900	GeneticVariation	disease	BEFREE	Fourteen patients with a clinical diagnosis (n=13) or suspicion (n=1) of MEN1 who had negative genetic sc…	31658439	2020
MEN1	0.900	Biomarker	disease	BEFREE	MEN1 is caused by inactivating mutations of the tumor suppressor gene MEN1 which encodes the protein menin.	31263451	2019
MEN1	0.900	GeneticVariation	disease	BEFREE	Genetic analysis of MEN1 and other ACC associated genes, loss of heterozygosity (LOH) of MEN1 locus, …	30721134	2019
MEN1	0.900	AlteredExpression	disease	BEFREE	We reported that inactivation of menin (the protein product of MEN1) increases activity of Dnmt1 and medi…	30190513	2019
MEN1	0.900	Biomarker	disease	BEFREE	Menin, a protein encoded by the MEN1 gene, suppresses cancers associated with multiple endocri…	31497350	2019
MEN1	0.900	Biomarker	disease	BEFREE	To date, a large number of frameshift, nonsense and missense mutations of MEN1 have been identified to…	31652443	2019
MEN1	0.900	GeneticVariation	disease	BEFREE	Multiple endocrine neoplasia type 1 (MEN1) is an autosomal dominant disease caused by mutations in the…	30339208	2019
MEN1	0.900	GeneticVariation	disease	BEFREE	Based on the array data, we concluded that the deletion was limited to the MEN1 gene and that the father…	30481156	2019
MEN1	0.900	GeneticVariation	disease	BEFREE	Pituitary adenomas (PAs) may rarely occur in well-defined hereditary conditions, like multiple endocrine ne…	30630164	2019
MEN1	0.900	Biomarker	disease	BEFREE	True MEN1 or phenocopy? Evidence for geno-phenotypic correlations in MEN1 syndrome.	31044390	2019
MEN1	0.900	Biomarker	disease	BEFREE	Menin is the protein mutated in patients with multiple endocrine neoplasia type 1 (MEN1) syndrome and th…	31211356	2019
MEN1	0.900	Biomarker	disease	BEFREE	All MEN 1 positive (MEN 1⁺, n = 63) and MEN 1 negative (MEN 1⁻, n = 75) de…	30368878	2019

*Esta tabla aparece ordenada por *Score* descendente por defecto, por lo que las mejores asociaciones siempre aparecen al principio.

¿Qué características tiene el gen?

Respuesta: 

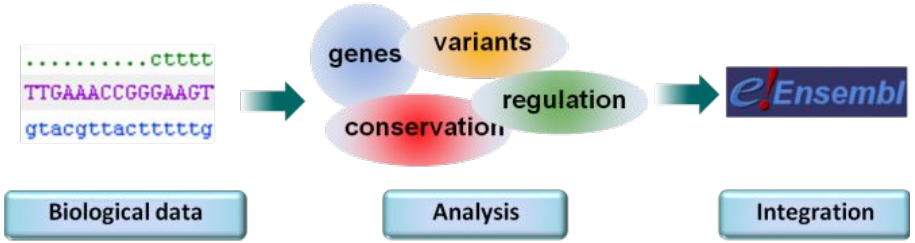
Enlace

Ensembl

Es un **navegador genómico** que actúa como punto de acceso único a datos procedentes de múltiples bases de datos y proyectos bioinformáticos.

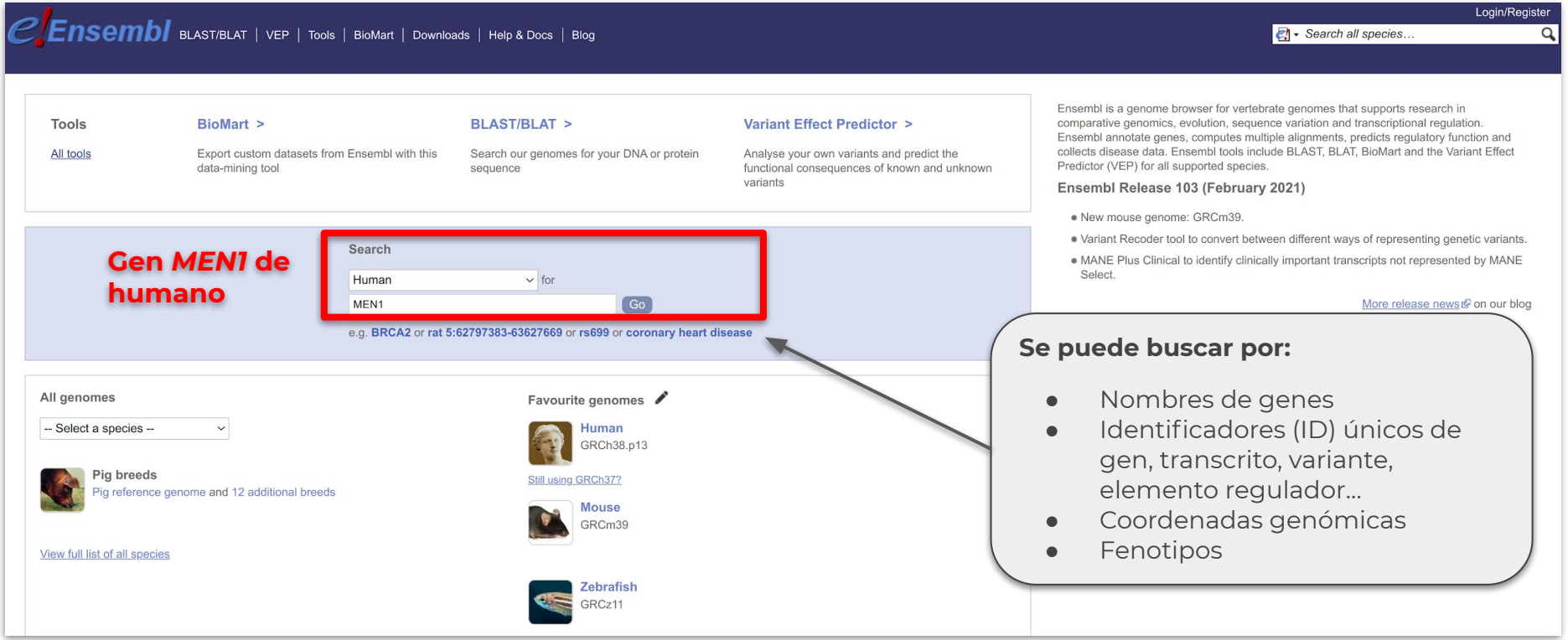
Contiene información de genes y transcritos a nivel de genoma, gen y proteína:

- Secuencias (ADN, ARN y proteína)
- Isoformas
- Variación genética
- Dominios proteicos
- Homología entre especies
- Elementos reguladores



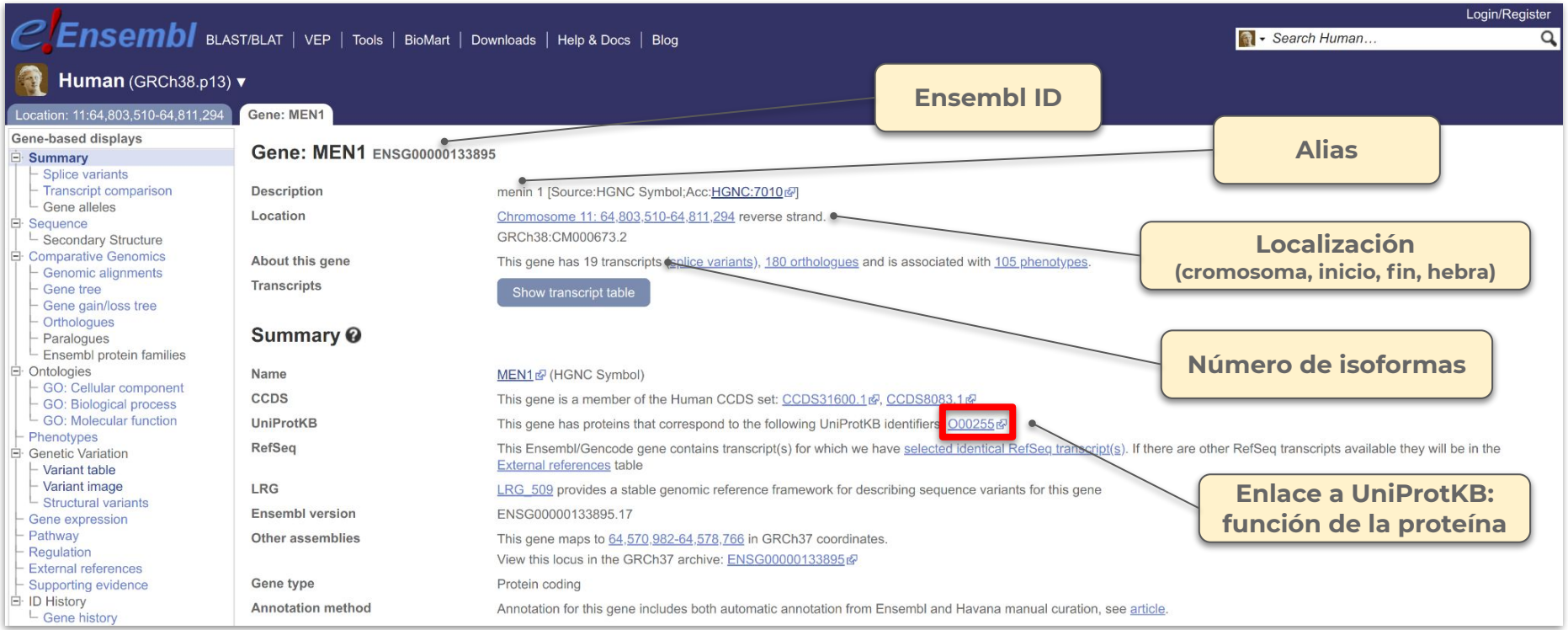
Otro navegador genómico interesante es UCSC Genome Browser.

Buscar en Ensembl



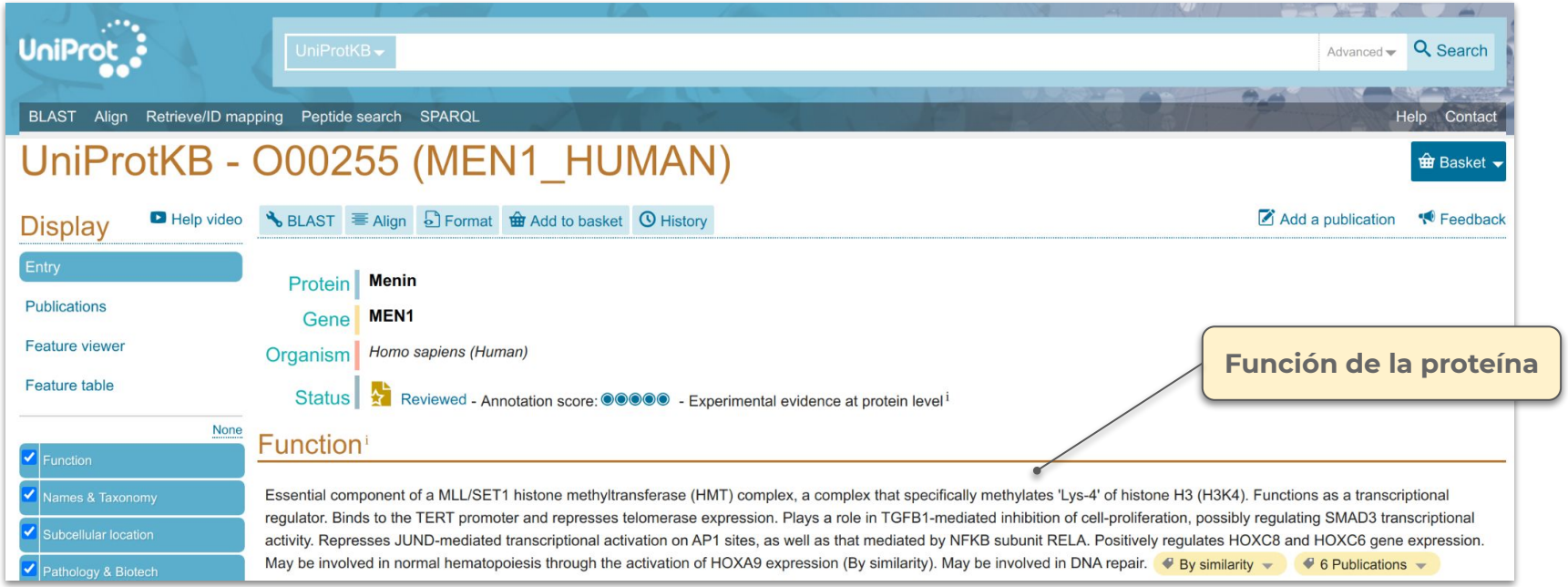
Buscar en Ensembl

Características del gen



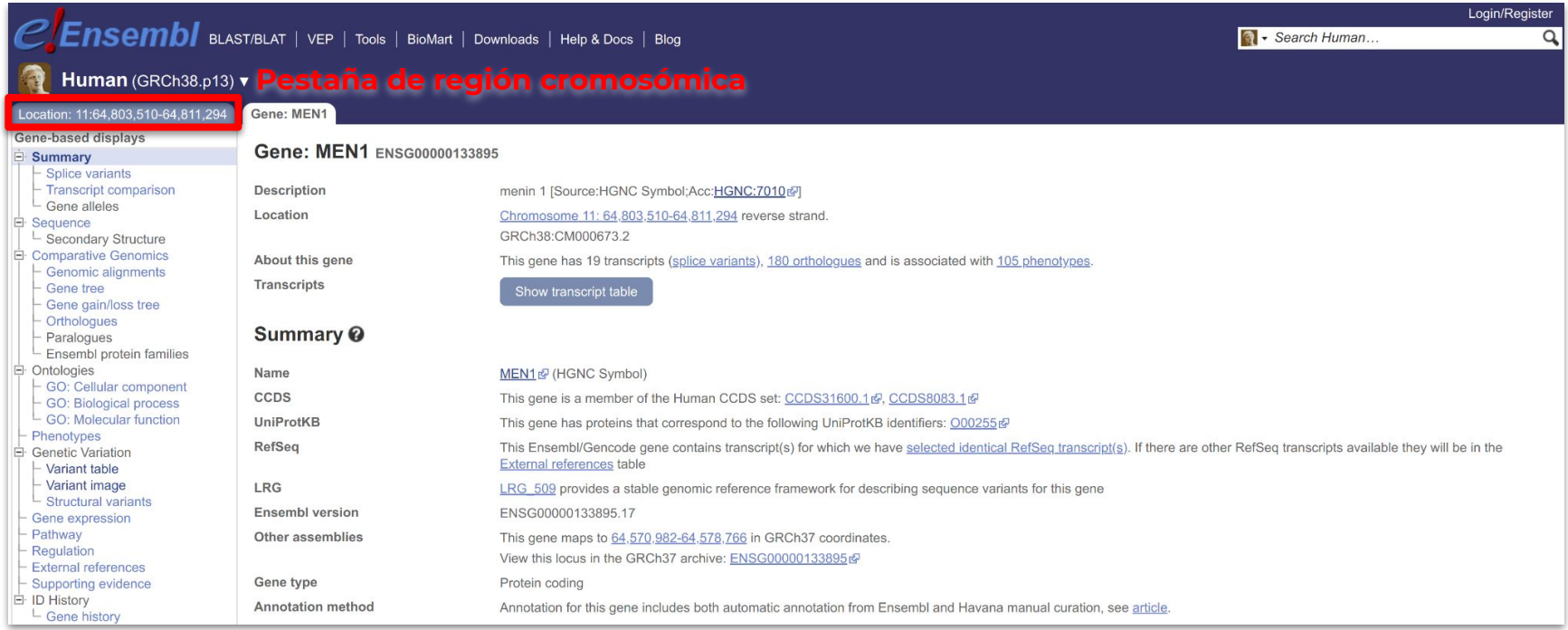
Función del gen (UniProtKB)

UniProtKB es una base de datos de secuencias de proteínas y sus anotaciones.

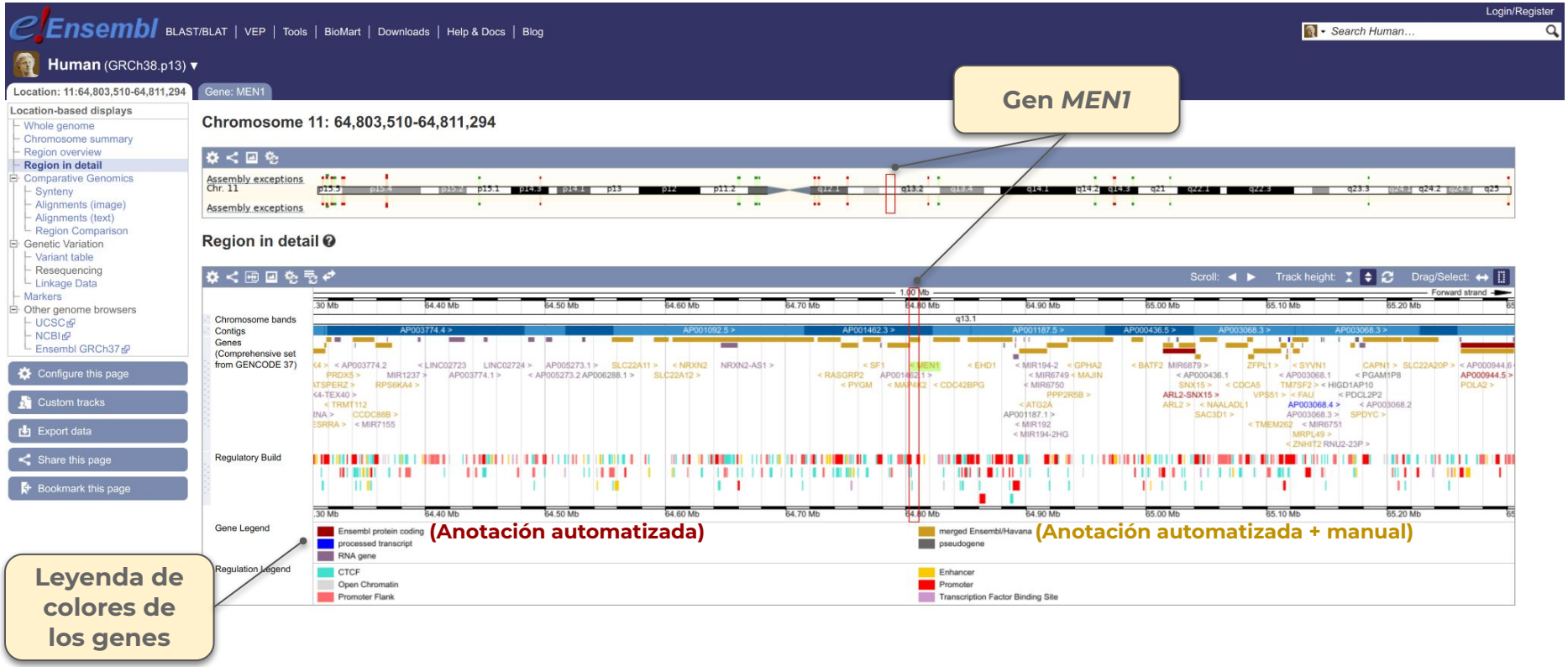


Genes vecinos

Volvemos a Ensembl:



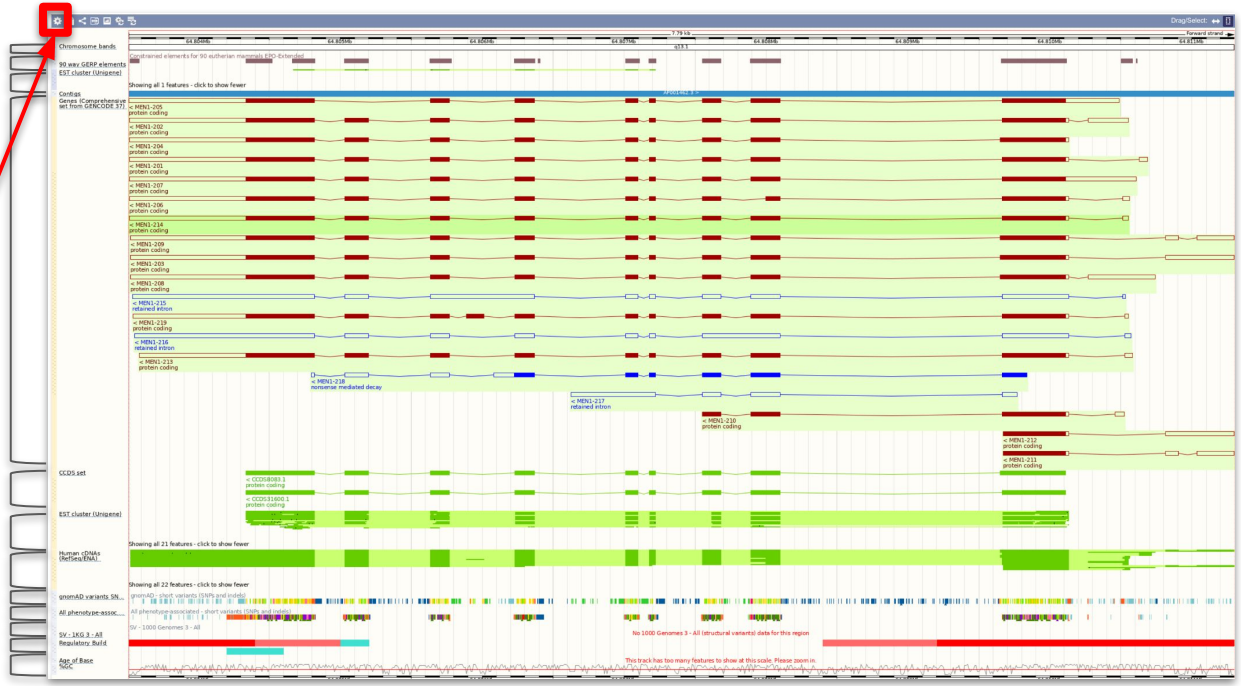
Genes vecinos



Pistas de Ensembl

Hacemos scroll en la pantalla anterior hasta llegar a este otro visualizador genómico:

- Información procedente de distintas bases de datos
- Podemos ver isoformas, variantes (SNPs e indels), regiones de regulación...
- Son **customizables**: permiten mostrar sólo la información que nos interesa



Variantes patogénicas

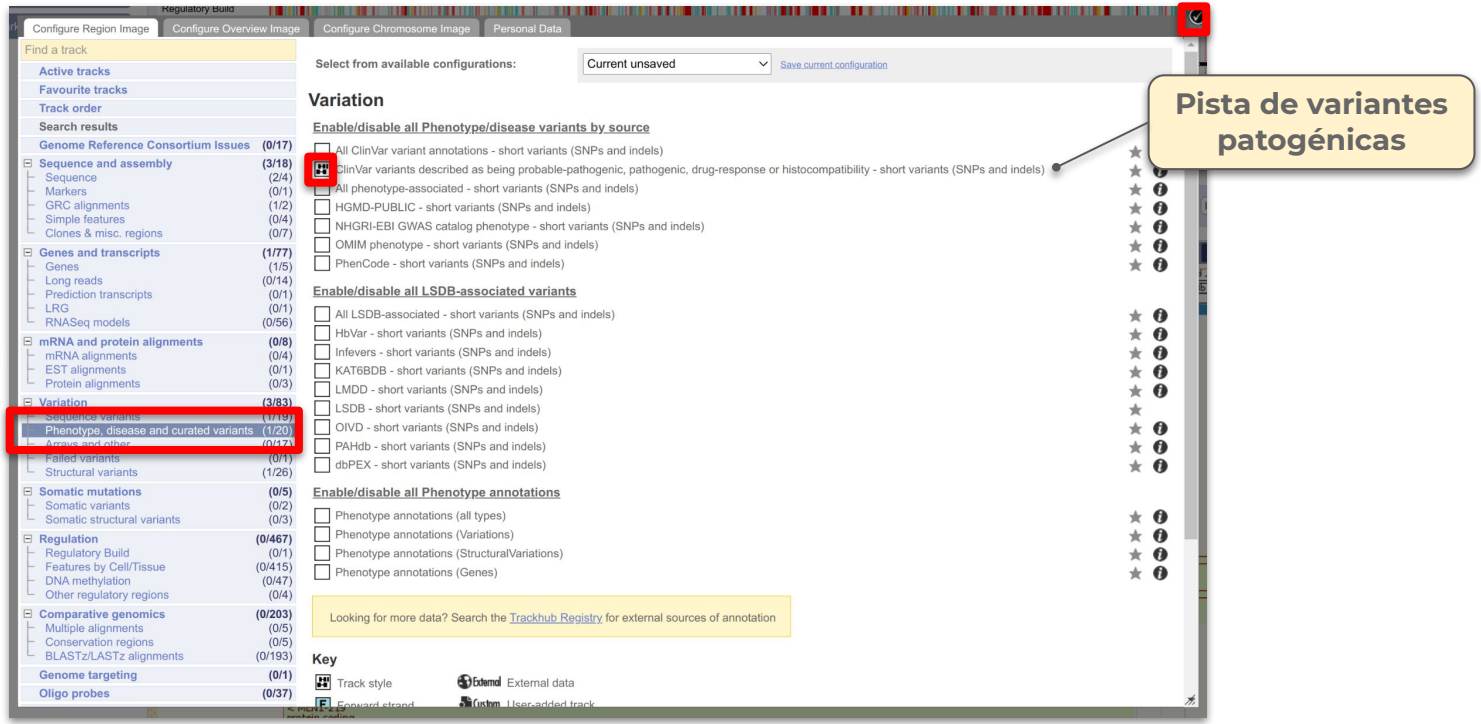
Queremos visualizar sólo las isoformas y las variantes:

Apagamos todas estas pistas (9)

Variantes patogénicas

Activamos la pista de variantes patogénicas

(en modo Normal)



Variantes patogénicas

Hay variantes en todo el gen, pero las **patogénicas** están principalmente en **exones**.

Leyenda:

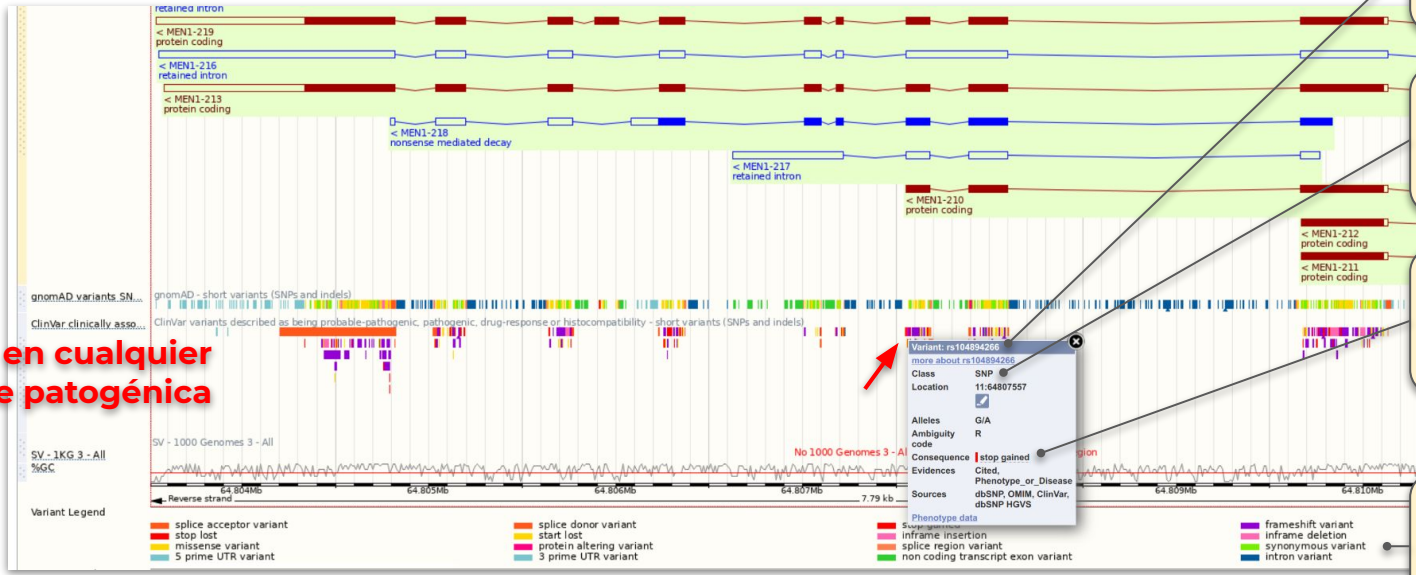
- UTRs
- Exones codificantes
- Intrones

Pista de isoformas

Variantes

Variantes patogénicas

Variantes patogénicas



Click en cualquier variante patogénica

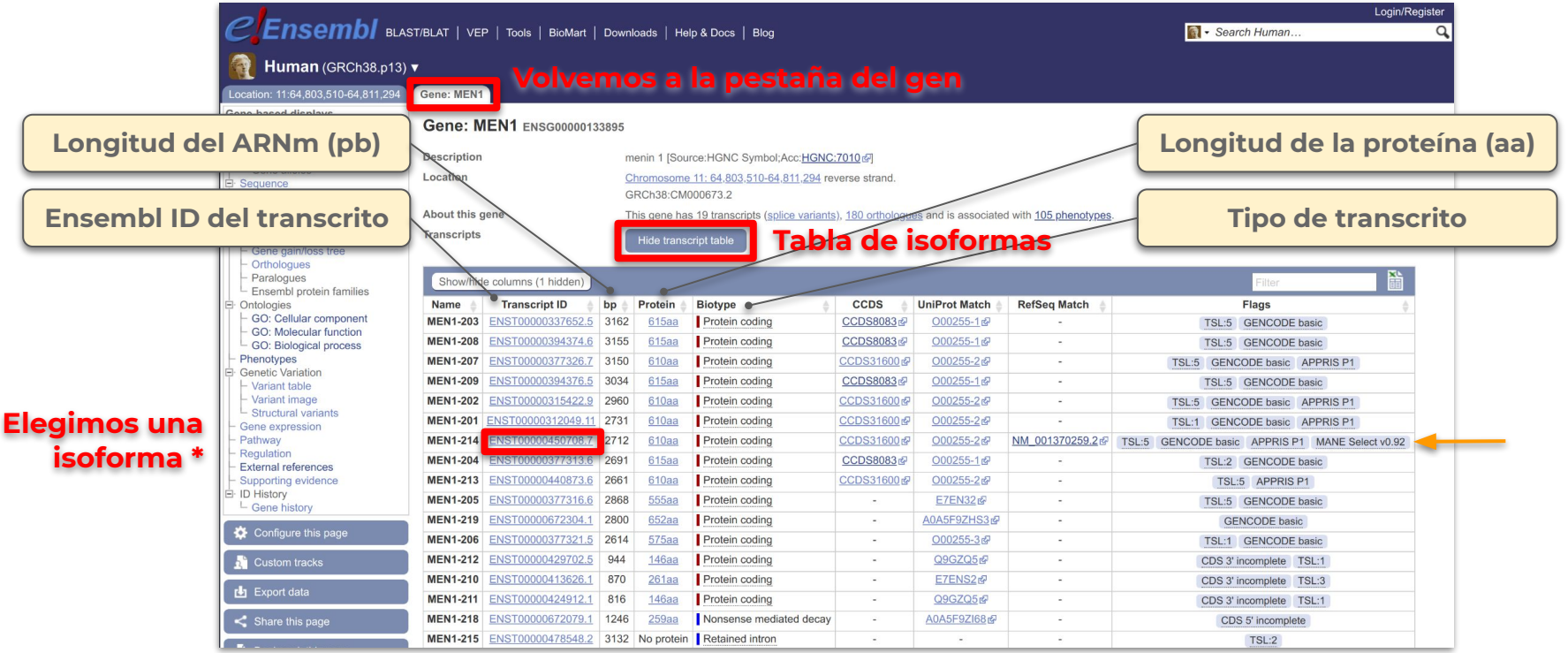
Identificador de variante

Tipo de variante (SNP, deleción, inserción...)

Consecuencia de la mutación

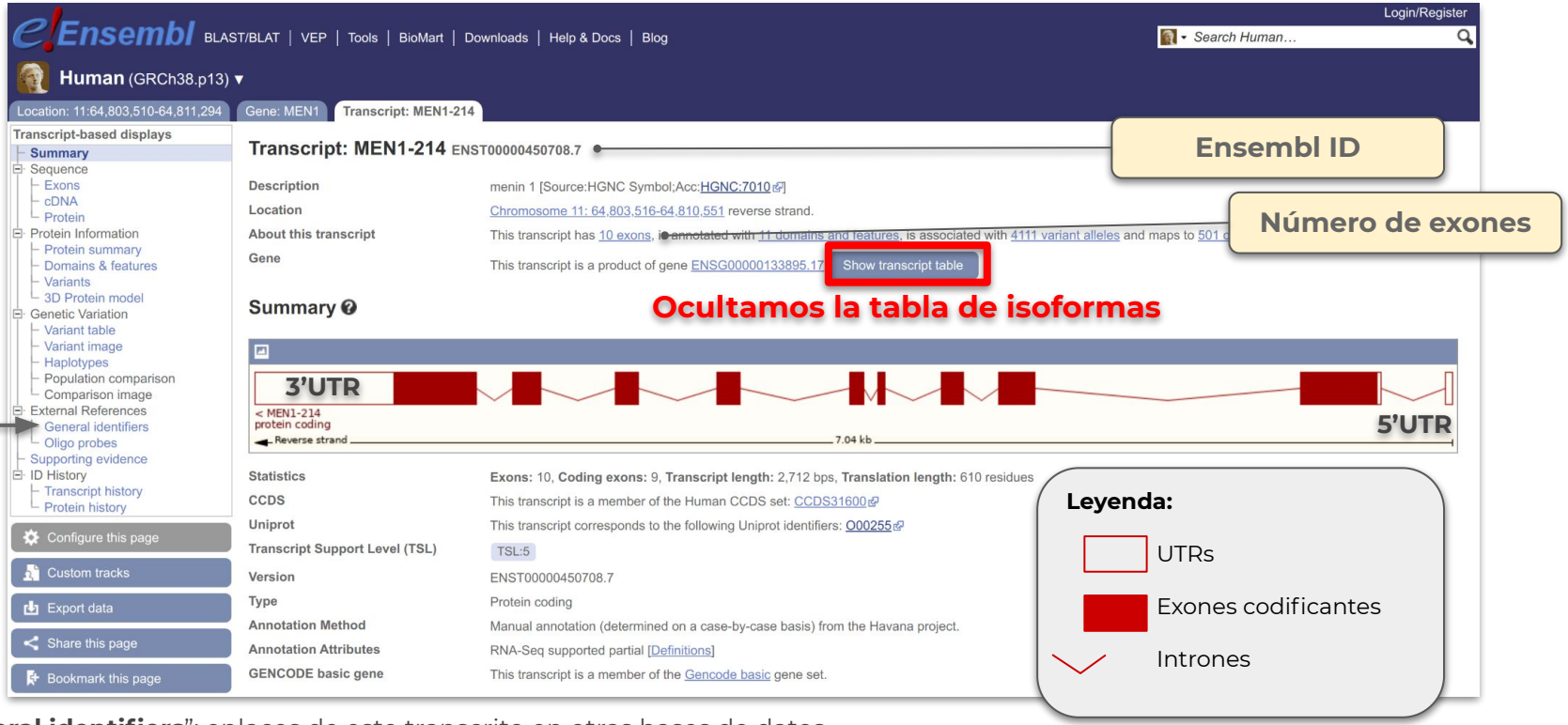
Leyenda de consecuencias de las variantes

Isoformas



***** Preferiblemente la isoforma con la etiqueta "**MANE Select v0.92**" (columna *Flags*), ya que ha sido escogida como isoforma de referencia del gen.

Información de la isoforma

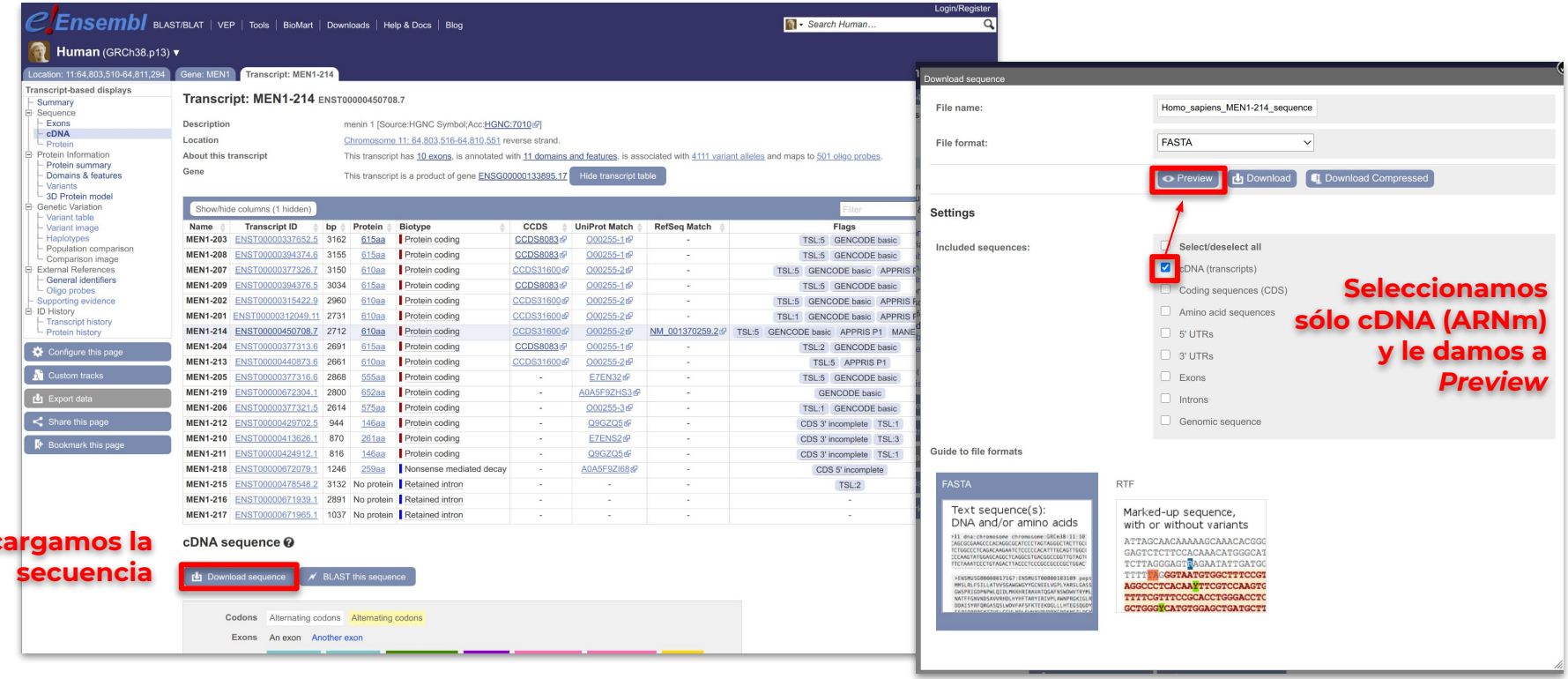


"**General identifiers**": enlaces de este transcrito en otras bases de datos.

Secuencia del ARNm

En la sección *Sequence* seleccionamos *cDNA*

Secuencia del ARNm

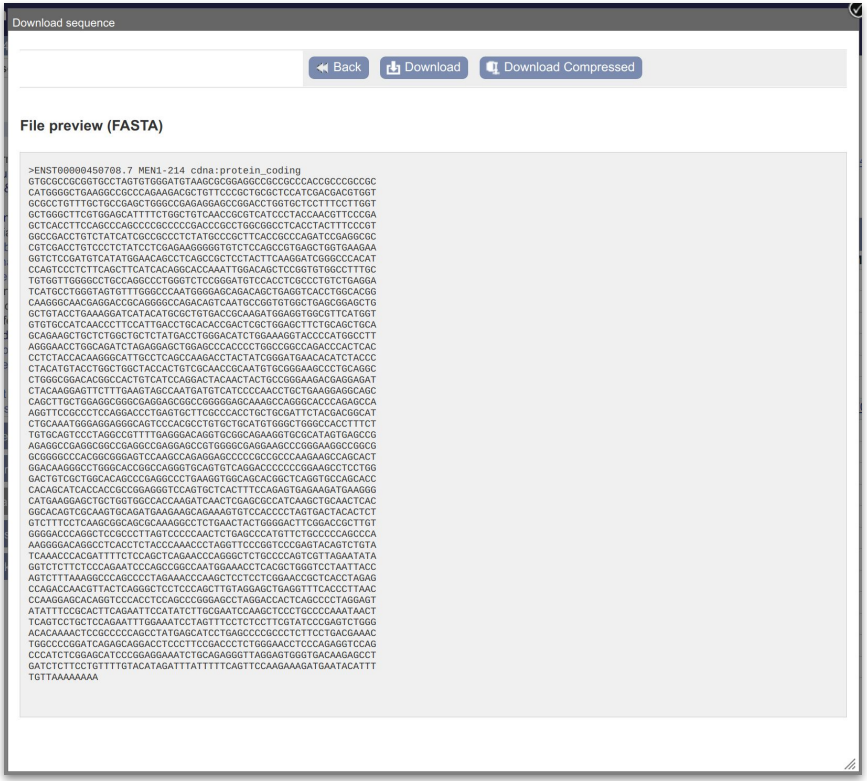


Secuencia del ARNm

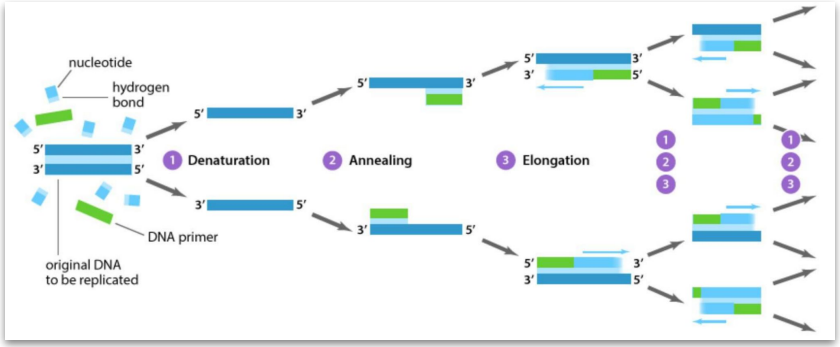
Formato FASTA

Formato para representar secuencias de nucleótidos y aminoácidos.

- Línea de **encabezado** que comienza por “>”. Contiene información sobre la secuencia (ID, nombre...)
- Líneas de **secuencia**



¿Cómo haría una PCR de este gen?

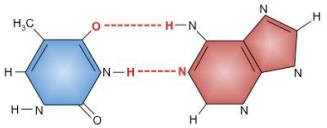


Respuesta: Primer-BLAST

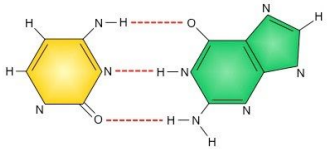
Enlace

Diseño de primers para PCR

- **Longitud:** 18-24 pb de longitud
- **Contenido en G/C:**
  - Contenido en G/C del 40-60%
  - 1-2 pares G/C al principio y al final
- **Temperatura de desnaturalización (Tm):**
  - 50-60ºC
  - Menos de 5ºC de diferencia entre primers
- **Secuencia:**
  - Sin regiones complementarias entre ellos: **dímeros de primers**
  - Evitar regiones que formen estructuras secundarias internas
  - Evitar hibridaciones inespecíficas con otras partes del genoma

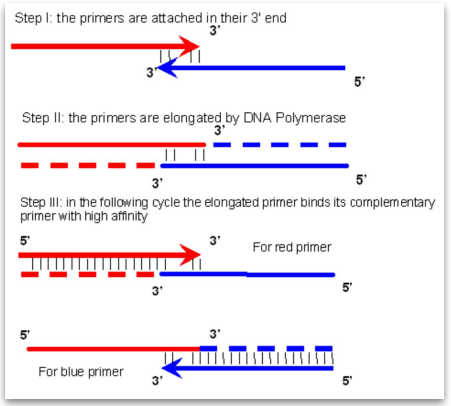


Thymine / **U**racil pairs with **A**denine
(2 hydrogen bonds)

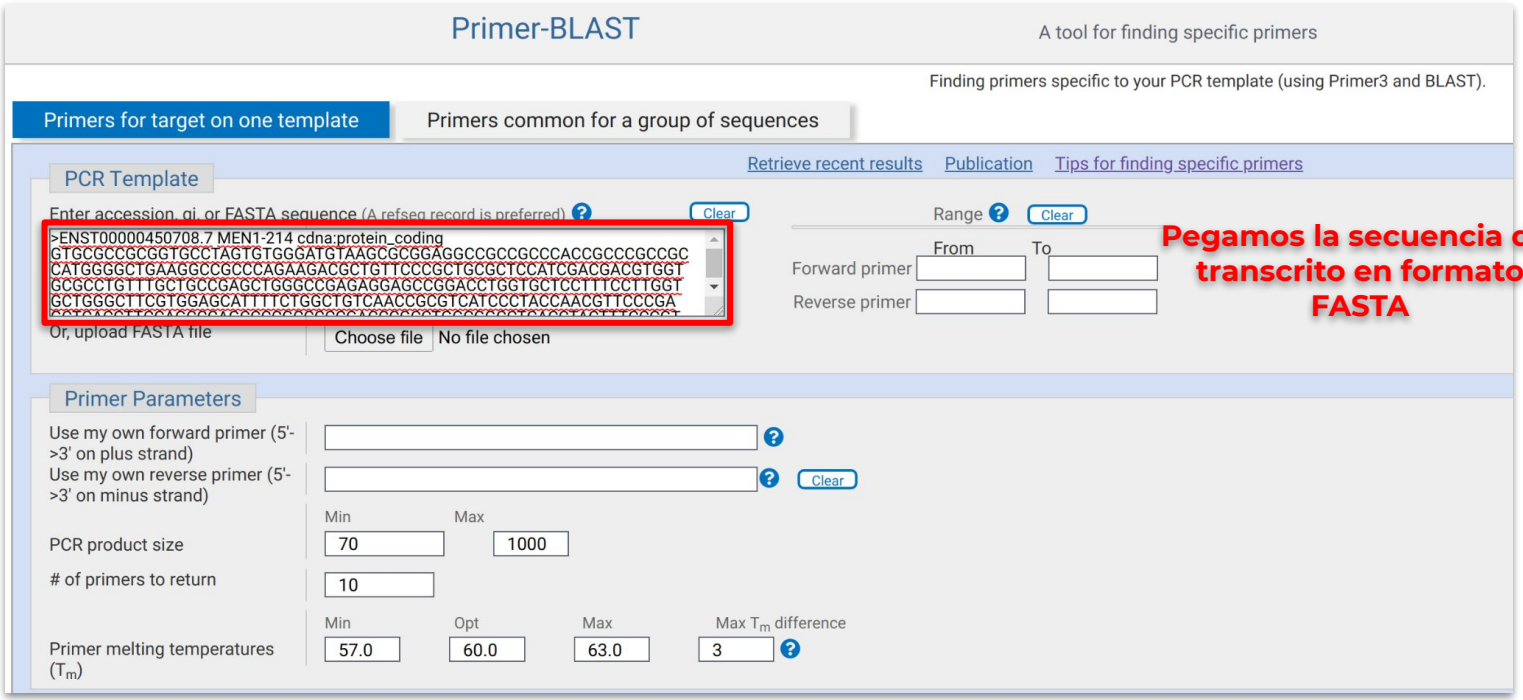


Cytosine pairs with **G**uanine
(3 hydrogen bonds)

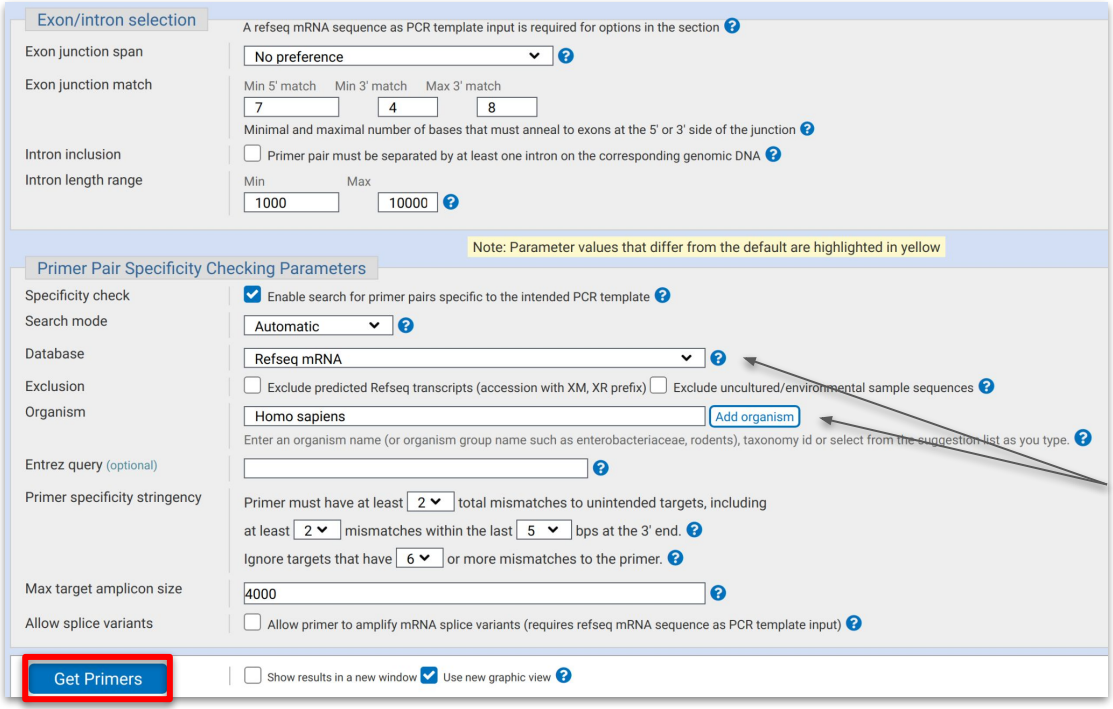
Formación de dímeros de primers



Primer-Blast



Primer-Blast



Primer-Blast tiene en cuenta las secuencias del **resto de transcritos de humano** para no generar primers con hibridaciones **inespecíficas**

Primer-Blast

Región codificante

ARNm (codificante + UTRs)

Parejas de primers diseñados

Secuencias de los primers *forward* y *reverse*

Longitud de la secuencia amplificada (pb)

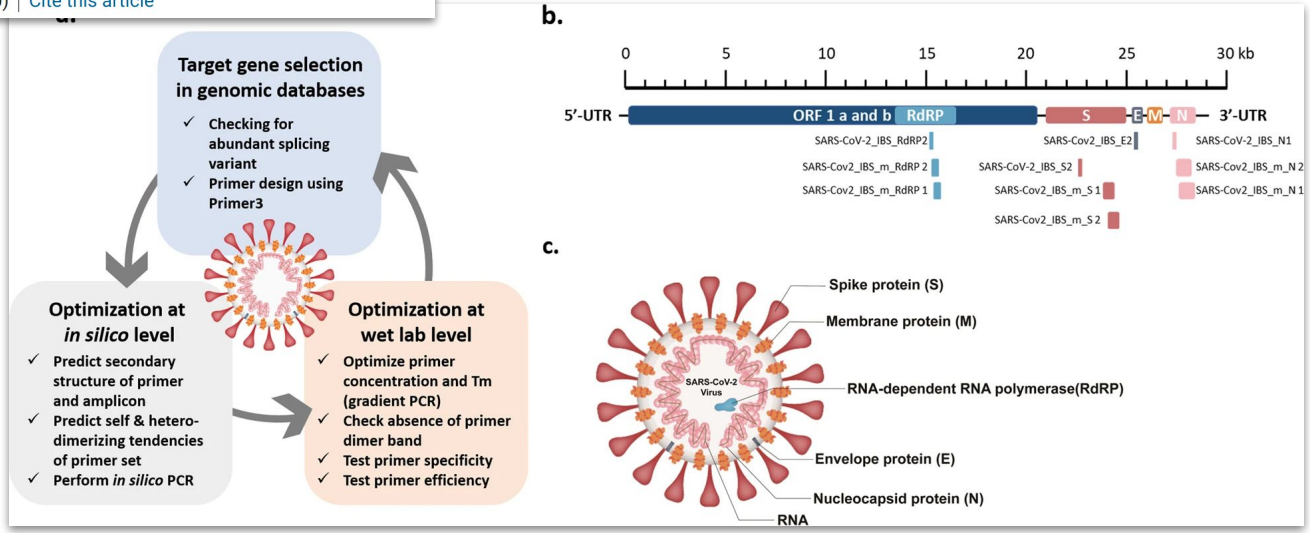
Diseño de primers para SARS-CoV-2

Article | Open Access | Published: 16 June 2020

Optimization of primer sets and detection protocols for SARS-CoV-2 of coronavirus disease 2019 (COVID-19) using PCR and real-time PCR

Myungsun Park, Joungha Won, Byung Yoon Choi & C. Justin Lee

Experimental & Molecular Medicine **52**, 963–977(2020) | Cite this article



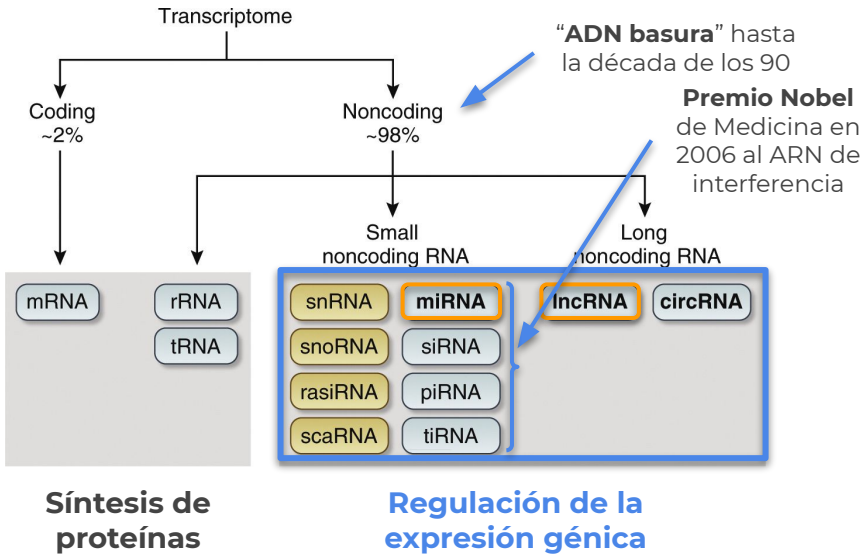
¿Existe algún microARN que regule al gen?

Respuesta: 

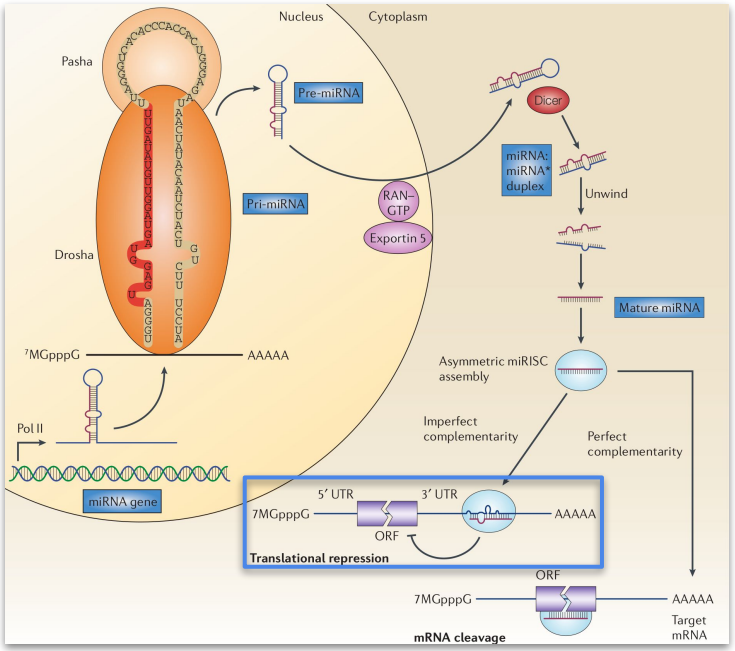
Enlace

ARNs no codificantes: miRNA y lncRNA

- No traducidos a proteína
- Proceden de “genes de ARN” (color **morado** en Ensembl)
- **En síntesis de proteínas**: ARN ribosómico y de transferencia
- **En regulación de la expresión**: miRNA, lncRNA, entre otros.



microRNAs



Biogénesis y actuación de miRNA

- ARN pequeños (18-20 nt)
- Regulan **negativamente** la expresión de genes
- Dos mecanismos de acción:
    - Inhiben la traducción del ARNm uniéndose a su **extremo 3'** (hibridación **imperfecta**, más frecuente)
    - Provocan la escisión del ARNm uniéndose a regiones **codificantes** (hibridación **perfecta**)
- Un miRNA puede controlar **varios genes** y un gen puede estar controlado por **varios miRNA**

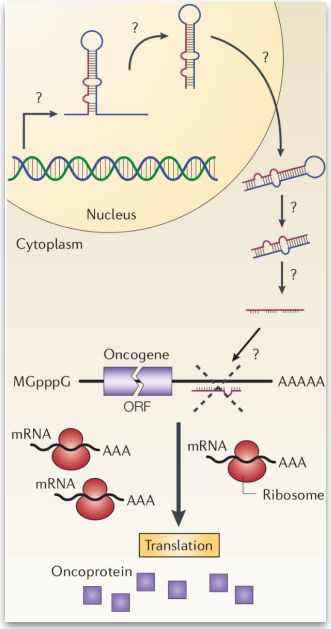
microRNAs y cáncer: Oncomirs

miRNA como **supresor tumoral**
(silenciando un oncogen)

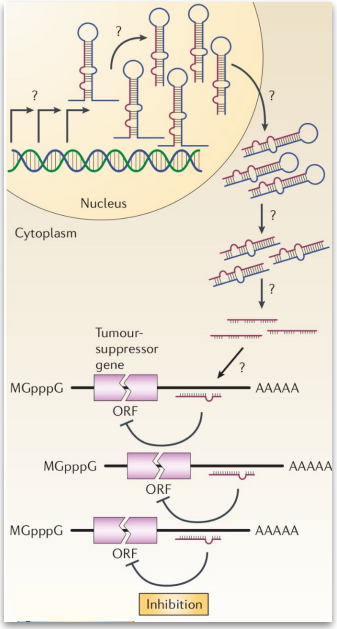
Si se infraexpresa

Si se sobreexpresa

miRNA como **oncogen**
(silenciando un supresor tumoral)

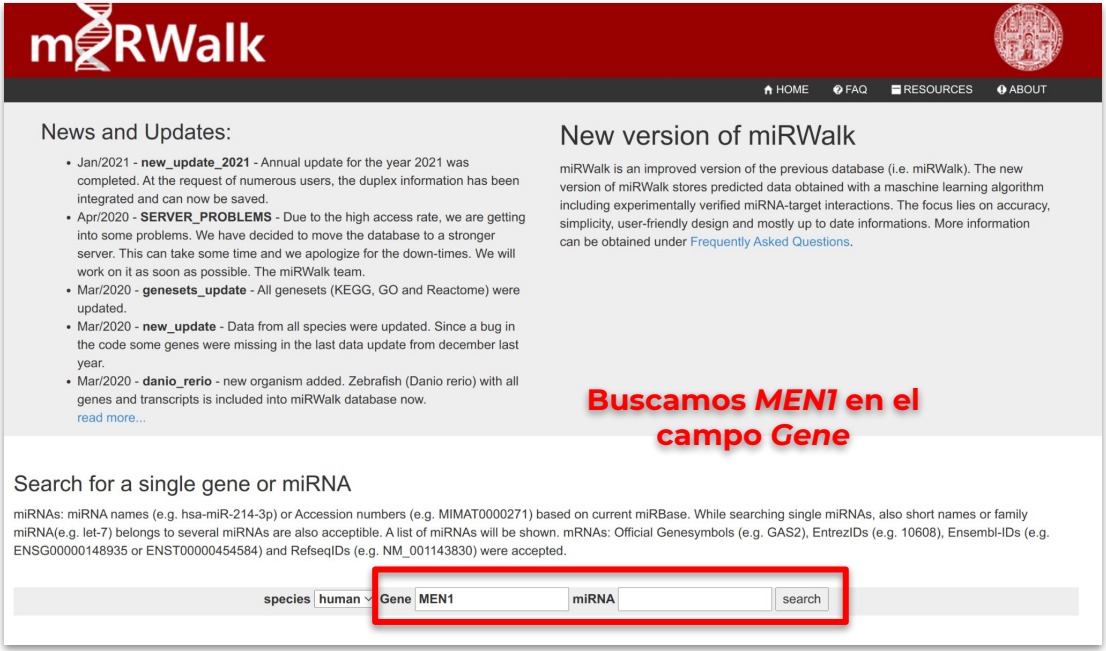


Formación de un tumor
↑ Proliferación
↑ Invasión
↑ Angiogénesis
↓ Muerte celular



miRWalk

Base de datos sobre relaciones entre miRNAs y genes.



miRNA asociados al gen

Identificador del miRNA

Compuesto de 4 partes:

- Especie: hsa es *Homo sapiens*
- Procesamiento: miR es miRNA maduro (mir sería el gen o pre-miRNA)
- Identificador numérico
- Brazo: 5' o 3' (sólo para miRNA maduros)

MEN1

Entrezid 4221
Genesymbol MEN1
Alias MEAI;SCG2
Description Homo sapiens menin 1 (MEN1), transcript variant 1, mRNA.
Chromosome 11
Ensemblid ENSG00000133895

Transcripts:

Refseq	Ensemblid	Length	Startcds	Endcds
NM_001370251	ENST00000672304	2828	50	2008
NM_001370259	ENST00000450708	2712	62	1894
NM_001370260	ENST00000440873	2702	50	1882
NM_001370261	ENST00000315422	2960	308	2140
NM_001370263	ENST00000377321	2609	62	1789
NM_130799	ENST00000312049	2770	111	1943
NM_130801	ENST00000377313	2736	62	1909
NM_130802	ENST00000394374	3172	498	2345
NM_130803	ENST00000337652	3179	505	2352
NM_130804	ENST00000394376	3051	377	2224

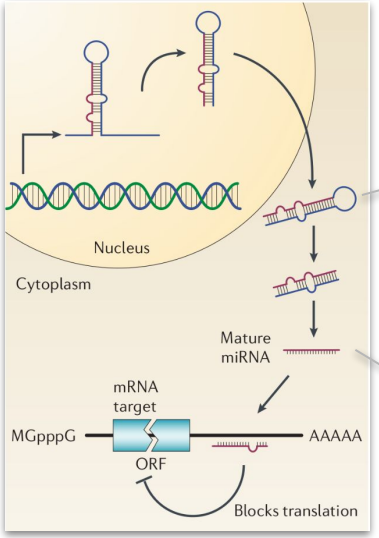
Región del gen donde se une el miRNA

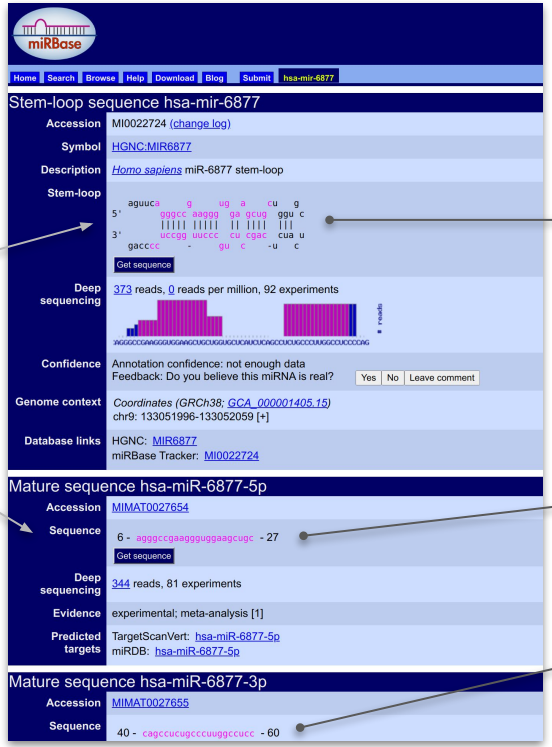
Seleccionamos un miRNA

Interactions:

Mirna	Refseqid	Genesymbol	Duplex	Score	Position	Binding Site	Au	Me	N Pairings	Targetscan	Mirdb	Mirtarbase
hsa-miR-6877-5p	NM_130799	MEN1	.((((..((((((((((..(((.&)))..)))))))))).)))).	0.80	3UTR	1976,1996	0.32	-7.238	16	—	—	—
hsa-miR-6877-5p	NM_130801	MEN1	.((((..((((((((((..(((.&)))..)))))))))).)))).	0.80	3UTR	1942,1962	0.32	-7.238	16	—	—	—
hsa-miR-6877-5p	NM_130802	MEN1	.((((..((((((((((..(((.&)))..)))))))))).)))).	0.80	3UTR	2378,2398	0.32	-7.238	16	—	—	—
hsa-miR-6877-5p	NM_001370259	MEN1	.((((..((((((((((..(((.&)))..)))))))))).)))).	0.80	3UTR	1927,1947	0.32	-7.238	16	—	—	—
hsa-miR-6877-5p	NM_001370260	MEN1	.((((..((((((((((..(((.&)))..)))))))))).)))).	0.80	3UTR	1915,1935	0.32	-7.238	16	—	—	—
hsa-miR-6877-5p	NM_001370261	MEN1	.((((..((((((((((..(((.&)))..)))))))))).)))).	0.80	3UTR	2173,2193	0.32	-7.238	16	—	—	—
hsa-miR-6877-5p	NM_001370263	MEN1	.((((..((((((((((..(((.&)))..)))))))))).)))).	0.80	3UTR	1822,1842	0.32	-7.238	16	—	—	—
hsa-miR-125b-2-3p	NM_130799	MEN1	.....(((((((((((...(((&)))...))))))))))).	0.81	3UTR	2676,2693	0.59	-15.567	14	—	—	—
hsa-miR-520e-3p	NM_130799	MEN1	...(((((((((.(((((((&)))))))...))))))))).	0.81	3UTR	2321,2341	0.4	-9.685	16	—	—	—
hsa-miR-4280	NM_130799	MEN1	((((.(((((((((((.(((.&)))..)))))))))).....)))).	0.81	3UTR	2105,2129	0.48	-9.6	17	—	—	—

Información del miRNA en miRBase





Si no se abre esta página de miRBase, copia el identificador del miRNA (**hsa-miR-6877-5p** en este caso), haz click **aquí** y pégalo en el buscador de la derecha ("Search by miRNA name or keyword")

Estructura que forma el pre-miRNA

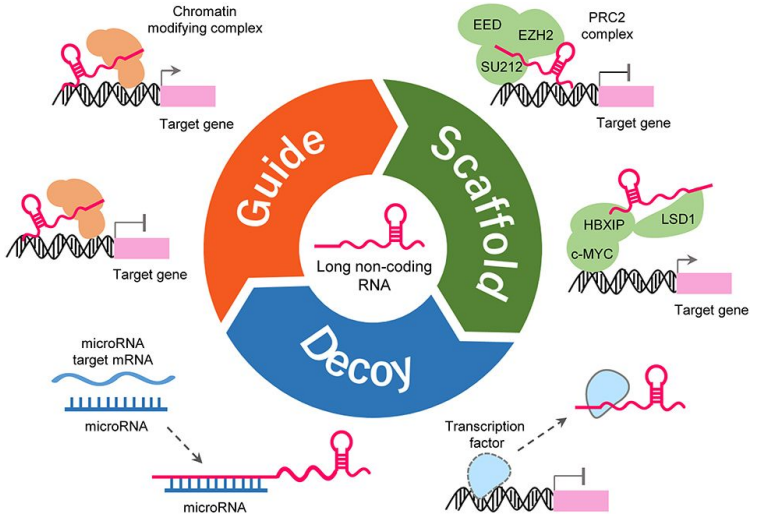
Secuencia del brazo 5'

(miRNA maduros)

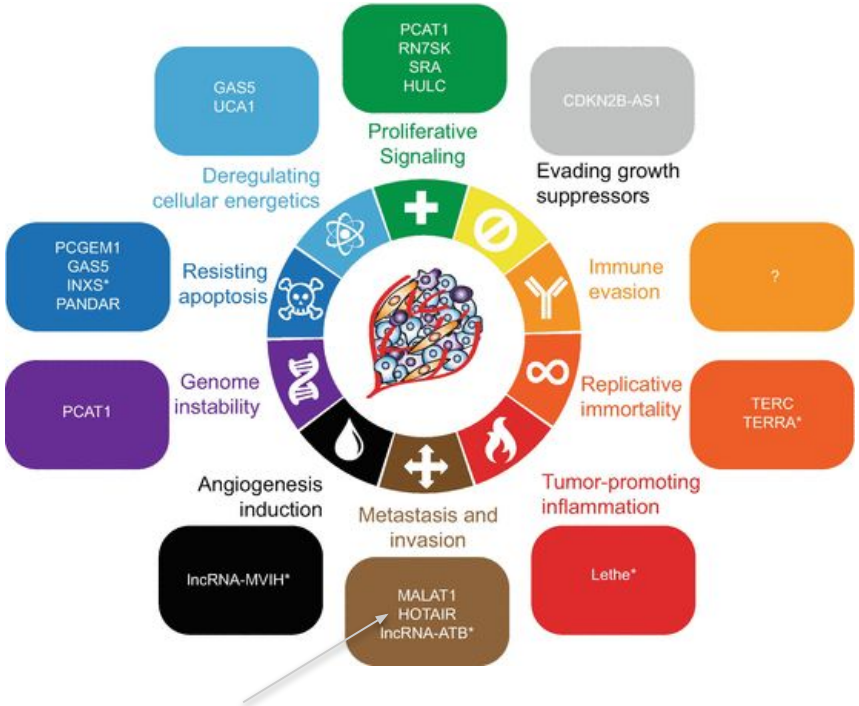
Secuencia del brazo 3'

lncRNA

- *Long non-coding RNA*
- **> 200** nucleótidos
- **Regulación genética** a varios niveles: epigenético, transcripción, post-transcripcional...
- Diversos mecanismos de acción:
  - **Guía**: unión a ADN
  - ***Scaffold*** (andamio): unión a complejos proteicos
  - **Esponja**: unión y secuestro de miRNAs, factores de transcripción...



lncRNA y cáncer



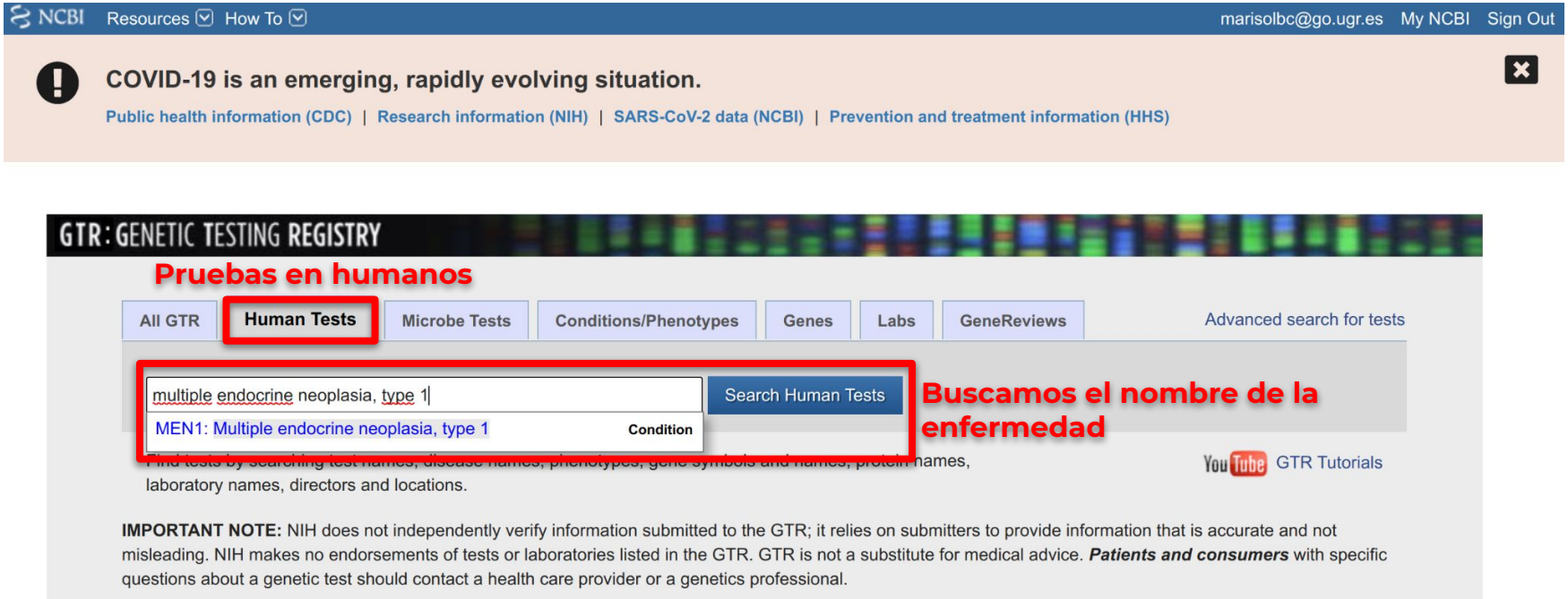
Entrada del lncRNA *HOTAIR* en la base de datos RNAcentral (enlace)

¿Qué laboratorios realizan el diagnóstico genético en España?

Respuesta: NCBI GTR Genetic Testing Registry

Enlace

Buscar en NCBI GTR



Tests genéticos de MEN1 en España

Seleccionamos *Human tests*

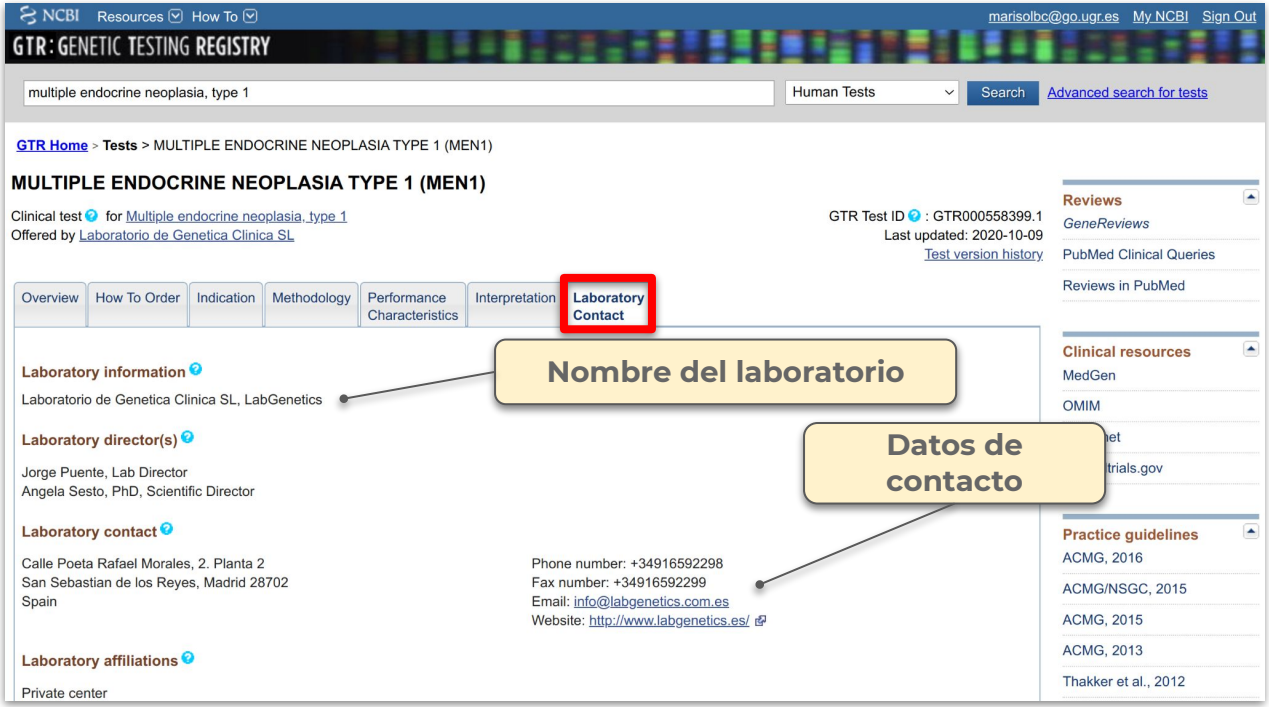
Aplicamos 3 filtros

Seleccionamos una prueba

Enfermedades que diagnostica la prueba

Tipo de prueba

Tests genéticos de MEN1 en España



Entrega en PRADO: 30/05/2021 - 23:55h

Rellenad individualmente la tabla de la derecha con otra enfermedad monogénica a vuestra elección.

Consejos:

- En Ensembl intentad elegir el transcrito con la etiqueta “MANE select v0.92” (transcrito de referencia). Podéis escoger cualquier variante patogénica.
- Elegid cualquier pareja de primers diseñados por Primer-Blast.
- Elegid cualquier miRNA en miRWalk.

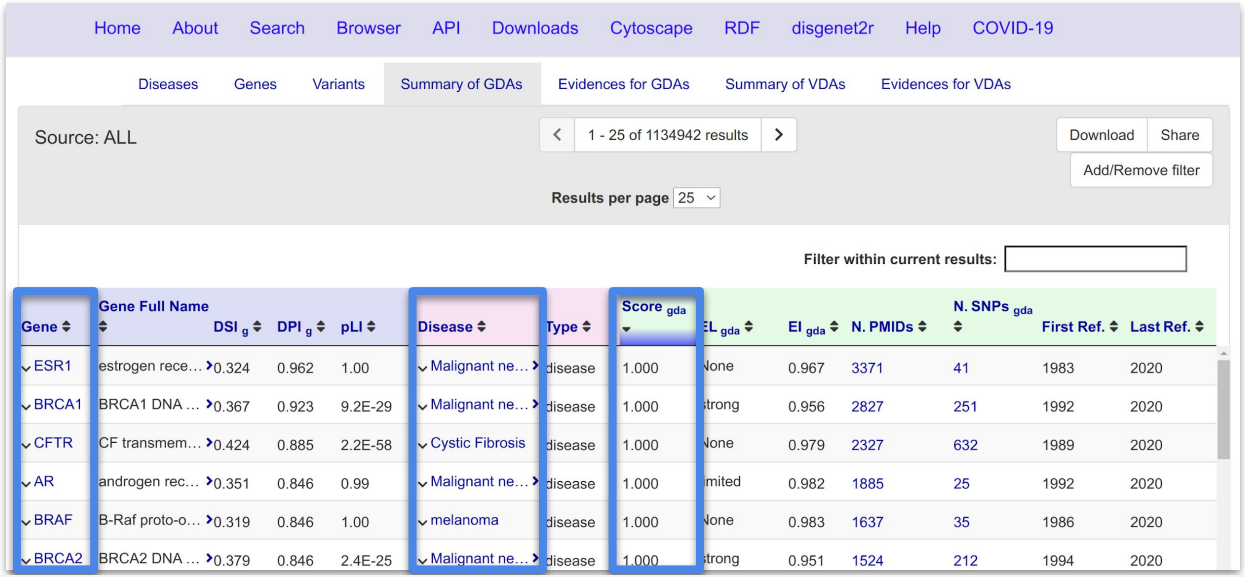
Cualquier duda, escribidme a marisolbc@ugr.es

PRÁCTICA II: BIOINFORMÁTICA	
Nombre y apellidos: María Soledad Benítez Cantos	
Enfermedad:	
Neoplasia endocrina múltiple de tipo 1	
Nombre oficial del gen responsable	*MEN1*
Alias	menin
Evidencia de la asociación gen-enfermedad	Frase: “We conclude that many "hyperplastic" parathyroid tumors in familial MEN-1 are in fact monoclonal and may progress or even begin to develop by inactivation of the MEN-1 gene (at 11q13) in a precursor cell.” Enlace a publicación: https://pubmed.ncbi.nlm.nih.gov/2568586/
Ensembl ID del gen	ENSG00000133895
Localización cromosómica (cromosoma, coordenadas y hebra)	Chromosome 11: 64,803,510-64,811,294 reverse strand
Genes adyacentes	MAP4K2, SF1, RASGRP2
dbSNP ID de una variante patogénica	rs104894266
Consecuencia de la variante	Codón de parada prematuro
Número de transcritos	19
Ensembl ID del transcrito seleccionado	ENST00000450708.7
Longitud del transcrito (pb)	3162
Número de exones	10
Longitud de la proteína (aa)	615
Secuencia del transcrito en formato FASTA	>ENST00000450708.7 MEN1-214 cdna:protein_coding GTGCGCCGCGGTGCCTAGTGTGGGATGTAAGCGCG GAGGCCGCCGCCCACCGCCCGCCGC (...)
Primers diseñados	Forward: CGTGAGCTGGTGAAGAAGGT Reverse: GCTGTCCAATTTGGTGCCTG
Identificador del miRNA que lo regula	hsa-miR-6877-5p
Secuencia del miRNA maduro	agggccgaaggguggaagcugc
Laboratorio que realiza el diagnóstico genético de la enfermedad (preferiblemente en España)	Laboratorio de Genética Clínica SL, LabGenetics (Madrid)

¿Cómo escojo un gen?

Si no conoces ninguna enfermedad monogénica, DisGeNet tiene una **tabla** con todas las asociaciones gen-enfermedad publicadas.

Seleccionar un gen cuya columna “Score” sea igual a 1 (la asociación es 100% fiable).



Diagramas biológicos con BioRender

BioRender es una plataforma online que cuenta con más de 20000 iconos relacionados con la Biología y que permite crear ilustraciones científicas de forma sencilla.

Útil para trabajos del grado, TFG, TFM...